

JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION 04.2017 N°45



RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

TRANSFORMATION

CAPCO

EMPOWERING THE FINANCIAL WORLD

Pushing the pace of financial technology, we help our clients solve technology challenges for their business – whether it's capital markets in Mumbai or community banking in Macon.

We leverage knowledge and insights from our clients around the world:

20,000 clients in towns everywhere are becoming more efficient, modern and scalable.

27 billion transactions processed help solve clients' challenges – big and small.

\$9 trillion moved across the globe in a single year empowers our clients' communities to build storefronts, homes and careers.

55,000 hearts and minds have joined forces to bring you greater capabilities in even the smallest places.

Empowering the Financial World

FISGLOBAL.COM



JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

Recipient of the Apex Award for Publication Excellence

Editor

Shahin Shojai, Global Head, Capco Institute

Advisory Board

Christine Ciriani, Partner, Capco

Chris Geldard, Partner, Capco

Nick Jackson, Partner, Capco

Editorial Board

Franklin Allen, Nippon Life Professor of Finance, University of Pennsylvania

Joe Anastasio, Partner, Capco

Philippe d'Arvisenet, Adviser and former Group Chief Economist, BNP Paribas

Rudi Bogni, former Chief Executive Officer, UBS Private Banking

Bruno Bonati, Chairman of the Non-Executive Board, Zuger Kantonalbank

Dan Breznitz, Munk Chair of Innovation Studies, University of Toronto

Urs Birchler, Professor Emeritus of Banking, University of Zurich

Géry Daeninck, former CEO, Robeco

Stephen C. Daffron, CEO, Interactive Data

Jean Dermine, Professor of Banking and Finance, INSEAD

Douglas W. Diamond, Merton H. Miller Distinguished Service Professor of Finance, University of Chicago

Elroy Dimson, Emeritus Professor of Finance, London Business School

Nicholas Economides, Professor of Economics, New York University

Michael Enthoven, Board, NLF, Former Chief Executive Officer, NIBC Bank N.V.

José Luis Escrivá, Director, Independent Revenue Authority, Spain

George Feiger, Pro-Vice-Chancellor and Executive Dean, Aston Business School

Gregorio de Felice, Head of Research and Chief Economist, Intesa Sanpaolo

Allen Ferrell, Greenfield Professor of Securities Law, Harvard Law School

Peter Gomber, Full Professor, Chair of e-Finance, Goethe University Frankfurt

Wilfried Hauck, Chief Financial Officer, Hanse Merkur International GmbH

Pierre Hillion, de Picciotto Professor of Alternative Investments and Shell Professor of Finance, INSEAD

Andrei A. Kirilenko, Visiting Professor of Finance, Imperial College Business School

Mitchel Lenson, Non-Executive Director, Nationwide Building Society

David T. Llewellyn, Professor of Money and Banking, Loughborough University

Donald A. Marchand, Professor of Strategy and Information Management, IMD

Colin Mayer, Peter Moores Professor of Management Studies, Oxford University

Pierpaolo Montana, Chief Risk Officer, Mediobanca

Steve Perry, Chief Digital Officer, Visa Europe

Derek Sach, Head of Global Restructuring, The Royal Bank of Scotland

Roy C. Smith, Kenneth G. Langone Professor of Entrepreneurship and Finance, New York University

John Taysom, Visiting Professor of Computer Science, UCL

D. Sykes Wilford, W. Frank Hipp Distinguished Chair in Business, The Citadel



Transformation

FinTech/RegTech

- 8 **Opinion: Open APIs and Open Banking: Assessing the Impact on the European Payments Industry and Seizing the Opportunities**
Thomas Egner
- 14 **Algorithmic Regulation: Automating Financial Compliance Monitoring and Regulation Using AI and Blockchain**
Philip Treleven, Bogdan Batrinca
- 22 **RegTech is the New Black – The Growth of RegTech Demand and Investment**
Kari S. Larsen, Shariq Gilani
- 30 **From “Blockchain Hype” to a Real Business Case for Financial Markets**
Massimo Morini
- 41 **Trade Finance Disrupted: A Blockchain Use Case**
André Brunner, Nourdine Abderrahmane, Arjun Muralidharan, Patrick Halfpap, Oliver Süme, Stephan Zimprich
- 49 **Towards a Standards-Based Technology Architecture for RegTech**
Tom Butler
- 60 **Machine Learning: A Revolution in Risk Management and Compliance?**
Bart van Liebergen
- 68 **Data-centered Dependencies and Opportunities for Robotics Process Automation in Banking**
Sandeep Vishnu, Vipul Agochiya, Ranjit Palkar

Investments

- 78 **John Bull Can’t Stand Two Percent: QE’s Depressing Implications for Investment**
Jason M. Thomas
- 90 **Do Credit Rating Agencies Inflate Their Ratings? A Review**
Kee-Hong Bae, Hamdi Driss, Gordon S. Roberts

- 101 **The Power of “Negative Beta”: Why Every Portfolio Should Include Private Equity**
Andrew Freeman, Iordanis Karagiannidis, D. Sykes Wilford
- 111 **Downside Risk Protection of Retirement Assets: A New Approach**
Atanu Saha, Alex Rinaudo
- 121 **The Asset Management Industry, Systemic Risk, and Macroprudential Policy**
Claude Lopez
- 129 **The Role of Asset Owners in the Market for Investment Research: Where Are the Fiduciary Capitalists?**
Alistair Haig, Neil Scarth
- 136 **Risk, Data, and the Barcodes of Finance**
Allan D. Grody

Banking

- 159 **Opinion: Risk Culture: Risk Prevention Starts With the Individual**
Ulrich Hunziker
- 164 **The Troubled Future of Global Banking**
Brad Hintz, Roy C. Smith
- 177 **Policy Response Asymmetry and the Increasing Risks From Rising Government Debt Level**
Blu Putnam, Erik Norland
- 187 **Public Disclosure and Risk-adjusted Performance at Bank Holding Companies**
Beverly Hirtle
- 207 **What do New Forms of Finance Mean for Emerging Markets?**
M. S. Mohanty

Dear Reader,

Welcome.

Heralded as the new FinTech, RegTech – regulatory technology – has recently begun to rise in prominence. The so-called ‘regulatory tsunami’ continues to set rigorous compliance challenges, with associated operational demands and expense. In this context, RegTech has started to develop a profile matching that of its big brother, FinTech.

RegTech’s core promises are to achieve high quality compliance, to lower costs to attain this, and to derive data-driven insights informing new products and services. Yet the key difference between traditional solutions and RegTech is agility – it’s a powerful proposition. As costs rise and uncertainties around some regulatory rules grow, RegTech will increase in appeal. In this edition of the Journal we’re featuring papers that

explore key aspects of RegTech’s evolution into a fully viable and highly productive discipline and service.

Meanwhile, FinTech is far from being eclipsed. We continue our examination of ways in which the early promise of blockchain may be translated into direct improvements to operational practice. We also look at instances where other technologies, including machine learning, are playing a central role in meeting the dual challenges of cost control and improved operations.

Of course, no aspect of successful financial technology develops in a vacuum. The economic and political context of our industry continues to have an immense impact on current practice and future direction. So, in the Banking and Investments sections of this edition, we consider a variety of influential factors from the behavior of

some credit ratings agencies, to the outlook for global banking.

The sheer diversity of issues for consideration emphasizes the need for effective and efficient responses to change. The technologies that best help financial institutions respond to these challenges will rise exponentially in demand.

I hope that the range of insights and expertise on offer in this edition of the Journal interest you and I wish you my very best for your own ongoing transformation journey.



Lance Levy
CEO, Capco

Transforming the Financial Services Industry

The financial services industry has undergone numerous transformational changes over the past few decades, from the introduction of derivative contracts to electronic trading and securitized assets. Each transformation has been more profound than its predecessor. However, the industry has never faced the kinds of transformational changes that it is currently undergoing.

New, increasingly powerful technologies are truly changing the industrial landscape, and both established players and new, nimble entrants are trying to find their feet in this new environment.

Many view the new technology-led entrants, often referred to as FinTech companies, as nothing more than hype. Others are confident that technologies such as blockchain can have truly transformational implications on the industry, both for established

players and the FinTechs. These effects could include disintermediating established players in clearing trades, using artificial intelligence to deal with clients, and complying with the never-ending stream of new regulations via RegTech technologies and companies. All these possibilities are forcing financial services firms to re-assess their businesses.

Meanwhile, the established players have to manage their current businesses, ensuring growth while coming to terms with their new competitors and determining whether they are foes or potential partners.

It is for this reason that this edition of the Journal focuses not only on the implications of new technologies on financial services firms by FinTech and RegTech companies, but also on how the financial institutions must re-assess their operating and business models to compete with their old

foes, and improve the way they manage client assets to ensure investors can retire comfortably.

The authors featured in this issue have ensured that while their contributions are of the highest scholarly standards, they are also of practical use for senior financial executives and the enterprises they manage.

We hope that you enjoy reading this edition of the Journal and that you continue to support us by sharing your best ideas with us.

On behalf of the Board of Editors,



Shahin Shojai



FinTech/RegTech

Opinion: Open APIs and Open Banking: Assessing the Impact on the European Payments Industry and Seizing the Opportunities

Algorithmic Regulation: Automating Financial Compliance Monitoring and Regulation Using AI and Blockchain

RegTech is the New Black – The Growth of RegTech Demand and Investment

From “Blockchain Hype” to a Real Business Case for Financial Markets

Trade Finance Disrupted: A Blockchain Use Case

Towards a Standards-Based Technology Architecture for RegTech

Machine Learning: A Revolution in Risk Management and Compliance?

Data-centered Dependencies and Opportunities for Robotics Process Automation in Banking

Open APIs and Open Banking: Assessing the Impact on the European Payments Industry and Seizing the Opportunities

Thomas Egner – Secretary General, Euro Banking Association (EBA)

European banks are at a critical juncture, with changes in the payment industry requiring major strategic decisions. They need to decide whether to become a banking service utility, supporting other providers in their customer-facing solutions, or play a central role in the daily lives of consumers. In this context, the regulatory requirements to open up payment accounts via application programming interfaces (APIs) play a major role, since they are expected to help drive the industry towards a new ecosystem shaped by the wider concepts of “Open Banking.”

The aim of this article is to try to explain the main aspects of this expected industry evolution and describe the impact and key strategic challenges and opportunities this will entail for banks in their role as “account-servicing

payment service providers” (AS-PSPs), in particular.

THE REGULATORY AND TECHNOLOGY DRIVE TOWARDS “OPENING UP” PAYMENT ACCOUNTS AND SERVICES

Since 2007, the European regulators have published two Payment Services Directives (PSDs) in order to create a pan-European legal framework for payments. PSD1 came into force in 2007 and paved the way for the introduction of SEPA in 2014, which aimed to harmonize payments processing. PSD2, which entered into force in January 2016, further looked into providing a regulatory framework and the necessary security requirements for an opening up

of payment accounts for third parties at the request of the customers holding the accounts. The intended effect of both sets of regulations, in conjunction with SEPA Regulation 260/2012 and previous legislations, has been to increase competition as well as to promote innovation and strengthen customer rights with regards to the use of payment and account-related services.

The introduction of this regulatory framework for payment account access will encourage new players to enter the payments market and existing players to revise and expand their service propositions. In their capacity as AS-PSPs, banks will need to offer and publish an interface for third-party providers (TPPs) to use to access the payment accounts of customers held with a respective bank. The Regulatory Technical

Standards (RTS), specifying the requirements for strong customer authentication and common and secure communication under the PSD2, which have a strong bearing on this interface topic, are expected to come into effect in late 2018 or early 2019. The common understanding in the industry seems to be that from this date on, the direct connection between the customer's bank and TPPs should be enabled via APIs.

Regarding TPPs, the PSD2 distinguishes between "payment initiation service providers" (PISPs) and "account information service providers" (AISPs). One key benefit for customers delivered by AISPs could be the ability to have access to their banking information from multiple providers in the same place. To this effect, it is expected that a wide range of aggregator websites and apps will emerge to provide this information to customers in an easy-to-use interface, giving customers a more complete view of all the accounts they hold with different banks, and helping them to better monitor and manage their finances.

In addition to the regulatory initiatives, advances in technology – including immediate payment infrastructures, blockchain, and the Internet of Things (IoT) – are creating new ways to pay in a digitized end-to-end value chain. Customer demands are also evolving as more and more transactions are initiated via mobile devices, demanding a real-time, personalized, and seamless payment experience.

Although the PSD2 regulates only the field of payment accounts and services, the opportunities that open APIs and Open Banking hold for the financial services industry in a digitalized market are very substantial and not limited to payments only. After all, outside the banking industry, some of today's globally operating corporate giants in the digital

space could not have grown so fast in the past decade without the business-accelerating capabilities of APIs.

Opening up to other market participants outside of one's own organization creates value for customers and benefits the surrounding ecosystem. Open APIs and Open Banking could change the way the banking industry thinks about products and distribution – two key dimensions in every business. APIs, and digitization in general, allow value to be created in a distributed fashion, through an ecosystem of partners. Co-creating value is likely to prove to be a major change and challenge for banks in the near future.

The changes triggered by the move to open APIs will also impact many of the traditional business models in the banking industry, in line with the experience that a variety of industries have already gone through following the large-scale adoption of open APIs. For banks, as well as other players in this space, the key to success will be to adapt to this changing landscape by re-conceptualizing their business models around the customer, and seize the opportunities of APIs as an enabler of new products and services.

RELEVANT CONCEPTS IN APIS

In order to fully understand the potential impact of APIs, and especially open APIs, on the payments industry, it is necessary to clarify a few relevant concepts relating to APIs. The significance of the degree of openness of APIs, how it relates to creating value through APIs, and the extent to which this value depends on the level of standardization are key aspects in this regard, as is a widening of the scope from open APIs to Open Banking.

To start with basic definitions, APIs can be seen as interfaces between software applications, both within as well as between organizations. More specifically, APIs enable communication between software applications where one application calls upon the functionality of another.

APIs represent a specific software-architectural approach that revolves around the view that interfaces should be scalable, reusable, and secure, while offering ease of use for developers through self-service. APIs, therefore, hold the promise to reduce cost and lead time of interfacing between systems, allowing faster, cheaper, and better innovation on a larger scale.

Various business dimensions of APIs can be identified, starting with the concept of "openness" in relation to APIs.

The level of API openness determines potential reach

APIs enable secure, controlled, and cost-effective access to data and/or functionality. If APIs can only be accessed within the boundaries of one organization, they are referred to as "closed APIs" or "private APIs." If they can also be accessed by third parties (outside of the organizational boundaries), they are referred to as "open APIs." It is relevant to stress that "open" does not mean that every third party can access a bank's system at their discretion. There will always be some form of control by the bank, in order to preserve security, privacy, and contractual conditions. This will be further detailed below.

In practice, different levels of API openness can be observed. This is important because the level of openness determines the potential number of parties with access and thus the potential reach of the functionality offered through an API. For the purposes of the

present article, the following levels of “API openness” are to be distinguished:

- **Private APIs:** private APIs are closed APIs, and, therefore, exclusively accessible by parties within the boundaries of the organization.
- **Partner APIs:** APIs that are open to selected partners based on bilateral agreements. Like Private APIs, Partner APIs are exclusively accessible at the discretion of the provider of the APIs. Bilateral agreements on specific data exchanges between, for instance, a bank and an enterprise resource planning (ERP) software provider is an example of a Partner API.
- **Member APIs:** this type of API is open to everyone who is a formal member of a community with a well-defined set of membership rules. When becoming a member of such a community, the API provider allows access to the community members who comply with community membership rules and regulations. Account information and payment initiation services as defined under the PSD2 fall in this category, since only authorized or registered TPPs can obtain access.
- **Acquaintance APIs:** this type of open APIs is inclusive, as they are open to everyone complying with a predefined set of requirements. Developer portals distribute this type of API, which also comes with some form of standardized agreements. Merchant access to point-of-sale (POS) APIs is an example in this category.
- **Public APIs:** Public APIs are inclusive and can thus be accessed by anyone, typically with some form of registration for identification and authentication purposes.

The levels of API openness are depicted in Figure 1.

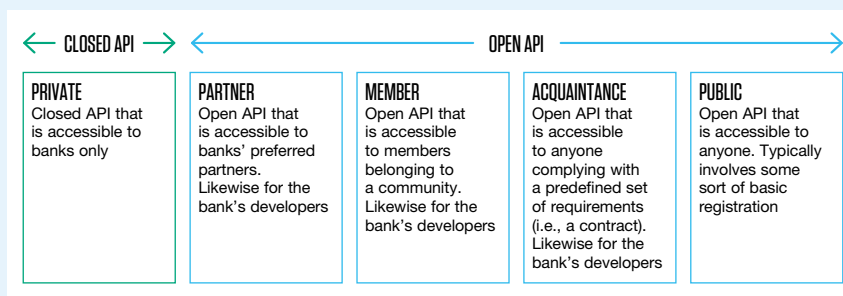


Figure 1 – Levels of API openness

Creating value with open APIs

Most digital players have used API technologies to meet their business objectives and ultimately create customer value. They have discovered that using APIs in opening up systems (to the outside world) is essential for driving traffic to one's assets, for co-creating end-customer value in the ecosystem, and for sharing the burden and benefits (including the profits) between the parties involved when unlocking new markets.

Value co-creation through APIs can be categorized as follows:

- **Enabling third parties to build applications “on top” of the platform:** examples include Facebook, Amazon, eBay, PayPal, Twitter, and Google. Developers can reuse existing functionality or use data sources to enrich their own applications. This lowers costs and speeds up time to market, but also creates additional dependencies on third-party developers. For API providers, this way of value co-creation provides a wider distribution network, creating traffic and minimizing innovation costs, which are carried by third-parties.
- **Social sharing for marketing purposes:** examples include Flickr, Delicious, Twitter, YouTube, LinkedIn, and Facebook. Social sharing is about sending, for example, photos,

videos, product recommendations, and website links to contacts within a social network. Social sharing is highly effective for branding and marketing purposes and for generating web traffic. Banks could use social sharing principles to build user communities, retain or increase brand awareness, and increase brand loyalty.

- **Syndicate products and services across different platforms:** examples include eBay and Google. Syndication occurs when multiple players work together to co-create and provide a service to a customer. Each player provides distinct features to create the value provided by the service. The fees paid by the customer are distributed amongst the syndicate membership.

The models described above not only enable and promote cooperation between different parties but also make it possible to create attractive and completely new value propositions. Ease of use, product and information aggregation, and direct communication are key for catalyzing this process.

Financial APIs require agreement on the scope and breadth of standardization

In the financial world, defining technical interfacing only is not enough for collaboration across organizations. Where

funds and sensitive data are involved, trust needs to be created. An additional dimension is provided by the various ways in which data may be handled, this includes different requirements for reading versus writing of data as well as for different types of data. Personal customer data require different provisions than bank data or aggregated (anonymous) customer data. A higher level of control is, therefore, needed. However, banks already have experience with controlled third-party access.

The financial services industry has a long tradition of applying control and standardization beyond technology, when creating infrastructures, such as for payments and securities, and when interfacing with clients and other third parties. In a fully digitized world (i.e., in a machine-to-machine environment without any foreseen human intervention), standardization becomes even more important. We distinguish four agreement and standardization dimensions (scope) in the banking industry:

1. **Legal:** rights and obligations of concerned parties for creating trust among the parties involved.
2. **Operational:** the agreements needed for running an API (after implementation): performance, uptime, service levels, support, etc.
3. **Functional:** aspects related to the user functionalities, data semantics, etc.
4. **Technical:** all aspects relating to technical implementation.

Industry infrastructures and business networks, such as today's payment systems (and financial systems in general), cannot function without agreements on all of these dimensions, either between individual banks or within communities. Consequently, financial APIs need at the very minimum a similar scope when it comes to agreements and standards.

The report by the U.K. Open Banking Working Group¹ provides explicit recommendations on the use of standards in the field of Open Banking.

In addition to the dimension of technical standardization, the governance in the field of standard setting is of utmost importance for the use of, and acceptance by, the users. The following governance levels could be distinguished:

1. **Organization:** this is the smallest unit of governance, as it concerns a single bank. Company policies, guidelines, and Member APIs fall under this category.
2. **Community:** standards are accepted and adopted by a group sharing common characteristics or interests, such as national communities, processors, banks, etc. The pan-European e-authorization solution MyBank is an example in this category, as well as the recent work done by the U.K. Open Banking Working Group.
3. **Industry:** standards are accepted and adopted by a complete industry on a regional or global scale. The SWIFT standards are an example of an industry standard. The SEPA Schemes also fall into this category.
4. **Universal:** standards are accepted and adopted by multiple industries around the world. Any standard defined by an international organization such as ISO, ITU or IEC fall into this category (HTTP/HTTPS used for Internet communication is one practical example).

From open API to Open Banking

From a strategic point of view, payments industry players in general – and account-servicing payment service providers, such as banks, in particular – should consider their use of (open)

APIs against the wider horizon of Open Banking and their positioning in that broader emerging ecosystem, which is facilitated by, but not limited to, a growing usage of open APIs across the financial services industry.

There are several definitions in use for Open Banking, coming, for example, from the Open Bank Project² and the U.K. Open Banking Working Group.³ In both cases, Open Banking revolves around the standardization of how banks share their own data, but also how they allow customers more choice and sharing of their data for use in third-party (FinTech) applications in a secure and resilient fashion. Open Banking can be characterized as a technology-driven evolution of banking, and this includes Open APIs. As such, Open Banking is a movement “bridging two worlds,” i.e., making it possible for customers to use their banking services in the context of other (FinTech) services, thereby combining innovative functionalities from banks and non-banks with reach through infrastructure.

Functionally, Open Banking is about how banks share their own products (i.e., services, functionality, and data) and how they enable their customers to share their data and account functionality with third-party (e.g., FinTech) applications in a secure and resilient fashion. As customers drive the actual uptake of such innovations, the concept of “customer ownership” or “product centric” is changing towards a concept of “customer centric” between banks and third-party developers.

1 <http://bit.ly/1nYWjv4>

2 <http://bit.ly/2mSAa1E>

3 <http://bit.ly/1nYWjv4>

The real challenge of customer centricity through Open Banking is to move away from the product-driven push and to develop the corporate capabilities to truly understand the customer's stated and tacit needs, to generate new ones, and to provide highly personalized solutions and experiences. Achieving customer centricity is a journey, an iterative one that may have significant impacts on the entire organization. It will require banks to holistically rethink the way the business is conducted, starting from its customers.

OPEN BANKING FROM A BANK'S PERSPECTIVE

Open Banking challenges traditional assumptions by creating new opportunities in product creation and distribution. Traditionally, banks have not only provided their customers with products but have also been responsible for the distribution of these products, i.e., the bank distributed its payment products through its own banking channels, such as mobile, web, and branches. In this traditional scenario, the bank controls the entire product and distribution chain.

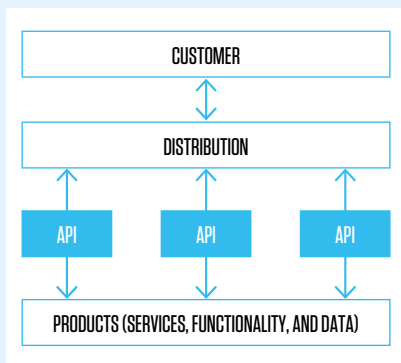


Figure 2 – APIs are the pivot between products and distribution

Open Banking redefines both product and distribution as the principles (re-usable, scalable, secure, self-service), technologies, and agreements of Open Banking allow for new possibilities. Figure 2 demonstrates where APIs fit in between products and distribution.

Using APIs for interfacing between product and distribution enables banks to decouple these functions. The combination of decoupling and opening-up allows banks to play different roles in the financial value chain with regards to the offering of products and the distribution of these products. Two fundamental strategic questions need to be looked at in this context:

1. Who distributes the products that are made accessible via an API to existing and new customers?
2. Who creates the products that need to be distributed to a bank's customer base?

Based on these two questions, four generic roles in the financial value chain may be defined as illustrated in Figure 3: integrator, producer, distributor, and platform.

Most of the larger financial institutions already play roles 1, 2, and 3 (integrator, producer, and distributor) at the same time (often assigned to different business lines or products), whereas role 4 (platform) is still at a very early stage of its development. The roles could also indicate the platformization levels of Open Banking. Each level correlates with increasing customer control when looking at things from the customers' perspective (Figure 4).

Embracing a new role in the financial value chain with a limited or extended level of platformization entails transformational challenges, as it requires changes in the business and operating

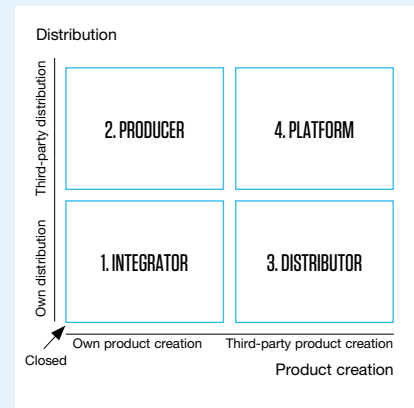


Figure 3 – Potential roles in the digital value chain

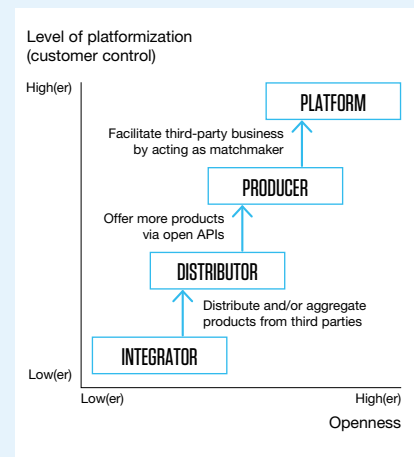


Figure 4 – The four platformization levels of Open Banking

model. Criteria to consider when evaluating the level of strategic change include customer choice and control, customer loyalty, market propositions, cost efficiencies, innovation culture, employer attractiveness, business and IT alignment, available means for investing, and possibly outsourcing.

Given this wide range of criteria and potential consequences to consider, decision-makers within banks are faced with significant strategic challenges. Opening up and giving customers more control can have a positive impact on revenues (and profits) as successful open

(platform) models in the non-financial industry have shown in the recent past. Risks, in terms of increased compliance challenges and increased competition, are mounting as well, all potentially to be mitigated and adequately managed by changes in the operating model. Each level of platformization comes with varying challenges, where the pursuit of a “doing nothing” strategy is not an option.

CONCLUSION

The above elaborations describe the strategic crossroads that every account-servicing payment service provider could face in the next one to three years, when confronted with choices on how to approach or embrace Open Banking and apply API technology. The minimum engagement in “opening-up” is what the PSD2 will prescribe in terms of access-to-account, i.e., a limited “producer role” and thus limited level of platformization, while the current FinTech and innovation initiatives pose growth challenges regarding business strategies for partnering and product propositions towards third-parties.

The ever-changing customer expectations – driven by the experiences customers make in their digital life – will increase this need for more advanced levels of platformization, enabling ultimate customer choice and control options. It will be up to each individual player in the market to rise up to the challenge of meeting these expectations and seizing the opportunities offered by the industry’s move towards open APIs and Open Banking.

Algorithmic Regulation: Automating Financial Compliance Monitoring and Regulation Using AI and Blockchain

Philip Treleaven – Professor of Computer Science and Head of Financial Computing Centre, University College London

Bogdan Batrinca – Honorary Research Associate in Financial Computing, University College London

Abstract

Efficient financial regulation is crucial to the future success of the financial services industry and especially the rapidly evolving new financial technology (FinTech) area. The concept of “algorithmic regulation,” modelled on “algorithmic trading systems” [Treleaven et al. (2013)], is to stream compliance, social networks data, and other kinds of information from different sources to a platform where compliance reports are encoded using distributed ledger technology and regulations are “codifiable” and “executable” as computer programs, using the same technology being developed for blockchain smart contracts. In this paper, five areas are discussed: a) an “intelligent regulatory advisor” as a front-end to the regulatory handbook; b) “automated monitoring” of online and social media to detect consumer and market abuse; c) “automated reporting” using online compliance communication and big data analytics; d)

“regulatory policy modeling” using smart contract technology to codify regulations and assess impact before deployment; and e) “automated regulation” employing blockchain technology to automate monitoring and compliance. We refer to algorithmic regulation for systems that facilitate compliance and regulation decision-making in financial services using advanced mathematical tools and blockchain technology.¹

¹ Tim O'Reilly originated the term more generally for “government by computer algorithms.” Algorithmic Regulation, Wikipedia, <http://bit.ly/2i1pT2O>

INTRODUCTION

There is a growing concern about regulation and compliance, which is increasingly perceived to have negative effects on the development of financial services, discouraging innovation by requiring an ever-growing amount of data reporting. Overcoming this impasse requires radical automation, especially for regulation of new FinTech entrants [Brummer and Gorfine (2014), PayPal (2013)].

This paper explores five regulatory technology (RegTech) areas ripe for automation in regulation using blockchain technology (see Figure 1):

- **Intelligent regulatory advisor:** an artificial intelligent front-end to the regulatory handbook to simplify registration.
- **Automated monitoring:** monitoring of online and social media, and using natural language processing and sentiment analysis to monitor consumer opinions, concerns, and level of trust and identify market abuses.
- **Automated reporting:** using the FinTech paradigms of online communication, big data analytics, and distributed ledger technology to automate compliance and regulation reporting [known as RegTech in the U.K.: U.K. Government Office for Science (2015)].
- **Regulatory policy:** using smart contract technology to codify regulations; and using computational modeling, such as agent-based systems, for assessing regulatory proposals' potential market impact before deployment (e.g., Basel IV, MiFID II, Solvency III).
- **Automated regulation:** the most interesting, using blockchain distributed ledger technology to record compliance reports and use smart contract technology [U.K. Government Office for Science (2016), Norton Rose Fulbright (2016)] to codify, computerize, and automate financial regulation and compliance (cf. algorithmic trading).

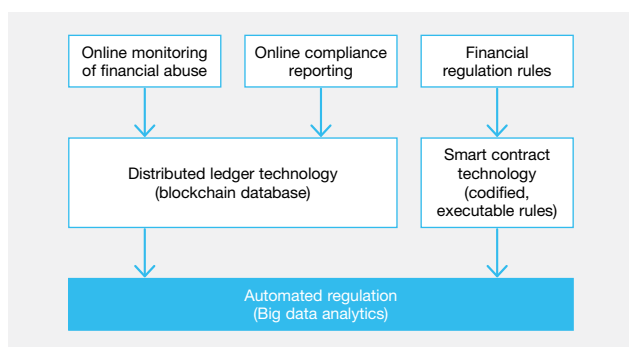


Figure 1 – Algorithmic regulation using blockchain technology

AUTOMATING REGULATION AND COMPLIANCE

Financial regulation is becoming increasingly burdensome. Research from the American Action Forum has suggested that as of July 2016 U.S. banks had paid U.S.\$24 bln and allocated 61 million employee hours to comply with Dodd-Frank Wall Street Reform and Consumer Protection Act, passed in the U.S. amid outcry over the financial crisis [Batkins and Goldbeck (2016)].

That said, financial regulation faces a myriad of pressures: political pressure to curb excesses (e.g., Libor); escalating international and European Union regulations (e.g., MiFID II); individual firms simultaneously regulated in multiple jurisdictions and with frameworks; and institutions asked to produce increasing amounts of financial, risk, and compliance data. All this pressure has generated the negative perception that data is being requested “speculatively” and not being used by the regulators. The challenge is to simplify and balance regulation while encouraging innovation for new FinTech alternative finance entrants, in rapidly changing environments [U.K. Government Office for Science (2015)].

In recent years, a number of technologies that can help handle this increased demand for detailed reporting have been developed and have reached commercial maturity:

- **Data scraping:** the technique in which a computer program extracts data from human-readable output coming from the Internet or another program. This involves scraping social networking sites such as Twitter, Facebook, etc., but also web pages, forums, blogs, RSS feeds, online newspapers, and product/service reviews or feedback.
- **Natural language processing:** content interpretation of natural language by means of algorithms mainly based on machine learning.
- **Sentiment analysis (or opinion mining)**²: the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer’s attitude towards a particular topic, product, etc. is positive, negative, or neutral [Medhat et al. (2014)].
- **Automated fraud detection:** identifying suspicious patterns in credit card transactions, identity theft, insurance claims, money laundering, insider dealing, etc.³

2 Sentiment analysis techniques, Wikipedia, <http://bit.ly/1e2zqkS>

3 Data analysis techniques for fraud detection, Wikipedia, <http://bit.ly/1OQnzX1>

- **Big data analytics:** the process of examining large data sets containing a variety of data types to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful business information.⁴

Potential solutions for automating regulation and compliance include an intelligent regulatory advisor, automated systems for monitoring and reporting, regulatory policy modeling, and ultimately an automated regulation system.

Intelligent regulatory advisor

A major challenge for new financial companies is navigating a regulator's handbook and completing the registration process. A solution is to provide an artificial intelligence front-end that supports the location of relevant information and guides that user through registration.

Automated monitoring

The monitoring challenges faced by regulators are illustrated by the U.K. Financial Conduct Authority (FCA). Previously, the FCA monitored 25,000 large and medium size firms. With essentially the same resources, the FCA now has to supervise an additional 21,000 small firms. The obvious solution is to monitor social media for financial abuses.

Developed for brand management and customer profiling, there are a number of sophisticated data scraping and sentimental analysis tools that can equally be deployed by regulators for automated monitoring. Examples include Adobe Social, Brandwatch, Google Alerts, and Mention [Batinca and Treleven (2015)].

Automated reporting

One of the recommendations of the U.K. Chief Scientist's review of the emerging new financial technology (FinTech) sector was the so-called RegTech [U.K. Government Office for Science (2015)], in order to use FinTech-style online analytics software techniques (cf. peer-to-peer) to improve compliance and regulation of FinTech companies. Regulation of major financial institutions is largely immutable, set by international, U.S., and E.U. authorities. In contrast, regulation of rapidly evolving FinTech companies arguably provides an opportunity to pioneer lightweight automated reporting.

The three key requirements for automating compliance are [Brunner and Gorfine (2014), PayPal (2013)]: a) **Reporting language** – employing a standard (XML) compliance reporting language, the emerging standard is ISO 20022⁵; b) **Reporting platform** – employing a standard, lightweight, client-side reporting platform that interfaces to industry standard accounting

systems, especially for small firms⁶; and c) **Regulatory analytics** – for transparency, employing standard compliance software applications, such as anti-money laundering (AML) or know your customer (KYC), used by both the reporting firms and the regulators.

Regulatory policy modeling

Another emerging area is the use of (agent-based) computational models to evaluate laws and regulations prior to deployment. For example, a number of the regulatory proposals considered after the 2010 Flash Crash (e.g., lodging algorithms with regulators, best price quotes, trading pauses, tick sizes, etc.), if implemented, may have actually increased systemic risk [U.K. Government Office for Science (2012)].

Automated regulation

Here, the concept – inspired by algorithmic trading systems – is a comprehensive automated system for compliance and regulation, where analytics is driven by regulations encoded as computer programs, leveraging blockchain smart contract technology. Below, as background, we explain in simple terms blockchain and distributed ledger technology in the context of cryptocurrencies, and then how the technology is being developed for smart contracts. Having laid this groundwork, we then discuss the possible design of an algorithmic regulation platform.

BLOCKCHAIN AND DISTRIBUTED LEDGER TECHNOLOGY

Blockchain⁷ [Lewis (2015)], originally conceived for Bitcoin and other digital currencies (or cryptocurrencies), is now recognized to have far-reaching potential in other areas, such as computer-executable contracts. People use the term “blockchain technology” to mean different things, and it can be confusing. Sometimes they are talking about the bitcoin blockchain, sometimes it is other virtual currencies or digital tokens, sometimes it is smart contracts, but mainly it is about distributed ledgers.

A distributed ledger is where all transactions are kept in a shared, replicated, synchronized, distributed bookkeeping record, which is secured by cryptographic sealing and made

4 Big Data Analysis, Wikipedia, <http://bit.ly/1FSZacQ>

5 ISO20022 Regulatory Reporting XML, <https://www.iso20022.org/>

6 OpenMRS and other medical records systems, <http://bit.ly/1p9IQWN>

7 Smart contracts, Wikipedia, <http://bit.ly/2im8wXZ>

hard to alter by a computationally costly proof-of-work. Every participant (node) has a ledger replica. Nodes synchronize the ledger periodically by approving blocks of transactions. The validity of a block is established by the next block attaching to it, forming, therefore, a chain. The blockchain is the chronological list of all blocks of transactions from the genesis block.

Bitcoin blockchain is a public ledger of all “coin” transactions. The proof-of-work is a crucial element. It is constantly growing as “completed” blocks are added. The blocks are added to the blockchain in a linear, chronological order (every 10 minutes). Each node (i.e., computer connected to the digital currency network) uses a client that performs the task of validating and relaying transactions. The blockchain has complete information about the addresses and their balances right from the genesis block to the most recently completed block. Consequently, in simple terms, a “block” is an encrypted, linked record, and a “blockchain” is a continuously growing list of data records held in a replicated, distributed database or ledger.

For regulatory reporting, the blockchain would be fundamentally a record of the transaction history, delivering a fully transparent, accessible transactional database for regulatory bodies.

Smart contracts

As discussed, a smart contract is a codified legal contract, executable as a computer program, which can initiate actions (e.g., payments). Smart contracts can interact with any software system including other contracts, and potentially highlight when they are no longer valid (e.g., due to changes in the law)⁸ [Oasis (2007)].

The potential benefits of smart contracts codified and executable by a computer include formal verification [Walker (1990)], lowering the cost of contracting for low-value transactions, automation, enforcement, and compliance.

Hence, smart contracts can define strict rules and consequences in the same way that a traditional legal document would, stating the obligations, benefits, and penalties that may be due to either party in various different circumstances. But, unlike a traditional contract it can also take information as an input, process that information through the rules set out in the contract, and take any actions required of it as a result. A contract could potentially recognize when it is no longer valid or legal.

Automated trading contract example

As a further illustration of smart contracts’ potential applications, consider a manufacturer in China shipping a product to a retailer in Europe. The manufacturer has a contract with the shipping agent, the agent with a shipper, the shipping agent with the receiving agent in Europe, in turn a contract with a haulier, a contract with the distributor, and lastly the distributor with a retailer. At each stage in the supply chain, the appropriate contract executes, the next stage is informed, responsibility is transferred, and the previous stage is paid.

Perhaps more interestingly, as currencies fluctuate, trade tariffs are applied, and laws change, the various contracts could automatically apply the new rates, alert their owners, or potentially reconfigure.

In preparing smart contracts for the above example, it is important to understand that supply chains are complex by their nature, with various parties involved, from manufacturers, shippers, distributors, and retailers, all the way to the consumer. This is especially true when the supply chain partners are in different countries and each partner is responsible for their own working capital and inventory. Trade is typically financed via a Letter of Credit (LC), which, although guarantees payment, is acknowledged to be costly (2%-4% on an annual basis), error prone, and necessitates intermediaries.

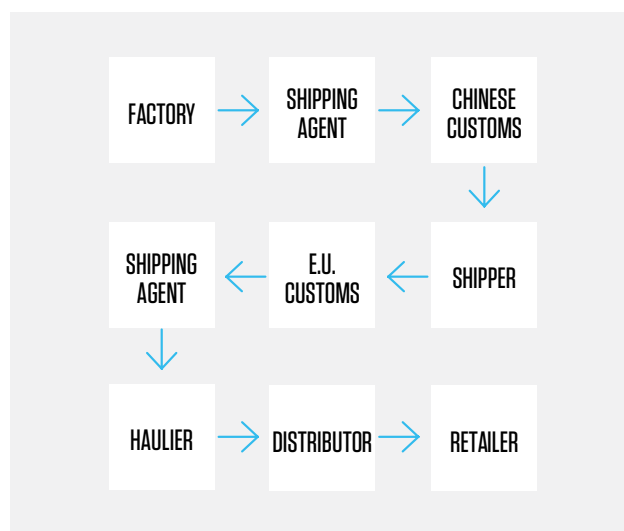


Figure 2 – Smart contracts in a supply chain

⁸ Computational law, Wikipedia, <http://bit.ly/2iSiHHV>

In summary, blockchain smart contract technology offers major potential for business efficiency: automation of back-office functions, increased control, reduction in errors, and a major reduction in cost. Currently, most smart contract developments focus on smart “financial” contracts between major financial institutions, notably the R3 Consortium,⁹ rather than general legal contracts or statutes.

SMART CONTRACT PROGRAM NOTATION

Smart contract platforms, such as Ethereum,¹⁰ have developed their own proprietary contract notation. The obvious question to ask is whether we can use traditional programming languages such as Haskell, Python, or Java, given their wealth of associated content, in order to code contracts. Notations broadly cover: a) declarative, functional, and logic languages that are mathematically concise (e.g. Haskell, F#, Prolog); and b) imperative languages (e.g. Python, Java):

- **Declarative languages** – a programming paradigm that expresses how to accomplish the problem or the logic of a computation without describing its explicit steps.
 - **Special-purpose languages** – a specification that describes the problem to be solved for a specific domain, such as database programming (e.g., SQL) or smart contracts (e.g., Ethereum).
 - **Functional languages** – a style of programming that models computations as the evaluation of expressions (e.g., Haskell, F#).
 - **Logic languages** – a programming paradigm based on formal logic, where a program is a set of sentences expressing facts and rules about some problem domain (e.g., Prolog).
- **Imperative languages** – a programming paradigm that uses statements that tell the computer what to do and that change a program’s state.
 - **Procedural languages** – a programming paradigm that specifies a series of well-structured steps and procedures to complete a computational task or program (e.g., C).
 - **Object-oriented languages** – a programming paradigm that defines not only the type of a data structure, but also the types of operations (functions) that can be applied to the data structure (e.g., Java, C++, Python).

The benefits of declarative languages are that they are more concise, and amenable to mathematical analysis and verification [Walker (1990)], but these languages are less popular

for general programming. In contrast, imperative languages are computationally powerful, efficient, and popular, but the semantics of a program can be more complex and difficult to prove due to so-called side-effects.

As an illustration of the possible use of traditional languages for programming smart contracts, we show some pseudo-code in a declarative subset of Python; a “multi-paradigm” language. Although programming in a declarative style of Python¹¹ may seem an odd constraint to work under, it brings a number of benefits. From a mathematical viewpoint, the benefits include formal provability, modularity, composability, and ease of debugging and testing, whereas pragmatically, the benefits consists of the wealth of associated code, and seamless analytics. Figure 3a illustrates Python pseudo-code for a simple smart contract.

Returning to markup languages, a significant feature is that they are translatable into a human-readable format (cf. HTML to web page), which could be a major smart contract benefit when collaborating with lawyers. Arguably, in addition to choosing a declarative programming notation for smart contracts, we should also ensure it is renderable into plain text. This is illustrated by Figure 3b.

In Figure 3a, code is in blue and black, and values in red.

As discussed, we believe algorithmic trading is an interesting model for the proposed fully automated algorithmic regulation systems.

9 R3 Distributed Ledger Consortium, <http://bit.ly/2jdOTBL>

10 Ethereum blockchain platform, <http://bit.ly/29sDD4H>

11 Kuchling, A., “Functional programming (in Python) HOWTO,” <http://bit.ly/2j0UBaE>

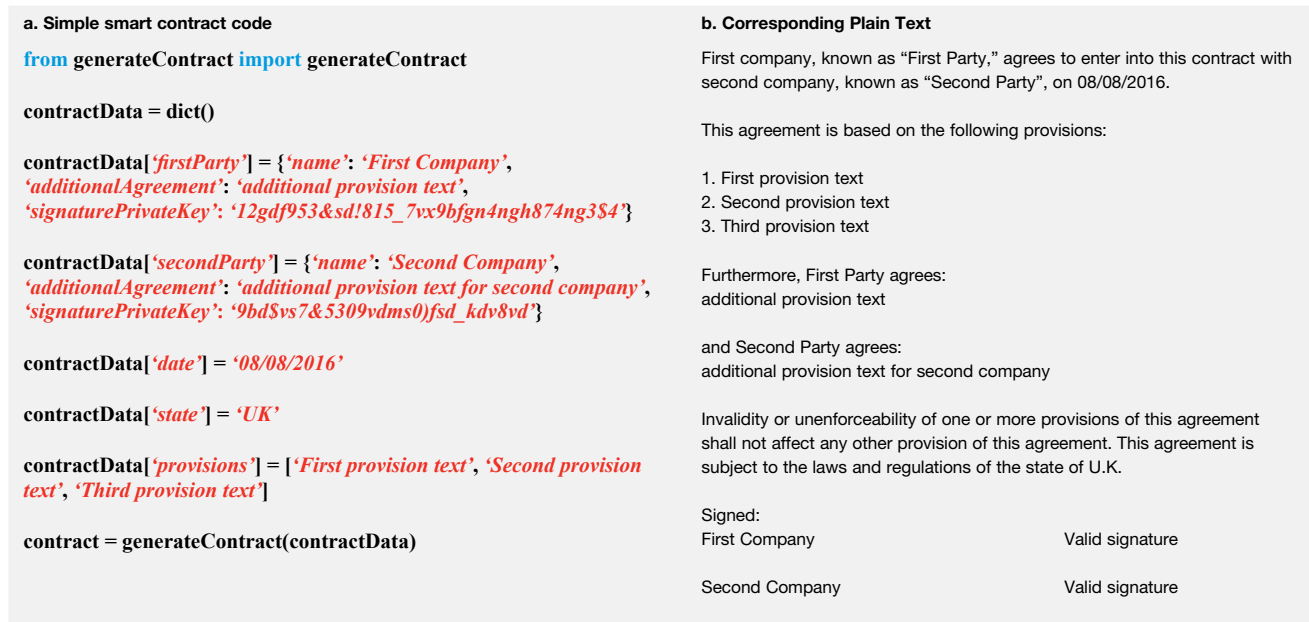


Figure 3 – Simple smart contract – declarative (Python) pseudo-code and corresponding Plain Text

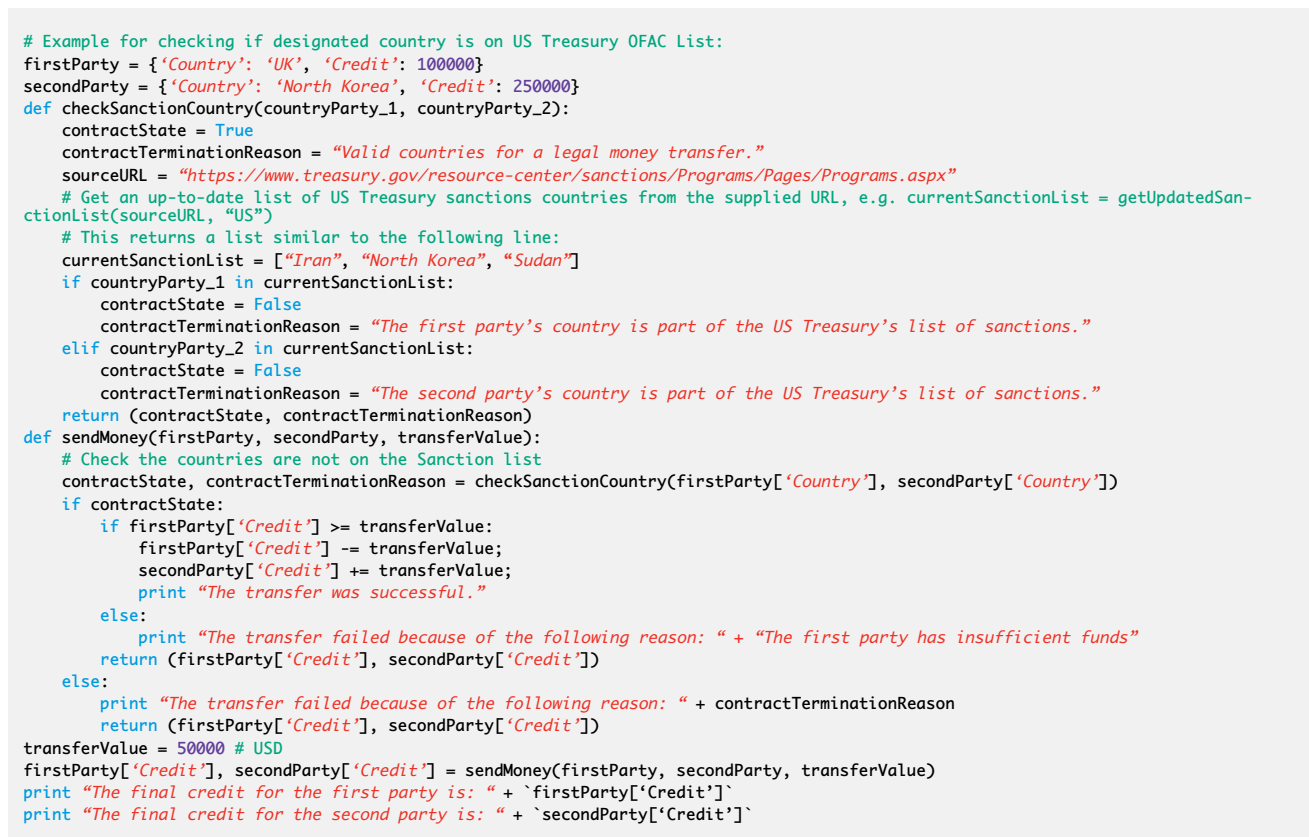


Figure 4 – Smart regulation notation for U.S. Treasury sanctioned countries

ALGORITHMIC REGULATION

We now explore the potential structure of algorithmic regulation systems built upon blockchain smart contract technology. We start by looking at smart “regulation” contracts notation, and then discuss the components of an algorithmic regulation system.

Smart regulation contract

As an example, Figure 4 shows the Python code to check a cross-border payment against an abbreviated list of U.S. Government sanctioned/embargoed countries (e.g., North Korea). The complete list of countries is on the U.S. Department of the Treasury Office of Foreign Assets Control (OFAC) website,¹² with levels of sanctions varying by country. This example is purely illustrative.

In Figure 4, code is in blue and black, comments in green, and values in red. The two principal routines are checkEmbargoCountry and sendMoney.

Algorithmic regulation system

The proposed system (see Figure 5) comprises five main components for: a) “intelligent regulatory advisor” front-end to the regulatory handbook; b) “automated monitoring” of online

and social media data to identify individuals’, firms’ and sector-wide potential abuse; c) “automated reporting” by regulated firms, notably FinTech companies; d) “regulatory policy” specified by international, government, and regulatory bodies; and e) “automated regulation” where regulations are codified, compliance reports are stored in a blockchain, and regulatory analytics is applied to identify abuse, regulatory breaches, and potential risks.

Automated monitoring: this covers scraping the web, social media sites, newspapers, blogs, and chat rooms, seeking to identify complaints about individuals and firms, and sector-wide abuse, such as the incorrect selling of Payment Protection Insurance (PPI) in the U.K. Although there is a number of commercial tools for harvesting web data, such as Adobe Social, Brandwatch, and Synthesio, identification of potential sources of online information remains a big challenge, since disadvantaged victims of small financial firms are unlikely to use Twitter or Facebook to air their grievances.

Automated reporting: as discussed, multiple E.U. and U.S. regulatory bodies are already adopting the ISO 20022 XML standard for reporting. The additional requirement is the need for a “light-weight” (open-source) platform using ISO 20022 XML for compliance reports for small financial companies.

Regulatory policy: for regulatory policy, firstly a declarative smart contract notation is required to encode regulations, and secondly the requirement to use agent-based modeling of proposed regulations for assessing the impact of proposals before deployment.

Automated Regulation: lastly, automation comprises five components: 1) the monitoring analytics component that uses sentiment analysis to identify individuals, firms, and sector-wide problems that may cause concern; 2) the compliance reports encoded using blockchain distributed ledger technology; 3) the compliance analytics component that seeks to identify regulatory breaches, AML, KYC, etc.; 4) the systemic risk component that seeks to identify major firms at risk (e.g., Solvency II); and 5) the regulatory rules component that contains codified regulations using Smart Contract technology.

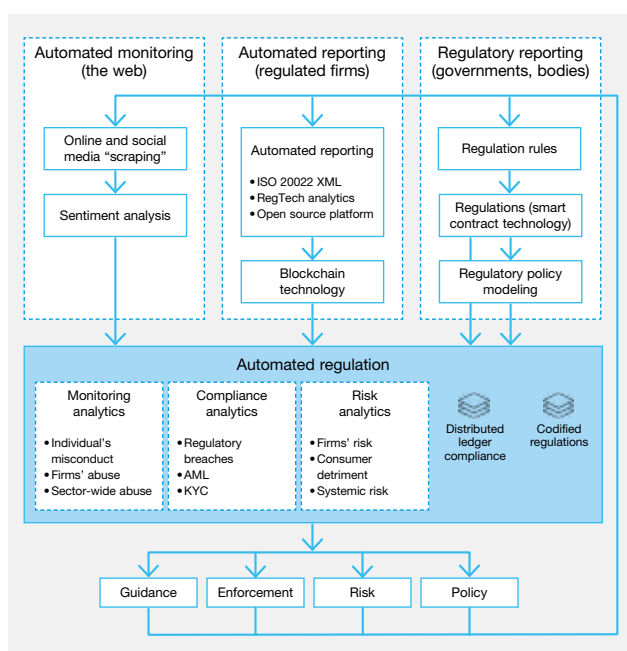


Figure 5 – Algorithmic regulation system

12 U.S. Department of the Treasury: Office of Foreign Assets Control, <http://bit.ly/23za3iL>

CONCLUSION

This paper presents the concept of algorithmic regulation modeled on the algorithmic trading paradigm, and employing technology under development for blockchain distributed ledgers and smart contracts. The five major components are: “intelligent regulatory advisor,” “automated monitoring” of abuse, light-weight “automated reporting” principally for FinTech companies, “regulatory policy modeling,” and “automated regulation.” As discussed, algorithmic regulation applied to finance builds on the pioneering work of the R3 consortium of banks in the area of smart “financial” contracts, and any results will be applicable to smart “legal” contracts in general, and the “algorithmic regulation” paradigm applied to government, as proposed by Tim O’Reilly, the founder and CEO of O’Reilly Media Inc. What is clear is that blockchain smart contract technology will have a more major “disruptive” effect on legal services than FinTech is having on financial services.

REFERENCES

- Batkins, S., and D. Goldbeck, 2016, “Six years after Dodd-Frank: higher costs, uncertain benefits,” The American Action Forum, July 20
- Batrinca, B., and P. Treleven, 2015, “Social media analytics: a survey of techniques, tools and platforms,” AI & Society 30:1, 89-116
- Brummer, C., and D. Gorfine, 2014, “FinTech: building a 21st century regulator’s toolkit,” Center for Financial Markets, Milken Institute, <http://bit.ly/1CRgQBP>
- Lewis, A., 2015, “A gentle introduction to blockchain technology,” BraveNewCoin, <http://bit.ly/2jdE8iz>
- Medhat, W., A. Hassan, and H. Korashy, 2014, “Sentiment analysis algorithms and applications: a survey,” Ain Shams Engineering Journal 5:4, 1093–1113
- Norton Rose Fulbright, 2016, “Smart contracts: coding the fine print,” <http://bit.ly/2iZCdz8>
- Oasis, 2007, “eContracts version 1.0: committee specification,” <http://bit.ly/2iM1SNZ>
- PayPal, 2013, “21st century regulation: putting innovation at the heart of payment regulation,” <http://bit.ly/2hZzNid>
- Szabo, N., 2002, “A formal language for analyzing contracts,” <http://bit.ly/2i2IRpB>
- Treleven, P., M. Galas, V. Lalchand, 2013, “Algorithmic trading review,” Communications of the ACM 56:11, 76-85
- U.K. Government Office for Science, 2012, “The future of computer trading in the financial markets,” a report of the U.K. Government Chief Scientific Advisor, <http://bit.ly/2j0LXZH>
- U.K. Government Office for Science, 2015, “FinTech futures: the UK as a world leader in financial technologies,” a report of the U.K. Government Chief Scientific Advisor, March 18, <http://bit.ly/1FCBDgS>
- U.K. Government Office for Science, 2016, “Distributed ledger technology: beyond blockchain,” a report of the U.K. Government Chief Scientific Advisor, <http://bit.ly/1WreVPL>
- Walker, H. M., 1990, “Program verification (tutorial session): techniques and uses,” SIGCSE '90 Proceedings symposium on Computer science education

RegTech is the New Black – The Growth of RegTech Demand and Investment

Kari S. Larsen – Counsel, Reed Smith LLP

Shariq Gilani – Associate, Reed Smith LLP

Abstract

The financial services industry has undergone significant changes over the past decade, due primarily to technological innovations and regulatory change. In recent years, financial services regulatory requirements have dramatically increased, and the costs of compliance have increased correspondingly. Various regulations in the U.S., Europe, and worldwide have together greatly increased capital, recordkeeping, reporting, administrative, and other compliance costs, significantly raising the barriers to entry.

FinTech is broadly defined as technological innovations that support or enable banking and other financial services, potentially disrupting the financial services sector and/or making it more efficient. RegTech is a subgroup of FinTech, described as technology that is providing solutions to companies across all sectors of financial activity to ensure that they are able to comply with regulatory requirements.

Increased regulatory burden has created high demand for new technological solutions to regulatory challenges. Whilst regulations are becoming increasingly prescriptive as to the result to be achieved (e.g., what details of trades must be reported), they generally are not prescriptive as to how to achieve the required result, so the methods of compliance can be varied. RegTech innovations have the potential to increase margins for companies subject to a myriad of multi-jurisdictional requirements and to allow for competition by creating a path for a less expensive entry for startups into heavily regulated industries. In this article, we discuss the potential regulatory burdens placed on financial institutions and startups, and the related costs, potential solutions being offered, recent investments in RegTech, and what we see for RegTech in an uncertain future.

INTRO TO FINTECH AND REGTECH

The financial services industry has undergone significant changes over the past decade, due primarily to technology innovations and regulatory change. Both have altered how financial services are provided to consumers and each affects the other.

FinTech is broadly defined as technological innovations that support or enable banking and other financial services, potentially disrupting the financial sector and/or making it more efficient. Originally, FinTech referred primarily to back-office technology but now the term applies to any innovation in how people transact business in the financial services sector, including trading, online banking, double-entry bookkeeping, mobile finance apps, peer-to-peer lending sites, digital wallets, and newly created crypto-currencies and digital assets.

RegTech is a more specific term coined to classify a group of companies that, by harnessing the capabilities enabled by new technologies such as cloud computing, big data, and blockchains, are devising solutions to help companies across all sectors of financial activity, ensure that they comply with regulatory requirements. In financial parlance, RegTech is deemed a subgroup of FinTech.

FinTech companies have experienced a massive influx of investment in the last few years. However, among the factors inhibiting further FinTech development and startups are the various regulatory regimes, which restrict the way in which financial services can be provided. These are especially burdensome on new entrants. Even for established companies used to dealing with complex regulations, compliance with the recent slew of regulations has become an even larger cost burden (in money and time) on businesses.

There is, therefore, an increasing demand to create new solutions to overcome the regulatory challenges. Whilst regulations are becoming increasingly prescriptive as to the result to be achieved (e.g., what details of trades must be reported), they are generally not prescriptive as to how to achieve the required result, so the methods of compliance can be varied.

RegTech offers the potential for smaller companies subject to significant regulatory requirements to expand quickly, by using new technologies such as machine learning, cloud computing, and blockchains to give them the know-how with regard to the regulatory environment, help them interact with it, and allow them to meet their obligations, without the need for a large and expensive operations and compliance infrastructure.

Similarly, RegTech innovations have the potential to reduce costs and increase margins for large banks and companies that are being challenged by FinTech startups. RegTech innovations also have benefits outside the financial services sector; for example, for companies that would benefit greatly from performing quick identity checks.

The potential solutions and cost savings that RegTech offers have caused an increasing proportion of FinTech investments to be made specifically in the RegTech area.

Below we discuss some of the potential regulatory burdens placed on financial institutions and startups, and the related costs. We next discuss how some RegTech companies are addressing those requirements and solving potential issues. We also discuss investments in RegTech and what we see for RegTech in an uncertain future.

THE COSTS AND ADMINISTRATIVE BURDEN OF RECENT REGULATORY REFORMS

In the last decade or so, financial services regulatory requirements have dramatically increased, and the costs of compliance have increased correspondingly. The 2007–2008 financial crisis affected markets worldwide. The G20 held a Summit on Financial Markets and the World Economy in November, 2008, in Washington, D.C. There was general agreement among the G20 on how to cooperate in key areas to strengthen economic growth and to deal with the financial crisis. With the subsequent G20 Summit in Pittsburgh in 2009 and implementation of the Basel III accords, the foundations were laid for reforms aimed at avoiding similar crises in the future.

The new regulatory requirements and laws implemented to give effect to the G20's plans imposed additional burdens on financial institutions that have dramatically raised the cost of doing business and making it increasingly difficult for new entrants to access the sector. The Dodd–Frank Wall Street Reform and Consumer Protection Act¹ (Dodd–Frank) in the U.S. and various regulations in Europe, including the Markets in Financial Instruments Directive (MiFID),² the Market Abuse

1 The Dodd–Frank Wall Street Reform and Consumer Protection Act (Pub.L. 111–203, H.R. 4173)

2 Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004 on markets in financial instruments amending Council Directives 85/611/EEC and 93/6/EEC and Directive 2000/12/EC of the European Parliament and of the Council and repealing Council Directive 93/22/EEC

Regulation (MAR),³ and the European Market Infrastructure Regulation (EMIR),⁴ have together greatly increased capital, recordkeeping, reporting, administrative, and other compliance costs, significantly raising the barriers to entry. At the same time, more onerous regulations have been introduced that not only focus on the stability of the financial sector, but also on areas such as data protection, cybersecurity, and increasingly stringent know-your-customer (KYC) requirements.

U.S. regulations

The extensive increases in regulation of the U.S. financial sector began in the early 2000s. The accounting malpractice scandals that affected companies like Enron Corporation and WorldCom resulted in a lack of investor confidence. Then, the financial crisis of 2007–2008 plunged the U.S. economy into a recession. As a result of the accounting scandals, Congress passed the Sarbanes–Oxley Act of 2002 (Sarbanes–Oxley) to protect investors against the potential for fraudulent corporate accounting activities. The recent financial crisis led to widespread calls for reform and resulted in Dodd–Frank, signed into law in 2010. At the same time, the increased use of and reliance on technology, and the corresponding threats, have led to new cybersecurity rules, such as those put in place by the National Futures Association⁵ and the U.S. Financial Industry Regulatory Authority.⁶

These new layers of regulation are intended to provide rigorous standards and supervision; to protect the economy as well as consumers, investors, and businesses; provide an advanced warning system to ensure the stability of the economy; enhance corporate governance; and provide transparency to markets and investors. While many argue that the greatly enhanced regulatory regime has made the markets and financial

institutions safer and more stable, there is no doubt that regulation has also brought significant increases in compliance costs to both large and small companies – much of which ultimately has to be borne by the customer. For example:

1. **Sarbanes–Oxley:** the costs for compliance with Sarbanes–Oxley continue to increase for many companies. Internal and external audit fees, the associated costs of man-hours, and the fees for IT processes and controls are still rising. A 2015 poll of audit executives and professionals revealed that 58% of large companies spent more than U.S.\$1 million each on Sarbanes–Oxley compliance in 2014.⁷
2. **Implemented Dodd–Frank regulations:** the Dodd–Frank regulations have imposed more than U.S.\$36 billion in costs on the economy since 2010 and have created approximately 73 million paperwork hours, according to a new report from the conservative American Action Forum (AAF).⁸ The largest sources of costs related to margin and capital requirements for swap entities (costing U.S.\$5.2 billion in the past year) and margin requirements for uncleared swaps (costing a further U.S.\$2.1 billion). Certain disclosure requirements have resulted in approximately U.S.\$3 billion in costs. During 2015–2016, costs resulting from rules implemented under Dodd–Frank – which reached its sixth anniversary on July 21, 2016 – totaled U.S.\$10.4 billion, the highest amount in any year since the introduction of the regulations (Figure 1).⁹
3. **Outstanding Dodd–Frank regulations:** the AAF report also estimated that there were 61 regulations remaining in the Dodd–Frank rulemaking mandates that could add an additional U.S.\$3.3 billion and 1 million paperwork hours to the regulations, although the change in the U.S. presidency may mean that some or all of these do not come to pass.¹⁰ However, any further changes to the regulations will bring additional, albeit perhaps only temporary, costs of compliance as the requirements change.

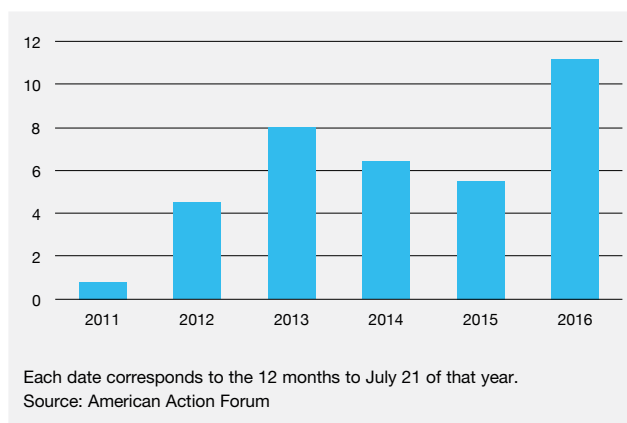


Figure 1 – Cost of Dodd–Frank to U.S. business (U.S.\$ bln)

3 Regulation (EU) No. 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (market abuse regulation) and repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC

4 Regulation (EU) No. 648/2012 of the European Parliament and of the Council of 4 July 2012 on OTC derivatives, central counterparties and trade repositories

5 <http://bit.ly/2ll2HuF>

6 <http://bit.ly/2eCt9h9>

7 <http://bit.ly/2mpKPDU>

8 <http://bit.ly/2ma1CdH>

9 <http://bloom.bg/2m4bBRr>

10 <http://bit.ly/2lvGCdK>

4. **Data breaches:** data breaches themselves can be extremely costly – according to the Ponemon Institute’s 2016 “cost of data breach study” (sponsored by IBM),¹¹ the average consolidated total cost of losing sensitive corporate or personal information was approximately U.S.\$4 million per breach in 2016 – up 29 percent since 2013. Consequently, it is understandable that spending on information security products and services globally will be approximately U.S.\$81.6 billion in 2016, which is an increase of 7.9 percent over 2015.^{12,13}

European Union regulations

In Europe, the European Commission (the E.U. Commission) has focused on overhauling the supervisory framework of the financial services sector. The European market infrastructure regulation (EMIR) aims to reduce the risks posed to the financial system by derivatives trades. It imposes obligations to report trades, clear trades, and take additional steps to mitigate the risks associated with OTC derivatives transactions. The Markets in Financial Instruments Directive (MiFID II) aims to improve the regulation of firms that provide services linked to financial instruments, and the venues where those financial instruments are traded, aiming to update the regulatory regime for investment services and activities in Europe.

MiFID II is set to come into full effect on January 3, 2018. It had originally been due to be implemented at the start of 2017 but was put back a year because of the time it was taking for the national regulators to build the necessary IT systems. In its press release relating to this delay, the European Commission stated that the European Securities and Markets Authority (ESMA) had to collect data from about 300 trading venues on about 15 million financial instruments. To achieve that result, ESMA needed to work closely with the competent national authorities and the trading venues themselves, but informed the E.U. Commission that neither the competent authorities, nor the market participants would be ready in time.¹⁴ MiFID II implementation is illustrative of the costs to businesses of preparing for compliance with legislation that continues to shift in scope and timing because of the complexity of the relevant markets and the technical challenges involved with compliance.

The enhanced European regulation regime has also imposed significant costs on industry. In 2014, the E.U. Commission estimated that one-off compliance costs would be between €512 million and €732 million, and ongoing costs between €312 million and €586 million per year.¹⁵

A recent report, however, forecasted that the top 40 global investment banks and the top 400 asset managers would spend

U.S.\$2.1 billion in 2017 in order to meet MiFID II guidelines.¹⁶ While many companies already comply with MiFID, the costs of compliance will only increase under MiFID II.

1. **EMIR:** EMIR includes similar requirements to Dodd-Frank, such as those relating to reporting and clearing. A member group that prepared data for the E.U. Commission’s impact assessment on EMIR, provided in December 2016, estimates the compliance costs for non-financial companies related to the ongoing reporting obligations to be up to €500,000 a year per company. Even for smaller companies the annual costs of reporting amount to €20,000 and upwards. The European Association of Corporate Treasurers has estimated that EMIR will cost non-financial companies between €50,000 and €200,000 to implement.¹⁷
2. **Market abuse:** another area of focus for regulators has been cracking down on market abuse and market manipulation. The E.U.’s Market Abuse Regulation (MAR) came into effect on July 3, 2016 and aims to increase the integrity of financial markets and reduce incidences of financial crime. The penalties for breach of MAR are significant; for example, in the U.K., the Financial Conduct Authority (FCA) can impose unlimited fines, order injunctions, or prohibit regulated firms or approved persons from participating in financial services. In addition, criminal sanctions for insider dealing and market manipulation can incur custodial sentences of up to seven years and unlimited fines.
3. **KYC and anti-money laundering (AML):** the E.U. adopted the fourth Anti-Money Laundering Directive in May 2015 as a means to combat money laundering and terrorism financing. The E.U. has explicitly raised the prospect of FinTech innovations, such as virtual currencies, being used to circumvent the traditional financial system and conceal illegal financial transactions carried out in an anonymous manner. This poses a risk to companies that create innovative financial products but do not have the technological capacity to meet their regulatory obligations. According to Thomson Reuters’ 2016 KYC survey, the average annual costs to banks of KYC compliance is U.S.\$60 million, with some banks spending as much as

11 Ponemon Institute, 2016, “2016 cost of data breach study: global analysis,” June

12 <http://bit.ly/2lvPvUD>

13 <http://gtnr.it/2b3s6l7>

14 <http://bit.ly/1V34Jfp>

15 <http://bit.ly/2lvHVtf>

16 <http://bit.ly/2d0JISO>

17 <http://bit.ly/2mHQ51S>

U.S.\$500 million on annual compliance.¹⁸ These costs are providing an opening for RegTech KYC solutions, whether they be blockchain, data aggregator, cloud, or artificial intelligence based technologies.

4. **Data protection:** another area of focus for regulators with obvious RegTech opportunities is data protection. As part of a comprehensive reform of data protection rules, the E.U. Data Protection Directive came into force on May 5, 2016 and will have to be transposed into national law by May 6, 2018 and the E.U. Data Protection Regulation¹⁹ came into force on May 24, 2016 and will apply from May 25, 2018, imposing various obligations on a range of firms including financial services companies, and stringent penalties for non-compliance. RegTech should be able to assist firms in areas such as identifying risks, auditing infrastructure, and reporting.

Global compliance

Companies that operate in the U.S., in Europe, and worldwide have the added burden of compliance with numerous regulatory regimes that frequently differ, often considerably. Maintaining policies, procedures, and systems to address different jurisdictional requirements places an increasing strain on global corporations' budgets. Due to the greater global regulatory demands placed upon managing risk, three-quarters of all firms surveyed by Thomson Reuters for its 2016 KYC survey expect the focus on managing regulatory risk and corresponding compliance costs to continue to rise, with 15% of companies expecting to spend "significantly more" on compliance in coming years.

Costs on government and on regulators

Regulatory requirements are, of course, not only expensive for companies. They are costly to government agencies responsible for oversight as well. IBM recently conducted a survey on potential implementation of blockchain technology in the public sector, polling 200 government executives from 16 different countries worldwide. 14% expect to utilize production-grade blockchains in 2017 and 48% anticipate launching some use of a blockchain between now and 2020. In its survey report, IBM stated: "For example, nine in ten government organizations plan to invest in blockchain for use in financial transaction management, asset management, contract management and regulatory compliance by 2018. And seven in ten government executives predict blockchain will significantly disrupt the area of contract management, which is often the intersection of the public and private sectors."²⁰

Following the release of a report in January 2016 by the U.K. Government Office for Science that called for the U.K.

government to experiment with distributed ledger technology within government operations, the U.K. Department for Work and Pensions launched an experimental blockchain system to distribute welfare payments partnered with Barclays, the U.K. arm of German energy firm RWE, FinTech startup GovCoin, and University College London.²¹

In the U.S., the State of Delaware is investigating using blockchain technology to store contracts and other essential corporate data on a distributed ledger. Delaware anticipates that this will allow companies and agencies to keep documents more secure in multiple locations and automate access by constituents, shareholders, and employees. Additional important benefits include lower costs and the capacity for longer document-retention, according to Caitlin Long, chairman and president of Symbiont. Symbiont provides smart contract and financial market distributed ledger technology and is working with Delaware on this project.²²

Potential cost savings for taxpayers are just as important a consideration for government as lower costs are for companies.

SPECIFIC AREAS OF REGULATION AND TECHNOLOGICAL SOLUTIONS

Reporting and recordkeeping

One of the main areas of focus of financial regulators worldwide is reporting of transactions, requiring companies' IT teams to dedicate an ever-increasing amount of resources to ensure regulatory compliance. This, though, is often not an option for smaller companies. There is, therefore, a significant demand for RegTech products that would make regulatory reporting easier and cheaper for market participants and allow them to meet changing regulatory requirements. This provides opportunities for RegTech companies entering this market like Cappitech, a privately held boutique that assists with full EMIR compliance by reviewing trade information, validating it, and then submitting it to the regulator.²³

18 <http://tmsnrts/2kRpQsD>

19 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regards to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC

20 Building trust in governments: study conducted by IBM Institute for Business Value (January 2017), <http://ibm.co/2kKYtMU>

21 <http://bit.ly/29p7oi4>

22 <http://bit.ly/2cEazX1>

23 <https://yhoo.it/2mHE4JK>

Regulatory reporting is also an issue beyond the U.S. and Europe. The Indian company Fintellix, for example, offers clients the ability to access global systems while allowing in-country teams to concentrate on local regulatory rules.²⁴

Another area of technological development is outsourcing recordkeeping to the “cloud.” As part of its Project Innovate, the U.K. FCA published guidance in July 2016 on outsourcing to the cloud and to other third-party services.²⁵ It recognized that the cloud can give firms greater flexibility in the services they receive, enabling innovation and reducing costs. Broadly speaking, it is supportive of the use of the cloud, provided that there is a clear business case for doing so, and that the outsourcing firm carries out sufficient due diligence of, and maintains oversight over, the service provider.

Blockchain solutions may also contribute to cost savings for companies. For example, rather than the traditional model in which firms collate and send data to the regulator to review, blockchain-based technology could provide regulators with almost instant access to transaction information. This could lead to significant cost savings by simplifying recordkeeping processes. Accenture recently published a report proposing that large investment banks could cut operational costs by as much as U.S.\$12 billion annually by implementing blockchains in their businesses. The report estimates that financial reporting expenses could fall by at least 70% and compliance costs by between 30% and 50%.²⁶

The potential of RegTech on regulatory reporting has also caught the attention of regulators. In November 2016, the U.K. FCA hosted a two-day “TechSprint” event aimed at unlocking regulatory reporting by finding collaborative solutions for the future.

RegTech is not just a means of making life easier for market participants, but can potentially also make government agencies’ own activities more efficient, by changing the way in which they receive and view data.

Monitoring and surveillance

Under MAR and other regulations, firms have an obligation to identify and reduce the risk of market abuse and report suspicious transactions to the relevant regulator; for example, MAR Article 16(2) places an obligation on anyone professionally arranging or executing transactions to establish and maintain effective arrangements, systems, and procedures to detect and report suspicious transactions.

A breach of MAR can lead to unlimited fines and the imposition

of restrictions on the company and individuals. The risks, therefore, are high.

Given the potential risks, financial firms employ large teams of compliance personnel to monitor the activities of their traders and other individuals. The use of technology based on artificial intelligence and behavioral software offers the potential to reduce the risks of market abuse, and reduce compliance costs.

While regulations prescribe the red flags that companies need to identify, they do not tell firms how to identify those red flags. Again, this provides opportunity for innovation. Sybnetix is a London-based company that offers market surveillance and compliance tools that works with companies to help them meet their regulatory obligations.²⁷ Ancoa, also based in London, is another example of a RegTech startup. It provides contextual surveillance and insightful analytics for exchanges, regulators, and buy- and sell-side firms. It can operate on the cloud, and can save smaller exchanges and firms significant deployment costs, allowing more competition among brokers and exchanges.²⁸

KYC

Innovators, such as virtual currency providers and online wallet providers, risk being subject to KYC regulations that would be onerous without technological solutions to aid compliance.²⁹ For example, London-based Onfido seeks to use machine learning technology to verify identities and carry out adverse history searches.³⁰

This is also an area where blockchains may improve compliance processes. For example, Singapore-based startup KYC-Chain aims to use distributed ledger technology as a basis for onboarding that allows sensitive information to be shared easily and securely.³¹

Cybersecurity

2015 and 2016 saw significant venture capital funding in the area of cybersecurity in the expectation that cybersecurity startups will only continue to grow. Regulators require more

24 <http://bit.ly/2m4aJvU>

25 <http://bit.ly/2czYQX5>

26 <http://reut.rs/2llyjNP>

27 <http://bit.ly/2ma2kaO>

28 <http://ancoa.com/>

29 <http://bit.ly/2l8x7oK>

30 <http://bit.ly/2l8HU28>

31 <http://bit.ly/2lv1bbL>

robust data storage, systems testing, and technical controls, resulting in smaller companies increasingly looking to out-source these duties and larger institutions trying to become more secure and efficient at the same time.

Startups are using big data and artificial intelligence to provide next-generation, platform-based solutions that may be responsive to the needs of both small and large companies. For example, public cloud-technology adoption is increasing, and is expected to begin to meaningfully reduce firewall spending by 2019.

Penalties high for non-compliance

By the end of 2015, U.S. banks had paid more than U.S.\$200 billion in fines since the financial crisis for non-compliance with regulatory requirements, and for problematic behavior.³²

There is no question that non-compliance with the various regulatory requirements can be very costly. The U.S. Commodity Futures Trading Commission (CFTC), for example, continues to prioritize the enforcement of reporting violations. In 2016, the CFTC issued more than twice the number of enforcement orders for reporting violations that it issued in the previous year. A number of these violations involved new reporting requirements under Dodd–Frank. For example, the Division of Enforcement recently fined two large banks U.S.\$560,000 and U.S.\$400,000, respectively, for violating the Swaps Large Trader Reporting Rule.³³

In another recent example, the CFTC fined two agribusinesses, a cooperative, and a processor/trader, jointly U.S.\$1 million³⁴ and Golden Agri U.S.\$150,000³⁵ for failing to file the appropriate reports. The CFTC also aggressively pursues sanctions against traders that fail to keep required records and/or file complete and accurate reports pursuant to its regulations, which has resulted in significant fines.

But it is not just the U.S. authorities who are taking a tough approach. In the U.K., the FCA also has imposed fines for reporting violations. Its heads of enforcement and financial crime have said that effective market surveillance is critical to maintaining the integrity of markets and depends on accurate and timely reporting of transactions. It has, therefore, taken enforcement action against firms failing to meet their obligations. The FCA has also recently carried out several high-profile enforcement actions against major global institutions for failing to reduce the risks associated with financial crime.

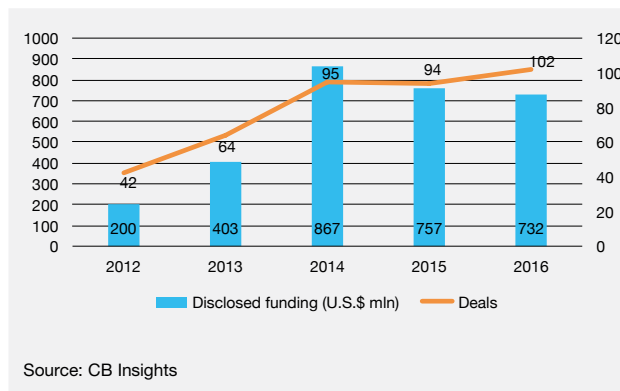


Figure 2 – RegTech annual global financing history

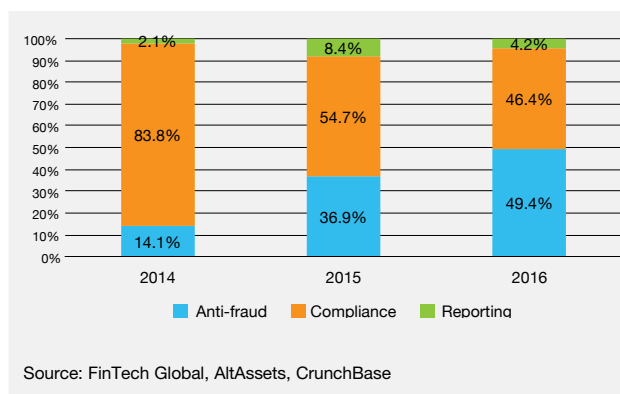


Figure 3 – RegTech investments by sub-sector, 2012-2016 (% of total amount invested)

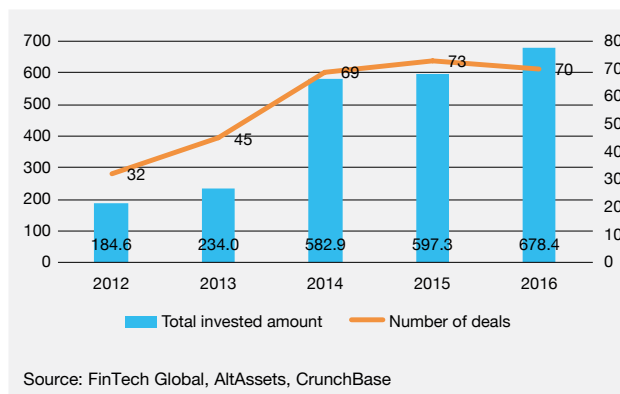


Figure 4 – RegTech investments, 2012-2016 (U.S.\$, number of deals)

32 <http://cnb.cx/1Q3HGsd>

33 <http://bit.ly/2mHCTtJ>

34 <http://bit.ly/1p8uaSe>

35 <http://bit.ly/2llg9i4>

REGTECH INVESTMENT GROWTH

Against this background, it is no surprise that more and more venture capitalists and private equity firms are looking to FinTech generally and to RegTech opportunities specifically. Since 2012, investments in RegTech have raised approximately U.S.\$2.3 billion in over 300 deals (Figure 2).³⁶ This activity has continued into 2017.

For example, Droit Financial Technologies, a New York-based firm specializing in trading compliance, raised U.S.\$16 million from investors including Goldman Sachs and Wells Fargo in its first fundraising in late 2016.³⁷ Droit aims to assist clients by analyzing whether or not they are up-to-date with regulatory regimes including Dodd–Frank and MiFID II.³⁸

In terms of startups that focus on monitoring, RedOwl analytics obtained U.S.\$17 million in funding in 2015.³⁹ It aims to capture data from multiple sources in order to detect and deter unwanted behavior that might otherwise result in regulatory enforcement action.

Significant deals in the KYC space include the British firm Onfido raising U.S.\$25 million in Series B funding,⁴⁰ and ComplyAdvantage, a RegTech startup that has built a proprietary database on AML risk, covering sanctions and watchlists, politically exposed persons, and adverse media, received U.S.\$8.2 million in Series A funding.⁴¹

Another interesting case study is Fundapps, a London-based RegTech company that assists clients with solutions in relation to investments restrictions and shareholding disclosure. Founded in 2010, it monitors U.S.\$2 trillion in client assets daily and has become profitable without any outside investment so far.⁴²

A provocative development in the artificial intelligence space has been the acquisition by IBM of Promontory Financial, a risk management and regulatory compliance consultancy, in November 2016. IBM's intention is to use Promontory's know-how to train IBM's AI computer system, Watson, to help firms meet a range of regulatory and compliance obligations, from financial risk modeling to AML and KYC.⁴³

Despite these examples, an analysis of investment in the sector has shown that the total investment in RegTech is still relatively small as compared to the total amounts spent on regulatory compliance in financial services, suggesting that there is a lot of room for increased investment in the RegTech sector in the coming years.⁴⁴ Prior to any capital raise, RegTech

startups should consult their counsel to ensure applicability of their technology to current, proposed or potential regulation to ensure the most effective growth.

LOOKING AHEAD

There is no doubt that the events of 2016, specifically Brexit and the U.S. election, have brought uncertainty to the future shape of U.S., U.K., and E.U. regulatory regimes, in particular. The new U.S. administration has stated the goal of reducing regulatory burdens on businesses, but it is unclear where and how that will be accomplished. The U.K. may be crafting regulations separate from the known and expected E.U. requirements. Navigating the morass of regulations in multiple jurisdictions often requires involving global legal and government experts who also are familiar with the company's subject matter and with potential technological compliance solutions.

Political events and technological developments likely will spur further regulatory changes worldwide. However, instead of relief, many commercial players are worried these changes will bring even more costs, on top of the existing investment in compliance. Companies should actively engage regulators in discussions regarding any potential regime changes to ensure their needs, and anticipated costs, are considered during the rule-making processes. Companies can then guide regulators to consider technological advances and potential alternative compliance methods while crafting new rules, as RegTech companies and consumer can provide a unique perspective to government agencies.

It remains essential that market participants have available to them convenient, cost-efficient regulatory compliance options in order to continue to compete in a global and fragmented regulatory environment. We, therefore, expect the significant investment in, and demand for, RegTech solutions to continue in 2017 and beyond.

36 <http://bit.ly/2f3xwEx>

37 <http://on.ft.com/2l15JPO>

38 <http://bit.ly/2f66aOl>

39 <http://bit.ly/2lvUcOf>

40 <http://tcrn.ch/29a5JSq>

41 <http://bit.ly/2m9WwOs>

42 <http://bit.ly/2m4tM9J>

43 <http://ibm.co/2m9WwCO>

44 <https://medium.com/@janmaartenmulder/regtech-is-real-and-120-startups-to-prove-it-6b396d94dd8c#.6jxf8shkr>

From “Blockchain Hype” to a Real Business Case for Financial Markets

Massimo Morini – Head of Interest Rate and Credit Models, IMI Bank - Intesa S.P. Group, and Professor of Fixed Income - Advanced Methods, Bocconi University¹

Abstract

There has been a huge amount of coverage in the press about the great potential uses of bitcoin-related technology for financial markets, such as improvements in efficiency. In addition to the supporters of blockchain, many have been critical of its real-life applications within the business world and suggest that what we are witnessing is nothing short of “blockchain hype,” and that this technology can only be applied to bitcoins. This paper will demonstrate that there are real business cases for improving financial markets based on the lessons learned from cryptocurrencies, but, unlike what the hype-enthusiasts suggest, they are not application of a technology to the existing business models within financial markets. They are reforms of the business model itself. What needs to be exported from the world of cryptocurrencies are aspects of the market organization, inspiration for a different accounting and legal system, and some aspects of the technology. These can result in a huge contribution towards more robust, efficient, and stable markets. However, the process cannot be immediate and effortless, and can only be achieved within a market-wide

strategic perspective. In this paper, I develop these concepts initially within a parallel analysis of cryptocurrencies and financial markets. Then, I will focus on a specific business case regarding the collateralization of financial derivatives, which will highlight quantifiable benefits in terms of reducing costs, capital, and risk. It is an example of a situation where the use of cryptocurrency technology is not more important than the business ideas developed in the analysis of cryptocurrencies; yet it was inconceivable prior to the advent of distributed ledgers, smart contracts, and oracles.

¹ Fruitful conversation with Robert Sams, Giacomo Zucco, and Alex Lipton is gratefully acknowledged. The second part of this paper is just an extension of “Smart derivatives can cure XVA headaches,” by Massimo Morini and Robert Sams, Risk Magazine, August 2015. I also thank all those – too many to be mentioned by name – that asked me the questions that form the backbone of this paper. This work expresses the views of its author and does not represent the opinion of his employers, who are not responsible for any use which may be made of its contents.

INTRODUCTION: BLOCKCHAIN HYPE VERSUS BLOCKCHAIN SECLUSION?

There has been a huge amount of coverage in the press about the great potential uses of bitcoin-related technology for financial markets, such as improvements in efficiency. In addition to the supporters of blockchain, many have been critical of its real-life applications within the business world and suggest that what we are witnessing is nothing short of “blockchain hype,” and that this technology can only be applied to bitcoins.

This paper will demonstrate that there are real business cases for improving financial markets based on the lessons learned from cryptocurrencies, but, unlike what the hype-enthusiasts suggest, they are not application of a technology to the existing business models within financial markets. They are reforms of the business model itself. What needs to be exported from the world of cryptocurrencies are aspects of the market organization, inspiration for a different accounting and legal system, and some aspects of the technology. These can result in a huge contribution towards more robust, efficient, and stable markets. However, the process cannot be immediate and effortless, and can only be achieved within a market-wide strategic perspective.

One crucial misunderstanding here is the idea that blockchain technology can be exported to financial markets “as they are” to make them more efficient. This is meaningless, since blockchain technology was created to change some trust-based business processes to make them less reliant on trust. Without structural changes, the best of blockchain technology is lost and we are left with the inefficiencies.

It should be added that suggesting that blockchain technology cannot be used outside of the bitcoin world is also incorrect. Bitcoin was created to create a level of independence from trust sufficient to allow players to be anonymous and do so without any legal protection. Other business solutions based on a level of trust intermediate between bitcoin and traditional finance can use similar technology and yet be very different from bitcoins. But we must ready to use the concept of trust in a totally different manner, as a way to analyze the different parts of a business process and the reasons for its current inefficiencies and risks.

In the next section, I will develop these concepts initially within a parallel analysis of cryptocurrencies and financial markets. Then I will focus on a specific business case regarding the collateralization of financial derivatives, which will highlight quantifiable benefits in terms of reducing costs, capital, and

risk. It is an example of a situation where the use of cryptocurrency technology is not more important than the business ideas developed in the analysis of cryptocurrencies; yet it was inconceivable prior to the advent of distributed ledgers, smart contracts, and oracles. In fact, it was first presented in Morini and Sams (2015), in an introduction to the blockchain innovation for the derivatives world.

THE MISUNDERSTANDING ABOUT “TRUST”

Notice that the term “trust” is often used in the bitcoin debate in a radical way, moving from a totally trustless anarchist model to a cooperative model based on absolute trust. None of them really exists. Even bitcoin has created peculiar elements of trust in some players, like a stable group of core developers or miners. And, financial markets have never been based on absolute trust in counterparties or central bodies. The radicalism of the debate has hidden the fact that different business models are associated with different levels of trust. Trust can be hidden in many passages in the working of a market, and can be eliminated or reduced in some without disappearing from others. More than a generic term for ideological debate, trust can be used as a precise concept to understand the features of a business model, and how that model can be positively reformed; without forgetting that any removal of trust creates some form of disintermediation, of some institutions or of some functions within institutions, and in this way it requires changes to the business model, and often to the legal, regulatory, and accounting frameworks.

An example of an unnecessary element of trust is the reliance on the agreement between two counterparties about the exact representation of a deal without any automation enforcing this agreement, not even in critical cases. Many markets are still crippled by this feature. This can be addressed with elements of distributed automation similar to those seen in cryptocurrencies.

WHAT ARE THE PROBLEMS OF FINANCIAL MARKETS THAT WE WANT TO SOLVE?

Financial market transactions are still based on the logic of “consensus-by-reconciliation.” Every player gives its own representation of a transaction in its own accounting systems (ledger) and its own IT systems. The only proof that this representation is correct is coincidence with the representation

given by the counterparties. The confidence in the legal validity of the contract in all its aspects is crucially dependent on trust in this coincidence, which in practice needs to be verified more than once. This verification requires a number of steps, such as confirmation, affirmation, communication to central bodies, and other reconciliation passages along clearing and settlement.

This is an objective bottle-neck towards more efficient and reliable markets. Current reconciliation steps slow the process down, even if the technology enables very fast communication. They also drive up costs.

Furthermore, the need for this kind of reconciliation leaves open the risk of disagreement and litigation, making the process uncertain and increasing the capital requirements for members. It is a system intrinsically inefficient that has not been seriously reformed in decades, for lack of incentives and no visibility of a technological and organizational stack suitable for a change. Even if many bits of the fundamental technology to solve it were already available in the past decades, just think of the internet giving a shared information platform, this had never been applied to changing the foundations of some transactions. Now there is visibility of a different business model in the cryptocurrency example, together with a full technology package enabling it.

WHAT EXACTLY ARE THE SOLUTIONS THAT MAY COME OUT OF THE BITCOIN EXPERIENCE?

Bitcoin and the other experiments of cryptocurrencies or crypto-transactions are based on a single accounting and reporting system, a distributed ledger (DL). With a DL, the reconciliation bottleneck is avoided since there is a consensus algorithm that verifies transactions and gives to them a unique representation on the ledger, collapsing all reconciliation steps into a single initial passage. Further reconciliation steps are unnecessary when there is a single authoritative deal representation for all the parties. It is this business model that makes transactions so fast for bitcoins, more generally than any specific piece of technology. This insight is useful for the use of ledgers in financial markets too, even if financial players may need DLs that are different from the blockchain, which is a peculiar implementation where all transactions are reported together, visible to all, and their time-order is defined through a sequence of blocks.

For advanced financial markets, distributed consensus can be

applied also to a deal made up of many payments, like a derivative or a bond, through the concept of a Smart Contract, which is a piece of program code, in a given computer language, executing the transaction agreed at inception between the parties. This guarantees the enforcement of consensus, namely that the deal will follow the agreement taken at inception between the parties. Bitcoin has only basic smart contracts, but other cryptocurrencies like Ethereum have smart contracts written in a Turing-complete language, which means it can do everything that a normal computer does.

Notice that this is a further step towards a different and more advanced model of the market. Not only do the accounting/reporting of the transaction move from individual representation to an authoritative distributed representation, but also the contract stops being two pieces of papers to be implemented and represented in separate ways but becomes a unique manager of the transaction signed (cryptographically) by the interested parties. Financial contracts are already translated by parties into software running on IT systems. What was missing were working examples of a technology where the piece of code could become the contract itself and not one of the many representations of it given by the parties. When the unique smart contract signed by the parties manages directly the flow of the transaction, there is a further reduction of delays and risks of disagreements and misalignments.

BUT ALL THESE GOALS CAN BE OBTAINED JUST VIA A CENTRAL DATABASE AND COMPUTING GRID ON ONE SERVER

For many of the above goals, the answer is: of course. But a computer/database shared among all the players of a market is a centralized solution, with all the well-known limits of centralization. These limits are a central topic in the state machine replication approach: centralized systems are usually more efficient from a technological point of view, but they are not fault-tolerant. In abstract terms, this means that failure of the central server is failure of the entire system. In economic terms, this unpleasant fact has additional consequences. In case of centralization, there will be an administrator of the database/hardware, and this institution would bear a great operational risk, the risk of the entire network, thus demanding an equally great power on controlling and unilaterally changing the rules. Centralized solutions create monopolies that drive the business costs up because the monopolist does not have the right incentives to contain them. Additionally, in finance centralized solutions also generate a concentration of financial

risk that drives up – correctly – both the regulatory burden and the amount of risk-management provisions, such as collateral.

A centralized database also raises the likelihood of legal disputes; it would be easy to accuse the administrator of tampering with the ledger. Since the ledger must report the situation of everyone and yet belong to no one, a DL appears a more natural solution. It avoids the need for a central body and also reduces the legal uncertainties. The ledger downloaded by one party is the official ledger as much as the version downloaded by someone else. They are all replications of the ledger, there is not one central database and many duplications, which is a lays the groundwork for uncertainty, reconciliation delays, and legal disputes.

AND WHAT IF THE DATABASE WAS “FULLY REPLICATED AND DISTRIBUTED”?

The technology of “distributed services” (DS) (and the state machine replication approach) that developed in the last decades are certainly a crucial part of the solution. There exist database technologies that try to keep away from the risks of centralization and predate DL technology. One can find works on fully replicated distributed databases that date back to the early 1990s, such as the evolution of the technology that has helped bring about well-known distributed solutions like DVCS (distributed version control systems), of which Linus Torvalds’ Git is the best known.

The bitcoin blockchain evolved in the same stream of technological advances, partially based on the same cryptographic solutions. It is a relevant example of radical economic application of this form of technology, and in this way it showed how this technology applied logically to a market brings about a fundamental change in market organization.

Bitcoin found a decentralized solution for chronological tracking and time-stamping that was suitable for its peculiar context of building a market from scratch based on pseudonymous players. Even if this solution cannot be exported rigidly to different contexts, like current and foreseeable financial markets, blockchain is the natural turning point of distributed technology to take inspiration from when building DLs for financial markets, without ideological distinctions between distributed ledgers with blocks and proof-of-work, and distributed ledgers that may be different in these respects. An additional reason to keep more than an eye on blockchain in evolving existing financial markets is to keep a standard compatible with other

DL solutions that have different privacy and validation requirements, cryptocurrencies included.

WOULD THE CURRENT TECHNOLOGY FOR DLS BE READY TO PROVIDE THIS?

No. First of all, there is a scalability issue. The logic of distributed consensus across the entire network limits the amount of transactions that can be managed in a block. Solutions can exist for financial markets, but they are not tested yet.

Furthermore, the most tested market, bitcoin, has only basic smart contracts. Large-scale application of smart contracts is exactly the test that DLs for financial markets need to perform.

Finally, neither bitcoin nor other solutions like Ethereum have a focus on privacy and identity, as needed for financial markets. Identity is an unavoidable issue for any legal recognition. Privacy is a concept that is evolving in financial markets, with regulators demanding increasing transparency. These privacy challenges might be solvable with solutions such as complex data-encryption, interlinked bilateral ledgers, or regulated exploiting of pseudonymity. However, it should be noted that these are all elements that prove that the process will take time.

SO FAR ONLY THE LEDGER, THE BLOCKCHAIN, IS USED FROM THE BITCOIN STACK. WHAT ELSE IS USEFUL?

In bitcoin there is also a fundamental use of cryptographic techniques, such as asymmetric cryptography and hashing, both for ledger management and inside the incentive/selection method called proof-of-work. Asymmetric public-private key cryptography is important also for extension to financial markets, as it is already in many fields. This form of cryptography can be used to eliminate a level of intermediation, for example bitcoin use it to disintermediate the role of banks as providers of cash deposits. In financial markets, the main players, including banks, have a different role as structurers, traders, issuers of deals and securities, lenders, and managers of credit and market risks. There is less fear of cryptographic disintermediation here, since the layers that can be eliminated or disintermediated in financial markets with no loss of security and a gain of efficiency and transparency are mostly not banks or their business counterparties.

Furthermore, cryptography may enable both identity and privacy at the same time. Other applications of cryptography are emerging now. For example, the use of cryptography by Oraclize² is interesting. They apply cryptography in order to provide a cryptographic guarantee that an operation has been executed. It is a way to enforce a contract with a computer or website and being guaranteed the contract has been executed exactly. That is another bit of technology, related to the concept of Oracles, that was developed around blockchain even if it is not part of it. This can be used, for example, to secure the process of importing data from outside the DL for internal use, something very common in financial markets. Excessive reliance on trust here would create a single point of failure outside the control of those that have a stake on the ledger. A similar logic is also behind the Intel Software Guard Extensions,³ with the additional feature that the logic is embedded in the hardware itself and not only in the software.

IS PROOF-OF-WORK TO BE EXPORTED TO FINANCIAL MARKETS?

Proof-of-work, as we see it in bitcoins, may not be applicable to financial market because it is designed to solve a specific problem: finding a way to make players update the blockchain in an honest way even if they are not forced to either by a reputation incentive (because they can be anonymous) or by any legal framework. This is a very extreme concept of disintermediation and lack of trust that does not apply in a context where players are not anonymous and where fraud is legally prosecuted. This is an important reason why proof-of-work might not be used in advanced financial markets: the core motivation for its use is missing.

But there is another reason why it makes sense for us to look at proof-of-work in more detail. The clever idea behind the mechanism is that it requires the participants that wish to receive remuneration for updating the blockchain (miners) to solve a complex computational problem. This forces them to make an off-ledger investment in energy and computational power that makes it uneconomical to fraud the system. In fact, double spending is the only fraud that miners could implement easily in bitcoins, since asymmetric cryptography and the public ledger protects, in its own peculiar way, past transactions and possessions. The loss of credibility of the network coming from a fraud would be, for those who have made the off-ledger investment in energy and computational power, greater than the easy gain from double spending. It is important here to understand a practical point not enough stated in theoretical

analyses: that the investment in computational power is dominant over the investment in energy, and that the former is more relevant also because it is a long-lasting one. Mining technology is very expensive and difficult to reuse for other purposes. This is crucial in helping explain why it makes frauds uneconomical, and also why alternatives like proof-of-stake did not work: they did not guarantee an off-chain, long-lasting, capital investment.

Retruning to today's financial markets, it should be stated that while proof-of-work is not a waste of resources in bitcoin, since it is the only off-chain long-lasting investment of the crucial players, in financial markets it would be a real waste since the existence of off-chain economic commitment for crucial players is already proved; they already have a strong incentive to maintain the credibility of the whole financial system. This state of affairs may not last for ever, but it is the reality we start from.

SO, SHOULD TRANSACTION VISIBILITY AND VALIDATION BE LEFT TO THE COUNTERPARTIES ONLY?

In principle, a basic extension of the current reality is a consensus algorithm where only the two parties involved sign the smart contract and validate the transaction, potentially on a private DL. This is already an improvement in terms of efficiency and finality of financial markets, removing some of the need for reconciliation and the risk of litigation. Considering also that currently visibility and some aspects of validation of a trade would involve a number of regulators, introducing a role for reliable third parties, this could be sufficient for most goals of the practical business case described.

Yet, it would be shortsighted to depart from the cryptocurrency experience to such an extent as to use a bilateral solution. The business case described below would work easily, from a technological point of view, in a multilateral public blockchain. From a financial point of view, a multilateral setting would pose some issues, but there are also many services benefiting from multilateral reliable and efficient distributed transaction validation and recording. For example, in some extensions of the business case described below, collateral may be provided or guaranteed by a third party. In this case, consensus, speed,

² <http://bit.ly/2mdCJ0M>

³ <http://intel.ly/2k9mUYr>

and transparency, like those allowed by a multiplayer ledger, are particularly useful. Other examples are the use of the techniques for compression of exposures that are possible on a network, see, for example, TriOptima,⁴ or the possibility to give regulators a broader and deeper vision of the market.

This can lead to a range of possible consensus algorithms, not excluding something more similar to what we have seen for cryptocurrencies. In fact, we can add that the validation algorithm used in the bitcoin world is mainly required to avoid double spending. This possibility might seem to be of marginal importance in financial markets with trusted, or at least known, members. But, in fact, the same economic problems take on a different appearance. Double spending, or spending of non-existing resources, also takes place in financial markets too and is considered the main risk by market participants and regulators: except that its called default.

Default risk is where we clearly see that trust is not unlimited in financial markets. Financial markets are made up of trustworthy parties, but none is completely riskless, since even the largest institutions/governments can default, as history shows.⁵ And, default is by definition a form of double spending, where a counterparty undertakes commitments that are greater than the funds they actually have available. Thus, thinking of methods for assessing fund availability within the network, beyond pure unilateral confirmation, is relevant, and this may support the case for more advanced validation, involving regulators or custodians or some other players not directly related to the transaction.

But a similar business case is still out-of-sight. Still difficult, but nearer, are solutions to some credit-related issues that can be solved using some aspects of the DL model that we discussed above, and that are robust enough for different choices about consensus and ledger visibility.

OK. CAN WE MOVE FROM GENERAL PROBLEMS TO A SPECIFIC ONE?

It is about time. Simply saying that a business reform might eliminate reconciliation or make settlement faster is not enough. There can be cases in which reconciliation steps are a real business need or faster settlement is prevented by regulators for good reasons. One also needs to demonstrate business cases for when these sorts of worries are outweighed by the risk and cost savings coming from less reconciliation and faster settlement.

As seen in Morini and Sams (2015)⁶ and later in AssiomForex,⁷ a relevant case concerns collateral and default management in the derivatives market, a market 7-8 times the GDP of the world in notional terms, and as large as the U.S. GDP or the global bond market in terms of value.

Credit risk is a central issue for over-the-counter (not listed) derivatives, which are the dominant part of the market. The issue reached dramatic levels after the financial crisis started in 2008. The Lehman Brothers' default marked a crucial change in the derivatives markets. From an aggressive market with high leverage, little attention to risk, and a disordered multiplication of complex payoffs, we moved to a market with strong standardization, heavy regulations, and potentially excessive attention to risks. This has made the financial world a safer place from the perspectives of many, but certain negative side effects are also becoming clear. Firstly, derivatives users, such as funds and corporates, are increasingly unhappy with a market in which prices do not express the intrinsic market risk of a financial product (interest rate risk, commodity risk, etc.), but are skewed by charges that are more or less related to default risk.

This includes credit valuation adjustment (CVA), a valuation adjustment made by financial dealers for the risk of default by banks' counterparties, an adjustment called FVA (funding value adjustment), which accounts for the funding cost of banks, which increased when the banks' default risk increased, and KVA (capital value adjustment), an adjustment for the amount of extra capital that banks need to hold to account for the increased default risk. Additional costs to users of derivatives come from the recent increase of the margin requirements for market players (these are part of funding costs and generate another value adjustment, the MVA (margin value adjustment), which is also in response to increased default risk).

Buy-side clients still need financial markets and derivatives for their investment and diversification needs, as well as for hedging their costs and risks, in terms of cashflows and from an accounting perspective. For these clients, the above transformation meant a sharp increase in costs.

4 <http://bit.ly/2ldvcPE>

5 Defaults by the likes of Latin American governments, Enron and Parmalat, and partially Lehman, illustrate this fact clearly.

6 Morini, M., and R. Sams, 2015, "Smart' derivatives can cure XVA headaches," Risk Magazine, August 27, <http://bit.ly/2mf2XjB>

7 <http://bit.ly/2mVzJU2>

IN THE CURRENT MARKET SITUATION, WHAT IS USED TO REDUCE CREDIT RISK? COLLATERAL?

The mainstream approach for reducing the size of these charges is to mitigate the losses in case of default, of banks or their counterparties, through collateralization.

Collateral for derivatives comes in two forms. First, we have the variation margin, which requires that the derivative is revaluated every day by party A using its pricing model f^A . It takes as input the current value M_t^A of the relevant market variables from the info provider chosen by A, and gives current derivative value $V_t^A = f^A(M_t^A)$

If V_t^A is negative for A, which means that A expects to pay to party B in the transaction more than what it will receive, so that A is a net debtor, A will ensure that an amount in cash (and sometimes using other assets, bonds or equities, with a haircut rule) equal to V_t^A is available for counterparty B in a collateral account. Party B does the same thing but with its model f^B and its data M_t^B . Hopefully $V_t^A \approx -V_t^B$ and the process proceeds smoothly. When there is a remarkable difference between V_t^A and $-V_t^B$, the two counterparties talk to each other for a reconciliation. In some cases, it is the net creditor that makes a margin call, but this does not change the general picture.

In many cases, particularly when a party is a non-financial corporate that has difficulties in moving cash quickly or computing quickly the right amount of collateral, this process of collateral update happens less often than daily. It may be that longer period is stated in the agreements, or that the agreements accept explicitly to leave a part of the exposure not collateralized (via defining thresholds or a minimum transfer amount). These are inferior variation margin agreements that contrast with the top-class agreements between banks, characterized by daily updates, no minimum transfer amounts, and zero threshold.

IS THIS ALL WE NEED AS MARGIN? OR ADDITIONAL MARGIN, SUCH AS THE INITIAL MARGIN, CAN BE USEFUL?

Even in case of variation margin, there is always an expected delay between the last collateral update and the closeout for liquidation of a defaulting counterparty’s positions, leaving risk of default still open. This delay is called “margin period of risk” (MPOR), and comes from combining the collateral frequency with the delay between default time and the computation of a closeout amount. The total delay is estimated to be rather large by regulators since, when a default happens, there is no

guarantee that the valuation of the residual derivatives, V_τ^A and $-V_\tau^B$, with τ being the default time, coincide for the two parties. The current process assumes disagreement and potential litigation, and a reconciliation procedure driven by the liquidators that involve asking various third parties to give a valuation of the residual deal before arriving at a closeout amount. This pushes MPOR to range from 5 to 40 days.

Thus, on top of variation margin, there can be an additional amount of collateral called initial margin to cover the risks due to the length of the MPOR. In an initial margin agreement, counterparties use their risk models to make a conservative estimate (worst case scenario or Value at Risk computation) of the difference between the amount of collateral available at the beginning of the MPOR (last collateral update) and the actual default closeout amount computed at the end of MPOR (closeout day). This computation needs to take into account the unpredictability of market movements along the MPOR and the uncertainties concerning how the closeout amount will be computed. Under a long MPOR, initial margin can be very high.

IS THIS SOLUTION FULLY SATISFACTORY TO MARKET PLAYERS?

It has some very relevant limits.

1. First of all, collateral management is not, in the current market, so easy for non-financial players. Computing, finding, and moving the necessary margin liquidity can be an obstacle even for agreeing on the valuation margin procedure.
2. Secondly, even a top-class variation margin procedure is tampered by uncertainty on the different valuation models, market data, computations, and accounting representations from the two parties, an uncertainty that can create misalignments and make the process never faster than daily.
3. Thirdly, the margin period of risk is very long. Combining collateral frequency and the period for the agreement on closeout can take as long as 10 days. It is a delay sufficiently long to result in high levels of credit risk and capital costs (KVA), even in the presence of VM.
4. Finally, initial margin on top of the variation margin can reduce these costs dramatically, but only at the cost of a fourth problem: setting up a conservative initial margin agreement is expensive. Initial margin stays in a secluded account and due to its size, which in turn depends on the length of the MPOR, drains a large amount of liquidity from institutions.

ANY ADDITIONAL SOLUTION IN PLACE TOWARDS A REDUCTION OF CREDIT RISK? WHAT ABOUT CCPs?

One solution is trading through central counterparties (CCPs), which can reduce credit risk through trade compression. Consider a situation where bank A owes 100 to bank B, bank B owes 100 to C, C owes 100 to A. If all the players trade through a central counterparty, the three above payments cancel out with each other, reducing settlement and credit risk.

CCPs do something even more important beyond compression. By pooling risks together, they reduce the size of potential losses through the netting effects. When a bank defaults, its obligations towards a counterparty are usually netted with those of the counterparty towards the bank. This reduces the closeout amount to be paid, thus reducing potential losses. When there is a unique counterparty like the CCP, this netting effect is stronger.

CCPs, however, are an intrinsically centralized solution. Centralization has the advantages just mentioned, but also symmetric disadvantages like creating a central institution whose default, however unlikely, would spread losses to the entire market at an unprecedented speed and scale^{8,9}. This also means that the regulatory burden is particularly high on such institutions, also increasing collateral cost and demand, since a CCP is such a single point of failure that it needs to be massively overcollateralized.

We also have to remember that a centralized body lacks some of the competitive pressures to optimize collateral costs to members. Excessive collateral demand does worry regulators, since it can strain the market's liquidity conditions.

Finally, CCPs, as a natural corollary to this business, decide unilaterally the rules for variation and initial margins. The rules are also changed unilaterally quite often, particularly for Initial Margin.

CAN WE THINK OF AN ALTERNATIVE OR COMPLEMENTARY SOLUTION WITHOUT THE COSTS OF CENTRALIZATION?

This is where DLs come into play, but they can be useful only if we are eager to take from cryptocurrencies not only some of the technology but also inspiration on how the process can be designed, making a change that needs to be technological, regulatory, legal, and organizational. DLs can be designed to

be an independent solution, or a solution to be adopted by a body like a CCP, as long as the CCP is eager, while reducing some of the possible shortcomings seen above. In order to achieve that, it needs to change its own business model and decentralize some of the actions and decisions that now are centralized, while remaining the facilitator of the smooth working of the market, and a possible counterparty of last resort in times of crisis.

On the technology side, smart contracts suitable for derivatives can be implemented within a DL system if the consensus algorithm contains what is known as a Turing-complete state-transition function – for example, it must support if-then-else-branching, enabling the conditional features of a derivative to be executed.

A smart contract transaction might, for example, instruct the network to transfer: $\max(S_{1Y} - X, 0)$ from account A to account B a year from now, where S_{1Y} is the price of a given security one year later, provided a certain sum – the value of the contract – is transferred from account B to account A of the distributed ledger now. This is a sketch of the implementation for a cash-settled call option.

Once knowledge of S_{1Y} is provided in real time to the smart contract through an oracle managing access to trusted data providers, the contract can take care of the terminal settlement, transferring the right amount of money automatically. The smart contract can be much more detailed than the simple example provided above, incorporating more complex contractual features such as breakups, American exercise, legal requirements, and International Swaps and Derivatives Association standards. And the smart contract can take care of collateral regulation.

8 Gregory, J., 2014, Central counterparties: mandatory clearing and bilateral margin requirements for OTC derivatives, John Wiley & Sons, Ltd

9 Duffie, D., and H. Zhu, 2011, "Does a central clearing counterparty reduce counterparty risk?," *Review of Asset Pricing Studies* 1:1, 74-95

WHAT ARE THE REAL SAVINGS OF SMART DERIVATIVE CONTRACTS ON A DL?

The main savings are seen when we consider collateral. The smart contract can include the implementation of a model that computes the amount of collateral to exchange, as a subroutine or from an external source communicating with the network through a precise, cryptographically signed agreement with the contract itself (an Oracle in the extended sense of Oraclize).

After the above reasoning, there is not much to say about which changes are brought about by a DL logic coupled with a smart contract that uses a single cryptographically secure implementation of a model.

All the uncertainties we have seen before are eliminated. There can be no differences because of the model, the data, the computation, or the accounting rules. The agreement was taken not on a generic paper contract, but on a single smart contract managing the quantification of the payments through a single model implementation, and recording the exchanges on a single ledger.

Since precise rules for collateral payments have been agreed and validated from the start on a DL, and then managed by a digital contract, the need for reconciliations and the risk of litigation are minimized. This reduces credit risk in two ways. Firstly, collateral can match exposures much more precisely than currently. Secondly, slashing the time required for reconciliation means that much faster collateral update becomes possible. Collateral exchange frequency can be reduced from the current 1-day delay to a fraction of an hour. This, eventually, makes collateral a real-time guarantee and eliminates problem 2) seen above.

Additionally, in an environment where transactions are naturally automated, and collateral is quantified and managed by the smart contract, problem 1) can also be reduced.

BUT THIS MEANS COUNTERPARTIES SHOULD AGREE ON A VALUATION MODEL, MOVING FROM f^A, f^B TO f^C

Extending the range of what is contractually agreed and validated at the beginning, reducing the scope for trust and future reconciliation, is a core point of this possible business evolution. It is the price to pay for efficiency, risk-reduction, and cost saving. But this specific price may not be seen as too high these days.

First of all, banks are already accepting, and in some cases they are even seeking, more consensus about models. Before the crisis, private valuation models were regularly used for complex payoffs, and valuation differences could be seen as drivers of value, as much as of risk. Today, the stress in regulations, margins, and credit risk has changed the picture, making risk the dominant effect, and valuation differences have already been minimized in many contexts. In the post-crisis years, regulations regarding CCPs have already led the market to accept external standardized valuations for margin purposes (initial and variation margin), and ISDA/IOSCO have led the market to agree on a common model for part of the margin (the initial margin), even for non-cleared products. However, this trend towards sharing calculation logic is not only regulations-led, and goes beyond the margins issue. Services, such as Markit Totem,¹⁰ are used by banks to also indirectly reach a general consensus on the pricing logic of complex, non-cleared products, that have gained importance in recent years.

Secondly, what would happen on a DL is much less restrictive than the model standardization banks are already accepting. This is because it is in principle a bilateral agreement between the same parties that have just agreed on a price (valuation model at time zero) and on a future collateral exchange (valuation model in future times); they do not need to accept the one-size-fits-all model of current CCPs or regulatory intervention.

Finally, we get the majority of benefits even if the counterparties just agree on a valuation model for all those cases where valuation becomes a payment in the contract, like collateral regulation (and potentially anticipated closeout, a topic addressed below). If the ledger is used just as a transaction report but is not the only accounting report, players can be left free to keep private models for valuation in their own accounting systems, while binding them to smart contract agreement when valuation is used to quantify payments with the original counterparty. Misalignments between the private accounting valuation model and the agreed collateral valuation model already exist in dealing with CCPs.

¹⁰ <http://bit.ly/2ldFYp4>

CAN WE DO SOMETHING MORE TO REDUCE CREDIT RISK IN COLLATERALIZED DERIVATIVES?

Faster and precise collateral would already reduce risks and associated costs, however, the full benefit would derive from extending the agreement on the revaluation model from collateral to default closeout. A mutually agreed valuation model can change the closeout process completely – reducing the MPOR to a few hours.

With collateral on a ledger, a missed collateral update (a default warning) is detected in real time, and can trigger the smart contract to terminate itself and provide immediately closeout valuation based on the agreed model. Suppose party B misses a collateral payment at t . We can agree on the smart contract that there is a grace period δ , say $\delta = \text{few hours}$, during which the contract waits for a collateral payment from B. At the end, the difference between deal value and collateral from A’s point of view is computed with the agreed model $\Delta = f(M_{t+\delta}) - f(M_{t-\epsilon})$, where ϵ is the interval between two collateral payments, so that $t - \epsilon$ is (conservatively) the time of the last collateral update. Party A becomes owner of the collateral, already in his/her possession, with only Δ to be paid by A to B if $\Delta < 0$ to A. Party B is left with the residual obligation, apart from Δ to be paid by B to A if $\Delta > 0$ to A.

In this way, the closeout amount is promptly computed in the network using the agreed bilateral model $f(M_{t+\delta})$, and the margin period of risk is slashed to the short $\epsilon + \delta$ time, that can realistically be few hours, an order of magnitude smaller than the current MPOR of several days. The discrepancy between the last collateral update and the closeout amount, Δ , will be as small as the change in few hours of a net present value, computed with a single model.

No longer will derivative users have to endure litigation and lengthy procedures involving multiple third parties to arrive at a closeout amount, solving problem 3) and reducing risk and associated regulatory capital. The gap Δ between collateral and close-out amounts can be reduced to much smaller levels. If we want to minimize even this risk, we can think of initial margin here too. It will have to cover Δ , and will be much smaller than it is now, creating less strain on the liquidity of financial players, which solves problem 4).

COULD FAST REACTION TO MISSED PAYMENTS CREATE MORE DEFAULTS FOR TEMPORARY LIQUIDITY PROBLEMS?

Not necessarily, because on a ledger we can reduce the gap between collateral and close-out amounts to levels sufficiently small to allow us to exclude “on-chain” default. A missed collateral payment can be treated as an unwinding that generates a small balance to be settled in the longer term, when temporary problems, if they were really the issue, will surely be over. Let us see how this can be done.

It is reasonable to worry that a market where everything is faster or more automatic creates more technical defaults, due to problems like a temporary lack of digital cash. But the procedure above for the case of a missed collateral payment need not be considered a default in the usual legal sense, since we can design it contractually. We increase the risk of “technical” defaults only if we ask B to pay Δ immediately after the grace period. However, since the payment is now determined by a precise contractual agreement and is likely to be small, being based on a MPOR of few hours rather than 10 days, we can postpone this payment to a later time, to allow the counterparty to get the necessary liquidity. Default in the legal sense is thus driven out of the ledger. If this happens, it will be driven by external reasons, and will affect the network only for the precomputed amount Δ .

The participants will still be unhappy when a counterparty defaults and, for example, a hedge is lost. But, at least players now have as soon as possible as much cash as possible to

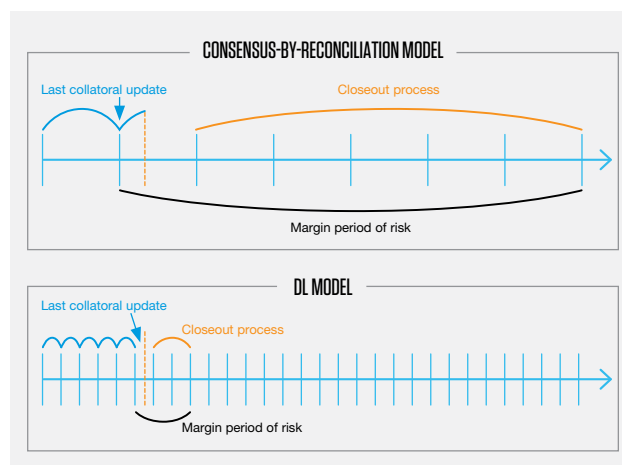


Figure 1 – Margin period of risk

find a new counterparty for the same deal. The long waiting times and discrepancies are cut out by design. The role of providing a counterparty in case of a crisis could be one of the roles that a CCP can take up if it reduces its direct exposure as counterparty of all trades in non-crisis times.

IT SEEMS MANY THINGS HAVE TO CHANGE TO ALLOW THIS: NEW (SMART) CONTRACTS, NEW (DISTRIBUTED) ACCOUNTING...

A lot of things have been to be made consistent with such a framework, including the regulatory framework. We also need money with digital representation, which can be a digital currency fully convertible in central bank accounts, an independent cryptocurrency, or just one or more currencies issued by banks redeemable with fiat currencies or other assets. The first choice is preferable, the other two choices have their own limitations, such as too much market risk, or volatility risk, for current cryptocurrencies,¹¹ too much credit risk, or default risk, for banks' money. The network needs to receive a number of inputs from outside, such calendar changes, fixings, data for valuation, and potentially valuation from an external engine. The technology for communication between ledger and the external world is the technology of Oracles. Standard contract specifications, including ISDA standards, will have to be expressed as template code.

DL technology is the natural way to get the cost and risk savings seen above for derivatives, not only because they enable faster clearing and settlement, but more importantly because they require first to move the market logic towards putting the on-ledger smart contract at the center of the transaction, as opposed to the current approach based on two different implementations and two different reports of a paper contract.

Legal and regulatory status could come earlier than expected if regulators see advantages in an architecture that is more transparent and creates less risk than most of the current solutions. This is why I think it useful that we continue with the analysis of advanced business cases: to show the possible advantages for financial markets, and to clarify the hard, but necessary, journey, moving once and for all beyond the false dichotomy between “blockchain hype” and “blockchain seclusion.”

¹¹ Morini, M., 2014, “Inv/Sav wallets and the role of financial intermediaries in a digital currency,” working paper, <http://bit.ly/2ldUqh2>

Trade Finance Disrupted: A Blockchain Use Case

André Brunner – Associate Partner, Capco

Nourdine Abderrahmane – Principal Consultant, Capco

Arjun Muralidharan – Senior Consultant, Capco

Patrick Halfpap – Partner – Corporate, Fieldfisher LLP

Oliver Süme – Partner – Technology, Outsourcing and Privacy, Fieldfisher LLP

Stephan Zimprich – Partner – IP Enforcement and Litigation, Fieldfisher LLP

Abstract

Blockchain holds considerable promise, but all too often it fails to find sustainable use cases. Trade finance is one significant exception. This traditional revenue source for banks is ready for disruption. A whole new approach, informed by blockchain logic, could bring the very transformation an established yet restricted business is crying out for. Placing the end-to-end trade finance process on a new platform would have a very clear objective: to offer all exporters and importers fast and easy access to credit issuance and advisory services. This is a radical shift but it is within reach. An open, automated, and transparent trade finance platform, which replaces cumbersome trust mechanisms with automatic checks and assurances, is a technical possibility right now. Barring certain technical and legal challenges, a product vision for blockchain as a disruptive force in trade finance is described and validated, awaiting adoption by a consortium of industry players willing to pilot a solution.

INTRODUCTION

Clients of trade finance services have been struggling for years with the risks involved in doing business with partners that are in other countries and whom they might not trust. Traditionally, the settlement risk has been lowered using letter of credits; however, the preparation and execution of these transactions have consumed a considerable amount of resources by the multiple parties involved.

Previous attempts of using technology to address the clients' pain points have not provided the desired breakthrough (we will be highlighting some of them later in this article). Would blockchain just be another technology not addressing the specific trade finance issues, does it have the power to finally disrupt the trade finance business, or may it be just another attempt designed for failure?

This article examines the current structure and state of the trade finance industry and summarizes previous efforts to address the needs of customers. It then introduces blockchain, and the potential of Distributed Ledger Technology (DLT), as a context to a product vision for a blockchain-backed trade finance platform, which we would like to introduce.

We conclude on further feasibilities and challenges or other potential roadblocks that could prevent this product vision from becoming successful, such as the legal challenges currently involved.

PROBLEM STATEMENT

The global trade of physical goods relies on the availability of credit, solid logistics, and transparent payments. Trade finance as a discipline addresses these challenges with well-established instruments to issue credit, document the transfer of exported or imported goods, and execute the subsequent payment. Our clients tell us that this business has become harder to manage than ever before.

Traditionally, the trade finance business has targeted exporters and importers of a certain size, with only mid-range to larger players being able to afford trade finance services. These include issuance of letters of credit (LC) and payment services on the buyer's side, as well as creation of bills of lading on the seller's side. The combination of the two, along with the respective negotiation, shipping, and payment processes, form the foundation of traditional trade finance.

There is room for growth [Clark (2014)]. Most crucially, today's trade finance business suffers from some key challenges:

- **Poor customer experience:** from a bank's point of view, the key customers, exporters, and importers often suffer from a poor end-to-end customer experience. Not only does the issuance of LC-backed trade finance transactions require high coordination efforts among exporters, importers, and issuing and advising banks, often represented by individual legal counsel, the settlement of the transaction follows overly complicated and manual processes, long waiting times with low transparency, and a relatively high residual settlement uncertainty. For example, importers still face the delivery risk of fraudulent shipments, even if the transaction is backed by an LC, and have to proactively and manually track all stages of the agreed delivery terms.
- **Increasing cost pressure:** issuance of LC is associated with high costs for both the bank and clients, while dispute resolution and limited scale create additional pressures due to a relatively high fraction of manual processing and a yet untapped customer base. Trade finance providers can gain strategic advantages if they succeed in finding fundamentally new approaches to delivering the service to their clients.
- **High regulatory burden:** management of geopolitical risks, such as sanctions and trade barriers, along with fraud prevention, KYC (know your customer), and AML (anti-money laundering) requirements are becoming a larger part of the trade finance business, further driving up operational overheads.

Since transactions may be highly complex, global, and expensive, the business is in dire need of both significant reductions in cost, as well as new sources of revenue to provide a broader, more stable footing for the business.

BLOCKCHAIN AND DLT

Overview

In 2008, Satoshi Nakamoto published the whitepaper concept for a new cryptocurrency called bitcoin, aimed to redefine how peer-to-peer transactions could work without clearing intermediaries. While the currency has fallen victim to countless journalistic critiques, the underlying mechanism of recording transactions without spending the same coin twice, namely blockchain, remained pertinent. Tapscott (2016) provides a succinct definition of the blockchain as "a vast, global distributed ledger running on millions of devices and open to anyone,

where anything of value – money, but also titles, deeds, identities, even votes – can be moved, stored and managed securely and privately. Trust is established through mass collaboration and clever code rather than by powerful intermediaries like governments and banks.”

Blockchain, or DLT, captured the hearts and minds of the very sector it was set to disrupt. Many existing financial services firms have reacted quickly and innovatively to this potential disruption, appearing to embrace its characteristics by launching joint ventures, creating industry alliances, joining consortia, and implementing proof-of-concept use cases. But will this be enough to combat the effects of disruption?

CHARACTERISTICS AND FEATURES

Characteristics

A blockchain is a permission-less distributed database that maintains a continuously growing list of transactional data records. The system’s design means it is hardened against tampering and revision, even by operators of the nodes that store data. The initial and most widely known application of the blockchain technology is the public ledger of transactions for bitcoin, but its structure has been found to be highly effective for other financial vehicles driving disruption in financial services. Some key characteristics of the technology are listed below.

- **Consensus building:** the ability for a significant number of nodes to converge on a single consensus of the most up-to-date version of a large dataset, such as a ledger.
- **Transaction validity:** the ability for any node that creates a transaction to determine whether the transaction is valid, able to take place, and become final (i.e., that there were no conflicting transactions).
- **Automated resolution:** an automated form of resolution that ensures that conflicting transactions (e.g., spend the same balance in different places) never become part of the confirmed dataset.
- **Generic adaptability:** originally applied to currencies, the chain can be applied to record transactions for any kind of assets (or even pure information), registering their existence, ownership, and changes thereof.

Features

While hailed as a new, disruptive technology, it remains important to note that blockchain builds on a range of existing concepts, weaving these into a new paradigm that applies ideas of distributed computing, encryption, and programmable

business logic into a singular concept. Among the key features of interest are blockchain’s sequential database, distributed nature, immediate processing, smart contracts, immutability, and security.

- **Database:** blockchains can store data in the form of ledger entries that are stored in strict sequence. Every participant in the network can see each other’s transactions.
- **Distributed:** many copies of the same data are stored with each participant of the network. This established trust and ownership defined by algorithms and no central party or system is needed.
- **Immediate:** settlement can happen quickly in a large network, and confirmed transactions are broadcast to the rest of the network.
- **Programmability:** smart contracts can store and execute defined business logic while using bespoke “coins” for handling custom assets.
- **Immutability:** a chain of transactions is posted sequentially in time-stamped blocks, so that amendment of a transaction retroactively is not possible unless the blockchain is overridden by the majority of the network or the it is designed to accept changes with a special type of hash that does not endanger the integrity of the overall chain.
- **Security:** the ledger is open to the network, yet encrypted with industry standard private and public keys, where each transaction is encrypted using a hash function depending on the previous transaction or block.

These features lead to the hypothesis that blockchain is a viable candidate to address some of the challenges faced in trade finance. Based on a review of the current market, along with our outlined overall vision for trade finance, this paper intends to confirm the viability and applicability of blockchain as a suitable technology.

PREVIOUS MARKET DEVELOPMENTS

Several efforts to innovate in the area of trade finance and central platform or peer-to-peer network approaches could be observed in recent history.

In the research world, several studies have examined the potential of blockchain and its specific features with regards to solution potential for specific industry issues. From a customer requirements point of view, there is research that underlines that any electronic settlement substitute for manual trade finance settlement must also be able to cope with “trust, risk,

legal recognition and multi-jurisdictional issues” [Dixon and Glasson (2017)]. Likewise, a case study highlighted that the main obstacle to adoption of electronic bills of lading solutions would be the lack of modern, public registries, to which the settlement parties would refer to [Dubovec (2006)].

From a solutions perspective, a study on the potential for blockchain to significantly disrupt supply chains between companies [Dahlberg et al. (2017)] concluded that distributed ledger and smart contract features can contribute to a significant cost reduction and disintermediation in supply chains.

There have also been several developments in the trade finance industry to connect a number of participants in the industry. In 2013, for example, Kyriba, a provider of cloud-based treasury management solutions, announced that it will integrate CGI’s Trade 360, a trade finance platform, into its offerings [Kyriba (2013)]. While it is unknown to us the extent to which this effort has addressed actual TF customer needs and what market share this solution has earned, Kyriba’s SVP Edi Poloniato pointed out that “the need to manage trade finance and cash management in a single platform has become a core requirement for our clients.”

A slightly different approach was announced in the same year by Bolero, a leading provider of electronic trade finance documentation management, who have partnered with China Systems’ Eximbills Enterprise to provide a seamless electronic integration of the Eximbills Trade Finance Back-Office solution into the Bolero network [Bolero (2013)]. What might sound like as simply another incremental software evolution, has, in fact, provided an immense automation potential for

banks’ back-offices to exchange electronic messages and documents via a platform. We judge this development as a clear sign that trade finance banks, the customers of Bolero, have also expressed that manual settlements of LCs as a clear pain point.

In 2016, Bank of America Merrill Lynch (BAML), HSBC, and the Infocomm Development Authority of Singapore (IDA) tried to bring paper-intensive LC transactions onto blockchain. Even the vibrant community around IBM Hyperledger have begun outlining the use case of an electronic ledger-based LCs [BAML (2016)]. Barclay’s claimed to have executed the first letter of credit transaction on a blockchain in late 2016 [Barclay’s (2016)]. All these efforts came on the back of a landmark paper released by the European Banking Association on applying crypto technologies to trade finance, which asserts that the potential use cases go beyond LCs and could also serve to enable more flexible financing, better exchange of information, and provide instant payment infrastructures on the back of smart contracts.

We believe that the potential for disruption in trade finance goes beyond simply transactions and could be approached more holistically. According to AT Kearney and the WHU Otto Beisheim School of Management, digitization of supply chains will see significant investment at almost three out of every four firms that trade goods globally [AT Kearney (2015)]. The enablement of a trade finance transaction relies on the end-to-end collection of information through sensors, smart input mechanisms, and real-time data processing. Consequently, we need to determine whether blockchain as a technology can solve the challenges of the trade finance industry, and how it can be used in the context of a broader platform solution.

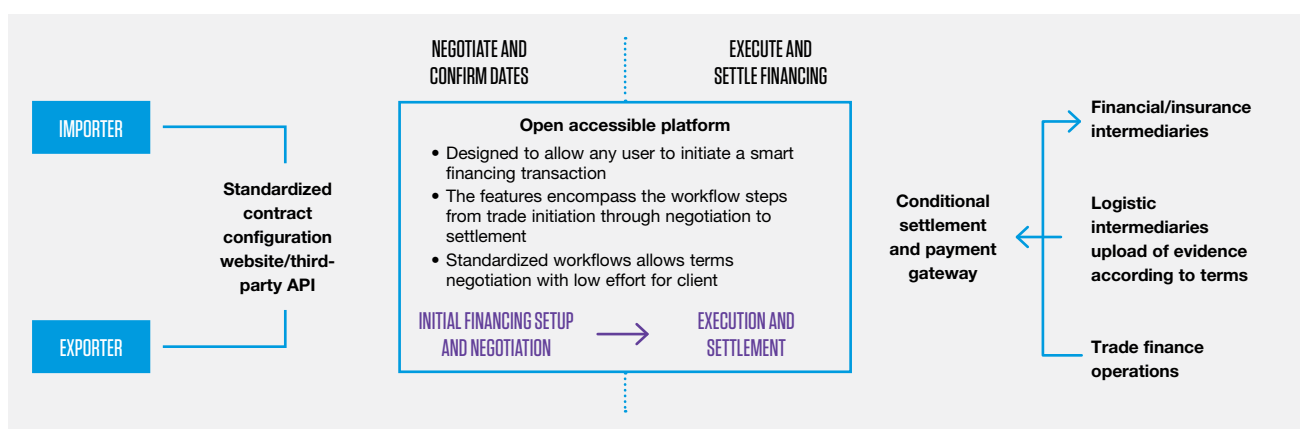


Figure 1

PRODUCT VISION

A potential solution lies in moving beyond trade finance providers’ traditional confines and thinking across the wider landscape. Hitherto closed systems need to open up if banks expect to harness untapped market potential that may, sooner or later by virtue of technological developments, be siphoned off by disruptive and niche upstarts. This entails a fundamental paradigm shift: rather than provide a bespoke service to a limited set of clients who can afford it, we envision a marketplace where any exporter or importer can easily and quickly have access to credit issuance or advisory services.

A new platform for executing trade finance transactions is needed. It needs to be open, automated, and so transparent that the parties do not have to rely purely on trust.

- **Trade finance operations teams** need to quickly process transactions and reduce overheads from daily operations to focus on handling a small number disputes.
- **Exporters and importers** need to be able to apply for credit and ensure payment without complicated agreements. They also need to track the status of a transaction at any point in time.
- **Logistics providers** need to understand when goods can be picked up, shipped, and delivered without violating contractual terms of their customers.
- And lastly, the involved **financing intermediaries** need full transparency to minimize risk.

A platform that is open to all involved parties would allow:

- **Self-service “à la carte” contracts:** the platform provides an API for standard contract creation. Exporters and importers can build their contracts through any website that uses the platform’s openly available API and use predefined templates or building blocks to ease the creation of contracts.
- **Conditional payment and settlement gateway:** the platform allows for financiers, logistic intermediaries, and bank operations to track and execute the transaction according to the contractual terms. Any software solution can be adapted to use the new gateway.
- Furthermore, such a platform should aim to implement **transparency** of trade finance transactions among the involved parties, in real-time, and,, therefore reduce the **requirement for trust** by minimizing counterparty risk from the outset.

Such a platform would be built by an incumbent or consortium of incumbents, and would scale easily given the openly

available APIs to adapt to existing solutions gradually. It could integrate with existing offerings by addressing specific market segments, such as mid-complexity deals, while legacy systems and processes continue to serve the traditional, highly-custom, and personalized LC business. While new market potential can be tapped through this platform, it would also increase competitiveness given the standardized protocol and ease of use, thereby adding pressure to incumbent banks to innovate beyond commodity trade finance services.

So how does blockchain solve any of these challenges? Fundamentally, blockchains are good for a variety of scenarios where a confirmed sequence of events is business critical, and predetermined business logic needs to be respected by multiple nodes in a network with no room for tampering.

Many start-ups have attempted to address trade finance with blockchain. However, we have found them to be lacking in two ways so far. Firstly, they usually aim to map the existing processes into a blockchain, rather than fundamentally rethinking the business model and information flows. This means you will find many solutions providing a “digital LC” but not questioning whether such a letter is needed when credit can be issued, confirmed, and payment initiated instantly. Secondly, they

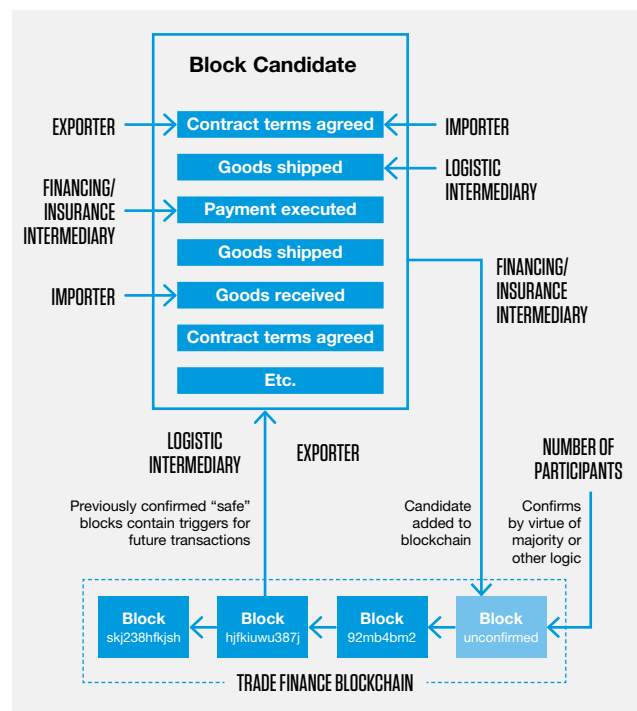


Figure 2 – Sample blockchain-based view

tend to map only a single, isolated piece of the value chain, focused usually on LCs or bills of lading, and much less on the issuance or advisory aspects that go on in-between. Importers and exporters pay for these expensive services but are often neglected and suffer from a poor customer experience.

In the context of a trade finance transaction, we can foresee blockchain solving three fundamental challenges:

- **Automating issuance, shipment tracking support, and payment:** assuming the various network players have adopted the proposed platform, they would record events such as issuance of credit, shipment of goods, or the payment for a transaction in real-time to the blockchain, which in turn broadcasts this out to the network. Given this public record of events, a paper-trail chasing operations team is no longer required and subsequent steps in the workflow can be reliably initiated on an automated basis.
- **Full audit transparency:** shipments are prone to fraud or illegal activity. A shipping container labelled as carrying bananas may be carrying freighting weapons, hence control is needed. Blockchain can support this through a network-controlled mechanism, where a “shared truth” is established by the majority. In addition, all events, such as term agreements, shipments, or payments can be traced back and the likelihood of erroneous events is minimized. Transparency and access to this information can be controlled and limited to approved network participants, resulting in a hybrid blockchain sharing traits of both private and public blockchains.
- **Improved advisory services:** by providing a transparent, up-to-date source of truth, trade finance operations teams are empowered to provide more accurate and relevant advisory services. Moreover, by taking over many of the generic, repeatable, and standardized elements of a transaction, the blockchain frees up resources that can be dedicated to more bespoke forms of advisory.

Beyond these two fundamental aspects, blockchain can enable development of client endpoints across the globe in a cost-effective manner. If the platform champions decide to make development open and the blockchain’s logic accessible, rather than just a workflow solution, you would create an entire ecosystem of trade finance apps, all backed by the same internal logic but tailored to the needs of a shipping coordinator, a bank employee, or a buyer of goods.

FEASIBILITY AND CHALLENGES

We acknowledge that a blockchain does not inherently replace your trade finance workflow, however, a platform backed by blockchain, along with other emerging principles of open architecture, collaborative and API-based computing, and real-time processing can provide the foundation for a potentially disruptive solution in this business. This solution, in contrast to previous attempts, would focus predominantly on the needs of trade finance customers, and would fundamentally challenge the banks’ man-in-the-middle approach implemented within the existing LC approach.

Putting this into practice will not be simple though. Our proposed vision assumes that a group of partners in the supply chain get on board initially with such a platform, and that adoption is made easy and cost-effective across the entire value chain. The platform comes to fruition by starting with a minimum viable product, such as a smart contract to replace an LC, and building this out to support both incumbent legacy systems as well as newly crafted applications.

Realistically, a transitional model would include smaller chains built by groups and consortia of trade finance parties, which in turn can be chained together down the road with an emerging technology called Sidechains. These allow moving transactions between blockchains if they agree to and adopt some common understanding of proof of work. Trade finance customers may be incentivized to use the new model through lower transaction costs.

Tokenization and ownership

A fundamental technical challenge to multi-party blockchains is posed by process ownership. Who is entitled to write to the blockchain in a series of logical events? How does the blockchain pass the torch from importer to exporter to issuing bank?

In basic terms, blockchains can handle the ownership of assets in the form of coins, as used in any cryptocurrency. Consequently, a paradigm in which the ownership of an ongoing trade finance transaction is handled as currency may be the solution. As “trade finance coins,” the handling of process ownership and entitlements to the blockchain can be ensured by executing a transfer of an authorization token, in the form of a coin, to the next party that needs to act in the process. For example, once an importer has received the goods, an authorization token is passed on to the importer’s bank, executing the payment to the exporter. In this step, it is ensured that the importer’s bank is notified and authorized only once the goods

have been confirmed as delivered by the importer, purely using the mechanics of a blockchain.

Gatekeeping and KYC concerns

A further consideration needs to be made with regards to how new actors can be admitted to and execute trade finance transactions on the proposed platform. In an initial model, we foresee an authorizing entity, such as a bank or consortium of trade actors, approving new importers, exporters, logistics intermediaries, and banks via traditional forms of KYC. Over time, these players will have established histories on the blockchain, facilitating automated KYC procedures for actors to cross-check each other using the available history of transactions on the chain.

Apart from issuing authorizations, safely revoking the same without endangering the integrity of the blockchain is a feature that established blockchain technologies, such as Hyperledger, have yet to implement.

Legal challenges

In order to fully exploit the potential of the blockchain technology, the legal and regulatory framework requires modifications, too. The blockchain questions fundamental assumptions underpinning most regulations, such as the idea that a certain individual or entity is accountable and responsible for a certain part of a multilateral system. In a fully-fledged, distributed blockchain system, neither the infrastructure nor the application itself is centrally operated or controlled, and the lack of central control or gatekeeper is a substantive part of the idea.

With regards to trade finance platform solutions discussed herein, there are various legal and regulatory issues that have to be kept in mind. As the platform would ultimately also affect payments of the parties involved, the provider of such a platform is likely to be held as a provider of payment services by the German Payment Institution Act (Zahlungsdienstleistungsgesetz – ZAG) or similar mandatory provisions of law under legal regimes of other jurisdictions. If the platform solutions also contains the conveyance of insurance products (i.e., such as import/export credit insurance products), the relevant provisions of the German Trade Regulation Act (Gewerbeordnung – GewO) or similar mandatory provisions of law under legal regimes of other jurisdictions will have to be checked in more detail. In case it is intended to offer even further financial services, the German Banking Act (Gesetz über das Kreditwesen – KWG) or similar mandatory provisions of law under legal regimes of other jurisdictions will have to be examined as well, as there is a catalogue of regulated activities that require a

license. Hence, when setting up the solutions discussed herein, it will be crucial to determine beforehand the exact scope of services provided and to assess if and to what extent regulated activities shall be carried out. If regulated activities, such as payment services, are to be provided, either the relevant regulatory permissions need to be obtained by the providing entity or cooperation with partners holding the required licenses, and willing to provide the discussed services in the jurisdictions relevant for the market, could be a solution. As a first step, it, of course, might also be an option to limit the functionality of the platform to a mere technical infrastructure portal and hence avoid any provision of regulated services.

The described trade finance platform could provide for contractual solutions based on “smart contracts,” which is based on the idea that a contract is self-performing. In general, this conflicts with the general legal assumptions that each and every legally relevant act must be initiated by a human being. It is easy to imagine smart contracts with autonomous decision-making, and without any human involvement at all. We have as yet no answer to the question of what such a decision might mean in legal terms – and we certainly would need to define rules to determine which decisions may be made by machines, and which require human intervention.

Further, issues could arise once there is a dispute between the parties of such a trade finance platform. There are also no rules regarding the question of how to use blockchain data in court. From a coder’s perspective, it is rather clear how a set of blockchain data must be interpreted – but a judge is not a coder, and for now courts rely on expert opinions to establish technical facts. In order to also preserve the efficiency gains of the blockchain in court proceedings, we would need rules and technical solutions that make blockchain data accessible and comprehensible for a court. And the rules should be such that there would not be a need to appoint an IT expert witness in each and every case, which usually results in significant delays before a decision is rendered.

Further questions are raised regarding the enforcement of claims. So far, if a claimant is awarded a title against a debtor, they could enforce it by several means, including seizing payments on the debtor’s bank accounts. In a blockchain-based system, any payment would require the active involvement of the debtor, as no third party would be technically able to prompt a payment from the debtor’s crypto-currency account without knowing the private key.

Another feature of the blockchain that challenges traditional thinking is the built-in transparency of a blockchain-based

system. While this is great in some aspects – you always have a bullet-proof audit trail – it raises numerous questions related to data protection regulation and the secrecy of transactions.

A blockchain-backed trade finance platform would, of course, concern the storage and processing of data on a large scale. Even if the trade data are encrypted, it is most likely that at least a significant part of this data has to be qualified as personal data according to the general data protection regulation (GDPR), which comes into force on May 25th, 2018.

Against this background, a blockchain-based trading system would have to comply with the requirements of this regulation, which will be binding and directly applicable in all E.U. member states. For any blockchain platform project, it is, therefore, key to anticipate, who will act as a “controller” according to the GDPR, as the controller is responsible and liable for the lawful processing of personal data. Unlike a public and permissionless blockchain, such as those used for cryptocurrencies, such as bitcoin, a blockchain-backed trade finance platform is likely to be a private and permissioned blockchain, where key functions are operated by one or more entities, who set up the terms of use. It is most likely that those players will have to qualify as “controllers” (or “joint controllers,” who are explicitly mentioned in the regulation) from the perspective of GDPR. The controller is obliged to implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with the GDPR. While even a private blockchain platform is basically ideally suited to protect personal information by its decentralized structure, the controller of the platform would have to document how the individual platform protects personal data and information of their users. The GDPR requires the controller to maintain a record of processing activities, which contains amongst other information the purpose of the processing of personal data and descriptions of categories of data subjects.

In addition, under the GDPR controllers will see their obligation to inform the data subject increased. They will have to notify those concerned of the anticipated retention period as well as about the right to withdraw the users consent and the right to lodge a complaint.

An appropriate and compliant privacy policy will, therefore, play an important role for a blockchain-based trading system. Given the fact that the GDPR can lead to significantly higher fines than the current legislation, compliance plays a very crucial role here and the impact of processing personal data on the trading platform should be considered in due time.

CONCLUSION

Trade finance is ready for a blockchain-driven disruption. The technology provides fundamental aids in alleviating concerns around cost, security, ease of use, and speed. Furthermore, a platform surrounding a trade finance blockchain would facilitate actors in executing transactions. Such a platform could be developed as a pilot among a selected consortium of trade actors.

While there remain technical and legal challenges to bring the proposed platform to fruition, a practical business case is likely. We foresee blockchain making large strides in trade finance, as is already happening with distinct pieces of the entire transaction chain. The fundamental question, therefore, is not whether trade finance will be disrupted, but by whom, and when the first successful end-to-end attempt will be made.

REFERENCES

- AT Kearney, 2015, “Digital supply chains: increasingly critical for competitive edge,” <http://bit.ly/2muqkpu>
- Barclay’s, 2016, “The blockchain revolution in trade finance”, <http://bit.ly/2mrNlsF>
- BAML, 2016, “BofAML, HSBC, IDA Singapore build pioneering blockchain trade finance app,” <http://bit.ly/2INbCoP>
- Bolero, 2013, “Bolero announces integration with China Systems’ Exim bills trade finance application,” <http://bit.ly/2mutut8>
- Clark, J., 2014, “Trade finance: developments and issues,” Bank of International Settlements, CSFG papers no. 50, 14 - 16
- Dahlberg, T., J. Hallikas, and K. Korpela, 2017, “Digital-supply chain transformation toward blockchain integration”, in “The digital supply chain of the future: technologies, applications and business models minitrack,” 4188-4189
- Dixon, M., and B. Glasson, 2017, “Electronic payment systems for international trade,” <http://bit.ly/2luVihu>
- Dubovec, M., 2006, “The problems and possibilities for using electronic bills of lading as collateral,” *Arizona Journal of International & Comparative Law* 23:2, 466
- Kyriba, 2013, “Kyriba and CGI to deliver seamless cloud-based trade finance capabilities,” <http://bit.ly/1pRSGWV>

Towards a Standards-Based Technology Architecture for RegTech¹

Tom Butler – Professor, GRC Technology Centre, University College Cork, Ireland

Abstract

This paper highlights the need for industry and technology standards in the design, implementation, and use of RegTech. Without such standards, RegTech may fail to fulfil its promise of facilitating “smart regulation.” It is well-accepted that RegTech has the potential to help financial enterprises address the following issues: (1) solve the regulatory interpretation problem; (2) develop compliant governance and business policies; (3) make regulatory compliance reporting more efficient and effective; (4) help firms perform better data governance and analytics; (5) enable integrated risk management; and (6) automate controls across the business. Two significant problems challenge the potential of RegTech. The first of these is the “translation problem,” which affects not only the design and implementation of RegTech, but also how it will be employed to close the gap in regulatory interpretation and understanding. The second is the “Tower of Babel” problem, which refers to the absence of a “common language” in the financial services industry. This paper discusses how semantic

standards can help solve potential problems with RegTech. Semantic technologies enable meaning to be attached to data – both structured and unstructured. RegTech solutions anchored on semantic standards can unpack regulatory requirements in complex and voluminous regulations. This will, we believe, require the use of standards-based regulatory and business ontologies. Semantic standards and technologies thus developed can enable RegTech solutions to help practitioners better navigate their digital labyrinths. Semantic technologies will, we believe, play a key role here, as without them the challenges arising from BCBS 239 cannot be addressed in a coherent, cohesive, and comprehensive manner.

¹ This work was supported by Enterprise Ireland and IDA under the Technology Centre Program [Grant TC-2012-009].

INTRODUCTION

According to a recent report by Bain and Co., Governance Risk, and Compliance (GRC) spend accounts for 15-20% of “run the bank cost,” and 40% of “change the bank costs” for major banks.² Bain and Co. contend that such costs will grow over the next five years, as banks continue to struggle with regulatory requirements.

There is broad agreement that banks could realize substantial benefits from innovations in RegTech in addressing this challenge.³ EY, for example, argue that “In the short term, adoption of RegTech will provide operational efficiencies and cost benefits when applied to current compliance and risk management practices.”⁴ Regulators appear to agree with and support the adoption of RegTech. In a speech delivered by Christopher Woolard, Director of Strategy and Competition at the Financial Conduct Authority (FCA), at London FinTech Week in July 2016,⁵ several use cases for RegTech were identified viz.

1. “First, making the business of complying with reporting requirements simpler – technology that allows more efficient methods of sharing information (for example: alternative reporting mechanisms, shared utilities and online platforms).
2. Second, technology that drives efficiencies in regulatory compliance by seeking to close the gap between the intention of regulatory requirements and the subsequent interpretation and implementation within firms. For example, we have seen a range of semantic technologies and significant enthusiasm for robo-advice style models to help firms understand their regulatory responsibilities.
3. Third, technology that simplifies and assists firms in managing and exploiting their existing data, supporting better decision-making and finding those who are not playing by the rules easier. This includes new data analytics technology, real-time compliance monitoring and trade surveillance systems.
4. Finally, technologies and innovations that allow regulation and compliance processes to be delivered differently and more efficiently. Here we see significant interest in distributed ledger technologies, automated compliance systems, machine-readable regulation and expanding use of biometrics for identity verification purposes”

The FCA’s Project Innovate incorporates TechSprint events, the focus of two of these has been RegTech and, in particular, the theme of “unlocking regulatory reporting.”⁶ The GRC Technology Centre and several of its industry members attended the most recent event in February 2017. While the focus was

on key aspects of regulatory reporting, the themes emerging from the discussion and presentations on nascent RegTech innovations mirror those found in thought leadership pieces and in technologies currently being deployed, such as: (1) fraud prevention and anti-money laundering (AML); (2) employee and third party surveillance; (3) regulatory and governance compliance and conduct risk assessment metrics; (4) predictive analytics; and (5) regulatory compliance and reporting support and automation.

There are varying degrees of maturity and market acceptance of these technologies. While there are clear benefits to the adoption of RegTech, there is also an unacknowledged downside. This is due primarily to the ad hoc way in which RegTech is being adopted across the industry. The key issue here is the need for standards in the design, development, and implementation of RegTech.⁷

THE PROBLEMS CONFRONTING THE SUCCESSFUL IMPLEMENTATION OF REGTECH

In his penetrating analysis of “technologies of compliance,” Kenneth Bamberger, states that “While these technology systems offer powerful compliance tools, they also pose real perils. They permit computer programmers to interpret legal requirements; they mask the uncertainty of the very hazards with which policy makers are concerned; they skew decision-making through an ‘automation bias’ that privileges personal self-interest over sound judgment; and their lack of transparency thwarts oversight and accountability. These phenomena played a critical role in the recent financial crisis.”⁸ One of the key issues identified by Bamberger is the problem of translation, which has several dimensions.⁹ There are, however, other problems.

² <http://bit.ly/2devi2n>

³ Arner, D. W., J. Barberis, and R. P. Buckley, 2016, “The Emergence of RegTech 2.0: from know your customer to know your data,” *Journal of Financial Transformation* 44, 79-86

⁴ <https://go.ey.com/24SGCnl>

⁵ <http://bit.ly/2m2UH54>

⁶ <http://bit.ly/2ffadWC>

⁷ See the following on the need for standards in GRC: Spies, M., and S. Tabet, 2012, “Emerging standards and protocols for governance, risk, and compliance management,” in *Handbook of research on e-business standards and protocols: documents, data and advanced web technologies*, IGI Global.

⁸ Bamberger, K. A., 2009, “Technologies of compliance: risk and regulation in a digital age,” *Texas Law Review*, 88:4, 669-739

⁹ Butler, T., and E. Abi-Lahoud, 2014, “A hermeneutic approach to solving the translation problem in designing ontologies,” 22nd European Conference on Information Systems (ECIS), Tel Aviv, Israel

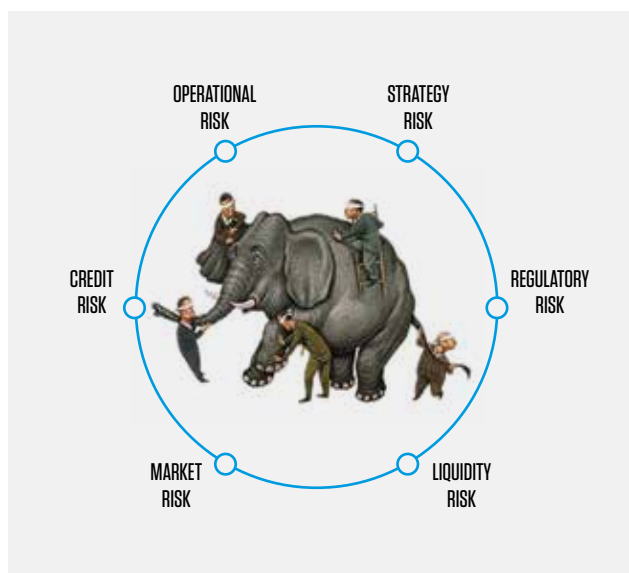


Figure 1 – Regulatory reporting and risk data

In 2013, Andrew Haldane, Executive Director for Financial Stability at the Bank of England, identified what he termed the “Tower of Babel” problem. He argued that the financial industry “has no common language for communicating financial information. Most financial firms have competing in-house languages, with information systems siloed by business line. Across firms, it is even less likely that information systems have a common mother tongue. Today, the number of global financial languages very likely exceeds the number of global spoken languages.”¹⁰ The scale of this problem is highlighted by the fact that our research indicates that a typical large international bank may have up to 70,000 information systems and over 250,000 spreadsheets. We have previously indicated a practical consequence of this problem, best illustrated through our application of the blind men and the elephant parable.¹¹ Figure 1 illustrates this from a risk perspective.

Figure 1 attempts to not only highlight the siloed nature of operational, regulatory, and other risk data, but also the fact that professional silos exist in financial services organizations themselves. As Andrew Haldane points out, people, processes, and technologies within the same organizations do not share a common language. Thus, not only do existing GRC systems suffer from translation problems, they also exhibit the “Tower of Babel” problem. Without standards, RegTech will simply mean that a business-as-usual approach will prevail, and the desired transformations will prove elusive.

Clarion calls for change

In January 2013, the Basel Committee on Banking Supervision (BCBS) issued its “Principles for effective risk data aggregation and risk reporting,” also known as BCBS 239. This came into effect for G-SIBS, or Global Systemically Important Banks, in January 2016. These new regulatory requirements are targeted at the manner in which financial institutions manage data aggregation and risk. Here again the need for standards is evident in that key requirements include: (a) harmonization of data definitions across information systems and lifecycles; (b) enhanced governance policies and the allocation of data ownership and accountability for the quality of risk data; and (c) improved data quality through the accuracy, completeness, timeliness, and adaptability of data infrastructures.

In November 2015, the Financial Stability Board’s (FSB) called for a common language or taxonomy with which to manage conduct risk.¹² It is worthwhile restating their requirements here: “The integration of conduct risk in all aspects of a firm’s business, in a manner that is consistent across the industry, requires the development of a consistent set of definitions, methods of assessment and measurement of conduct risk.” Of course, what the FSB is really requesting is a standard.

The diversity of data formats and the absence of modeling and reporting standards is also of concern to the European Commission (E.C.). The department for financial stability and capital markets (DG FISMA) is responsible for the E.C.’s policies on banking and finance. In 2016, it instituted the Financial Data Standardization Project.¹³ Specifically, this is looking to (a) implement financial data standards for messaging; (b) semantic standards for data dictionaries/ontologies/classification; (c) legal and other business identifiers, specifically entities, products, and transactions; (d) reporting and business domain standards; and (e) business contract standards.

Thus, we argue that RegTech solutions providers and adopting financial institutions need to be aware of the need for standards-based approaches to the above problems, if RegTech is not to become part of the problem itself.

¹⁰ Haldane, A. G., 2012, “Towards a common financial language,” presentation at the Securities Industry and Financial Markets Association (SIFMA) “Building a global legal entity identifier framework” Symposium, New York, 14 March

¹¹ Butler, T., and E. Abi Lahoud, 2014, “Applying semantic technologies for risk data aggregation,” Consortium for System Risk Analytics (CSRA) meeting, MIT Sloan Center for Finance and Policy, Cambridge, MA, December

¹² <http://bit.ly/2m3g4Si>

¹³ <http://bit.ly/2lIZFvg>

USING STANDARDS TO HELP SOLVE POTENTIAL PROBLEMS WITH REGTECH

Semantic technologies (SemTech) have been identified as a means to help solve enduring problems of regulatory compliance in the financial services industry. The recognition that SemTech could be of benefit to the industry was contemporaneous with an important and generally unnoticed paradigm shift in the IT industry with the emergence of NoSQL (Not only SQL) solutions, such as graph data stores.¹⁴ The emergence of this new paradigm has generated new possibilities for managing, mining, and processing of structured and unstructured data. However, the de facto and de jure standards that developed around semantic technologies help address the various problems with RegTech.

What is SemTech?

Semantic models and related technologies enable unstructured and structured data to be endowed with meaning; something which is not possible with traditional technologies based on relational “structured query language” (SQL) databases or web pages based on HTML. At one level, a semantic model enables human communication. At another level, a semantic model enables heterogeneous data to be linked and data in siloed SQL databases to be federated and integrated. In addition, SemTech can make unstructured data, such as text-based documents, such as regulatory texts, machine readable using domain ontologies, thereby enabling information extraction into a knowledge base.

The World Wide Web Consortium (W3C) is establishing recommendations that have become de facto standards for supporting machines in processing data on the WWW, which includes data in databases. Using de facto standards ensures trust and enables trusted interactions between applications in computer networks. The primary use case for Semantic Web technologies is to enable developers to store data on the Web, to build vocabularies, and to write rules for handling data. There are several core technologies that are represented in Figure 2.¹⁵ At the bottom of the stack is “uniform resource identifier” (URI), which is a string of characters used to identify resource in a network. Above it is XML (extensible markup language), which defines a set of rules for structuring data and documents in a human-readable and machine-readable format. The upper layers of the stack are built on top of XML. For example, RDF (resource description framework) is one of the three foundational Semantic Web technologies, the other two being SPARQL and the “web ontology language” (OWL). RDF is the data modeling language for SemTech. OWL is the knowledge representation language. SPARQL, or the SPARQL Protocol and RDF Query

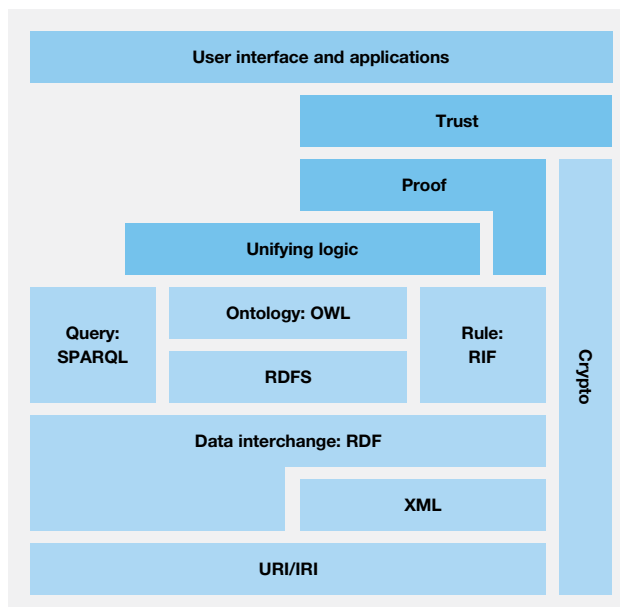


Figure 2 – W3C semantic web stack

Language, is, as its name indicates, the query language for the Semantic Web and siloed and distributed networked systems.

An ontology expressed in OWL provides additional semantics for data models, in that knowledge of objects and their relationships is more richly expressed. Triple stores are essentially graph stores based on RDF/RDFS, and while they more expressive than a relational data store they are less expressive than OWL. Both can be used to capture knowledge about a domain, such as operational risk.

An ontology describes a conceptual model about a problem domain, which is in effect metadata. This can also be expressed in RDF/OWL and may be persisted in the same RDF triple store as the instance data. Thus, both metadata and data can be queried.

The power of ontologies is that they enable reasoning or inferring in RDF triple stores. The advantage is that a reasoner may infer new/additional triples or relationships – that is add new knowledge – based on the asserted knowledge or axioms about classes and instance data in the ontology.

¹⁴ McCreary, D., and A. Kelly, 2013, Making sense of NoSQL, Manning Publications, Greenwich

¹⁵ A complete overview of the W3C Semantic Stack may be found at <http://bit.ly/2mAGEVT>

The W3C semantic technology stack provides an ideal platform to create extensible, standards-based RegTech platforms. This paper explains how these and related technologies can be employed for standards-based risk and compliance data aggregation in an upcoming section. However, we first address the considerable challenge of unpacking regulations and related rules into human and machine-readable formats requiring additional support for end-users. Here, again, semantic standards will play a key role.

Using standards to unpack regulatory requirements

In a perfect world, legislators and regulators would publish regulations and rules in an unambiguous, easy to interpret human- and machine-readable format. However, we do not live in such a world, and firms in the financial services industry face a Herculean task. It is estimated that 50,000 regulatory texts were published by G20 members since 2009. There is an average of 45 new documents each week.¹⁶ MiFID II (Markets in Financial Instruments Directive) has recently led to over 30,000 pages of text being generated in all aspects of its implementation.

Current approaches to unpack regulatory requirements are labor intensive and have a lot in common with the classical Greek myth of Sisyphus. Sisyphus, the King of Ephyra, was punished for his cunning and deceitfulness by the Greek God Zeus and condemned to roll an enchanted boulder up a hill. However, Zeus' spell ensured that it rolled back down again, leading to Sisyphus repeating the task. This destined Sisyphus to an eternity of futile, fruitless, repetitive activity.

It appears that the financial services industry has been so condemned, given the volume, variety, velocity, and complexity of regulations drafted since 2008, and the responses being taken to deal with the problems of regulatory compliance. As with Sisyphus, organizations typically reach the top of the hill and perform regulatory change management with boulder-sized regulations such as MiFID, for example, only to have to begin the process all over again when MiFID II came along. Generally speaking, organizations appear to be starting from scratch each time they do regulatory change management, as any previous knowledge they gained in interpreting and making sense of prior regulations has not been codified and captured in an organizational knowledge base. Dealing with regulatory rules spawned by the likes of Dodd Frank involves similar trips up and down the regulatory compliance mountain, with equally problematic outcomes for knowledge acquisition and institutional learning outcomes.

Using standards-based regulatory and business ontologies

Ontologies can help legal and business practitioners make sense of a wide and complex spectrum of legislation and regulations and to provide financial services organizations, GRC and RegTech vendors, and others in the ecosystem, with the ability to (1) query legislation, regulations, and other texts in order to identify compliance imperatives; and (2) identify changes to existing legislation and regulation introduced by amendments to existing law or new law. Thus, standards-based ontologies should inform the architecture of, or be incorporated into, RegTech solutions.

For example, a variety of upper-level ontologies may be used (i.e., accessed via URIs) to map, integrate, semantically enrich, and categorize lower level concepts and help increase overall reasoning and inferencing accuracy. URIs (universal resource identifiers) are globally unique, permit data elements (objects, classes, entities, concepts, relationships, attributes) to be identified, and link data from different sources and merge them with accuracy. Thus, concepts from core ontologies, such as the Financial Industry Business Ontology (FIBO), can be linked with those defined ontologies used to develop RegTech solutions.

In addition, general concepts in such ontologies can be imported from taxonomies published by the International Accounting Standards Board (IASB)/International Financial Reporting Standards (IFRS) and the U.S. Financial Accounting Standards Board (FASB) directly, if URIs are available, or indirectly imported as concepts. Concepts and elements from FIBO, IFRS, and FASB-GAAP could form the basis of the top-half of the domain specific ontologies. In this scenario, a RegTech operational or domain-specific ontology will contain core ontology concepts and relationships and firm-specific concepts and relationships. The latter may be generated using readily available technologies from the relational schemas in operational and risk data stores, Excel schemas, or objects and relationships in unstructured data such as texts.

¹⁶ <http://bit.ly/2lCoQVv>

USE CASE FOR REGULATORY ONTOLOGIES

In keeping with the objective of standards for RegTech, the GRCTC (Governance Risk and Compliance Technology Centre at University College Cork) developed the “financial industry regulatory ontology” (FIRO), an open standardized model of regulations. The FIRO semantic framework is composed of four modular ontologies: FIRO-H (high-level), FIRO-S (structural), FIRO-D (domain-specific), and FIRO-Op (operational). The FIRO-H ontology describes high-level concepts and their relationships, which are applicable across the regulatory domain. This includes concepts, such as obligation, prohibition, exemption, or sanction. FIRO-S ontology models the formal structure of parliamentary, legislative, regulatory, and judicial documents. FIRO-D describes domain-specific concepts and their relationships.

FIRO underpins the development of a suite of RegTech applications currently under R&D. In terms of use for regulatory compliance it can achieve the following:

- Reason on rules that are exceptions to other rules because they allow a subset of the conditions forbidden by another rule.
- Reason on business rules that ensure compliance with legal rules because they require a subset of the conditions required by another rule.
- Classify data (e.g., transactions) as “relevant” to a certain rule (legal statement) and further distinguish between “relevant and compliant” and “relevant and in breach of” the legal statement.

As regulatory rules reference financial processes and products, there is a necessity to have a business equivalent – and here is where business natural languages come in.

USE CASE FOR BUSINESS ONTOLOGIES

The financial services industry faces system and data integration problems that are unique in nature. Business processes and transactions span multiple entities and functions and sophisticated supply chains, with several trading entities and data being exchanged in a range of formats and message protocols. Add to this a multiplicity of systems involved in risk and compliance management, general ledger, reporting, and so on.

The major problem here is that the same data is defined differently across systems, with divergent data models and

database schemes. It was with this in mind that the Enterprise Data Management (EDM) Council decided to commission a semantics model and repository for security terms and definitions to help begin to address the aforementioned problems with multiple meanings of data stored in heterogeneous databases. This would then be extended into other areas. Thus, the EDM Council recognized that the major problem facing the industry was not, necessarily, the huge volumes of data, but the different meanings attributed to the real-world objects and data entities that represent them both within and across a multiplicity of organizational information systems. Hence, in order to begin to manage the mountains of data effectively, it was recognized that the first task would be to provide a common language for the industry globally – a semantic approach was, therefore, adopted in order to arrive at unambiguous concept and relationship definitions for all financial industry data. In FIBO, concepts are defined at the business level and represented in OWL. Significantly, FIBO references other standards such as FpML, FIX, ISO, MISMO, MDDL, and XBRL.

The development and application of FIBO, as indeed FIRO, has confirmed, from both business and regulatory perspectives, the relevance of SemTech.

A standards-based approach to capturing regulatory and business vocabularies and rules

Given the ambiguity and complexity of legal and regulatory texts, “natural language processing” (NLP), “machine learning” (ML), and “artificial intelligence” (AI) are not yet up to the task of unpacking regulations. Hence, the lawyer or legal subject matter expert (SME) must bear the burden of responsibility.

Our ground-breaking R&D identified a standards-based approach that helps lawyers and legal SMEs to unpack regulations into both a human-readable and machine-computable format. The core semantic technologies we identified are based on the Object Management Group’s (OMG) semantics of “business vocabulary and business rules” (SBVR) specification – this is a de facto standard. SBVR is a specification for capturing and expressing a business vocabulary (e.g., at base a taxonomy) and business rules in a business natural language. It is grounded in ISO common logic and expresses rules in Deontic and Alethic Logics. SBVR was designed with business SMEs in mind, not computer scientists, who use “controlled natural languages.”

Researchers at the GRCTC build upon SBVR to permit a lawyer or SME capture regulatory semantics and rules in a “regulatory natural language” (RNL). We call this Mercury. This RNL is not the controlled natural language of the computer scientist.

Rather, the RNL is logical, clear, unambiguous, and comprehensible by a computer programmer, while representing the regulatory semantics and rules in a human readable format. It could then be employed by the computer programmer as a specification guiding the technical implementation – avoiding the translation problem. We position Mercury as a potential de facto standard and have opened it accordingly.

An SBVR-compliant semantic repository or knowledge base typically includes a “terminological dictionary” and “rulebook.” The terminological dictionary contains the vocabulary made up of noun concepts and verb concepts but also contains definitional rules that constrain the meaning of the entries. The rulebook is a set of regulatory requirements in the form of behavioral and constitutive rules that capture the regulatory intent of legal texts. We also adapted SBVR and extended it to enable legal experts to perform the interpretation of regulations and capture these using our Mercury RNL. We refer to this extension as Mercury-SE (structured English). This enables the smart storage of legal interpretations in a knowledge base. Our SBVR-based approach also makes it possible for business SMEs to draft business vocabularies and rules on the same platform.

Together, the GRCTC’s FIRO, Mercury-SE, and its related XML schema, Mercury-ML (HgML), are implemented in a web-based software application prototype called Ganesha. This application is developed in Java on the server side, and Angular JS on the client side, and the latter communicates with the server through RESTful APIs, where the vocabulary and rulebook are persisted in SQL, XML, and RDF/OWL (resource description framework/web ontology language 2) data stores.

It is clear from a wealth of industry feedback gained from our field research and views voiced at the recent FCA TechSprint, that standards-based RegTech architectures, such as those described above, are required.

Navigating the digital labyrinth with RegTech

In the Myth of the Labyrinth a Minotaur lay in wait to devour his victims. Ariadne, Mistress of the Labyrinth, helped Theseus overcome the Minotaur by providing him with a sword and a ball of golden thread – the former to slay the Minotaur, the latter to navigate his way through the maze. The myth is instructive given the significant challenge that financial institutions face in navigating through digital structured and unstructured data labyrinths without an Ariadnean Golden Thread to guide them and with the Minotaur of regulatory sanctions lying in wait.

It is evident that many financial organizations are blindly and

SemTech and risk data aggregation

“The foresight required for long-term sustainability of smart data lakes is embedded within a semantic model, which provides conceptual descriptions of data via ontologies and visually represents them, their attributes, and their relationships to other data via graph technologies. These descriptions and different data elements are useful for meta-data management, mapping, and linking data as needed, and provide the foundation for ensuring governance protocols, data discovery, preparation processes, and more. The graph-based model and detailed descriptions of data elements they enable substantially enhance integration efforts, enabling business users to link data according to relevant attributes that provide pivotal context across data sources and business models. The result is considerably decreased time to a more profound form of analytics, in which users can not only ask more questions more expediently than before, but also determine relationships and data context to issue ad-hoc queries for specific needs.”¹⁷

mechanically navigating their way through the digital maze due to the limitations of traditional data management tools and techniques. Organizations cannot solve the problems they created using siloed SQL technologies in a piecemeal fashion by applying yet more SQL-based approaches, which do nothing to semantically enrich data or provide the capabilities to dynamically link it with other siloed internal or external data. Thus, financial enterprises continuously repeat labor-intensive processes of manually curating and integrating regulatory risk and compliance data at significant cost to the bottom line – however, the Minotaur that is BCBS 239 also awaits the unwary and unprepared.

Our research identified how financial organizations can transcend the limitations of siloed SQL data stores and repositories of unstructured data by using standard semantic and No-SQL technologies to virtualize structured data and unlock unstructured data stored in verbatim reports, text fields, and documents; thereby presenting them for semantic querying, inferencing, and in-depth analysis.

¹⁷ Szekely, B., 2015, “Avoiding three common pitfalls of data lakes,” <http://bit.ly/2mARrQ9>

In many organizations, data capture and aggregation processes that integrate structured and unstructured data from multiple siloed sources are imprecise, relatively immature, and lack the exactness to perform good data governance, let alone proper data management for risk assessment. As indicated above, organizations need to navigate a complex digital labyrinth of heterogeneous structured and unstructured data to identify, extract, transform, and load data into a target platform for interpretation, analysis, and reporting. It is standard practice for the majority of firms in the industry to manually curate, cleanse, and reconcile data, typically using spreadsheets, prior to the creation of aggregated management and regulatory compliance reports.

The solution to the problem of the digital labyrinth is technically feasible and practically possible, although there are few players in the market providing comprehensive solutions for the financial services industry. One approach that is receiving much attention is “data virtualization.” This approach provides access to data directly from one or more disparate data sources, without physically moving the data, and presenting it in a form that makes the technical complexity transparent to the end-user. There is broad agreement across industry sectors that semantic metadata is required to make data virtualization and other NoSQL approaches work.

In commenting on extant approaches, Richard Robinson states that “What has been missing is the centralized semantical business context, and intelligent metadata usage to create a tightly coupled, but still independent and flexible, data architecture.”¹⁸ However, while Brian Stein and Alan Morrison of PwC argue that the “means of creating, enriching, and managing semantic metadata incrementally is essential,”¹⁹ there is a general paucity of information on the creation of semantic metadata models. All this certainly provides an opportunity for RegTech companies, particularly in light of the BCBS 239. However, compliance with BCBS 239 aside, there are compelling business drivers for effective data aggregation, which provide additional opportunities for the sector.

Figure 3 presents our proposed solution. While there are many tools to help knowledge engineers create an integrated semantic metadata model, we advise a semi-automated approach that involves the business SMEs building the metadata model according to the Object Management Group’s SBVR standard. Remember, the objective here is to create a common language to express the meaning of organizational data – only then can the apparent heterogeneity of structured and unstructured data be reconciled. RegTech applications can help achieve this if they are designed to help SMEs build both business and regulatory vocabularies and rules.

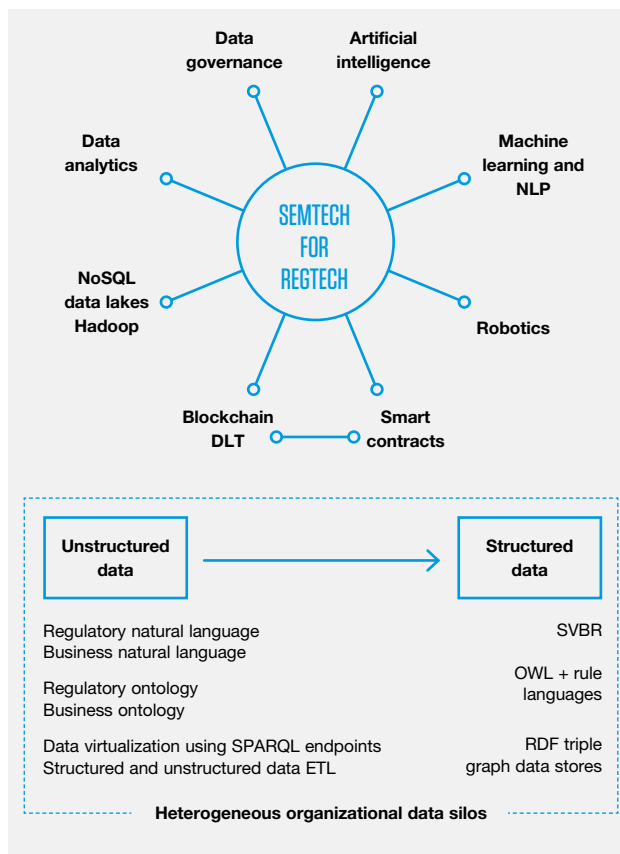


Figure 3 – SemTech for RegTech

We are not alone in arguing that it is the business, and not IT, that needs to take responsibility for its data and the meanings it accords to them. Thus, business needs the tools to semantically enrich its data so that IT can then virtualize it. The next step then is to transform the business meanings to a machine readable semantic data modeling language, such as OWL/RDF. This will then form the basis of the “integrated semantic metadata model” through which the structured and unstructured data may be queried and an associated “risk data knowledge base” populated.

One of the clear benefits of such a model is that the semantic metadata model expressed in both SBVR and OWL/RDF can be linked seamlessly (using URIs) with related semantic models like FIBO and any other standards-based knowledge

18 <http://bit.ly/2mj1g3V>

19 <http://pwc.to/2lmkUwT>

base. In addition, unlike traditional SQL-based approaches, the model can be extended easily. Further, adopting such an approach avoids the double whammy of the fate of Sisyphus and the danger of being lost in the Digital Labyrinth.

USING SEMTECH STANDARDS FOR REGTECH

Figure 3 presents a standards-based model that may be employed as a frame of reference for the development of RegTech solutions. While comprehensive, it requires further elaboration and extension by the industry and its regulators. Support for the model's contention that semantic technology or SemTech provides the necessary and sufficient conditions for RegTech to succeed comes from Mark Robinson,²⁰ who states that "semantic models provide the underpinning of all these technologies. They facilitate the communication between databases, applications, documents and people in extracting data from one point and transferring that message to another point – in a language that can be understood...In the RegTech world, these models can, and have, produce a semantic ontology that links the words used across regulations to describe the specific classes of requirements and how they apply to a particular regulation."

While the above RegTech article references AI, data analytics, distributed ledger technologies, and so on, the left hand side of Figure 3 was developed based on the themes that emerged from the Financial Information Management Europe (FIMA) Conference (November, 2016). It was significant, for example, to find that the industry has yet to derive benefits from data analytics, as fundamental issues of data governance have still to be resolved. Peter Serenita, Group Chief Data Officer (CDO), HSBC, pointed out that the industry had yet to go beyond CDO 1.0 (Governance) to reach CDO 2.0 (Analytics). The panel on the implications of AI, machine learning, and robotics for financial data management confirmed the pivotal role that such technologies will play in the FinTech and, particularly, RegTech domains. However, Adrian Weller, Faculty Fellow, Alan Turing Institute, stated that the real benefits of AI, in terms of unsupervised learning, are still some way off. Nevertheless, it is clear that ontologies, machine learning, and natural language processing technologies are being used effectively in the RegTech space by, for examples, RegDelta, Palantir, and others.

The points being made above by Ben Szekely in relation to the application of SemTech for enhanced data analytics and risk data aggregation in the context of smart data lakes are

highly relevant, as practitioners at FIMA felt that the business benefits were neither clear nor proven. The missing ingredients in this new paradigm are a semantic layer and NoSQL technologies, such as Graph or Triple Stores, as indicated by Gregory Goth.²¹

As expected, blockchain and distributed ledger technologies (DLT) figured greatly at FIMA, as it does at most business meetings and conferences. However, the implications of DLT and smart contracts for RegTech is receiving attention by regulators, such as the Financial Conduct Authority. Similar to what Szekely stated, DLT and smart contracts were proposed as a potential solution for regulatory reporting.

This last point brings us to the right-hand side of the model – the SemTech stack. It was accepted by participants at the FCA's TechSprint event that regulators, lawyers, and business professionals would need an intermediate format – a regulatory, legal, and business natural language to draft smart contracts. Thus, in addition to the arguments made earlier, this is further corroboration of the need for a standard specification, such as the OMG's SBVR, as a basis to develop practitioner-facing controlled natural languages.

One point that needs to be made here is that whether regulation is principles- or rules-based, regulators need to step up and draft regulations and rules in a human and machine readable way. Hence, the upper level of the SemTech stack falls within their area of responsibility. Financial enterprises will need to map these into governance and business policies based on a business natural language.

Both OWL and RDF are knowledge representation languages. An ontology expressed in OWL provides additional semantics for data models and representations, in that knowledge of objects and their relationships is more richly expressed through, for example, axioms. An ontology describes a conceptual model of a problem domain – viewed from another perspective, it contains metadata. Ontologies expressed in OWL may be persisted in the same RDF triple store as related instance data. Thus, both metadata – the ontology – and data – instances of classes/objects – can be queried. In addition, rule languages, such as "Semantic web rule language" (SWRL), may be employed to add expressivity to OWL models.

²⁰ Robinson, M., 2016, "The RegTech marketplace: in depth analysis," <http://bit.ly/2m3J26b>

²¹ Goth, G., 2016, "The data lake concept is maturing," <http://bit.ly/2mTPSc2>

Graph stores based on RDF/RDFS or other languages, while more expressive than a relational data store, may become more expressive, when augmented with ontologies written in OWL, in capturing knowledge about a domain, such as fund management and related topics like regulatory or operational risk.

The power of ontologies is that they enable reasoning or inferencing in RDF triple stores. The advantage here over a graph store, for example, is that a reasoner may be used to infer new/additional triples or relationships – that is add new knowledge – based on the asserted knowledge or axioms about classes and instance data in the ontology. The question here is how to get the data into an RDF triple store from SQL and other data stores, and from unstructured data in text documents or spreadsheets.

There are two approaches:

1. Structured data in relational databases and Excel spreadsheets are extracted, transformed, and loaded (ETL) into an RDF triple store or graph database. Several readily available tools perform this function. Unstructured data from text or XML documents may also be semantically enriched and mapped into an RDF triple store.
2. Structured and unstructured data in relational databases and other sources may be accessed by what is known as SPARQL endpoints. Here, the data stays where it is. An endpoint is a service that permits applications to query a relational database using SPARQL, the RDF query language. Thus, SemTech accesses relational databases as virtual, read-only RDF graphs. SemTechs offer the full power of RDF-based access to data in relational databases without having to replicate it into an RDF store. Thus, for many the preferred solution is not to transform the source data into RDF, but provide the answer to the target semantic query directly from the original source data.

Typically, SemTechs field SPARQL queries, access the relevant data stores, extract the data, transform it into RDF, and then load the RDF data into an in-memory RDF triple store for semantic querying and inferencing. Note that RDF is not the only standard format supported. Notation 3 (also known as N3) is a W3C assertion and logic language that is a superset of RDF. It extends the RDF data model by adding formulae, variables, logical implication, functional predicates, and other features. It is being used instead of RDF for certain applications.

One of the key challenges for RegTech is to transform unstructured data into structured data. SemTech-based solutions for

this are already in use in financial service organizations. NLP technologies may be used to help semantically tag and enrich content and load it into an RDF triple store for querying. SemTechs that also use a combination of machine learning and domain ontologies to query texts as unstructured data are also available. Use cases for RegTech include regulatory change management, risk management, and compliance reporting. Absent regulatory participation at the production end of the “smart regulation,” RegTech has the capabilities to make regulations smart.

It is clear from this brief overview that there is a wealth of approaches that enable standards-based technologies to apply the power of SemTech to achieve the promise of RegTech.

CONCLUSIONS

It is well-accepted that traditional technologies are not up to the task of dealing with the volume, variability, and velocity of unstructured and structured regulatory compliance and risk data. This paper highlighted the urgency for industry and technology standards for RegTech. Without comprehensive standards, RegTech may not be the silver bullet that many perceive it to be in order to help financial enterprises solve the regulatory interpretation problem and enable them to develop compliant business models, processes, and products. Standards are also vital if RegTech is to make regulatory compliance reporting more efficient and effective. Likewise, standards will play a key role if RegTech is to have the ability to help firms perform better data governance and analytics. Standards-based RegTech can bring automation to risk identification assessment and controls, and with enhanced capabilities to detect and prevent breaches of regulatory rules. Perhaps the greatest opportunity for RegTech, however, is to enable regulators to draft smart regulation.

The achievement of these goals is, nevertheless, hampered from the outset. A number of problems exist that may have an impact on the successful adoption and use of RegTech. The first of these is the “translation problem.” Evidence has been adduced to the effect that the translation problem impacts not only the development of RegTech itself, but also the manner in which it is employed to close the gap in regulatory interpretation and understanding. The second problem – the “Tower of Babel” issue – is more important. This refers to the lack of a “common language” in the financial services industry. The lack of progress in arriving at shared business and regulatory terminological dictionaries, thesauri, and taxonomies will not

only imperil successful RegTech initiatives, it will impede the creation of a financial data standard.

This paper discussed how SemTech-based standards can address these problems. It illustrated how RegTech can be used to enable legal and financial industry experts to transform complex legislation, related regulatory rules, and other text containing principles and standards/guidelines into a regulatory natural language (RNL). The same standards can be used to develop business natural language (BNL). The use of SemTech means that both the BNL and RNL are expressed in human- and machine-readable formats.

Of course, regulators and lawyers need to leverage the power of SemTech (ontology-enabled machine learning and NLP) to become more productive. This is the basis for smart regulation, at least from the consumption side of the equation. Such solutions provide a standardized, scalable, systematic approach that overcomes the limitations of current ad-hoc proprietary solutions, which see financial institutions effectively “reinventing the wheel” in terms of understanding regulatory imperatives and developing related governance policies, risk management strategies, and compliance reporting solutions, whenever new legislation is published or regulations applied to industry.

Semantic technologies permit meaning to be embedded in data, whether it is structured or unstructured. RegTech solutions anchored on SemTech standards can facilitate the development and use of standards-based regulatory and business ontologies and their integration with industry standard taxonomies, such as IASB/International Financial Reporting Standards (IFRS) and the U.S. Financial Accounting Standards Board (FASB). As daunting as this task may seem, SemTech can now enable the semi-automatic development and enrichment of both business and regulatory ontologies.

Putting it all together, a combination of SemTech and RegTech can enable regulators and practitioners to achieve the goal of smart regulation, so that they can be more effective and efficient in performing regulatory compliance, and accomplish all data related activities, from aggregation to analytics, in a manner that is compliant with regulations, such as BCBS 239, MiFID II, and so on, and acts as a strategic enabler.

Machine Learning: A Revolution in Risk Management and Compliance?

Bart van Liebergen – Associate Policy Advisor, Institute of International Finance

Abstract

Machine learning and artificial intelligence are big topics in the financial services sector these days. Financial institutions (FIs) are looking to more powerful analytical approaches in order to manage and mine increasing amounts of regulatory reporting data and unstructured data, for purposes of compliance and risk management (applying machine learning as “RegTech”) or in order to compete effectively with other FIs and FinTechs. This article aims to give an introduction to the machine learning field and discusses several application cases within financial institutions, based on discussions with IIF members and technology ventures: credit risk modeling, detection of credit card fraud and money laundering, and surveillance of conduct breaches at FIs.

Two tentative conclusions emerge on the added value of applying machine learning in the financial services sector. First,

the ability of machine learning methods to analyze very large amounts of data, while offering a high granularity and depth of predictive analysis, can improve analytical capabilities across risk management and compliance areas in FIs. Examples are the detection of complex illicit transaction patterns on payment systems and more accurate credit risk modeling. Second, the application of machine learning approaches within the financial services sector is highly context-dependent. Ample, high-quality data for training or analysis are not always available in FIs. More importantly, the predictive power and granularity of analysis of several approaches can come at the cost of increased model complexity and a lack of explanatory insight. This is an issue particularly where analytics are applied in a regulatory context, and a supervisor or compliance team will want to audit and understand the applied model.

INTRODUCTION

In recent years, machine learning and artificial intelligence have seen increasing interest and popularity in the financial services community, as hopes are that they can dramatically improve analytical capabilities and streamline and automate all kinds of business lines including credit underwriting, compliance, interaction with clients, and risk management. The Institute of International Finance (IIF) has previously written about the use of machine learning/AI as “RegTech” in banking, and in the new business models of FinTech.¹

In past years, the amounts of data gathered in financial institutions (FIs) have increased significantly as the detail of reporting requirements has mushroomed and digitalization of services is creating a large amount of high-frequency, unstructured consumer data. As a result, FIs have a clear need for more powerful analytical tools to cope with large amounts of data of all kinds of sources and formats, while maintaining or improving granularity of analysis. Machine learning is widely seen in the financial services sector as a technique that may deliver that analytical power. It is a subfield of statistics that quickly gained prominence in the tech community in recent years. While elements of machine learning go back to the early 20th century, widespread use picked up as computing innovations and greater availability of high-frequency data allowed it to model complex, non-linear relationships, while making machine learning much easier to be applied.

This article aims to shed more light on the concept of machine learning and its uses within financial services: machine learning’s links with other types of statistical analysis, its possibilities, and its limits. It will also briefly touch on deep learning, a form artificial intelligence that has its roots in machine learning. Thereafter, applications within banking will be discussed through three use cases of machine learning: credit risk modeling, detection of fraud and money laundering, and surveillance of conduct breaches and abusive behavior within financial institutions.

BACKGROUND TO MACHINE LEARNING

Machine learning comprises a broad range of analytical tools, which can be categorized into “supervised” and “unsupervised” learning tools. Supervised machine learning involves building a statistical model for predicting or estimating an output based on one or more inputs (e.g., predicting GDP growth based on several variables). In unsupervised learning,

a dataset is analyzed without a dependent variable to estimate or predict. Rather, the data is analyzed to show patterns and structures in a dataset.²

Machine learning is a particularly powerful tool for prediction purposes. By identifying relationships or patterns in a data sample, it is able to create a model incorporating those relationships that lead to the most powerful out-of-sample predictions. Such a model is created by running variables and the model on subsamples of the data to identify the most powerful predictors, and then testing the model on many different data subsamples.³ This can be done thousands of times so that the model can “learn” from the data and improve its predictive performance. Due to its reliance on large datasets and heavy computing power, machine learning is closely associated with the “big data revolution.” In all, “[t]he extraordinary speed-up in computing in recent years, coupled with significant theoretical advances in machine-learning algorithms, have created a renaissance in computational modeling.”⁴

The accuracy of some supervised machine learning approaches is further augmented through their ability to conduct non-parametric analyses, which can flexibly fit any model to estimate the data. This is in contrast to some conventional statistical approaches that start out by making an assumption about the relationship between the dependent and independent variable. Linear regression, for example, assumes that this relationship is linear, while this does not necessarily need to be the case. Several machine learning approaches, in contrast, are also able to infer non-linear relationships, which makes them better able to fit the data.

1 See IIF, “Regtech in financial services: technology solutions for compliance and reporting,” March 2016, at <http://bit.ly/2mD0q39>. Also, IIF, “Digitizing intelligence: AI, Robots and the future of finance” at <http://bit.ly/1WIEbE6>.

2 James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013, *An introduction to statistical learning: with applications in R*, Springer Texts in Statistics. The difference between both methods has also been described as the supervised ML being based on “labeled” data to train the algorithm, while unsupervised ML lacks training data with such labels and has to determine correlations by itself. However, this is the same as having a dependent variable or not: labels in the training data are values of the dependent variable.

3 Large datasets are typically divided into several separate samples to estimate a model (training), to choose the model (validation), and to evaluate how well the chosen model performs (testing).

4 Khandani, E. A. J. Kimz, and A. W. Lox, 2010, “Consumer credit risk models via machine-learning algorithms,” *Journal of Banking & Finance* 34, 2767–2787

Machine learning methods

The machine learning spectrum comprises many different analytical methods, whose applicability varies with the types of statistical problem one might want to address. Broadly speaking, machine learning can be applied to three classes of statistical problems: regression, classification, and clustering. Regression and classification problems both can be solved through supervised machine learning; clustering is an unsupervised machine learning approach.

Regression problems involve prediction of a quantitative, continuous dependent variable, such as GDP growth or inflation. Linear learning methods try to solve regression problems including partial least squares⁵ and principal component analysis; non-linear learning methods include penalized regression approaches, such as LASSO and elastic nets.⁶ In penalized approaches, a factor is typically added to penalize complexity in the model, which should improve its predictive performance.

Classification problems typically involve prediction of a qualitative (discrete) dependent variable, which takes on values in a class, such as blood type (A/B/AB/O). An example is filtering spam e-mail, where the dependent variable can take on the values SPAM/NO SPAM. Such problems can be solved by a decision tree, “which aims to deliver a structured set of yes/no questions that can quickly sort through a wide set of features, and thus produce an accurate prediction of a particular outcome.”⁷ Support vector machines also classify observations, but by applying and optimizing a margin that separates the different classes more efficiently.⁸

In clustering, lastly, only input variables are observed while a corresponding dependent variable is lacking. An example is exploring data to detect fraud without knowing which observations are fraudulent and which not. An anti-money laundering (AML) analysis may nonetheless yield insights from the data by grouping them in clusters according to their observed characteristics. This may allow an analyst to understand which transactions are similar to others. In some instances, unsupervised learning is first applied to explore a dataset; the outputs of this approach are then used as inputs for supervised learning methods.⁹

Table 1 classifies popular machine learning approaches according to their (un)supervised learning character, and the types of problems they can be applied to.

Prediction versus explanation

Machine learning’s ability to make out-of-sample predictions does not necessarily make it appropriate for explanation or

	Linear methods	Non-linear methods	
Problem type	Supervised		
	Regression	<ul style="list-style-type: none"> • Principal components • Ridge • Partial least squares • LASSO 	Penalized regression: <ul style="list-style-type: none"> • LASSO • LARS • elastic nets Neural networks and deep learning
	Classification	Support vector machines	Decision trees: <ul style="list-style-type: none"> • classification trees • regression trees • random forests Support vector machines Deep learning
	Unsupervised		
	Clustering*	Clustering methods: K- and X-means, hierarchical Principal components analysis Deep learning * Since unsupervised methods do not describe a relation between a dependent and interdependent variable, they cannot be labelled linear or non-linear.	

Table 1 – Overview of machine learning methods

inference as well, as statistical methods are typically subject to a trade-off between explanatory and predictive performance. A good predictive model can be very complex, and may thus be very hard to interpret.¹⁰ For predictive purposes, a model would need only to give insight in correlations between variables, not in causality. In the case of credit scoring a loan portfolio, a good inferential model would explain why certain borrowers do not repay their loans. Its inferential performance can be assessed through its statistical significance and its goodness-of-fit within the data sample. A good predictive model, on the other hand, will select those indicators that prove to be the strongest predictors of a borrower default. To that end, it

5 PLS is used to find the fundamental relations between two matrices through linear regression.
 6 LASSO stands for least absolute shrinkage and selection operator. LASSO and elastic nets both perform variable selection, yet apply different types of penalties for model complexity.
 7 Tiffin, A., 2016, “Seeing in the dark: a machine-learning approach to nowcasting in Lebanon,” IMF Working Paper WP/16/56, March
 8 Auria, L., and R. Moro, 2008, “Support vector machines as a technique for solvency analysis,” Deutsches Institut für Wirtschaftsforschung, discussion papers no. 811, Berlin.
 9 Bholat, D., S. Hansen, P. Santos, and C. Schonhardt-Bailey, 2015, “Text mining for central banks: handbook,” CCBS Handbook No. 33, Bank of England
 10 Tiffin (2016)

does not matter whether an indicator reflects a causal factor of the borrower's ability to repay, or a symptom of it. What matters is that it contains information about the ability to repay.

Tackling overfitting: bagging and ensembles

Excessively complex models can also lead to "overfitting," where they describe random error or noise instead of underlying relationships in the dataset. Model complexity can be due to having too many parameters relative to the number of observations.¹¹ In machine learning, overfitting is particularly prevalent in non-parametric, non-linear models, which are also complex by design (and therefore also typically hard to interpret). When a model describes noise in a dataset, it will fit that one data sample very well, but will perform poorly when tested out-of-sample.¹²

There are several ways to deal with overfitting and improve the forecast power of machine learning models, including "bootstrapping," "boosting" and "bootstrap aggregation" (also called bagging).¹³ Boosting concerns the overweighting of scarcer observations in a training dataset to ensure the model will train more intensively on them. For example, one may want to overweight the fraudulent observations due to their relative scarcity when training a model to detect fraudulent transactions in a dataset. In "bagging," a model is run hundreds or thousands of times, each on a different subsample of the dataset, to improve its predictive performance. The final model is then an average of each of the run models. Since this average model has been tested on a lot of different data samples, it should be more resilient to changes in the underlying data. A "random forest" is an example of a model consisting of many different decision tree-based models.

Econometricians can take this concept even further by combining the resulting model with a model based on another machine learning technique. The result is a so-called ensemble: a model consisting of a group of models whose outcomes are combined by weighted averaging or voting.¹⁴ It has been shown that averaging over many small models tends to give better out-of-sample prediction than choosing a single model.¹⁵

A theory-free approach to analysis?

Due to a typical lack of explanatory power and inherent complexity of machine learning models, the discipline has been criticized by some as "a theory-free analysis of mere correlations," which is "inevitably fragile."¹⁶ Machine learning relies on found in-sample (past) correlations to predict out-of-sample (future) correlations, without always offering an understanding of the relationship analyzed. In that sense, it is as much a backward-looking way of prediction as other statistical

approaches. It can only be more accurate at inferring those correlations. However, one observer has noted, "[i]f you have no idea what is behind a correlation, you have no idea what might cause that correlation to break down."¹⁷

Deep learning and neural networks: from machine learning to artificial intelligence

So far, discussion has focused on "classic" machine learning methods that are applied to statistical problems with well-defined and structured datasets. Additionally, machine learning approaches have been advanced and combined to solve all kinds of complex problems, functioning as "artificial intelligence." One of the dominant approaches is deep learning, a learning approach that can be based on both supervised and non-supervised methods; all are non-linear.¹⁸

In deep learning, multiple layers of algorithms are stacked to mimic neurons in the layered learning process of the human brain. Each of the algorithms is equipped to lift a certain feature from the data. This so-called representation or abstraction is then fed to the following algorithm, which again lifts out another aspect of the data.¹⁹ The stacking of representation-learning algorithms allows deep-learning approaches to be fed with all kinds of data, including low-quality, unstructured data; the ability of the algorithms to create relevant abstractions of the data allows the system as a whole to perform a relevant analysis. Crucially, these layers of features are not designed by human engineers, but learned from the data using a general-purpose learning procedure.²⁰

11 For example, R^2 , a goodness-of-fit indicator, tends to increase (and cannot decrease) with any variable that is added to the model, whether or not it makes sense in the context. See Ramanathan, R., 2002, "Introductory econometrics with applications, South-Western

12 James et al. (2013), p. 22.

13 In regression models, overfitting is also mitigated through "ridge regression" or "LASSO," both of which add a factor penalizing complexity from having too many variables.

14 Tiffin (2016).

15 Varian, H., 2014, "Big data: new tricks for econometrics," *Journal of Economic Perspectives* 28:2, 3-27

16 Harford, T., 2014, "Big data: are we making a big mistake?" *Financial Times*, March 28

17 Ibid.

18 Najafabadi, M. M., F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, 2015, "Deep learning applications and challenges in big data analytics," *Journal of Big Data* 2:1

19 LeCun, Y., Y. Bengio, and G. Hinton, 2015, "Deep learning," *Nature* 521:28, 436-444

20 Ibid.

Deep learning is being applied to a wide range of uses. The ability to crunch large amounts of raw data and to identify complex patterns in it makes it particularly well-placed to analyze “big data,” such as the user datasets of tech giants, such as Google, Microsoft, and Amazon.

Given that it was partly developed by the U.S. National Security Agency, it is perhaps unsurprising that deep learning has proved to be very proficient at face recognition and natural language understanding, including question answering and language translation. Upon “overhearing” a discussion, it is able to classify the topic of discussion and the sentiments of the speakers.²¹ While some conventional machine-learning approaches can be equipped to solve non-numeric problems as well (for example, x-means clustering has been applied to text mining), deep learning has often proved to be more accurate. However, a typical deep-learning system is extremely complex and requires a dataset with hundreds of millions of labeled observations only to be trained. In many fields, availability of sufficient data for such extremely large datasets is hardly a given.

Application within financial services

In past years, the amounts of data gathered in financial institutions (FIs) have increased significantly as the details of reporting requirements have mushroomed and digitalization of services is creating a large amount of high-frequency, unstructured consumer data. As a result, FIs have a clear need for more powerful analytical tools to cope with large amounts of data of all kinds of sources and formats, while maintaining or improving granularity of analysis.

After the financial crisis of 2008-09, many new regulations and supervisory measures were introduced that required FIs to report more detailed and more frequent data on more aspects of their business models and balance sheets. Under the new capital regime, banks report large exposures, liquidity measures, collateral, and capital levels. Stress tests are based on all kinds of firm data including loan-level balance sheet data and qualitative aspects of the business model. The Federal Reserve’s CCAR exercise requires FIs to consider the impact of more than 2000 economic variables on their business. For insurers, Solvency II has dramatically increased reporting requirements.

These processes create large amounts of reporting data that need to be well-defined and structured, aggregated across the group, and delivered in-time with supervisors. Regulators have, therefore, introduced numerous initiatives to improve the quality of supervisory data and the ability of financial

institutions to deliver these data. The Basel Committee’s Principles for Risk Data Aggregation (Basel 239) sets standards for G-SIBS to improve their IT systems and reporting structures. IFRS 9 aims to improve the quality of supervisory data.

Apart from reporting data, FIs are increasingly able to gather large amounts of low-quality, unstructured, high-frequency data. These include outputs from consumer apps and other digital interactions with clients, metadata from payment systems, and external data sources, such as social media feeds, which can be mined to gauge insights on market sentiment. This type of data is typically called “big data.”

With practically all aspects of FI’s business model regulated and supervised with detailed risk metrics, running a bank, insurer, or asset manager is increasingly becoming a matter of optimization within hundreds of constraints. To compete effectively, they need to find this optimum while also mining consumer data for detailed insights on client preferences and behavior.

The extensive set of machine learning approaches is well situated to deliver this analytical power in different contexts due to its ability to cope with (or better said, its need for) extremely large datasets and the granularity of analysis. For the mining of high-quality, structured supervisory data, more conventional machine learning techniques are typically applied. To mine high frequency, low quality “big data” sources, Google-like deep learning and neural network techniques are applied, which cope with these data due to their representation learning abilities.

Below, the state of play in three use cases of machine learning is being discussed: the modeling of credit risk, detection of fraud and money laundering, and the detection of conduct risk and abusive behavior within financial institutions.

THREE USE CASES

Credit risk and revenue modeling

Since the early 2000s, an extensive academic literature on the use of machine learning methods to model credit risk has developed. To give just a few examples, Angelini et al. (2007) apply a neural network approach to model SME credit risk on

²¹ Ibid.

a small dataset of Italian SMEs. Auria and Moro (2008) assess company solvency using support vector machines, and find that they produce more accurate out-of-sample predictions than existing techniques. Khandani et al. (2010) apply generalized classification and regression trees (CART) to a large dataset of a commercial bank to build consumer credit risk models. These combine traditional credit factors, such as debt-to-income ratios, with consumer banking transactions, which greatly increases the predictive power of the model.

FIs have traditionally used linear, logit, and probit regressions to model credit risk for capital requirements, stress-testing, and internal risk management procedures.²² Recently, many have started to experiment with the application of machine learning methods to improve financial risk predictions. Unsupervised methods are typically used to explore the data, while regression and classification methods (trees, support vector machines) can predict key credit risk variables as probability of default or loss-given default. Banks normally have extensive records of loan-level data to serve as inputs.

Banks have sometimes also experienced that machine learning can be hard to apply, as methods can be complex and models sensitive to overfitting the data. Thereby, the quality of data within banks is not always fit enough for advanced statistical analysis, while banks are not always able to consolidate the data from across the financial group, among others, due to inconsistent data definitions across jurisdictions and the use of multiple systems. Non-parametric and non-linear approaches (support vector machines, neural networks, and deep learning) and ensembles are so complex that they are practically “black boxes” that are hard, if not impossible, for any human to understand and audit from the outside. That makes these models hardly useful for regulatory purposes, such as the development of internal models in the Basel Internal Ratings-Based approach. Financial supervisors typically require risk models to be clear and simple in order to be understandable and verifiable and appropriate for validation by them.

That does not, however, rule out the use of machine learning to optimize parameters and models with a regulatory function. Linear and simple non-linear machine learning approaches can be applied and still perform better than similar non-machine learning approaches. Machine learning can also be applied to select variables and optimize parameters in existing, linear regulatory models. Khandani et al (2010) stress that CART (tree) models produce easily interpretable decision rules whose logic is clearly laid out, despite their non-linear character. Indeed, there have been examples already of banks applying machine learning in a regulatory context. In a public

example, Citigroup hired an external vendor to build a revenue forecasting model for the 2015 CCAR exercise.²³

Fraud

One area in which machine learning has been applied for more than a decade and with significant success is the detection of credit card fraud. Banks have equipped their credit card payments infrastructures with monitoring systems (so-called workflow engines), which monitor payments for potential fraudulent activity. Fraudulent transactions can then be blocked in real-time. The fraud models used by these engines have been trained on historical payments data.

The high frequency of credit card transactions provides the large datasets required for algorithm training, back testing and validation. Furthermore, since banks are able to verify unambiguously which transactions were fraudulent and which were not, they can construct clear historical data with relevant fraud and non-fraud labels to train classification algorithms. The historical transaction datasets showcase a wide variety of pre-determined features of fraud, which distinguish normal card usage from fraudulent card usage, ranging from features from transactions, the card holder, or from transaction history.

The detection of money laundering and terrorism financing through payments systems stands as a contrast to machine learning’s long-standing record in credit card fraud. Many banks are still relying on conventional rules-based systems, which focus on individual transactions or simple transaction patterns. These systems are often unable to detect complex patterns of transactions or obtain a holistic view of transactions behavior on payment infrastructures. Due to their coarse selection methods, the number of false positives created by these systems is substantial. As a result, significant human capacity is required for the assessment of alerts and filtering false positives from actual suspicious observations. In addition, impediments to data sharing and data usage, as well as long-established regulatory requirements, have complicated innovation in the AML/CFT area.²⁴

22 In a probit model, the dependent variable is binary (can only take two values); in a logit model, the dependent variable is categorical.

23 Ayasdi, “CCAR stress test,” <http://bit.ly/2m5n4y2>, undated; and “After yesterday, CCAR less stressful for Citigroup,” March 6, 2015, <http://bit.ly/2mmbfph>

24 See the IIF’s forthcoming report on the use of “regtech for AML” and submissions to FATF and the BCBS for more information on data sharing issues in AML/CFT on www.iif.com.

Machine-learning systems have the potential to improve detection of money laundering activity significantly, due to their ability to identify complex patterns in the data and combine transactions information at network speed, with data from many other sources to obtain a holistic picture of a client's activity. Indeed, these systems have already been shown to bring false positives down significantly.²⁵

However, application so far in the AML space has lagged for several reasons. First, money laundering is hard to define. There is no universally agreed definition of money laundering and financial institutions do not receive feedback from law enforcement agencies on which of their reported suspicious activities have turned out to be money laundering. It is, therefore, more difficult to train ML-detection algorithms using historical data, because an incidence of money laundering typically is not firmly established. As a second-best, FIs are optimizing ML detection algorithms using lower-level suspicious activity reports as a depending variable for classification – using classification between alerts that the bank could classify as false alerts, and those that moved on to be submitted as SARs to law enforcement agencies.

Unsupervised learning methods are also applied to AML/CFT as they “learn” relevant patterns from the data by clustering transactions or client activity. This yields additional insights, since laundering methods take all kinds of form and develop on a continuous basis.

An example of such unsupervised learning is clustering. Clustering requires large datasets where it can automatically find patterns within the data without the need for labels. Clustering works by identifying outliers as points without any strong membership in any one cluster group, thus finding anomalies within subsets of the data. In AML, clustering is one of the methods used to group together data; using other analytics, such as topological data analytics and dimensionality reduction, machine learning can reduce the significant amounts of false positives often associated with alternative methods.

Surveillance of conduct and market abuse in trading

A third area in which machine learning is increasingly being applied within financial institutions is the surveillance of conduct breaches by traders working for the institution. Examples of such breaches include rogue trading, benchmark rigging, and insider trading – trading violations that can lead to significant financial and reputational costs for FIs. In the last couple of years, automated systems have been developed that monitor the behavior of traders in multiple ways and with increasing accuracy.

The capabilities of the first generation of these surveillance systems were limited to monitoring trading behavior, and only through assessing single trades. However, the improved ability of machine learning approaches to identify large, complex patterns in data has allowed a new generation of systems to analyze entire trading portfolios. These systems are also able to link trading information to other behavioral information of a trader, such as e-mail traffic, calendar items, building check in and check out-times, and even phone calls. Technologies, such as natural language processing (typically based on deep learning) and text mining (which can be based on several learning algorithms²⁶), have made those sources machine-readable and suitable for automated analysis. The outputs of the trading behavior and communications of one or multiple traders are then integrated and compared to a profile of “normal” behavior. When a trader's behavior or trading performance deviates from what is deemed normal, the system will send an alert to the FI's compliance team.

There are several challenges to applying machine learning in this space. First, there are typically no labeled data to train algorithms on, as it is legally complex for financial institutions to share the sensitive information on past breaches with developers. Supervisory learning approaches are, therefore, hard to apply. Second, a surveillance system needs to be auditable for supervisors and for compliance officers, and needs to be able to explain to a compliance officer why certain behavior has set off an alert. For systems that are entirely based on machine learning, that can be difficult due to the “black box” character of learning approaches. In order for an alert to be interpretable and actionable for compliance teams, it should ideally be linked to detection of a specific kind of behavior, rather than based solely on a statistical correlation in the data.

These issues can be addressed at least partly by founding the learning system on a behavioral science-based model, which incorporates human decisions and behavioral traits. In a way, such a model addresses the lack of explanatory power of machine learning approaches. Any alerts from the system will be based on deviations it has identified from the model. However, the inclusion of machine learning approaches on top of the model creates a feedback loop in the system through which it can adapt to evolving behavior, and “get to know” a

25 Adamson, D., 2016, at “Machine learning – the future of compliance?” panel discussion at Sibos conference, September 28

26 Bholat et al., 2015.

trader as it ingests more data. That is a crucial difference with previous rules-based systems, which are unable to tailor their surveillance methods to changed probability distributions and correlations. Consequently, these systems are typically based on more conventional types of machine learning, which can be audited and explained more easily than complex types, such as neural nets and deep learning.

A practical barrier to the implementation of automated surveillance systems is the fragmentation and complexity sometimes found in FI's IT systems. To gain a perspective on a trader's behavior, surveillance systems require information from many sources, which are likely to be found in different systems that can be mutually incompatible or slow to deliver.

CONCLUSION

Machine learning and artificial intelligence are big topics in many fields of business these days, including the financial services sector. FIs are looking to more powerful analytical approaches as they need to manage and mine increasing amounts of regulatory reporting data and unstructured data, either for compliance purposes or in order to compete effectively with other FIs and FinTech's. There seems to be no aspect of the FI business model that is not impacted in some way by machine learning and artificial intelligence: it could improve insights into client preferences, risk management, the detection of fraud, and conduct breaches, and automate client support or allow for automated identity verification when coupled with biometrics.

This article has given an introduction to the machine learning field and has discussed several cases of application within financial institutions, based on discussions with IIF members and technology vendors: credit risk modeling, detection of credit card fraud and money laundering, and surveillance of conduct breaches at FIs. Two tentative conclusions emerge on the use of machine learning in the financial sector – tentative, because the field is developing fast and many FIs are still experimenting with machine learning in some spaces.

First, machine learning comprises a range of statistical learning tools that are generally able to analyze very large amounts of data while offering a high granularity and depth of analysis, mostly for predictive purposes. The ability of some approaches to infer non-linear relationships and to conduct data analysis without making assumptions about the shape or form of the relationship between variables (i.e., non-parametric) increases

the detail with which data can be analyzed and outcomes predicted. Unsupervised approaches allow for exploration of data without a dependent variable. Running algorithms thousands of times on training data and combining models improves their predictive power while limiting overfitting and maintaining analytical granularity.

Such improved, often automated, analytical capabilities allow FIs to gain better insights in business processes such as lending, risk management, customer interaction, and payments. With ever more data produced in these processes, machine learning can discover richer, more complex patterns and relationships as in the analysis of transactions or credit risk, or by connecting different datasets to draw more accurate overarching conclusions, as in the monitoring of conduct breaches.

Second, the application of machine learning approaches within the financial sector is highly context-dependent. Ample, high-quality data for training or analysis are not always available in FIs. More importantly, the predictive power and granularity of analysis of several approaches can come at the cost of increased model complexity and a lack of explanatory insight. This is an issue particularly where analytics are applied in a regulatory context, and a supervisor or compliance team will want to audit and understand the applied model. Fortunately, simpler machine learning approaches do exist, combining non-linear analysis with simplicity. Indeed, vendors of machine learning analytics in finance typically aim to combine machine learning's depth of insight with model simplicity, or add factor models to improve the auditability of their products. As it seems, there is an algorithm for every problem.

Data-centered Dependencies and Opportunities for Robotics Process Automation in Banking

Sandeep Vishnu – Partner, Capco

Vipul Agochiya – Head of Enterprise Data Strategy and Analytics, Bank of the West

Ranjit Palkar – Partner, Capco

Abstract

Across the industry, financial institutions and financial technology (FinTech) companies are exploring the potential for transformative technology by implementing Robotic Process Automation (RPA) to improve service quality, reduce cost, and increase operational efficiency and effectiveness. Essentially, RPA is a software robot that mimics human functions through user interfaces. These robots interpret third-party applications and are configured to execute data and process flows identical to that of a human user. Traditional automated solutions typically require a higher level of programming knowledge; however, RPA solutions can be handled by non-technical business users. A subject matter expert or business operations

employee would walk through the required process on an RPA interface, and code would be generated automatically. RPA uniquely disregards the need for programming skills, unlike traditional business process management tools. A key factor in enabling RPA implementation is the underlying data setup and management. Well-defined data structures are needed for effective automation – the easier the codification, the easier it is to create the underlying data flow, and automate it. This paper explores the centrality of data in enabling RPA, presents frameworks to identify and evaluate candidate-RPA functions, and provides examples of data-centric activities to implement RPA.

RPA AND DATA – A SYMBIOTIC RELATIONSHIP

The Institute for Robotic Process Automation (IRPA) defines RPA as “the application of technology that allows employees in a company to configure computer software or a robot to capture and interpret existing applications for processing a transaction, handling data, triggering responses and communicating with other digital systems.”

Typically, RPA, or “Robomation,” is applicable where there is a high volume of repetitive tasks. Such tasks are generally more prone to human error because of their monotonous nature. “Robomation” serves as a good solution to automate these tasks because it offers:

- Improved efficiency and execution quality of tasks.
- Increased oversight and control while executing tasks.
- Utilization of existing systems/application interfaces.

In many ways, RPA is a data-enabled, machine-centric mechanism for aligning process and technology. It is a component of a spectrum of intelligent automation technology geared to improve service delivery. There is, however, one key distinction between RPA and other automated process solutions: the ability to “do” versus “think.” Systems, such as RPA, are oriented to execute tasks through defined and structured inputs and outputs; on the other hand, artificial intelligence and machine learning serve to “think” in judgment-based processes or solutions. Central to developing this capability is a robust data supply chain that presents appropriately defined, structured, and cleansed data. Data is the fuel that powers the process automation engine.

Since the early 1990s, banks have been increasing their investments in technology and process improvement to harness economies of scale. With the widespread adoption of virtual banking, banks must find innovative ways to deliver the best possible customer experience while trying to minimize cost, follow security standards, and meet regulatory and compliance requirements. Optimizing operations and improving efficiencies means more than just upgrading systems or outsourcing processes – the objective is to improve the speed and accuracy of core business processes, and RPA offers a potential solution to achieve this goal.

Financial institutions generate high volumes of documents across their operations. These are typically managed through a combination of legacy systems, manual processes, and emerging technology, which simultaneously create adoption, integration, and retrieval challenges. Everything from the initial

Barclays uses RPA in its back offices to automate a range of processes, ranging from fraud detection and risk monitoring, to the automation of account opening. This enables Barclays to rapidly scale its ability to process customer requests and growing business needs while maintaining quality. Another example is U.K.-based Co-operative Banking Group which has automated over 130 processes with robotic automation including complex CHAPs processing, VISA chargeback processing, and other back-office processes. Source: Ovum (2015)

application for account opening to deposits, withdrawals, loan documents, and a whole myriad of other day-to-day transactions inevitably generate documentation. Banking professionals struggle to connect the many legacy systems being used to manage and retrieve the information, which at times creates workarounds and inefficiencies. For example, one financial institution found that it was using 10-12 FTEs to collect data from one system, interpret and transform it, and then input it into another system. Substantial M&A activity has only added to this complexity by increasing the number of systems that need to be either linked or streamlined. In all such situations, it is the data flow and the structure of the data supply chain that directly impacts efficiency and effectiveness.

Technology disruption in the economy continues to be widespread, and business and operating models will likely be further disrupted by uncertainty in the geopolitical environment and industrialization of the financial services industry. The rise of virtual banking is making it increasingly difficult for many financial institutions to remain competitive in a saturated market. Customers have more options than ever before, and are demanding the best possible user experience. FinTech is eroding the space previously reserved for banks and financial institutions, who are being challenged to maximize efficiency, ensure the highest possible level of security and data integrity, and do this cost-efficiently. RPA is a powerful and effective mechanism to meet such demands. Michio Kaku suggests that “The job market of the future will consist of those jobs that robots cannot perform. Our blue-collar work is pattern recognition, making sense of what you see. Gardeners will still have jobs because every garden is different. The same goes for construction workers. The losers are white-collar workers, low-level accountants, brokers, and agents.”¹

¹ <http://bit.ly/2mdwHL8>

Benefits

Potential core benefits of RPA include cost-reduction, improved quality, faster outputs, and the ability to integrate with legacy systems. This helps create a more uniform approach to data management without having to start from scratch. Examples include²:

- **Improved operational agility:** robots can be “trained” quickly and hence can respond much faster to changing requirement and business needs.
- **Improved scalability:** quick and easy enterprise level scalability as robots can be scaled up, as well as down, quite quickly as the business needs vary.
- **Increased speed:** the processing speed of the robots is at least 2-3 times higher (may be more in some cases).
- **Improved quality:** more consistent and predictable output. Dramatic reduction in error rate. Another aspect that leads to cost reduction.
- **Improved governance structure:** collaboration between IT and business since IT supports/governs it and business controls it.
- **Improved business planning and forecasting:** “robomation” can make data gathering, organizing, and analyzing much faster and easier, thus helping the organization to plan better for future business needs, trends, and opportunities.
- **Improved compliance:** every action is traceable and available for audit and reporting.

- **Enhanced customer experience:** an automated solution model with 24/7/365 availability.
- **Better labor management:** makes manual workforce available for other non-repetitive or knowledge-based tasks that need judgment/interpretation.
- **Geography independence:** can provide a single, centrally located geography independent solution to businesses which have a global presence.
- **Cost reduction:** automation costs are significantly lower than the costs associated with FTEs. Not only are there payroll and HR savings, it is also possible to reduce infrastructure cost as “robots” do not need space, desks, machines, etc.

EVALUATING CANDIDACY FOR RPA

Process and implementation complexity are two key dimensions to consider when evaluating the potential for RPA, and to understand whether the focus will be at a desktop, enterprise, or at a higher cognitive (thinking) level. Figures 1 and 2 provide a structure to consider RPA potential.

To determine the best uses of RPA, we need to consider the nature of the activity being considered. Figure 2 provides a conceptual framework to understand the applicability of RPA.

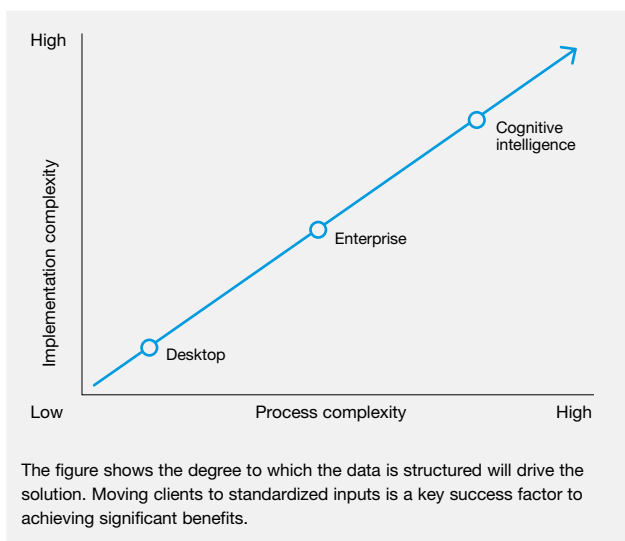


Figure 1 – Process and implementation complexity

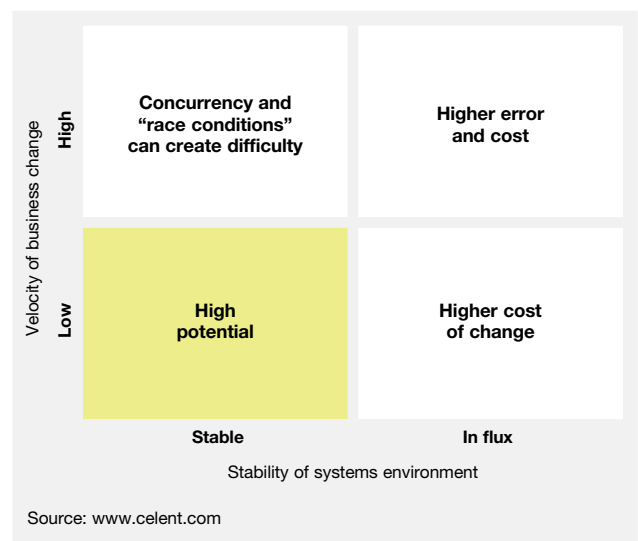


Figure 2 – RPA scenario feasibility matrix

2 Source: www.blueprism.com; www.automationanywhere.com

In essence, RPA works best when the velocity of business change is low and changes to underlying systems are infrequent. If the data is not changing and the system is not changing, robotic process activities can most effectively be applied. The BPO industry has the highest adoption of RPA technology. Many of these processes occur with legacy systems that are not regularly updated, and the input data (format, standards, type, etc.) tends to be relatively static after the initial definition.

Business considerations

Four important considerations with examples of features that may help identify RPA-candidate activities are:

- **System characteristics:** data is entered in three or more applications (more than two duplicate data entry steps); dedicated full-time employees (FTEs) with 50% or higher administrative time for one process cycle; and percentage of manual decisions higher than 15% or fewer than five automatic validations.
- **Human characteristics:** people need to manually enter one or more documents; they print or sign more than three documents and capture one or more; and they make three or more handoffs and two or more back office tasks to complete.
- **Customer experience:** multiple authentications and sign-ins with duplicative information; long wait times for task execution and attention; and inconsistent outcomes.
- **Cognitive integration:** processes have a high degree of machine-to-machine interaction; high repetition of activity and data flow, which can help create learning points for cognitive integration; and processes have sufficiently discrete components to enable experiential learning.

Once potential RPA activities and functions are identified, next steps include:

1. **Capture data entry and review tasks that span multiple platforms:** more channels, products, and codification of systems can lead to more complex customer service or back/front office tasks. RPA robots would enter and review all the captured data required to complete tasks that would involve interaction with multiple systems. This reduces the amount of time required to train staff on multiple systems and allows robots to handle the transition process across platforms.
2. **Articulate processes at a detailed level:** to successfully program a bot, you need to understand where to grab a particular field on a screen, and which events may trigger an action. If a screen changes X and Y coordinates, often-times the robot will not be able to complete the task with

the orientation change. However, cognitive platforms and machine learning allow the robots to reprogram themselves.

3. **Design with practitioners and experts:** programs should be planned in tandem with employees who understand the processes at a granular level. They have key subject matter expertise and know which systems work best and the shortcuts required for certain tasks. However, programming RPA with employees whose tasks are potentially being replaced will require a delicate balance of training and explanation on the use of RPA.
4. **Validate with compliance:** process changes should be validated by business and technology stakeholders, as well as compliance. RPA is efficient in compliance reporting by pulling information from desktops, web-based apps, and core systems. However, RPA programming should keep in mind the regulatory requirements to ensure completeness in addition to efficiency. Replacing human functions with robots provides a new level of risk in terms of understanding what is “right” and “wrong.”
5. **Install airbags:** when connecting RPA to analytics, RPA will require cognitive support. The system should be designed to eventually allow room for machine learning so that the robots may take advantage of their self-correcting process and become more efficient over time. However, robots have the capability of learning “bad behavior” and would require protective cushions and domain knowledge to mitigate the risk of exacerbating incorrect behavior.

By shifting manual tasks from humans to machines, banks have been able to significantly reduce the need for manual intervention, which has had a direct impact on everything from performance and efficiency levels to staffing issues and expenses. With humans at the helm, errors are inevitable – some of which could prove incredibly costly to the institution, both financially and in terms of the bank’s reputation. By automating back-office functions, delays and errors can largely be eliminated, thereby creating a more productive, efficient, and accurate process.

RPA can also help with compliance. Since regulatory changes occur so frequently, banks and financial institutions must stay up to date to accommodate these changes. RPA’s ability to alter and adjust rapidly makes it useful in addressing situations where changes happen frequently.

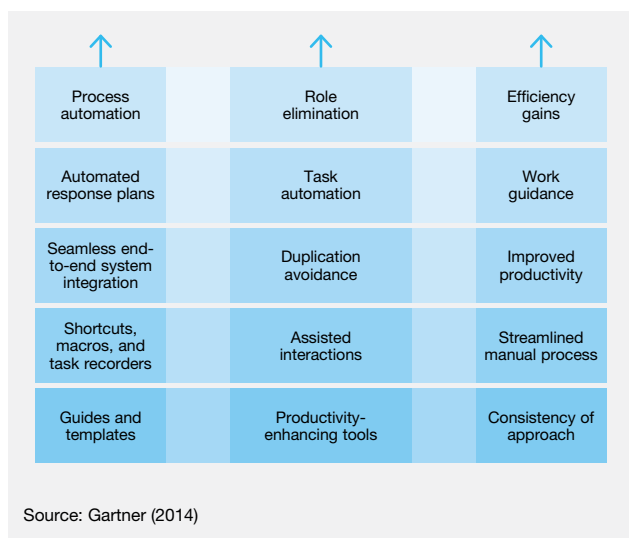


Figure 3 – Building adoption drivers by showing the value of automating processes

FRONT OFFICE	BACK OFFICE
Call center <ul style="list-style-type: none"> • Process information • Pre-segment clients • Address change • Credit card lost 	Branch <ul style="list-style-type: none"> • Automatic calculations of terms and conditions • Sales/recommendations of financial products – cross selling • Applications for loans and deposits
Online/mobile <ul style="list-style-type: none"> • Chat bots • Prefill forms • Capture data to answer instant requests • Voice Payments 	Background processes <ul style="list-style-type: none"> • Form filling • Data validation • Client data updates • Data migrations • Report and auditing • Claims processing • Inactivate/close account • Quality assurance

Source: www.celent.com

Figure 4 – Front and back office applications

RPA IN BANKING

RPA offers different opportunities for harnessing efficiencies and reducing errors across the front-, middle-, and back-office. Figure 4 shows a set of examples of front- and back-office functions that could be RPA candidates.

In general, the high degree of slow and costly back office manual processing present in banks can lead to inconsistent results and high error rates. Thus, there is significant opportunity to increase the levels of automation in back offices and reduce unnecessary errors and costs. By reworking their IT architecture, banks can reduce the size of operational units, and run additional value-adding tasks, such as deal origination and financial reviews. Operations can be improved both by automating specific processes (allowing for the reduction of paper, digitization of work flows, and automation of decision making) and by using IT solutions to manage residual operations that must be carried out manually. According to a report by McKinsey, by taking full advantage of this approach, banks can generate efficiency improvements of over 50% in productivity and customer service.

Certain industry leading banks have already taken key steps towards harnessing the considerable potential of this technology. For instance, one large global bank categorized its 900-plus end-to-end processes into three ideal states: fully automated, partially automated, and “lean” manual. The bank determined that 85% of its operations, accounting for 80% of

its current FTEs, could be at least partially automated. At the time of this analysis, fewer than 50% of these processes were automated at all. According to a study by McKinsey [Lhuer and Willcocks (2016)], almost 50% of current FTE positions could be automated if automation programs were successful.

While this scenario is very attractive, implementation poses a unique set of challenges. The bank mentioned above did some due diligence to determine whether there was a viable business case for process automation in an acceptable time frame. It found that only half of the efficiency gains, measured by the automation business cases performed on manual processes, could be captured. Understanding and validating organizational capabilities will help clarify how robotic automation solutions can fit the needs of banks across various stakeholder interactions, such as:

- **Bank to bank** – new accounts, treasuries management, loan origination, wire transfer, and compliance filing.
- **Bank to consumer** – New accounts, change requests, problem resolution, loan servicing, consumer communications, and marketing.
- **Bank to government** – compliance filings (SEC filings like S1, 10K, 10Q, SARS), student loan servicing, FDIC reporting, NASD filings, and compliance with legislation (Patriot Act, Gramm-Leach-Bliley Act)
- **Bank to employee** – performance management, HR benefits enrollment, employee change of status, procurement applications, and leave/travel requests

Here are some of the areas within financial institutions where RPA can play an important role: data movement and multiple entries (account entry across systems); duplication/data movement (a/c reconciliation); cross-system report generation; eForms (data extraction and systems entries); support for accruals; mortgage approval process (data movement and automatic calculations); alerts and notifications (emails and letters to clients); credit card order processing; fixed asset amortization; foreign exchange/bad debt accounting/write-offs; performing calculations and entries for pricing reviews; account purge activities; client onboarding checks like required documentation; account set up operations across trading, settlement, and other systems; KYC/AML authentication process; legal and compliance process like credit checks, identification checks; data mapping across systems; activity tracking and fraud detection; reconciliation processes; and collection and distribution of payments (dividends, interests...). Examples of successful implementation of RPA in the above functions are shown in Figure 5.

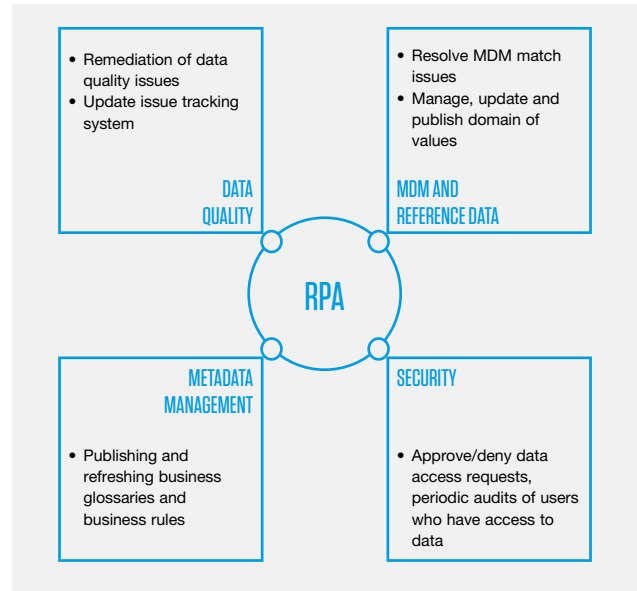


Figure 6 – RPA application within the data governance functions

INSTITUTION	CHALLENGE	IMPLEMENTATION
1 \$100+ billion asset bank United States	<ul style="list-style-type: none"> Migrate loan documents to one system Reduce cost and complexity of migration process 	<ul style="list-style-type: none"> Built RPA bot to take inventory of all content and its source Extracted all data and meta-data and cleaned it according to business process rules Loaded data into ECM system
2 \$100+ billion asset bank Australia	<ul style="list-style-type: none"> Deal with lack of agility in operational processes Handle seasonal variable volumes 	<ul style="list-style-type: none"> Used RPA for institutional and retail banking Automated sub-processes like transaction investigations, tracing funds, recalling funds, audit certificates
3 \$100+ billion asset bank United States	<ul style="list-style-type: none"> Lower call center costs Improve customer satisfaction 	<ul style="list-style-type: none"> Built RPA to work with Alexa open APIs and development toolkits to give direct access to accounts through voice recognition
4 \$20+ billion asset bank United Kingdom	<ul style="list-style-type: none"> Improve customer service Move staff from time consuming activities to customer-facing 	<ul style="list-style-type: none"> Automated 10 processes: automated, direct debt cancellation, account closures, foreign payments, audit reports

Source: www.celent.com

Figure 5 – Examples of successful adoption of RPA in banking and financial services

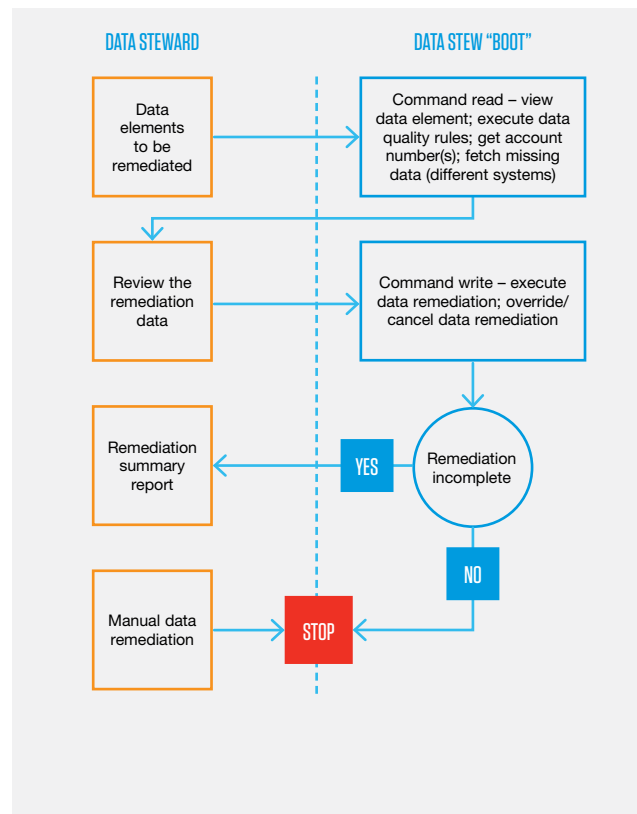


Figure 7 – Illustration of RPA assisting data steward in DQ remediation

DATA FUNCTIONS AND ACTIVITIES CAN BE EARLY ADOPTERS FOR RPA

Key to any successful RPA implementation is the right selection of tasks to automate. When selecting a task to automate, it is important to select those that have a clearly defined rule-based process and are repetitive in nature. One such use case that fits these criteria covers tasks performed by a data steward within the data governance organization. Data stewards in the financial services industry are not only responsible for data management and governance activities, but they also support business and technology users during any regulatory audit. In some cases, their workload can become significant and important tasks do not always get completed on time. A lot of the tasks are repetitive and manual in nature and make a good use case for automation (Figure 6). It is important to note that a RPA “boot” (Figure 7) is not a way to replace data stewards, but more of an automated assistant to take over repetitive manual tasks so that the steward can shift focus on other more important data governance activities which in turn will increase efficiency and reduce cost.

Data-centric activities for RPA

Data preparation and management activities provide good use cases for implementing RPA. Below are some examples of data-centric activities that can serve as early adopters of RPA and provide quick wins.

- 1. Back-office service request:** information is received via email and needs to be transcribed into a structured form for input into other systems
- 2. Loan onboarding data input:** collect information, capture patterns/ interpretations, transform and load data
- 3. Data validation:** improve coverage, speed and accuracy of data quality checks on loan onboarding data
- 4. CCAR reconciliation:** execute scripts, compare results, identify gaps/variances
- 5. New loan validation:** examine, identify, and extract data from pdf files and upload into lending workflow systems requiring structured data input
- 6. Manual file movement:** download and upload files from one directory to another. Manual handling of data received from a third-party source into one repository, and then downloaded, saved, and transferred to another repository
- 7. Preparing and filing regulatory schedules:** download and upload files; extract data, populate work papers, and submit schedules
- 8. Data stewardship:** execute data quality scripts, update metadata repositories, manage domain values, resolve MDM match issues

Using the RPA-candidacy structures discussed earlier, we can examine these 8 use cases as shown in Figure 8.

In both frameworks, activities that involve the manual movement of information from one structure to another emerged as the initial candidates for RPA-focus. The more complex activities are likely to require Cognitive Automation.

Cognitive automation

Cognitive Process Automation (CPA) takes the concept of RPA a step further. If RPA is, at a high-level, about automating repetitive high volume tasks like entering data from one application to another, CPA is more knowledge-based work, like extracting information from unstructured sources, and is all about enhancing decision making.

The cognitive agents act and learn from experience, from human trainers, and even on their own, thereby developing the ability to effectively deal with their environment. While RPA can help with eliminating inefficiencies, CPAs (like machine learning, chat-bot technology, artificial intelligence, natural language processing, big data analytics, evidenced-based learning, computer vision technology, and speech recognition) can help with work requiring judgment and perception. This has the potential of taking RPA to a new level.

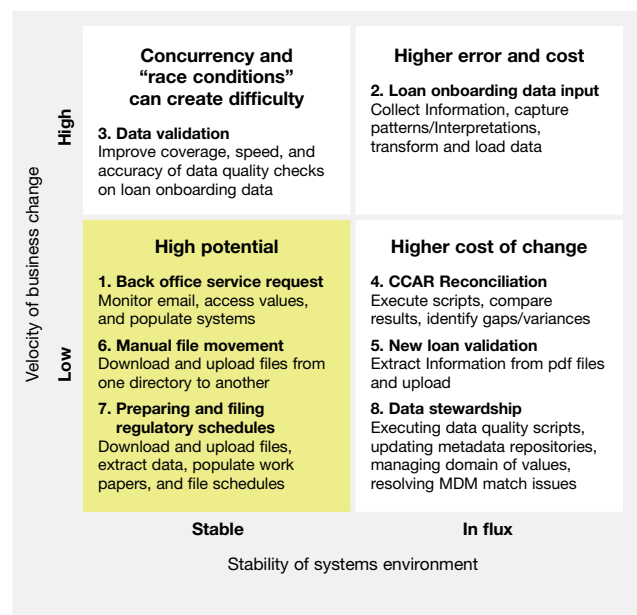


Figure 8 – RPA scenario feasibility matrix

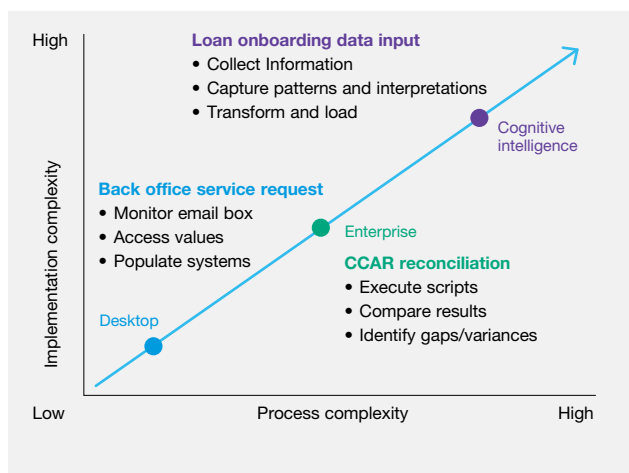


Figure 9 – RPA scenario feasibility

Cognitive RPA has the power and potential to deliver business results, such as greater customer satisfaction and increased revenues, by going beyond basic RPA. For example:

- Machine learning can make predictions about process outcomes by identifying patterns and helping RPA to prioritize actions.
- Unstructured data, like speech audio, text, or images, can be converted into structured information by bots, which can be passed to the next step of the process.

Organizations are just beginning to grasp the use and importance of robotic process automation. Combining RPA with cognitive technologies helps provide a more strategic perspective. One example of cognitive automation and RPA is shown in the document management use case below.

Document process flow use case

1. Collect documents from various sources
2. Process documents (classify, prioritize, etc.) based on type
3. Extract information from documents
4. Information validation against business rules
5. Populate downstream system with data

In this typical flow, RPA and cognitive automation can be combined effectively to increase process efficiency – RPA can be deployed for steps 1 and 5 while cognitive automation be used for steps 2, 3 and 4.

At a leading global bank, cognitive RPA was used to automate its payments business in the area of foreign trade finance. Highly unstructured datasets (comprised of invoices, bills, declarations, certificates, and letters), were one of the main challenges of automating this process end-to-end. A high daily volume of transactions requiring same-day processing, complex business processes, and the need to interface with multiple core systems were some of the other challenges. Instead of employing only RPA and partially automating the process, the bank took the approach of a combined solution. By combining traditional RPA techniques with a host of cognitive technologies that could automate most steps in the process, the bank was able to reduce the number of FTEs required to perform the process by nearly 60%.

Another instance involves a U.S. bank using cognitive RPA to automate its billing system. Like many organizations, it was facing a problem of revenue leakage due to mismatches between rate cards and client invoices. Contracts and client invoices were in paper form or PDFs. Additionally, they were written in multiple languages, which further complicated the matter. The reconciliation between paper documents was labor intensive and prone to error resulting in revenue leakage. The bank utilized NLP (Natural Language Processing) techniques to scan fee schedules and invoices. The bank also translated process requirements into an automated, executable business process workflow, identifying billing opportunities and chargebacks. Through this process, the bank recovered revenue leakage of about 10%.

CONCLUSION

Financial institutions continue to explore and expand their use of technology to improve customer experience and service, drive operational efficiencies and reduce cost – all of which can be addressed through RPA.

RPA is still in its infancy within financial services, however, many activities and functions offer the potential for RPA adoption. In particular, data preparation and data management activities are particularly attractive for RPA implementation because they span the spectrum of automation. Some that involve extraction, transport, and load (ETL) activities offer easy potential for immediate automation. Others that require transcription, interpretation, and synthesis offer the possibility of exploring cognitive automation through Artificial Intelligence and Machine Learning.

As RPA matures, the efficient and effective handling and presentation of data will become increasingly important, if not critical. A comprehensive and robust data supply chain is typically the foundation for enabling machine-to-machine interactions and realizing the potential of automation.

REFERENCES

- Barnett, G., 2015, "Robotic process automation: adding to the process transformation toolkit," OVUM
- Beattie, C., and S. Greer, 2016, "Robo Op: how robotic process automation is applied in banking operations," Celent
- LeClair, C., and A. Cullen, 2016, "Digitization leaders share robotic process automation best practices," Forrester Report
- Lhuer, X., and L. Willcocks, 2016, "The next acronym you need to know about: RPA (robotic process automation)," McKinsey & Co.
- Tornbohm, C., 2016, "Market guide for robotic process automation software," Gartner



Investments

John Bull Can't Stand Two Percent: QE's Depressing Implications for Investment

Do Credit Rating Agencies Inflate Their Ratings? A Review

Available online:

The Power of "Negative Beta": Why Every Portfolio Should Include Private Equity

Downside Risk Protection of Retirement Assets: A New Approach

The Asset Management Industry, Systemic Risk, and Macroprudential Policy

The Role of Asset Owners in the Market for Investment Research: Where Are the Fiduciary Capitalists?

Risk, Data, and the Barcodes of Finance

John Bull Can't Stand Two Percent: QE's Depressing Implications for Investment

Jason M. Thomas – Managing Director and Director of Research, The Carlyle Group¹

Abstract

Much of the existing literature misunderstands “reach for yield” behavior as an increase in risk-taking in response to low interest rates. By focusing on common stocks – where dividend yields are inversely related to systematic risk – I demonstrate that “reach for yield” instead reflects an increase in the marginal utility of current income in response to low interest rates. The monthly returns of a long-short portfolio that buys the highest-yielding 10% of stocks and sells the lowest-yielding decile increase by 1.4% for every 1% decline in two-year interest rates. These effects are three times as large when the decline in interest rates is attributable to a fall in the term

premium, which suggests unconventional monetary policies may generate especially large increases in the marginal utility of current income. By increasing the market value of current income relative to future returns, unconventional policy may lead corporate managers to boost shareholder distributions at the expense of capital accumulation.

¹ The views expressed in this paper are those of the author and do not necessarily reflect the views of The Carlyle Group L.P. or any of its affiliate entities.

INTRODUCTION

Much of the existing literature misunderstands “reach for yield” behavior as an increase in risk-taking in response to low interest rates. I demonstrate that the “reach for yield” instead involves portfolio shifts towards assets that generate more current income. This is an important distinction, as yields and expected holding period returns can differ substantially. When the utility function of the representative investor includes a preference for current income, portfolio choice is not limited to the marginal rate of substitution between mean (expected return) and standard deviation (risk), but also the substitution between assets that offer higher yields today relative to those with higher expected returns over the entirety of the holding period.

Evidence of a “risk-taking” channel of monetary policy comes predominately from fixed income markets where yield is a function of the conditional volatility of returns. A reduction in real interest rates appears to stimulate a willingness among investors to accumulate riskier securities to maintain a return target, but this relationship does not always hold. I demonstrate that the risk-taking channel disappears when the portfolio choice problem is opened to asset classes where yields and conditional volatility are not correlated, like common stocks. Contrary to the predictions of the “risk-taking” channel, investors respond to low rates by increasing exposure to low beta stocks, reducing systematic risk in the search for additional yield.

It is well known that some investors, such as seniors, prefer assets that generate current income (coupons, dividends, rents) to those assets with higher expected returns [Miller and Modigliani (1961)]. I demonstrate that low real interest rates change relative prices in the aggregate by increasing the marginal utility investors derive from current income. I show that the relative price of dividend-paying stocks depends on the level of real interest rates and monthly returns on high-yield stocks vary in response to changes in policy-sensitive Treasury yields. The higher the dividend yield on a portfolio of stocks, the greater the sensitivity of its monthly returns to variation in interest rates. The monthly returns of a portfolio of the highest-yielding 10% of stocks increase by 0.76% for every 1% decline in two-year interest rates, after controlling for known risk factors. A long-short portfolio that buys the highest-yielding 10% of stocks and sells the lowest-yielding decile generates monthly returns that increase by 1.4% for every 1% decline in two-year interest rates.

Interestingly, when the decline in two year rates is attributable to a fall in the term premium, the increase in the return on the long-short portfolio is over three-times as large. Monthly returns on

the long-short portfolio rise by 4.2% for every 1% decline in the term premium, as measured by Adrian et al. (2013). Unconventional monetary policies that suppress term premia, like quantitative easing (QE) and forward guidance, may generate especially large increases in the marginal utility of current income.

If “reach for yield” involves a preference for current income rather than a change in attitudes towards risk, unconventional monetary policy could potentially depress business investment by increasing the market value of shareholder distributions relative to the expected returns from long-lived capital. Some commentators have suggested that unconventional monetary policy makes business managers more inclined to repurchase stock rather than invest in productive capital [Spence and Warsh (2015)]. Unfortunately, explanations for this behavior rely on assumed frictions that somehow make corporate equities less risky than the underlying corporate assets, or generate otherwise inexplicable departures from the standard results of state-based asset pricing models.

I demonstrate that one does not need to rely on fantastical assumptions to understand why unconventional monetary policy may depress business investment. Production-based asset pricing models in the spirit of Cochrane (1991, 1996) make no distinction between real and financial assets. The corporate manager is assumed to pursue an investment policy that maximizes the present value of the stock price of the business, which is tied through arbitrage to the state-based payoffs of its assets. If a negative shock to real interest rates increases the representative investor's marginal utility of current income, the corporate manager would be expected to reduce planned investment in favor of higher current shareholder distributions. Such a result would be consistent with Baker and Wurgler (2004), who find that the decision to pay dividends is driven by investor demand.

“THE REACH FOR YIELD” IN THE LITERATURE

Beginning with Rajan (2005) and Borio and Zhu (2008) researchers have observed that low interest rates provide incentives for investors, banks, and intermediaries to assume incremental risk to achieve nominal holding period return targets. This phenomenon has become known as the “reach for yield,” which Becker and Ivashina (2015) define formally as “the propensity to buy riskier assets in order to achieve higher yields.” From their perspective, the positive relationship between risk and expected return implies that increased demand for higher yielding assets necessarily involves increased risk-taking.

Central Banks, like the U.S. Federal Reserve, closely monitor financial markets for evidence of “reach for yield” behavior. According to Stein (2013), if low policy rates increase investor demand for riskier instruments in finite supply, the expected returns on such assets must fall, which reduces the compensation investors receive for bearing risk and leads to systemic mispricing. Under certain conditions, such mispricing can increase systemic fragility. Yellen (2015) cites the “compression of spreads on high-yield debt” as evidence of dangers introduced by “a reach for yield type of behavior.” Martinez-Miera and Repullo (2015) offer a theoretical model of this phenomenon.

There is a tendency in this literature to conflate “yield” with “expected return.” Perhaps that is because empirical studies tend to focus on fixed income markets like corporate bonds [Becker and Ivashina (2015); Choi and Kronlund (2015)], leveraged loans [Aramonte et al. (2015)], and bank lending [Morais et al. (2015)] where the two concepts are practically indistinguishable. Hanson and Stein (2015) is the rare exception. In their model, a portion of investors care about current portfolio income and respond to a decline in short-term rates by increasing allocations to long-term bonds to keep the total yield on their portfolio from declining “too much.” The buying pressure on long-term bonds increases their price relative to short-term bills, which lowers the real term premium, or compensations investors earn for bearing duration risk.

The existence of yield-oriented investors helps to explain how unconventional monetary policies like QE are transmitted to the real economy. Empirical research finds that by reducing the duration risk borne by private balance sheets, QE shrinks the term premium [Gagnon et al. (2011); Wu (2014); Abrahams et al. (2013)]. Estimates of negative term premiums are not uncommon post-2010 [Adrian et al. (2013)], implying that investors are willing to accept future market value losses, in expectation, to increase current coupon income.

The suppression of risk premia is not a byproduct, or side-effect, of unconventional monetary policy, but rather a conscious objective of the policy [Bernanke (2013)]. To the extent that QE succeeds in reducing risk premia, it should increase investment demand and consumption through a decline in external finance costs [Bernanke and Gertler (1989)]. While QE has been an apparent success in boosting asset prices, the unresponsiveness of business investment to the substantial increase in business net worth has been a puzzle of the post-crisis period.

IS “RISK” A CONFOUNDING VARIABLE?

There is not always such a close correspondence between yields – defined as the current income generated by an asset or portfolio – and expected returns. As a result, yield is not always increasing in conditional volatility (i.e., risk), as observed in fixed income markets. When the portfolio choice problem is opened to more assets and asset classes, one can conceive of any number of ways an investor (or her agent) can augment the current income of a portfolio without an increase in “risk,” whether defined as the portfolio’s total variance or its covariance with the market portfolio or stochastic discount factor. This possibility is largely ignored in the literature. Even Hanson and Stein (2015) restrict their model to two assets, which ensures that an increase in current income can be obtained only through an increase in risk-taking.

It is well understood among practitioners that declines in interest rates increase demand for “yield products,” or securities and funds for which a large share of total returns come through cash distributions. One routinely sees articles in the popular press discussing strategies to combat low yields by diversifying into dividend-paying stocks, MLPs, REITs, leveraged mutual funds and ETFs, and “business development companies” (BDCs) [Norris (2014)]. Implicit to these articles is the understanding that risk-adjusted holding period returns are not the sole determinant of investor utility. Retirees, family offices, endowments, or pension funds often require a certain level of current income to fund retirees’ consumption, cover expenses, or meet legal or investment policy distribution requirements. Low rates are more likely to lead these investors to rethink overall allocation targets rather than simply ramp-up risk-taking in the fixed income portion of their portfolio.

Portfolio rebalancing of this sort does not really concern substitution between “risk” and “return,” but rather an increase in the marginal utility of current portfolio income relative to expected holding period returns.² The sale of an emerging market stock position to finance the purchase of an investment grade corporate bond would likely increase the yield of a portfolio without increasing its variance. A more common

2 Perhaps substitution towards assets that pay a larger share of total returns in dividends is better described as a “search for yield,” as investors seek assets capable of supplementing the decline in coupon income. “Reaching” for yield implies a more conscious decision to assume more risk in the hope of higher returns. Yet, to my knowledge, there is no formal distinction between the two, with “search” and “reach” used more or less interchangeably in the literature. Rajan (2005), Borio and Zhu (2008), Aramonte et al. (2015), Martinez-Miera and Repullo (2015), and Buch et al. (2014), all use “search for yield” to describe the ways low rates influence risk-taking incentives of investors, banks, and other intermediaries.

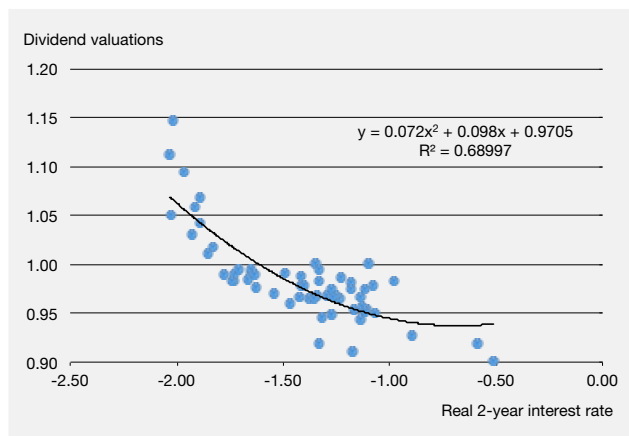


Figure 1 plots the relationship between the relative valuation of dividend stocks and the real two-year interest rate between 2011 and 2015. The relative valuation is the difference between the trailing twelve months' GAAP P/E ratio of the top 100 dividend yielding stocks in the S&P 500 relative to the P/E ratio of the aggregate U.S. stock market. The real two-year interest rate is calculated as the two-year constant maturity Treasury yield net of the annual change in the core consumer price index. Stock data come from S&P Capital IQ Database. The yield data come from the Federal Reserve, H.15.

Figure 1 – Relative valuation ratio of dividend stocks and real interest rates

“yield-increasing, risk-decreasing” portfolio shift would involve the sale of a “high beta” growth stock to finance the purchase of a “low beta” dividend-paying stock. Available evidence suggests these kinds of portfolio shifts happen routinely in response to low rates.

Figure 1 captures the relationship between the relative price of high-yield stocks and real interest rates. The relative price of the dividend stock index – i.e., its trailing P/E ratio scaled relative to that of the S&P 500 – rises nonlinearly as real rates decline. A 100bp decline in two-year real yields is associated with a 7% increase in the relative price of dividend stocks. The price of high-yield stocks responds to the variation in rates, consistent with practitioners’ experience.

What it means to be a high-yield stock in a given year also depends on the level of real interest rates. When sorting stocks annually by dividend yield, the yield of a stock at the 90th percentile of the distribution (i.e., a stock with a dividend yield higher than 90% of other stocks that year) exhibits a sensitivity to changes in the level of real interest rates that is not observed among lower-yielding stocks. Figure 2 plots the sensitivity of dividend yields, sorted by percentile, to annual changes in two-year real interest rates.

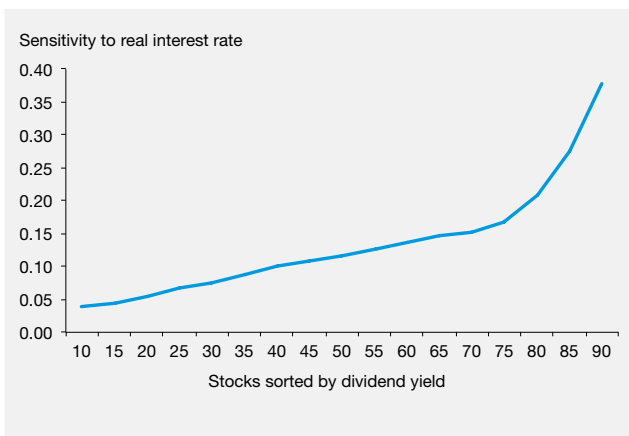


Figure 2 plots the sensitivity of stock valuations to real interest rates. Stocks are sorted into percentiles by dividend yield. The sensitivity is measured by a linear regression of the real two-year interest rate on the dividend yield of a stock at a given percentile. For example, every 1 percentage point decline in the two-year real yield generates a 0.4% decline in the dividend yield of a stock at the 90th percentile of the distribution. By contrast, the yield of a stock at the 10th percentile would decline by just 0.04% in response to a 1% fall in two-year yields. Data come from the CRSP and Federal Reserve (H.15) and cover 1976-2015.

Figure 2 – Stock Valuations and Real Interest Rates

When real two-year yields rise, high-yield stocks appear to fall out of favor with investors and their prices decline (dividend yields rise); when real rates fall, net demand for high-yield stocks increases and their prices rise (dividend yields decline). There is no similar price effect on low-dividend yield stocks, which reinforces that the observed variation is due to shifts in the net demand for current income, not the result of broader changes in discount rates or risk appetite. The dividend yield of the highest-yielding 10% of stocks is about ten-times more sensitive to changes in real interest rates than stocks in the bottom quintile.

THE MARGINAL UTILITY OF CURRENT INCOME IN THE CROSS-SECTION

This tendency for high-yield stocks to appreciate in relative terms suggests that the marginal utility of current income may shift predictably through time in response to real interest rates. If a negative interest rate shock leads to states of the world where the marginal utility of current income is high, assets that appreciate in relative terms following a negative interest rate shock should earn lower returns on average, and vice

versa. That is, a “reach for yield” factor must be priced in the cross-section of assets. Otherwise, the observed preference for yield may disappear in the presence of other factors known to explain returns, or low rates may create arbitrage opportunities for “smart” investors to sell (temporarily overvalued) high-yield stocks, buy (temporarily undervalued) low-yield stocks, and fund current income needs through asset sales.

Shifts in investor preferences for current yield differ from the intertemporal marginal rate of substitution, which relates expected returns to consumption growth. I am not seeking to determine the yield on a portfolio that makes an investor indifferent between saving and consumption. Instead, I focus on the utility derived from that portion of the expected return that comes in the form of cash distributions. Retirees, pension funds, foundations, and other institutions may derive additional utility from current income because of the difficulty in calibrating asset sales (portfolio withdrawals) to fund consumption in the presence of longevity, market, and liquidity risks.

To test whether a “reach for yield” factor is observed in the cross section, I perform ordinary least square regressions on the monthly returns of stocks sorted annually by dividend yield into three, five, and ten portfolios, in addition to a portfolio of common stocks that pay no dividend. Data come from CRSP (via Ken French). I assume the expected return of each portfolio is linearly dependent on four risk factors: the CAPM market risk premium, the Fama-French book-to-market factor or “value premium” (HML), the Fama-French small stock, or “size premium” (SMB), and a momentum factor.

To test whether interest rates provide any residual explanatory power, I add the monthly return on the two-year Treasury note as an independent variable. Interest rate data are obtained through the Federal Reserve (H.15). The gross monthly return is calculated using the reported yield each month R_t as $(1 - R_t)^2 / (1 - R_{t-1})^2$. The two-year is the most “policy-sensitive” Treasury yield, which is both influenced by Fed policy and contains macro information likely to influence such policy [Piazzesi (2005)]. The two-year yield could be thought of as the “connective tissue” that links the money and bond markets and its variation is likely to be especially significant for portfolio allocation decisions.

I also include an independent variable that captures the portion of the monthly two-year Treasury return attributable to the change in the two-year term premium. Monthly estimates of the term premium are obtained through the Federal Reserve Bank of New York website as estimated by Adrian et al. (2013). The return from the term premium π is calculated in the same manner as the returns on the two-year yield $(1 - \pi_t)^2 / (1 - \pi_{t-1})^2$.

Since unconventional policy aims to reduce the term premium, isolating the response of stock returns to variations in that premium may help to identify the impact of unconventional policy on asset prices.

Table 1 reports the results of the regression of the six factors on the monthly returns of four portfolios: (1) non-dividend paying stocks; (2) the lowest-yielding 30% of dividend-paying stocks; (3) the middle 40% of dividend-paying stocks; and (4) the highest-yielding 30% of dividend paying stocks. Tables 2 and 3 summarize the results of regressions of the same six factors regressed on the returns of five and ten portfolios of dividend-paying stocks, respectively, sorted by dividend yield. Finally, Table 4 reports the results from regressions of the same six factors on the returns of four long-short portfolios.

As shown in the Tables, the variation in interest rates influences the returns of high, low, and zero dividend yield portfolios to an economically and statistically significant degree. (A positive “interest rate beta” indicates that returns on the stock portfolio increase when interest rates fall, as the price of the two-year note rises as yields decline.) When controlling for other factors, a 100 basis point decline in two-year yields would be expected to increase returns by 0.76%, 0.63%, and 0.54% for portfolios of the highest-yielding 10%, 20%, and 30% of stocks, respectively. Just as significantly, the same 100bp decline in rates would be expected to reduce the monthly value-weighted return on the zero yield portfolio by 0.67%, and shave 0.65%, 0.46%, and 0.39% off of the returns of the portfolios of the lowest-yielding 10%, 20%, and 30% of stocks, respectively. These data provide clear support for the existence of a “reach for yield” factor that causes demand for high (low) yield assets to increase (decrease) when interest rates fall.

The Tables also reveal that yield does not depend on condition volatility. A portfolio’s market beta declines as dividend yield increases. The zero yield portfolio has a market beta of 1.2, while the highest yielding decile has a beta of just 0.7. Across the ten dividend portfolios, the correlation between the interest rate and the market beta is -0.94. Contrary to predictions of a “risk-taking” channel, a decline in rates in this context induces portfolio shifts that reduce systematic risk. Results in fixed income markets do not seem to be generalizable to broader allocation decisions.

Not surprisingly, no-and-low-yield stocks tend to be smaller (higher SMB beta) and more growth-oriented (lower HML beta). High-yield stocks tend to have a high loading on the value factor (HML) and the HML beta is nearly perfectly correlated with the interest rate beta across portfolios. It is no

	Portfolio			
	No dividends	Lowest 30%	Middle 40%	Highest 30%
Value-weighted portfolio				
Interest rate beta	-0.31 -(3.8)	-0.18 -(2.6)	0.12 (1.8)	0.25 (3.5)
Term premium beta	-0.10 (-0.5)	-0.15 (-0.8)	0.15 (0.8)	0.62 (3.3)
Market beta	1.20 (63.0)	1.12 (70.0)	0.95 (62.3)	0.80 (48.0)
SMB beta	0.56 (19.3)	-0.04 (-1.6)	-0.21 (-9.2)	-0.18 (-7.1)
HML beta	-0.38 (-12.5)	-0.04 (-1.6)	0.19 (8.0)	0.50 (19.0)
Mom beta	-0.09 (-4.1)	0.02 (0.8)	0.09 (5.2)	0.08 (4.2)
Elasticity of monthly portfolio return				
-100bp 2-year yield	-0.67%	-0.39%	0.26%	0.54%
-100bp term premium	-0.21%	-0.30%	0.30%	1.26%
Equal-weighted portfolio				
Interest rate beta	-0.52 -(4.3)	-0.01 (-0.2)	0.15 (2.3)	0.38 (5.7)
Term premium beta	-0.03 (-0.1)	0.35 (1.9)	0.34 (2.0)	0.64 (3.7)
Market beta	1.01 (37.0)	0.98 (60.4)	0.84 (56.8)	0.69 (45.5)
SMB beta	1.22 (29.3)	0.54 (22.0)	0.47 (20.6)	0.42 (18.1)
HML beta	0.11 (2.5)	0.33 (13.1)	0.50 (21.4)	0.58 (24.2)
Mom beta	-0.11 (-3.4)	0.10 (5.4)	0.13 (7.3)	0.09 (5.3)
Elasticity of monthly portfolio return				
-100bp 2-year yield	-1.10%	-0.02%	0.31%	0.81%
-100bp term premium	-0.06%	0.70%	0.70%	1.29%

Table 1 reports the results of regressions of six factors on the monthly returns of four portfolios sorted annually by dividend yield. Data are monthly and come from the CRSP via Ken French and Federal Reserve (H.15) and cover 1976-2015. The gross monthly return is calculated using the reported yield each month R_t as $(1 - R_t)^2 / (1 - R_{t-1})^2$. A positive "interest rate beta" indicates that returns on the stock portfolio increase when interest rates fall, since the price of the two-year note rises as yields decline. Parameters of interest significant at the 5% confidence interval are bolded; t-statistics are in parentheses.

Table 1 – Returns of Four Dividend Yield Portfolios

	Quintile portfolio				
	1	2	3	4	5
Value-weighted portfolio					
Interest rate beta	-0.21 -(2.7)	0.00 (-0.1)	0.13 (1.6)	0.14 (2.0)	0.30 (3.2)
Term premium beta	-0.16 (-0.7)	0.02 (0.1)	0.12 (0.5)	0.21 (1.1)	0.96 (3.9)
Market beta	1.15 (63.5)	1.04 (56.6)	0.95 (49.9)	0.91 (56.5)	0.77 (35.7)
SMB beta	0.02 (0.8)	-0.16 (-5.7)	-0.23 (-8.1)	-0.21 (-8.6)	-0.14 (-4.2)
HML beta	-0.08 (-2.8)	0.08 (2.8)	0.18 (6.0)	0.35 (13.7)	0.60 (17.8)
Mom beta	0.01 (0.4)	0.07 (3.3)	0.08 (3.5)	0.09 (4.9)	0.09 (3.4)
Elasticity of monthly portfolio return					
-100bp 2-year yield	-0.46%	-0.01%	0.29%	0.30%	0.63%
-100bp term premium	-0.31%	0.05%	0.24%	0.42%	1.95%
Equal-weighted portfolio					
Interest rate beta	-0.06 (-0.9)	0.08 (1.2)	0.15 (2.2)	0.26 (3.9)	0.41 (5.4)
Term premium beta	0.28 (1.4)	0.34 (1.9)	0.39 (2.2)	0.34 (1.9)	0.82 (4.1)
Market beta	1.01 (58.7)	0.91 (57.4)	0.85 (55.6)	0.76 (48.3)	0.67 (38.0)
SMB beta	0.55 (21.0)	0.50 (20.7)	0.47 (20.2)	0.44 (18.6)	0.40 (15.1)
HML beta	0.30 (10.9)	0.45 (18.0)	0.50 (20.8)	0.52 (21.1)	0.61 (22.1)
Mom beta	0.09 (4.2)	0.14 (7.4)	0.12 (6.7)	0.12 (6.8)	0.08 (4.0)
Elasticity of monthly portfolio return					
-100bp 2-year yield	-0.14%	0.18%	0.31%	0.56%	0.88%
-100bp term premium	0.56%	0.70%	0.79%	0.68%	1.65%

Table 2 reports the results of regressions of six factors on the monthly returns of five portfolios sorted annually by dividend yield. Data are monthly and come from the CRSP via Ken French and Federal Reserve (H.15) and cover 1976-2015. The gross monthly return is calculated using the reported yield each month R_t as $(1 - R_t)^2 / (1 - R_{t-1})^2$. A positive "interest rate beta" indicates that returns on the stock portfolio increase when interest rates fall, since the price of the two-year note rises as yields decline. Parameters of interest significant at the 5% confidence interval are bolded; t-statistics are in parentheses.

Table 2 – Returns of five dividend yield portfolios

	1	2	3	4	5	6	7	8	9	10
Value-weighted portfolio										
Interest rate beta	-0.30 -(3.1)	-0.11 -(1.3)	-0.13 -(1.3)	0.12 (1.3)	0.02 (0.2)	0.19 (2.1)	0.06 (0.7)	0.20 (2.3)	0.28 (2.9)	0.35 (2.7)
Term premium beta	-0.20 -(0.8)	-0.11 -(0.5)	-0.11 -(0.4)	0.18 (0.7)	0.24 (0.9)	0.06 (0.3)	0.19 (0.9)	0.26 (1.1)	0.31 (1.3)	1.87 (5.3)
Market beta	1.20 (53.2)	1.09 (53.9)	1.08 (48.7)	0.99 (46.9)	1.01 (42.8)	0.92 (43.9)	0.97 (50.4)	0.86 (44.1)	0.81 (36.7)	0.70 (22.9)
SMB beta	0.04 (1.1)	0.01 (0.4)	-0.17 -(5.0)	-0.14 -(4.4)	-0.19 -(5.4)	-0.25 -(7.9)	-0.21 -(7.3)	-0.21 -(7.0)	-0.15 -(4.5)	-0.12 -(2.6)
HML beta	-0.11 -(3.1)	-0.02 -(0.6)	0.05 (1.3)	0.13 (4.0)	0.24 (6.6)	0.18 (5.5)	0.35 (11.3)	0.37 (11.9)	0.51 (14.6)	0.75 (15.6)
Mom beta	-0.01 -(0.3)	0.04 (1.6)	0.02 (0.7)	0.11 (4.6)	0.08 (3.0)	0.06 (2.5)	0.11 (4.9)	0.08 (3.7)	0.07 (2.9)	0.11 (3.1)
Elasticity of monthly portfolio return										
-100bp 2-year yield	-0.65%	-0.24%	-0.27%	0.25%	0.05%	0.40%	0.13%	0.42%	0.59%	0.76%
-100bp term premium	-0.41%	-0.22%	-0.22%	0.36%	0.49%	0.13%	0.38%	0.52%	0.64%	3.79%
Equal-weighted portfolio										
Interest rate beta	-0.17 -(2.0)	0.03 (0.4)	0.10 (1.4)	0.07 (0.9)	0.13 (1.8)	0.16 (2.3)	0.20 (2.8)	0.33 (4.6)	0.33 (4.8)	0.52 (5.0)
Term premium beta	0.26 (1.2)	0.24 (1.2)	0.47 (2.5)	0.21 (1.1)	0.45 (2.3)	0.34 (1.8)	0.35 (1.9)	0.33 (1.7)	0.63 (3.5)	1.03 (3.8)
Market beta	1.08 (54.8)	0.95 (54.0)	0.90 (54.3)	0.91 (53.4)	0.86 (50.9)	0.84 (52.6)	0.78 (48.1)	0.73 (44.5)	0.68 (43.2)	0.64 (26.6)
SMB beta	0.57 (19.0)	0.53 (19.9)	0.52 (20.4)	0.48 (18.3)	0.49 (19.2)	0.45 (18.3)	0.44 (17.7)	0.44 (17.8)	0.40 (16.5)	0.40 (11.0)
HML beta	0.26 (8.6)	0.33 (11.9)	0.41 (15.5)	0.49 (18.1)	0.50 (18.7)	0.50 (20.0)	0.51 (20.1)	0.52 (20.2)	0.59 (23.7)	0.63 (16.7)
Mom beta	0.07 (2.9)	0.10 (4.8)	0.14 (6.9)	0.14 (6.9)	0.12 (5.9)	0.12 (6.6)	0.13 (6.7)	0.12 (6.2)	0.10 (5.5)	0.06 (2.1)
Elasticity of monthly portfolio return										
-100bp 2-year yield	-0.36%	0.07%	0.21%	0.15%	0.28%	0.34%	0.43%	0.70%	0.70%	1.11%
-100bp term premium	0.54%	0.48%	0.95%	0.43%	0.92%	0.68%	0.70%	0.66%	1.27%	2.09%

Table 3 reports the results of regressions of six factors on the monthly returns of ten portfolios sorted annually by dividend yield. Data come from the CRSP via Ken French and Federal Reserve (H.15) and cover 1976-2015. The gross monthly return is calculated using the reported yield each month R_t as $(1 - R_t)^2 / (1 - R_{t-1})^2$. A positive "interest rate beta" indicates that returns on the stock portfolio increase when interest rates fall, since the price of the two-year note rises as yields decline. Parameters of interest significant at the 5% confidence interval are bolded; t-statistics are in parentheses.

Table 3 – Returns of ten dividend yield portfolios

surprise that value stocks tend to be higher dividend payers, on average. These firms tend to have more assets-in-place and greater cash flows. What deserves attention is that the interest rate beta remains statistically significant in the presence of the value factor. The portion of high-yield stock returns unexplained by HML appears related to the "reach for yield" dynamic, as a decline in rates increases the relative price of value stocks that distribute more of their income.

While changes in the term premium only influence the returns on the highest-yielding portfolios, the returns on high-yield stocks are far more sensitive to variation in the term premium than to changes in the expected path for short-term interest rates. If the entire 100bp decline in two-year yields is attributable to a decline in the term premium, the return on the highest-yielding 10% of stocks would be expected to increase by 3.79%, nearly five-times larger than the baseline response of

Long-short portfolio				
	Highest 30 - lowest 30	Highest 20 - lowest 20	Highest 10 - lowest 10	Highest 10 - no dividends
Value-weighted portfolio				
Interest rate beta	0.44 (3.8)	0.51 (3.6)	0.66 (3.5)	0.67 (4.0)
Term premium beta	0.77 (2.6)	1.12 (3.0)	2.08 (4.2)	1.97 (4.5)
Market beta	-0.32 (-12.2)	-0.38 (-11.7)	-0.50 (-11.7)	-0.50 (-13.1)
SMB beta	-0.14 (-3.6)	-0.16 (-3.2)	-0.16 (-2.4)	-0.68 (-11.7)
HML beta	0.54 (13.1)	0.68 (13.3)	0.86 (12.8)	1.13 18.66
Mom Beta	0.07 (2.2)	0.08 (2.1)	0.12 (2.4)	(0.2) 4.54
Elasticity of monthly portfolio return				
-100bp 2-year yield	0.93%	1.09%	1.40%	1.42%
-100bp term premium	1.56%	2.27%	4.20%	3.99%
Equal-weighted portfolio				
Interest rate beta	0.39 (4.9)	0.48 (4.7)	0.69 (5.1)	1.04 (6.7)
Term premium beta	0.29 (1.4)	0.54 (2.0)	0.77 (2.2)	1.06 (2.6)
Market beta	-0.29 (-15.7)	-0.35 (-15.0)	-0.44 (-14.0)	-0.37 (-10.4)
SMB beta	-0.12 (-4.5)	-0.15 (-4.3)	-0.17 (-3.5)	-0.82 (-15.1)
HML beta	0.24 (8.4)	0.31 (8.5)	0.37 (7.5)	0.52 (9.4)
Mom beta	-0.01 (-0.4)	0.00 (-0.2)	-0.01 (-0.2)	0.17 (4.0)
Elasticity of monthly portfolio return				
-100bp 2-year yield	0.83%	1.01%	1.48%	2.22%
-100bp term premium	0.59%	1.09%	1.55%	2.26%

Table 4 reports the results of regressions of six factors on the monthly returns of four long-short portfolios sorted annually by dividend yield. Data come from the CRSP via Ken French and Federal Reserve (H.15) and cover 1976-2015. The monthly return on the portfolio is the difference between the return of the high-yield and low-yield portfolio. The gross monthly return is calculated using the reported yield each month R_t as $(1 - R_t)^2 / (1 - R_{t-1})^2$. A positive "interest rate beta" indicates that returns on the stock portfolio increase when interest rates fall, since the price of the two-year note rises as yields decline. Parameters of interest significant at the 5% confidence interval are bolded; t-statistics are in parentheses.

Table 4 – Returns of four long-short portfolios

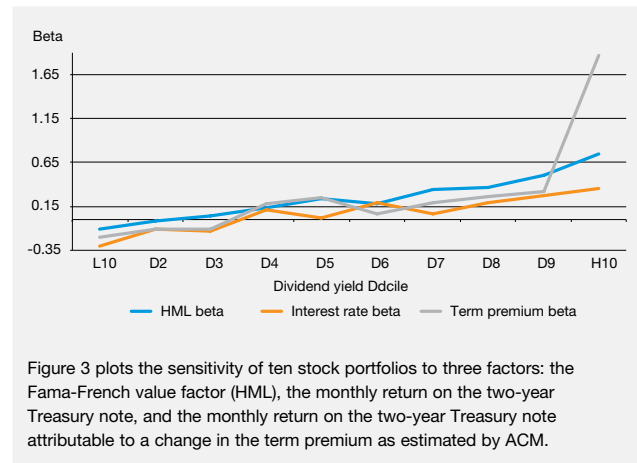


Figure 3 – Return sensitivity of ten dividend yield portfolios to three factors of interest

0.76%. For the highest-yielding 20% of stocks, the expected response is 1.95% or three-times larger; and for the highest-yielding 30% of stocks the expected response is 1.26% or 2.3-times larger than expected for a decline in the two-year yield as a whole.

The results provide strong support for the proposition that the marginal utility of current income increases as interest rates fall and that the relevant price ratios reflect the marginal rate of substitution between states. When rates fall, returns on no-or-low yield stocks decline, after controlling for other factors, as investors sell these stocks, on the margin, to diversify into high-yield alternatives. As high-yield stocks possess greater value in states when the marginal utility of current income is high, average returns are lower, after controlling for other factors, on average. The “reach for yield” involves the substitution between current income and higher expected holding period returns.

The results also suggest that the “reach for yield” is amplified by unconventional monetary policy. Woodford (2012) argues that the term premium depends on investor expectations about the operative monetary policy feedback rule. If QE or forward guidance convinces investors that rates will remain lower for longer, the term premium naturally declines to reflect the diminished risk that incoming data will cause the central bank to tighten policy. The increased probability (at least in a risk-neutral sense) that rates will remain at lower levels increases the marginal utility of current income.

IMPLICATIONS OF A PRODUCTION-BASED ASSET PRICING MODEL

If investor preferences for current income impact asset prices, such preferences should also enter businesses' first order conditions for optimal investment demand [Cochrane (1991)]. Specifically, "reach for yield" behavior should create incentives for businesses to increase distributions (dividends, share repurchases) at the expense of fixed investment because of the higher market value assigned to current income. The "catering" theory of Baker and Wurgler (2004) also anticipates that corporate managers would increase payouts if low rates increase investors' demand for distributions.

To derive a producer's first order conditions, I assume the arrival each period t of an endowment stream (net operating income) y_t and depreciated capital stock δk_{t-1} , which I assume is illiquid and cannot be sold. At the end of period t the firm can either reinvest the endowment stream in additional capital according to $k_t = I_t + \delta k_{t-1}$ or distribute the proceeds to shareholders d_t . In a dynamic setting, the firm wishes to choose an investment plan $\{I_t\}_{t=0}^{\infty}$ to maximize the discounted present value of all future dividends

$$\text{Max } E_0 [\sum_{t=0}^{\infty} m_t d_t], \quad (1)$$

Subject to

$$Y_t = d_t + I_t, \quad (2)$$

And

$$y_t = f(k_t), \quad (3)$$

where m_t is the stochastic discount factor. The depreciated capital the firm inherits from $t - 1$ is a state variable. The investment I_t chosen in period t together with income stream y_t are the control variables whose level determines the production in $t + 1$. The intertemporal separability of the objective function and budget constraints allows (1) to be converted into a two-period problem where the discounted present value of dividends can be expressed in terms of a value function

$$V(k_t) = \text{Max}(d_t + E_t[m_{t+1}V(I_t + \delta k_t)]), \quad (4)$$

Where E_t is the expectation conditional on all information available at time t . I assume returns are normally distributed and investors have standard preferences regarding risk and return. This yields stochastic discount factor

$$m_{t+1} = \beta_t(E_t[R_{t+1}] - \frac{1}{2}\sigma_t^2), \quad (5)$$

with the expected gross return and variance of the firm's investment opportunities represented as $E_t[R_{t+1}]$ and σ_t^2 , respectively. As addressed below, the discount factor $\beta_t \leq 1$ depends on time preference and the marginal utility of current income at time t . When setting $Y_t = 1$ for convenience, and multiplying I_t through (5), the firm's first order conditions become

$$\partial V(k_t)/\partial I_t = 1 = \beta_t(E_t[R_{t+1}] - I_t\sigma_t^2), \quad (6)$$

which is the arbitrage free equation $1 = E_t[m_{t+1} R_{t+1}]$, consistent with production-based asset pricing theory. Simplifying yields, an Euler equation for investment I_t^* equal to

$$I_t^* = E_t[R_{t+1}]/\sigma_t^2 - 1/\beta_t\sigma_t^2 \quad (7)$$

According to (7), the optimal level of investment equals the difference between the expected risk-adjusted return on new investment and the reciprocal of the product of the discount factor and the conditional variance of the firm's investment return. If we assume that β_t is the reciprocal of the gross real interest rate ρ influenced by the central bank, (7) restates the standard neoclassical investment model

$$E_t[R_{t+1}] = \rho + I_t^* \sigma_t^2 \quad (8)$$

The firm continues to invest until marginal product $E_t[R_{t+1}]$ equals marginal cost $\rho + I_t^* \sigma_t^2$. Expected returns are a linear function of the quantity of risk I_t^* and the price of risk σ_t^2 .

Despite a two percentage point fall in real yields ρ and a 55% decline in the VIX – a proxy for the conditional variance of returns σ_t^2 – investment has remained weak while distributions d_t have hit record levels (Figure 4). It may be that low inflation expectations and the effective lower bound on nominal rates combine to keep real rate ρ too high [Summers (2014)]. Alternatively, $E_t[R_{t+1}]$ may have declined markedly due to slower potential GDP growth stemming from a negative productivity shock [Gordon (2015)]. It is also possible that the conditional variance of stock returns differs from that of the returns on the underlying business capital, as posited by Spence and Warsh (2015).

The results in the prior section suggest a fourth possibility: the variation in the marginal utility of current income enters the stochastic discount factor and therefore influences optimal investment policy.

Campbell and Cochrane (1999) introduce a "habit formation"

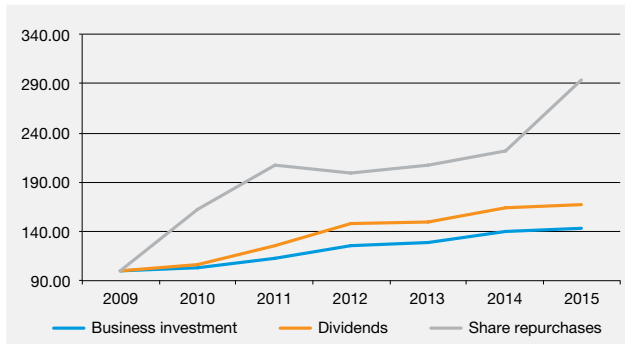


Figure 4 plots the scaled uses of nonfinancial corporate cash flow as measured by F.103 of the Federal Reserve between 2009 and 2015. All data are in nominal terms, scaled to 100 as of June 30, 2009.

Figure 4 – Scaled uses of nonfinancial corporate cash flow, 2009-2015

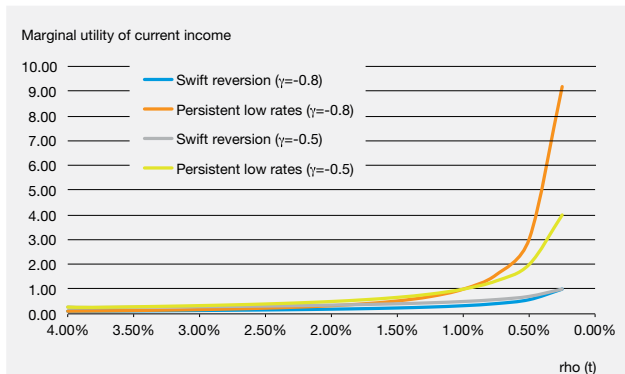


Figure 5 plots the marginal utility of current income estimated from (9) under two scenarios: $E_t \rho_{t+n} = 4\%$ and $E_t \rho_{t+n} = \rho_t$ and two γ parameters: 0.8 and 0.5. In all cases $X_t = X_{t+n} = 1\%$.

Figure 5 – Marginal utility of current income under two scenarios

model where investor utility depends on the difference between current consumption and a “subsistence level” that varies slowly through time. It may be that the marginal utility of current income that determines β_t depends on the level of real interest rates relative to some slow-moving “subsistence yield.” Walter Bagehot’s aphorism, “John Bull can stand many things but he cannot stand two per cent,” captures savers’ presumed refusal to accept low yields. As yields fall to subsistence levels, and are expected to remain there, the prices of high-yield assets adjust upward as investors “reach for yield.”

To formalize this intuition, I assume that investor utility depends on the ratio of the real yield ρ_t relative to a subsistence yield X_t that may evolve slowly through time. I assume that allocation decisions depend not only on yields at time t , but also

on expectations for yields over the entirety of the investment horizon $t + n$. The utility of current income d_t can be expressed as

$$U(d_t) = 1/(1 - \gamma) [(\rho_t/X_t) (E_t \rho_{t+n}/X_{t+n})]^{1-\gamma} \tag{9}$$

where power parameter γ captures the sensitivity of utility to changes in yields relative to subsistence levels. If we assume, as in Hanson and Stein, that only share α of all investors derive utility from the portion of expected returns that comes in the form of current income, β_t in (7) can be expressed as

$$\beta_t = \delta(1 - \alpha)^\lambda, \tag{10}$$

where δ is the subjective time discount factor and λ is the marginal utility of current income calculated from (9). With $r = 1/\delta$, the Euler equation for optimal investment becomes

$$I_t^* = E_t[R_{t+1}] - r(1 - \alpha)^\lambda/\sigma_t^2 \tag{11}$$

Figure 5 graphs estimates of the marginal utility of current income for two values of γ under two scenarios: (1) ρ_t declines but is expected to revert to its prior level as $t \rightarrow t + n$; and (2) ρ_t declines and the negative shock is expected to persist throughout the investment horizon. In both cases, marginal utility rises nonlinearly as yields decline, but the effect is much greater when the negative shock is expected to persist. In that scenario, the magnitude of the decline in marginal utility is squared when $\rho_t < X_t$. The model predicts an especially large increase in marginal utility when sizeable declines in rates interact with the expectation that rates will remain at depressed levels over the entirety of the investment horizon.

The model helps to explain why the returns of high-yield stocks are so sensitive to variation in the term premium. The term premium is the compensation investors earn for the risk that short-term rates may rise faster over the holding period than currently anticipated. Any policy that aims to suppress this risk necessarily involves convincing market participants that rates will remain lower for longer. A decline in the term premium provides information about the persistence of the rate shock, which generates the observed increase in the marginal utility of current income. For this reason, the graphic relationship between the two scenarios in Figure 5 closely resembles that of the “interest rate beta” and “term premium beta” in Figure 3.

Figure 6 graphs estimates of discount factor β_t from (10) for the same two scenarios and for two values of α . The graphs demonstrate the extent to which interest rate shock reduce the expected discounted value of fixed investment projects.

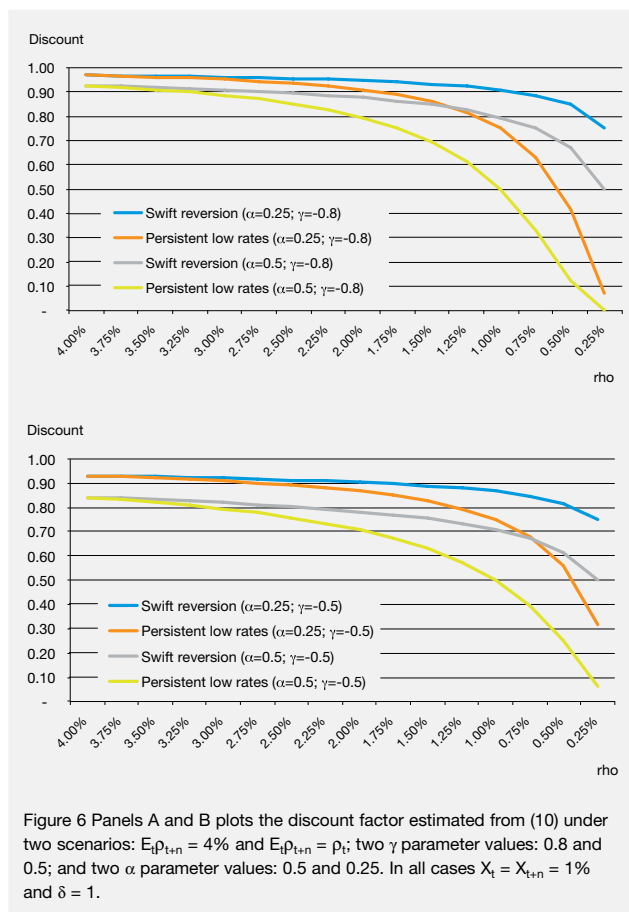


Figure 6 – Implied discount factor

The model predicts that as rates decline, the utility investors derive from illiquid capital declines nonlinearly relative to current income d_t .

When $\alpha = 0.25$, $\gamma = 0.5$, and $2\% < \rho_t - X_t$, a 100bp decline in ρ_t generates a 0.8% increase in $r(1 - \alpha)^{-\lambda}$, the reciprocal of β_t and the effective “hurdle rate” on investment introduced by investors’ preference for current income. When the rate shock is permanent, the effective hurdle rate rises by 1.9%. As $\rho_t \rightarrow X_t$, the same 100bp decline increases the effective hurdle rate by 1.6% when rates revert and by 4.9% when low rates persist. The magnitude of the modeled interest rate response is similar to the results obtained in the empirical section. The predicted variation of the effective hurdle rate – the discount applied to illiquid capital relative to current income – generally tracks the long-short portfolio returns reported in Table 4.

In cases where $\rho_t < X_t$ and yield-oriented investors account for

a large share of the total ($\alpha = 0.5$), the model suggests that the effective hurdle rate on new investment would become nearly insurmountable. The model may have important implications for economies where societal aging has increased the share of investors dependent upon current income to fund consumption in retirement. In these cases, α and γ are likely to be highly correlated, which could render monetary policy ineffective, as investment demand would be expected to fall in response to further declines in rates.

The close relationship between the interest rate and HML betas provides clues about the types of businesses likely to optimize investment in the manner akin to that predicted by the model. Value firms tend to have more assets-in-place, higher depreciation expenses, and greater operating cash flows to distribute to investors. As a result, their investment policy is likely to be more responsive to variation in the marginal utility of current income. By channeling increased distributions into share buybacks (which raise dividend yields by reducing shares outstanding), the firm retains greater flexibility to reduce shareholder distributions in the future when current income is less valued [Jagannathan et al. (2000)]. Growth businesses, by contrast, are generally unable to adjust investment policy in response to negative rate shocks despite the decline in market values.

CONCLUSION

The “reach for yield” is misunderstood. Low rates cause investors to rebalance portfolios towards assets that generate more current income. Portfolio rebalancing of this sort does not really concern substitution between “risk” and “return,” but rather an increase in the marginal utility of current income relative to expected holding period returns. This is an important distinction because “yield” is not an increasing function of conditional volatility when the portfolio optimization problem is opened beyond fixed income. I demonstrate that systematic risk (market beta) actually decreases with yield in the cross-section of stocks. In this case, investors “reach for yield” by bidding up the price of low-beta stocks.

I demonstrate that the marginal utility of current income varies in response to interest rates and term premia: a 100bp decline in the two-year yield increases returns of the highest-yielding 10% of stocks by 0.76%; a 100bp decline in the term premium increases returns on this portfolio by 3.79%. When measured relative to returns on the lowest-yielding 10% of stocks, the increase in returns is 1.4% and 4.2%, respectively.

If business managers seek to maximize the value of their firm's stock price, they will respond to an increase in the relative value of current income by increasing shareholder distributions and reducing investment. I introduce a model where the effective hurdle rate on new investment increases in response to a negative interest rate shock. With plausible parameter values, the model's predictions are close to the observed increase in the relative returns on high-yield stocks. The sharp increase in shareholder distributions relative to investment since the global financial crisis may be partly explained by this phenomenon.

REFERENCES

- Abrahams, M., T. Adrian, R. Crump, and E. Moench, 2013, "Pricing TIPS and Treasuries with linear regressions," FRB of New York Staff Report no. 570
- Adrian, T., R. Crump and E. Moench, 2013, "Pricing the term structure with linear regressions," *Journal of Financial Economics* 1, 110-38
- Aramonte, S., S. J. Lee, and V. Stebunovs, 2015, "Risk taking and low longer-term interest rates: evidence from the U.S. syndicated loan market," *Finance and Economics Discussion Series 2015-068*. Washington: Board of Governors of the Federal Reserve System
- Baker, M., and J. Wurgler, 2004, "A catering theory of dividends," *Journal of Finance* 3, 1125-1166
- Becker, B., and V. Ivashina, 2015, "Reaching for yield in the bond market," *Journal of Finance* 5, 1863-1902
- Bernanke, B., 2013, Speech at the Annual Monetary/Macroeconomics Conference: The Past and Future of Monetary Policy, Federal Reserve Bank of San Francisco, March 1
- Bernanke, B., and M. Gertler, 1989, "Agency costs, net worth, and business fluctuations," *American Economic Review* 1, 14-31
- Borio, C., and H. Zhu, 2008, "Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism?" Bank for International Settlements working papers no. 268
- Campbell, J. Y., and J. H. Cochrane, 1999, "By force of habit: a consumption-based explanation of aggregate stock market behavior," *The Journal of Political Economy* 2, 205-251
- Choi, J., and M. Kronlund, 2015, "Reaching for yield by corporate bond mutual funds," Mimeo
- Cochrane, J. H., 1991, "Production-based asset pricing and the link between stock returns and economic fluctuations," *Journal of Finance* 46 1, 209-237
- Cochrane, J. H., 1996, "A cross-sectional test of an investment-based asset pricing model," *Journal of Political Economy* 104 3, 572-621
- Gagnon, J., M. Raskin, J. Remache, and B. Sack, 2011, "Large-scale asset purchases by the Federal Reserve: did they work?" *Economic Policy Review* May, 41-59
- Gordon, R., 2015, "Secular stagnation: a supply-side view," *American Economic Review: Papers & Proceedings* 2015, 105, 54-59
- Hanson, S., and J. C. Stein, 2015, "Monetary policy and long-term real rates," *Journal of Financial Economics* 115, 3, 429-448
- Jagannathan, M., C. Stephens, and M. Weisbach, 2000, "Financial flexibility and the choice between dividends and stock repurchases," *Journal of Financial Economics* 57, 355-384
- Martinez-Miera, D., and R. Repullo, 2015, "Search for yield," Centre for Economic Policy Research discussion paper no. DP10830
- Miller, M., and F. Modigliani, 1961, "Dividend policy, growth and the valuation of shares," *Journal of Business*, 34 4, 411-433
- Morais, B. J. L. Peydro, and C. Ruiz, 2015, "The international bank lending channel of monetary policy rates and QE: credit supply, reach-for-yield, and real effects," *Federal Reserve Board of Governors International Finance discussion papers no. 1137*
- Norris, F., 2014, "Searching for yield, at almost any price," *The New York Times*, May 1, <http://nyti.ms/2l2vPGZ>
- Piazzesi, M., 2005, "Bond yields and the Federal Reserve," *Journal of Political Economy* 113:2, 311-344
- Rajan, R., 2005, "Has financial development made the world riskier," *National Bureau of Economic Research working paper no. 11728*
- Spence, M., and K. Warsh, 2015, "The Fed has hurt business investment," *Wall Street Journal*, October 26, <http://on.wsj.com/2ltRLhQ>
- Stein, J. C., 2013, Speech at the Banking, Liquidity and Monetary Policy Symposium Sponsored by the Center for Financial Studies, Frankfurt, Germany, September 26
- Summers, L. H., 2014, "U.S. economic prospects: secular stagnation, hysteresis and the zero lower bound," *Business Economics*, 49:2, 65-73
- Woodford, M., 2012, "Methods of policy accommodation at the interest-rate lower bound," *Federal Reserve Bank of Kansas City, The Changing Policy Landscape: 2012 Jackson Hole Symposium*.
- Wu, T., 2014, "Unconventional monetary policy and long-term interest rates," *International Monetary Fund working paper no. WP/14/189*
- Yellen, J., 2015, At the "Finance and Society," a conference sponsored by Institute for New Economic Thinking, Washington, D.C., May 6

Do Credit Rating Agencies Inflate Their Ratings? A Review

Kee-Hong Bae – Professor of Finance and Bob Finlayson Chair in International Finance, Schulich School of Business, York University

Hamdi Driss – Assistant Professor of Finance, Sobey School of Business, Saint Mary's University

Gordon S. Roberts – Professor Emeritus and Senior Scholar, Schulich School of Business, York University¹

Abstract

In this paper, we review the academic evidence on the roles and quality of credit ratings and structure our review around questions that are of interest to academics, professionals, and regulators alike. We review the evidence on how ratings affect market prices and corporate policies and discuss how incentive problems arising from the unique structure of the credit rating industry can adversely affect ratings quality. In particular, our discussion focuses on the issues of conflicts of interest, competition, and ratings shopping and their implications for ratings inflation. Our review identifies opportunities for future research on credit ratings.

¹ Driss gratefully acknowledges support from Saint Mary's University. Roberts thanks the Social Sciences and Humanities Research Council of Canada for support for this research.

INTRODUCTION

The financial crisis of 2007–2009 demonstrated dramatically the shortcomings of ratings for structured products. Severe conflicts of interest, competition among credit rating agencies (CRAs), and ratings shopping by issuers appear to have conspired to create a systematic upward bias in ratings. To illustrate the severity of ratings inflation in structured product markets, White (2010, p. 221) notes that “As of June 30, 2009, 90 percent of the collateralized debt obligation tranches that were issued between 2005 and 2007 and that were originally rated AAA by Standard & Poor’s had been downgraded, with 80 percent downgraded below investment grade.”

Do these severe limitations also apply to the ratings of corporate debt? Despite the travails of bond ratings during the crisis, participants in the corporate bond market continue to attach weight to corporate ratings: corporate bond issuers undertake restructuring to prevent downgrades, and ratings are built into regulations reinforcing investors’ preferences for bonds with investment grade ratings. Are these practices well-founded in light of the evidence of financial research on the role of CRAs, possible conflicts of interest, and possible ratings inflation? Of interest to potential users of ratings – bond issuers and investors – this question is the focus of the current survey article.

We begin in Section 2 with a brief review of the three principal roles that CRAs play in capital markets. First, CRAs produce information about the credit quality of bond issues and issuers. Second, their ratings have a regulatory impact on the investment choices of institutional investors and financial institutions. Third, CRAs play a certification role as possession of a favorable bond rating opens the door for issuers to raise debt funding in capital markets.²

In performing the three roles described above, CRAs may be subject to conflicts of interest arising from their issuer-pays business model, which could produce an incentive to cater to issuers by producing overly optimistic ratings. As will be discussed in Section 3, this potential conflict of interest is constrained by a desire to preserve reputation capital necessary for future business. Significant evidence of catering to issuers has been uncovered in the concentrated and lucrative market of structured products, whereas such catering seems far less prevalent in the less lucrative market of corporate bonds.

Closely related to conflicts of interest is the question of how increased competition from the entry of new CRAs impacts ratings quality. Section 4 addresses this issue beginning with the trade-off discussed earlier between pressures to cater to

issuers versus the desire to preserve reputation capital. Here, the evidence points to a dual role for competition. In the structured products market, increased competition among CRAs is strongly associated with a positive bias in ratings. For corporate bonds, there are results on both sides, and the issue remains open for future research.

In Section 5, we take up the topic of ratings shopping under which issuers seek ratings from several CRAs and select the most favorable. While there is widespread anecdotal evidence that such shopping occurs, the overall take-away from the empirical literature is weak, likely due to limitations of research design. Researchers cannot observe ratings that were requested by issuers but dropped because they were not the most favorable.

In the conclusion, we examine implications for financial best practices.

THE ROLES OF CRAS

CRAs specialize in gathering and analyzing public and private information and offering expert opinions about the creditworthiness of debt securities and their issuers. They play a central role in capital markets by helping to bridge the information gap between investors and issuers. Ratings reveal credit-relevant information that influences the prices of debt securities. Aside from their impact through an information channel, ratings can also affect market prices through a regulation channel because they form a focal point for investment rules and regulations that restrict the investment activities of certain institutional investors. For example, banks and insurance companies typically face higher capital requirements for holding debt securities with lower ratings; mutual funds and pension funds are often restricted in the amount of speculative (BB rating or lower) debt they can hold.³

In this section, we provide a brief review of the literature on the roles of CRAs in capital markets. We begin by reviewing the main findings on their information production function. Next, we discuss key papers investigating the regulatory impact of

2 Driss et al. (2016) review studies on the impact of bond ratings on financing.

3 See, for example, Kisgen and Strahan (2010) for a detailed description of ratings-dependent investment rules and regulations.

ratings. Finally, we review recent papers on the certification role of CRAs – whereby CRAs facilitate issuers' access to capital markets.

The question of whether bond ratings have information value has been the subject of extensive research. An early strand of literature investigates the relation between yield spreads and bond ratings, controlling for issue- and issuer-level characteristics [e.g., Ederington et al. (1987); Liu and Thakor (1984)]. The general picture that emerges from this literature is that bond ratings have power in explaining the cross-section of yield spreads, consistent with the ability of ratings to classify credit risk. Using an event study approach, another strand of literature examines investors' responses to the announcements of rating changes in the context of bond and/or stock markets [e.g., Dichev and Piotroski (2001); Goh and Ederington (1993); Grier and Katz (1976); Griffin and Sanvicente (1982); Hettenhouse and Sartoris (1976); Hand et al. (1992); Ingram et al. (1983); Weinstein (1977)]. The consensus in this literature is that rating downgrades are associated with a statistically significant and economically large negative market reaction, whereas the positive market response to rating upgrades is generally smaller and less significant. For example, using Moody's bond ratings changes between 1970 and 1997, Dichev and Piotroski (2001) report a three-day abnormal average return of -1.97% (0.48%) for downgrades (upgrades). However, there is one important caveat associated with the results of this literature. It is unclear whether the documented market reaction to rating changes exclusively reflects incremental information specific to these rating changes or simply captures concurrent public information that affects market prices.

Two studies attempt to address this identification issue by exploiting exogenous shocks to the information content of ratings. Kliger and Sarig (2000) use Moody's 1982 refinement of its ratings system (e.g., a refinement upgrade from A to A1 or a refinement downgrade from A to A3) and argue that this refinement was not accompanied by any fundamental change in the issuers' risks. They show that "debt value increases (decreases) and equity value falls (rises) when Moody's announces better- (worse-) than expected ratings". In another study, Jorion et al. (2005) exploit the implementation of Regulation Fair Disclosure (Reg FD) on October 23, 2000 to study the information content of credit ratings. Reg FD prohibits U.S. public companies from making selective disclosure of non-public information to market participants, such as institutional investors and equity analysts. However, an exemption was granted to CRAs, which allowed credit analysts to have access to confidential information that is no longer made available to other market participants. Consistent with the information function of CRAs,

Jorion et al. find that both rating downgrades and upgrades have become more informative following Reg FD.

Ratings can also affect market prices through a regulatory channel, as shown in Kisgen and Strahan (2010) and Bongaerts et al. (2012). In 2003, DBRS was certified by the Securities and Exchange Commission (SEC) as a Nationally Recognized Statistical Ratings Organization (NRSRO), thereby qualifying DBRS's ratings to be used in ratings-based investment rules and regulations. Kisgen and Strahan (2010) examine issuers whose ratings were in place prior to DBRS's certification and find that when DBRS rated bonds higher than Moody's and S&P, the rated issuers' cost of debt declined, particularly around the investment grade boundary of BB+/BBB-. However, in cases in which a DBRS's rating was the same or lower than those of competing CRAs, no yield impact occurred. Because the effect only works in one direction, i.e., when the DBRS's rating is higher, it is consistent with a regulatory effect, but is inconsistent with DBRS enjoying better reputation following its SEC's certification. Bongaerts et al. examine multiple credit ratings and, in particular, the role of Fitch as the third opinion provider after Moody's and S&P. They document that on average, Fitch's ratings are more optimistic, consistent with earlier research by Cantor and Packer (1997). Focusing on the demand for multiple ratings by issuing firms, Bongaerts et al. find support for their regulatory certification hypothesis: for cases in which Moody's and S&P's ratings are split and on opposite sides of the investment grade boundary of BB+/BBB-, a Fitch's rating acts as a tiebreaker and likely improves the issuer's standing with regulators.

Several studies provide evidence that CRAs play the role of certifiers in credit markets, thereby facilitating firms' access to debt financing. Driss et al. (2016) examine Moody's issuer-level credit watches with direction downgrade over the period 1992 to 2014 and offer evidence consistent with the certification role of CRAs. A credit watch with direction downgrade occurs when a CRA observes a deterioration in a rated issuer's credit quality and announces that it is monitoring the issuer with a plan either to confirm the rating if the deterioration is reversed or to downgrade the issuer. In other words, during a credit watch a CRA apparently influences a borrower to address issues weakening credit quality and assigns a confirmed rating if such actions occur. Since a confirmed rating constitutes a certification of the borrower by the CRA, credit watches afford an opportunity to study whether such certification works in practice to facilitate access to borrowing. Driss et al. find that in 27% of the cases Moody's confirmed the issuer's rating after an average watch period of 142 days, and in the remaining 73% of the watches the issue was downgraded after a mean period of 93 days.

They examine corporate characteristics for four quarters before and after the watch period. Supporting the view that confirmed ratings constitute valuable certification, firms with confirmed ratings after the watch period achieve significant growth in long-term debt financing and total investment expenditures – a finding that does not apply to downgraded firms. Further, firms with confirmed ratings outperform firms with downgraded ratings exhibiting higher operating income and return on assets ratios after the watch period. In addition, this effect is stronger for firms with greater financial constraints or enhanced information asymmetry, indicating that CRA certification is more valuable to such borrowers.

Other prominent studies on CRA certification include Sufi (2009), who exploits the introduction of syndicated bank loan ratings by Moody's and S&P in 1995 and shows that CRA certification has real effects on corporate financing and investment policies. Specifically, Sufi finds that the introduction of bank loan ratings caused an increase in the use of debt financing and investment activities of the firms that obtain a rating. Tang (2006) uses Moody's 1982 refinement of its ratings system as in Kliger and Sarig (2000) to show that firms with higher refined ratings (e.g., refinement of Baa rating to Baa1 as opposed to Baa3) have better access to credit markets and invest more capital, consistent with a CRA certification effect. Finally, Faulkender and Petersen (2006) show that firms with a bond rating choose significantly higher levels of debt financing than non-rated firms, indicating that CRA certification can effectively facilitate firms' access to debt financing.

In summary, the literature provides consistent evidence that credit ratings influence market prices not only because they contain credit-relevant information but also because they restrict institutional investment choices through ratings-based investment rules and regulations. Further, the certification services of CRAs facilitate firms' access to credit markets and have a real positive impact on the economy. Despite these valuable functions, CRAs have come under intense scrutiny, particularly in the wake of the 2007-2009 financial crisis. They have been accused of exacerbating the financial crisis and misleading investors by offering unduly favorable ratings to highly risky mortgage-related securities. Below, we discuss how conflicts of interest and competition among CRAs can affect the quality of credit ratings and potentially result in inflated ratings.

CONFLICTS OF INTEREST

What does economic theory tell us about the role of information producers? In an ideal world, information producers would provide valuable and unbiased information to help alleviate information asymmetries about economic variables whose values are unknown ex-ante [Diamond (1984); Millon and Thakor (1985); Ramikrishnan and Thakor (1984)]. Central to this result is the assumption that information producers have a compensation scheme that depends on the quality of the information produced. In reality, CRAs are compensated differently. Most CRAs adopt the issuer-pays business model, whereby they earn their income from the issuers seeking ratings for the securities they sell. This compensation structure creates an obvious conflict of interest and provides CRAs with an incentive to cater to issuers to attract business. Offsetting this potential conflict of interest, CRAs have their reputation capital at stake. They face an incentive to provide unbiased ratings to avoid irrevocable damage to their reputation in the long run. The trade-off faced by CRAs between maintaining reputation to increase future rents and catering to issuers to increase current rents is the subject of several papers.

Covitz and Harrison (2003) examine one important way that CRAs could cater to issuers – delaying rating downgrades. This delay preserves the issuer's cost of funding, avoids possible covenant triggers, and gives the issuer time to restore its credit quality. However, it is possible that investors have the ability to foresee and incorporate delayed downgrades into their bond pricing, in which case we would observe higher spreads prior to any CRA action. To test the degree to which CRAs act in the interest of issuers by delaying downgrades, two cases in which the incentive to delay could be significant are examined: (a) the case of a CRA receiving important fees from large issuers with many bonds outstanding and (b) the case in which a potential downgrade moves a bond's rating from the investment grade to the high yield category (a fallen angel). Using a database of around 2,000 rating changes by S&P and Moody's between 1997 and 2002, including 773 upgrades and 1,234 downgrades, Covitz and Harrison find that anticipation is significantly less for large issuers and fallen angels. These results are not consistent with the conflicts of interest hypothesis but rather support the view that reputation effects dominate and are in keeping with popular views about CRAs during the 1990s: "Indeed the major complaint about the rating agencies during this era was not that they were too compliant to issuers' wishes but that they were too tough and too powerful. This view was epitomized by the New York Times columnist Thomas L. Friedman's remarks in a Public Broadcasting System (PBS) "News Hour" interview on February 13,

1996: ‘There are two superpowers in the world today in my opinion. There’s the United States, and there’s Moody’s Bond Rating Service. The United States can destroy you by dropping bombs, and Moody’s can destroy you by downgrading your bonds. And believe me, it’s not clear sometimes who’s more powerful’ [Quoted in White (2010)].”

The issuer-pays model emerged in the 1970s. According to White (2010), there were several reasons for this shift from the earlier approach of investor-pays model. The first reason arose from changing technology: with the introduction of Xerox machines, investors did not need to buy ratings manuals but could substitute photocopies. This made it more attractive to sell ratings to issuers. Further, when Penn Central Railroad went bankrupt in 1970, bond issuers became focused on the need for ratings to underwrite the quality of their debt, and CRAs sought to capture the resulting rents. Jiang et al. (2012) exploit a quasi-natural experiment and examine how S&P’s ratings for corporate bonds changed when S&P made the shift in its business model in 1974. Benchmarked against Moody’s, which was already working on an issuer-pays basis, S&P’s ratings were lower before the shift and became similar to Moody’s’ thereafter. Consistent with significant conflicts of interest in the issuer-pays model of CRAs, Jiang et al. find that S&P’s ratings increased more for lower credit quality bonds or bonds issued by larger and more frequent bond issuers.

Using a sample between 1999 and 2009, Strobl and Xia (2012) compare U.S. corporate bond ratings from two competing CRAs with different business models: S&P with an issuer-pays approach versus Egan-Jones Rating Company (EJR), which collects its fees from investors. The authors identify three measures of conflicts of interest faced by S&P and examine the impact of such measures on ratings inflation measured by the extent to which S&P’s ratings are more optimistic. Their first metric for conflicts of interest is the amount of outstanding short-term debt, with larger values indicating greater likelihood of future debt issues that could bring more business to the CRA. A second measure is based on past ratings business and its concentration with S&P. Issuers with lower concentration may be more likely to switch CRAs, and thus the researchers hypothesize that S&P is more likely to issue favorable ratings to retain their business. Third, conflicts of interest may be heightened for issuers with recently-appointed CEOs and CFOs, as such firms may be more likely to switch CRAs. Consistent with the predictions, Strobl and Xia find that each of their three measures of conflicts of interest is associated with significantly higher ratings by S&P, averaging one-fifth of a rating notch. Further, the research finds no evidence that investors adjust for S&P’s rating bias in their bond pricing,

indicating that they are unaware of S&P’s incentives to assign friendly ratings due to conflicts of interest.

Market concentration could create an incentive for a CRA to inflate ratings [Frenkel (2015)]. In a market, such as structured products, with few issuers that repeatedly interact with a CRA, severe conflicts of interest could influence the CRA to seek a private reputation for leniency with issuers. Unlike the corporate bond market, which has many individual issuers, the structured products market is an oligopoly dominated by a few investment banks working closely with CRAs in designing the products [White (2010)]. These securitizing investment banks could exert pressure on CRAs for higher ratings that would facilitate the profitable distribution of structured products, as well as threaten to shift business to a competing CRA if displeased. In support of this view, Griffin and Tang (2012) document positive adjustments to credit ratings for CDOs by a major CRA. They show that such adjustments were quite common with only a 0.49 correlation between the percentage of published AAA ratings and the percentage of AAA ratings that would be obtained under the CRA’s credit risk model. These adjustments and other forms of over-optimism by CRAs on structured products led to 60 percent of global structured products being rated AAA in 2007, while only 1 percent of corporate issues qualified for the top rating [Coval et al. (2009)].

In the market for corporate ratings, a similar effect occurs in which CRAs provide more favorable ratings for long-standing customers. Mählmann (2011) finds that ratings inflation for corporate bonds increases with the duration of the relationship between a CRA and a rated firm over the sample period 1986-2005. To illustrate, his study shows that an issuer with an 11-year relationship with a CRA enjoys a rating of approximately 0.6 notches higher than that of a similar firm with only a one-year relationship with the CRA. Follow-up tests on default rates show that this higher rating is not associated with stronger credit quality; on the contrary, bonds issued by companies with long-standing relationships with CRAs have higher credit risk in yield spread tests. The research identifies a “dark side” to long-standing relationships between issuers and CRAs in that ratings may be inflated.

Other researchers have also identified a motivation for a CRA to become more generous with ratings. Mathis et al. (2009) model CRA incentives in the face of reputation effects. Limited to a monopoly CRA, but taking into account reputation effects, they identify reputation cycles in which CRAs become more optimistic over time and predict that CRA incentives to be truthful weaken when there is more business in a given product. As expressed by White (2010): “It is not surprising

that the members of a tight, protected oligopoly might become complacent and less worried about the problems of protecting their long-term reputations". He et al. (2012) provide empirical support for this prediction. They find that over the period 2000-2006 large issuers of mortgage-backed securities (MBS) received AAA ratings for greater proportions of their issues and that these larger issues had poorer ex-post performance, as measured by subsequent price drops.

Overall, there is significant evidence that CRAs catered to issuers by providing favorable ratings for structured products leading up to the recent financial crisis. In the corporate bond market, however, catering to issuers appears to be far more moderate.

COMPETITION

Since the onset of the 2007-2009 financial crisis, the issue of whether competition among CRAs improves the quality of credit ratings has often been raised in policy debates. In economics, a competitive product market generally leads to lower prices and/or better product quality. Despite the virtues of competition, its impact on ratings quality is conceptually unclear due to the peculiar structure of the credit rating industry. Most CRAs adopt the issuer-pays business model, whereby they are paid by the firms they rate. The conflict of interest inherent in this business model gives issuers an incentive to shop around for favorable ratings. In turn, this issuer behavior pressures CRAs into offering friendly ratings to quickly ramp up their market shares.

Offsetting this potential conflict of interest, CRAs have reputation concerns. They face an incentive to provide unbiased ratings to avoid damage to their reputations in the long run. The trade-off faced by CRAs between maintaining reputation to increase future rents and catering to issuers to increase current rents makes the impact of competition among CRAs on ratings quality ambiguous. On the one hand, competition can reinforce the disciplining role of reputation due to the potential loss of market share in the future, which increases reputational costs and leads to improvement in ratings quality. On the other hand, competition may erode future profits leading to more focus on short-term profits and thereby result in greater ratings inflation.

Consistent with the latter view, several theoretical studies show that CRAs are more likely to issue inflated ratings in response to competitive pressure [Bolton et al. (2012); Camanho

et al. (2012)]. Other studies further show that CRAs' incentives to issue inflated ratings are stronger when (a) the rated securities are complex [Mathis et al. (2009)], (b) ratings are issued during a boom period [Bar-Issac and Shapiro (2013)], (c) the incentives for regulatory arbitrage exist [Archarya and Richardson (2009); Acharya et al. (2013); Opp et al. (2013)], or (d) issuers shop around for favorable ratings [Bongaerts et al. (2012)]. Frenkel (2015) presents a theory predicting that a CRA has an incentive to inflate ratings in a market with few issuers who repeatedly interact with the CRA (a concentrated market). In a concentrated market, such as the market of structured products, conflicts of interest are so severe that a CRA may be incentivized to develop a private reputation for rating leniency among issuers.

Consistent with these theories, ratings inflation has been shown to be particularly serious in the structured products market, wherein CRAs have systematically made upward adjustments beyond their rating models to gain market share. Numerous studies provide compelling evidence on upward biased ratings in the structured bond markets [Ashcraft et al. (2010); Baghai and Becker (2016); Benmelech and Dlugosz (2009); Cohen and Manuszak (2013); Coval et al. (2009); Griffin et al. (2013); Griffin and Tang (2012); He et al. (2011, 2012)].

While the evidence on the negative impact of competition on ratings quality in the structured bond markets is well-documented, there remains ongoing controversy over whether competition affects ratings quality in the corporate bond market. If ratings inflation exists in the corporate bond market, its extent could differ from that in the structured bond markets for several reasons. First, unlike structured bonds, corporate bonds are simple in their features and their ratings are mainly determined by issuers' fundamentals. These differences make the rating process of corporate bonds more transparent and easier for investors to understand than that of structured bonds. Consequently, ratings inflation in the corporate bond market, if it exists, can be more easily detected, making CRAs less likely to inflate their ratings for corporate bonds. Second, while the corporate bond market has many issuers, the structured bond market is dominated by a small number of large financial institutions. The corporate bond market tends to be less concentrated than the structured bond market, so the bargaining power of corporate issuers against CRAs is smaller than that of issuers in the structured bond market. These arguments imply that ratings inflation should be more prevalent in the structured bond than in the corporate bond market.

The empirical evidence on the relation between competition and ratings quality in the corporate bond market is limited and

mixed. Supporting the view that competition among CRAs results in ratings inflation, Becker and Milbourn (2011) find that Fitch's market share – their measure of the competitive pressure Fitch exerts on the two incumbent CRAs, S&P and Moody's – in a particular industry is positively correlated with the incumbents' ratings for firms in that industry during the period 1995–2006. They interpret this result as evidence that increased competition from Fitch led S&P and Moody's to offer friendly ratings.

Challenging the findings by Becker and Milbourn (2011), Bae et al. (2015) find no relation between Fitch's market share and ratings. They argue that Fitch's market share is subject to an endogeneity problem. Since Fitch's market share is an industry-level variable capturing Fitch's presence in a particular industry-year, the regression model likely suffers from an omitted variable problem in that certain industry characteristics affect both Fitch's market share and credit ratings. In support of their argument, they find that the positive effect of Fitch's market share on the level of ratings issued by S&P and Moody's disappears once the endogeneity bias caused by unobservable industry effects is controlled for. This result suggests that competition does not cause ratings inflation.

The size of a CRA may play a role in how it trades off upholding a reputation for quality ratings against catering to issuers with friendly ratings. For instance, the current profits foregone from lost market share are likely bigger than the loss in reputation capital in the future for a small CRA facing severe competition from larger CRAs. In a recent paper, Bae et al. (2016) test this prediction using Canada-based DBRS, which competes against the big three U.S. CRAs (S&P, Moody's, and Fitch). They show that competition from the big three in the Canadian corporate rating market appears to incentivize DBRS to assign favorable ratings to Canadian bonds over a sample period from 2004 to 2012. The competition effect is particularly stronger for bonds of issuers relying heavily on debt financing, having greater concerns about their ratings, or for which DBRS faces stronger conflicts of interest. Their credit spread analysis shows that for Canadian bonds, investors are less responsive to DBRS's ratings than to the U.S. CRAs' ratings, particularly when competition from the U.S. CRAs is intensive. Their evidence supports the view that reputation concerns are not an effective disciplining mechanism for small CRAs facing competitive pressure from their larger peers.

A few articles document the benefits of competition on ratings quality. Doherty et al. (2012) examine the market for insurance ratings. Using S&P's entry into the insurance ratings market previously covered by a monopolist, A.M. Best, they test the

impact of entry (i.e., competition) on the quality of ratings. They find that S&P required higher rating standards to assign a rating similar to the one assigned by A.M. Best and that higher-than-average quality insurers in each rating category of A.M. Best chose to receive a second rating from S&P. Their findings indicate that insurers of the same quality received a lower rating by the new entrant relative to the incumbent, suggesting that CRAs deflate rather than inflate their ratings in response to competitive pressure.

In a similar vein, Xia (2014) examines how the quality of ratings issued by an incumbent issuer-paid CRA (S&P) responds to the entry of Egan-Jones Rating Company (EJR), an investor-paid CRA. He finds a significant improvement in the quality of S&P's ratings following EJR's rating initiations: S&P's rating levels are shifted downward, ratings are more responsive to market-based risk measures, and S&P's rating changes are associated with stronger market reactions. These findings suggest that increased competition among CRAs in fact improves ratings quality.

Overall, there is strong evidence that competition among CRAs led to a systematic upward bias in the ratings of structured products. By contrast, the literature offers conflicting results on the impact of competition in the corporate rating market. These inconclusive results suggest that the impact of competition on ratings quality in the corporate bond market remains an open empirical question that warrants further research.

RATINGS SHOPPING

Ratings shopping is a practice whereby an issuer solicits ratings from multiple CRAs and then selects the CRA(s) that will assign the most favorable rating(s) relative to its (their) competitors. There is a widespread belief among market participants, investors, and regulators alike that ratings shopping has been a pervasive practice in the credit rating industry, particularly for structured finance products. Anecdotal evidence abounds. For example, Luchetti and Ng (2010) note that "Real-estate investment firm Redwood Trust Inc. approached two credit-rating firms early this year to rate a new mortgage-bond offering. It was an important deal, the first of its kind in two years. One of the firms, Standard & Poor's, expressed reservations about parts of the deal. Redwood chose Moody's Investors Service – and in April sold more than \$200 million of bonds carrying Moody's top rating of triple-A, without a hitch." In another example, "Fitch Group's new chief executive said Credit Suisse Group AG dropped the firm's rating from a mortgage-backed

security because Fitch took a harsher view than two rivals that assigned triple-A ratings to the deal.” [Neumann (2012)].

Both ratings shopping (by issuers) and competition among CRAs can result in inflated ratings. However, unlike the competition story, ratings shopping does not necessarily require that CRAs be subject to incentive problems that may cause them to inflate ratings. Ratings shopping reflects selection bias in ratings caused by issuers’ strategic choices rather than by perverse incentives on the part of CRAs. Even when CRAs maintain their ratings standards and offer unbiased ratings on average,⁴ ratings inflation may arise because of the ability of issuers to obtain ratings from multiple CRAs and then to select and disclose the most favorable rating(s). In this scenario, the published ratings will be upward biased relative to the ratings we would observe in a shopping-free market.

The extant literature attempts to answer a number of important questions regarding ratings shopping. How relevant and pervasive is this phenomenon? How does it relate to the complexity of rated assets or the business cycle? Do firms benefit from engaging in ratings shopping? Do investors account for ratings shopping in their pricing of rated assets? In what follows, we seek to answer these questions in the context of the existing theoretical and empirical literature.

Numerous theoretical studies model ratings shopping in different settings. Skreta and Veldkamp (2009) develop an equilibrium model in which investors do not rationally account for an upward bias in published ratings due to the ability of issuers to choose among potential raters. They show that an issuer’s incentive to shop strengthens with the complexity of rated assets – when the potential for disagreement on the same rating for a given asset is greatest among CRAs. Their theory predicts that ratings shopping is pervasive and that ratings exhibit a systematic bias in markets of complex credit products, such as mortgage-backed securities (MBSs) and collateralized debt obligations (CDOs), but is somewhat limited in traditional bond markets.

Bolton et al. (2012) develop a rich theoretical model in which ratings inflation emerges from a sufficiently high fraction of naïve investors who take ratings at face value. The model shows that in equilibrium, two distortions of market efficiency may occur. First, the presence of multiple CRAs (e.g., a duopoly versus a monopoly) facilitates ratings shopping, which in turn may lead to ratings inflation. Second, ratings inflation is more likely to occur during economic booms as more investors are likely to accept ratings uncritically and/or when CRAs are less likely to be concerned about their reputation costs in

the form of lower future profits associated with ratings mistakes. This prediction is consistent with the theoretical modeling by Mathis et al. (2009) and empirical results in Ashcraft et al. (2010).

More recently, Sangiorgi and Spatt (2016) present a model of ratings shopping without making the restrictive assumption of naïve investors. They show that even under rationality, ratings inflation can emerge when an issuer selectively discloses to the market a subset of the solicited ratings. In turn, selective disclosure of ratings causes uncertainty in the market about whether there are any undisclosed ratings, which results in inefficient investment decisions and misallocation of resources in the economy.

The empirical evidence on ratings shopping in the corporate bond rating market is mixed. Becker and Milbourn (2011) investigate changes in the quality of bond ratings from S&P and Moody’s in response to the material entry of Fitch to the competitive landscape during the 1995-2006 period. Although their overall results point to a competition story as discussed above, they present one piece of evidence that lends support to the presence of ratings shopping in the corporate rating market. They find that Fitch tends to rate bonds with a low rating from the incumbents, suggesting that issuers disappointed with their existing ratings from S&P and/or Moody’s tend to solicit an additional rating from Fitch.

In an early study, Cantor and Packer (1997) document that bond ratings from Fitch and Duff & Phelps are higher than those assigned by Moody’s and S&P. They argue that the observed bond rating differences result from divergence in rating scales across CRAs rather than sample selection bias. In a follow-up study, Bongaerts et al. (2012) examine corporate bond issuers’ demand for an additional rating from Fitch, conditional on already having a rating from Moody’s and S&P, and test three possible explanations for this phenomenon: (a) information production – an additional Fitch’s rating adds value-relevant information, (b) ratings shopping – issuers disappointed with their existing ratings from Moody’s and S&P shop for better ratings from Fitch, and (c) regulatory certification – Fitch plays the role of a tiebreaker at the high-yield (HY)–investment-grade (IG) boundary. Bongaerts et al. find evidence that adding a Fitch’s rating is not associated with lower bond yields unless Fitch rates the issue IG when Moody’s and S&P’s

⁴ The CRAs provide unbiased ratings on average but their models are noisy in a way their ratings reflect some positive and negative errors.

are on opposite sides of the HY-IG boundary, in which case the yield improves by about 40 basis point. This evidence is consistent with Fitch's ratings having regulatory certification effects, but is inconsistent with the information production hypothesis. Further, Bongaerts et al. show that Fitch's ratings are particularly optimistic for bonds rated just below IG or bonds for which Fitch plays the role of a tiebreaker around the HY-IG boundary. Although the evidence seems to be consistent with the regulatory certification hypothesis, it could also be interpreted as being supportive of the ratings shopping explanation because ratings shopping incentives are expected to be stronger around the HY-IG boundary than elsewhere.

A possible explanation for the weak empirical evidence on ratings shopping for corporate bonds is sample selection bias. Most empirical studies focus on a sample of ratings in which issuers solicit an additional rating from a third CRA (typically Fitch), conditional on already having ratings from Moody's and S&P. Such a sample suffers from selection bias because issuers who engage in ratings shopping and end up with only one published rating (from Fitch), after hiding the less favorable ratings (from Moody's and S&P), will not be covered. Thus, any shopping-related bias in published ratings will be difficult to detect and likely understated.

To the best of our knowledge, Kronlund (2016) is the only exception to date, examining ratings shopping in the sense that issuers shop for multiple ratings, withhold lower ratings, and only publish the higher ones. Kronlund finds that for issuers with only one rating, a CRA that rated the issuer's bonds higher than the other CRAs last year is more likely to be the CRA that the issuer solicits, suggesting that the published ratings are more representative of the favorable opinions. Consistent with the asset complexity prediction of Skreta and Veldkamp (2009), Kronlund shows that this bias is strongest among junior and long-term bonds, which are more complex to rate. Kronlund also finds that bond investors account for shopping-related bias in ratings and demand higher yields, a finding that is at odds with the naïve investor assumption made by Skreta and Veldkamp (2009) and Bolton et al. (2012), but is consistent with the rationality assumption of Sangiorgi and Spatt (2016). Finally, Kronlund investigates ratings shopping motives and finds strong evidence that issuers engage in ratings shopping primarily for regulatory arbitrage purposes.

There are two recent major studies investigating the phenomenon of ratings shopping in the structured product domain: (a) Griffin et al. (2013), who provide evidence against ratings shopping in the CDO market; and (b) He et al. (2015), who offer evidence in support of ratings shopping in the MBS market.

Using a sample of CDOs from 1997 to 2007, Griffin et al. document that nearly 85% of all AAA CDO capital with a rating from either Moody's or S&P also receives a rating from the other CRA. For dual-rated tranches with at least one AAA rating, they find that more than 96% of the capital receives identical AAA ratings from Moody's and S&P. Thus, it appears that dual certification and agreement on the same rating is the norm for highly-rated CDO tranches. This is inconsistent with ratings shopping, which posits that only the most favorable rating(s) will be purchased and reported. Griffin et al. conduct a more direct test of ratings shopping by comparing the yield spreads and default rates of AAA CDO tranches rated by both Moody's and S&P with those of AAA CDO tranches rated by only one of them. They find that although the latter are associated with larger yield spreads, they actually experienced fewer defaults, implying that the ratings provided by only one CRA are not of lower quality, inconsistent with a ratings shopping story.

He et al. use a sample of MBS deals originated and issued between 2000 and 2006 and find evidence in favor of ratings shopping for non-AAA rated MBS tranches. They first document that non-AAA rated tranches are significantly more likely to receive only one rating than AAA rated tranches, which suggests that ratings shopping appears more likely among non-AAA rated tranches. Consistent with this observation, they find that single-rated tranches experience larger losses than multi-rated ones. Interestingly, investors seem to recognize and take ratings shopping into consideration when pricing non-AAA rated tranches. He et al. find that initial yields predict future losses for non-AAA rated tranches, but not for multi-rated ones.

Overall, the evidence on ratings shopping seems to be inconclusive not only in the traditional corporate bond markets, for which the potential for ratings shopping is presumably limited, but also in the more complex structured product markets, despite the extensive anecdotal evidence at hand. In our view, this mixed evidence likely reflects the empirical difficulties encountered by researchers. At the center of these difficulties is the opaque nature of the solicitation process. For example, researchers typically do not observe unpublished ratings that were solicited by issuers. Worse, researchers do not observe how the CRAs that were not solicited would have rated, so they cannot determine whether the published ratings are higher than those that were unsolicited. In our view, future research on this topic will primarily focus on addressing these identification issues.

CONCLUSION

Ratings have a dual role: they affect market prices not only because they provide credit-relevant information but also because they form a focal point for investment rules and regulations that restrict the investment activities of certain institutional investors. Further, the certification services of CRAs appear to facilitate firm access to capital markets and to have positive real effects on the economy.

However, a recent strand of literature focuses on CRAs' incentive problems arising from the peculiar structure of the credit rating industry and offers several alarming findings. Reputation concerns appear insufficient to offset the severe conflicts of interest inherent in the issuer-pays business model, whereby CRAs are paid by the issuers they are supposed to rate objectively. Increased competition among CRAs worsens this perverse incentive as CRAs seek to cater to issuers' preferences to gain market share. Aside from CRAs' incentive problems, the ability of issuers to shop around for ratings from multiple CRAs and to select and disclose the most favorable one(s) leads to a systematic upward bias in published ratings. While there seems to be consensus in the literature that these findings hold true in the context of structured product markets, it is premature to conclude that they also apply to the realm of corporate bonds. Indeed, the evidence on inflation in corporate bond ratings due to conflict of interests is mixed and inconclusive. Further, it is unclear whether competition among CRAs improves or worsens the quality of corporate bond ratings. In our view, future research on these topics should focus on resolving identification issues, such as how to correctly measure competition among CRAs and how to resolve empirical difficulties related to the opaque nature of the ratings solicitation process.

These conclusions suggest that when corporate CFOs engage in financial restructuring and other actions to preserve favorable bond ratings, they are acting in the interests of shareholders to avoid higher costs of capital and to preserve access to debt financing. Investors and regulators correctly regard bond ratings as useful information about credit risk. Nonetheless, bond investors remain somewhat skeptical about ratings quality: they discriminate among CRAs when ratings are split and appear to price bonds based on independent research as well.

REFERENCES

- Acharya, V. V., and M. Richardson, 2009, "Causes of the financial crisis," *Critical Review* 21, 195–210
- Acharya, V. V., P. Schnabl, and G. Suarez, 2013, "Securitization without risk transfer," *Journal of Financial Economics* 107, 515–536
- Acharya, V. V., T. Philippon, M. Richardson, and N. Roubini, 2009, "The financial crisis of 2007-2009: causes and remedies," *Financial Markets, Institutions & Instruments* 18:2, 89-137
- Ashcraft, A. B., P. Goldsmith-Pinkham, and J. I. Vickery, 2010, "MBS ratings and the mortgage credit boom," unpublished working paper, Federal Reserve Bank of New York staff report no. 449
- Bae, K., H. Driss, and G. S. Roberts, 2016, "The effect of competition among rating agencies on ratings quality: evidence from the Canadian corporate bond market," unpublished working paper, Saint Mary's University and York University
- Bae, K., J. Kang, and J. Wang, 2015, "Does increased competition affect credit ratings? A reexamination of the effect of Fitch's market share on credit ratings in the corporate bond market," *Journal of Financial and Quantitative Analysis* 50, 1011–1035
- Baghai, R., and B. Becker, 2016, "Non-rating revenue and conflicts of interest," unpublished working paper, Swedish House of Finance research paper no. 15-06
- Bar-Isaac, H., and J. Shapiro, 2013, "Ratings quality over the business cycle," *Journal of Financial Economics* 108, 62–78
- Becker, B., and T. Milbourn, 2011, "How did increased competition affect credit ratings?" *Journal of Financial Economics* 101, 493–514
- Benmelech, E., and J. Dlugosz, 2009, "The alchemy of CDO credit ratings," *Journal of Monetary Economics* 56, 617–634
- Bolton, P., X. Freixas, and J. Shapiro, 2012, "The credit ratings game," *Journal of Finance* 67, 85–112
- Bongaerts, D., K. M. Cremers, and W. N. Goetzmann, 2012, "Tiebreaker: certification and multiple credit ratings," *Journal of Finance* 67, 113–152
- Camanho, N., P. Deb, and Z. Liu, 2012, "Credit rating and competition," unpublished working paper, Catholic University of Portugal, London School of Economics & Political Science, International Monetary Fund, and Bank of England.
- Cantor, R. M., and F. Packer, 1997, "Differences of opinion and selection bias in the credit rating industry," *Journal of Banking & Finance* 21:10, 1395–1417
- Cohen, A., and M. D. Manuszak, 2013, "Ratings competition in the CMBS market," *Journal of Money, Credit and Banking* 45, 93–119
- Coval, J. D., J. W. Jurek, and E. Stafford, 2009, "The economics of structured finance," *Journal of Economic Perspectives* 23, 3–25
- Covitz, D. M., and P. Harrison, 2003, "Testing conflicts of interest at bond rating agencies with market anticipation: evidence that reputation incentives dominate," unpublished working paper, Federal Reserve Board, Washington, D. C.
- Diamond, D. W., 1984, "Financial intermediation and delegated monitoring," *Review of Economic Studies* 51, 393–414
- Dichev, I. D., and J. D. Piotroski, 2001, "The long-run stock returns following bond ratings changes," *Journal of Finance* 56, 173–203
- Doherty, N. A., A. V. Kartasheva, and R. D. Phillips, 2012, "Information effect of entry into credit ratings market: the case of insurers' ratings," *Journal of Financial Economics* 106, 308–330
- Driss, H., N. Massoud, and G. S. Roberts, 2016, "Are credit rating agencies still relevant? Evidence on certification from Moody's credit watches," *Journal of Corporate Finance*, Forthcoming
- Ederington, L. H., J. B. Yawitz, and B. E. Roberts, 1987, "The informational content of bond ratings," *Journal of Financial Research* 10, 211–226
- Faulkender, M., and M. Petersen, 2006, "Does the source of capital affect capital structure?" *Review of Financial Studies* 19, 45–79

- Frenkel, S., 2015, "Repeated interaction and rating inflation: a model of double reputation," *American Economic Journal: Microeconomics* 7:1, 250-280
- Goh, J. C., and L. H. Ederington, 1993, "Is a bond rating downgrade bad news, good news, or no news for stockholders?" *Journal of Finance* 48, 2001-2008
- Grier, P., and S. Katz, 1976, "The differential effects of bond rating changes among industrial and public utility bonds by maturity," *Journal of Business* 49, 226-239
- Griffin, J. M., and D. Y. Tang, 2012, "Did subjectivity play a role in CDO credit ratings?" *Journal of Finance* 67, 1293-1328
- Griffin, J. M., J. Nickerson, and D. Y. Tang, 2013, "Rating shopping or catering? An examination of the response to competitive pressure for CDO credit ratings," *Review of Financial Studies* 26, 2270-2310
- Griffin, P. A., and A. Z. Sanvicente, 1982, "Common stock returns and rating changes: a methodological comparison," *Journal of Finance* 37, 103-119
- Hand, J. R. M., R. W. Holthausen, and R.W. Leftwich, 1992, "The effect of bond rating agency announcements on bond and stock prices," *Journal of Finance* 47:2, 733-752
- He, J., J. Qian, and P. E. Strahan, 2011, "Credit ratings and the evolution of the mortgage-backed securities market," *American Economic Review* 101, 131-135
- He, J., J. Qian, and P. E. Strahan, 2012, "Are all ratings created equal? The impact of issuer size on the pricing of mortgage-backed securities," *Journal of Finance* 67, 2097-2137
- He, J., J. Qian, and P. E. Strahan, 2015, "Does the market understand rating shopping? Predicting MBS losses with initial yields," *Review of Financial Studies* 29, 457-485
- Hettenhouse, G., and W. Sartoris, 1976, "An analysis of the informational value of bond rating changes," *Quarterly Review of Economics and Business* 16, 65-76
- Ingram, R. W., L. D. Brooks, and R. M. Copeland, 1983, "The information content of municipal bond rating changes: a note," *Journal of Finance* 38, 997-1003
- Jiang, J. X., M. H. Stanford, and Y. Xie, 2012, "Does it matter who pays for bond ratings? Historical evidence," *Journal of Financial Economics*, 105, 607-621
- Jorion, P., Z. Liu, and C. Shi, 2005, "Informational effects of regulation FD: evidence from rating agencies," *Journal of Financial Economics* 76, 309-330
- Kisgen, D. J., and P. E. Strahan, 2010, "Do regulations based on credit ratings affect a firm's cost of capital?" *Review of Financial Studies* 23, 4324-4347
- Klinger, D., and O. Sarig, 2000, "The information value of bond ratings," *Journal of Finance* 55, 2879-2902
- Kronlund, M., 2016, "Do bond issuers shop for favorable credit ratings?" unpublished working paper, University of Illinois, Urbana-Champaign
- Liu, P., and A. V. Thakor, 1984, "Interest yields, credit ratings, and economic characteristics of State Bonds: an empirical analysis: note," *Journal of Money, Credit and Banking* 16, 344-351
- Luchetti, A., and S. Ng, 2010, "Ratings shopping' lives as Congress debates a fix," *Wall Street Journal*, May 24
- Mählmann, T., 2011, "Is there a relationship benefit in credit ratings?" *Review of Finance* 15, 475-510
- Mathis, J., J. McAndrews, and J.-C. Rochet, 2009, "Rating the raters: are reputational concerns powerful enough to discipline rating agencies?" *Journal of Monetary Economics* 56, 657-674
- Millon, M. H., and A. V. Thakor, 1985, "Moral hazard and information sharing: a model of financial information gathering agencies," *Journal of Finance* 40, 1403-1422
- Neumann, J., 2012, "Fitch ditched in Bond Dispute," *Wall Street Journal*, April 4, p. C1
- Office of the New York State Attorney General, 2008 "Attorney General Cuomo announces landmark reform agreements with the nation's three principal credit rating agencies," Press Release,
- Opp, C. C., M. M. Opp, and M. Harris, 2013, "Rating agencies in the face of regulation," *Journal of Financial Economics* 108, 46-61
- Ramakrishnan, R. T. S., and A. V. Thakor, 1984, "Information reliability and a theory of financial intermediation," *Review of Economic Studies* 51, 415-432
- Sangiorgi, F., and C. Spatt, 2016, "Opacity, credit rating shopping, and bias," *Management Science*, Forthcoming
- Skreta, V., and L. Veldkamp, 2009, "Ratings shopping and asset complexity: a theory of ratings inflation," *Journal of Monetary Economics* 56, 678-695
- Strobl, G., and H. Xia, 2012, "The issuer-pays rating model and ratings inflation: evidence from corporate credit ratings," unpublished working paper, Frankfurt School of Finance and Management and University of Texas at Dallas
- Sufi, A., 2009, "The real effects of debt certification: evidence from the introduction of bank loan ratings," *Review of Financial Studies* 22, 1659-1691
- Tang, T. T., 2006, "Information asymmetry and firms' credit market access: evidence from Moody's credit rating format refinement," *Journal of Financial Economics* 93, 325-351
- Weinstein, M. I., 1977, "The effect of a rating change announcement on bond price," *Journal of Financial Economics* 5, 329-350
- White, L. J., 2010, "Markets: the credit rating agencies," *Journal of Economic Perspectives* 24, 211-226
- Xia, H., 2014, "Can investor-paid credit rating agencies improve the information quality of issuer-paid rating agencies?" *Journal of Financial Economics* 111, 450-468

The Power of “Negative Beta”: Why Every Portfolio Should Include Private Equity

Andrew Freeman – Risk fellow, Centre for Risk Studies, Judge Business School, University of Cambridge

Iordanis Karagiannidis – Assistant Professor of Finance, The Citadel

D. Sykes Wilford – Hipp Chair Professor of Business and Finance, the Citadel and Senior Advisor to Access Corporate Finance¹

Abstract

Building on previous work we analyze the option-like characteristics of investment in private equity. While the main academic focus has been on the disputed ability of this asset class to produce above-average risk-adjusted returns, our focus is on the underappreciated role played by volatility in private equity (PE) performance. Our conclusion is that PE is a much more attractive asset class (lower risk) than commonly believed. In contrast to most approaches, we focus on the influence of the options built into the private-equity investment business model primarily from the perspective of the fund manager, or General Partner (GP). As the owner of call options

on the underlying investors’ capital commitments initially and later of complex put options as assets can be retained before being sold, the GP is well placed to take advantage of market volatility, particularly during bad times. We posit the existence of “negative beta” as a reason for large investors to make much bigger allocations to PE than are typical.

¹ The opinions presented in this paper reflect those of the authors and not necessarily those of the respective institutions noted. The authors can be contacted respectively at af489@cam.ac.uk, iordanis@citadel.edu, wsykes@dsykeswilford.com. The authors wish to thank Juan Montalvo Bressi for his comments.

INTRODUCTION

In recent decades, it has become fashionable to include an allocation to so-called alternative assets in most professionally managed investment portfolios. These assets have included hedge funds, venture capital, and private equity, as well as occasionally exotics such as art and real estate. The rationale for such investments has typically been that they provide the only free lunch in investing where assets in public markets represent the principal portfolio components – the benefits of non-correlated returns via diversification. While relatively small allocations in the 2-5% range have been the norm, there have been occasional examples of institutions, notably the Yale endowment, taking larger exposures, thereby attracting a mixture of approval and opprobrium [Provence (2008); Considine (2010); Kaplan and Sensoy (2014)].

Private equity (PE) is generally considered as the critical “alternative” asset class [Bain (2016)]. Like other alternative investments, PE offers benefits in the form of diversification from public markets, deploying information advantages and a variety of approaches to create value from corporate assets [Kaplan and Schoar (2005); French et al. (1987)]. Investors have traditionally paid high fees in return for access to the asset class, but this has not been without controversy. Investors remain doubtful about the efficacy of the PE model on a risk-adjusted basis.

However, we identify a much more intriguing argument for including PE in an asset allocation. Building on earlier work, it is suggested that PE offers unique option-like characteristics in the form of embedded “negative beta.” Simply put, in a financial crisis PE will perform so differently from publicly traded assets that for a typical large investor it would constitute a buffer against a downturn that is in proportion to the relevant portfolio’s allocation. As Bain (2016) stated: “The global financial crisis claimed many victims but barely put a dent in PE assets under management.” We believe that this can be at least partly explained by the risk control that is intrinsic to the PE business model. However, we go further in arguing that even today’s PE assets are underappreciated by most investors.

The starting point for the argument can be found in Freeman and Wilford (2016). They identified and analyzed the options embedded in the fund structures of a hypothetical PE firm; during its investment period, a PE fund owns a series of call options that allow it to demand investment capital from investors who have signed up at the beginning of the life of the fund. For four or five years after a fund’s creation, the PE firm can use guaranteed liquidity to buy assets, albeit highly idiosyncratic

ones.² The option value of this liquidity rises sharply if financial markets are stressed, so a fund that owns calls can buy assets cheaply in bad times. In addition, a fund that is fully invested undergoes a neat reversal of its former position. From being long the right to call capital, it is now the owner of complex put options. Because the fund’s objective is now to dispose of assets at or above some notional exit multiples, it can judge market conditions and decide when and how to exercise its right to sell. Crucially, the typical fund has the option of waiting for quite long periods in order to avoid selling if markets are distressed. It is in its, and its investors’, interests to delay until the price is right. Ownership of this complex option constitutes a key element of the overall “negative beta” of PE.³

Freeman and Wilford (2016) provided a simple initial simulation of PE option values and their response to movements in the volatility of public markets. This supported the suggestion that using embedded options as a way to analyze PE risk from a risk management perspective might be beneficial. Tentative conclusions were offered about the possible positive influence of PE embedded options on systemic risk, as well as on the possibility that large investors might be underexposed to this asset class because of a failure to appreciate its portfolio risk-management potential.

This paper further explores the rationale for including PE in an asset allocation on the explicit grounds that it has unrivaled risk-return characteristics, crucially once the risk elements in the equation are better understood. Arguments against PE typically cite high management fees as unjustified given the level of performance once adjusted using conventional risk-return analysis [SEC (2015); Sorensen et al. (2014)]. This objection fades in light of our insight that the embedded option values, intrinsically difficult to model though they are, mean that PE assets are in reality very different from their typical characterization in the literature. In this paper, we extend the original analysis by undertaking simple simulations to show the effect of optionality on the volatility of PE assets compared to a standard equity market investment.

² A typical buy-out fund, for example, will make between 6 and 12 investments over 5 years

³ Thomas Meyer recognized as long ago as 2007 that real options could offer new insights into private equity. See Mathonet and Meyer (2007) as well as Meyer (2014). However, this considers real options from the perspective of the investor. Our focus is on the GP and the interaction of optionality with both investors and the market. Separately, Chen et al. (2008) focused on the optionality of a GP’s investment portfolio. Thanks to Thomas Meyer for the references to his work.

PE AS A PORTFOLIO RISK MITIGATION TOOL

A useful starting point for evaluating the value that PE can bring to a portfolio as a risk mitigation tool is a simple side-by-side comparison. Taking the S&P as our base case diversified portfolio, the analysis builds simplified portfolio allocations utilizing monthly data for the period January 2005 to December 2014. These data encompass the 2008 – 09 crisis and the volatility spikes that occurred. Further, initial modeling controls for performance by assuming that it is equal to the S&P for each period except for the implicit value of the options in the PE fund. In this way, the focus is placed directly on volatility. For example, if the S&P goes up from a base of 100 to 110, but the fund has not yet invested, then the value of that investment remains 100 plus the value of the option to time the investment by calling capital. If the funds were already invested and the S&P goes up by 10% then the value of the investment is assumed to rise by 10%; however, this value is not the overall value of the PE investment. As shown in Freeman and Wilford (2016), the actual value includes the value of the right to put the component parts of the PE investment portfolio to the market at a chosen time, or opportunistically if the price is right. Thus, the value of the PE investment is the combination of the call and put options and the invested capital as indicated by the S&P. Simply using a standard option pricing model,⁴

PE fund value = call option value (C) + investment growing at the market rate + put option value (P), where

Call option value (C) = $S \times N(d_1) - X \times e^{-rT} \times N(d_2)$

where $d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right) \times T}{\sigma \times \sqrt{T}}$ and $d_2 = d_1 - \sigma \times \sqrt{T}$

and C = call option value, S = the amount still to be called, X = the amount still to be called, R = 2%, T = time to maturity (time to reach month 60) in years, and sigma = the monthly volatility (VIX)

And,

Put option value (P) = $C + X \times e^{-rt} - S$, however now

S = current investment value, and X = forward value of that investment at time 120 using a 2% rate or a 25% of the original investment as a cap.⁵

Put option value remains zero if no investment has been called.

As one can observe from the calculation process above for P

one must be careful in understanding the concept of valuing the put. This analysis makes the simplifying assumption that the put option's strike price is a forward value at each date of the option's price calculation unless the forward is greater than 25% of the original investment (this will be modified for examples where the takeout target is greater).⁶ In reality, there is not a true strike price to calculate against at each calculation date, suggesting that our measure will have errors. Also, the option value will tend to move with the spot price (as the S&P moves) not just from spot movements but also from changes in the strike price implied by our assumption of how the forward is calculated. Such an analysis does take away the subjectivity implicit in alternative measures of the strike that may be considered, and will be altered if the takeout target is altered for the experiment. More importantly, the impacts of time and volatility on the value of the right to time taking an investment to market are consistently considered along the path of pricing.

In making the assumptions concerning the strike of P, the goal is not to bias the measures of risk that will be estimated using this options approach. Ultimately, due to the methodology used, all of the investments will have a return similar to that of the S&P (in some cases leveraged) or less, depending on the timing of investment, which is generated randomly to avoid unintended bias. Over time, however, the PE fund will take a significantly different path to that end. The call options will go from valuable to worthless once exercised. The put options will first fade and finally cease as investments are placed in the market. During the lives of the options the value of the fund and the S&P will differ significantly.

4 The standard references are Black and Scholes (1973) and Merton (1973).

5 For example: If S = \$1,200,000 at time 20, then $X = \$1,200,000 \times (1+0.02)^{8.3}$, where 8.33 is the time remaining in years (100 months /12). If S = \$1,300,000 at time 20, then $x = 1,250,000$. If S = 900,000 at time 20, then $x = 900,000 \times (1+0.02)^{8.3}$, where 8.33 is the time remaining in years.

6 We acknowledge that the put option takes the form of a complex set of real options rather than a pure financial option, hence the need to simplify in the way we propose to calculate the strike price at each point in time. The 25% cap is arbitrary and based upon our modeling the date of when an investment goes to the market, which is set at a 25% gain (no leverage version) in the value of the investment. One can correctly argue that without a set strike for the put option it cannot be truly valued as a typical financial option. We attempt to deal with this problem by biasing its "value" downward, in a sense, through the use of the forward and cap technique; this is not a well-defined "closed form" solution to the problem of valuing a set of real options, however it highlights the critical issues underlying any option value – time until expiry and volatility – because without volatility the exercise of pricing any option is meaningless.

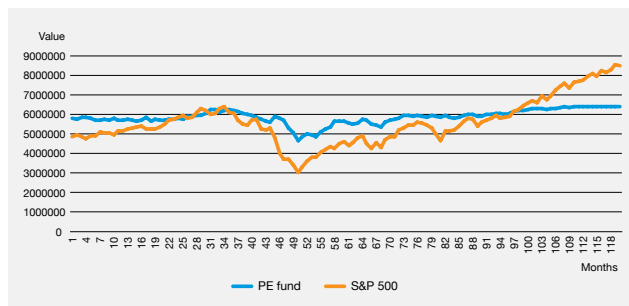


Figure 1 – Path of \$5 million investment

	Call 1	Call 2	Call 3	Call 4	Call 5
Time	9	12	32	39	53
	Put 1	Put 2	Put 3	Put 4	Put 5
Time	98	99	109	104	63

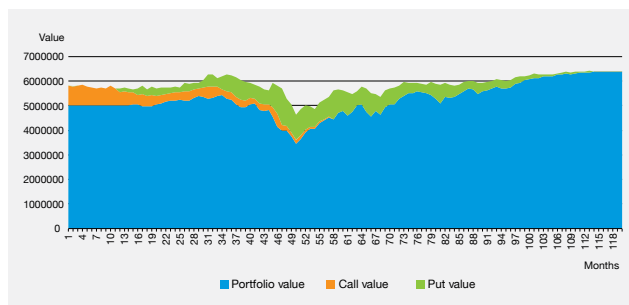


Figure 1a – Value of each component

Typical PE allocations are intended to produce returns superior to those available from public markets, but our analysis begins with the assumption that outperformance is not the goal. The focus is on volatility, and this allows us also to generate the correlation of returns to the S&P. During a long period of low volatility and good performance of the S&P, the path of returns for the fund will differ based upon time of entry into and exit from an investment plus how the option value changes as volatility (even if small) moves.

The interesting case from a risk diversification perspective is what happens to that path of returns when there is a volatile market, a crisis, a rise in fear and consequent plunge in value. In these instances, the value of the embedded options rises significantly, offsetting the mark-to-market plunge in value that would normally (directly in our simulations) be assumed to occur if the embedded options in the fund are ignored.⁷ The path of the value of the fund will be different from that of the S&P, even if in the end the overall return is the same and, therefore, the correlation of the value of the fund to the S&P should not

be perfect. The degree to which it will vary will be sensitive to entry and exit of investments (creation and extension of the options) and the movement of market volatility. Aware of its limitations, we use the VIX as the best available representative measure for expected volatility.

Figure 1 depicts the path generated for the S&P and one of many possible simulation paths. The simulations assume that the call is made by the PE fund randomly with all calls made within five years; many possible paths can thus be generated. Since all the calls are made randomly during the first five years of the life of the fund, the value of the calls has been dissipated before the depth of the recession. In these simulations, much of the call value expired before the worst period during the life of the data. Calls are made between the 9th and 53rd months. Interestingly, the last call occurs at one of the low points of the recession and will be the first investment to be monetized just 10 months later.⁸

Offsetting the loss of the value of the call, the put option value commences after a call is made. As the crisis volatility increases, we begin to observe the impact of volatility on the value of the PE fund. Through the recession, as expressed in lower stock market prices, the fund loses money at a much slower pace, even as the shortening time to maturity impacts the option value negatively. Finally, our decision to cap the value of the put as a measure of the forward rate comes into play during the deep downturn by limiting the value of the put due to the treatment of the strike price for calls made at much higher levels of the S&P. In general, however, the put values impact the overall volatility of the fund to offset the slide in the S&P (the underlying).⁹

The valuation impact of the options is more easily observed in Figure 1a. Initially, all of the value is in the call, then at 9 months and 12 months two of the calls disappear (the investor commitments are called) while the puts now become valuable.

7 A mark-to-market mentality is, of course, the antithesis of the PE approach and is resisted fiercely by the industry. Our analysis is a rare and admittedly stylized example of where the direct comparison in fact does not harm the underlying justification for PE investing.

8 See bottom of figure for a list of put and call timings. For example, the first call is made month 9 of the 5-year period, chosen randomly, and that investment is not taken to the market until month 98, when its value is now 1.25% of the original investment.

9 Since the "correct" value for the put is always in doubt due to the methodology of estimating the financial put, one may wonder in what ways the volatility of the PE fund may be biased if alternative estimation procedures were used. We believe this methodology does not bias the volatility of the fund downward, and in some instances perhaps the opposite is true.

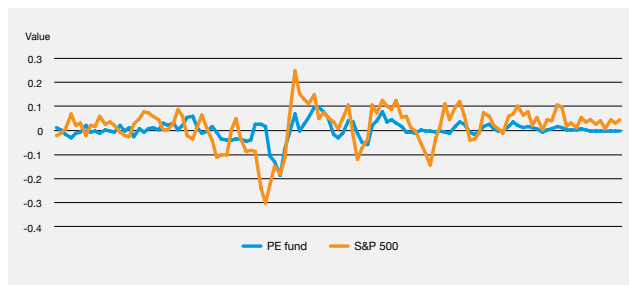


Figure 2 – 3-month rolling returns

The bulk of the value remains in the underlying that varies depending on the movement of the S&P. While the S&P is losing value quickly as the crisis occurs, the path for the PE fund is smoothing out the impact resulting from the rise in the value of the puts; volatility is rising. As noted, the first put value drops to zero as it is cashed out on month 63, after which 4 puts remain, slowly losing value until the underlying investments are cashed out in months 98, 99, 104, and 109. At this juncture, all of the put values are gone from the estimates and the cash is left sitting in the fund ready for distribution to investors.¹⁰

With these caveats in mind, we observe biases in returns and volatility that may be introduced due to the methodology chosen for simulations. Still, the positive impact of the options during the crisis is evident in the form of obviously lower volatility, with any bias occurring toward the tail of the 10-year period.

In Figure 2, we look at the volatilities of the fund and S&P more closely. To take advantage of the full dataset while making the numbers a bit more meaningful the data are smoothed by using 3-month rolling average index levels of value for both the PE fund and the S&P. Specifically, the formula is:

$$\text{PE fund return}_t = \frac{\text{PE fund value}_t - \text{PE fund value}_{t-3}}{\text{PE fund value}_{t-3}}$$

$$\text{S\&P return}_t = \frac{\text{S\&P index value}_t - \text{S\&P index value}_{t-3}}{\text{S\&P index value}_{t-3}}$$

It is clear that the level of the volatility for the fund is systematically lower than that of the S&P. However, these observations are for one path (one we will hold in subsequent figures for comparison purposes). Because the calls are generated randomly, other paths could look entirely different. For example, if the random generator selected all the calls in months 50 through 60 the cash out would occur quickly, thereby implying an even lower volatility.

	PE fund		
	Mean	Min	Max
PE fund risk – standard deviation (returns)	9.01%	6.71%	15.78%
S&P risk – standard deviation (returns)	14.62%	14.62%	14.62%
Correlation (PE fund, S&P)	0.39	0.15	0.57
Ratio S&P risk/PF fund risk	1.62	2.18	0.93

Table 1 – Average volatilities

We next ran multiple simulations to calculate the average volatility of the simulations, as well as a maximum and a minimum volatility paths. Just as with the “average” path volatility pattern, we can calculate the correlation of the PE fund to the S&P. Table 1 clearly shows that the volatility of the PE fund is much lower than that of the S&P in the maximum, minimum, and average case simulations. The implication that PE is a highly attractive asset class should be clear even from this simple analysis and its assumptions that the call and put options have value for the PE fund – despite the fact that this is not a typical way to view the asset class today.

Table 1 presents the data needed to create an optimal portfolio based upon the noted paths, assuming the same expected return for both the S&P and the PE Fund. We can observe that the volatility ratios show that the embedded PE options dampen the volatility significantly, with the random call process implying a correlation, in all three cases, significantly less than one. In the maximum path with the highest standard deviation the correlation is still significantly less than one, while in the average path the correlation is only 0.39. This is important because the base for the investment is the S&P. We could be satisfied at this juncture that PE fund investments do indeed provide diversification, even in this simplistic form.

¹⁰ In effect, we can ignore the tail of the figures. In reality, cash is distributed to investors as the puts are realized, so the fund’s assets naturally shrink. Occasionally funds for distribution are held in escrow for future investment by the PE fund. The eventual returns received by investors are a function of how much, and when, their cash is put to work. More analysis needs to be done on the effects of cash flow movements in PE investment cycles, but this is not necessary for our purposes in this article.

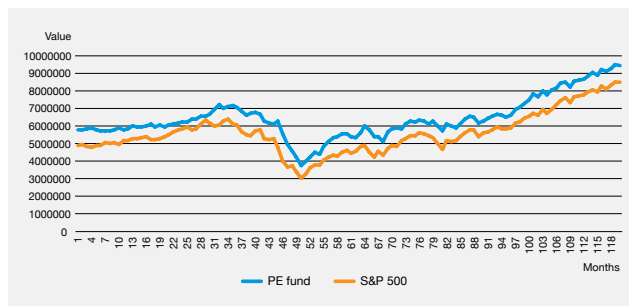


Figure 3 – Path of \$5 million investment

	Call 1	Call 2	Call 3	Call 4	Call 5
Time	9	12	32	39	53
	Put 1	Put 2	Put 3	Put 4	Put 5
Time	106	106	119	109	72

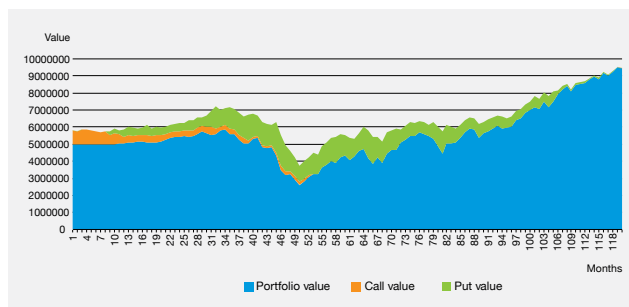


Figure 3a – Value of each component

	Mean	PE fund	
		Min	Max
PE fund risk – standard deviation (returns)	14.12%	12.25%	15.60%
S&P risk – standard deviation (returns)	14.62%	14.62%	14.62%
Correlation (PE fund, S&P)	0.69	0.48	0.75
Ratio S&P risk/PF fund risk	1.04	1.19	0.94

Table 2 – With leverage average volatilities

ADDING REALISM TO THE ANALYSIS

A knowledgeable PE investor might scoff at our simple model above. After all, as noted above, what happens to the money once any cash out occurs could bias the risk, albeit only at the tail in the example. Clearly it matters when funds are returned to the investor. Obviously there would be leverage. Typically, the reason to invest with a PE fund is to take advantage of the manager’s knowledge to choose a segment of the market in which to invest, determine what company might be undervalued in that segment, and to make efficient use of leverage. To address some of these challenges, we now assume the following:

1. The fund is leveraged by 50%. The borrowing rate is 2% above Libor, and the borrowing occurs at the time of investment.
2. Once the PE fund returns 150% of the investment it is cashed out.
3. Any cash in this example is now invested in the S&P, which likely biases the correlation up and increases the volatility.

With these assumptions, the PE fund’s performance versus the S&P is presented in Figures 3 and 3a. The volatility of the PE fund is higher due to leverage.¹¹ Summary statistics are provided in Table 2. In this case, the volatility of the PE fund is higher, although it is similar to that of the unleveraged S&P.

From Figures 3 and 3a, which have the same call dates as Figures 1 and 1a, we see a different alignment and can easily observe the volatility due to leverage affecting the put values. The call values are not impacted since the leverage kicks in when the call is made and shows up in the put. Leverage also shows up in the return pattern influenced by the S&P’s movement because the size of the investment has now increased by the leverage factor.

We can compare the ratios in Tables 1 and 2. Volatilities are higher for the PE fund as would be expected due to leverage. However, the reinvestment into the S&P does not alter the correlations as much as one might expect, ex-ante, given that for much of the period the bulk of the PE fund takes exactly the

¹¹ Clearly, these paths depend on the various assumptions we have made. In future work we hope to use actual cashflows of a PE fund to determine whether the theoretical conclusions are indeed robust.

same pattern as the S&P, due to the reinvestment into the S&P as investments are taken to market. If we were comparing for performance we believe the returns would differ, probably in favor of the PE fund, but for our purposes this is not relevant.

Given the correlation and volatilities suggested by this new methodology, we can argue that an optimal portfolio should have a considerable amount allocated to the PE fund.¹² Even with the restrictions imposed, PE allocations clearly provide ample diversification. Given that many institutional investors and foundations have PE allocations of 5% or less, it is highly likely that, if our analysis is correct, most of them are sorely underallocated to PE [Ennis and Sebastian (2005) and CFA Digest (2005)].

ASSUMING SUBJECTIVE INVESTMENT DECISIONS

We have taken care not to bias our results significantly in favor of the PE fund’s lower volatility. In reality, investors move assets into a PE fund if they believe in the manager’s superior ability in some way. Superiority could take the form of asset selection, financial expertise and knowledge of capital structures, or ability to influence management behaviors at the level of portfolio companies. Again, we choose to simplify by giving to PE managers the ability to make investments when they observe an opportunity and decide to exercise the option to call investors’ funds. Our intention is to highlight the call and put option elements that we believe are fundamental to a PE fund. To do so, we move to a different 10-year period, commencing in January 2007.

More of the calls now occur during the crisis (call dates are again held the same at month 9, 12, etc.). The impact on the volatility measure shows up differently. The timing on the cash out for the investments – the puts – will differ accordingly (the 150% return is met at very different times than in the earlier analysis given the starting points of the investment).

Again, due to the nature of the investment, the values end up in similar positions. Since calls are partially made during, rather than before, the crisis, it is interesting to see how the put values adjust and how quickly a cash out occurs.

The most interesting results, however, show up in the ratios of risk for the PE fund and the S&P. The reduction in risk relative to the S&P is apparent in Table 3. In all cases, the PE fund is much less volatile, even with 50% leverage. Moreover, the correlations are extremely low, but the difference here from Table 2 is that the general methodology favors a higher correlation,

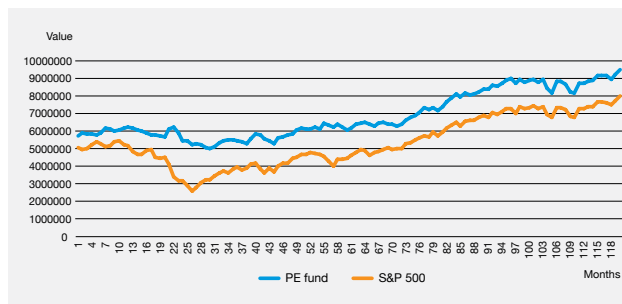


Figure 4 – Path of \$5 million investment

	Call 1	Call 2	Call 3	Call 4	Call 5
Time	9	12	32	39	53
	Put 1	Put 2	Put 3	Put 4	Put 5
Time	119	100	71	79	87

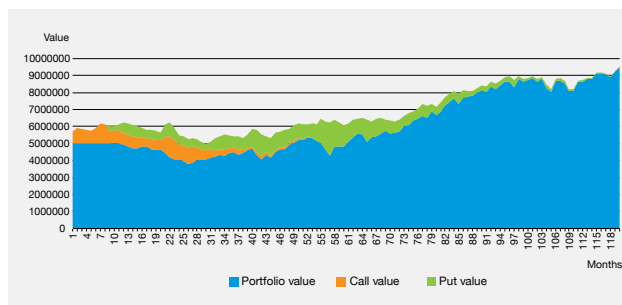


Figure 3a – Value of each component

	PE fund		
	Mean	Min	Max
PE fund risk – standard deviation (returns)	12.11%	11.30%	13.39%
S&P risk – standard deviation (returns)	15.24%	15.24%	15.24%
Correlation (PE fund, S&P)	0.37	0.09	0.50
Ratio S&P risk/PE fund risk	1.26	1.35	1.14

Table 3 – Actual crisis with leverage average volatilities

12 Our approach here would violate the theoretically correct way of determining an optimal portfolio [see Wilford (2012)]. However, our point is not to show a suggested level of allocation to PE, but rather to establish that for most large investors it would be significantly higher than allocations based on current understanding of risk-adjusted returns from PE.

as noted above. Still, the mean correlation is 37% and the minimum path is only 9%.

Freeman and Wilford (2016) noted that many believe that crises are the life-blood of the PE firm (fund). This view is supported by Table 3, not only in the much lower volatilities of the leveraged PE fund, but also in the very low correlation measures.

REALISM AND MARGINAL RISK REDUCTION – “NEGATIVE BETA”

The set of options that exist for the PE fund manager are multiple and can become complex. As one adds realism to the process, more interesting aspects of decision-making can be considered. What type of investment is made and when (do we need a crisis to buy optimally, for example)? When does a PE firm take any particular investment to market? And for the PE fund investor, what is the marginal impact of any PE fund investment on the portfolio? Above, we have made simplifying assumptions and as the research progresses some of these can be modified to consider the complications that better reflect the reality of a PE fund.

EXTENDING THE LIFE OF A FUND – VALUING THE PUT

Consider one of the important elements in risk reduction available to the PE Fund/GP (management firm). Critical to the value of the investment is the ability of the PE fund to delay taking an investment to the market (we may call this action the “extension right”). To see the value of this option, which may be exercised due to a crisis or simply due to conditions related to the market’s desire for the type of investment that is made by the fund, we create an exercise where there is a crisis that extends the implied value of the put. In this instance, the PE fund delays taking its remaining investment/s to market. In order to model this, we modify our method in several ways, while acknowledging again that this is a necessarily stylized version of reality. Here, the “put” option is in fact much closer to a set of real options than to a pure financial option.

- Using the basic model a crisis is created in the data, allowing the S&P to fall. An artificial crisis in 2012 is assumed by allowing another 2007-type downturn, and then the dataset is extended from that now artificial date onward. All of the calls have happened as above, however instead of the PE fund being forced to close out straight away the

GP can choose to extend the period for three years.

- In this artificial path only one of the five investments has been returned to investors and cashed out (here again all proceeds are invested into the S&P as in Figure 4 once this occurs).
- After the crisis occurs, the S&P is allowed to recover as it did in reality, however the dates are artificial, pushing out to 2019 (thus 2019 data are actually 2016, etc.).
- The PE fund manager is able to extend the holding period of the investments that remain by exercising the fund’s right to extend the investment period by three years after witnessing four successive months of negative returns while also recognizing that the normal life of the fund is about to end.

Extension increases the life of the put option, which creates a new path for the overall risk measure and valuation. Now the question becomes how valuable is this “crisis option”? From our modeling, we can see that the ability to extend the life of the put was very valuable. In reality, most PE firms will have

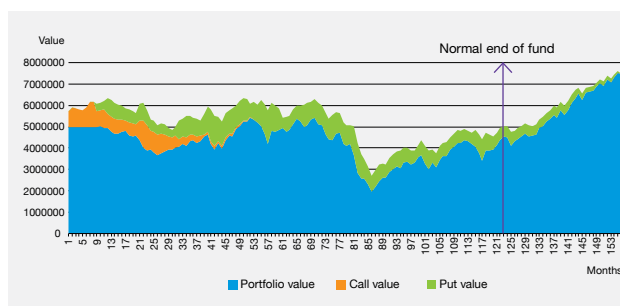


Figure 5 – Value of each component

	Call 1	Call 2	Call 3	Call 4	Call 5
Time	9	12	32	39	53
	Put 1	Put 2	Put 3	Put 4	Put 5
Time	156	156	69	145	156

	PE fund		
	Mean	Min	Max
PE fund risk – standard deviation (returns)	15.08%	14.14%	17.18%
S&P risk – standard deviation (returns)	17.13%	17.13%	17.13%
Correlation (PE fund, S&P)	0.62	0.39	0.73
Ratio S&P risk/PF fund risk	1.14	1.21	1.00

Table 4 – Simulated crisis average volatilities

the right to extend the life of a particular fund by one year in the first instance – this is typically written into the governing documents of the fund and may or may not be contingent upon a vote of approval from the fund’s advisory board. Equally significant for our analysis, extensions can often be made in yearly increments for up to three years and depending on the fund there might even be an option to ask the Advisory Board for a further exceptional additional extension in circumstances where the interests of the manager and investors are clearly aligned. No one wants to be a forced seller.¹³ Many GPs will reduce any remaining management fees at this point in order to cement that alignment.

To compare the value of the put option we assume that the PE fund manager goes to market at whatever price is available at the end of the period of recovery. If this extension option had not existed the fund would have been extinguished at the end of year 10 with losses; this equates to the line drawn after 120 months on the figure. The implied losses relative to the return that was accepted after the extension is some \$3.5 million or about half of what was returned to the client. The put option’s value would be extinguished and its value reduced to zero, forcing the firm to take the loss not only in opportunity but in actuality as well. Of the original \$5 million invested, \$1.5 million or some 30% of the original investment would have been lost. Dates of when the investments are actually taken to the market are noted under Figure 4, with the extension during the artificial crisis. As such, the puts extinguish near to or at the end of the extra years, not during a crisis.

The value of the option to extend is very clear from our results as presented in Figure 4 (our focus here is upon the “put” area in particular).

In Table 4, the data are once again presented for comparison. The volatility ratios are again greater than one, suggesting the PE fund was able to enjoy lower volatility in this circumstance and, just as importantly, the correlations of returns are low.

In summary, our initial assumption that capital must be returned at year 10 implies in our modeling that the PE fund would be forced to sell into a down market. With the ability to extend the life of the fund, the manager is given the freedom to continue to wait for better conditions to take the investment to market. This right makes the fund much more valuable during a crisis. The risk mitigation during a crisis is expressed in large part by the value of that option to wait, providing a risk profile for the PE fund that is significantly less volatile than that of the market as a whole, but meanwhile the fund is providing the returns sought by the investor.

CONCLUSION

A significant conclusion of our work is that most investment funds are systematically underexposed to PE if standard risk parameters are utilized during asset allocation decisions. This holds in normal times, when well-managed PE should be expected to offer outperformance even after fees. However, the real justification is found in bad times, when the “negative beta” effect of PE assets is to mitigate the effects of any downturn in public markets. Depending on the specific PE assets, this mitigation can be significant.

- Private equity firms draw down their investor commitments via capital calls.
- The right to call money at any time, including during market distress, is a powerful option,
- PE firms can also delay selling assets during a crisis, which is a valuable, if complex, put option.
- This combination of options makes PE investments perform with powerful “negative beta” during bad times.
- Institutional investors, particularly those with long time horizons such as foundations, have traditionally focused mainly on PE fees and may, therefore, be systematically underexposed to this asset class

REFERENCES

- Bain, 2016, “Global private equity report 2016,” Bain & Company, January, <http://bit.ly/21hj9dz>
- Black, F., and M. Scholes, 1973, “The pricing of options and corporate liabilities,” *Journal of Political Economy* 81:3, 637–654
- CFA Digest, 2005
- Chen, A. J. Conover, and J. Kensinger, 2008, “Private equity arrangements as real options,” *Midwest Finance Association 2012 Annual Meetings Paper*
- Considine, G., 2010, “Lessons from Yale’s endowment model and the financial crisis,” *Advisor Perspectives*, April 20, <http://bit.ly/2IZQIGZ>
- Ennis, R. M., and M. D. Sebastian, 2005, “Asset allocation with private equity,” *Journal of Private Equity* 8:3, 81–87
- Freeman, A., and D. S. Wilford, 2016, “Private equity capital commitments: an options-theoretic risk management approach,” *Journal of Financial Transformation* 43, 106–116
- French, K., W. Schwert, and R. Stambaugh, 1987, “Expected returns and volatility,” *Journal of Financial Economics*, 19:1, 3–29
- Kaplan, S. N., and B. A. Sensoy, 2015, “Private equity performance: a survey,” *Annual Review of Financial Economics* 7, 597–614
- Kaplan, S. N. and A. Schoar, 2005, “Private equity performance: returns, persistence, and capital flows,” *Journal of Finance* 60, 1791–1823

¹³ A logical extension of this is for the PE manager to consider making certain funds into quasi-permanent vehicles where the nature of the underlying assets means that pay-offs are bond-like or of very long duration, infrastructure being the obvious example.

- Mathonet, P-Y., and T. Meyer, 2007, J-curve exposure: managing a portfolio of venture capital and private equity funds, Wiley
- Merton, R. C., 1973, "Theory of rational option pricing," *The Bell Journal of Economics and Management Science* 4:1, 141-183
- Meyer, T., 2014, *Private equity unchained: strategy insights for the institutional investor*, Palgrave-Macmillan
- Provence, L., 2008, "Missing billion: how UVA's investment strategy worked... until it didn't," *The Hook*, November 20, <http://bit.ly/2I6kbQh>
- SEC, 2015, "Blackstone charged with disclosure failures," Securities and Exchange Commission press release no. 2015-235, <http://bit.ly/1L06uGk>
- Sorensen, M., N. Wang, and J. Yang, 2014, "Valuing private equity," *Review of Financial Studies* 27:7, 1977-2021
- Wilford, D. S., 2012, "True Markowitz or assumptions we break and why it matters," *Review of Financial Economics* 21, 93 – 101

Downside Risk Protection of Retirement Assets: A New Approach

Atanu Saha – Chairman, Data Science Partners

Alex Rinaudo – Chief Executive, Data Science Partners¹

Abstract

Over the past few decades, 401(k) plans, IRA accounts, and other self-directed investment vehicles have become the most important pool of retirement savings, leaving retirees exposed to the risk of outliving their assets, a hazard largely absent from traditional pension plans. Prior research has examined asset diversification, annuities, put options, and dynamic withdrawals as ways to mitigate this risk. This study proposes an alternative: explicit downside risk protection (or DRP) at the individual account level. The proposed DRP takes the following form: in years the account suffers a loss, that loss is capped at a predetermined amount. In return for this protection, the account holder gives up a portion of the gains only in years where the account's performance is positive. The net effect of this protection is to reduce the retirement account's downside risk, significantly reducing the likelihood of early account

depletion. A Monte Carlo simulation demonstrates that the chance of outliving one's assets over a retirement horizon of 45 years drops from nearly 15% without DRP to about 4% with DRP. Furthermore, by eliminating extreme negative outcomes, DRP has the potential to increase the average portfolio return (even accounting for the cost of protection) while simultaneously reducing the portfolio volatility. This paper also demonstrates that DRP can be profitably offered by a financial institution. It provides lower bound estimates of the rate of return a financial institution is likely to earn by offering DRP to retirement accounts.

¹ Sonya Rauschenbach and Akhil Shah have provided invaluable research assistance for this article. The authors also thank Rob Jones for his comments on an earlier version of this paper.

INTRODUCTION

The past few decades have seen a dramatic shift in how most Americans hold their wealth when entering retirement. While forty years ago pension plans dominated retirement assets, for future retirees 401(k) assets or similar retirement assets are by far the most important pool of retirement savings. However, most 401(k) plans do not provide adequate tools to manage downside risk. As a result, many retirees face a significant risk of early asset depletion, i.e., outliving their retirement assets. This paper proposes a new approach for providing downside risk protection for retirement portfolios.

There has been a marked asset shift in the types of retirement accounts in the past four decades. According to Investment Company Institute data, in 1974, 82% of U.S. retirement assets were in pension funds; by 2014 it had nearly halved to 42%, with a majority of retirement assets residing in 401(k) plans, IRA accounts, and other self-directed investment vehicles for retirement assets.²

Looking at changes in the mix of retirement assets over time masks the fact that a majority of the pension assets are held by older individuals. According to the Employment Benefit Research Institute, in 2010 only 30% of current private sector plan participants had access to pension plans; in 1979 this figure was higher than 80%. Based on these data, it is reasonable to expect that the share of retirement assets in 401(k) plans, as opposed to pension plans, is likely to increase substantially in the coming years.

Compared with pension plans, 401(k) plans confer certain benefits, such as portability and investment control. These are particularly valuable before retirement (the accumulation phase), as they give individuals more control over their ability to change employment and over the composition of their portfolio. However, during retirement (the withdrawal phase), 401(k) plans expose retirees to risks that are largely absent in pension plans. In particular, 401(k) plans, when compared with pension plans, have two key areas of risk: longevity and income risk.

Longevity risk refers to the unknown amount of time in retirement over which any individual will require income. In a 401(k) plan, retirees have assets that they can choose to invest in and withdraw from as they see fit. However, there is no guarantee that these assets will last for the entire retirement time horizon. This is in contrast to a typical pension plan that provides a predetermined income stream until death, eliminating longevity risk for an individual.

Income risk refers to the uncertainty of the income stream during retirement. 401(k) account holders have a specific balance from which to draw funds and are, therefore, impacted by investment performance. Poor investment performance (downside risk) accelerates the decline in retirement assets, reducing the asset balance left to draw upon.

In light of these risks, both the academic literature and the popular press have paid considerable attention to the issue of risk mitigation. This paper proposes a new approach of providing downside risk protection for a retirement account. The proposed downside risk protection (DRP) takes the following form: in the years where the retirement account has positive returns, the account holder pays a portion of the amount gained and in years where the account suffers a loss, that loss is capped at a predetermined amount. The net effect of this protection is to reduce the retirement account's downside risk, significantly reducing the likelihood of early account depletion.

A Monte Carlo simulation of 50,000 different portfolio return outcomes suggests that the proposed downside risk protection approach markedly reduces the likelihood of early asset depletion. This risk reduction is far more effective than the widely-recommended diversified portfolio of equity and fixed-income assets. Furthermore, downside risk protection, by eliminating extreme negative outcomes, has the potential to increase the average return (even after accounting for the cost of protection) and reduce the volatility of portfolio returns. This paper concludes by demonstrating that downside risk protection can be provided by a financial institution both effectively and profitably with a reasonable cost to investors.

THE RELEVANT LITERATURE

There is a wide body of literature discussing different strategies that could be implemented by investors or their financial advisors to mitigate premature asset depletion risk for retirement portfolios. These include asset allocation, annuities, and usage of derivative instruments such as put and call options.

Asset allocation

Many studies have discussed static asset allocation strategies. Blanchett (2007) compared a constant allocation strategy to various dynamic strategies and concluded that constant

² In the rest of the paper, in the interest of brevity, the term 401(k) plan will be used to denote all such self-directed retirement accounts.

allocation strategies are reliably efficient, recommending a 60-40 stock-bond allocation. Israelsen (2015) argued that the classic 60-40 stock-bond allocation might not serve retirement investors, since if interest rates rise, bond returns would be low. He found that having a diversified portfolio with seven asset classes (large- and small-cap U.S. stocks, non-U.S. developed market stock, real estate, commodities, U.S. bonds and cash) was optimal. Ameriks et al. (2001) found that an aggressive portfolio with an 85-15 equity-bond allocation performed well. However, given the high risk of this aggressive portfolio, the authors suggested purchasing fixed-life annuities as well. Lemoine et al. (2010) demonstrated that an aggressive portfolio, with 100% in equities, coupled with a fixed annuity purchased when the portfolio was deemed to be sufficiently large, had the highest chance of success for meeting investment goals.

Other papers have proposed dynamic asset allocation strategies and used glide paths to describe the changing allocation of stocks and bonds. Bodie et al. (1992) argued that younger investors were able to hold more of their wealth in risky assets because of their greater labor market flexibility. This argument led to the traditional glide path, introduced by Bengen (1996), which began with higher equity exposure during the asset accumulation phase and became more conservative by increasing bond allocation when approaching retirement. Milevsky (2012) elaborated on this topic by describing equations that factor in the valuation of human capital to an investor's allocation of stocks. In contrast to Bengen et al. (2014), Delorme (2015) found that rising equity glide paths, where investors gradually increased their equity exposure as they approached retirement, were more successful. Kingston and Fisher (2014) argued that investors should have a "V-shaped" lifetime glide path, where the share of equity investments fell over the asset accumulation phase but then rose during retirement. Blanchett (2015) compared decreasing, increasing, V-shaped, and inverted-V-shaped glide paths, where allocation change was made slowly or quickly. Out of the eight different scenarios tested in many different market environments, he found decreasing fast glide paths had the highest chance of being optimal.

Recent studies have highlighted the importance of market environments in determining optimal glide paths. Kitces and Pfau (2015) argued that investors should factor in how the market was valued when they began investing for retirement. If retirement investors started saving in an overvalued environment, Kitces and Pfau recommended using a rising equity glide path. However, they also found that a static 60% equity allocation was effective for retirees who did not choose dynamic strategies. Blanchett (2015) pointed out that the differences in the findings of the various retirement asset allocation

studies were due, in part, to the differing return assumptions for stocks and bonds. His study found increasing equity glide paths to be more successful in higher-return environments, but decreasing equity glide paths were better in lower-return environments.

Annuities

Annuities are another commonly-discussed risk mitigation instruments and have been gaining popularity, particularly since the 2008 financial crisis. A recent study by Allianz reported that 61% of baby boomers feared "outliving my money in retirement" more than death [Bhojwani (2011)]. Retirees' fear of running out of money has bolstered the growth of annuities in the mainstream retirement marketplace.

Various studies have analyzed the benefits of different types of annuities. Scott (2015) found that allocating 10-15% of a retirement portfolio to longevity annuities was comparable to allocating 60% to immediate annuities; hence longevity annuities were better suited for retirees. Finke and Pfau (2015) found that deferred income annuities, which are similar to longevity annuities but with a shorter deferral period, were a good choice for retirees seeking stable income. Horneff et al. (2015) demonstrated that although variable annuities with guaranteed minimum withdrawal benefit riders were expensive, they still improved a retiree's income, especially when purchased before retirement.

Wasik (2015), who described the state of the variable annuity market, found that there were over 220 different products in the variable annuity market with an average fee of 1.4%. He found that guaranteed minimum income benefits raised the annual expense by an additional 1% - 1.15%. In addition, lifetime income benefits added 0.35%-1.25% to the annual expense. Blanchett (2013) highlighted that intermediate fixed annuities were particularly costly in low interest rate environments.

Annuities are often purchased for the security they offer, and despite being thought of as a retirement investment vehicle with a secure return, annuities are not riskless instruments. Xiong and Idzorek (2012) pointed out that annuities, like other financial products, had risks, including default risk and illiquidity risk. The authors argued that these risks should be weighed against the cost of the annuity.

Derivatives

A few studies have focused on options as risk mitigation tools for retirement assets. For example, Simonian (2011) argued that tail risk hedging using put options was necessary for retirement investing because capital preservation was almost as

important as return generation for retirees. However, Basu and Drew (2014) used historical data to show that purchasing put options to hedge tail risk was not worth the cost for active or passive mutual fund retirement portfolios. Johnston et al. (2013) also found that, since put options often expire out of the money, writing a call option worth 1% - 3% of portfolio value offered higher returns than a portfolio of fixed income securities and put options. However, put and call options are almost always unavailable in retirement accounts, particularly in 401(k) plan offerings. Even assuming one had the means to purchase put options in a separate account, using put options to properly hedge a portfolio on an ongoing basis can be quite expensive and complex.

Loss protection

Our paper is related to a recent study by Miccolis et al. (2015), which focused on how much a retiree should be willing to pay, in basis points (bps), to “buy” risk-managed investing (RMI), which provides loss protection for retirement assets. Their study examined historical S&P 500 returns to model RMI’s costs for various levels of loss protection. They found that the break-even cost ranged between 145 bps and 1,130 bps. The paper also analyzed the 29 instances in the past when the S&P 500 dropped more than 10% and found that had RMI covered half of the losses beyond 10%, it would have provided value to the retiree as long as its cost did not exceed 410 bps. The authors stressed that their cost estimation was conservative since it did not account for other factors, like peace of mind.

Our paper builds on and extends the Miccolis et al. (2015) study in several ways. First, we propose a concrete, implementable strategy through which retirees’ assets could have exact, and not approximate, downside risk protection. For example, the RMI strategies discussed in Miccolis et al. (2015) include: tactical allocation of capital based on fundamental analyses of markets; investing in funds whose strategy is to provide equity exposure with less volatility; quantitative, momentum-based strategies that provide a signal to move in and out of certain asset classes or sectors; tail risk hedging, including investing in volatility derivatives, put spreads, etc.; and combinations of these strategies, none of which provide exact downside risk protection. These strategies are complex and require retirees’ assets to be managed by a skilled manager. Furthermore, none of these strategies are available as an investment option in a typical 401(k) plan.

Second, our paper turns the conceptual framework of RMI into a viable investment product. In particular, it discusses how downside risk protection can be provided by a financial institution that is managing a 401(k) plan or by a separate financial

institution. Importantly, we demonstrate that it can be done both effectively and profitably.

Third, while we use historical data to illustrate the importance of downside risk protection, the analysis and quantification of the probabilities of asset depletion risks are undertaken through Monte Carlo simulation. This quantitative technique provides rigorous estimates of the likelihood of early asset depletion in the uncertain future with and without downside risk protection.

SEQUENCE RISK: AN ILLUSTRATION

While negative returns are always a contributor to asset balance reduction, they are particularly harmful if they occur early in the retirement phase, during which assets are withdrawn each year. The ability for early negative returns to disproportionately affect the value of a portfolio is commonly referred to as sequence risk. This risk has received considerable attention in the literature [see, for example, Kitces (2008), Frank et al. (2011), Basu (2011), Basu et al. (2012), Guyton (2013), Pfau (2014), and Miccolis et al. (2015)].

A hypothetical example illustrates the importance of sequence risk. Consider two accounts that are invested in the same assets but enter the retirement phase one year apart. Figure 1 shows the annual returns of the two retirement portfolios over time and the two starting points of withdrawals.

Although the two accounts start the withdrawal phase one year apart, they are the same in all other aspects: both have the same starting balance when withdrawals begin and both make the same dollar amount of withdrawals each year. Figure 2 shows the outcome for these two accounts. In this figure, the value of the retirement account balance is indexed to 100 at the beginning of the withdrawal phase. It shows that the one-year delay in withdrawal commencement makes all the difference. Account 1 is fully depleted by year 14, while Account 2 provides income through retirement. It is worth noting that while Account 2 provides sufficient retirement income, it also experiences some significant declines in the first few years and does not rise to a level consistently above its initial account balance until year 22 (not shown in the figure).

While one might be inclined to think that the retirement portfolio returns utilized in this example are fabricated, these scenarios are based on actual historical equity returns. In this case, the holder of Account 1 had the misfortune of retiring just as the Great Depression started in 1929, while the holder

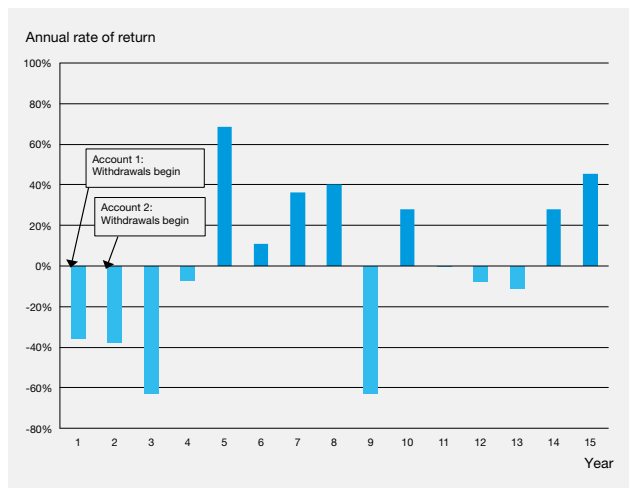


Figure 1 – Annual rates of return of retirement assets over time

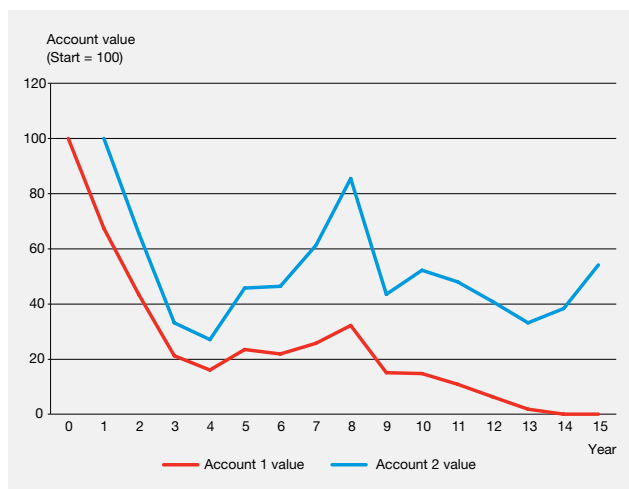


Figure 2 – The retirement account balance over time

of Account 2 retired one year later in 1930. Admittedly these two start years and the associated market returns are extreme examples, but we chose them to illustrate the issue of asset depletion risk as it relates to the timing of negative returns.

One commonly suggested remedy for the retirement asset depletion problem (illustrated in Figure 2 by Account 1) is dynamic withdrawals, as described by Stout and Mitchell (2006), Kitces (2008), Frank et al. (2011), Basu (2011), Blanchett (2013), Guyton (2013), Pfau (2014), and Delorme (2015). In the example of the two accounts discussed above, the withdrawals were set at a constant dollar amount each year. Under dynamic withdrawals, the account holder sets a percentage of retirement assets to withdraw each year rather than a set

dollar amount; this allows the retiree to reduce the withdrawal dollar amount when the retirement assets shrink as a result of negative returns.

While it is true that this approach would allow Account 1 to survive the entire retirement horizon (assuming the percentage is set low enough), consider the implications over the first 15 years: under dynamic withdrawals, in 9 of the first 15 years the withdrawal amount has to be less than half of the constant dollar amount, and in no year is the amount greater than the constant dollar amount in the first year.

While it is easy to say that retirees should adjust their spending in accordance with dynamic withdrawals if their portfolio performs poorly, in reality this can be quite difficult given that a majority of retirees’ costs are likely to be fixed for necessities such as housing, medical expenses, food, and transportation. As a result, most retirees may think of their withdrawal from retirement assets as a specific dollar amount, and not as a percentage of their assets. While dynamic withdrawals as an abstract concept seems reasonable, it is likely not a practical solution to the asset depletion problem.

Indeed, early asset depletion – not having enough assets to last through retirement – is a key issue on the minds of many retirees. And this issue, in turn, hinges on the risks of negative retirement portfolio performance, particularly large negative returns in the early years of the retirement horizon.

Our paper proposes a new approach to address the early asset depletion risk: explicit downside risk protection at the individual account level. The goal of this protection is to eliminate large negative returns of a retirement portfolio. As discussed in the next section, this type of protection can substantially reduce the risk of asset depletion while maintaining virtually all the benefits of 401(k) plans.

DOWNSIDE RISK PROTECTION: A NEW APPROACH

The DRP proposed in this paper takes the following form: in years where the account has positive returns, the account holder pays a portion of the amount gained; conversely, in years where the account suffers a loss, that loss is capped at a predetermined amount. The net effect of this protection is to reduce the retirement account’s downside risk, significantly reducing the likelihood of early account depletion.

To illustrate the effect of DRP on the performance of retirement

portfolios, there needs to be a basis for modeling outcomes in an uncertain future. Incorporating the assumption that the market's past performance over many years is a reasonable basis to model a likely range of outcomes for the future, we use equity and bond returns from 1926 through 2014. Over these 89 years, a portfolio comprised of 50% large-cap U.S. equity and 50% small-cap U.S. equity, has an average annual return³ of 10.9% and a volatility of 23.2%. Of the 89 annual returns, 25 are negative; 9 are worse than -15%.

The proposed DRP has the following specific structure: a retirement portfolio's returns cannot be worse than -15% in any given year. To pay for this protection, the account holder gives up 10% of gains in years with a positive investment return.⁴ To illustrate, consider an account which starts with \$100 and has an investment performance of -20% in that year. In this case, DRP would kick in and instead of losing \$20, the account balance would fall only by \$15, the portfolio receiving \$5 from the DRP provider. Conversely, an account with \$100 that experiences a return of +20% would only go up by \$18 with the remaining \$2 being paid to the provider of DRP.

As discussed earlier, negative returns can have a disproportionately adverse effect when they occur in the early years of the retirement time horizon. Specifically, the impact of the sequence of large negative returns matters because of dollar withdrawal, as opposed to a withdrawal based on a percent of retirement assets. For example, Figure 3 revisits the scenario discussed earlier in which two similar accounts with different retirement start dates had very different outcomes. In the figure, the dashed lines show the original account balances using actual returns. The solid lines show the balances with DRP. As the figure clearly shows, DRP has a significant positive impact on both accounts. In particular, Account 1, which was fully depleted by year 14 of retirement (dotted red line), is no longer depleted with DRP and, in fact, shows healthy account balance growth for the entire retirement horizon (solid red line).

As discussed earlier, even a single difference in the sequence of negative returns can make a significant difference in the ultimate outcome. At the outset it is not known when negative returns will impact a portfolio; hence, it is necessary to model the impact of negative returns, particularly large negative ones, at various points in time during the retirement period. A Monte Carlo simulation based on actual historical equity returns is used to create 50,000 different sequences of equity returns. Each iteration of the simulation draws a random set of returns from the set of 89 annual historical equity returns.

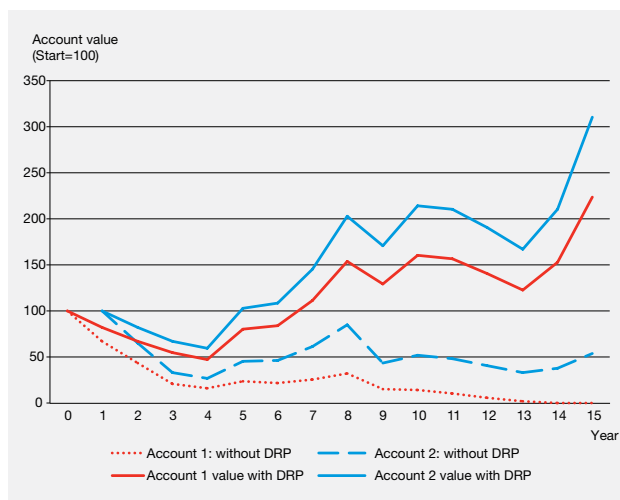


Figure 3 – Impact of DRP

As a result, each simulation iteration is based upon actual returns data, but the sequence of the returns is different for each iteration. This simulates the varying effect of negative returns early or late during a retirement horizon. Some iterations have few negative annual returns, others have a mix of positive and negative returns, and in some cases most of the negative returns occur very early on in the retirement period. It is this latter sequence of returns that leads to the highest risk of asset depletion.

Before running the simulation, two more inputs are required: the time horizon of the portfolio and withdrawals for retiree's expenses.

Retirement time horizon

According to the Social Security Administration, the average 60-year-old can expect to live a little over 22 years. But this means that 50% of individuals will live beyond 22 years after turning 60. Since no one knows ahead of time whether they will live more or less than average, the retirement time horizon used here is double the average, 45 years.⁵ This virtually guarantees that the Monte Carlo simulation has accounted for

3 All annual returns discussed in this article are based on logarithmic returns.
 4 These cutoffs can and should vary depending on the return and volatility characteristics of the portfolio. For example, downside protection of -15% on a bond portfolio makes little sense. We will also explore the impact of downside risk protection with a different threshold on a mixed stock-bond portfolio below.
 5 While this parameter is held constant in the Monte Carlo simulation, the authors have also evaluated random time horizons based upon probabilities in the SSA life tables and found this did not affect the results. According to the SSA life tables, less than 0.5% of people live more than 45 years after age 60.

the entire period over which a retiree may need to draw upon assets from the retirement account. In other words, this simulation eliminates longevity risk from the analysis. Thus, for each iteration of the simulation 45 annual stock market returns are randomly drawn (with replacement) from the 89 data points (with year-matched inflation rates).

Retiree’s expenses

The most commonly suggested sustainable withdrawal amount is 4% of assets at the start of the retirement period, first approximated as a general rule by Bengen (1994). Our simulation starts with the assumption that the account holder will require 4% of their account balance in the first year and then that dollar amount will grow by the (year-matched) inflation rate each year.⁶ That is, in the Monte Carlo simulation, in each iteration, the equity return and inflation rate are drawn from the same year. For the reasons discussed earlier, we do not consider dynamic withdrawals, where the amount withdrawn each year fluctuates with the account value.

Our paper’s main inquiry is the likelihood of early asset depletion; that is, given a fixed starting balance and an inflation-adjusted annual withdrawal amount, what is probability that a retirement account with uncertain investment performance will be depleted before the end of the retirement horizon of 45 years?

Table 1 shows the results for early asset depletion probability across 50,000 simulations for an all-stock, an all-bond, and a stock-bond hybrid portfolio. It is clear from the results in Table 1 that an all-bond portfolio’s risk is unacceptable – the likelihood of early asset depletion is extremely high, at 50%. This is largely explained by the lower average returns of bonds.⁷ The best performing portfolio is a mix of stocks and bonds.

To illustrate the outsized effect negative returns can have in the first few years, consider two different scenarios, one where DRP is utilized only for the first 10 years of the retirement period and the other where DRP is in place for all 45-years. Table 2 shows the estimated probability of asset depletion over time for an all-equity portfolio, with and without DRP.

The first row shows the estimated asset depletion probabilities at years 25 and 45 assuming no DRP. The results across 50,000 simulations show that by year 25 there is an 8.4% probability of asset depletion; by year 45 this probability is 14.6%. In the partial DRP scenario – with DRP in place for the first 10 years – the probability of asset depletion by year 45 is cut by nearly half to 7.9%. In the full DRP scenario, in which the protection is in effect for the entire 45-year period,

Asset allocation	Estimated probability of asset depletion	Annual return	Annual standard deviation
All stocks	14.6%	10.9%	22.9%
60-40 bond-stock split	8.8%	8.5%	10.3%
All bonds	50.0%	5.7%	8.0%

Table 1 – Effect of asset allocation on early asset depletion risk

Downside risk protection	At year 25	At year 45
None	8.4%	14.6%
Partial (DRP years 1-10)	2.7%	7.9%
Full (DRP for all 45 years)	1.6%	4.3%

Table 2 – Estimated probability of early asset depletion (all stocks)

Downside risk protection	Annual average return	Annual volatility
None	10.9%	22.9%
Partial (DRP years 1-10)	11.1%	21.6%
Full (DRP for all 45 years)	11.6%	16.8%

Table 3 – Portfolio performance

the probability of asset depletion at year 45 is cut again nearly by half to 4.3%. Notably, this probability is also less than half of the asset depletion probability of the 60-40 bond-stock hybrid portfolio, which is 8.8%, as shown in Table 1. Thus, DRP provides a far superior protection than the widely-used recommendation of retirement asset diversification.

Importantly, DRP not only reduces the risk of early asset depletion but it also improves the risk-return characteristics of the portfolio. Table 3 shows the estimated average annual returns and volatility of the returns of retirement assets under the three scenarios. Since the returns are logarithmic, the averages shown in this table are geometric and not arithmetic means. The first row provides the benchmark, with no DRP. In this case, the average annualized return is 10.9% and the annualized volatility is 22.9%. As shown in the table, the addition

⁶ Inflation data from Ibbotson.

⁷ The average returns for the various assets shown in Table 1 (and in subsequent tables) are based on logarithmic returns; as a result, the averages are geometric and not arithmetic means.

Downside risk protection	Median balance	5th percentile
None	65x	-12.5x
Partial (DRP years 1-10)	75x	-2.9x
Full (DRP for all 45 years)	97x	1.2x

Table 4 – Portfolio balance at year 45

of DRP has the dual benefit of increasing average return while decreasing volatility. The difference of 0.7% in average annual returns between the “no DRP” and “full DRP” scenarios might not seem much, but over a 45-year horizon this small difference has a cumulative impact of over 37%!

Thus, in addition to allowing the account holder to maintain all the benefits of a 401(k) account, and providing explicit rather than approximate downside protection, DRP provides a net benefit to the account holder in terms of higher average returns. This is in sharp contrast to other available options discussed earlier, in many of which annual costs are 1% of assets or more.

Table 4 illustrates the impact of this benefit by showing the account balances at year 45; the account balance at year 45 is expressed as a multiple of the initial balance. As one would expect, given the higher average returns and lower volatility, the median outcome and 5th percentile outcome (across 50,000 outcomes) are both markedly better for the portfolios when DRP is in place.

The impact of DRP on a stock-bond mixed portfolio

A stock-bond mixed portfolio typically will have lower volatility (and lower downside risk) than the stock-only portfolio considered so far. As a result, for DRP to provide meaningful downside risk protection, the threshold needs to be different depending on the mix of stocks and bond assets in the portfolio. Monte Carlo simulation can also be used to demonstrate the efficacy of DRP for a portfolio that has 60% bonds (equal weighting of corporate and government bonds) and 40% stocks. For this portfolio, we examine DRP with the following structure: portfolio losses are capped at -6% and in return the account holder gives the DRP provider 5% of the gains in years when the portfolio earns a positive return. Monte Carlo simulation demonstrates that DRP provides a meaningful reduction in early asset depletion probability: from 8.8% without DRP to 5.6% with DRP. The portfolio also earns a higher average annual return and experiences a lower return volatility with DRP in place. These results illustrate that by modifying the downside cap and the upside payment, DRP can work in portfolios with different mix of asset types.

CAN DRP BE OFFERED PROFITABLY?

We have modeled DRP, for a stock-only portfolio, with a cost of 10% of gains in years with positive returns. In this section, we show that this cost is sufficient to create reasonable profits for the provider of DRP.

We begin by modeling the DRP provider’s cost of hedging the exposure it faces from offering downside protection. For an all-equity portfolio, the asset holder gives up 10% of the positive returns and portfolio’s losses are capped at -15% in any given year. It follows, therefore, that the DRP provider’s payoff structure is the converse of this: it gains 10% of the positive returns and faces the full amount of the loss beyond -15%. This return profile is shown in Figure 4.

In this figure, the value of the portfolio for which protection is provided is denoted by S , and its current value is S_0 . When the value of the portfolio exceeds S_0 , the DRP provider gets 10% of the gain (note the dashed line is a 45° line). When the portfolio value falls below $0.85 \times S_0$ the provider faces the full amount of the loss beyond $0.85 \times S_0$. The DRP provider’s payoff structure is depicted by the thick solid line. It has three separate segments: when $S > S_0$ it has an angle of 4.5°, (i.e., one-tenth of 45°); between S_0 and $0.85 \times S_0$ it is flat with zero payoff; and when $S < 0.85 \times S_0$ it is negative at a 45° angle.

Thus, the DRP provider’s payoff is identical to one faced by an investor who is holding a portfolio comprised of long call and short put, with the call being at the money and the put 15% out of the money, and the ratio of put-to-call is 10-to-1; that is, there are 10 out-of-the-money puts for every at-the-money call. The DRP provider can fully hedge this exposure by buying 10 puts (15% out of the money) and selling an at the money call. The DRP provider’s net cost of hedging depends on the

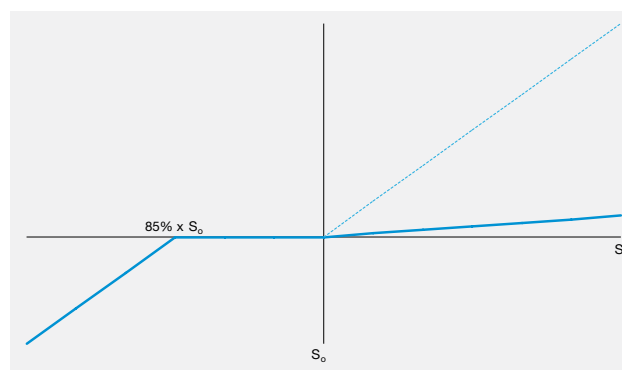


Figure 4 – The DRP provider’s payoff structure

Underlying volatility	Call revenue	Put revenue	Net revenue
10%	\$0.50	(\$0.12)	0.38%
15%	\$0.70	(\$0.74)	-0.04%
20%	\$0.89	(\$1.79)	-0.89%
25%	\$1.02	(\$3.44)	-2.42%

Assumptions: 1-year options, 2% risk free rate. Call, at the money, covering 10% of portfolio value. Put, 15% out of the money, covering 100% of the portfolio value.

Table 5 – Net hedging cost for DRP provider

volatility of the underlying portfolio. These hedging costs, at various levels of volatility, are shown in Table 5.

In this table, the column “call revenue” shows the proceeds from selling the call (which is one-tenth the price of the call) and the “put revenue” reflects the cost of buying the put. Table 5 shows, when the underlying annualized volatility is 25%, the hedging cost is 242 bps points. In this table, the options have been priced utilizing the Black-Scholes option model. We have also examined market prices for one-year-out, exchange-traded options for SPY, an ETF that tracks the S&P500 index. The observed market prices of the options suggest that the DRP provider’s hedging cost is consistent with those shown in Table 5; it is approximately 252 bps.⁸

The foregoing option-based analysis is conservative as it does not account for the various forms of additional revenue the DRP provider might receive from holding the assets of the DRP purchasers. For example, the DRP provider might be able to lend out securities being held in the DRP accounts and earn security lending fees. These lending fees could potentially offset the DRP provider’s hedging costs. Furthermore, in most years, the DRP provider would not have to pay out any money (since annual returns worse than -15% are relatively rare), but would instead be receiving cash flows from DRP purchasers equal to the 10% of the gains in positive-return years. The cash flows the DRP provider receives could be invested in relatively safe assets and returns of this investment could also defray its hedging costs.

To examine the DRP provider’s most likely returns under various market conditions, we undertook a Monte Carlo simulation using the same method described earlier in this paper: random sequences of equity returns are drawn 50,000 times from the set of 89 historical annual equity returns. The DRP provider’s pay-off structure is assumed to be identical to the one depicted in Figure 4. We have ignored the security lending income, but have assumed that the DRP provider’s account

balance earns the return it receives from the DRP purchaser plus the annual T-bill rate. In the Monte Carlo analysis, we find that the DRP provider earns, on average, an annualized return of 240 bps, which is approximately equal to the hedging cost of 242 bps, noted above.

The fact that the DRP provider likely earns a positive return is intuitive and consistent with historical data on market performance. In the past 89 years, there were 64 years where the market returns were positive; this suggests, on average, a DRP provider would be receiving a positive cash flow from DRP purchasers in 72% of the years. There were only nine years with returns worse than -15%; this means that the DRP provider would, on average, be paying out to DRP purchasers in 10% of the years.

Like any insurance, the proposed DRP approach works for two key reasons: (a) extreme negative events, where the insurance provider has to make payments, are rare; and (b) the insurance purchasers are diversified. For the case at hand, even assuming all accounts hold identical assets, the accounts would be diversified by anniversary dates, thereby creating multiple cut-offs with different payouts each year. However, in our Monte Carlo analysis of the profitability of the DRP provider, we have assumed that all retiree accounts have identical equity portfolios with identical anniversary dates. This is clearly unrealistic in assessing the profitability of the DRP provider because typically the accounts will not have identical anniversaries; as a result, the severity or the frequency of the losses would vary from one retirement account to the next. This diversity can only improve the cash flows for the DRP provider.

CONCLUSION

Our study proposes explicit downside risk protection of retirement portfolios. The proposed DRP ensures that a portfolio’s annual return can never be worse than a floor chosen by the protection provider, thereby eliminating extreme negative outcomes and protecting the portfolio against sequence risk. In return, the portfolio pays a portion of the gains to the protection provider only in years with positive returns. The Monte Carlo simulation has shown that with DRP, the chance of

⁸ The market prices of options reflect volatility skew, while Black-Scholes model assumes identical volatility for put and calls. The market prices of the options reflect an average implied volatility of approximately 25% for puts and about 13% for calls.

premature retirement asset depletion drops from 14.6% to 4.3%. DRP also reduces volatility of returns and increases the average return (net of protection cost) of the portfolio. Furthermore, we also show that DRP can be provided profitably by a financial institution.

The results of our paper should be of interest to financial institutions offering wealth management services for several reasons. First, they underscore how early asset depletion risk is enhanced by losses in a retirement portfolio, particularly when the losses occur in the early years of a retirement horizon, reinforcing the importance of downside risk protection. Second, our paper provides a viable alternative to the existing risk mitigation products and tools available in the marketplace by proposing a method for explicit and exact downside risk protection rather than approximate and expensive alternatives. Third, because we demonstrate that DRP can be provided profitably by a financial institution, it is indeed a win-win situation: retirees have the benefit of markedly lower risk of outliving their assets and of potentially earning higher average returns even after paying for the protection, and the financial institution managing the retirement assets can enhance the cash flows they receive from asset management by providing DRP to the retirees. Because this financial product is currently not provided by institutions, DRP presents an opportunity to create a differentiated and profitable service that could increase the likelihood of retention and growth of assets for the financial institutions.

REFERENCES

- Ameriks, J., R. Veres, and M. J. Warshawsky, 2001, "Making retirement income last a lifetime," *Journal of Financial Planning* 14:12, 60-76
- Basu, S., 2011, "Dynamically managing the sequence risk of retirement income," *Journal of Financial Service Professionals* 65:1, 18-20
- Basu, A. K., and M. E. Drew, 2014, "The value of tail risk hedging in defined contribution plans: what does history tell us," *Journal of Pension Economics and Finance* 17:1, 240-265
- Basu, A. K., B. Doran, and M. E. Drew, 2012, "Sequence of returns risk: the worst returns in their worst order," *In Finance: The Magazine for Finsia Members* 126:3, 36-38
- Bengen, W. P., 1994, "Determining withdrawal rates using historical data," *Journal of Financial Planning* 7:4, 171-181
- Bengen, W. P., 1996, "Asset allocation for a lifetime," *Journal of Financial Planning* 9:4, 58-67
- Bhojwani, G. J., 2011, "Rethinking what's ahead in retirement," Allianz Life Insurance Co., <http://bit.ly/2emqLPe>
- Blanchett, D. M., 2007, "Dynamic allocation strategies for distribution portfolios: determining the optimal distribution glide path," *Journal of Financial Planning* 20:12, 68-81
- Blanchett, D. M., 2013a, "Examining the benefits of immediate fixed annuities in today's low-rate climate," *Journal of Financial Planning* 26:1, 42-50
- Blanchett, D. M., 2013b, "Simple formulas to implement complex withdrawal strategies," *Journal of Financial Planning* 26:9, 40-48
- Blanchett, D. M., 2015a, "Revisiting the optimal distribution glide path," *Journal of Financial Planning* 28:2, 52-61
- Blanchett, D. M., 2015b, "Initial conditions and optimal retirement glide paths," *Journal of Financial Planning* 28:9, 52-60
- Bodie, Z., R. C. Merton, and W. F. Samuelson, 1992, "Labor supply flexibility and portfolio choice in a life-cycle model," *Journal of Economic Dynamics and Control* 16:3, 427-449
- Delorme, L. F., 2015, "Confirming the value of rising equity glide paths: evidence from a utility model," *Journal of Financial Planning* 28:5, 46-52
- Employee Benefit Research Institute, "FAQs about benefits – retirement issues," <http://bit.ly/2g1hxlz>
- Finke, M. S., and W. D. Pfau, 2015, "Reduce retirement costs with deferred income annuities purchased before retirement," *Journal of Financial Planning* 28:7, 40-49
- Finke, M. S., W. D. Pfau, and D. Blanchett, 2013, "The 4 percent rule is not safe in a low-yield world," *Journal of Financial Planning* 26:6, 46-55
- Fisher, G. S., 2014, "The price of downside protection," *Forbes*, August 6, <http://bit.ly/2fAFP9b>
- Frank, L. R., J. B. Mitchell, and D. M. Blanchett, 2011, "Probability-of-failure-based decision rules to manage sequence risk in retirement," *Journal of Financial Planning* 24:11, 44-53
- Guyton, J., 2013, "Sequence-of-returns risk: gorilla or boogeyman?" *Journal of Financial Planning* 26:10, 41-43
- Horneff, V., R. Maurer, O. S. Mitchell, and R. Rogalla, 2015, "Optimal life cycle portfolio choice with variable annuities offering liquidity and investment downside protection," *Insurance: Mathematics & Economics* 63, 91-107
- Investment Company Institute, 2015, "The U.S. retirement market, third quarter 2015," (December), <http://bit.ly/2fUU6kR>
- Israelsen, C. L., 2015, "When a 60/40 mix falls short," *Financial Planning* 45:6, 101-102
- Johnston, K., J. Hatem, and E. Scott, 2013, "Individual investors: asset allocation vs. portfolio insurance (puts or calls)," *Financial Services Review* 22:3, 291-310
- Kingston, G., and L. Fisher, 2014, "Down the retirement risk zone with gun and camera," *Economic Papers* 33:2, 153-162
- Kitces, M. E., 2008, Kitces report, May, <http://bit.ly/2fFUwqy>
- Kitces, M. E., and W. D. Pfau, 2015, "Retirement risk, rising equity glide paths, and valuation-based asset allocation," *Journal of Financial Planning* 28:3, 38-48
- Lemoine, C., D. M. Cordell, and A. W. Gustafson, 2010, "Achieving sustainable retirement withdrawals: a combined equity and annuity approach," *Journal of Financial Planning* 23:1, 40-47
- Miccolis, J., G. Chow, and R. Eggidi, 2015, "Quantifying the value of downside protection," *Journal of Financial Planning* 29:1, 50-57
- Milevsky, M. A., 2012, *The 7 most important equations for your retirement: the fascinating people and ideas behind planning your retirement income*, John Wiley & Sons
- Pfau, W. D., 2014, "The lifetime sequence of returns—a retirement planning conundrum," *Journal of Financial Service Professionals* 68:1, 53-58
- Pfau, W. D., and M. E. Kitces, 2014, "Reducing retirement risk with a rising equity glide path," *Journal of Financial Planning* 27:1, 38-45
- Scott, J. S., 2015, "The longevity annuity: an annuity for everyone?" *Financial Analysts Journal* 71:1, 61-69
- Simonian, J., 2001, "Mind the tails! Anticipatory risk management for target-date strategies," *The Journal of Risk* 13:3, 45-54
- Stout, R. G., and J. B. Mitchell, 2006, "Dynamic retirement withdrawal planning," *Financial Services Review* 15:2, 117-131
- Wasik, J. F., 2015, "Dissecting annuity fees," *Financial Planning* 45:11, 68-72
- Xiong, J. X., and T. M. Idzorek, 2012, "Estimating credit risk and illiquidity risk in guaranteed investment products," *Journal of Financial Planning* 25:7, 38-47

The Asset Management Industry, Systemic Risk, and Macroprudential Policy

Claude Lopez – Director, International Finance and Macroeconomics, Milken Institute¹

Abstract

In the aftermath of the 2007-2008 financial crisis, new legislation and regulations have pressured banks and insurance companies to reduce their size, leverage, and riskier lines of business in order to avoid another too-big-to-fail debacle. Nonbank financial intermediaries have naturally taken up some of that slack, and, not surprisingly, regulatory scrutiny has turned toward these intermediaries to evaluate whether they could pose similar risks to financial stability that banks did pre-crisis. This article explores whether there is a demonstrable link between the asset management industry and systemic risk.

¹ The article is a shorter version of the Milken Institute report titled “The asset management industry and systemic risk: is there a connection,” co-authored with D. Markwardt and K. Savard.

In the aftermath of the 2007-2008 financial crisis, new legislation and regulations have pressured banks and insurance companies to reduce their size, leverage, and riskier lines of business in order to avoid another too-big-to-fail debacle. Nonbank financial intermediaries have naturally taken up some of that slack, and, not surprisingly, regulatory scrutiny has turned toward these intermediaries, especially asset managers, to evaluate whether they could pose similar risks to financial stability that banks did pre-crisis.

Yet, most of the existing literature and regulatory tools on financial stability focus on the banking system and overlook the fact that the asset management industry and its subsectors are different from that system and perform vastly different roles. The challenge is to define appropriate framework that would provide the appropriate safeguard when it comes to the asset management industry. As a result, the appropriate macroprudential framework would require a significant departure for the current one; because asset managers do not present the same risks as banks. Yet, as discussed in FSOC (2016), FSB (2016), and FSB (2017a), they might possess other dynamics that could contribute to the transmission of – or even amplification of – systemic risk.

This article analyzes and assesses the ways in which the asset management industry might act as a catalyst or contributor to systemic risk. It proceeds as follows: Section 1 recalls the definition of financial stability and systemic risk before turning to those risks specific to the asset management sector that are of concern from a macroprudential perspective; Section 2 discusses the pertinence of the current framework in regulating asset managers from a financial stability perspective; Section 3 explores the necessity of such a role, highlighting the differentiating factors between traditional targets of macroprudentialism (banks) and asset managers; and Section 4 concludes.

FINANCIAL STABILITY AND SYSTEMIC RISK

Reforms since the financial crisis have focused on financial stability and systemic-risk mitigation. While these two notions play a key role in the current regulatory environment, defining them in a tractable, time-sensitive, and relevant manner remains a challenge.

Financial stability often is defined in terms of “its ability to facilitate and enhance economic processes, manage risks, and absorb shocks” [Shinasi (2004)]. It is worth emphasizing that such a definition does not imply protecting badly run firms or

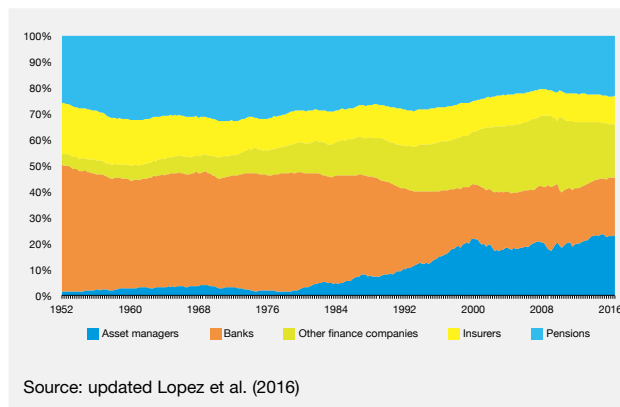


Figure 1 – U.S. financial assets by industry

creating a risk-free environment. Ensuring such stability is a complex, difficult task that requires identifying commonly agreed-upon objectives as well as their unintended consequences among regulators, firms, and clients/investors.

Conceptually, once agreed upon, these financial-stability objectives should be used to define, measure, and monitor the aspects of systemic risk deemed pertinent and “anticipatable.”² Ultimately, the relevant mix of macroprudential and microprudential tools should be used to mitigate it. Unfortunately, there are no hard boundaries between systemic and nonsystemic risk, and the ever-evolving financial landscape requires regular assessment of both objectives and how to achieve them. In other words, monitoring systemic risk and operationalizing a policy response to it remain a challenge because only the outcome of the risk, not the risk itself, can be directly observed.

Asset management: a segmented industry

Figure 1 shows the increased importance of asset managers, as they now oversee nearly a quarter of domestic financial assets, up from less than 3% in 1980.

Broadly defined, asset managers provide investment services as fiduciary agents for their clients, using a wide variety of specific asset management models. A summary of the major fund families’ characteristics and risk profiles can be found

² Systemic risk is usually defined as a “risk of disruption to financial services that is caused by an impairment of all or parts of the financial system and has the potential to have serious negative consequences for the real economy” (IMF, FSB, BIS). Yet Reinhart and Rogoff (2009) suggest that more than 50% of the financial crises come from the real side of the economy.

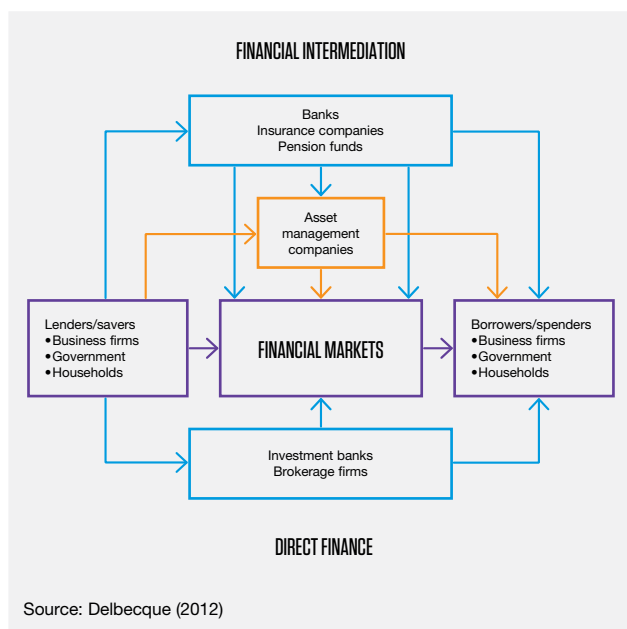


Figure 2 – Flow of funds in the financial system

in IMF (2015). They complement existing financial players in their function, as shown in Figure 2: these fund families service not only households, businesses, and governments, but also other categories of financial intermediaries, including banks, pension funds, and insurance companies.

Overall, asset managers are engaged in activities occurring either at the management-company level or at the fund level. Management-company activities include administration, centralized execution of trades, risk management, and market research, while fund-level activities include overall asset allocation, selection of specific securities, and liquidity management. Fund shareholders receive any profits or losses while the asset managers’ primary source of revenue is from fees for services.³ Furthermore, the separation between the custody and the management of assets protect investors from the risk of default of the asset manager.

From financial to systemic risk

The Basel III framework of financial reforms identifies two dimensions across which financial agents create or amplify systemic risk: either through exacerbating the extremes of the financial cycle (procyclical risk) or increasing fragility across financial sectors or institutions (contagion risk). Activities and incentives built into the asset management industry could transmit or potentially amplify risk across both these dimensions.

Theoretically, asset managers do not face the same risks as banks and insurers (other than operational risks). Yet, their fiduciary obligation exposes them to some financial risks. As a result, the question is whether the individual risks can become systemic and, if so, via which channels. This section provides a closer look at two types of risks – herding and liquidity risks – that stand out as specific to asset managers, particularly among the “plain vanilla” investment funds, such as mutual funds and ETFs.

Herding and procyclical risk

The fund management industry has traditionally operated with managers actively selecting securities on behalf of their investors. Competing for clients based on relative performance, fund managers are measured against a comparable benchmark. For portfolio managers who are risk-averse or face career risk when falling in a lower percentile of performance, there are incentives to “herd” into positions similar to those of their peers and not stray too far from the benchmarks. This can create strong disincentives for a manager to take counter-cyclical positions, resulting in “chasing yield” during upswings in the financial cycle and herding to sell positions during cycle downswings, thus exacerbating financial bubbles and the devastation of their fallout [Feroli et al. (2014)]. The IMF’s recent Global Financial Stability Report notes that U.S. mutual funds now exhibit significantly more herding behavior than in 2009, just after the crisis [IMF (2015)].

It is unclear to what extent these herding dynamics contribute to financial bubbles or if they are merely symptomatic. Equally unclear is what, if anything, can be done to mitigate these potentially destabilizing incentives. Figure 3 shows that both retail and institutional end investors appear to be moving toward cutting active managers out of the investment process and self-directing investment decisions using passive indexes.

The rise of passively managed funds – those that track indexes without fund managers actively selecting securities – introduces new potential consequences for the financial cycle and stability. The majority of passive funds buy or sell securities based on the market capitalization weights of their respective indexes. This can lead to a “momentum bias” where fund managers must buy (or sell) the fastest-appreciating (or depreciating) index constituents, again exacerbating the highs and lows of financial asset price cycles [Jones (2015)].

³ Private funds, such as hedge funds, are a partial exception to this rule, as they are not subject to restrictions on receiving performance fees, which gives the management company a direct stake in the performance of the funds.

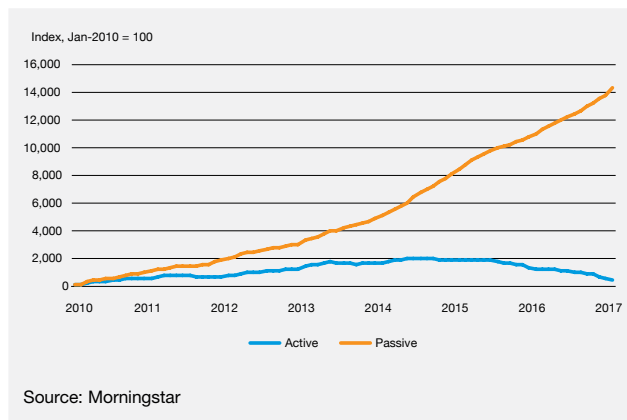


Figure 3 – U.S. fund cumulative flows

While it generally is accepted that limits to arbitrage exist that could lead to unconstrained asset price bubbles, it is less obvious that anything could reasonably be done to mitigate these unmeasurable impacts. Potential reforms such as introducing alternative benchmarks or altering investor-manager contract designs with stronger emphasis on long-term performance appraisal are unlikely to be adopted by the industry en-masse and would be difficult to enforce on a regulatory basis. Regulatory attention instead is turning primarily toward the other major perceived risk emerging from the asset management industry: liquidity mismatches in investment funds.

Liquidity and contagion risk

The implementation of the Dodd-Frank Act following the financial crisis placed greater constraints on the ability of banks and dealers to engage in various risky activities, including warehousing bond risk on their inventories (Figure 4). The result has been a sharp decline in the ability of dealers to offer two-way quotes (an offer to buy or sell a given security). While bonds have always been more difficult to trade compared with equities, given their size and lack of standardized exchange, the diminishing role of dealers in the bond market has led many fund managers to complain that bonds – corporate bonds in particular – have become increasingly illiquid.

This refers to market liquidity, the ability to trade securities without creating adverse price movements. As bond market liquidity and broker-dealer bond inventories have declined, investment funds’ ownership of corporate debt securities has risen substantially, in part displacing previous broker inventories but also in response to greater demand for corporate bond mutual funds and ETFs. Notably, as sluggish global growth and easy monetary policies have pushed interest rates to lows not witnessed in recent decades, there has been an

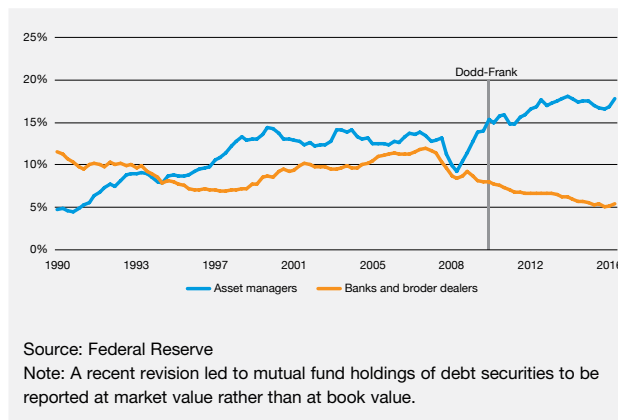


Figure 4 – Share of corporate and foreign bond ownership

increased appetite for higher-yielding instruments, such as emerging-market bonds, leveraged loan funds, and domestic high-yield corporate bonds.

While many of these higher-yielding securities have grown increasingly illiquid (and owe part of their additional yield to the illiquidity factor), the proliferation of mutual funds and ETFs providing exposure to these securities continues to offer end investors very liquid redemption terms: investors can easily buy and sell the funds on a daily basis without meaningful gates or fees. This contrast between highly liquid redemption terms and the illiquid underlying securities that the funds invest in creates a liquidity mismatch, a concern for regulators and many in the industry.

Liquidity mismatches on a large scale are of concern to financial-stability monitors because of their ability, in a worst-case scenario, to cause a “death spiral” of mass investor redemptions, causing fire-sale asset prices, which leads to further investor withdrawals. Studies find that funds investing in less-liquid corporate bonds experience disproportionately high outflows in response to bad performance and that these outflows can create destabilizing financial shocks even in the absence of significant leverage or actions by leveraged intermediaries [Goldstein et al. (2016); Feroli et al. (2014)]. Mancini et al. (2012) found that funds holding illiquid bonds during the market turmoil of the global financial crisis were forced to sell higher-quality investment-grade bonds to raise cash, thus “propagating the crisis” across the entire corporate bond sector, suggesting the potential for cross-sector contagion.

To some extent, this fire-sale scenario is analogous to countless historical examples of bank runs in which depositors rushed to withdraw their funds before the bank ran out of

money, or, more recently, the “breaking of the buck” in money market mutual funds that sparked extreme fears in the aftermath of Lehman Brothers’ collapse. Unlike banks or money market funds, investment funds do not guarantee investor balances; rather, they float with the net asset value (NAV) that provides an up-to-date cash value of the fund’s underlying investments. Nonetheless, they can still be vulnerable to redemption runs when investors have a “first-mover advantage,” as is the case with mutual funds. Focusing on the high-yield sector, Lopez et al (2016) illustrates how major disruptions to the sector’s funding environment could have a significant impact on the real economy.

THE CURRENT U.S. MACROPRUDENTIAL POLICY FRAMEWORK

The initial targets of the Basel III and Dodd-Frank reforms were banks or institutions presenting similar transmission channels in terms of systemic risk, mostly based on leverage. As discussed previously, this framework identifies two risk dimensions that may threaten the stability of the entire financial system: across institutions (contagion risk, mostly using the SIFI denomination) or across the financial cycle (procyclical risk). Both dimensions are closely linked and their problems often accumulate at the same time.⁴ This section compares the current framework with the risks it should be assessing.

Systematically important financial institutions (SIFI)

The SIFI denomination relies on the size of an institution. This proxy seems adequate when assessing the amplitude of risk that banks can generate to the system. By contrast, most fund managers tend to have simpler funding mechanisms: Figure 5 shows that they incorporate little or no leverage, while Table 1 compares the potential solvency risks banks and asset managers might experience during crisis periods when asset prices fluctuate.⁵ It also shows that some asset managers are divisions of institutions already identified as SIFIs.

If the definition of systemic risk focuses on the possibility of disruption to the real economy and the dislocation of markets, then the main concern related to the size of asset managers is the potential for direct wealth loss. However, this issue fades in importance when considered in conjunction with the interconnectedness and substitutability of an institution. Interconnectedness measures the potential of one firm to transmit financial distress to others. The more a firm is able to transmit distress, the greater potential impact its own distress can have.

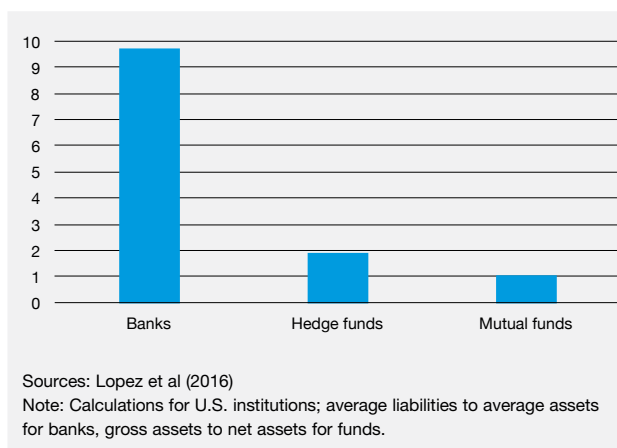


Figure 5 – Median leverage ratio (2016)

	Total assets (U.S.\$ bln)	Financial leverage
JPMorgan Chase	2,423	10.7
Bank of America	2,185	9.2
Wells Fargo	1,788	10.6
Citigroup	1,731	8.5
U.S. Bancorp	422	10.4
	Assets under management (U.S.\$ bln)	Gross fund leverage
BlackRock	4,652	1.1
Vanguard	3,148	1.1
State Street Global Advisors*	2,448	1.0
Fidelity Investments	1,974	1.1
BNY Mellon Wealth Management*	1,710	n/a

Source: Lopez et al (2016)
 * Asset managers that are divisions of SIFIs (insurers or banks)

Table 1 – Largest U.S. banks versus asset managers

4 While the size of the banks’ balance sheet and degree of leverage have been identified as potential contagion risks, Shin and Shin (2011) and Lopez et al. (2015) have shown that looking at funding sources provides information regarding procyclical risk, especially banks’ excessive reliance on “noncore” liabilities – short-term funding in particular.

5 Hedge funds often make use of short-term funding arrangements and achieve leverage synthetically through the use of derivatives, but on average they are not leveraged to the same extent as banks.

Substitutability focuses on the critical functions performed by an asset manager and the extent to which other firms could provide similar services at a similar price in a timely manner in the case of its failure. The asset management industry is an intensely competitive business with relatively low barriers to entry, hence substitutability from the perspective of investors in the market for investment management services is of limited concern. However, it is important to consider the degree to which the manager or its funds are a hard-to-replace source of financing for certain businesses or sectors of the economy. Due to both interconnectedness and substitutability, the effects of asset managers on the economy depend on the asset classes, while the channels of risk transmission (and their complexity) depend on the instruments used and how they are combined.⁶

Liquidity risks

The financial crisis has shown that a family of funds, such as money market funds, could lead to a systemic crisis via two channels: liquidity risk and connections between lightly regulated businesses and banks. As a direct response to the first issue, the Securities and Exchange Commission in 2014 adopted a set of rules that “require a floating net asset value (NAV) for institutional prime money market funds that allows the daily share prices to fluctuate with changes in the market-based value of fund assets and provide nongovernment money market fund boards new tools – liquidity fees and redemption gates – to address runs.”⁷ More recently, the SEC has proposed rules for mutual funds and ETFs to set up programs for managing liquidity risks and broaden disclosures about their liquidity and redemption practices. Furthermore, the Dodd-Frank Act requires the SEC to run stress-tests on asset managers with more than U.S.\$10 billion in assets. Since, as previously discussed, banks’ and asset managers’ business models are significantly different, the methodology needs to be adjusted. So far there is no consensus on how to define and measure the concepts of liquidity and leverage that matter in the context of systemic-risk buildup within the asset management industry. In 2017, the Financial Stability Board (FSB) requested that the International Organization of Securities Commissions (IOSCO) provides appropriate measures for liquidity by the end of 2017 and for leverage by the end of 2018.

Dodd-Frank addresses the second issue by requiring central clearing of standardized derivatives transactions. The resulting strengthening of central clearing counterparties (CCP) or clearinghouses comes with a trade-off. It makes the credit chains more transparent, providing a foundation for centralized risk-management and data-processing operations. However, it also concentrates credit, liquidity, and operational risk

within the CCPs. The Commodity Futures Trading Commission (CFTC) is also required to implement stress-tests on CCPs in order to monitor potential systemic-risk buildup, but it runs into difficulties similar to those at the SEC.

The challenges faced by the SEC and the CFTC led to the creation of a working group within the FSOC to investigate these issues, including counterparty exposures, margin investing, trading strategies, and possible standards for measuring leverage.⁸ These discussions and consultations are part of a broader international program led by the FSB [FSB (2016), FSB (2017a, b)].

Herding

Basel III is, by design, unable to discourage herding behavior because it relies on the Asymptotic Single Risk Factor Model to compute capital requirements for the monitored institutions. The model assumes that all financial institutions have a diversified portfolio and are all exposed to the same single risk factor. Wagner (2010) discusses the trade-off between ensuring that they all have the same prudent behavior and encouraging heterogeneity in risk-taking: recent reforms could encourage more correlation across banks and financial institutions. Similar reasoning would hold for asset managers if stress tests were to assess their reaction to a common shock.

MACROPRUDENTIAL POLICY FOR ASSET MANAGEMENT?

The asset management industry encompasses a wide variety of business activities, ranging from traditional asset management to alternative investing and direct lending. In other words, it is a highly-segmented industry with minimal information available to regulators attempting to monitor it. Little is known about the importance of portfolio size compared with the possibilities of nonlinear and threshold effects given the strategic situations of the institutions involved. Furthermore, given the absence of clear regulatory leadership, designing a coherent body of rules would require a significant amount of coordination among the different institutions, such as the SEC and CFTC.

6 Roncalli and Weisang (2015) generate a set of simulation to illustrate this point.

7 SEC website

8 UCITS and European alternative investment funds have been subject to such requirements and have had access to a range of liquidity management tools for some years.

While asset managers have not been the primary focus of recently introduced macroprudential policy, they continue to be affected by it. Basel III and, for the U.S., Dodd-Frank moved riskier activities (proprietary trading) off banks and onto non-bank intermediaries. New regulations are still being implemented, including the Department of Labor's fiduciary rule and the "living wills" of large banks.⁹ Furthermore, the regulatory and political momentum that followed the financial crisis is fading, leading to some questioning of the current framework and its potential expansion to the asset management industry.¹⁰ So far, regulators seem mostly focused on identifying the largest potential sources of systemic risk rather than the likelihood of a systemic shock originating from a specific institution.¹¹ This approach captures the functional risk of banks where size is an appropriate proxy of importance when it comes to systemic risk. "However, in the case of asset managers, it would confuse large institutions with systemically strategic institutions, giving wealth loss too much importance over the potential for broader economic disruptions and market dislocations." [Roncalli and Weisang (2015)]

The noted segmentation of the asset management industry explains in large part the industry's resilience as a whole, as well as its usefulness to the real economy. It is, by business design (low cost of entry, fiduciary activity), a dynamic industry that evolves and adjusts to new conditions (direct or indirect regulations, technological progress, or very low interest rates) and passes all asset-value fluctuations to its clients. As a result, monitoring and regulating the asset management industry is quite challenging. One approach suggested by both market participants and regulators is to regulate specific type of activity that provides an economic function and which, if failing, would trigger systemic crises [BlackRock (2017); FSB (2017a); ESRB (2016)]. Then, the appropriate resolution strategies should be designed to avoid such chaos. This approach implies an iterative process or rounds of communication among all parties (regulators, firms, and their clients) to secure the buy-in of all sides. Cooperation among all parties is required to minimize unexpected consequences such as pushing risky activities into a more shadowy environment or generating unrealistic expectations among investors. It would also reduce the risk posed by layers of uncoordinated regulations due to the numerous institutions overseeing part of the same industry. The current setup of the FSOC could facilitate such a process as long as it remained politically independent and a lead institution was identified to oversee the asset management industry.

Moving forward in setting the regulatory agenda, the FSB identified in its latest report four aspects of asset management activities that could potentially threaten financial stability:

liquidity mismatch, leverage within investment funds, operational risk and challenges under stress, and security lending activities. Most of the FSB's recommendations are at the fund level and rely on IOSCO to operationalize them. They suggest strengthening transparency and microprudential guidance by enhancing and standardizing data collection across jurisdictions, improving best practices, especially in terms of liquidity risk management, and stress testing at the fund level. In contrast, few recommendations focus on the stability of the financial system. They advocate for system-wide stress-tests and a risk management framework linked to asset managers' potential to disrupt the financial system. It is worth noting that these recommendations, while using Basel III's keywords stress test and orderly resolution, rely on a framework that is currently being developed while the financial system is still adjusting to sweeping post-crisis regulatory changes.

CONCLUSION

This article has highlighted the challenges of a system-wide monitoring of asset management and have questioned such an approach. The advocated alternative is to regulate by function, imposing similar regulations for institutions performing similar tasks (for example, depository institutions and money market funds) and setting requirements consistently across markets and institutions.¹²

Yet, it also seems necessary to take a step back and remind ourselves of the required, but not sufficient, elements for the successful use of prudential regulation in mitigating systemic risk. First and foremost, prudential policies are complements to – not substitutes for – proper macroeconomic policies (monetary, fiscal, structural). The current global monetary policy stance with pervasive low or negative interest rates and continued divergence among major central banks could generate financial instability that prudential policies would be unable to fix. Second, many financial markets and actors are

9 The Department of Labor's fiduciary rule is not part of the Dodd-Frank Act but an initiative competing with the SEC fiduciary rule.

10 Lopez and Saeidnezhad (2016) provide an assessment of the implementation of Dodd-Frank.

11 The SIFI denomination ignores whether the scenarios suggested in the stress-tests are likely or not.

12 See Richardson (2014): "If the risk of the underlying loans is the same, it should not matter how those loans are sliced and diced through securitization in terms of determining the required capital buffer of banking institutions."

international. As a result, successful toughening of prudential requirements necessitates international coordination, yet the political momentum for such efforts has significantly weakened in recent years.¹³ Third, the financial world is highly complex, whether due to business models or extremely integrated activities across different industries. Therefore, it is rather unlikely that any datasets will provide a complete understanding or mapping of all the risk profiles. As a result, limitations should be clearly accounted for when designing regulations and their goals.

Looking ahead, it will be important for political decision-makers and regulators to realize that the nature of systemic risk will change with the evolution of the financial landscape. Hence, the rules or policies should be targeted sufficiently to strengthen resilience of the desirable economic functions (such as lending to firms) but simple enough to limit regulatory avoidance.

The center of power in finance is shifting to the buy-side. As assets under management rise toward U.S.\$100 trillion by 2020 (according to some projections), the buy-side is poised to replace banks as the major source of funding for deals and underwriting. In the post-crisis world, regulators have as much, if not more, power as shareholders. Using this power wisely to simplify rules and minimize complex regulatory changes to the financial system, while providing the right incentives for the private sector to adopt proper governance and monitoring, seems to be the best way to achieve long-term financial stability and benefits to the real economy.

REFERENCES

- BlackRock, 2017, "Macroprudential policies and asset management," BlackRock Viewpoint, February
- Delbecque, B., 2012, "Key functions of asset management," VOXEU, March
- ESRB, 2016, "Macroprudential policy beyond banking: an ESRB strategy paper", European Systemic Risk Board, July
- Feroli, M., A. K. Kashyap, K. Schoenholtz, and H. S. Shin, 2014, "Market tantrums and monetary policy, the initiative on global markets," Chicago Booth Research Paper No 14-09
- Frankel, J. A, 2016, "International coordination," NBER working paper no. 21878
- FSB, 2016, "Proposed policy recommendations to address structural vulnerabilities from asset management activities," Financial Stability Board, June
- FSB, 2017a, "Policy recommendations to address structural vulnerabilities from asset management activities," Financial Stability Board, January
- FSB, 2017b, "Guidance on central counterparty resolution and resolution planning," Financial Stability Board, February
- FSOC, 2016, "Update on review of asset management products and activities," Financial Stability Oversight Council April
- Goldstein, I., H. Jiang, and D. T. Ng, 2016, "Investor flows and fragility in corporate bond funds," Wharton working paper
- IMF, 2015, "Global financial stability report, 2015," International Monetary Fund
- Jones, B., 2015, "Asset bubbles: re-thinking policy for the age of asset management," IMF working paper no. 15/27
- Lopez, C., and E. Saeidinezhad, 2016, "Dodd-Frank: Washington, we have a problem," Milken Institute
- Lopez, C., D. Markwardt, and K. Savard, 2015, "Macroprudential policy: what does it really mean," Banking and Financial Services 34:10, 1-10
- Lopez, C., D. Markwardt, and K. Savard, 2016, "The asset management industry and systemic risk: is there a connection," Milken Institute
- Reinhart, C. M., and K. S. Rogoff, 2009, "The aftermath of financial crises," American Economic Review 99:2, 466-472
- Richardson, M., 2014, "Regulating Wall Street: the Dodd-Frank Act, Dodd-Frank Wall Street Reform and Consumer Protection Act," World Scientific-Now Publishers Series in Business 2, 33-56
- Roncalli, T., and G. Weisang, 2015, "Asset management and systemic risk," working paper
- Shin, H. S., and K. Shin, 2011, "Procyclicality and monetary aggregates," NBER working paper no. 16836
- Shinasi, G. J., 2004, "Defining financial stability," IMF working paper no. 04/187
- Wagner, W., 2010, "Diversification at financial institutions and systemic crises," Journal of Financial Intermediation 19:3, 373-386

¹³ Frankel (2016) provides several reasons why such coordination remains a challenge, as shown in recent G-20 and G-7 summits.

The Role of Asset Owners in the Market for Investment Research: Where Are the Fiduciary Capitalists?

Alistair Haig – University of Edinburgh Business School

Neil Scarth – Frost Consulting Limited

Abstract

Fiduciary capitalists, such as leading pension plans and endowments, can be influential in aligning the interests of asset management firms with their clients. In the market connecting investment professionals with the information they need to meet client goals, we identify numerous conflicts of interest, but find little action has been taken by asset owners. Interest in the obscure practices surrounding the use of dealing commissions for research has heightened since 2014 due to regulatory scrutiny in the U.K. and the impending implementation of the second Markets in Financial Instruments Directive (MiFID II) in Europe. The authors make recommendations to guide asset managers and asset owners through a complex information market during this time of dramatic change.

INTRODUCTION

Financial capitalism, the prevalent operating system behind global financial markets, has been highly criticized in the years since the 2008 financial crisis. Despite this, we can find examples where the collective power of asset owners has succeeded in improving end-investor outcomes. Rogers (2014), citing Hawley and Williams (2000), observes the emergence of an alternative system driven by asset owners acting as “fiduciary capitalists” to improve alignment with end-investors’ long-term goals.

Fiduciary capitalists select asset managers in order to meet investment goals. Research will only be of value if it helps asset managers to meet these goals. As a result, asset owners are in a strong position to call for higher standards of transparency and objectivity.

We briefly review the forces of change in the information market connecting fund managers to external sources of investment research. Despite its obscurity, this market provides an important link between investment management firms and thousands of research providers, such as brokers and independent firms, around the world. It is also large: estimated to turn over in excess of U.S.\$20bn per annum according to Frost Consulting Estimates. The U.K.’s Financial Conduct Authority (FCA) estimates the U.K. market to be £1.5bn.

Our ongoing work with firms, regulators, and industry and professional bodies, combined with evidence collected through surveys and interviews, reveals little evidence of fiduciary capitalism in this market to date. Asset owners, who represent end-investors and could, therefore, be expected to mitigate conflicts of interest, have been quiet in this debate. Regulators and entrepreneurs, and indeed the buy-side and sell-side firms themselves, appear to be the agents of change.

ASSET MANAGERS REMAIN HUNGRY FOR INFORMATION

Rogers (2014) cites the shift to lower cost index-based strategies as a result of fiduciary capitalism. Passive index strategies have grown since the introduction of index funds in the 1970s and now account for some 14% of assets under management (AUM) [BCG (2015)], and some commentators speculate that index funds and ETFs could make up 50% of AUM by 2024 [Rivas (2017)]. Rogers notes the important role played by asset owners in shifting to such strategies to meet investor goals rather than overpaying for the hope of short-term performance.

Despite the ascent of index investing, active management remains the prevalent type of equity fund management around the world. Indeed, it will do so even if passive management doubles in size. Active management is likely to remain an important segment of equity ownership for decades, much as it prevails in most other asset classes.

Active managers need research in order to make decisions in the face of uncertainty to meet investor goals. Consequently, buy-side firms have to either produce their own research or buy it from third parties. Most choose ingredients from both sources and the recipe will depend on the availability, quality, trustworthiness, and costs associated with each source.

WHAT EXACTLY IS RESEARCH?

Investment research comprises much more than written analyst reports [CFA Institute and CFA Society of the U.K. (2014)]. Customized analysis, quantitative models, and analyst time are just some of the services that investment managers value. Despite frequent claims to the contrary, most asset managers remain heavily dependent upon broker research.

In the U.S., asset managers can use commissions to purchase data and gain access to company management in addition to procuring financial analysts’ research. In the U.K., only the latter is permitted, and buy-side firms must also pay for raw data and corporate access with their own, not their clients’, money. Definitions of research in other markets tend fall between these two markets.

Unlike some economic goods, the value of research is in the eye of the beholder. Consequently, by definition, there is no “right price.” Regulators are in no position to tell an asset manager that a particular product or service is not “substantive” in relation to their investment process.

Many asset owners would also find it difficult to evaluate research efficiency, but this is largely due to lack of information. Research is procured to improve the chances of meeting investor goals. While this may be consistent with asset owners prioritizing investor goals over short-term alpha [Rogers (2014)], it is perhaps puzzling that scrutiny on research costs, or at least demand for attempts to value research, has not been higher in the past. This puzzle can be explained at least in part by a lack of understanding of this complex market, which is briefly explained in the following section.

HOW DOES THE RESEARCH MARKET WORK?

The means of paying analysts for investment research is strikingly different to markets for most other professional services. Investment management companies can pass on the cost of research to the funds they manage, meaning that their clients in fact pay for the research. This is done using research commissions that are paid to brokers when shares are bought or sold. Unsurprisingly, most research has typically been purchased using commissions because this way the fund management company does not bear the cost. Additionally, it is most unusual to find contractual arrangements between the investment management companies and research providers, for such things as billable hours, service levels, or specified deliverables.

Fund managers decide how to reward analysts for various research services on an ex-post basis, i.e., after consumption. A typical broker vote process would involve fund management staff deciding how to allocate commissions at the end of each period, typically six months. For example, an equity fund manager might pay a given brokerage firm 7% of its firm's total commission allocation as payment for research. This information would be translated into a target allocation for the buy-side dealers to execute in the coming period. As a result, research would be paid in arrears. Detailed analysis of a U.S. broker vote process can be found in Maber et al. (2014), but such processes no longer comply with U.K. regulation today or European regulation post MiFID II. In 2006, U.K. regulation created a payment mechanism that allowed research commissions to be paid to other research providers, thus ending the one-to-one mapping between execution and research relationships. It also paved the way for hundreds of independent research providers. This mechanism, analyzed by Haig and Rees (2017), and usually called the Commission Sharing Arrangement (CSA), has equivalents in the U.S. and other markets. Figure 1 shows estimated CSA adoption aggregated across U.S. and European markets.

PROBLEMS ASSOCIATED WITH THE BROKER VOTE

The traditional broker vote process has a number of problems [CFA Institute and CFA Society of the U.K. (2014)]. First, because the vote payments are percentages of commission paid, which outside the U.S. is typically determined by the trade value rather than number of shares, the price of a certain service in dollar terms can fluctuate from year to year due to changes in AUM (which is affected by market prices of underlying securities, fund performance, and fund flows), as illustrated in

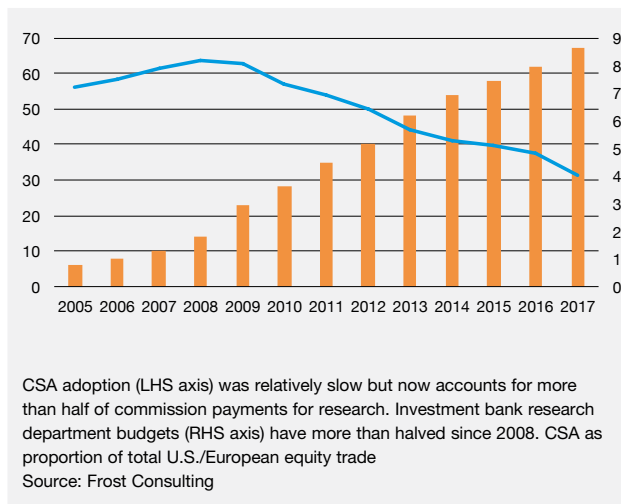


Figure 1 – Adoption of CSAs and investment bank research budgets (2005-2017)

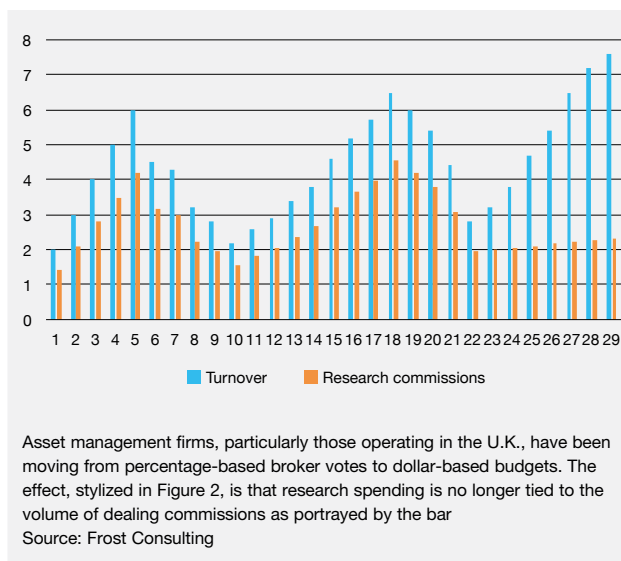


Figure 2 – Research commissions before and after the introduction of research budgets

Figure 2. The fund manager would be charged more for exactly the same research just because of an increase in stock prices or fund inflows. Second, the fund manager needs to trade in order to pay commissions to the broker, which creates the incentive to trade even if transactions are not required. Traditionally only the executing broker could be paid for research and brokers competed for bundled commissions on the strength of their analyst research. Third, broker votes have often failed to provide useful feedback to brokers regarding the services

required. We have strong evidence [Extel (2011)] that the process has been lacking in detail, accuracy, and timeliness. The FCA view the broker vote as “inherently flawed” [FCA (2012)].

CURRENT STATUS: A MARKET IN FLUX

Some 15 years after the influential Myners report [Myners (2001)], which proposed to ban research commissions in the U.K., the FCA has finally succeeded in elevating the importance of the research market. The FCA's 2011-12 thematic review and subsequent consultation lead in 2014 to clarification on the definition of research and the requirement for the CEO of the largest 200 asset management firms operating in the U.K. to make a personal attestation regarding the use of commissions for research. By interpreting research as an inducement to trade under MiFID II, it also supported a ban on research commissions commencing in January 2018. As a result of London's scale in global investment management, research payment has ascended the “to do” list for asset management firms around the world. The regulatory spotlight on this area has intensified and is unlikely to diminish.

The U.K. and some other European regulators have sought to break the link between turnover and research payments. Rising equity markets led to larger AUM, higher share prices, and often increased turnover typically result in larger research commission payments even if most asset management organizations consume similar levels of research service from one year to the next. Going forward, payment for a similar service level is likely to vary much less over time.

MiFID II will require investment managers who wish to pay for research using dealing commissions to create Research Payment Accounts (RPAs) based on a research budget that is to be set in advance. The research budget must be independent of trading, thus removing any incentive for fund managers to trade excessively in order to purchase research.

By mandating finite monetary (rather than percentage-based broker vote) research budgets, and encouraging managers to adopt board-level research budget approval processes, European regulators have largely achieved the FCA's aim of breaking the link between equity turnover and research payments. The outcome echoes Myner's (2001) call for fund management firms to compete by using research efficiently to meet client objectives. Yet, the impact is now far wider than Myner's U.K. remit. Research consumers and producers around the world have tightened up policies in this area.

Given the vast change in regulatory environment, and the resulting change in the economics of the research industry, asset owners should now question how their underlying managers are responding to these industry changes. Most asset owners routinely and systematically measure the impact of their managers' trading decisions via trade cost analysis. The efficiency of execution commissions has been regularly reported to asset owners since MiFID (2007) or before. In contrast, research commissions have typically not been reported. Ironically, the performance impact of sub-optimal execution, which could exceed 100 basis points per annum in only the most extreme events, is dwarfed by the impact of sub-optimal use of research: poor asset allocation or stock selection decisions could easily lead to underperformance of 100 basis points per annum, or even 1000 basis points, depending on the strategy.

Many investment management firms have collected insufficient information on their use of research commissions and as a result have been unable to measure the return on investment of their research spend. As a result, few have been able to present such information to end-investors. Senior officials at investment management firms consistently report that clients remain generally uninterested in valuing research.

WHY HAVE FIDUCIARY CAPITALISTS NOT BEEN MORE VOCAL?

We believe that the following reasons have impeded asset owners from demanding clear and transparent information on the cost and efficiency of research purchased with their money.

First, other regulations aimed at improving alignment with end-investor goals, such as the 2012 U.K. Retail Distribution Review, have been taking effect. Similar initiatives have taken place in other markets. Investment managers and advisers have been right to focus on implementation of these high-profile regulations.

Second, the opaque nature of the payment mechanism made it hard to see the costs involved. Limited awareness even of the existence of research commissions is perhaps understandable given that few buy-side firms presented research costs at all.

Third, low awareness of the mechanics of research commissions provided media and the public with limited understanding of the issues. The FCA's 2013 Thematic Review changed this and specialists within the financial press now keenly study the issue on both sides of the Atlantic and even globally.

Fourth, the 2008 financial crisis and resulting gyrations in equity markets required asset owners to focus on other priorities in order to survive long enough to consider this issue of longer-term consequence.

It remains unclear whether asset owners have a fiduciary responsibility to monitor their managers' research spending and its relationship to fund returns. CFA members will recognize their responsibility to meet CFA soft dollar standards, which provide guidance on how to use client brokerage ethically. The standards recognize the possible conflict of interest between the buy-side firm and its clients that arises from the opportunity for an investment management firm to offset some fixed costs through the use of services paid for via client commission. The standards seek to require members to manage that conflict appropriately through their own actions and by providing clients with the information that they might need to monitor their managers' behavior.

Note that fund managers can buy whatever research they want if they pay with their own money and asset owners should also consider procurement in their evaluation.

WHAT CREATES THE BEST OUTCOME FOR INVESTORS?

Is the lowest possible research cost in the best interest of the asset owner if it results in sub-optimal research provision and investment decision making? We believe that the goal should be to maximize efficiency rather than minimize spending. Asset managers should be expected to align the research budget with the investment strategy, investible universe, and expected returns at the fund level.

As always, there are likely to be costs to regulation as well as benefits. Close relationships with sell-side analysts provided fund managers with tailored information, thus providing the best shot at market outperformance, an outcome which is entirely in the end-investor's interests. Cross-subsidies between business units at banks provided a model that allowed fund managers to benefit indirectly from expertise and services beyond those of the equity research department. Investment banks struggled to limit the dissemination of research and much was often available to smaller fund managers, thus helping them to compete against larger firms. Given the social complexity and economic dynamics of the interface between buy-side and sell-side experts, it seems unlikely that more rigid regulation could not come without costs to the end-investor. This key point is frequently lost in the debate.

The original MiFID II proposal to require asset owners to approve their asset manager's proposed research budgets would directly involve asset owners in the research funding discussion. U.K. pension trustees are frequently not investment professionals and, therefore, not usually qualified to judge complex and variable research budget proposals from widely differing investment strategies. They face questions such as the following: is the same research budget appropriate for a distressed debt fund and a highly leveraged emerging market equity hedge fund? What is the "right" price for research? What is the relationship between research budget spending and end-investor's outcome (expected returns)?

It is likely that multiple answers will emerge. Different firms representing different sets of asset owners should be encouraged to articulate the best practice to suit their end-investors. The U.K. National Association of Pension Funds (NAPF) has recognized the need for a principle-based approach balancing the appropriateness and alignment of the research budget with the underlying investment strategy and expected returns.

HOW MIGHT ASSET OWNERS EFFECT CHANGE?

While acknowledging that there are different ways to succeed in aligning research procurement with client interests, we identify several ways that influential asset owners, such as sovereign wealth funds and pension plans, could effect change.

First, fiduciary capitalists will lead the efforts to compare research costs to investment goals and will demand information to be presented in their preferred format and frequency. Asset managers will then be required to provide such information in the course of client reporting and when competing for mandates. International regulatory coordination on research procurement has typically been limited. Major asset owners have the power to improve the practices of investment management groups worldwide. This could avoid damage to competition between geographic investment management hubs due to regulatory arbitrage. Although MiFID II provides the opportunity of consistent regulation across one continent, therefore reducing the risk of regulatory arbitrage, a relatively stricter interpretation and enforcement of the delegated acts in some European markets could discourage fund managers from operating in those locations. Reduced competition has also been argued to risk the loss of high quality fund management jobs in countries where research payments are most restricted.

European regulators have stepped back from banning the

use of commission for research, thus averting a significant trans-Atlantic non-tariff barrier in international capital flows (use of commission for research is enshrined in 28(e) of the Securities and Exchange Act of 1934, a venerable statute that would have been unlikely to change). Yet, differing national interpretations of the same MiFID II text mean that the playing field across Europe may not be entirely even.

Second, asset owners need to be aware that unbundling could lead to potential concentration in the investment management industry. Bundled commissions supported smaller buy-side firms: effectively they were subsidized by larger buy-side competitors. Although undesirable in terms of fiduciary responsibility, this acted to level the playing field. Start-up asset management firms often seek access to investment bank research in their early days while operating on seed funds. Unbundling, therefore, presents a higher barrier to entry to new fund managers and may encourage a further shift in power to larger asset management groups.

Third, asset owners should demand that investment managers adopt the following practices.

1. Research budgets should be set based on an independent review rather than by portfolio managers. Aggregate research commissions should require board approval. Ongoing internal consistency checks under the oversight of the investment management firm's Chief Financial Officer or equivalent should be reviewed in an annual audit.
2. The firm's compliance team, not the portfolio management team, should manage the process. Portfolio managers may, however, shape the design of the policy within their firm.
3. Appropriate records of research consumption should be maintained to the highest regulatory requirements globally. In most firms this will require improved accounting practices.
4. Provide clear and consistent feedback to research providers as to what products/services are valued.
5. In time, research budgets should be monitored against quantitative benchmarks. Such benchmarks are likely to emerge and become available by the end of the decade. In the interim, a clear comparison with previous years will allow asset-owners to evaluate research efficiency.

The focus may move beyond equities to fixed-income research, where commission is not paid. Moves to bring fixed income markets into line will present a major change for many bond fund managers that were not able to use CSAs but will be required to initiate RPAs as they move to price research.

In particular, multi-asset managers may be asked to present research costs for equities, bonds, and other types of investments. We believe that asset-owners may be more effective than regulators in improving research procurement practices in non-commission markets.

CONCLUSION

Research procurement has seen a murky past. Some ten years after the 2006 introduction of CSAs in major equity markets we observe that transparency is improving. Investment managers are moving towards better practices.

Significant improvements in the first half of this decade largely stem from the responses of research consumers and producers to U.K. regulatory change. Such practice has been mirrored around the world to varying degrees. The spotlight has been directed at research procurement and as a result the topic has moved up significantly on the "to do" list for those managing investment firms. The issue is here to stay. End-investors stand to benefit.

Yet, fiduciary capitalists appear to remain largely silent on issues relating to the use and payment of research. Like other participants, asset owners will have been watching the interplay between regulators, government agencies, firms, and bodies representing firms and professionals in the lead-up to MiFID II. The interpretation by regulators in Europe and other important markets, and the response from firms developing global policies, will take longer to emerge. In the coming years, from 2017-2018, research valuation information will become more available for asset owners who will then aggregate and compare research costs to custodial and other costs. More informed asset owners will become more vocal and will perform an important monitoring function.

In this paper, we provide recommendations to assist them to make this important change. We expect the level of scrutiny of research procurement to reach a higher bar within the next few years. Compliance, transparency, and fiduciary responsibility is likely to increase and compliance departments the world over will be busy ahead of MiFID II taking effect from January 2018.

The impact has spread well beyond the U.K. The CSA mechanism allowed independent research providers to enter the market in greater scale. Buy-side research budgets are expected to fall as poorly justified elements of bundled research

are removed. However, despite the challenges to research budgets, there is sufficient commercial demand for independent research to fuel innovation both directly and indirectly.

Rather than acting to minimize the cost of research, we recommend aligning the research budget with the investment strategy, universe, and expected returns at the fund level.

Asset managers have fiduciary responsibility to act in their clients' interests. When paying for research there should be a clear demonstration of the expected value of that research in obtaining the investors goals. Research consumers and producers have been vocal in providing feedback to proposed regulations. Asset owners, in contrast, have been watching quietly. Given that they may have a fiduciary responsibility to evaluate research spending, we expect this group will be the next to take action to further improve the lot of the active investor. Most likely, this shift will occur once MiFID II has been integrated into member state regulations. Fiduciary capitalists will then use their power to improve alignment of investment manager action with end-investor goals.

REFERENCES

- BCG, 2015, "Global asset management report," Boston Consulting Group
- CFA Institute and CFA Society of the U.K., 2014, "Investment research valuation approaches: a framework for investment managers and asset owners," <http://cfa.is/2kD7ZTq>
- Extel, 2011, "Buyside broker reviews: commission or omission?" <http://bit.ly/2lkVtbj>
- FCA, 2014, "The use of dealing commissions regime," Financial Conduct Authority, discussion paper no.: 14/3
- Haig, A., and W. Rees, 2017, "The changing market for investment research: the impact of innovations in the payment mechanism on market structure and information flows," working paper, <https://tinyurl.com/jpmhrtc>
- Hawley, J. P., and A. T. Williams, 2000, *The rise of fiduciary capitalism*, University of Pennsylvania Press.
- Maber, D., B. Groysberg, and P. Healy, 2014, "The use of broker votes to reward brokerage firms and their analysts' research activities," Working Paper
- Myners, P., 2001, "Institutional investment in the United Kingdom: a review," H.M. Treasury, <http://bit.ly/2kO8D2G>
- Rivas, T., 2017, "Passive investing could claim the lion's share of assets as soon as 2021," *Barron's*, <http://on.barrons.com/2kDbr0G>
- Rogers, J., 2014, "A new era of fiduciary capitalism? Let's hope so," *Financial Analysts Journal* 30:3, 6-11

Risk, Data, and the Barcodes of Finance

Allan D. Grody – President, Financial InterGroup Holdings Ltd

Abstract

The failure of Lehman Brothers uncovered a shocking fact that after six decades of automation there are no global standards for the identity of financial market participants nor their financial products. Bankruptcy attorneys and forensic accountants tried to understand Lehman's exposure to others and others' exposure to Lehman. There was no consistency in identifying Lehman as a counterparty; no understanding of what relationships Lehman had with others; and no mechanism to associate all of Lehman's products and businesses into a holistic view of the exposure others had to Lehman should it fail. In effect, no one, be they regulators, creditors, or counterparties, could see into Lehman's exposure to risk. This paper illuminates a fundamental component of the financial system that goes largely unrecognized as a key pillar of finance, non-standardized financial transaction data. Financial transactions lack unique, universal, and unambiguous identifying codes for the

supply chain of financial market participants and the products they manufacture, issue, own, process, and trade. Imagine if every supermarket had a different barcode for the same product on its shelves or a different code for the producer or supplier of the product. Walmart, FedEx, and Amazon could not exist. This failure has exasperated researchers, analysts, financial institutions, and regulators, who are forced to map and scrub this data before aggregating financial transactions for performance, risk, and regulatory reporting. Considerable risks, costs, and delays in receiving payment are also inherent in this reconciliation process. This paper explores the history, status, issues, work yet to be done, and recommendation by the author to create financial industry identity standards. The "barcodes of finance" will enable an automated means to aggregate risk data so that firms can reduce risk and costs, and regulators can oversee the largest systemically important global financial firms.

INTRODUCTION

There has been a long standing and unfulfilled global interest amongst members of the financial services industry to create a unique, unambiguous, and universal identification scheme for financial market participants and the instruments and contracts they manufacture, issue, trade, own, and process. While praised initially by global standards setters as a transforming pillar of financial reform, regulators who are working through the first implementations of a global financial industry identification scheme are already expressing caution. A “hurry-up to get the regulations finished” approach is dominating the regulatory agenda without considering its consequences for effective and efficient implementations and fitness for purpose.

Long overdue and now driven by the needs of financial regulators, the “barcodes of finance” (a globally unique transaction identifier – UTI, product identifier – UPI, and legal entity identifier – LEI) are expected to provide financial regulators with an automated means to observe the buildup of enterprise risk across silos of businesses within each financial institution and systemic risk across the global financial system. For industry members they are expected to allow for straight-through-processing (STP) and cost efficiencies as the barcodes of commerce had done for the commercial and retail trade supply chain.

These issues and regulators’ responses go back to the Group of Thirty’s (G30’s) study of the 1987 global market disruption, which resulted in the recognition of an interconnected global financial system. The G30 is a private think tank made up of retired central bankers and financial regulators, and chaired by Paul Volker, the retired U.S. Federal Reserve Chairman. The G30’s study offered many recommendations, amongst them a need for global standards of identification and their associated reference data.

In 2006, after twenty years of monitoring implementation, the G30 monitoring committee concluded that the industry had not solved the problem of non-standard identifiers and reference data. The financial crisis began in 2007 and reached its catalytic event in 2008 with the Lehman failure. It was at that point that regulators recognized that they could not wait for the industry to subordinate their own interests in furtherance of solving the industry’s collective action problem and fixing its own plumbing. Regulatory compulsion was applauded and an integrated identification system was requested, one which satisfied the industry’s desire for STP and the regulators’ need to observe accumulating systemic risks. Implementation in

one area, swaps transaction reporting to swaps trade repositories in the U.S. began in 2012 and in the E.U. in 2014.

Praised initially by global standards setters as a transforming and necessary pillar of financial reform, regulators are already issuing new public consultations on components of the financial barcodes and their associated reference data. These new consultations were necessitated by the recognition of the haphazard and premature implementation under prior reform regulations. Regulators and industry members who are living through the first implementations of these identification schemes are already expressing caution as documented in this paper. Others, not yet in implementation mode are moving ahead, possibly unaware of the expedient legacy solutions that are being patched into the existing infrastructure.

These Rube Goldberg or Heath Robinson implementations, incrementally complex machine constructions designed to accomplish a simple objective, are creating multiple additional layers of financial data intermediaries and identifiers that are reinforcing the high cost, high risk data mapping exercises that are duplicated across the financial supply chain. This layering-on of additional technology infrastructure is occurring as more is learned from the tens of billions of transactions containing these financial barcodes now sitting in swaps data repositories with no computerized means of accessing them or any means of aggregating them for systemic risk analysis. This later point being the first objective for their use.

A clear path toward eliminating risk and excessive costs was the promise regulators made to the industry in embarking on a new global identification scheme.

The intent of this paper is to assess the current state of implementation of the barcodes of finance and their associated reference data and to propose solutions to the known issues that have arisen. This reassessment is necessary to understand the impact of what may prove to be the early dysfunctional starting point of the journey, a U.S. regulator’s premature adaption of a LEI code while it was still under construction and which subsequently changed, and the expectations that set regulatory compulsion as a necessary enabler and global acceptance as its ultimate goal. Additionally, by examining similar programs of global identification standards successfully implemented in other industries, notably the global commercial barcodes and the domain names used in email and Web services, these known issues may be resolved and yet unknown issues prevented.

BACKGROUND

The global financial system is under pressure to adjust to a new regulatory order and to the new technologies of the digital age. The lessons of the financial crisis taught us that global financial institutions transcended sovereign boundaries of regulation. Another lesson was that the ability of regulators to observe risk building up in the financial system is critically dependent on accurate, timely, and aggregated financial transaction data. A more fundamental observation is that the discipline of risk management had for too long neglected improvements in data management.

The G20's new global standards setter, the Financial Stability Board (FSB), has been tasked with first creating data identification (ID) standards for uniquely identifying participants, products, and transactions in the financial supply chain. We refer to these ID standards as the barcodes of finance. They are then to be embedded in financial transactions and used by both regulators and industry members in automating regulatory reporting and in STP. STP has long been the unfulfilled vision of the financial services industry, described as the means to completely automate the life cycle of a financial transaction.

To accomplish STP, standard identifiers must be associated with standard reference data, the terms and conditions (or "metadata") that impart meaning to the identifiers (data elements such as dates, prices, rates, etc.). Another component, standard data "tags," are necessary to accompany the standard IDs and standard reference data to enable each ID and data element to be accessed, understood, and processed by computers.

Initially, these barcodes are to be incorporated into the reporting and recordkeeping of swaps transactions, the first globally constructed and newly regulated financial market to result from remedies from the financial crisis. In turn, global banking's risk standards setter, the Basel Committee on Banking Supervision (BCBS), has advocated their use in aggregating data for the reporting of risk. Some sovereign and regional regulators have incorporated these IDs into their own regulations. Many, however, have made them provisional, awaiting the finalization of the standards, the associated reference data, and the systems to operationalize the codes issuance and use.

The objective for the deployment of this global ID regime, as stated by regulators, is to aggregate financial transactions to observe a single firm's enterprise risk and multiple firms' systemic risk across the globe. To this end, financial transactions can be thought of as a set of computer encoded data

elements. These data elements collectively represent standard identification data: identifying the transaction uniquely (the unique transaction identifier – UTI) with a specific instrument or contract (unique product identifier – UPI) bought by a specific business entity (legal entity identifier – LEI); identifying variable transaction data, such as quantity and amount; and identifying associated referential data such as price data, credit ratings, and other types of fundamental data. Analogous to specific component items of a manufactured product, reference data also defines the products' changing specifications (periodic or event driven corporate actions such as mergers, acquisitions, and spin-offs), occasional changes to sub-components (calendar data, reset dates, credit ratings, historical prices, betas, correlations, volatilities) and seasonal incentives or promotions (dividends, capital distributions, and interest payments).

The first test of standard identifiers and reference data is in its use for data aggregation of swaps transactions reported to newly created trade repositories. This has not yet proven successful even though billions of transactions are already being reported with these codes. Recent regulatory and industry initiatives are focusing on incrementally adjusting the coding scheme and associated reference data in an attempt to correct known problems as regulators continue to attempt to rollout the ID system across the globe. The Bank of England recently reported on its attempt to use the reported data in just one market, foreign exchange derivatives, and found significant data quality issues with newly created UTIs, UPIs, and LEIs.¹

It may well be that in the haste to get the regulations passed in the newly regulated swaps markets, regulators implemented a coding scheme not fit for all its intended purposes. It is apparent from regulators' own words, as documented in this paper, that the issues already discovered in swaps data reporting needs to be fixed before their further consideration of use elsewhere.

Historically, and to this day, and as planned in the "new" identification system, an assortment of private and public data suppliers, many being for-profit data and technology vendors, stand between original sources of data generated by financial market participants (corporate issuers of securities, derivative contract creators, mortgage originators, loan and securitization dealers) and their interpretation into computer readable form. This additional layer of data intermediaries between

¹ Bank of England, 2017, "Gauging market dynamics using trade repository data: the case of the Swiss franc de-pegging," Financial Stability Paper no. 41, <http://tiny.cc/zv2kiy>

originators and ultimate consumers, while necessary in the past, with new technologies now available only serves to persist past “best practices” that have failed to attain high quality standardized data.

A new generation of digital technologies, such as reporting and product data tagging taxonomies [eXtensible business reporting language (XBRL) and financial product markup language (FpML) being working examples in finance] are already providing standardized means of identifying source data that can be directly input into financial systems. Further, the new technologies of cloud computing, distributed ledger technology (DLT) – a component of Blockchain technology – and smart contracts are ushering in an even newer digital era where collaboration, shared utilities, and a new vision for STP by eliminating data intermediaries and market utilities is being contemplated and experimented with. In this evolving technology environment, new thinking is being applied to the way in which the “new” identification system is being constructed, particularly in applying DLT concepts to the LEI.²

Notwithstanding these promising new technologies, industry infrastructure entities supported by their trade association constituents and compliant regulators are creating and interposing another set of data intermediaries and market utilities into the already complex, costly, and risk prone industry infrastructure. If this approach persists, it will perpetuate the practice of proprietary and duplicate codes and poor quality reference data arising from using secondary sources and multiple interpretations of these same sources, and from the timing differences in updating data. In the case of the LEI, it will continue the need for risk prone mapping services for parent/control/ownership hierarchies and linkages to the UPI where issuer, obligor, and guarantor relationships must be understood for analyzing risk. Rather than improve the situation it will leave the financial services industry where it has come to today, with additional processing intermediaries and unnecessary infrastructure costs for reconciling poor quality data.

How the financial system got so far before it nearly collapsed without a global set of standardized identifiers and their associated reference data for contracts, instruments, counterparties, and financial market participants is not the intent of this paper to describe. These issues and regulators’ responses go back to the Group of Thirty’s (G30’s) original study of the 1987 global market disruption. Chaired initially by then Chairman of Citibank, John Reed, the study resulted in the recognition of an interconnected global financial system and a need for adoption of uniform messaging standards and communication protocols.

In 2006, after twenty years of monitoring, the G30 study concluded that the issue had not been resolved. The G30 stated, “While all large financial intermediaries have moved to adopt common international standards, most infrastructure providers still operated proprietary standards.”³

The financial crisis began in 2007 and reached its catalytic event in 2008 with the Lehman failure. The public record and the academic literature since then describes the failure of Lehman Brothers and its revelations that drove the leaders of the G20 nations to place a global identification system on the regulatory agenda. Regulators recognized that they could not wait for the industry to fix its own plumbing, they had been given two decades to do so by the G30 and failed to make meaningful progress. Regulatory compulsion was applauded and an integrated identification system was requested, one which satisfied the industry’s desire for STP and regulators’ need to observe accumulating systemic risks.

After the financial crisis, regulators charged with implementing the barcodes of finance, after almost an additional decade of trying, summed it up succinctly: “Data standards are not an issue to set the pulse racing. In addition, the technical challenges involved in arriving at a global common financial language are nothing short of daunting. But if the financial crisis taught us anything, it is that the prize could not be larger.”⁴

Fundamentally, the barcodes of finance are tools of information and communication technology. Technology would be the enabler of these regulations. The barcodes needed to be understood through the lens of reengineering of financial institutions’ risk and technology ecosystems. It also needed to be understood in the context of the reengineering of its supply chain of financial market participants and the reconstruction of existing business processes to leverage its technical features. These have yet to be fully understood by regulators and industry members alike. This may partially account for the dysfunction observed presently with the initial rollout of the IDs.

Another complicating factor is that these IDs needed to be retrofitted onto individual firm’s legacy systems to interface

2 Wolf, S., 2016, “Identifier verification: evaluation of blockchain and alternative technologies,” Global Legal Entity Identifier Foundation (GLEIF), <http://tiny.cc/g42kiy>

3 Group of Thirty, 2006, “Global clearing and settlement, final monitoring report,” <http://tiny.cc/u62kiy>, at page ix

4 Haldane, A., A. Schubert, and R. Berner, 2015, “Knowledge needed to prevent Lehman repeat – global financial language essential to close dangerous data gaps,” Financial Times, January 14, <http://tiny.cc/382kiy>

with the global financial industry's technical ecosystem. This ecosystem evolved over nearly six decades of incremental systems development across six generations of technology innovation. It now looks more like a Rube Goldberg or a Heath Robinson contraption than anything well thought through around good systems design.

With a long history of industry failure and the current dysfunction observed with this new regulatory driven attempt at a global identification scheme, industry groups such as ISDA (International Swaps and Derivatives Association) and the Japanese Exchange Group (JPX) are calling for a deeper look into the regulatory, operational, and technical ecosystem that is to be supported by this global identification scheme. JPX has stated recently in exploring distributed ledger technology (DLT) that "We would also like to highlight that the DLT could bring innovation by reconstructing existing business processes to leverage its technological features. If existing entities knowledgeable in the business processes will lead the discussion, they need to take care not to eliminate the technical advantages by focusing too much on existing work flow."⁵ ISDA, similarly, has stated, "If DLT does take a more prominent role in the derivatives ecosystem, or in capital markets more broadly, there may be a requirement for regulators to reconsider some of their existing regulation. Therefore, collaboration between regulators and industry participants will be an important element of any deployment in this field."⁶

The barcodes of finance were expected to be transformational. Universal codes fit for all reporting purposes in all markets. Rather than disappoint politicians, who had the need to tell their constituents that all was well after the financial crisis of 2007-2008, regulators began the first use of these codes in swaps reporting with an incomplete identification scheme and incomplete, almost non-existent, reference data. They also began with an incomplete knowledge of the swaps markets, the operational components of the interconnected global capital markets, and the technical ecosystem that they operate within.

BARCODES AND SWAPS DATA REPORTING

Standardized global identifiers, to the extent they now exist, are being placed in swaps transactions at the latest point in the trade's life cycle, that being just before submitting them to newly developed trade repositories. However, what was expected but failed to materialize before their use was the vetting of the standardized definitions of the codes and their supporting reference data. Missing was standardized reporting of

data elements comprising a swap transaction and a means to aggregate the reported data in and across trade repositories. To this point, Timothy Massad, CFTC Chairman, stated in his preamble to the CFTC's consultation on standardized swaps reporting, "In our original rules, we purposely didn't prescribe exactly how each field should be reported – for two principal reasons. First, when the agency issued the reporting rules, we didn't yet have any data to inform our views. And second, we needed the industry to take coordinated steps toward standardizing its reporting. That, unfortunately, has not happened."⁷

After nearly four years of reporting swaps data in the U.S. and two and a half years in the E.U., regulators have encouraged industry members to support a myriad of new infrastructure entities: to issue new ID codes – 29 for issuing LEIs (referred to as LOUs, local operating units that act as facilities operators that organize the local LEI registries and maintain business card data on each legal entity); to maintain swaps trades in new data repositories (currently 25 new facilities); and additional regional facilities to aggregate trade repository data. In this later regard, E.U. regulators have assigned the task to the European Securities and Markets Authority (ESMA) to build the facility to aggregate data from the six trade repositories that now exist in the E.U.⁸ The U.S. has five such trade repositories and the CFTC has discussed designating the National Futures Association (NFA) as the aggregator of this data.⁹ Other sovereign jurisdictions, about ten in total at present, will have to be aggregated with the others in a still undetermined way.¹⁰ Recently, a number of trade associations in a joint statement voiced concern on the need to conform identification standards and harmonize data requirements for this all to work.¹¹ Notwithstanding this call for action going forward, there is still no understanding of what to do with the tens of billions of non-standardized transactions already in trade repositories.

5 Japan Exchange Group, 2016, "Applicability of distributed ledger technology to capital market infrastructure," <http://bit.ly/2ch11jU>

6 ISDA, 2016, "The future of derivatives processing and market infrastructure," <http://bit.ly/2jJHGua>

7 CFTC, 2015, "Request for comment, draft technical specifications for certain swap data elements," Press Release <http://bit.ly/1JpZC7G>

8 ESMA, 2015, "ESMA launches centralized data projects for MIFIR and EMIR," Press Release, April 1, <http://bit.ly/2iPJW4B>

9 NFA, 2010, "Possible role for NFA as a utility for swap transactions," <http://bit.ly/2iHRjrd>

10 Euractiv.com, 2015, "Regulators warn over-the-counter derivatives are out of control," April 14 <http://www.euractiv.com/sections/euro-finance/regulators-warn-over-counter-derivatives-are-out-control-313723>

11 Joint Association letter on global trade reporting and data harmonization, 2015, "Key principles to improve global trade reporting and data harmonization," June 11, <http://bit.ly/1fcd94Z>

It is no wonder that the premature rollout of the codes and their continually evolving features have compromised regulatory reporting. Incrementally failed implementations are forcing regulators and the industry to make costly compromises that are negating both the vision and the usefulness of a global identification system for regulators and industry participants.

It is in stark contrast that such formidable pillars of business as Wal-Mart, Federal Express, Amazon, and other participants in global commerce have a common standard for identifying their supply chain participants and the products they manufacture, transport, and sell – it is found in the ubiquitous barcode. It can be thought of as the enabler of global commerce in the digital age.

Similarly, the domain names of the internet provide unique identification across the vast interconnected set of communication networks that support unique email addresses and websites. It too can be thought of as the enabler of global commerce in the digital age.

Financial regulators have now set their own agenda for creating such digital enablers for the financial services industry, the LEI, UTI, and UPI. Regulation in the E.U. is now poised to impose these codes on all banks and other financial market participants; initially having started with reporting and recordkeeping of swaps transactions, but soon for all regulatory reporting requirements under MiFid II's (Market in Financial Instruments Directive) regulations scheduled for implementation in 2018.¹² However, with the codes already proving dysfunctional in their use in reporting swaps to swaps data repositories (SDRs), caution in extending their use should prevail.

To this end, Timothy Massad, Chairman of the CFTC, stated in his testimony before the U.S. Senate Committee on Agriculture, Nutrition & Forestry, on December 10, 2014, that "The proliferation of data repositories across various jurisdictions makes moving forward in this area more important than ever. We are leading an international harmonization effort to achieve consistent technical standards and identifiers for data in SDRs ... Standardizing the collection and analysis of swap market data requires intensely collaborative and technical work by industry and the agency's staff. We have been actively meeting with the SDRs on these issues, getting input from other industry participants and looking at areas where we may clarify our own rules ... In short, the data collection issues will take time, but we are making progress. Going forward, it must continue to be one of our chief priorities."

Most significantly, the LEI code, a key pillar of swaps data

reporting and touted as an example of industry and regulatory consensus, must still be proven to work for both aggregating data in trade repositories and within a global LEI system (GLEIS) for other market participants and in other markets. Even though 481,522 codes have already been issued, it represents only 10% of the expected total. Furthermore, 29% of the LEIs issued have not been renewed. Annual renewal of LEIs is a requirement in order to maintain a valid LEI, although no regulator has yet to mandate renewing LEIs.¹³

Most critically, the GLEIS has yet to be completed. Regulations require that each LEI be identified as to its ownership and place in the control structure of its ultimate parent. That the framework for this endeavor remains only partially defined while LEIs are being issued may prove a fatal flaw. Most importantly, the opportunity to compel both the registration of the ultimate and immediate parent of the registered counterparty (referred to collectively as relationship reference data) simultaneously at the initiation of the LEI registration process for swaps market participants has, in the cases of already issued LEIs, been lost.

Recently the FSB's Regulatory Oversight Committee (ROC), the overseer of the LEI initiative, published a public consultation on relationship data in which it stated, "Each potential parent entity, in conjunction with the work of its external auditors, determines whether it controls another entity and should consolidate it under applicable accounting standards."¹⁴ Matthew Reed, Chairman of the ROC, in an interview, in anticipating the public consultation, commented, "We expect that we will view the LEI file as complete only when certain corporate information is revealed with respect to hierarchy information."¹⁵

In March 2016, the ROC published its response to the comments received in its public consultation on relationship data.¹⁶ Its overreaching recommendation requires entities that either possess or acquire an LEI to report their "ultimate accounting consolidating parent," defined as the highest level legal entity

12 ESMA, 2016, "Technical standards under Directive 2004/39/EC (MiFID I), Directive 2014/65/EU (MiFID II) and Regulation (EU) No 600/2014 (MiFIR)," <http://bit.ly/2iPMJjUQ>

13 Financial InterGroup, 2017, "The global LEI initiative – a 2016 year-end review of progress and issues," Research Note, <http://bit.ly/2iAtQgy>

14 LEI ROC, 2015, "Consultation document on collecting data on direct and ultimate parents of legal entities in the global LEI system," September 7, <http://bit.ly/1KIDMg5>

15 Roland, N., 2015, "LEIs seen as key to understanding firms' overall risk exposure," FIA SmartBrief, <http://bit.ly/2jMZ4Op>

16 LEI ROC, 2016, "Collecting data on direct and ultimate parents of legal entities in the global LEI system – phase 1," March 10, <http://bit.ly/2iBgCAn>

preparing consolidated financial statements, as well as their “direct accounting consolidating parent” to the LOU maintaining their LEI. In both cases, the identification of the parent is to be based on the accounting definition of consolidation.

The ROC is referring to the next steps process as a six-month pilot, after which they will evaluate outcomes. This process of soliciting ultimate and direct parents will begin in 2017. The ROC has recognized that this falls short of complete hierarchies as requested by the FSB, which they expect to be addressed in further consultations. To this end, the FSB in their 2016 annual report addressed the need to expand the LEI by stating that “This unique identifier is used in two-thirds of FSB jurisdictions to support regulatory activities, for instance in connection with financial reporting. Additional uses are contemplated, such as in the area of correspondent banking. Further adoption of the LEI by legal entities worldwide and its use by authorities for a wider range of regulatory purposes is essential to fully reap its collective benefits.”¹⁷

The UTI, already in use but without consistent definition, is also contained in the same tens of billions of transactions reported to swaps trade repositories as is the LEI. It too is in need of a global standard before it can be used to match buys and sells or pays and collects of the same swaps transaction. Already underway is a public consultation sponsored by IOSCO and the Committee on Payments and Market Infrastructure’s (CPMI’s) Board of IOSCO that is attempting to do just that, referring to existing dysfunctional UTIs as “legacy UTIs.”¹⁸

Recommended in the IOSCO/CPMI proposal is the use of the LEI as a construction element (a prefix) for the UTI. Industry commentators, however, have noted issues with such use, such as the length of the LEI code, which is too long to fit into data fields of legacy systems and the timeliness of acquiring a LEI.

The UPI, like the LEI and UTI, is also incomplete and being reported without consistent ways of constructing it. It too needs to adhere to a global standard in order for it to be fit for all its intended uses. In its UPI consultation paper, IOSCO/CPMI proposes standard UPI reference data elements.¹⁹ The UPI code construction itself is to be the subject of further IOSCO/CPMI consultations in 2017.

To this end, ANNA (Association of National Numbering Agencies) in consultation with ISDA have advanced the idea of a UPI based on the use of ANNA’s ISIN (International Securities Identification Number) infrastructure, generating UPI codes for each contract from templates submitted by requestors. The

code will have twelve alphanumeric characters, like existing ISINs, but, unlike them, will be generated in near real-time. ANNA will establish the ANNA Derivatives Service Bureau, proposed as a centralized facility to generate ISIN’s for derivatives contracts. The requestor would be validated, presumably by inputting a valid LEI, but further input data elements would not be validated.²⁰

The FSB in referring to their expectations for the global ID system of the LEI stated that “Such a system would provide a valuable ‘building block’ to contribute to and facilitate many financial stability objectives, including: improved risk management in firms; better assessment of micro and macro prudential risks; facilitation of orderly resolution; containing market abuse and curbing financial fraud; and enabling higher quality and accuracy of financial data overall. It would reduce operational risks within firms by mitigating the need for tailored systems to reconcile the identification of entities and to support aggregation of risk positions and financial data, which impose substantial deadweight costs across the economy. It would also facilitate straight through processing (STP).”²¹

With regard to the ultimate use of these identifiers in data aggregation of financial transactions and in STP, the BCBS states, “Many banks lacked the ability to aggregate risk exposures and identify concentrations quickly and accurately at the bank group level, across business lines and between legal entities. Some banks were unable to manage their risks properly because of weak risk data aggregation capabilities and risk reporting practices. This had severe consequences to the banks themselves and to the stability of the financial system as a whole.”²²

The necessary next steps to complete the barcodes of finance may be to take a step back and start with a global committee looking at the entire range of needs, not just by mixed

17 Financial Stability Board, 2016, “Implementation and effects of the G20 Financial Regulatory Reforms 2nd Annual Report,” August 31, <http://bit.ly/2cgh6aM>, at page 16

18 Committee on Payments and Market Infrastructures Board of the International Organization of Securities Commissions, 2015, “Harmonization of the unique transaction identifier,” consultative report, <http://bit.ly/2jZPV11>

19 BIS, 2015, “Harmonization of the unique product identifier,” Bank for International Settlements, December, <http://bit.ly/2iJ535b>

20 FinOps, 2016, “OTC derivatives players: the ISIN engine emerges,” <http://bit.ly/2chWjVB>

21 Financial Stability Board, 2012, “A global legal entity identifier for financial markets,” June 8, <http://bit.ly/1ESyEkC>, at page 1

22 BCBS, 2013 “Principles for effective risk data aggregation and risk reporting – final document,” Basel Committee on Banking Supervision, January 9, <http://bit.ly/2iQQkJO>

committees formed in multiple silos of sovereign regulation, markets, or businesses, as is the situation at present with the singular focus being on the swaps markets. An example is ISDA recently calling for expanding the LEI, UPI, and UTI and creating a central data dictionary of terminology and reference data, but just for derivatives trades.²³

The G30 study, concluded in 2006 and referred to earlier, had recognized that “implementing reference data standards is a complex and long-term project. There is no dispute that the diversity of coding systems and the difficulty of translating between them is a core reason for the currently high processing costs of global investment. More significant effort is needed to develop global standards in this area. There is not yet a clear global owner of reference data and there is friction between the needs of the domestic and cross-border market user. Senior level support for standardization will be vital.”²⁴

Perhaps it is time for the G30, a neutral private, nonprofit, international body composed of very senior representatives of the private and public sectors and academia to reassert itself into fixing the infrastructure of the global financial system at this vital juncture by examining the choices available to market practitioners and policymakers.²⁵

CURRENT STATE OF THE LEGAL ENTITY (LEI) IDENTIFICATION STANDARD

The method of universally identifying financial supply chain participants, after earlier industry attempts going back over two decades, remained unresolved when the financial crisis erupted in 2007-2008. This failure left data vendors and financial market utilities in charge of supplying their own proprietary and quite different codes and reference data for the same legal entity or product. Financial institutions, too, created their own codes and tried to cleanse multiple versions of reference data supplied from data vendors. The bridge between the external and internal codes was left to computer driven tables used to map each of the many codes representing the same company or product. These mapping tables are populated by comparing the alphabetic names associated with each proprietary code to determine a match, not a very exact science considering different spellings appear frequently for what is the same product or entity requiring manual intervention and research.

These mapping tables have to be continually maintained as companies and products are ever changing. Some vendors and financial market utilities do not change their information

in the same timeframes as others; some miss these changes altogether and some interpret the changes incorrectly. Each financial institution similarly makes its own determination as to changes to be made and their timing. The result has been an error prone and costly process to establish which part of each company is each financial institution’s counterparty and the product or contract being traded. This point was made in a recent Dun & Bradstreet report quoting a senior data manager at a Tier-1 global bank: “At the moment, the LEI is a mapping exercise. Only a limited number of LEIs have been issued and my database holds over one million identifiers. The LEI can provide value in niche corporate markets, but for us it is a big mapping exercise and getting bigger.”²⁶

In the aftermath of the financial crisis, regulators proposed solutions and sought consensus from the industry, focusing initially on legal entity identification. Two different approaches emerged to register codes and their legal entity data, either by intermediaries or via companies registering their own data directly. Self-registration was approved by the FSB to register a unique code accompanied initially by business card data (principally name and address data). Later more significant data, such as relationship information amongst legal entities and other operational reference data would be registered.

Assignment of preliminary codes (LEIs) through “self-registration” would later be modified to allow them to be assigned by data intermediaries (LOUs) as “third-party registrants.” This third-party registration has become the de-facto method of registering information in the interim global LEI system. This approach was condoned so that regulators could accommodate the early issuance of codes by the CFTC under the CFTC’s own mandates, using both its own proprietary legacy codes and LEI issued codes (the CICI – CFTC Interim Counterparty Identifier) of its first designated LOU, the Global Market Entity Identifier (GMEI) utility organized by DTCC and SWIFT. These CICI codes were issued before the FSB was tasked with the global LEI implementation. This early mover approach was cautioned against in the earlier G30 study, which stated, “First mover implementation of global standards should not be mistaken for the first mover setting the standard.”²⁷ Subsequently,

23 ISDA, 2015, “Improving regulatory transparency of global derivatives markets: key principles,” International Swaps and Derivatives Association, February, <http://bit.ly/2k4e3Hs>

24 See footnote 3

25 Group of Thirty, <http://group30.org/members>

26 Dun & Bradstreet, 2016, “Deriving true business value from the global LEI,” October, <http://bit.ly/2jrkwZL>

27 See footnote 3

the standard had to be reset by the FSB, the consequences of which resulted in thousands of codes being rescinded by the CFTC; different formatted codes populating multiple LOUs' LEI registries; and high lapse rates for annual code renewals, currently 29% of total issued LEIs, with the GMEI accounting for 37% of all lapsed LEIs.²⁸

At this time, the LEIs being assigned in the interim GLEIS (global legal entity identifier system) have no association with their registering parents. Rather, they are associated with a prefix embedded at the front of the code that recognizes the intermediary facility operators (pre-local operating units or "pre-LOUs") that initially assigned the codes. The FSB, and subsequently the ROC, formed of regulatory authority representatives to oversee the implementation of the GLEIS, assigned these unique prefixes to each pre-LOU as the first four digits of the LEI codes. This was done so that each pre-LOU could then assign codes to each legal entity at their choosing in order to make all the codes globally unique. In contrast, the GS1 system, the commercial barcode assignment system, assigns a globally unique company prefix to each company to create global uniqueness. The company itself then uses this prefix to self-register separate codes for locations (branches), legal entities,²⁹ and products by affixing a suffix (at the companies' choosing) to the company prefix. For smaller entities, GS1 assigns both the company prefix and the suffix.

According to Grody and Smucker (2011), "There are a few approaches to assigning a LEI. One suggested in the legislation itself and supported by industry trade associations is self-registration whereby corporations and others who participate in financial markets assign their own company identifiers under assignment rules administered by a global standards registration authority. This procedure resembles the process used in industries such as food, healthcare and consumer packaged goods, whereby manufacturers self-register their products and locations under a standard system administered by a global standards body, GS1."³⁰

Another global identification system, which is used for the Internet for email and World-Wide-Web (WWW) addresses, uses the same company-determined self-registration method as GS1 to establish unique codes (domain names), register business card data, and company email addresses.

Similar to GS1's assignment of the unique company prefix, ICANN (Internet Corporation for Assigned Names and Numbers) assigns domain names to communities of large users, such as universities and companies that, in turn register multiple users, such as their students or employees' email names

by extending the main domain name to include a unique prefix for each student or employee. Telephone companies worldwide follow similar protocols, assigning a main telephone number to a company and then allowing the company to assign extensions for each employee or department.

As stated by Haldane in his 2012 SIFMA symposium presentation on the LEI in comparing GS1 and the internet to the financial services industry "Yet despite these similarities, finance lags by a generation both products and information in the management of its network. Today's financial chains mimic product supply chains of the 1980s and the information chains of the 1990s. For global supply chains and the Internet, a common language transformed their fortunes. This enabled them to become global in scale and scope and highly adaptive to new demands and technologies. They are astonishing success stories."³¹

Noting the similarities of both GS1's and the internet's assignment protocols, both allowing for a code for each registering parent and leaving the choice of specific sub-codes to each large enterprise, the FSB, and later the ROC, chose instead to give the power of the entire code creation and assignment to intermediary facilities' operators. They, in turn, are permitted to embed their own assigned code (the LOU prefix, positions 1-4) in all the LEIs they issue. This left a large contingent of government and commercial enterprises – patent offices, depositories, business registrars, payment system operators, central bankers, custodian banks, software companies, and data vendors – to intermediate the process of code assignment and issuance as data vendors had done in the earlier era of proprietary codes. An unintended consequence was that it prevented the code from being created in real-time at the source of the entity's creation nor could it be used directly in aggregating financial transaction data. The LEI, if each registering parent entity of a financial market participant was to obtain its own prefix and embed it as part of the code, would allow for direct aggregation of financial transactions to its parent entity.

28 Financial Inter Group, 2016, "The Global LEI Initiative – moving in the right direction but still another month of lapsed LEIs exceeding new LEIs", <http://bit.ly/2kFgqgK>

29 GS1, global location number (GLN), <http://bit.ly/2IEDNKd>

30 Grody, A. D., and T. Smucker, 2011, "Legacy main street solution proposed for Wall Street," CEO Magazine, May 17, <http://bit.ly/2jOMLkP>

31 Ali, R. D., A. G. Haldane, and P. Nahai-Williamson, 2012, "Towards a common financial language," Presented at the Securities Industry and Financial Markets Association (SIFMA) "Building a global legal entity identifier framework" Symposium, New York, March 14, <http://bit.ly/2kFwwXO>

The “Report on OTC derivatives data reporting and aggregation requirements” by the Committee on Payment and Settlement Systems (CPSS) and the Technical Committee of IOSCO, supported such a prefix, stating, “The federated approach could also be extended to parent companies within a national jurisdiction. If this approach were followed, the national authority would issue a range of LEI codes to a parent company, which would in turn issue individual LEIs to legal entities within the parent company. The characters in the LEI code immediately following the first characters (the LOU prefix) that identify the national authority would identify the issuing parent company (without embedded intelligence).”³²

Parent entities self-registering each of their financial market participants is still a possibility so that the LEIs could be tied directly to the outside LEI coding scheme, thus eliminating mapping issues and allowing for internal data aggregation using their assigned codes. Parent entities of mutual fund families could self-register their individual funds under a single fund-family prefix rather than as it is done now, using a reference data element in the GLEIS, referred to as an Associated Entity, to populate the LEI of the parent entity. The ROC is planning to study the current way relationships of funds are recorded in the GLEIS.³³ The way this can be accomplished without violating the “no-intelligence in the code” construction rule is described more fully in an earlier paper by the author³⁴ and summarized later in the conclusion section of this paper.

Regulators and industry members have yet to decide on an integrated system of standards for both global identification of financial market participants and the products they trade. Connecting counterparties and their aggregated position of transactions is critical to evaluating risk. Still to be implemented just for the interim GLEIS are: the access method for more timely issuance of LEIs and more timely access of the entire database of LEIs; support of the entire hierarchies of LEIs for the parent company, many of which are now registered as LEIs in separate sovereign jurisdictions; timely updating of the information for corporate events that affect the ownership of the legal entities globally; identifying the ownership and organizational hierarchies of the legal entities; masking legal entities as required by those domiciled in privacy jurisdictions; and assuring that the code and the system are fit for all legal entity identification purposes, beyond its initial pilot test for counterparty identification for swaps data reporting.

Such broader use was intended when the FSB defined which legal entities must acquire a LEI “...the term ‘legal entity’ refers to a legal person or structure organized under the laws of any jurisdiction. Legal entities include, but are not limited to,

unique parties that are legally responsible for the performance of financial transactions or have the legal right in their jurisdiction to enter independently into legal contracts, regardless of whether they are incorporated or constituted in some other way (e.g., trust, partnership, contractual, etc.). It excludes natural persons, but includes governmental organizations; and supranationals, defined as governmental or non-governmental entities established by international law or treaty or incorporated at an international level. Examples of eligible legal entities include, without limitation: all financial intermediaries; banks and finance companies; all entities that issue equity, debt, or other securities for other capital structures; all entities listed on an exchange; all entities that trade stock or debt; investment vehicles, including mutual funds, pension funds, and alternative investment vehicles constituted as corporate entities or collective investment agreements (including umbrella funds as well as funds under an umbrella structure, hedge funds, private equities, etc.); all entities under the purview of a financial regulator and their affiliates, subsidiaries, and holding companies; and counterparties to financial transactions.”³⁵

Other remaining challenges for the LEI include using it in the development of the UTI for each financial transaction so that buyers and sellers, and payers and receivers of interest on financial products can be matched. Here the LEI, as earlier described, is being considered as a prefix for the UTI.

PROGRESSING THE BARCODES OF FINANCE (LEI, UTI, AND UPI) INITIATIVES

The first use of the financial supply chain’s new global coding schemes are being tested by swaps market participants across the globe in their new responsibility to record-keep and report swaps transactions to trade repositories. While all would agree that considerable global cooperation has been achieved and progress has been made, it is apparent that the identification codes are not yet up to industrial strength nor is their use in risk data aggregation yet functioning.

32 CPSS and IOSCO, 2012, “Report on OTC derivatives data reporting and aggregation requirements – final report, annex 3,” January, <http://bit.ly/2iMKu7Y>, at reference 111 page 65.

33 ROC, 2017, “Update on the LEI ROC activities,” January 12, <http://bit.ly/2joJLOZ>

34 Grody, A. D., P. J. Hughes, and D. Reininger, 2015, “Final report on global identification standards of counterparties and other financial market participants,” *Journal of Risk Management in Financial Institutions – Special Issue on Counterparty Risk* 5:2, March 10, 2015, <http://ssrn.com/abstract=2016874>

35 See footnote 20

David Wright, Secretary General of IOSCO, believes that “there is a general data issue...I think we don’t have a sufficient understanding of market-based financial [data].”³⁶ Recently, ESMA published updated reporting guidelines on the LEI and the UTI and on the technical standards for reporting under EMIR using the LEI and UTI.³⁷ EIOPA published reporting guidelines for the LEI.³⁸ ISDA published updated reporting guidelines for the UTI.³⁹ IOSCO/CPMI has published a recent public consultation on the UTI⁴⁰ and the UPI.⁴¹

These five recent releases are an attempt to bring clarity to the identification regime first proposed by the SEC, CFTC, and Office of Financial Research (OFR) in 2010 and subsequently transferred to the Financial Stability Board (FSB) in 2011. Since then, the FSB has organized the ROC, a group of 70 regulators from 40 countries, and the GLEIF, its board made up of 16 industry practitioners and academics to implement one part of the global identification scheme, the GLEIS. Other regulators and trade associations have opined previously on the identification regime, including the EBA,⁴² Singapore Monetary Authority,⁴³ the FSB,⁴⁴ and the CFTC.⁴⁵

ISDA’s CEO, Scott O’Malia, an earlier Commissioner of the CFTC whose remit was to oversee the U.S.’s first versions of the counterparty (LEI), transaction (UTI) and product (UPI) identifiers, commented recently on needed improvements. “Plans for a global snapshot of risks in the financial derivatives market are a ‘dream’ that must not detract regulators from tackling discrepancies in trade reporting ... Where regulators need to focus right now is working together to harmonize the convention of reporting, making sure we are doing an apples-with-apples comparison with the data”⁴⁶

The LEI is particularly important for the implementation of the global swaps risk regimes as it is intended to uniquely describe the counterparties and, potentially, reference entities in swaps transactions.

The LEI is also to be used in certain jurisdictions to construct the UTI. The IOSCO/CPMI proposal suggests it be used universally to do so. Finally, an LEI is to be assigned to identify all financial supply chain entities involved in the life cycle of swaps transactions and, eventually, in the life cycle of all financial transactions. Most importantly, the LEIs are to be chained together to aggregate data up to the controlling or parent entity and to aggregate data across multiple trade repositories. This is the first attempt to develop by regulatory fiat, albeit in cooperation with industry members, a global identification scheme and a new global market infrastructure. Timothy Massad, the Chairman of the CFTC, called it “a huge information technology challenge.”⁴⁷

The CFTC, after initiating trade reporting rules, subsequently recognized their inability to aggregate swaps data being reported to multiple trade repositories in the U.S. and requested a review of their swaps data reporting and recording keeping rules. These rules were dependent on the global identification system of the LEI, UPI, and USI (unique swaps identifier, later changed to the UTI) to provide for data aggregation. Many of the questions posed in the review were related to improvements to these identifiers as well as the data tagging language used to describe other reference and value data elements for inclusion in swaps transaction reporting.⁴⁸ The CFTC has only partially responded to the many comment letters received, responding primarily to cross border issues of reporting obligations of cleared swaps but not to the remaining issues of data standards still to be resolved.⁴⁹

The CFTC also earlier enlisted the Office of Financial Research (OFR) to assist in the data issues that had surfaced due, in part, to the early embrace of the identification schemes by the CFTC that proved to be premature as it had yet to be vetted by the FSB.

Richard Berner, Director of the OFR, stated in a speech at a CFTC Technology Advisory Committee meeting, that “We live in a world of global markets and global institutions and there’s no escaping the fact that, if we don’t standardize data and

36 Wright, D., 2014, “Iosco warns on data challenge to risk oversight,” Financial Times, October 29, <http://on.ft.com/2iEDVJn>

37 ESMA, 2014, “Review of the technical standards on reporting under Article 9 of EMIR,” consultation paper, November 10, <http://bit.ly/2kFoRZr>

38 EIOPA, 2014, “Final report on public consultation no. 14_037 on the proposal for guidelines on the use of the legal entity identifier (LEI),” September 11, <http://bit.ly/2jbl28G>

39 ISDA, Unique Trade Identifier (UTI): generation, communication and matching (updated as of 20 July, 2015), <http://bit.ly/2jbTeFQ>

40 Committee on Payments and Market Infrastructures Board of the International Organization of Securities Commissions, 2015, “Harmonization of the unique transaction identifier,” consultative report, <http://bit.ly/2jZPV11>

41 BIS, 2015, “Harmonization of the unique product identifier,” Bank for International Settlements, December, <http://bit.ly/2iJ535b>

42 EBA, 2014, “Recommendation on the use of legal entity identifier (LEI), EBA/REC/2014/01,” European Banking Authority, January 29, <http://bit.ly/2iEDRJO>

43 Singapore Monetary Authority, 2013, “Swaps regulations,” <http://bit.ly/1q3v5Pt>

44 See footnote 21

45 CFTC, 2012, “Swaps data and recordkeeping requirements, final rule,” Federal Register, Vol. 77, no. 9, January 13), <http://bit.ly/2iMBfQ>

46 O’Malia, S., 2014, “Derivatives industry calls global reporting plan a dream,” Reuters, September 23, <http://reut.rs/2jAXeDg>

47 Massad, T., 2014, “Bank scandal sparks CFTC outrage, aggression,” CNBC, November 12, <https://yhoo.it/2iVJDVU>

48 CFTC review of swap data recordkeeping and reporting requirements, March 26, 2014, <http://bit.ly/1hggugF>

49 CFTC, 2015, “Notice of proposed rulemaking on amendments to swap data recordkeeping and reporting requirements for cleared swaps,” Statement of Commissioner Christopher J Giancarlo, August 19, <http://bit.ly/2kLeAdy>

harmonize them across those borders, then we won't be able to use them ... The ability to compare and aggregate those data across the SDRs and across borders is absolutely critical to monitor those threats ... The implementation reflects the need to use standards for entity identification (LEIs) ... Obviously those are important. Equally important will be the use of instrument and product identifiers ... and the use of hierarchies to organize those data in a coherent framework so that we can compare and aggregate similar, but not exactly alike, either entities in a particular industry segment or instruments in a particular asset class, and both with respect to entities and instruments."⁵⁰

The FSB, in an attempt to finalize the issues still to be resolved around OTC derivatives, issued their "Recommendations on aggregation of data for OTC derivatives."⁵¹ In their recommendations, they challenge themselves and the industry to define an aggregation method for data accumulating in trade repositories; complete the global identification scheme to include ownership and control issues; come to some global consensus on UTI and UPI construction; and resolve the issues of counterparty identification in privacy jurisdictions.

In its "Feasibility study on approaches to aggregate OTC derivatives data," the FSB stated that "The report notes that, amongst these steps, it is critical for any aggregation option that the work on standardization and harmonization of important data elements be completed, including through the global introduction of the Legal Entity Identifier (LEI), and the creation of a Unique Transaction Identifier (UTI) and Unique Product Identifier (UPI). The report also indicates, in broad terms, the types of legal and regulatory changes that would be needed to allow a central mechanism to access the necessary data from trade repositories and to aggregate the data for authorities. While further work in both of these areas will be challenging, progress will be essential for a global aggregation mechanism to be effective."⁵²

The OFR in their 2014 annual report to Congress reported on the global standards initiative, the LEI in particular, and the success it had in building global consensus amongst regulators. While exemplary in its consensus building, it is suffering from the dysfunctional implementation in its first use test with swaps data reporting to SDRs.⁵³

In describing the role the OFR was playing in the U.S. with the CFTC in resolving these issues, the report describes the need to assess and improve the quality of data collected. The report stated, "Members of the CFTC Technology Advisory Committee concluded at a meeting on February 10, 2014, that

missing, incomplete, and inaccurate data made SDR data unfit to use in regulatory oversight. The committee said the CFTC's definitions for SDR reporting were not sufficiently precise and that standards must be applied when data are collected instead of trying to harmonize data later in the process. The OFR and the CFTC are collaborating to address these data quality issues with the data already collected."⁵⁴ In a subsequent 2016 report to Congress, the OFR stated, "The problem the LEI addresses – the precise identification of counterparties – remains unresolved."⁵⁵

The FSB in reporting on its overall standard's progress to its G20 members recognized the considerable effort still left in the first use test of the identification scheme in reporting swaps transactions to trade repositories. In his letter to the G20 finance ministers and central bank governors, Mark Carney, Chairman of the FSB, stated, "With the support of the G20, FSB members have made major investments to reduce the opacity of OTC derivative markets. We now need to make trade reporting truly effective. There currently are significant legal and other blockages to the reporting, sharing and aggregation of key information regarding trades and these must be removed."⁵⁶

Of significance, what remains to be accomplished at the conceptual design level are some prominent features of a global identification scheme. The first is that of the relationship of the manufacturer or issuer of a product or contract and that relationship as a counterparty, especially when the two roles are performed by the same financial market participant. Another is the collective relationship of LEIs that are under common ownership or control.

Matthew Reed, Chairman Emeritus of the ROC's Legal Entity Identifier initiative asks: "Who is who?" "Who owns whom?" "Who owns what?"⁵⁷ These two "left to last" attributes (who owns whom and who owns what) of the global identification

50 CFTC, 2013, "Technology Advisory Committee, transcript, remarks," September 12, <http://bit.ly/2jHDBcJ>

51 FSB, 2014, "Feasibility study on aggregation of OTC derivatives trade repository data," September 19, <http://bit.ly/2iJKM4t>

52 FSB, 2014, "Feasibility study on approaches to aggregate OTC derivatives data," September 19, <http://bit.ly/2j2qjGL>

53 OFR, 2014, "Annual report 2014," Office of Financial Research, November 30, <http://bit.ly/2k8qu1U>

54 See footnote 53

55 Office of Financial Research, 2016 Financial Stability Report, <http://bit.ly/2jV39jQ>

56 FSB, 2015, "Financial reforms – finishing the post-crisis agenda and moving forward," Financial Stability Board, February 4, <http://bit.ly/1Mc95xU>

57 Reed, M., 2015, "Remarks at the data transparency coalition," March 11, <http://bit.ly/1FNgWND>

scheme, along with the ability to identify and maintain changes to global identifiers, are the most critical components of the system. Solutions achieved by defining such characteristics in the code construction itself will be discussed in the Conclusion section of this paper.

To the point of the hierarchical relationships between counterparties, a recent paper by the then two leaders of the BIS secretariat supporting the FSB's LEI initiative specifically focused on such relationship data. Leonova and Jenkinson state, "In the financial context, the relationships may be determined by accounting rules set, for example, by IFRS or US GAAP, as well as regulatory requirements in areas of risk management, market integrity, know-your-client, network analysis and statistical consolidation. The financial industry and regulators have spent countless hours arguing and debating the definition of ownership. The problem lies in the question itself. We suggest that as part of any relationship data system the best approach is to put the question aside and avoid a conceptual and practical quagmire. Rather, we recommend collecting and storing less-subjective granular data on the actual legal and economic relationships between firms, which provides a flexible framework from which any user can answer the question on corporate relationships he or she determines is appropriate at a given time. Encouragingly, technological solutions are available to accommodate this multiplicity of requirements in a single solution..."⁵⁸

Unfortunately, in such a mix-and-match solution as advocated above, one which the industry has already been operating with, with all its inherent mapping issues no matter the technology deployed, there is no way to achieve a consistent view of systemic risk. A coherent, consistent view should be available to transcend each company's, or sovereign regulators', or data vendors' own organizational construction to determine the risks assumed by each parent of a legal entity and, in turn, throughout the financial system. However, given the advocacy and acceptance by the FSB that account consolidation rules should prevail in establishing a first set of standard control/ownership relationships, it follows that GAAP and IFRS consolidation rules should be adhered to, certainly for an initial common global benchmarking view of counterparty control for regulators.

The interpretation of GAAP and IFRS rules, where necessary, can be made by accountants as the accepted method of establishing the relationship standard for LEIs. Accountants and auditors are the most accepted and trusted interpreters of such issues. Their existing activities in viewing legal entity authorization documents for their "materiality attestation" role

can be extended to registering relationship information into the GLEIS under authority of a legal entity's self-registration requirement for the LEI. Materiality is determined by legal entity, requiring each legal entity to be identified in order to view the materiality of the overall parent's financial situation. Materiality, as defined by the International Accounting Standards Board (IASB), is "an entity-specific aspect of relevance based on the nature or magnitude or both of the items to which the information relates in the context of an individual entity's financial report."

According to Grody and Hughes, "The authors speculate on an expanded role for auditors to include support for functions of the FSB's newly proposed agent, the Trusted Third Party (TTP), and to validate legal identities and their ownership structures within the FSB's new global legal entity identification (LEI) system. That system requires the control and ownership structures to correspond to accounting principles and standards as applied in the preparation of consolidated financial statements."⁵⁹

DATA AGGREGATION – THE KEY TO RISK ANALYSIS

Aggregation of transaction data using the global coding scheme can be seen as a unique requirement for global financial industry codes and their associated reference data. For example, cash flow and position data, such as total notional value of swaps or holdings of a particular security, are accumulated from multiple transactions in each product in the hierarchical chain of control of each counterparty. These positions are summed together and the aggregate value used to describe the risk to the controlling parent entity that may potentially be putting the entire financial system at risk. This can be the result of a single firm accumulating exposures beyond its limits of capital, risk concentration, or liquidity that goes undetected by regulators. To this later point, no single regulator or financial institution at this time has the capability to see such systemic risk building up across multiple disbursed components of a counterparty's legal entities nor the myriad of financial products these entities own that create exposure to such risk. This capability was desired by regulators and was the impetus for mandating the creation of the barcodes of finance.

58 Leonova, I. S., and N. Jenkinson, 2014, "Relationship data: the missing link of the current financial infrastructure," October, <http://bit.ly/2k8JbBz>

59 Grody, A. D., and P. J. Hughes, 2015, "The global risk regime – new roles for auditors," December 21, <http://bit.ly/2jHCR7t>

To this end, the identification scheme ideally will provide the means to look through a single financial transaction to its ultimate owner or beneficiary. In the case of the UPI, its controlling party is the contract or instrument's issuer, obligor, or guarantor. In the case of the LEI, it is the entity owning or controlling the counterparty based upon accounting control standards. Eventually, LEIs will be assigned to identify all issuers, obligors, guarantors, counterparties, and beneficial owners. Finally, the UTI will allow for an audit trail throughout a transaction's life cycle. It will also allow for the component parts of a two-sided (buy-sell, pay-collect) transaction to be identified so that the same counterparty transacting in the same product can be separately identified. Examples include swaps reporting to trade repositories and trades reported to clearing entities.

If the codes themselves do not lead directly to such beneficial interest, ancillary databases will have to be accessed whereby some will be unavailable as they reside in privacy jurisdictions and others will have to wait until batch processes are run for next day availability, perhaps longer. Surely, global mapping issues will delay the aggregation of multi-market contracts and instruments, and multi-counterparties belonging to the same business entity. Real-time data aggregation, or even real-time risk exposure flagging, important in a real-time financial transaction system, may never be possible.

The BCBS, recognizing the data aggregation issue, has published a risk data aggregation and risk reporting paper, known as BCBS239, and mandated its implementation beginning January 2016.⁶⁰ A number of objectives are anticipated, including a comprehensive assessment of risk exposures at the global consolidated level. To this point, BCBS anticipates merging, in the case of the largest financial institutions, the risk exposures of hundreds, even thousands of legal entities that comprise a consolidated financial entity.

To accomplish this, BCBS is requiring controls surrounding risk data to be as robust as those applicable to accounting data (a risk control equivalent to the U.S.'s Sarbanes Oxley's financial control and auditing requirement). Further, that risk data be reconcilable to accounting data to ensure risk data accuracy, and that a financial institution should strive towards a single authoritative source for risk data.

To assist in this data aggregation effort, the U.S. and other sovereign regulators are assuming that global data standards will find their way into risk systems across the business silos of individual financial institutions. In turn, these data and identification standards will be carried through to aggregate

risk data across multiple financial institutions to allow the FSB and other regulators to observe the contagion of systemic risk. Regulators initiated reviewing the way financial institutions are implementing BCBS239 beginning in January 2016 after initiating surveys of preparedness, which showed little progress.⁶¹

Toward this goal, the very first initiative announced by the FSB and referenced in BCBS239 was the creation of the standardized global identification system for legal entities, the LEI initiative. The FSB suggests, "The financial crisis has provided a renewed spur to the development of a global LEI system. International regulators have recognized the importance of the LEI as a key component of necessary improvements in financial data systems. To provide additional impetus, the FSB was given a mandate by the G-20 to lead the co-ordination of international regulatory work and to deliver concrete recommendations on the LEI system ..."⁶²

The LEI was to be a unique, unambiguous, and universal standard identifying every financial market participant throughout the global financial supply chain. It was intended to enable regulators to aggregate and analyze risk data across an enterprise and to facilitate observing emerging systemic risks across the financial system.

In Europe, the LEI has been provisionally mandated for supervisory reporting purposes for every credit and financial institution in the E.U. The intent is to have unique, unambiguous, and universal codes embedded in all reports to regulators and in all financial transactions. While adhering to a "no LEI no trade" rule, the rule does not require the validation of the LEI at each use "... there is no requirement to ensure that an LEI for a client or a counterparty has been renewed."⁶³ With 29% of the issued LEIs not renewed, this may be considered a showstopper until some remedy is found.

To emphasize this point Rodrigo Buenaventura, Head of the Markets Division of European Securities and Markets Authority (ESMA), stated earlier that "... generating common identifiers is not only a legal obligation, it's also essential for the quality of the data. No matter what method counterparties choose, they

60 See footnote 22

61 Grody, A. D., and P. J. Hughes, 2016, GARP – BCBS 239: Heightened standards and operational risk, <http://bit.ly/2jHDgXx>

62 See Footnote 21

63 ESMA, 2015, Consultation on transaction reporting, reference data, recordkeeping and clock synchronization under MIFIR, December 23, <http://bit.ly/1sOnkDo>, at page 20

need to agree on a single identifier that is common for that transaction ... That is the main challenge, the main element that we are now working on. We are conducting a number of initiatives.”⁶⁴

A recent joint consultative paper on the reporting of risk in intergroup transactions of financial conglomerates was proposed by ESMA, EIOPA (European Insurance and Occupational Pension Authority), and the EBA (collectively, the joint committee of the European Supervisory Authorities). It requires breakdowns of risk by counterparties and groups of interconnected counterparties using the LEI.⁶⁵ The objective for the deployment of this global ID regime by regulators is to exercise their obligations to the public to provide assurances that aggregate risk data observed in a single firm’s enterprise risk management system and in calculating multiple firms’ systemic risk exposures are accurate and reliable. For banks and other financial intermediaries, it is also about cost reduction, operational risk mitigation, and the long sought efficiency of STP.

FIT FOR PURPOSE?

Regulators are noticeably confused about what exactly has been accomplished and whether the coding scheme rushed into use initially by a single agency, the CFTC, will be fit for all its intended purposes. Today, what is operational is only the local issuance and maintenance of standardized LEI codes by 29 LOUs. Another three LOUs are still in the formation stage. In fact, the CFTC has been tentative in supporting its own designated LOU, the GMEI (Global Markets Entity Identifier) facility, first providing a two-year mandate and then only extending its mandate year-by-year until the LEI system becomes “fully operational.”⁶⁶ What fully operational means, remains to be determined. Certainly, it would seem to mean using the LEI to aggregate financial transaction data from all its consolidated entities up to its ultimate parent, a far off but primary objective.

EIOPA’s final report on its public consultation proposal for guidelines on the use of the LEI states, “The Global LEI System (GLEIS) is not yet fully operational but a number of entities, sponsored by national authorities, have already started to issue LEI-like identifiers (LEIs) in order to satisfy local reporting requirements.”⁶⁷

The FSB’s report on correspondent banking, specifically its section on the use of the LEI in payment messaging, does not support any thought as to the replacement of the BIC (Banking

Identification Code) with the LEI, rather it supports a BIC-LEI mapping convention. This assumes, falsely, that there is a one-to-one correspondence between BIC and LEI codes (right now BIC codes identify financial institutions and their branches involved in the payment system, LEIs identify counterparties in swaps transactions). It also assumes that mapping is a desirable feature of any identification solution. However, as has been discussed in this paper, it is fraught with risk as no two identifiers are updated simultaneously owing to different change notifications methods and timing updates. The report further suggests that by adding the LEI to the payment message it would achieve some benefit “... adding the LEI may reduce the number of requests for additional information by correspondent to their respondents.”⁶⁸ This benefit is hardly the transformational benefit of a universal identification scheme intended by regulators.

The Board of Governors of the Federal Reserve System, a very early supporter of the LEI, moved cautiously in recommending its reporting entities to register for LEIs, stating “... The Federal Reserve is only proposing requiring the reporting of an LEI if one has already been issued for the reportable entity at the time of collection. At this time, the Federal Reserve is not requiring an LEI to be obtained for the sole purpose of reporting the LEI on the FR Y-6, FR Y-7, and FR Y-10.”⁶⁹

The Investment Company Institute (ICI) also moved cautiously in endorsing the LEI, stating “The use of LEIs is evolving, and as such, until companies adopt it, and there are service providers that can report LEIs along with their securities identifiers, as well as systems built to support fund reporting of LEIs, funds face significant challenges in obtaining and correctly identifying LEIs.”⁷⁰

64 Bloomberg Brief, 2014, “EMIR special report,” August, <http://bit.ly/2kiNk9C>

65 Joint Consultative Paper, 2014, “Draft regulatory technical standards on risk concentration and intra-group transactions under Article 21a (1a) of the Financial Conglomerates Directive,” July 24, <http://bit.ly/2jGY6pq>

66 CFTC, 2016, “CFTC extends designation of DTCC-SWIFT as LEI provider,” July 18, <http://bit.ly/2jHa16P>

67 See footnote 38

68 FSB, 2016, “Progress report to G20 on the FSB action plan to assess and address the decline in correspondent banking,” August 25, <http://bit.ly/2bMysuC>

69 Federal Reserve, 2015, “Federal Reserve Board announces proposal requiring banking organizations to include existing Legal Entity Identifiers on certain regulatory reporting forms,” Press release, March 16, <http://bit.ly/2jJRYMH>

70 Investment Company Institute, 2015 response to the proposed Investment Company Reporting Modernization and Amendments to Form ADV and Investment Adviser Act Rules, August 11, <http://bit.ly/2jx9XEG>

It was a further disappointment that the Securities Industry & Financial Markets Association (SIFMA), one of the very first trade association supporters of the LEI, declined to endorse the LEI for its non-swaps reporting members. When asked recently by the Securities & Exchange Commission (SEC) whether its broker-dealer members should be compelled to obtain an LEI for use in uniquely identifying themselves in the newly planned Consolidated Audit Trail System (CATS) it declined to do so.⁷¹ In a final action, the SEC declined to make the LEI mandatory.⁷² In similar manner, the U.S. Treasury declined to make the LEI mandatory for much of its Qualified Financial Contract Recordkeeping requirements.⁷³

Continuing the theme of caution, such a cautious approach was recently and explicitly requested by ISDA and GFMA (the Global Financial Markets Association).⁷⁴ In a joint letter to the ROC and GLEIF, they asked to modify the finalized international/foreign branch policy document on LEI issuance, stating, “If the framework is drafted in a hurry, we risk ending up with a system that is not practical and useable.”⁷⁵

A presentation by the OFR’s Chief Data Officer, Cornelius Crowley (the OFR having been the early lead regulator on the LEI), stated, “The OFR has also seen that even though industry participants praise standards, without regulatory mandates, they may not adopt them. An incentives mismatch remains. Firms have demonstrated that they will not spend money on data-related issues to keep their own houses in order without significant public-sector involvement, as the LEI and SDR (Swaps Data Repository) experiences show. Neither group can solve this mismatch separately, presenting a challenge that must be addressed if the industry and public sector are to use the same underlying microdata to improve risk management and reporting.”

Crowley further stated that “Resolving that challenge requires that regulators continue outreach to the industry as well as participation in multinational standards-setting initiatives. It also requires that industry participants collaborate in joint standards development efforts, then adopt and use those standards. Development and adoption of standards obviously is neither fast nor easy. The result, though, should be improved data quality and lower cost for both regulators and industry, with reduced reporting burden for the industry.”⁷⁶

This lack of follow through, both by industry trade groups and regulators in the U.S., prompted the U.S. Congress to author a bill that would hold the OFR responsible for the progress of the LEI initiative. It would require that the OFR report on regulations mandating the use of the LEI to ensure the adoption

of the LEI by primary financial regulators.⁷⁷ It should be noted that even though the LEI was first championed by the OFR, and that it did work on its deployment initially as a U.S. undertaking, it did so without any reference to the LEI in the Dodd-Frank legislation. This amendment in some respects creates the justification for the OFR’s pursuit of the LEI. Right now, the OFR is pursuing the LEI under Dodd-Frank authority to pursue “other necessary data,” after direct reference to establishing an LEI was removed from early drafts of Dodd-Frank.

REFERENCE DATA UTILITY

Identifiers and reference data about a counterparty or product should be consistent across each financial transaction’s life cycle and throughout the financial supply chain. Maintenance data about changes to either should similarly be consistent across the financial supply chain. However, poor quality and duplication of this data is pervasive in large financial enterprises, in data vendors that supply proprietary codes and associated data, amongst financial market utilities, and throughout the industry, leading to significantly higher risk and operational costs. When identification codes and reference data that should be identical are not, it causes miscalculated values, misidentified products and counterparties, and involvement of multiple supply chain partners (trade repositories, custodians, paying agents, etc.) to resolve the problem. Inappropriate transactions and individual transaction failures cause monetary loss, higher labor costs, fines, and the potential for systemic failure.

Shared utilities have become a model for collaboration, including shared reference data utilities (RDUs). However, there is a fundamental misunderstanding of both the past and current attempts at establishing an RDU. Multiple sourced, multiple

71 SIFMA, 2016, “Comments to SEC on File No. 4-698: Joint industry plan; notice of filing of the National Market System plan governing the consolidated audit trail,” July 18, <http://bit.ly/2a7xk1D>

72 SEC, 2016, “Statement on the joint industry plan on the Consolidated Audit Trail (“CAT”), Commissioner Kara M. Stein, November 15, <http://bit.ly/2jV2mzd>

73 U.S. Treasury, 2016, “Qualified financial contracts recordkeeping related to orderly liquidation authority, final rule,” Federal Register, vol. 81, no. 210, Rules and Regulations, October 31, <http://bit.ly/2k8Rld7>

74 International Swaps and Derivatives Association, 2016, “Joint letter to the LEI ROC policy document regarding international branch LEIs,” Global Financial Markets Association, August 30, <http://bit.ly/2jxbHxD>

75 LEI ROC, 2016, “Including data on international/foreign branches in the Global LEI System, July 11, <http://bit.ly/2jlxwBR>

76 OFR, Cornelius Crowley, 2016, “Setting standards for microdata,” July, <http://bit.ly/2j2vPsX>

77 Bill H. R. 3340, In the Senate of the United States, 2016, “Financial Stability Oversight Council Reform Act,” SEC. 5 (V), April 18, <http://bit.ly/2jHGw53>

copies of what is intended as golden copies of the same data cannot solve the STP issue, even when all are using the same transmission standards, standard data tags, standard IDs or standard reference data, or when everyone has one of their own golden copies in their own firms or in each central securities depository or clearing facility, or in collective facilities that serve multiple firms.

Financial transactions would not match more than occasionally within swaps data repositories and within the global payment, clearance, and settlement system. Further, collateral would still be valued differently at times and customers and traders would from time-to-time be improperly notified of corporate events, if notified at all, with monies received or positions adjusted incorrectly. Mappings and data transformations would still be necessary and the STP vision would remain unrealized.

Regulatory reporting would be inconsistent with different relationships of legal entities caused by different hierarchical constructions subject to financial firm and data vendor interpretations.

The mistiming of changes to product and legal entity data would result in different product or legal entity codes and differences in reference data, resulting in deterioration of data quality over time. Systemic risk and excessive cost would still be built into the industry's infrastructure due to this still unmitigated risk and duplicated costs.

To summarize, multiple versions of identification and reference datasets, whether in central data warehouses of each financial firm or available from multiple outsourced facilities will be ineffective because of:

- The limited availability of budgets to source data from multiple vendors
- Different vendors chosen for each firm or existing infrastructure facility thus imbedding a variance in the datasets maintained by each firm and each outsourced facility
- Each firm/facility with different rules for accepting "best of-breed" data
- Duplicated activities and costs for each firm/facility essentially trying to do the same thing
- Regulators and firms still dealing with faulty definitions of aggregated risk for a counterparty whose hierarchies and definitions of business entities are determined separately by each firm/vendor
- Firms still only finding out data faults when they try to send a transaction through its settlement process and it fails to complete

- The industry still lacking the ability to accommodate STP in any time frame approximating trade date settlement, let alone real-time settlement
- Regulators still rejecting electronically filed regulatory reports because they could not match incoming data sent electronically from firms to regulators' databases
- Regulators accepting electronically filed reports because they did match incoming data from firms, but the regulators' databases had different meanings (descriptions of business entities, instrument identities, data attributes, etc.) for the matched data elements.

Notwithstanding this, multiple industry led initiatives are being pursued. Some are focused on the data and documentation required for legal entities under various money laundering, know your customer regulations, and new derivatives regulations; others are focused on issued securities and their price and corporate event data.⁷⁸ Trade groups (XBRL International, EDM Council, and ISDA) are focused on data tags to allow, respectively, at source reporting of corporate event and LEI data, semantic ontologies, and use of FpML for UPI taxonomies.

To further these efforts, a collaborative effort is necessary to promulgate and maintain such standards and to support any resultant common shared utility to replace proprietary and duplicative repositories within a single enterprise, amongst multiple data vendors, throughout financial market infrastructure utilities, and at multiple regulators. To this end, many attempts at such shared market utilities for common identifiers and an associated reference data utility (RDU) have been made in the last 25 years, as have many attempts at unifying standards groups for the common goal of setting universal standards. The former has still to be accomplished and progress on the latter, through the incentive of oversight by regulators, has been reported on in this paper. Similar regulatory incentives will be required for a universal RDU.

In fact, such a universal RDU can start with a shared LOU. Under current rules, each large registering parent entity could become an LOU and maintain its own LEI registries and the eventual requirement to maintain the hierarchies of ownership. This is not unlike the webmail servers that the larger companies maintain for the assignment and maintenance of the email addresses of its employees.

⁷⁸ PWC, 2016, "Market utilities in financial services: what role will you play?" March 22, <http://pwc.to/2jGWpIy>

⁷⁹ See footnote 13

A collaboratively shared LOU could be utilized starting with all the designated systemically important financial institutions (SIFIs). This approach would serve to facilitate the reconfirmation of the LEI data required each year. Nearly one-third (29%) of the 481,522 LEI codes issued as of year-end 2016 have not been reconfirmed and are now in a “lapsed” or non-validated state.⁷⁹ This is occurring even though many of these LEIs are presumed to have been used in swaps data reporting to trade repositories where they now sit, corrupting counterparty identification in active swaps transactions and potentially masking reconciliation, fraud, and default issues.

A recent research paper by a Federal Reserve analyst, a very early supporter of the LEI and a member of the LEI Regulatory Oversight Committee (ROC), addressed the lapsed issue, stating that “At least some entities that are marked as lapsed appear to be no longer in existence. Absent external motivation, LOUs appear to have little incentive to research such entities and, generally, users of LEI data should tend to have fewer interactions with data on defunct entities that might lead to a formal challenge to the data. Additional work is needed to determine the most efficient means of addressing this problem.”⁸⁰

While some of these legal entities may no longer be in existence, others may be active. Have they been assigned to entities not compelled to have an LEI, and by which LOU and for what purpose? Did nefarious actors obtain one and are hiding in plain site by not renewing given they have had a full year to renew? How many lapsed LEIs have actually been used in trade reporting and what are the implications for fulfillment of their contractual commitments? Or, is it simply, as the research notes, that some legal entities are not prioritizing renewals given there is no regulatory compulsion to do so? These and other issues need to be attended to.

In addition, a new Q&A issued by ESMA in December 2016 further clarifies the acquisition and renewal of the LEI of an issuer of a financial instrument.⁸¹ While noting a new concept created by the GLEIF, a Registration Authority as a new intermediary to assist in acquiring an LEI,⁸² it continues the theme of so many other regulatory mandates that the financial institution (in this case, the trading venue or systemic internalizer) is not responsible for assuring the renewal of the LEI, this time of the issuer. This, even though the ROC has specifically stated that “lapsed LEIs should not be used, be it in regulatory reporting or more generally by market participants: the associated reference data may not be up-to-date anymore.”⁸³

The thirty globally designated SIFIs are the most obvious to

collaborate on a shared LOU to register LEIs for their own legal entities. Their size, complexity, and their thousands of legal entities could be more easily managed and, in conjunction with their auditors, are better able to define and maintain the hierarchies of these LEIs. In fact, in an interview, the previous Chairman of the ROC reflected on these larger financial institutions’ ability to accelerate this effort. He was quoted as saying that “We’d love to see large players, particularly banks here in New York and around the world who are global, take the LEI and push it down through their family tree.”⁸⁴

In the U.S., where eight of the 30 SIFIs are domiciled, they already have to report their company’s subsidiaries in their financial reporting to the SEC. They also have to report in XBRL format, a computer readable language, although they do not report the subsidiary information in this language. The GLEIF has taken the first steps to organize a working group with XBRL International to record these LEIs in the XBRL format.⁸⁵ Tied to the audit process for financial reporting, this could be a great leap forward in direct input to the GLEIS.

Shared interest of these complex multi-LEI organizations, expressed through an advisory board to the GLEIF, could propel the adaption of the LEI much more quickly than regulators in each sovereign jurisdiction. It could also prompt the FSB to place the mandate for the UPI, and the UTI, with the GLEIF to design a common standard for the barcodes of finance and their reference data. Perhaps, SIFIs can also be given incentives for operational capital relief. While they have been identified by the global regulators as systemically important for stabilizing the global economy, they are also systemically important to fixing the plumbing of the global financial system.

80 Kennickell, A. B., 2016, “Identity, identification and identifiers: the global legal entity identifier system,” Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C., November 8, <http://bit.ly/2jHrvjG>

81 ESMA, 2016, “Questions and answers on MiFIR data reporting,” The European Markets and Securities Authority, December 20, <http://bit.ly/2kk7wbR>

82 GLEIF, Registration Agents, Global Legal Entity Identifier Foundation, <http://bit.ly/2eXxgVh>

83 ROC, 2017, “Update on the LEI ROC activities,” Jan. 12, <http://bit.ly/2joJLOZ>

84 Roland, N., 2015, “Banks’ adoption of LEI stalls, Treasury official says”, FIA SmartBrief, July 15, <http://bit.ly/2kiWna9>

85 GLEIF, 2016, XBRL International Forms Joint Working Group with GLEIF to Address Legal Identity in Digital Business Reporting, <http://bit.ly/2jJMMa3>

REDUCING INFRASTRUCTURE COSTS

Standardizing on a common dataset of IDs and associated reference data, on variable transaction data, and on corporate actions and defining those data elements in a common tagging language would solve some long standing problems for the financial industry. Problems such as: the systemic risk caused by mismatched counterparty transaction failures; redundant costs for sourcing, on-boarding, and maintaining the fairly static referential data that comprise 70% of a financial transaction over its trade life cycle; unnecessary costs for reconciling, mapping, transforming, and securing this data; and failures from improperly and inconsistently aggregating data for reporting of performance and risk, both internally and for regulatory purposes. In the end, it would save U.S. \$2 billion annually for each of the largest financial institutions⁸⁶ and mitigate fines that have, to date, reached as high U.S. \$9 billion for a single institution.

These issues and regulators' responses were taken up by the earlier referenced Group of Thirty (G30) study of the 1987 global market disruption, whose monitoring was ended in 2006. That study resulted in the recognition of an interconnected global financial system, the need to shorten the trade date to payment date cycle, and a need for global standards of identification and associated reference data.

In 2006, after twenty years of monitoring, the G30 study concluded that "The implementation of reference data standards has proven difficult. With no global owner of reference data and friction between the needs of the domestic and cross-border market users, progress has been slow. Future progress will require greater efforts by market infrastructure operators and international institutions with global reach."⁸⁷

The benefit of a global owner of reference data – a global "data counterparty" for setting data standards, for example, for the LEI and UPI, and maintaining their associated reference data (i.e., one golden copy) – is transformational for both industry and regulators. The cost savings of a single virtual database distributed as nodes across a peer-to-peer network versus multiple golden copies is significant. Even if each firm had one centralized data warehouse (the Enterprise Data Management or EDM model), or multiple ones shared by multiple firms/facilities in multiple outsourced facilities, multiply sourced, multiple copies of these golden copies cannot solve the STP problem.

SHARED TECHNICAL MODEL

The technical model for the RDU, as is with the GLEIS, is proposed as an intelligent federated network – a secure virtual private network (VPN) overlaid on the internet, federating sovereign databases as a single virtual view, geographically distributed, organized across individual firms and regulator sponsored facilities, and regional compacts of either or both. As stated earlier, this facility could be formed initially by the largest financial institutions (the 30 SIFIs) as an industry sponsored, government regulated, and mutually shared LOU, later to be advanced as the RDU, built around DLT concepts. This collaborative industry/government mechanism has precedent as the industry's proven way of providing assurances to each participant that the use of the datasets from such a facility will be accepted as a faultless standard, both from a regulator's perspective and within the global payment, clearance, and settlement mechanism.

The private sector, initially financial institutions, will benefit through stripping its own infrastructure of the technology, people, and data costs of duplicate identifiers and reference data and multiple mappings of identifiers. Instead, financial institutions will be able to access "component parts" in the external data management layer of the shared "parts and supply chain participant catalogues" and build business applications on top of them, rather than incorporating such catalogues in each business application.

This technical model has similarities in design to the GLEIS as proposed to the FSB in its finalized form and to the most recent manifestation of technical innovation, the much touted immutable distributed database ledger technology of the blockchain and its associated "smart contracts." All commentators and collaborators, and there are many now in financial circles, are supporting experiments in blockchain technology. While a diverse set of objectives for first implementations are being considered, they all have one thing in common, a recognition of the needed prerequisite of a universal set of financial product and financial supply chain participant identification standards and associated reference data, what is being referred to as "smart contract taxonomies."

However, only a few of these blockchain visionaries and collaborators are placing the needed priority on globally unique

86 See footnote 32

87 See footnote 3

identifiers and standard reference data, the essential building blocks of smart contracts.^{88,89} Most are in denial of the existing mature technologies that can already support their vision. That vision is the displacement of financial infrastructure, such as post trade clearing, settlement, and payment mechanisms; and real-time finality of financial transactions from order placement to posting to digital ledgers.

To this end, a first industry collaboration is needed around the current efforts of the G20's Financial Stability Board to bring unique, universal, and unambiguous identification standards into existence. As discussed earlier, this effort is now bogged down in the one market it is being tested in, the global swaps market. The true test, thereafter, is the global financial industry's willingness to cooperate further around the promise of distributed database technology. This distributed capability exists and was in use, albeit not in finance, long before blockchain incorporated such techniques.

This technology, in whatever form, can be used to establish the one missing global utility to make all the blockchain global visions practicable. That utility is the universal product and participant catalogue, a facility that has been described as a reference data utility or "golden copy" of global identifiers and associated reference data. Along with standard data tags and common datasets that describe financial transactions, a distributed ledger utility can be created to underpin all subsequent legacy systems and infrastructure reengineering promised by blockchain visionaries. We refer to this facility as the "central counterparty for data management."⁹⁰ Without it, no consequential global industry transformation can take place, as is the collective vision of blockchain enthusiasts in this digital era, a vision shared by the Group of Thirty, albeit in the earlier information era.

MORE TO DO

Mark Carney, Chair of the FSB, in his 2014 Monetary Authority of Singapore Lecture stated, "From next year, the FSB will further enhance this reporting, through an annual reporting process on implementation. This will seek to highlight both shortcomings and good practice, and will seek to assess whether reform measures are having unintended effects and must therefore be adjusted ... Initiatives to collect and share data are important – whether it be the hub built at the BIS for sharing data on the balance sheets of cross-border banks, the global aggregation of trade repository data in markets such as derivatives or repos, the global legal entity identifier, or

enhanced operation of supervisory colleges and crisis management groups for systemic firms."⁹¹

The general themes common to all of the recent consultative papers and regulations related to use of global identifiers in reporting are summarized below. The key regulatory advocates of the positions are noted in parenthesis and further details on their positions described in the earlier references associated with each:

- The identification systems and coding schemes as currently implemented are not yet functioning as intended. Swaps transaction data is being sent to regulators with standards applied inconsistently, if they exist at all. Transaction data can neither be matched nor aggregated for risk assessment across trade repositories, nor up through a hierarchy of ownership to an ultimate parent (OFR).
- With some regulators, the interim LEI codes are to be used as the counterparty code until an officially approved global LEI system is operational (EIOPA). With others, the LEI code is presumed to be finalized and, therefore, should be used as the counterparty code (ESMA). And, still others have incorporated the terms "eligible to become a LEI" (EBA) and "transition to a LEI" (FSB) as suggestive of appendages and/or modifications that might be possible to the code, to the reference data, to the methodology of self-registration, etc., as more uses of the LEI are contemplated beyond swaps counterparty identification.
- With some regulators, as no LEIs were available for international branches, a BIC (SWIFT issued "banking industry code") was to be used (Singapore). With others, it is presumed that the LEI code is finalized and should be used as the branch code (ESMA) (note: there is a recent recommendation by the ROC that LEIs are to be issued for international branches).⁹²
- While there had been no provision for LEIs to be available for individuals who transact in swaps markets (or any other market), an internal customer number was allowed in some

88 Scheufen, J. H., 2016, "Legal entity global ledger – how to liberate public data (with chains)," April, 8, <http://bit.ly/2jrSgW9>

89 Sopek, M., G. Życiński, 2016, "Using blockchain for digital identifiers: improving data security and persistence for DOI (digital object identifier) and LEI (legal entity identifier)," Mako Labs, November 8, <http://bit.ly/2kFTMK7>

90 Grody, A. D., 2009, "Infrastructure issues in the financial services industry: a case for a central counterparty for data management," *Journal of Securities Operations & Custody* 2:3, 202-224

91 Carney, M., 2014, "The future of financial reform," 2014 Monetary Authority of Singapore Lecture, November 17, <http://bit.ly/1vksnL5>

92 ROC, 2015, "Consultation document on including data on branches in the Global LEI System," October 19, <http://bit.ly/2knf89f>

jurisdictions and not in others. Some regulators had been silent on what code to use for sole-proprietors that transact in swaps markets (CFTC). Recently, the ROC has allowed an LEI to be issued for those sole-proprietors that register as a business and appear in a business registry.⁹³

- Various construction “themes” of a number of regulators can be used to construct the UTI until an official “global” UTI construction method is determined. These vary from counterparties themselves agreeing on the UTI construction (ESMA) to counterparties using a portion of the LEI in its construction (ISDA). ISDA proposes a shortened (10 character) random regeneration of the 20-character LEI; the IOSCO/CPMI proposal suggests a reverse string of the entire 20 character LEI code.
- To construct a UPI for swaps data reporting, regulators either do not make any suggestion (CFTC) or had previously suggested using any available interim UPI taxonomy until a global one is determined (ESMA). ESMA has recently proposed the ISIN code be used, even though ISINs are not yet assigned to swaps.⁹⁴
- To define underliers or index products, regulators have decided to use ISIN codes. In the E.U., an ISIN repository for trade reporting is being organized for such purpose (ESMA).
- There are no provisions for identifying reference entities uniquely or consistently, whether they can be assigned an LEI or not. In the E.U., ISIN codes are to be used (ESMA). Some regulators allow for proprietary codes such as Markit’s Red code (Singapore).

For aggregating the financial data associated with these identifiers, the ROC still has to decide on the mechanism for registering and maintaining organizational hierarchies encompassing parent or controlling entities and the interrelationships of related multiple LEIs. This is no small feat, as the largest financial participants will have to register thousands of individual LEIs. Also, to be done is the mechanism to make global changes to LEIs as corporate events such as mergers, spin-offs, acquisitions, and bankruptcies occur.

To this later point, the ROC stated recently in its paper on assigning LEIs to international branches that dealing with corporate actions remains to be done, noting similar issues as with international branches when subsidiaries of legal entities and branches of legal entities are reorganized. “Independent of the implementation of this policy, the ROC intends to conduct a more comprehensive review of the effect of corporate actions on the reference data in the GLEIS, which would encompass the effects on subsidiaries and branches. Such actions could include cases when the LEI of an international branch’s head office expires (as opposed to simply lapsing), when an

international branch or a branch network is acquired by another firm (either foreign or domestic), and when an international branch becomes a separate legal entity apart from its head office.”⁹⁵

The ROC has now placed these issues as its next priority, along with other work to be done on improving the way relationships of fund families are recorded in the GLEIS and whether status as a registered entity in a financial regulator’s database should be the determinant of issuance of an LEI for a sole-proprietor or determined by presence in a business registry. The LEI ROC expects to launch a public consultation on these issues in the first half of 2017.⁹⁶

The GLEIF has still to decide on the mechanism to federate all the disparate LEI registries into a “logical” database using an internet-like federation mechanism for a single view of the entire set of LEIs. The FSB signed off on such a system,⁹⁷ although the current version of the implementation is a physically centralized database.⁹⁸ An RFP (request for proposal) had been anticipated from the GLEIF for some time aimed toward fulfilling some or all of these expectations.⁹⁹ The RFP has still to be issued.

The current approach being followed, of consolidating multiple LEI registries daily into one centralized database by downloading data to the GLEIF, might be expedient in the interim but in the longer term will leave the GLEIS vulnerable to hackers and a single point of failure. That was the reason for the internet-like federation and logical versus physical database design proposed to and accepted by the FSB – to parallel the internet’s resilience and use of its aggregation capabilities while leaving LEI registries’ data in its original physical (country or region) space.

The FSB also signed off on a plug-in network architecture (understood to be a virtual private network using internet protocol standards for interoperability) and, thus, allow virtual

93 ROC, 2015, “Statement clarifying the conditions under which individuals acting in a business capacity are eligible to obtain LEIs,” September 30, <http://bit.ly/2jTm7tB>

94 ESMA, 2015, “Final report: draft regulatory and implementing technical standards MiFID II/MiFIR,” September, <http://bit.ly/2kcoDZv>

95 LEI ROC, 2016, “Including data on international/foreign branches in the Global LEI System,” July 11, <http://bit.ly/2jlxwBR>

96 See footnote 33

97 See footnote 21, reference 3 at page 4 and page 19

98 GLEIF, “GLEIF Concatenated file specification and user manual,” GLEIF website, <http://bit.ly/2kK0cWR>

99 GLEIF, 2015, “Board meeting – meeting minutes,” June 24, <http://bit.ly/2julqC2>, at page 2

aggregation not physical centralization.¹⁰⁰ Finally, the FSB still has to decide on how to mask counterparties that transact through privacy jurisdictions while being able to aggregate their transactions for risk analysis. The CFTC, as noted earlier and other regulators require an operational LEI system before the LEI is considered final under their regulations.

Thereafter, still left to be done is standardizing on the nomenclature and data elements associated with each unique identifier (standard reference data), initially in global swaps market for swaps products (UPIs) and supply chain participants (LEIs), but eventually for all financial products in all financial markets globally.

CONCLUSION

This is not yet the success story regulators and industry members hoped for. Neither is it the nod to a unique, unambiguous, universal, and integrated set of identification codes the industry and regulators asked for. Also, the global identification scheme, the “barcodes of finance,” is not yet useable nor scalable for its intended purpose both in the immediate term (counterparty risk data aggregation within and across swaps trade repositories globally) and in the longer term (STP and risk data aggregation to support operational efficiencies and global risk analysis across all financial market participants and all the products they trade). Indeed, we are not yet there on the short term and a long way from the longer term. Most significantly, the original recommendations to the FSB on the LEI system, to operate a “virtual” database of LEI registries as an internet-like federated mechanism utilizing a plug-in architecture and network cards, has yet to be fulfilled.

To accelerate these transformational initiatives, a move to organize SIFIs, the most put upon and most significant industry participants could facilitate the establishment of a global LOU and RDU. Already they come together, albeit in varied groupings, to accomplish some of these tasks in collaborative undertakings. Organizing SIFIs for such a global market utility would permit huge infrastructure cost savings as well as individual firm’s savings approximated at U.S. \$2 billion annually. It would accommodate data aggregation without mapping to ancillary databases; ease global access to data stored in LEI and UPI registries, and in trade repositories; and permit rapid global updating of corporate events.¹⁰¹ These are all key functions yet to be implemented that are so critical to the BCBS’s risk data aggregation principles and the FSB’s mandate to observe the contagion of systemic risk building up across the

global financial system. Neither of these objectives have yet been met.

Such a collaboratively shared utility could build upon the financial industry’s excitement over DLTs and smart contracts, and reignite enthusiasm first generated by the G30 starting in the aftermath of the 1987 market crash for solving this long intractable reference data issue. And it need not wait for DLT and smart contract proofs of concept, the technology of distributed databases and placing business intelligence in computer “objects” (referred to in the context of blockchain’s “smart contracts”) have been available and in practice for some time, albeit with different names and in different industries.

It is expected that in further probing into the mechanics of implementation, where the rubber hits the technology road, the frameworks establishing the barcodes of finance and their associated reference data will meet their ultimate test.

¹⁰⁰ See Footnote 21

¹⁰¹ Grody, A. D., P. J. Hughes, D. Reininger, 2012, “Global identification of counterparties and other financial market participants,” *Journal of Risk Management in Financial Institutions* 5:2



Banking

Opinion: Risk Culture: Risk Prevention Starts With the Individual

The Troubled Future of Global Banking

Policy Response Asymmetry and the Increasing Risks From Rising Government Debt Level

Available online:

Public Disclosure and Risk-adjusted Performance at Bank Holding Companies

What do New Forms of Finance Mean for Emerging Markets?

Risk Culture: Risk Prevention Starts With the Individual

Ulrich Hunziker – Board Member, S.U.P. (Societät für Unternehmensplanung, Basel and Frankfurt)

THE BIGGEST RISK FACTOR IS THE HUMANS

There are two reasons for regulating the behavior of international financial dealers and the conduct of international financial markets. One is to moderate and restrain greed and the other is to moderate and restrain fear. Greed and fear are the two human emotions most evident in the day-to-day behavior of the international financial system today. The result is mad money. Dealers are either drawn by greed to take excessive risks with their own – or, more often, with other people’s – money or they are overcome by fear that they will be caught out by the risks they have taken. In their rush to escape the consequence of greed, they may start a chain reaction, an avalanche of panic that carries away the innocent along with the guilty.

There are two interesting facts that make this statement interesting, if not remarkable. The first is that it was not

made in the context of the latest financial crisis, but 18 years ago by Susan Strange.¹ The second is that it became a bitter reality. Decades of deregulations stimulated the globalization of the financial markets – and unfortunately its excesses.

The dramatic events surrounding bank bailouts and failures and the impact on national budgets have led to a political rethinking, and in certain circumstances banking laws that had been abolished or suspended in order to support the development of financial markets were reintroduced. In addition, international organizations, regulators, and governments have successfully pushed for the adoption and implementation of many new regulations that will have major implications for the structure, size, and business strategy of financial institutions. A stringent risk culture in financial services organizations will, in the longer term, also need to encourage behavioral change, as a result of which the world of finance becomes more predictable and customers are better protected.

But is this sufficient? The benefits of a safety net are greatest when it is woven as tight as possible. Official regulations, laws, supervisory checks and transfers are undoubtedly essential elements. One cannot deny that the pendulum of financial regulation is currently set too “high.” But how much energy had to be spent in order to get to this point despite all that we have learned from the crisis, and where will this policy-influenced-pendulum stop when we are in the next cycle of boom and prosperity?

The most significant and sustainable solution lies with the individuals themselves. Regulations should ensure that the known risks are treated in a compliant manner. They should also help ensure that missteps, which are usually identified and punished by ex-post controls, are prevented. However,

¹ Strange, S., 1998, Mad money: when markets outgrow governments, University of Michigan Press

individuals with strong characters also behave honestly and make the correct decisions in unpredictable circumstances provided they have the necessary characteristics.

After focusing on the drivers and the impact of the financial crisis and the importance of effective and sustainable risk culture, we shall return to this question.

DEREGULATION AND ITS CONSEQUENCES

In the last 40 years, a few selected events and factors have contributed very much to the liberalization, and consequently the globalization, of world markets and in particular the financial markets. Not only were the individual events critical per se, but their interactions were also of great importance.

The first was the abolition of the Bretton Woods system in March 1973. The founding treaty was established in July 1944, with the war still raging, by 44 nations for the purpose of Europe's recovery as an economic center and trading partner of the U.S. It forced the central banks of the Member States to maintain the movements of their currencies within specified limits. Compliance with these limits has been ensured through appropriate interventions in the currency markets. Although Europe in the 1950s and 1960s experienced an "economic miracle," some shortcomings in the system became clear. Particularly serious were the dwindling possibilities of autonomous monetary policies of the Member States and the absence of a mechanism for balance of payment adjustments. Finally in 1973, the Bretton Woods system was terminated and partly, through regional arrangements (such as in 1979 the European

Monetary System), replaced. Ultimately, this provided much support for globalization – the free movement of capital.

The second important event was the Big Bang, which brought financial deregulation to the City of London. Some of the important implications of the Big Bang were initially not recognized, and the move itself was deemed by some to be merely technical and insignificant. Two main drivers eventually made the liberalization of trades in the City of London successful. One the one hand was the fact that Margaret Thatcher's political reform program "privatization and deregulation" also encapsulated the financial markets, and on the other was the fact that fee cartels with fixed rates and the separation of dealers (jobbers) and intermediaries (brokers) were of great concern for the British competition authorities. To avoid the risk of larger legal dispute, the LSE negotiated a compromise with the competition authorities. By promising to lift the restrictions on competition, an investigation by the Office of Fair Trading was dropped. Introduced as a Big Bang on 27th October 1986, these changes had groundbreaking effects: many foreign brokers and banks got access to London's stock market or took over long-established "gentlemen-capitalism" companies. The simultaneous introduction of electronic trading – beyond the trading floor and around the clock – additionally accelerated the upward trend. London rose within a few years to become a major global financial center.

The third and most important event was clearly the lifting of the two-tier banking system in the U.S., the Glass-Steagall Act. What had been introduced in response to the Great Depression in 1933, and had significantly helped prevent banking crises for 66 years, was suspended in November 1999 under

pressure from globalization and competition. However, before discussing what happened after the Act was repealed, it is important to discuss how we got to that point.

The significant increase in production in the 1920s had generated an unequal distribution of wealth; consumer loans on a large scale helped to provide some compensation (consumer loans exploded from U.S.\$100 million in 1919 to U.S.\$7 billion a decade later.) The market crash of 1929, predominantly caused by speculation, not only resulted in a massive loss of confidence in the economy, but also in the banks. Many banks fell into bankruptcy. By combining the traditional lending business with the securities business, they were doubly vulnerable; credit losses on the one hand and price drops in the securities business on the other brought the banks under immense pressure.

The Glass-Steagall Act (named after Senator Carter Glass and Congressman Henry B. Steagall) decreed the two-tier banking system. Thus, the institutional separation of the deposit and loan businesses from the securities business was demanded – the main objective was the elimination of proprietary trading by commercial banks. The Banking Act signed by President Franklin D. Roosevelt on June 16th 1933 was intended to prevent repetition of such crises. During the 1970s and 1980s there were repeated attempts to revise the restrictive law or to override it. Competition considerations increasingly encouraged these efforts. Consequently, in the 1980s, the U.K. and Canada lifted their two-tier banking system. The U.S. banks fell further behind in the international rankings: in 1960, six U.S. domiciled banks were ranked in the top ten, by 1980 there were two and by 1989 there were no longer any U.S. banks among the world's top 25.

The pressure from the banking lobby and politicians increased. The Banking Law was considered “out of line with reality.”² “It was now the time for laws to adapt to changes in the international financial system.”³ Finally, the Act was repealed in 1999 by an overwhelming majority in Congress and the Senate. The normative power of the facts had been successful.

As a result, the deregulated, global financial markets developed a great “momentum” and there was some decoupling from the real economy. Terms such as “Casino Capital” and mad money found their way into the economics literature. Susan Strange commented as follows:

Bankers therefore are still with us but they are not what they used to be. Intermediation – taking in deposits and making loans – which was their traditional function, is no longer the name of the game. Commercial banks have become investment banks and are increasingly tempted into proprietary trading – that is, betting their own capital in the casino.⁴

The compensation system was a consequence of the fact that banks could make money out of money. The increasingly exorbitant remuneration, particularly in the investment banks, was increasingly becoming a socio-political problem. In the boom years of the new millennium (2003 to 2006) the growth was extrapolated – the rapid development of financial markets were carried forward into the future. Getting oneself into debt, with the use of, in particular, derivative financial instruments being developed on an almost daily basis spiraled out of control. This development found an abrupt end with the subprime crisis in the U.S.; in September 2008, Lehman Brothers went bankrupt. Worldwide governments had to step in

to save the banks and protect the financial system from collapse.

THE HIGHLY REGULATED POST-CRISIS FINANCIAL WORLD

The fact that the efforts to introduce more stringent rules during the deregulation phase were extremely difficult is illustrated by another quote from Susan Strange:

By 1996, the BIS had virtually thrown in the towel on capital adequacy rules. It abandoned, in effect, the whole idea of agreed common standards of banking supervision. This virtual U-turn is not easily perceived by reading its annual reports, which naturally concentrate on the institution’s positive achievements. The BIS general manager, Andrew Crockett, however, in an unofficial study has explained why a policy that had been developing for twenty years was finally abandoned (Crockett 1997).⁵ No regulatory system, he observed, was perfect, and applying standard rules to banks in very divergent national banking systems encountered all sorts of difficulties and dilemmas.

It was as a consequence of the crisis that the worst suspicions were confirmed and the weaknesses of the regulation mercilessly exposed. Financial institutions had to be rescued on a large scale; a redistribution of the debt burden and responsibility of private institutions in the U.S. was the result. Politicians and regulators appeared prominently on the scene – it was time for more stringent measures. Although the financial industry is already highly regulated in many areas, the recent financial crisis has been used as an opportunity to strengthen the regulation

significantly. Individual countries have unilaterally added various restrictions to national laws. Internationally, banking regulation is driven by the Basel Committee on Banking Supervision and the Financial Stability Board. In particular, the reform package “Basel III” should resolve many of the previous weaknesses. The fact that the regulation is now more comprehensive and complex becomes vividly apparent when one considers that the Basel Convention of 1988 comprised 30 pages, the succession plan “Basel II” had 347 pages and “Basel III” is 616 pages.

Ironically, the two-tier banking system, which was abolished in the U.S. in 1999, is highly topical in Europe: at a press conference on February 6th, 2013 Wolfgang Schäuble, the German Finance Minister, declared the announcement of the design for two-tier banking system:

We know that excessive deregulation was a mistake. Back then, we were blinded by technological innovations, new financial products and rapid developments in the financial markets and the realization is that we have to – also for the financial markets – create an open framework to enable financial markets to function as a market and not destroy themselves.

An essential element characterizes the regulations according to the

2 Taylor, A., 1995, “Banking: Jim Leach gently rocking the boat,” CQ Weekly (Washington, D.C.): 1162, April 29

3 Glater, J. D., 1995, “Rubin urges changes in U.S. banking laws; Treasury chief would end curbs on competition,” Washington Post, February 28, p. C.01

4 Strange, S., 1998, page 9

5 Crockett, A. D., 1997, “Why is financial stability a goal of public policy?” Economic Review, issue Q IV, 5-22

repercussions of the recent financial crisis: the focus is not only the safety of individual banks (or their customers) but also the implications for the entire financial system and the macro economy. To prevent future contagion and minimize the impact of future crises, special arrangements for so-called “systemically important banks” were adopted. In the current context, it refers in particular to implementation of Basel III and the guidelines for how to manage too-big-to-fail institutions, the execution of which is in full swing within most, if not all, of the major financial centers. The text and implications of these guidelines are currently available in detail from numerous professional publications, magazines, and newspapers.

A SUSTAINABLE RISK CULTURE NEEDS TO BE DEVELOPED AND “CULTIVATED”

The implementation of, and compliance with, statutory-regulatory rules and requirements must be ensured in individual banks. The fact that this leads to larger projects and requires an army of specialists in the legal departments (compliance), is shown in the following figures: in 2011, banks worldwide adopted about 14,000 new global regulatory changes, with over 40 new regulations being implemented in the compliance department of a global bank each day. Based on this, instructions are issued and their compliance regularly checked. However, this is not enough – the correct behavior must be embedded in the organization’s risk culture.

An effective risk culture facilitates compliance with all regulations and laws and can make a significant, positive contribution to improvements on and retention of corporate value. This is

particularly important for financial institutions because taking and managing risks is the core of the business. The recent financial crisis demonstrated how quickly reputation and credibility can be lost – with severe consequences. However, for an effective functioning of a sustainable risk culture some essential factors are “conditio sine qua non” in order to bring the written guidelines to life.

First and foremost is the role model. Each employee can read numerous documents on how to behave ethically and comply with the codes, and it is essential that these expectations are continuously communicated. However, the decisive factor is ultimately that the execution is exemplified consistently by the top management, and whether this is also recognized in the organization.

Of particular importance are the structural and organizational parameters. The provision of clear guidelines on responsibilities and efficient processes should ensure that the risks are systematically identified, weighted and provided to the relevant decision makers at the right time with appropriate proposals for management. In large institutions it is especially important not to leave this process entirely with the specialists. Each individual is a risk manager in their own environment, while the specialized units focus “full time” and independently on risk control. Risk committees at various levels should fulfil their responsibility to take appropriate action and monitor these actions.

The fact that deep technical expertise and continuous training are necessary is obvious. In addition, however, other skills and talents are in demand. A good understanding of the probable risks and a cross-departmental view (front-to-back) are also important requirements for rapid risk identification.

This is especially important for operational risks, since they usually cannot be measured with quantitative models and require qualitative assessment capabilities.

Just as performance targets and contributions to consolidated earnings are considered in the evaluation and compensation processes, risk behavior must also be an integral and mandatory part of this process. Risk adjustments should not only be made with regards to financial results, but also with regards to remuneration. There are excellent models, with a sufficiently high number of well-defined KPIs (Key Performance Indicators), for measuring risk behavior that can be incorporated into the annual performance review. These models do have high success rates, but require that they are applied consistently.

However, an effective risk culture is one that not only facilitates an open, critical environment for expression and discussion, but demands it. Risk control must not be limited to mechanically checking off individual control steps. Complex mathematical models and stress tests are essential tools but provide ultimately only “raw results.” To ensure that the right decisions are taken it is necessary that critical, qualitative assessments are accompanied by a dose of common sense. It was, for example, already much too late when the highest (AAA) ratings of subprime securities and their issuers were questioned in the recent financial crisis.

THE INDIVIDUAL IS AT THE HEART OF THE PROCESS

The introduction of a dense network of regulations has certainly made the financial markets more secure. Their forced implementation within financial

services organizations has once again led to greater focus on risk culture. Regulators and politicians will ensure that the pressure is maintained for even more stringent measures in this area. But can the safety net be so tightly woven that it can prevent future crises and thus restore the reputation and trust in banks? Even the most intensive regulations are only as effective as they are understood, accepted and complied to – voluntarily or involuntarily.

The greatest risk factor is the human being. Consequently, the real question is how can this important risk be managed. The evaluation criteria employed in the financial services industry focus primarily on professional expertise, performance, potential and motivation. Social skills have been largely neglected, as has character building. Current assessment methods have avoided this issue. The crucial question of character, however, hinges on the understanding of virtue of the individual. This question is so important since in today's world qualities that are often expressed as vices are considered desirable – for example greed.

If there are methods that can be used for the identification and analysis of the characteristics of employees and managers, then you should use this methodology because there are numerous indications that sustainability and prosperity are particularly linked to the question of character. Research shows that the modern crisis managers who are in great demand today are those characterized by outstanding character values.

As a result, a very important element in instituting an effective risk culture is the ability to judge the character of one's staff. Interestingly, this crucial component does not as yet appear on the radar of most businesses. Integration of

a systematic "character assessment" in the top management selection process could lead to a key competitive advantage. With a proactive analysis of attitudes on the topic of "character assessment" in connection with management's portfolio planning it is clear that a key competitive advantage could be achieved. Or in economic terms: if only a small fraction of what was lost in losses and fines during the financial crisis would be invested in such a methodology, a remarkable return on investment could be achieved.

The Troubled Future of Global Banking

Brad Hintz – Adjunct Professor of Finance, Stern School of Business, New York University

Roy C. Smith – Kenneth Langone Professor of Finance, Stern School of Business, New York University¹

Abstract

After the financial crisis of 2008, global capital market banks have been the focus of a battery of new regulatory initiatives coming from international organizations and national regulators. Assertive supervision, limitations on permissible activities, higher capital, and improved liquidity standards were intended to reduce systemic risk to the global financial system and make it far less likely that banks will need to be assisted by governments in the future. As a result of these changes, stability has returned to the global banking industry. But the regulatory measures combined with the slow global economic recovery have led to a prolonged decline in the performance of the capital markets business. Indeed, the increased regulatory burden has rendered the banks themselves economically unviable and, therefore, considerably less safe than they

were. Capital market banks, therefore, face the painful task of changing their business strategies and component configurations, a task that most have avoided addressing meaningfully. This paper discusses the evolution of bank regulation through the financial crisis and demonstrates how it has affected the market leaders that have been unable for several years to achieve returns on equity equal to the cost of that equity, and whose stock prices currently average only 77 percent of book value. It also discusses the strategic change options available to the banks.

¹ Brad Hintz was until 2015 Senior Banking Industry Analyst at Sanford Bernstein. Roy C. Smith is a former General Partner of Goldman, Sachs & Co.

INTRODUCTION

After the great crash of 1929, Wall Street was pilloried for its excesses and blamed for the Great Depression that followed, despite plenty of evidence that missteps by the Federal Reserve were at least partly responsible. But the blame brought with it the most extensive financial reforms ever enacted by any country. These included the omnibus Banking Act of 1933, which reformed and fortified the Federal Reserve System, introduced federal deposit insurance, and prohibited commercial banks from participating in most capital market activities through its Glass Steagall provisions.

Glass Steagall remained in place for sixty-six years. For the first fifty of those years, it helped to keep the U.S. banking system safe and sound, and there were very few bank failures. Most banks were content to follow the rules, to pay decent dividends, and to grow only at about the same rate as the economy.

But this cautious operating policy changed in the 1970s. Walter Wriston, CEO from 1967-1984 of First National City Bank of NY (later Citicorp) introduced the idea that banks too could be “growth stocks” if they were well managed and captured opportunities to expand overseas and beyond traditional regulatory limitations in the U.S. Growth stocks were then thought to be companies that could generate annual profit growths of 15% or more indefinitely.

Growth became the mantra of Citicorp and its many banking competitors. Lending rapidly expanded into recycling “petrodollars” into large loans to less developed countries, and to commercial real estate and oil and gas production across the U.S., invigorating a boundary dispute between banks, thrifts, securities firms, and insurance companies as the largest banks attempted to expand their business footprint.

A FORGOTTEN BANKING CRISIS

Continental Illinois Bank and Trust, the seventh largest bank in the US, aggressively followed the Citicorp model. For it to grow at 15% per year, however, it would have to double earnings every five years, which meant doubling the size of the balance sheet that the bank had taken almost a hundred years to assemble. To do this, the bank would have to waive traditional credit concerns in the interest of booking the new loans. Soon, the bank’s capabilities to manage and control credit risk went by the board. “Our systems broke down. It was a terrible

mistake,” said Roger Anderson, Continental Illinois’ chairman, when describing the extraordinary growth of the bank’s exposure to failing energy loans in 1984 that ultimately led to a run on its institutional funding sources and its takeover by the Federal Deposit Insurance Corp. and the Federal Reserve [Lascelles (1982)]. After its collapse, it soon became apparent that dozens of other important commercial banks had similar loan portfolio problems and had to be shored up, merged with others, or subjected to special regulatory attention in order to avoid failure [Smith (1993)].

For the next decade, U.S. commercial banks were in the penalty box, unable to grow their balance sheets and losing market share to investment banks that had devised a number of products to enable corporations to bypass the bank’s wholesale lending business and finance in capital markets. The SEC’s Rule 415 (shelf registration) considerably eased the issuance of corporate bonds; commercial paper (short term promissory notes) displaced bank working capital loans; medium term notes replaced bank term loans; and “securitization” enabled the sale of long-term bonds backed by packages of mortgages or other assets [Smith (1993)].

REPEAL OF GLASS-STEAGALL

By the early 1990s, the banks claimed that they had reformed and returned to the basics of “good” banking. They were financially solid once again, they said, but capital market disintermediation and competition from investment banks and foreign banks not subject to Glass-Steagall were killing their businesses, and they needed relief. They wanted Glass-Steagall to be repealed so that they could freely compete in capital markets with the others.

Opposition to repeal gradually melted away as regulatory policy shifts occurred. In 1987, the Federal Reserve allowed limited underwriting activity by banks under provisions of the Glass Steagall (Section 20) that allowed for exceptions. By 1988, Alan Greenspan, Chairman of the Federal Reserve, was supporting the call for deregulation, stating that the “... near-complete repeal of the Glass-Steagall Act allows for a market-driven evolution of financial services and products” [Berry (1988)]. Greenspan and others believed that repeal would increase competition in financial services and that market forces would carefully monitor banks and punish any unwise or unsafe activity.

Over the next decade, banks gained support for repeal within

the Executive branch and in Congress. In 1995, U.S. Treasury Secretary Robert Rubin announced his support for the end of Glass Steagall [Roberts (2014)]. The catalyst for final removal came from the announcement in 1998 that Travelers Insurance (which had previously acquired Smith Barney, a broker, and Salomon Brothers, a major investment bank) would merge with Citicorp. The transaction was prohibited by Glass Steagall, but the parties structured their deal so that it would be undone if the law was not repealed, which, it was.

The Citicorp/Travelers merger created Citigroup, the world's largest "diversified financial services company," one that was active in banking, insurance, capital markets, and asset management, at both retail and wholesale levels, in the U.S. and overseas.

The merger triggered a number of similar cross-industry combinations in the U.S. and in Europe. One by one the leading banks decided to deemphasize the stodgy ways of commercial banking and to recreate themselves as capital markets financial "conglomerates."

THE CAPITAL MARKETS EXPANSION

As regulations relaxed, the large banking institutions expanded their nascent capital markets franchises through aggressive pricing and cross selling of new investment banking products to corporate "relationship" clients. By conditioning the continuation of long established credit relationships on the achievement of a bank's profitability target, clients were encouraged to include banks as co-managers of mergers or stock and bond issues, despite the fact that doing so was technically prohibited by existing regulations.²

This strategy was successful. By the end of the 1990s, commercial banks had joined the ranks of leading U.S. debt and equity underwriters and were successfully competing with investment banks around the world.

The technology market collapse of 2000-2002, which resulted in three years of record bankruptcies and a multitude of credit losses and underwriters' liability claims, did not change the goals of the largest banks. The lure of capital markets revenue and market share growth remained a strategic imperative. CEOs, corporate strategy departments, consultants, and equity analysts all believed that the growing capital markets were the future of banking and that quantitative risk management techniques combined with the "great moderation" brought

about by modern central banking policies supported this business shift.

The most aggressive banks were characterized as "flow monster," able to sit astride and closely monitor market orders (flows) around the world, and to position themselves accordingly. They were also able to leverage their trading books considerably, even within the context of the Basel I minimum capital adequacy rules imposed in 1992. Thus, these firms were able to capture substantial trading profits that accounted for more than half of total revenues for some.

In the period leading up to the financial crisis of 2007-2008, banks pursued a capital markets strategy based on massive balance sheet commitments to support trading, the manufacturing of innumerable structured and synthetic securities using derivatives, and positioning investment banking coverage teams globally in order to serve clients anywhere in the world. By 2008, after several investment banks had been acquired by large banks, eight of the top ten capital market firms (ranked by origination of transactions) were diversified financial services companies.³

THE 2008 CRASH AND ITS CONSEQUENCES

The financial crisis of 2008 caught the banks largely unaware: liquidity disappeared as markets seized up and risk models proved inadequate. By the end of that year, AIG, Bear Stearns, Fannie Mae, Lehman Brothers, Merrill Lynch, and Washington Mutual were no longer independent firms, and Goldman Sachs and Morgan Stanley, who were, were no longer investment banks but bank holding companies (BHCs). Many of the banking conglomerates, both American and European, had to be bailed out with taxpayer funds to avoid failure. The 2008 crisis led to an economic recession and slowdown that reduced world economic output by 3.6%, and from an annual growth rate of 5.3% in 2006 to 3.3% in 2015 [IMF (2010, 2015)].

² Provision prohibiting credit tying arrangements by national banks are part of section 106 in the Bank Holding Company Act Amendments of 1970.

³ The authors have ranked the banks by the total value of transactions originated (ranked by full-credit to "book-runners") in four categories reported by Thomson Reuters: global bonds, global stocks, global M&A, and global syndicated bank loans. In 2008, the rankings were as follows: 1) J.P. Morgan/Bear Stearns, 2) Goldman Sachs, 3) Citigroup/Salomon/Smith Barney, 4) Bank of America/Merrill Lynch, 5) Morgan Stanley, 6) UBS/SG Warburg, 7) Deutsche Bank/Morgan Grenfell/Bankers Trust, 8) Credit Suisse/Donaldson Lufkin Jenrette, 9) Barclays/Lehman Brothers, and 10) BNP/Paribas. All but Goldman Sachs and Morgan Stanley were diversified financial service companies.

A consequence of this recession and slowdown was a 25% reduction in capital market new issue volume, from a peak of U.S.\$10.2 trillion in 2006 to an average of U.S.\$7.7 trillion over the next ten years, and a 16% reduction in completed merger and acquisition transactions, from U.S.\$2.1 trillion in 2006 to a ten-year average of U.S.\$1.7 trillion.⁴ New issues and merger transactions represent a substantial portion of investment banking revenues for the major firms.

The 2008 crisis was also followed by the most extensive regulatory reform and tightening since 1933. This was achieved by a substantial modification of the Basel minimum bank capital adequacy accord (to “Basel III”), the passage of the Dodd Frank Wall Street Reform Act (2010), the creation of the Financial Stability Board of the G-20 group of countries and the European Banking Commission, and a series of new regulatory powers being ceded to the European Central Bank.

COMPLEXITIES AND CONSEQUENCES OF LARGE BANK RE-REGULATION

As a result, thirty designated “global systemically important banks” (G-SIBs), including all of the global capital market banks, have been the focus of a battery of new regulatory initiatives coming from international organizations and national regulators. Assertive supervision, limitations on permissible activities, higher capital, and improved liquidity standards were intended to reduce systemic risk to the global financial system and make it far less likely that banks will need to be assisted by governments in the future.

As a result of these changes, stability has returned to the global banking industry: the banks’ credit default spreads, which soared during the crisis, have returned to near pre-crisis levels, and the “betas” (stock price volatility) of the largest banks have declined from peak levels. But the regulatory measures combined with the slow global economic recovery have led to a prolonged decline in the performance of the capital markets business.

The new regulatory rules form a labyrinth of constraints on the capital markets banks. They materially restrict the amount, type, and the riskiness of assets that a bank may hold for a given amount of capital, and thus narrow potential business strategies that a bank may pursue. Nevertheless, within the many regulatory constraints there is presumed to be a “safe harbor” – an area that is allowed under the regulatory standards in which a bank can operate freely and base its future activities upon.⁵

Logically, a bank should be able to choose a business strategy that optimizes its balance sheet and maximizes its return on equity (RoE) under existing regulations. Such an optimal mix, or the RoE “sweet spot,” will typically exist along the edges of the regulatory constraints that form the safe harbor.

But that sweet spot has proven to be elusive for G-SIB banks. The safe harbor is too small to support an RoE that can provide long-term economic viability for them.

In addition to well-defined capital constraints, risk weightings and “prudential” cushions, banks face a set of annual “stress-tests” that form the true binding constraint on their business activities. These derive from annual stress-tests required by the U.S. Federal Reserve, the U.K., and European authorities.

The U.S. test, the Comprehensive Capital Analysis and Review, or CCAR, is different from past quantitative measures because the regulators now adjust stress assumptions and measure outcomes based on qualitative policy and risk scenarios that are not communicated to banks in advance. The tests, therefore, are more than just a test of bank capital, and can be difficult to predict. The regulators also use the tests to ensure that there is no gaming of capital rules by banks (as happened extensively before 2008) and that potential financial bubbles can be addressed.⁶

The consequences of failing to pass an annual stress-test can be severe. In the U.S., dividend payouts and stock buyback plans, and a variety of other issues in which regulatory consent may be sought, can be denied.

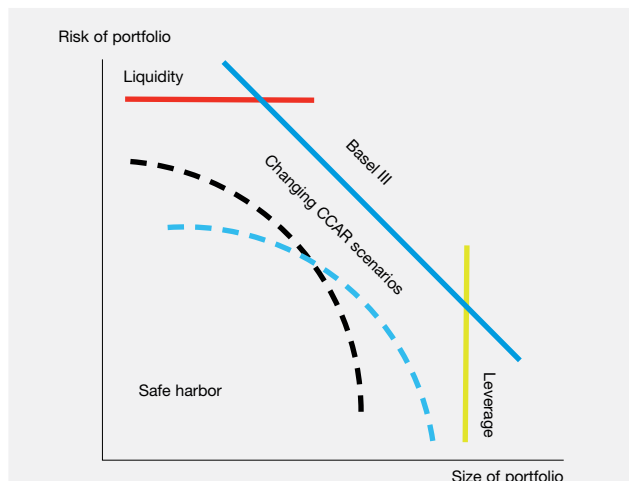
The uncertain frontiers of the stress-tests in effect reduce the size of the safe-harbor further, and that obstructs banks’ ability to fully maximize opportunities under the official regulatory constraints to reach an optimal RoE sweet spot. To be certain that an institution will pass the test, a bank must maintain a safety cushion well above the published minimums (Figure 1).

And the stress-tests continue to tighten. The Federal Reserve has proposed increasing the minimum capital limit by including the G-SIB surcharge, a further capital cushion that ranges from 1% to 3.5 percent, a “stress capital” buffer, and

4 Data from Thompson Reuters.

5 “Safe Harbor” is a legal term used by the SEC, but not as yet by bank regulators.

6 Some observers assert that the CCAR is in violation of the Administrative Procedures Act of 1946 that requires government agencies to be transparent and publicly accountable [see Scott and Gulliver (2016)].



The figure illustrates the situation facing a capital market bank. The vertical axis is the risk of a banking portfolio and the horizontal axis is the size of the portfolio. Given a fixed amount of capital and core funding, there are three constraints that limit a portfolio: a liquidity constraint, a Basel III risk-weighted assets limit, and a leverage limit. These three constraints form a frontier along which a bank is employing all its capital while meeting all regulatory requirements. Along this frontier, there is a point that defines an optimal portfolio that will maximize a bank's RoE. But the CCAR process makes it impossible for a bank to reach this optimality. Moreover, the bank cannot move to another RoE optimality along the CCAR frontier because the CCAR test is dynamic and thus the frontier is constantly changing. This forces banks to maintain business portfolios well inside the limits of any CCAR test. Furthermore, it traps unused banking capital between the annual CCAR stress-test frontier and the regulatory minimum targets of core tier-1 capital ratio (CET1), liquidity, and leverage. This is a prescription for weak equity returns.

Figure 1 – Operating constraints of banks

a “countercyclical capital” buffer. These would further narrow the safe harbor of large bank corporate strategies and complicate defining an optimal balance sheet mix.

Adding to this operating problem is the fact that new, prescriptive regulatory rules continue to be introduced by regulators that require changed business procedures, additional regulatory capital, increased liquidity reserves, and expanding compliance and control systems.

In the U.S., the Federal Reserve is expected to soon adopt Basel/G-20 rules on a liquidity measure called the “net stable funding ratio.” This will impose on BHCs four interrelated requirements consisting of a liquidity coverage ratio, a net stable funding ratio, a further G-SIB capital buffer, and a comprehensive liquidity assessment.

Further, regulators are continuing to boost large banks’ “total

loss-absorbing capital” by insisting that banks increase their issuance of long-term debt (so as protect balance sheets from overdependence on short-term bank deposits), and that long-term debt investors understand that they will be expected to participate in any losses resulting from any future bank restructuring efforts. This emphasis, of course, further affects the banks’ debt ratings and interest rates paid on the debt and their regulatory and compliance costs have increased markedly as a result of these continuing changes.

In addition, there are unknown regulatory changes that may be on the horizon. Neel Kashkari, the President of the Minneapolis Federal Reserve Bank, is calling for the legal conversion of large banks into public utilities. And American politicians on both sides of the political spectrum support the system-wide breakup of large banks so that they can no longer be considered “too big to fail.”

FALLING DOMINOS: LITIGATION, LIQUIDITY, TALENT, AND COMPETITION

Other changes have also occurred that were not foreseen when the new regulations were introduced. Key among these was an enormous wave of litigation that held the banks to be responsible for the economic harm imposed by the crisis, settlements of which further depleted their capital by approximately U.S.\$200 billion by 2015 and which further increased during the following year as the U.S. Justice Department worked through its list of banks yet to settle charges of mis-selling mortgage-backed securities before and during the crisis of 2007-2008.⁷

Competitive changes also occurred – first was a migration of talent from G-SIBs to unregulated alternative asset managers, such as hedge funds (for traders) and private equity firms (for deal makers). Next was an increase in the number

⁷ These settlements resulted from litigation from regulators or other government agencies, predominantly in the U.S., and from some smaller class action settlements. They involved only about twenty banks. In the 2002-2005 period, following the failures of Enron, WorldCom, and many other technology firms, similar bank settlements of litigation totaled less than about U.S.\$30 billion. The size of settlements in the recent period reflects the frustration by government officials and much of the public in the U.S. and Europe that banks should be punished, and their officials held responsible for their contribution to the economic hardship associated with the Great Recession. To date, many banking officials were investigated, but no one who was an officer or director of a bank during 2007-2010 was charged with an offence.

of independent investment banking “boutiques” set up by star bankers able to attract corporate clients and avoid the regulatory burden of the big banks. Since 2008, boutiques have doubled their share of the merger advisory business, to about 18% of deals, according to Dealogic.

Finally, other large banks, such as HSBC, Wells Fargo, and Royal Bank of Canada (RBC) that had previously not been considered capital markets leaders, increased their shares of the market for capital market activity. By 2014, Wells Fargo and HSBC ranked ninth and tenth among originators, respectively, and RBC had built strong franchises in North America and the E.U. These new participants did not attempt to duplicate the global strategies of the conglomerate banks, but rather chose to pursue a strategy of selective competition. They offered limited high margin capital markets services to particular clients in targeted regions and industries.

Beginning in 2010, it was clear that the capital market banks would have a hard time earning an appropriate return on their equity capital. By 2015, it was also clear that the top ten originators of capital market transactions were losing market share. Together, the top ten accounted for an 81% market share in 2009, but this share was reduced to 67% in 2015.⁸ On the basis of revenues from investment banking activities Dealogic reported a similar loss of market share among the top ten ranked banks.

Further, secondary market trading in fixed income, currencies and commodities, and in equities declined considerably as banks withdrew capital from this activity in order to reduce risk-weighted-assets that were subject to greatly increased capital requirements. Markets, thus, were deprived of important liquidity. The previously prized, leveraged trading business model, characterized by former Morgan Stanley CEO John Mack, as “credit was free and you were paid to take risk” is gone [Mack (2009)].

DOUBTFUL ECONOMIC VIABILITY

Nearly a decade after the crisis, the large conglomerate banks are still standing, but investors do not believe that they are capable of delivering adequate returns over the next five years [Broadridge and Institutional Investor (2015)].

As a result, these banks face a grim future of forced restructuring and change. Capital requirements have depleted much of the value in large balance sheets dedicated to trading and flow

monster business models. Banks will have to be smaller and more manageable, at least until they settle in to a new economic model that works. They will also be less leveraged, less dependent on trading, and are prohibited from making large acquisitions to achieve growth targets.

After a one-year RoE rebound in 2009, the major capital market origination banks have been unable to consistently earn a return on their equity capital greater than the cost of that capital. And, while the large capital market banks are certainly financially stronger because of capital increases and required risk-reduction efforts, it is arguable that the massive integrated global capital markets business that was built over twenty-five years is no longer viable, and its leading firms are no longer safe.

We judge this through a simple calculation of “economic value added” (EVA),⁹ the difference between reported RoE and the cost of equity capital.¹⁰ The average EVA for the top ten capital markets banks ranked by origination and advisory volume¹¹ for the eight-year period 2008-2015 was -8.6%. The average EVA was -23.6% in 2008, the worst year, but was still -9.2% in 2015 (Figure 2).

As Figure 2 illustrates, the banks have suffered from unusually high costs of equity capital, largely because of the high volatility of the bank’s stock (its beta), a measure of the riskiness of future earnings. Despite a clear reduction in balance sheet risk as a result of the regulatory changes, and presumably because of strategic and other uncertainties of the future, these betas rose after the crisis to levels close to 2.0. In the past, it was thought that a regulated bank’s stock beta ought to be close to 1.0, or about the same as the volatility of the entire market. The average beta for the top ten capital market originators in 2015 was 1.8 (Morgan Stanley’s was 2.3, Citigroup’s was 1.9).

8 Based on Global Capital Market originations in 2009 and 2015, prepared by the authors.

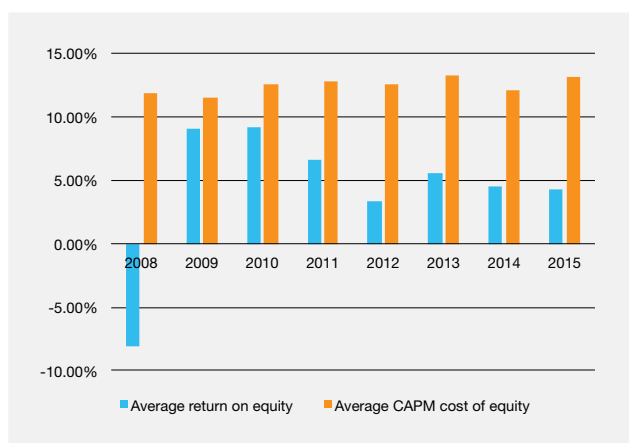
9 EVA is a registered service mark of Stern Value Management, an affiliate of Stern Stewart & Co.

10 For our purposes, RoE is the reported return on all equity capital, and the cost of equity capital is determined by the Capital Assets Pricing Model (i.e., the “risk free rate” plus the product of the equity risk premium of the market and the banks’ own beta).

11 The capital markets banks that originated and advised on the largest dollar value of transactions over the period 2008-2015, based on Dealogic data, were Barclays, Bank of America, BNP-Paribas, Credit Suisse, Citigroup, Deutsche Bank, Goldman Sachs, J.P. Morgan, UBS, and Morgan Stanley. On an annual basis, this ranking is different. On a revenue basis, the top ten ranking is also somewhat different.

Concern over strategic uncertainties is justified. Changed capital requirements and business prohibitions have turned the large balance sheets dedicated to trading into stranded assets. Proprietary trading has been banned. The model of market making in which liquidity was provided to clients to capture flow for positioning is largely gone. Leverage limitations and

regulatory oversight have scaled back prime brokerage and matched-book repo customer lending. The move to central clearing of derivatives has compressed profits that previously flowed from the natural turnover and modification of bilateral swaps books-of-business. And, “skin in the game” rules and continuing customer skepticism about opaque product structures has constrained securitization volumes. These factors together suggest that longstanding business models in the industry need to be changed (Table 1). The uncertainties are no longer over whether to do this, but how and to what extent.



Source: Mergent, Inc. Data: Summary financial data and ratios for the period 2008-2015 (2016). Mergent Online database. University of Arizona Library, 4 August, 2016. Analysis by Hintz.

Figure 2 – Average RoE of top 10 global capital market banks as ranked by origination and advisory activity level

	2008-2015 origination and advisory activity rank	2015 origination and advisory activity rank	8-year average EVA [®] spread (%)	2015 EVA [®] spread (%)
J.P. Morgan	1	2	-0.9%	-2.3%
Bank of America	2	1	11.2%	-6.3%
Goldman Sachs	3	3	-0.9%	-5.4%
Citigroup	4	6	-13.7%	-6.2%
Morgan Stanley	5	4	-6.6%	-7.9%
Barclays	6	5	-13.7%	-13.5%
Deutsche Bank	7	7	-12.7%	-23.1%
Credit Suisse	8	8	-7.5%	-18.9%
UBS	9	10	-13.3%	-1.0%
BNP Paribas	10	11	-5.6%	-6.3%
Average of top 10			-8.6%	-8.9%

Source: Thomson, Dealogic, Mergent, YCharts, Aswath Damodaran, and analysis by Hintz and Smith.

Table 1 – Capital efficiency of the top 10 global capital markets banks as ranked by origination and advisory activity level

With cost of capital still high and RoEs still low, the banks have also suffered from low price-to-book value ratios;¹² the stocks of the top ten capital market banks have averaged a 0.77 price-to-book ratio for 2015, with half below 0.70 (The average was 0.78 in November 2016). Banks sell at less than book value because investors either doubt the value of their assets or the future efficacy of their strategy. That is, if the bank’s basic business strategy is in doubt, then investors will not want to pay full liquidation value for it.

RETHINKING BUSINESS STRATEGIES

Thus, the global banks are caught in a global regulatory dilemma that has limited their freedom and undermined their business performance. Today, many bankers believe their “number 1 client is the government,” [Wall Street Journal (2013)] and this seems unlikely to change.

This fact makes the choice of a new business strategy that will generate reasonable long-term returns for a global capital markets bank a challenging task. The inherent volatility of the capital markets business, the differing impact of new regulation on each sector of the business, the strengths and weaknesses of product-line market shares, and the geographic strengths and shortcomings of each firm preclude a standardized approach to strategy. Still, there are several observations regarding strategies and change that external observers of the industry can make.

Capital markets mix

Banks that have higher-margin businesses within capital markets have an advantage. That is, banks with a greater portion

¹² Book value is the accounting estimate of the liquidation value of the bank – its assets are recorded at either market value or values reflecting reserves for losses, and its liabilities are at face value.

of equity underwriting, high margin fixed income underwriting, and mergers and acquisitions advisory revenues have the flexibility to improve performance by adjusting their capital markets business mix and constraining the growth of trading. Goldman Sachs and Morgan Stanley have the industry's highest percentage of high-margin banking businesses relative to total capital markets revenues.

Reliance on trading revenues

Banks that have relied more on sales and trading revenues, and particularly on lower margin fixed income revenues, are at a disadvantage. The capital rules implemented over the last eight years have reduced the RoE that is achievable from trading to single digits.

The operating prohibitions of the Volker anti-proprietary trading rule has changed the business model of trading, sharply curtailing the risk taking previously employed in the business. Leverage and liquidity rules have constrained balance sheets, increased funding costs, and reduced the ability of the banks to provide customer financing. Essentially, every new regulation that has been implemented since 2008 is telling the global capital market banks to constrain or shrink their trading units.

Diversification

With the sole exception of Goldman Sachs, over-reliance on capital markets remains a weakness for any bank. Banks with profitable revenue sources outside of capital markets are now able either to deliver better RoE and/or to weather what is likely to be a prolonged period of adjustments in the market making business until profitability is returned.

Among the global capital markets leaders, J.P. Morgan, Citigroup, HSBC, and RBC have capital markets units that now represent 25% or less of total revenue.

Those banks most exposed to trading, notably to fixed income trading, may not have the flexibility to wait for the eventual re-pricing and restructuring of the trading businesses and may need to exit before any restoration of market making profitability fully occurs. Moreover, several of these banks have highly profitable businesses that still generate reasonable returns, such as credit cards, mortgage origination, retail deposit, and middle market lending that may allow the bank to continue a war of attrition or last-man-standing strategy in their capital markets franchises.

Among capital markets banks, Goldman Sachs, Credit Suisse, and Barclays are most reliant on capital markets activities, with nearly 60% of revenue generated from trading and investment

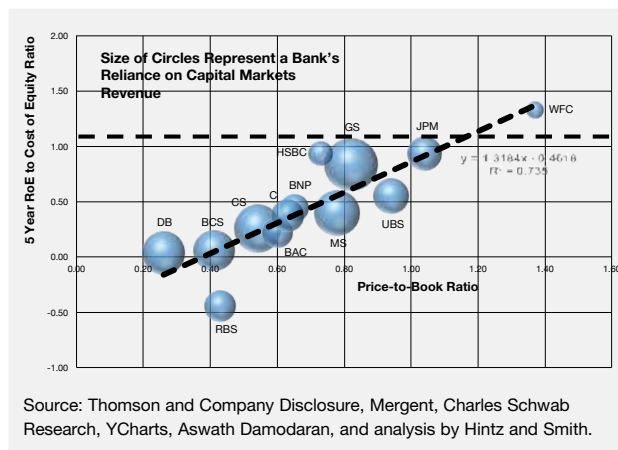


Figure 3 – Reliance on capital markets and RoE (price-to-book ratio versus 5-year average RoE/cost of equity ratio)

banking. Barclays and Deutsche Bank are most reliant on fixed income, currencies and commodities (FICC) businesses, with over 35% of total revenues from this source.

Execution market share

Trading is currently a significant drag on capital markets performance but it is difficult for any capital market bank to totally exit the business. Superior market share within trading units enhances operating leverage, allowing banks to efficiently spread the cost of trade execution technology and risk management systems across multiple trading desks and geographies.

Clients still demand secondary liquidity, consequently many executives within the industry believe that pricing in both fixed income and the institutional equity business will eventually adjust to absorb the cost of balance sheet constraints and regulatory changes on the market makers.

But the repricing of liquidity has not happened yet, partly because of the number of banks still seeking to maintain market share despite a limited future for them in the business.

And, it's not for everybody. Those banks with smaller market shares (and particularly those banks that have not invested in electronic execution systems and front-to-back office processing) run the risk of negative operating leverage as they attempt to adjust their revenue mix in capital markets.

There is a "just-right" level of market share in trading. If it is too large, the necessary large balance sheet commitment to a low return business will pull down total returns. But negative operating leverage can also be costly to players with too small a share.

J.P. Morgan, Citigroup, Barclays, and Deutsche Bank have the largest market shares in FICC. Those banks with large combined market positions in both equities and fixed income will have the option of sharing technology between trading units. Goldman Sachs and J.P. Morgan have the highest combined market share in fixed income and equities.

Geography

Capital market revenues are closely correlated with both global wealth and the global GDP. The most rapidly growing economies in the world are in the emerging markets, while the world's wealth is still largely located in the developed markets.

Banks with strong regional market shares will continue to profit from the growth of the emerging markets by capturing new issue flows and by participating in the maturation of Asian and other markets. But the new resolution rules and subsidiary capitalization and liquidity standards around the world are making it costlier and operationally challenging to maintain a global platform. A strategy of fully global banking no longer works for most capital markets banks. For many, a strategy of better-than-average market share in targeted high growth markets will likely be an acceptable alternative, particularly if the bank has some important legacy associations with these markets.

In developed markets, banks with strong wealth management or asset management franchises will be able to capture and retain profitable client relationships. These franchises can generate asset management revenues, incremental execution volume, and risk management services from capital markets units.

It is also useful to note that the large lending banks, with powerful syndicated loan market shares and those banks with strong operational businesses, such as wire transfer, cash management, and trade finance, can leverage these relationships with corporate clients to capture capital markets mandates.

Goldman Sachs and Morgan Stanley have the strongest global capital markets franchises. Deutsche, Barclays, and J.P. Morgan have powerful global fixed income franchises. The largest global loan syndicators are J.P. Morgan, Bank of America, Deutsche Banks, BNP-Paribas, Barclays, and Wells Fargo. Banks also have different strengths in regional markets – in the end they need to trade off market share, achievable margins, and regional comparative advantages to their best advantage. But that best advantage may be elusive for some, and the banks may be better off closing down operations in areas or product markets in which they won't be able to achieve satisfactory returns.

Expense control

Expense control remains key to achieving reasonable returns from capital markets. In the last eight years, banks have reduced head count significantly. Managing director pay has been reduced. Trading desks have been resized. But a “meat ax” approach to cost control is no longer effective. Despite success in controlling compensation ratios, non-compensation expenses remain stubbornly high. RoEs are still not adequate and there is little fat that is left to cut in front-office compensation.

In a survey of European investors, over seventy percent of respondents believed that banks have not been aggressive enough in reengineering their processes and two-thirds believe that banks have not invested enough in new technology [Broadridge and Institutional Investor (2015)].

This should not be a surprise to industry insiders. Few large banks have focused management attention on business unit redesign or back office rethinking. Expense allocations are negotiated between units rather than quantitatively determined. And in technology, costly legacy systems are allowed to run long past their expected lifespan, while a “not invented here” mentality frequently lead units to build rather than buy new applications.

The successful banks will focus on bottom line performance and count pennies. Industrial engineers, cost accountants, and technologists must rethink the business models and support systems of capital markets. While management teams need to consider participation in industry utilities and outsourcing of generic business activities.

Culture

Historically, the common idea of the culture of a commercial bank was that of a colorless, bureaucratic, risk-averse, hierarchical institution that was committed to reproducing in the future all that had worked in the past. The common portrait of an investment bank was of a small, flexible, opportunistic enterprise dedicated to making money for its owners by taking short-term risks and pursuing innovative products and solutions. Over time these two cultures were brought together through the many mergers that formed the large banks of today.

The sheer size of these organizations and their periodic need for large-scale layoffs over time weakened the loyalty employees have felt to their firms. A bureaucratic mindset led to weak risk management and financial controls, and performance expectations from the top to generate revenue led to an “eat

what you kill” culture at some organizations. And it was culture drift that facilitated the overaggressive activities that had to be paid for by loan write-offs, penalties, or litigation since the end of the financial crisis.

Regulators in the U.S. and Europe seem now to regard managers and employees of large banks as untrustworthy remnants of a corrupted subculture of big business. This view appears to be widely shared by the general public, and even among corporate and institutional clients and counterparties of large banks. This loss of reputation for integrity has been expensive to the banks in many ways, including loss of political support and weakened, less loyal and trusting client relationships.

According to Charles D. Ellis, a financial writer with a lifetime interest in what makes great firms what they are, those firms that rise to the top of their fields have distinctive cultural characteristics that stand out. These characteristics vary, but seem to share some important values that are inculcated down through their organizations by a chain of leaders that have devised and supported the values over a long period. These include a commitment to excellence in whatever they do, comparative disinterest in size per se, the importance of teamwork, and a mutual exchange of loyalty to employees by the firm, and to the firm by employees.

To return to a normal, sustainable business will require that banks get the economic model right, but it will also require a major cultural transformation. In global banking, this begins with a reputation for competence, excellence in service, and reliability. Much of this relates to the bank’s ability to process transactions efficiently, legally, and by the rules, with high importance devoted to enforcing ethical standards. For such a business, teamwork and organizational integrity is more important than the contributions of a few star individuals, no matter how entitled to stardom they may be. But individuals are important as team leaders, and need to be trained and motivated to be the best managers they can be. Good managers act as on-the-scene referees, spotting and stopping infractions before they occur.

RIGHT-SIZING

The search for a different business strategy must begin with the recognition that all banks will have to change and many will have to make massive changes. Rethinking the future has to start with realistic assessments of their upward potential from where they are.

One obstacle to many is size. Many banks have proven to be too big or complex to manage. Shrinking and simplifying seems a simple solution, but banks have been reluctant to pursue such a course, despite its obvious appeal. So far, markets have punished banks for their lack of effort with extremely low stock prices and EVAs. These performance factors will ultimately force boards of directors, or activist investors to consider more severe restructuring approaches.

So far only two banks have undertaken significant strategic moves to change their business mix:

In 2012, Morgan Stanley acquired the brokerage business of Citigroup (Smith Barney) as a strategic move that would divide the firm into two approximately equal parts, one a large, low-risk broker-dealer and other a leading investment bank subject to capital market volatility. However, Morgan Stanley has continued to struggle to demonstrate that this strategy is viable: in 2015 its price-to-book ratio was 0.73 and EVA -7.9%.¹³

UBS has sharply reduced its reliance on trading businesses while retaining a portion of its high margin investment banking revenues. The bank announced in 2014 that it was abandoning FICC proprietary trading and would limit market-making activities to the needs of its core clients, and would shift the bank’s strategic emphasis to asset management and wealth management. This was an easy call because UBS’ wealth management franchise was so vast, and it has paid off. In 2015, UBS stock was trading at 1.34 times book value, with a dividend yield of 3.6% that would increase further when the bank reached its near-term goal of a 50% dividend payout. Its EVA was +1.0% in 2015, though in 2014 it had fallen out of the top ten originators.

Some other European capital market banks have also said they would significantly shrink their trading businesses, though they do not have the extensive and profitable nonbanking business that UBS has to fall back on.

Citigroup has also followed this approach, having reduced its balance sheet assets by more than 20% since 2007, but so far without convincing the markets that its basic commitment to capital markets has changed meaningfully.

So far only UBS has convinced the markets that it has moved on to a more viable long-term strategy. The market leaders,

¹³ In July 2016, ValueAct, an activist hedge fund, acquired a 2% interest in Morgan Stanley with the idea of “working with management” to improve performance.

J.P. Morgan and Goldman Sachs continue to struggle to trade above book value and to post positive EVAs. As Exhibit 3 illustrates, all the rest have settled into price-to-book and EVA levels that are too low to be acceptable.

Some have considered selling their investment banking units, but there appears to be few willing buyers of such risky and constrained businesses, and, in any event, bank buyers would have to be approved by regulators, which may not be something they could count on.

SPIN-OFFS

The step that none of the banks have been prepared to take so far is to separate their businesses into commercial and investment banking units through a spin off operation in which present shareholders would own a stake in each business. Such an operation would be complicated, and almost certainly involve recapitalizing the investment bank so it could access credit markets on its own (probably requiring at least a Baa level bond rating). In 1994, American Express spun off its Lehman Brothers investment banking unit to its shareholders, after which the share prices of both units increased significantly. We believe that such an exercise could be successful for large capital market banks trading well below book value.

Investors would see the main, commercial banking activity being liberated from the hazards of risky capital market activities, and made more manageable as a single purpose enterprise. The capital markets unit, on the other hand, would be free to return to the more opportunistic and flexible ways of an investment bank. Each could perform according to its comparative advantages.

Regulators should also prefer the separate version of the two units: it would self-effect a breaking up of the banks that many regulators have preferred. In the case of Barclays, U.K. banking rules already require that it separate its U.K. retail business from its non-U.K. and wholesale activities in 2019.¹⁴ The two units would be separately capitalized and have separate boards of directors. The further transition into two separately owned companies by spinning off shares of the investment bank to the present shareholders of Barclays does not seem to be a large step.

When large, complex companies appear to languish for a time in a business strategy that seems to be falling short, investors lose confidence in management and the board to make any

sort of radical change. The companies don't like to admit that their longstanding business strategy has failed, or necessarily want to break up the business into smaller, less important units. Such changes may be what are needed but they can be risky and difficult to execute, so they get deferred.

For many years, General Electric struggled with the weight of an underperforming GE Capital Corp., its large finance subsidiary that was significantly affected by the crisis of 2008. Efforts at a gradual adjustment of the size and influence of GE Capital on GE by spinning off its consumer financing business (now called Synchrony, which it did in 2013), and selling a substantial portion of its real estate holdings and some other businesses (which it did in 2014 and subsequently), made little difference to GE's stock price. The message did not get through to investors until the announcement by chief executive Jeffrey Immelt on April 10, 2015 that GE would get rid of all but a few necessary customer-financing parts of GE Capital. The GE Capital share price quickly gained 10%, and outperformed both Honeywell and the S&P 500 over the following year. Clearly the market liked the idea of getting out of the dangerous finance businesses that contributed more problems than value for most of the prior eight years, and avoiding the slow, piecemeal approach management had earlier indicated it preferred.

In October 2016, Metropolitan Life, the largest U.S. life insurance company, announced that it would spin off to its shareholders 80% of its retail life insurance business, BrightHouse Financial, for strategic reason. BrightHouse is not only the firm's original whole life underwriting and sales business formed in 1868, it still represents about a third of MetLife's total assets under management.¹⁵

¹⁴ "Ringfencing" is a U.K. requirement for dividing the bank into two separately capitalized units, a domestic bank and a global and wholesale one, with significant restrictions on funding the global one with resources of the deposit insurance-protected domestic one.

¹⁵ Earlier in the year, MetLife won a lawsuit in federal court that reversed a Financial Stability Oversight Council ruling that it should be designated one of four "systemically important non-banks," making it subject to capital adequacy and other regulation by the Federal Reserve. The U.S. government has said it will appeal the ruling.

FACING REALITIES

No large capital market bank wants to renounce its integrated business model of operating in banking and financial markets. To spin off or rid itself of one of their primary revenue producing units is thought to be a step backward from where modern finance is headed. And, to separate a unit for spin off would involve a painful process of allocating debt, capital and personnel between existing units, and being sure that both the spun off unit and its former parent would be adequately capitalized and equipped with technology and other resources.

The reality is that the increased (and not yet fully applied) regulatory burden for G-SIBs may be too great for some banks to manage while attempting to remain an economically viable enterprise. The EVAs of the industry demonstrate the point; but it is really the markets' reaction to years of work-around efforts to return to "normal" that indicates the extent to which confidence has been lost in the integrated banking business model.

A few banks may be able to manage their way back to normal – by further reducing RWA and related exposures to a modest portion of the balance sheet, but others that trade at prices reflecting exceptionally low levels of market confidence will not be able to. In September 2016, the price-to-book ratio of Deutsche Bank had sunk to 0.31, Barclays' was at 0.43, Credit Suisse's was at 0.59, and Citigroup's and Bank of America's were just a little higher. At these levels a break up valuation of the banks' separate parts would surely indicate a post-spin off combined market value far greater than the banks' present market values.

In the case of Barclays, for example, its UK business (based on RWA) is only about one-third the size of its global capital markets unit. If we applied the price-to-book ratios of the peers of the two units proportionately (e.g., Lloyds and Royal Bank of Scotland in the U.K., and, say, Morgan Stanley globally) we get an estimate of what the combined market value of the two units might be. Using 1.1 for the U.K. unit and 80% of Morgan Stanley's price-to-book ratio, instead of Barclays' 0.43 ratio, the new combined market capitalization is more than doubled. From this amount must be subtracted some transactional expenses and provisions for additional capitalization of the capital markets unit, but the net result should still represent a potential increase in shareholder value from a spin-off of 50% to 70%. Such a potential increase in market value should appeal to shareholders (and activist investors), to regulators, and to credit rating agencies.

Some banks may be hoping that some substantial regulatory

relief will result from the election of Donald Trump as President of the U.S. It might. Mr. Trump said during his campaign that we "have to get rid of Dodd-Frank," it was holding back lending necessary for economic recovery because "the regulators are running the banks" [Schlesinger (2016)]. There is an argument for repealing Dodd-Frank based on the fact that capital and other requirements of Basel III and the Financial Stability Board, together with the stress-tests and other requirements of the Federal Reserve, provide adequately for bank safety, and that Dodd-Frank (passed before these other measures were in place) only extends the burden and cost of regulation to such an extent as to threaten the long-term viability of the banks. Whether Mr. Trump could garner enough votes in the Senate to repeal the law is questionable, however. If not, he might be able to amend the law significantly and to replace members of the board of the Federal Reserve with others more amenable to a lighter touch. It will probably take a year or more before we know what the outcome is, but even if Dodd-Frank were repealed, the ensuing regulatory relief may still not be enough to allow banks to escape having to address the strategic realities we have discussed.

Indeed, financial historians know that the current state of the global banking industry is the fourth major wave of transition since the 1930s. These transitions have been caused by regulatory actions and the market forces resulting from them. Not all of the regulatory changes have been wise or efficient, but nevertheless they have forced an element of Schumpeter's "creative destruction" on the industry. After Glass Steagall, firms like Morgan Stanley were formed by spin offs from powerful commercial banks. In the 1980s, regulators intervened to prevent banks from failing after a competitive lending binge, forcing many mergers among large players. After the repeal of Glass Steagall in 1999, many more mergers occurred as banks sought to keep up with Citigroup, the new colossus of the capital markets. The present wave of reorganization, forced after 2008 by the globally concerted effort to de-risk the banking industry, is only in its beginning stage.

What historians also know is that through each of these transitional waves, the leading players change almost totally. Of the top 80 banks and investment banks in business in 1990, all but two or three have disappeared into mergers or have failed. During this period, however, the market capitalization of the world's tradable debt and equity securities grew from U.S.\$54 trillion to U.S.\$300 trillion [McKinsey Global Institute (2016)]. Capital markets continue to grow and expand globally, but the principal competitors in the market turn over continually.

The current transition will take several years to complete.

During this time different ideas and structures for the industry may be tested but the markets themselves will determine the winners and thus the next world order in the global banking industry.

REFERENCES

- Berry, J. M., 1988, "Greenspan urges House to give banks authority to enter new business lines," Washington Post, May 13, available at <http://wapo.st/2eDnOoS>
- Broadridge and Institutional Investor, 2015, "Restructuring for profitability: analysts predict opportunities and challenges for global capital markets institutions through 2020," December, available at <http://bit.ly/2fCT2kf>
- IMF, 2010, World Economic Outlook, International Monetary Fund, April
- IMF, 2015, World Economic Outlook, International Monetary Fund, July
- Lascelles, D., 1982, "Bank admits terrible mistake," Financial Times, August 4
- Mack, J. J., 2009, CNBC.com, September 11
- McKinsey Global Institute, 2016, "Mapping Global Markets," McKinsey & Co., May 17
- Roberts, D., 2014, "Wall Street deregulation pushed by Clinton advisers, documents reveal," The Guardian, April 19, available <http://bit.ly/2fCl2DD>
- Schlesinger, J. M., 2016, "Donald Trump took aim at Dodd Frank on the stump," The Wall Street Journal, Nov. 9, <http://on.wsj.com/2fmRiuS>
- Scott., H., and J. Gulliver, 2016, "The Fed's stress tests needs to be transparent," The Wall Street Journal, September 17
- Smith, R. C., 1993, Comeback – the restoration of American banking power, Harvard Business School Press
- Wall Street Journal, 2013, "The government won on financial reform: the financial system is more regulated than ever, but also no safer," September 22, available at <http://on.wsj.com/2eDwoE0>

Policy Response Asymmetry and the Increasing Risks From Rising Government Debt Level

Blu Putnam – Chief Economist, CME Group¹

Erik Norland – Senior Economist, CME Group

Abstract

The pendulum is shifting away from a dependence on monetary policy and a switch to greater use of fiscal policy in the U.S., and the U.K., and Europe may result. At the same time, there is a robust debate about how fast mature industrial economies can grow, giving their aging demographic challenges. If significantly more rapid real economic growth can be achieved with fiscal stimulus or with continued monetary accommodation, then the currently high public and private debt loads may be manageable. This research argues that the ability of fiscal and monetary policy to push real economic growth rates higher is markedly reduced in periods of even modest economic growth, as compared to recession periods. And, stimulative fiscal policies that aim for unattainable economic growth targets risk pushing the national debt over the tipping point from which debt loads become destabilizing to the economy. That is, once economies have exited a recession and are growing again, stimulative policies may contain more downside risks than upside growth potential once the feedback loops from

rising debt are considered. We live in an asymmetric world full of unintended consequences and powerful indirect effects. Our analysis strongly suggests that since unconventional monetary policy largely impacted asset prices and not real GDP growth, assets may be entering a period of greater risk than historical measures of volatility might suggest, as unconventional monetary policy is ended. And, the starting point for fiscal expansion already embodies high debt-to-GDP ratios. If the expansionary fiscal policies fail to generate the hoped-for real economic growth, then the unanticipated increases in inflation and interest rates may bring difficult challenges with debt management to the fore.

¹ All examples in this report are hypothetical interpretations of situations and are used for explanation purposes only. The views in this report reflect solely those of the authors and not necessarily those of CME Group or its affiliated institutions. This report and the information herein should not be considered investment advice or the results of actual market experience.

Presidents, Prime Ministers, politicians of all stripes, regulators, and central bankers, just to name a few, are often focused on what policies they can put in place to encourage more rapid economic growth. Our analysis suggests that increasingly, policy-makers are having to confront the uncomfortable reality of asymmetric policy responses. Specifically, the economic responses to fiscal and monetary policy strongly suggests that for economies in recession, there are a number of policy options that can assist in bringing the economy back into positive economic growth. Unfortunately, the analysis also argues that there are meaningful limits as to the extent to which real economic growth can be encouraged to expand at a more rapid pace once a reasonably sustainable activity level has been reached. And, perhaps equally important, from a starting point of modest economic growth path, policy mistakes can easily lead to sub-par economic growth or even recessions. This means that there may be more downside risks than upside potential for major fiscal or monetary policy shifts made during periods of economic growth, despite the good intentions to push economic growth ever higher. And finally, having a realistic perspective of what a sustainable growth path might look like is essential, since an overly optimistic view of the growth potential may be a recipe for policy mistakes leading to debt expansion, which in turn triggers damaging economic consequences.

What we are talking about are the non-linear economic responses to policy decisions depending on the starting point. If the initial conditions are recessionary, then there is considerable potential for policy actions to improve economic growth. If the initial conditions are around or close to a sustainable economic activity environment, then policy actions are highly limited in their ability to encourage more rapid economic growth and the possibility of policy mistakes leading to sub-par economic growth are quite significant. To appreciate our general case for the asymmetry of policy outcomes for economic growth and critical importance of the initial conditions, we will work through the theory related to a set of selected and relevant examples from (a) fiscal policy and then (b) monetary policy.

Our analysis will start with fiscal policy, examining tax rate reductions, government spending increases, and the destabilization potential if debt levels grow to high too fast. With regards to monetary policy, we will study unconventional monetary policy, dividing the analysis into the immediate recession period and the later growth recovery phase. Our concluding section brings together the lessons from fiscal and monetary policy in recessions versus growth periods. Our objective is to lay the intellectual framework for a general non-linear and asymmetric theory of the economic growth responses to monetary and fiscal policy changes that depends critically on the

initial conditions. The implications for policy-makers are significant. The idea that “more of a good thing is always an even better thing” simply does not stand the test of analysis. And, moreover, during periods of even modest economic growth, policy shifts are not likely to have much impact on economic activity while they carry meaningful risks of causing dislocations and weakening economic activity. It is an asymmetric world full of unintended consequences and powerful indirect effects. Linear, or as some might say, “flat-earth thinking,” can be especially dangerous for economic policy.

FISCAL POLICY

As noted in the introduction, our analysis of fiscal policy will focus first on the economic impact of tax rate reductions and spending increases. Then, we will consider the national debt and how rising debt to GDP ratios can influence economic activity.

How likely is it that reductions in tax rates will stimulate economic growth?

Any discussion of the economic impact of tax rate reductions must consider what became known in the U.S. as the “Reagan Revolution.” The Reagan era was propelled in no small way by an economic idea called the “Laffer Curve,” named after Art Laffer² who is considered the father of supply-side economics. The “Laffer Curve” was a highly intuitive and appealing theory (Figure 1).

The “Laffer Curve” argued that there was an optimal marginal tax rate that could produce the most revenue for the government. If the marginal tax rate on the highest and last unit of income was set too high, economic growth was damaged and tax revenues would not be as high as they could be. It was also possible to set the marginal tax rate too low, so that government revenues from tax would fall as economic growth simply did not respond sufficiently to offset the sharply lower tax rates.

As is the case with all economic theories, they come with some heroic, embedded, and usually ignored assumptions that can turn out to make a very big difference when analyzing the economic outcomes from a policy change based on the theory.³ In the case of the Laffer Curve, the critical assumption

2 For the interested reader, an updated treatise on the Laffer Curve and related ideas: Canto, V. A., D. H. Joines, and A. B. Laffer, 2014, *Foundations of supply-side economics: theory and evidence*, Academic Press.

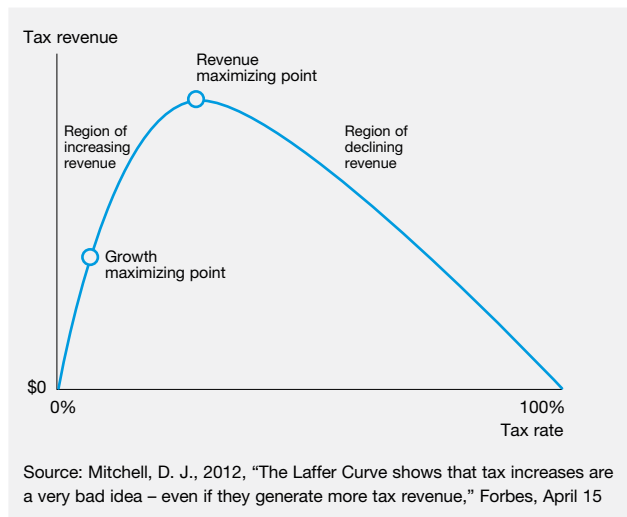


Figure 1 – The Laffer Curve

is that business and personal investment, savings, and consumption decisions will be primarily influenced by the marginal tax rate they pay on the last unit of income. This is a perfectly reasonable assumption; however, what it means is that if the tax code is rife with special exceptions and loopholes, then the marginal tax rate may not matter all that much for economic activity. Instead, the tax loopholes and special rules will dominate decisions about investment, savings, and consumption.

Given the complexity of the tax code in most countries, especially the U.S., the relevance of the “Laffer Curve” may be quite limited unless tax rate reductions are also accompanied with major tax reform that simplifies the tax code and puts the focus back on marginal tax rates and not on credits for medical expenses, deduction for mortgage interest, different tax rules for dividends and interest, special credits for certain types of investments and not others, special taxes on imports versus exports, differences in how wages or capital gains are taxed, differences in tax rates depending on investment holding periods, charitable deductions, etc., etc., and the list goes on for tens of thousands of pages, at least in the U.S.

During the 1980s, when Ronald Reagan was President, there were two tax cuts and significant tax reform. During the years 1983-1989, the growth years of the Reagan era, U.S. real GDP averaged 4.45% per year, a little bit higher than 4.26% of the Carter growth years, and a little lower than the 4.77% of the Kennedy-Johnson growth years. Government revenues remained in a tight range, 18.50% of GDP in 1979 (last year of economic growth before the recession of 1980-1982) to

18.43% in 1989. Unfortunately, there was little to no accompanying discipline on the spending side by the U.S. federal government, and the national debt rose from 33% of GDP in Q1-1981, as Mr. Reagan was coming into office, to 54.7% in Q1-1989 as he was leaving. By the end of President George Herbert Walker Bush’s Administration in Q1-1993, Federal debt had reached 68.3% of GDP, despite a modest (and unpopular) hike in the top marginal tax rate from 28% to 33%. We note that with lots of caveats and torturing of the data, some analysts have argued that the “Laffer Curve” worked as expected, but to our eyes the results on tax revenue and economic growth were inconclusive and the impact on the national debt was decidedly negative.

In addition, we would caution against parallels from 1981-1988 with 2017-2024, simply because the starting points differ and the likelihood of future tax simplification is a very large question mark. From a marginal tax rate perspective, in two steps the Reagan tax cuts took the highest marginal rate down from 70% in 1979 to 28% by 1986. This was a huge change. For 2017, the starting point is 39.6% as the highest federal marginal tax rate on personal income and 35% on corporate income. The tax rate reductions are simply not going to be as large as in the Reagan era, with or without tax simplification, hence the impact on real GDP may be smaller too.

As for economic growth, the 1983-1989 period benefited in a large way from the bounce back after the recession of 1980-82. To curb inflation then running above 10%, the Federal Reserve (Fed) had pushed short-term market rates toward 20%, long-term Treasury bonds reached yields of 14%, and the unemployment rate had jumped to over 10%. In the post-recession period, inflation subsided rapidly, bond yields and short-term interest rate fell sharply, and unemployment declined. One could argue that the bounce back from the recession was going to happen regardless of tax policy changes, especially since it was well underway before some of the tax changes were enacted into law or even known. For the case of the U.S. in 2017-2024, the economy has been growing modestly yet at a relatively steady pace of around 2% real GDP rate since 2010, while the unemployment rate has declined from 10% to below 5%, and short-term market interest rates were near zero and set to rise, albeit slowly. This is a vastly different starting point from the Reagan era.

3 The role of hidden assumptions in economic theory is a common source of erroneous analysis. See: Putnam, B. H., G. McDannel, and V. Shah, 2016, “Digital finance: at the cusp of revolutionizing portfolio optimization and risk assessment systems,” *Journal of Financial Transformation* 45, 35-42.

The size of the tax cuts discussed by the new U.S. Administration in Washington more closely resemble the relatively modest tax changes at the beginning of George W. Bush's Administration in 2001. These changes brought the top rate down from 39.6% to 35% and in 2003 the Administration convinced Congress to cut the dividend tax as well. Productivity growth was solid during the Bush Administration but overall GDP growth was weak and Federal debt expanded much faster than the economy as a whole. In Q1-2001 when Mr. Bush arrived in the White House the national debt was 49.4% of GDP and in Q1-2009 when he left office it had grown to 73.7%.

And there is another challenge for the 2017-2024 period – namely demographics. The labor force was growing at a healthy pace and population was relatively young during the 1981-1989 period. By 2016, labor force growth was down to less than 1% per year and retiring baby boomers represented a major cohort of the population likely to be cutting spending in their retirement years compared to their expansive consumption habits in their younger and highly productive working years.

Finally, there is one other aspect of the tax policy that we would like to consider in terms of estimating the economic impact of a reduction in tax rates. If the tax cuts primarily benefit wealthier individuals, as opposed to typical hourly wage-earners, then there can be a big difference in the savings and consumption outcome. Tax cuts for the relatively high earners produce fewer increases in consumption and more increases in savings. This means that if there is a tilt in the tax cuts favoring high incomes, then the impact can work to stimulate equity prices while not doing much for economic growth. One may remember many analysts (present company excluded) in the U.S., Europe, and Japan cheering the massive oil price decline in late 2014 as likely to work like a tax cut and lead to stronger economic growth. The extra growth never happened because a meaningful chunk of the fuel cost reductions to households went into savings or to pay down debt, instead of discretionary consumption. So, while there was a temporary boost to consumer discretionary spending in the months following the collapse of oil prices, it was just enough to offset the negative GDP impact of lower oil prices on the energy industry itself.

The bottom line is that there will be no following winds for U.S. tax reform in the 2017-2024 period like there were in 1981-1988. Caution is advised on just how aggressive to expect the real GDP impact to be from a given tax cut, given the initial conditions.

Will government spending increases lead to more real economic growth?

Our discussion of government spending starts with the arithmetic of national accounting and nominal GDP. Specifically, nominal gross domestic product is the sum of consumption (C) plus investment (I), including changes in inventories, plus government spending (G) plus the current account balance, which consists of exports (X) of goods and services minus imports (M). That is,

$$\text{Nominal GDP} = C + I + G + (X - M)$$

Other things being equal, any increase in government spending will work to increase nominal GDP. That is just arithmetic. Whether an increase in government spending works to advance real GDP growth, just raises inflation, or even increases imports to the detriment of nominal GDP, depends on a myriad of additional considerations. And, with all the feedback loops between markets and the economy, other things are never equal in predictive economic analysis.

If the starting point is a recession, then the impact on real GDP can be quite meaningful – the Keynesian view. And, John Maynard Keynes was not even concerned as to what the spending was – that is, wasteful spending was just as effective in the short-term at stimulating the economy as investment spending. Keynes viewed the Great Depression of the 1930s from the perspective that economies around the world were in disequilibrium.⁴ His solution was to suggest a role for government to step in, start spending, and not worry about the consequences for rising national debt until the economy was growing again. The key point was for government to spend while consumers and businesses were cutting back in fear of the economic situation deteriorating even further. Government spending could short-circuit the vicious cycle of fear gripping the economy. One can recall the famous quote from U.S. President Franklin D. Roosevelt⁵: “The only thing we have to fear is fear itself.”

If the starting point for increased government spending is an economy that is already growing, even at a modest pace, then the analysis gets considerably more complex. The type of new spending will matter, and positive impacts may come with a long time lag. Whether there is a long-term positive impact on real GDP is critically linked to whether the new spending can be expected to increase labor productivity.

4 Keynes, J. M., 1936, *The general theory of employment, interest, and money*, London: Macmillan & Co.

5 Roosevelt, F. D., 1933, U.S. Presidential Inaugural Address, March 4

If we return to arithmetic for the moment, one can look at real GDP growth as the sum of labor productivity growth plus labor force growth. Generally speaking, to have an expectation of raising labor productivity, one would be looking at whether the new spending was an investment in capital, such as improving the country's infrastructure, or whether the new spending was mostly subsidies or payments to individuals that might increase consumption but not contribute to improving labor productivity. And, since labor force growth is dependent on demographic patterns that evolve very slowly over time, there is no influence from fiscal policy.

Let us turn to labor productivity first and look at the case of China as an example. The Chinese focus on infrastructure building in the 1980s, 1990s, and early 2000s, provides a powerful explanation for why the country was able to modernize as rapidly as it did and achieve stellar real GDP growth rates for as long as did. Once the country reached a relatively mature degree of modernization, however, the gains to labor productivity from further infrastructure spending hit the point of diminishing returns and the economy began to decelerate toward lower real GDP growth rates. The U.S. spending on its interstate highway system in the 1950s and 1960s is another example of the type of investment spending that contributed to rising labor productivity and higher real GDP growth.

Military spending is a controversial category. Some military spending can influence innovation, such as advancing aerospace or satellite technology, and thus contribute to long-run gains in labor productivity. The vast amount of military spending from buying planes to building more missiles or a new aircraft carrier to adding soldiers are not likely to influence labor productivity even if they add to nominal GDP.

Even if the increased government spending has the potential to increase labor productivity through capital additions (infrastructure) or incentives for innovation and technological improvements, there are likely to be time lags. Infrastructure projects can take a long time to complete and then the knock-on indirect economic impacts can take even more years. And, innovation dances to its own tune, making it nearly impossible to attribute gains in labor productivity to specific government spending that may have occurred years before.

Consequently, when increased government spending hits an economy that is already growing there can be minor increases in labor utilization, so some increases in real GDP may be possible. Unfortunately, most of the increase in nominal GDP from the increased spending hitting an already growing economy is likely to lead to either increased imports or inflation. Other

things are never equal in economics, however, so the any increased imports or higher inflation may create opposing market reactions, particularly from higher interest rates and/or a restrictive monetary policy, as well as from exchange rates reacting to the interplay of trade and capital flows. In short, market prices move in response to the policy shift to increase spending during times of economic growth, but real GDP may only accelerate very little, and that will come with a long time lag.

With regards to labor force growth, it is all about demographic patterns, and it is the work force between the highly productive ages of 25 to 55 that really matters. If the prime age labor force is growing very slowly, as in the U.S., or not at all, as in Japan, real GDP growth expectations will need to be adjusted downwards, when compared to previous decades characterized by much faster labor force growth.

For the long-term analysis of labor force growth, there are three major factors that must be monitored: birth rate, rural to urban migration, and immigration.

For example, China's one-child policy to control population growth worked perfectly to reduce the population growth to zero. The unintended consequence a few decades later was the slowing of the labor force growth and then the aging of the population with steady increases in the over-65 cohort. Even ending the one-child policy will make little difference for a few decades. It takes 25 years to make a 25-year-old. And Chinese young adults that grew up in one-child families may well prefer to have only one child as they consider parenthood.

Rural to urban migration is also a common pattern as a country industrializes. And, since industrial output per person is generally higher than agricultural output in less developed countries, there are typically labor productivity gains associated with rural to urban migration as a country develops. This occurred in the U.S. in the early 20th century. It was a major factor in Japan and in the former Soviet Union in the 1950s and 1960s. And, it is still a factor in China, where upwards of 15 million people move from rural to urban communities each year. When this migration ends in the 2020s, China will face its most serious challenge to its economic growth model, not unlike the former Soviet Union did in the 1970s and 1980s.

Finally, immigration can be a source of an expanding labor force. Immigrants are often of prime working age, seeking a better life in a new country. The U.S. has been a major beneficiary of real GDP growth through embracing immigration in its past. Equally, Australia has enhanced its real GDP prospects through immigration. Japan eschews immigration and has had

to adjust to a reality of no population growth and an aging population much sooner than immigration-oriented countries.

The bottom line here, though, is that demographic forces are very slow moving and not at all influenced by fiscal policy. They are a given that must be considered as part of evaluating how rapidly an economy can grow.

At what point does a rising debt/GDP ratio destabilize an economy?

Any positive impact on real GDP from tax rate reductions or spending increases may happen with a lag, so the national debt as a percent of GDP is likely to rise in the short-term. The long-term effect on the debt-to-GDP ratio depends on how the policy mix influences real GDP, imports, and inflation. With the debt-to-GDP ratio heading higher from more stimulative fiscal policies until the long-term influences are realized, it raises the question as to whether the debt level itself could have destabilizing effects on the economy.

Specifically, we would argue that at some level of national debt-to-GDP, the debt overhang becomes a negative factor for the economy, as higher interest rates reflect rising inflation expectations and rising interest expense raises the risk of debt. Indeed, there is the possibility of capital outflows and less willingness by international investors to fund the rising debt.

From an economic indicator perspective, though, it is not the ratio of debt-to-GDP that makes the difference. The real issue is the long-run expected cash flow relative to debt service. A robust economy with solid growth prospects can handle more debt and a higher debt-to-GDP ratio than a slow-growing economy prone to setbacks or bouts of inflation. Economies that are adding debt from tax cuts and spending increases will eventually hit the point of destabilization, but it depends greatly on where they started. During the Reagan era, the starting point for the U.S. was a 30% debt-to-GDP ratio and relatively high growth expectations. In the U.S. in 2017, the starting point is a debt-to-GDP ratio of over 100% and a slower expected growth path due to demographic challenges not faced in the 1980s.

Higher levels of public and private sector debt inevitably put pressure on central bankers to maintain interest rates lower than inflation might suggest. The only way in which heavy public and private debt burdens can be financed is through relatively low interest rates or extremely fast growth. For mature economies, the latter is not usually an option, so the former becomes a priority.

Take the case of Japan. Japan reached 270% of total public and private sector debt-to-GDP in 1990, and rates hit zero by

1998 and the economy has continued to lever up with debt-to-GDP reaching 400%. The leveraging up has had little appreciable impact upon GDP growth, which has been mired at around 1% (in the good years) owing in large part to Japan's stagnant demographics.

What all this means is that there is no magic tipping point before the debt-to-GDP ratio starts to destabilize an economy. At some point, market participants around the world grow wary of the debt loads and the debt service requirements. Moreover, the process never happens smoothly. Everything seems to be OK, even if the danger from too much debt is well known and regularly debated, and then an event or a catalyst occurs that abruptly changes market behavior and the vicious cycle begins. Once it starts, it is like a snow ball rolling down a hill – getting bigger and bigger and much harder to stop. One generalization that we can make, however, is that the higher the level of debt, the lower the level of interest rates necessary to tip the economy into a recession. This is why it might be difficult for the Fed to put rates up to 3% by the end of the decade as it suggests that it will do in its “dot plot.”

MONETARY POLICY

The Great Recession of 2008-2009, triggered central bankers to launch an impressive experiment in unconventional monetary policy, which coincided with a pre-recession pattern of financial regulation putting more and more emphasis on capital adequacy as the method of controlling financial risks. And, it is important to note that the Great Recession only increased the desire of regulators to impose tighter capital adequacy rules, which we argue had an impact on how unconventional monetary policy worked.

In this examination of monetary policy, we want to focus on what we have (or have not) learned from the experiments with unconventional monetary policy. After all, while equity prices in the U.S., Europe, and Japan have soared from the low points during the Great Recession, real GDP growth has been quite modest by the standards of previous economic recoveries.

How did quantitative easing (QE) actually work?

To understand the impact of central bank balance sheet expansion on economies and asset prices one has to divide the 2008-2016 period into two parts.⁶ First, there was the

⁶ Putnam, B. H., 2013, “Essential concepts necessary to consider when evaluating the efficacy of quantitative easing,” *Review of Financial Economics* 22:1, 1-7

immediate central bank response to the financial crisis triggered in September 2008 by the way the U.S. authorities mis-handled the extremely messy bankruptcy of Lehman Brothers and the bailout of AIG. This initial period was very different in character and in policies to the later experiments in monetary policy by central banks once the Great Recession had ended and growth had resumed.

Starting with the financial crisis period, the U.S. Federal Reserve (Fed) and the European Central Bank (ECB) took different paths. In Q4-2008, the Fed bought U.S.\$1 trillion of distressed assets (aka, toxic waste) and removed them from bank balance sheets. The ECB chose a different approach and made available €1 trillion of low rate, term liquidity loans, which effectively backstopped the European banking system against a liquidity crisis and runs on the banking system. In hindsight, we know that the Fed’s approach, by removing the overhang of distressed assets, allowed for a much faster and more robust recovery by U.S. banks than the ECB’s approach of emergency liquidity loans. The ECB left the distressed assets on the books, including sovereign debt of very weak countries. This meant that the weakest banks had to be divided into “good” and “bad” banks, with governments bailing out the “bad” banks. And still in 2017, some of the overhang still exists, especially in the weaker banking systems of southern Europe, which has worked to delay a more robust economic recovery in Europe compared to the U.S.

The key takeaway is that it matters how the banking system is backed-stopped. If the central bank serves only as the lender of last resort, the liquidity crisis is prevented and there is no downward spiral into another Great Depression. But by removing the overhang of distressed assets from the financial system, the Fed went a step further and removed a critical barrier to a more rapid recovery. Credit goes to Fed Chair Benjamin Bernanke for appreciating that getting to the source of the problem (i.e., removing an overhang of bad assets) was as important to the recovery as just preventing a liquidity crisis (i.e., lender of last resort).

The second phase of unconventional monetary policy occurred after the recession had ended and growth had resumed. It is highly unusual for central bankers to add stimulus once an economy is growing again. Usually, once growth has resumed, the debate turns to how soon should the emergency monetary accommodation of the recession period be withdrawn. Two years into the economic recovery, however, the Bernanke-led Fed embarked on new rounds of asset purchases with the intent of encouraging a more rapid economic expansion. In this and subsequent rounds of QE, the Fed purchased only

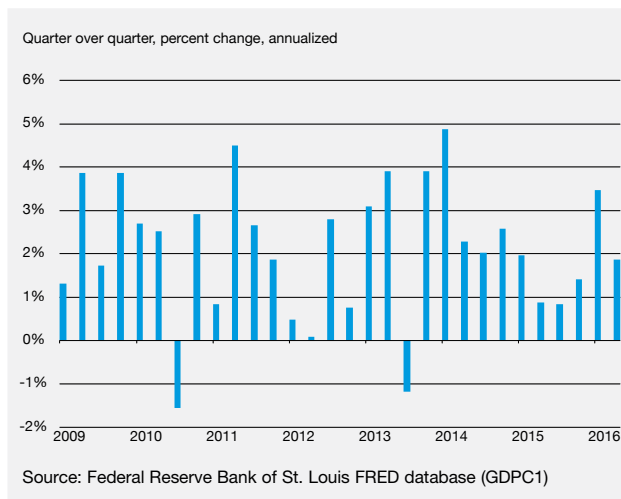


Figure 2 – U.S. GDP growth, Q3-2009-Q4/2016

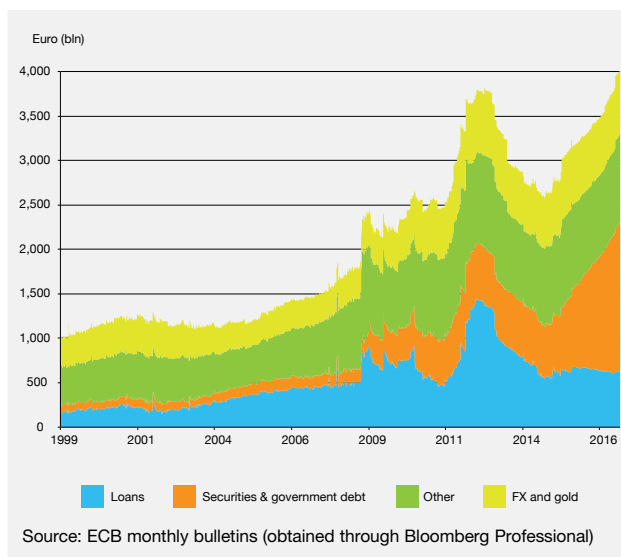


Figure 3 – ECB assets

high quality securities – U.S. Treasury securities and mortgage-backed securities.

If the criteria for success was the stated objective of more rapid economic growth, then this part of the Bernanke QE experiment was a total failure. See Figure 2 showing U.S. real GDP growth for 2010-2016 and see if you can find any difference in the pattern before or after QE. There was no real GDP response. If, however, the criteria were the intermediate impact on asset prices, then one can definitely see the response, in both government bond yields and equity indices. In short, the

Fed's experiment with QE in times of economic recovery went straight to asset prices and not to real GDP growth or even inflation.

The situation in Europe was more complex than in the U.S. because of the overhang of bad debt, especially sovereign debt from the weaker European Union (E.U.) countries. The original source of this problem was the Maastricht Treaty that created the single currency. A single currency for all of the E.U. would be a huge benefit to the stronger countries, such as Germany. The economic strength of Germany had led to currency appreciation and served as a restraint on exports to the rest of Europe, especially southern Europe. A single currency would remove the potential for exchange rates to adjust within and among Euro-Zone member countries. To provide an incentive for weaker countries to agree to join the Euro-Zone, the compromise was that the E.U. financial regulators would treat all sovereign debt as the same high quality credit risk, whether it was German government debt or Italian government debt or Greek government debt. Once this sovereign debt equivalence for capital adequacy regulations was agreed, there was a lending spree to take advantage of the higher rates offered by weaker sovereign debt issuers. This allowed bond yields to converge and the weaker countries to take advantage of lower capital costs than their inherent riskiness would suggest was reasonable. The chickens came home to roost with the Greek debt crisis of 2011, which also impacted Ireland, Portugal, Spain, and Italy.

Again, instead of dealing with the problem of the overhang of weak sovereign debt on bank balance sheets, the E.U. provided the weakest countries with some much needed funds in exchange for severe austerity measures that further drove down GDP and failed to contain debt ratios. The E.U. also decided for a stricter round of bank stress-tests and assigned the task to the ECB. Note that previous stress-tests were conducted by the E.U. and were largely viewed as publicity stunts, since everyone, weak or strong passed the tests. As the time (October 2014) for the ECB-conducted stress-tests neared, banks cleaned up their balance sheets as best they could. They focused on paying back the emergency liquidity loans to the ECB, since it was felt that these loans were a sign of weakness and could result in a bad stress-test credit score. The result was the ECB's balance sheet declined, and the credit markets stopped functioning properly while banks paid back the liquidity loans. This was an unintended "own goal" by the ECB, which continued to view the problem as one of capital adequacy instead of focusing on removing the overhang of bad debt from the system.

How did negative rates work?

After the stress-tests were over, the ECB changed direction and embraced Fed-style QE, purchasing only high-quality securities, and later introduced negative interest rates on bank deposits held at the central bank. The negative rate policy was controversial from its introduction. The idea was to provide an incentive for banks to take their deposits at the central bank and lend them out. Of course, at a time of stringent capital ratio rules, this was not possible. So, the main impact was to hurt bank profitability, since the ability of banks to pass their increased costs to their own depositors was extremely limited. If you are keeping score, this was another "own goal."

Worth highlighting in this discussion is the non-linearity in economic responses to negative rates. There is significant asymmetry in financial markets regarding responses to interest rates as they approach zero. Behavioral finance has shown that most investors get considerably less satisfaction from gains compared to their dislike of losses. The tax code in most countries focuses on taxing gains and strictly limits the deductibility of losses and limits loss carry-forward provisions. Many labor contracts and corporate incentive programs have zero bounds. It is relatively easy for corporations to let wage growth lag inflation, but nearly impossible to cut nominal wages should deflation occur. Executive stock options become worthless once the stock price drops below the strike price with little prospects of recovery. The reality is that lower interest rates become increasingly less effective as a tool for economic stimulus as they approach zero, and they may actually harm the economy if they go negative by hurting bank profits, by reducing banks' capability to lend.

As a final note on monetary policy, while we have focused on the Fed and the ECB, it is worth mentioning that the Bank of Japan (BoJ) did not participate in round one of QE immediately after the financial crisis started, as its banks were in better shape. The BoJ did eventually join the QE party after Prime Minister Abe was elected in late 2012 on a platform of getting growth going again. One of his three arrows for igniting economic growth was Fed-style asset purchases, and the BoJ balance sheet expanded in an explosive manner. The initial results hit the currency, a weaker yen. A weaker currency did restore Japan to positive nominal GDP growth but the pace of the expansion has been modest and inadequate to reduce Japan's debt burden. So, the BoJ added equity Exchange Traded Funds (ETFs) to its buying list, and managed to goose equity prices even higher. The BoJ also tried negative rates, with the same impact as the ECB – weakening bank profits, and disrupting the functioning of credit markets. The BoJ then altered its QE program to target fixing the 10-year government bond yield at zero.

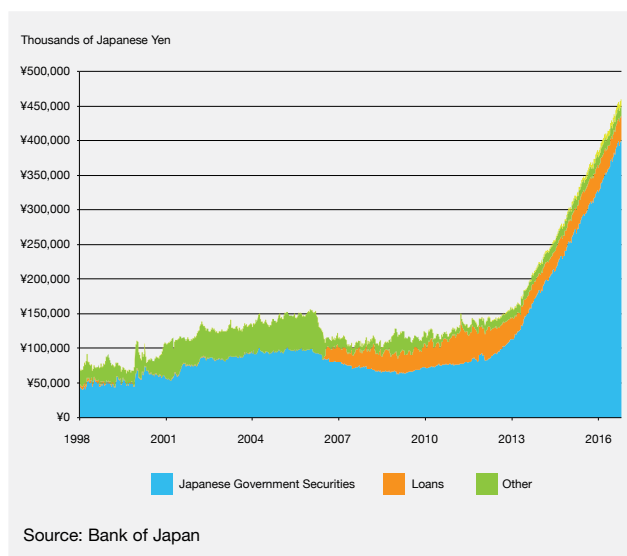


Figure 4 – Bank of Japan assets

GENERAL ASYMMETRIC THEORY OF ECONOMIC RESPONSES TO FISCAL AND MONETARY POLICY SHIFTS

Our examination of fiscal and monetary policy leads us to a set of observations, of which one common theme is that the analysis of economic responses to policy shifts is highly dependent on whether the starting point is a recession or whether the economy is growing, even just modestly. Our summary includes the following observations:

- The effectiveness of tax rate reductions to increase long-run, sustainable real GDP growth is dependent on tax simplification, such that marginal tax rates and not loopholes are the main incentives influencing consumption, savings, and investment decisions.
- The short-term effectiveness of government spending to increase real GDP is highly dependent on the state of the economy, such that policies that can help exit a recession do not work to increase growth once the recovery has begun.
- Moreover, the long-term effectiveness of government spending to increase real GDP is highly dependent on whether the increased spending can work to accelerate labor productivity growth; otherwise the impact is likely to lead to more imports and/or more inflation rather than more real GDP growth.

- Regarding monetary policy, in a crisis, central bank purchases of distressed assets can accelerate the recovery of the banking system and the economy compared to just “lender of last resort” approaches.
- In times of economic growth, central bank asset buying programs aimed only at high-quality securities largely impact asset prices and not economic activity.
- Economic responses to interest rate policy is highly non-linear, as rates approach zero, or even turn negative, they may actually hurt economic activity by damaging bank profits, which constrains bank lending and disrupts the efficient working of credit markets.
- Total levels of private and public sector debt play a significant role in determining at what level of interest rates the economy will become vulnerable to a recession. Generally speaking, the higher the level of leverage in the economy, the lower interest rates need to be in order to sustain growth and the less impactful fiscal and monetary stimulus will be.

The asymmetry of economic responses to fiscal and monetary policy depending on the initial conditions in the economy is not a new idea. Nevertheless, policy response asymmetry may be a more important consideration both for policymakers and for risk managers in the current environment because the sustainable growth path for mature industrial countries is much slower than it once was due primarily to challenging demographic trends. Policy attempts to push an economy toward faster than reasonable growth rates may well end in tears because the debt levels expand to a degree that leads to economic instability. In short, the probability of damaging policy mistakes rises as an economy approaches its sustainable growth path. For aging economies with little labor force growth, the ability to service debt payments has to confront the slower sustainable growth path. If policies designed to push real growth higher instead result in rising inflation and rising debt loads, then the higher interest rates that accompany the higher inflation may trigger a viscous cycle of debt default.

Of course, the warning over too much debt has been heard before. Too much debt has bitten many a developing country, and more recently some mature European ones. As debt levels rise relative to the cash flow of the economy to support them, rising interest rates will take a greater toll and do it faster. This is why the starting point is so important. The U.S., Europe, and Japan all experienced near-zero short-term interests in the 2010-2016 post-recession economic recovery period. The combination of aging demographic challenges and the possibility of higher rates in the later years of the economic expansion put even more of an emphasis on understanding the ramifications of policy shifts. And in many countries, not just

the U.S., the policy pendulum is swinging away from a dependence on monetary policy and a greater willingness to expand fiscal policy, which implies more debt.

Our analysis strongly suggests that since unconventional monetary policy largely impacted asset prices and not real GDP growth, assets may be entering a period of greater risk than historical measures of volatility might suggest, as unconventional monetary policy is ended. And, the starting point for fiscal expansion already embodies high debt-to-GDP ratios. If the expansionary fiscal policies fail to generate the hoped-for real economic growth, then the unanticipated increases in inflation and interest rates may bring difficult challenges with debt management to the fore.

Public Disclosure and Risk-adjusted Performance at Bank Holding Companies

Beverly Hirtle – Senior Vice President, Federal Reserve Bank of New York¹

Abstract

This article examines the relationship between the amount of information disclosed by bank holding companies (BHCs) and the BHCs' subsequent risk-adjusted performance. The key finding is that more disclosure is associated with higher risk-adjusted returns. This result is strongest for BHCs where trading represents a large share of overall firm activity. More disclosure does not appear to be associated with higher risk-adjusted performance during the financial crisis, however, implying that the findings are a “business as usual” phenomenon. These findings suggest that greater disclosure is associated with more efficient risk taking and thus improved risk-return trade-offs, a channel for market discipline that has not been emphasized previously in the literature.

¹ This article is republished with the permission of the Economic Policy Review and the Federal Reserve Bank of New York. Full reference of the original article is: Hirtle, B., 2016, “Public disclosure and risk-adjusted performance at bank holding companies,” Federal Reserve Bank of New York Economic Policy Review 22:1, https://www.newyorkfed.org/research/epr/2016/epr_2016_public-disclosures_hirtle. The author thanks Sarita Subramanian, Matthew Botsch, Ging Cee Ng, Peter Hull, Vitaly Bord, Eric McKay, and Bryan Yang for excellent research assistance in constructing the dataset used in this article and Robert DeYoung, Mark Flannery, Donald Morgan, Christophe Pérignon, Philip Strahan, and Til Schuermann for helpful comments and suggestions. The views expressed in this article are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.

The author declares that she has no relevant or material financial interests that relate to the research described in this article. The views expressed are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

INTRODUCTION

Market discipline has occupied an increasingly prominent position in discussions of the banking industry in recent years. Market discipline is the idea that the actions of shareholders, creditors, and counterparties of banking companies can influence the investment, operational, and risk-taking decisions of bank managers [Flannery (2001); Bliss and Flannery (2002)]. Bank supervisors have embraced market discipline as a complement to supervisory and regulatory tools for monitoring risk at individual banks and for limiting systemic risk in the banking system. For instance, the Basel Committee on Banking Supervision says “the provision of meaningful information about common risk metrics to market participants is a fundamental tenet of a sound banking system. It reduces information asymmetry and helps promote comparability of banks’ risk profiles” [Basel Committee on Banking Supervision (2015)].²

For market discipline to be effective, market participants must have sufficient information to assess the current condition and future prospects of banking companies. This fact has prompted a range of proposals for enhanced public disclosure by banks. Many of these proposals have focused on disclosure of forward-looking risk information, such as value at risk (VaR) for trading portfolios or model-based estimates of credit risk exposure. In the words of a major international supervisory group, disclosure of VaR and other forward-looking risk measures is a means of providing “a more meaningful picture of the extent and nature of the financial risks a firm incurs, and of the efficacy of the firm’s risk management practices” [Multidisciplinary Working Group on Enhanced Disclosure (2001)].

But to what extent does such information result in meaningful market discipline? Is risk taking or performance affected by the amount of information banks provide about their risk exposures and risk management systems? This article explores these questions by examining whether the amount of information disclosed by a sample of large U.S. bank holding companies (BHCs) affects the future risk-adjusted performance of those banking firms. We focus, in particular, on disclosures made in the banks’ annual reports about market risk in their trading activities. Following previous work on disclosure [Baumann and Nier (2004); Nier and Baumann (2006); Pérignon and Smith (2010); Zer (2013)], we construct a market risk disclosure index and ask how differences in this index affect future performance. Drawing on data from the banking companies’ regulatory reports, we examine each BHC’s returns from trading activities and, using equity market data, we examine returns for the firm as a whole.

The main finding of this analysis is that the disclosure of more information is associated with higher risk-adjusted trading returns and higher risk-adjusted market returns for the bank overall. This result is strongest for BHCs whose trading represents a large share of overall firm activity. The results are both statistically significant and economically meaningful, with a one standard deviation increase in the disclosure index leading to a 0.35 to 0.60 standard deviation increase in risk-adjusted returns. The positive relationship between disclosure and risk-adjusted performance is much less evident during the financial crisis period, however, suggesting that the findings reflect business-as-usual behavior. Finally, while higher values of the disclosure index are associated with better future performance, being a leader or innovator in disclosure practices seems to be associated with lower risk-adjusted market returns. This finding suggests that there may be a learning process in the market such that disclosure “first movers” – those banks that provide new types of information – face a market penalty.

Overall, the results suggest that increased disclosure may be associated with more efficient trading and an enhanced overall risk-return trade-off. These findings seem consistent with the view that market discipline affects not just the amount of risk a BHC takes, but how efficiently it takes that risk. This interpretation highlights the importance of examining returns, as well as risk, when assessing the effectiveness of market discipline.

An important question in interpreting these results is whether greater disclosure leads to enhanced market discipline and thus better performance, or whether some other channel is at work. Specifically, banks with better risk management systems may be able to trade more efficiently and, in a more general sense, be able to achieve a better risk-return trade-off. The same risk management systems that produce better risk-adjusted performance may also generate the information needed to make more detailed risk disclosures, which may be used by the bank as a public signal of its superior risk management abilities. Fang (2012) finds a correlation between VaR disclosures and measures of effective corporate governance, consistent with this channel. While this conclusion may not be the traditional view of market discipline, it is in keeping with the idea that the role of public information is to provide incentives for managers to optimize overall performance. This interpretation suggests that there are many potential channels for the exercise of market discipline on firms.

² The Basel II/III regulatory capital regime incorporates market discipline as the “third pillar,” along with minimum capital standards and supervisory oversight [Basel Committee on Banking Supervision (2004)].

The remainder of this article is organized as follows. Section 2 reviews previous work on the impact of disclosure in the banking industry and discusses how this article fits into that literature. Section 3 describes the empirical approach and data used in this analysis, with particular emphasis on the market risk disclosure index. The results are presented in Section 4, while the final section contains a summary and conclusions.

DISCLOSURE AND BANK PERFORMANCE

A number of previous papers have examined the impact of disclosure in the banking industry. The key idea is that disclosure of information about banks' current condition and future prospects will facilitate market discipline of risk-taking behavior. As argued in Flannery (2001) and Bliss and Flannery (2002), market discipline requires that investors and creditors have the ability to monitor and assess changes in bank condition and to influence management behavior. Both components are affected by the amount and quality of information disclosed. In theory, greater disclosure provides investors and creditors with more information on which to base their assessments of firm condition, which in turn makes a significant market reaction to an adverse change in condition – and subsequent management response – more likely and immediate.

The influence of market discipline on bank behavior may occur not only through a bank's response to a market reaction but also its anticipation of one. That is, market discipline may also work by affecting management behavior ex-ante so as to prevent a negative outcome and consequent market reaction. In this sense, greater disclosure can serve as a kind of commitment device by providing sufficient information to the market about a bank's condition and future prospects that the bank is constrained from altering its risk profile in a way that disadvantages either investors or creditors [Cumming and Hirtle (2001)]. Banks' ability to shift assets and risk positions quickly has been cited as one of the key sources of opaqueness in the banking industry [Meyers and Rajan (1998)]. In fact, several studies have found evidence of greater opaqueness at banks with higher shares of liquid assets, including, especially, trading positions [Morgan (2002); Iannotta (2006); Hirtle (2006)].³ In a related vein, Bushman and Williams (2012) find that loan loss provisioning practices intended to smooth earnings inhibit risk-taking discipline by making banks more opaque to outsiders.

Underlying much of this discussion is the idea that greater disclosure and enhanced market discipline will lead to reductions

in bank risk. Enhanced market discipline would mean that the costs of increased risk would be more fully borne by the bank and would, therefore, presumably play a larger role in its risk-taking decisions. More risk-sensitive market prices could also provide signals to regulators that might induce or influence supervisory action [Flannery (2001)]. While greater disclosure is likely to lead to a reduction in bank risk, it might also have some offsetting negative outcomes. More information reduces the likelihood that the bank would face an excessive (undeserved) risk premium or that market prices would overreact to news about the firm because of uncertainty about its true condition and prospects – an effect that could lower the bank's funding costs and increase the range of viable (positive net present value) investments, some of which could be riskier than its current portfolio. The net impact of all of these influences is an empirical question.

Most of the previous empirical work on market discipline has focused on how disclosure affects bank risk taking. For instance, several papers examine market price reaction to changes in bank condition or to differences in risk profiles across banks. Some of these papers have found that bond spreads increase with bank risk exposure, especially following the early 1990s reforms associated with the Federal Deposit Insurance Corporation Improvement Act. Morgan and Stiroh (2001) find that banks with riskier assets (such as trading assets) pay higher credit spreads on newly issued bonds. Similarly, Covitz et al. (2004a, b) and Jagtiani et al. (2002) find evidence that subordinated debt spreads increase with banking company risk. In related work, Goyal (2005) finds that riskier banks are more likely to have restrictive debt covenants in their publicly issued debt. However, more recent work [Balasubramanian and Cyree (2011); Acharya et al. (2014); Santos (2014)] suggests that the bonds of the largest banking companies are less sensitive to risk than bonds issued by smaller BHCs, presumably because the larger firms are regarded by market participants as "too big to fail." These papers call into question the efficacy of market discipline, at least for the very largest and most complex BHCs.

³ In contrast, Flannery et al. (2004) find no evidence that bank assets are more opaque than the assets of nonfinancial firms.

Performance variables	Mean	Median	Standard deviation	Minimum	Maximum
Risk-adjusted trading return	3.063	2.330	3.033	-5.428	21.501
Risk-adjusted market return	0.083	0.082	0.138	-0.333	0.0371
Alpha	0.046	0.025	0.483	-1.992	4.034
Disclosure variables					
Disclosure leader	0.072	0	0.260	0	1
Aggregate disclosure index	5.769	5	4.653	0	15
First principal component	0.014	-0.650	2.660	-3.018	5.692
BHC characteristics					
Asset size	415.2	169.7	573.3	25.1	2457.9
Risk-weighted assets divided by total assets	0.758	0.795	0.174	0.309	1.144
Common equity divided by total assets	8.271	8.248	1.950	3.235	15.696
Trading assets divided by total assets	0.073	0.029	0.103	0.001	0.490
Non-interest income divided by operating income	0.524	0.466	0.160	0.018	0.996
Revenue source concentration	0.406	0.404	0.063	0.249	0.654

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: The sample consists of 293 annual observations for a sample of thirty-six BHCs with trading assets exceeding U.S.\$1 billion (in 2013 dollars) at some point between 1994 and 2012. BHC characteristics and trading revenue data are from the Federal Reserve Y-9C reports. Disclosure data are from the BHCs' annual reports. Market price data are from CRSP. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. Risk-adjusted market returns is the annual average of weekly equity price returns divided by the standard deviation of weekly returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Trading return is annual trading revenue divided by trading assets. Market return is the annual average of weekly equity price returns. Disclosure leader is a dummy variable that indicates whether a BHC is the only one to report a given disclosure item in a given year. Aggregate disclosure index is the value of the market risk disclosure index. First principal component is the first principal component of the eighteen individual data items that comprise the aggregate index.

Table 1 – Basic statistics of the regression sample

In a somewhat different vein, several papers have examined the impact of disclosure on risk taking using equity trading characteristics – such as bid-ask spreads or price volatility – as proxies for risk.⁴ Many of these studies focus on non-financial firms [for example, Bushee and Noe (2000); Luez and Verrecchia (2000); Linsmeier et al. (2002)], but some examine the link between disclosure and market volatility in the banking industry. Baumann and Nier (2004) and Nier and Baumann (2006) construct a disclosure index based on the number of balance sheet and income statement items reported by a cross-country sample of banks. They find that stock price volatility decreases and capital buffers increase as the amount of information disclosed increases, consistent with the idea that greater disclosure enhances market discipline. Zer (2013) constructs a disclosure index using balance sheet information from BHC 10-K filings submitted to the U.S. Securities and Exchange Commission and shows that BHCs with higher values of the index have lower option-implied default probabilities and stock price volatility.

Fewer papers have examined the relationship between disclosure and performance – that is, whether banking companies that disclose more information have better subsequent operating or stock market performance. Several papers have examined this relationship for nonfinancial firms. Eugster and Wagner (2011) construct an index of voluntary disclosure by Swiss companies and demonstrate that firms with higher voluntary disclosure have higher abnormal stock returns, though this effect is evident predominantly for more opaque companies. Barth et al. (2013) find that firms with more transparent earnings have a lower cost of capital.

In the banking industry, Ellul and Yerramilli (2013) find that

⁴ Using a very different approach, Kwan (2004) examines the impact of market discipline on bank risk taking by comparing the risk profiles of publicly traded and non-publicly traded BHCs. He finds that publicly traded banks take more risk than non-publicly traded institutions, which he interprets as being contrary to market discipline.

banks with stronger risk management have higher operating profits (return on assets) and stock return performance. While that paper focuses on risk management rather than disclosure per se, it measures risk management strength based on an index constructed from 10-K filings – an approach similar to the one used in this article and others focusing on disclosure. Ellul and Yerramilli is also relevant because risk management and disclosure are linked, in that enhanced risk management systems generate the kind of forward-looking risk information disclosed by some BHCs. Consistent with this idea, Fang (2012) finds a positive correlation between the amount of information BHCs disclose about VaR and measures of effective corporate governance. Fang also finds that more disclosure is correlated with a lower cost of capital, when cost of capital is measured using equity analyst forecasts.

The analysis in this article is complementary to previous work on disclosure in that it examines the impact of enhanced disclosure on both operating and stock market performance for large U.S. bank holding companies. In particular, it investigates whether enhanced disclosure is associated with higher subsequent risk-adjusted performance. The analysis thus assesses whether disclosure affects the efficiency of risk taking, rather than whether enhanced disclosure is associated with higher or lower risk per se. As noted above, the theoretical relationship between disclosure and risk taking is not straightforward and there likely is considerable endogeneity between disclosure and subsequent risk.⁵ While the extent of both risk taking and disclosure are decisions made by each banking company, risk-adjusted performance is an outcome that is less directly under a firm's control. By examining performance, we gain an additional window into the ways that market discipline may play out at banking companies, because investors and creditors presumably care not only about the level of risk but also about how efficiently a bank translates its risk exposures into profits and returns.

Like much of the prior work, the analysis in this article is based on a disclosure index constructed from information reported by these banks in their annual reports or 10-K filings with the SEC. However, rather than constructing a disclosure index based primarily on balance sheet and income statement variables – which tend to be backward-looking – the disclosures we track are forward-looking risk estimates made by the banking companies.⁶ The index focuses specifically on disclosures concerning the market risk in banks' trading and market-making activities.

We focus on market risk in trading activities because trading is a well-defined banking business activity with distinct

regulatory and financial statement reporting. BHC annual reports have specific sections for reporting about market risk, and regulatory reports contain trading return information that can be linked directly to these activities. Thus, we can examine the impact of disclosure on overall firm performance and on the specific activities that are the focus of the disclosures. Previous work has also found that trading activities are associated with greater opaqueness and risk, so this is an area of banking for which disclosure might be particularly influential.

DATA AND EMPIRICAL APPROACH

Because we are interested in determining the impact of disclosure on BHC risk and performance specifically as it relates to market risk in trading activities, we begin by constructing a sample of U.S.-owned BHCs that appear to be active traders. We limit the sample to BHCs with significant trading activities because those are the firms that are most likely to make disclosures related to market risk in their annual reports. BHCs that are relatively active traders are also more likely to be engaged in purposeful risk management of their trading positions than they are to be using the trading account simply to book a limited number of mark-to-market positions.

To identify those BHCs with significant trading account assets, we use information from the Consolidated Financial Statements for Bank Holding Companies, the FR Y-9C quarterly reports filed by BHCs with the Board of Governors of the Federal Reserve System.⁷ Overall, relatively few BHCs report holding any assets in the trading account: at year-end 2013, only 164 (of more than 1,000) large BHCs reported holding any trading account assets, and only 18 of these held trading assets exceeding U.S.\$1 billion. Our sample consists of all U.S.-owned BHCs with year-end trading account assets exceeding U.S.\$1 billion (in 2013 dollars) at some point between 1994 and 2012.⁸ We include a BHC in the sample starting with the first year in

5 Ellul and Yerramilli (2013) and Zer (2013) use instrumental variable techniques to address this endogeneity.

6 As explained in Section 3, the index is similar to the one constructed in Pérignon and Smith (2010).

7 The FR Y-9C reports are available at <https://www.chicagofed.org/applications/bhc/bhc-home>.

8 We exclude foreign-owned BHCs because the U.S. activities of these institutions represent only a part of the banks' overall activities and because many of them do not submit 10-K filings with the SEC, which we need to construct the market risk disclosure index. In addition, two U.S. BHCs whose activities are primarily nonbanking in nature – MetLife and Charles Schwab – are omitted from the sample.

which its constant-dollar trading assets exceed U.S.\$500 million. The resulting sample consists of 293 observations from 36 BHCs over the years 1994 to 2012.⁹

The estimates consist of a series of regressions of risk-adjusted performance measures in year $t + 1$ on BHC characteristics and disclosure during year t :

$$Y_{i,t+1} = \beta_1 \text{Disclosure}_{i,t} + X_{i,t}\Gamma + \varepsilon_{i,t+1},$$

where $Y_{i,t+1}$ is the risk-adjusted performance measure (discussed below), $\text{Disclosure}_{i,t}$ is the index of market risk disclosure, and $X_{i,t}$ is a vector of BHC control variables. Both the disclosure index and the control variables are lagged one year to avoid endogeneity with the performance measures. Thus, disclosure data and control variables from 1994 to 2012 are paired with performance data from 1995 to 2013.

The control variables include measures of institution size (the log of assets), risk profile (the ratio of risk-weighted assets to total assets and the ratio of common equity to total assets), revenue composition (non-interest income as a share of operating income), and revenue concentration (Herfindahl-Hirschman Indices based on sources of revenue).¹⁰ The regressions also include the ratio of trading assets to total assets as a measure of the extent of the institution's trading activities. All BHC data are from the Y-9C reports. The regressions also include BHC fixed effects and year dummies. Table 1 reports the basic statistics of the regression data set.

The key variables in the estimates are the measures of risk-adjusted performance and the market risk disclosure index. The risk-adjusted performance measures are based on two distinct sets of information. The first is derived from accounting data on BHCs' trading activities. Specifically, BHC regulatory reports contain information on quarterly trading revenues: the gains and losses on the firms' trading activities, including commission, fee, and spread income. We collect trading performance data from the first quarter of 1995 to the fourth quarter of 2013. Using these data, we calculate quarterly trading returns as trading revenues in a quarter as a percentage of beginning-of-quarter trading assets. Trading volatility is then calculated as the standard deviation of quarterly trading returns within a year, and trading returns are calculated as the annual average of quarterly trading returns. Finally, we compute risk-adjusted trading returns as trading returns divided by trading volatility (essentially, the trading revenue "Sharpe ratio"). Since this measure reflects risk and return on the BHCs' trading activities, it is tied directly to the disclosure information covered in the market risk disclosure index.

Category	Data items
Overall value at risk (VaR)	Holding period and confidence interval
	Annual average VaR
	Year-end VaR
	Minimum VaR over the year
	Maximum VaR over the year
	VaR limit (dollar amount)
VaR by risk type	Histogram of daily VaR
	Annual average VaR by risk type
	Year-end VaR by risk type
	Minimum VaR by risk type
Backtesting	Maximum VaR by risk type
	Chart of daily trading profit and loss versus daily VaR
	Number of days that losses exceeded VaR
Returns distribution	Histogram of daily trading profit and loss
	Largest daily loss
Stress-testing	Mention that stress-tests are done
	Describe the stress-tests qualitatively
	Report stress-test results

Table 2 – The market risk disclosure index

The second set of measures is derived from firm-wide equity prices. Specifically, we use stock return data from the University of Chicago's Center for Research in Security Prices (CRSP) for the BHCs in our sample. For each year between 1995 and 2013, we cumulate daily returns from CRSP to form weekly returns, and then calculate annual average weekly returns, expressed at an annual rate. We also calculate the standard deviation of weekly returns within each year, and generate risk-adjusted market returns as the ratio of average returns to the standard deviation of returns. As a second measure of risk-adjusted market performance, we include in the dataset the "alpha" (intercept term) from the three-factor Fama-French model, where the model is estimated annually for each BHC using weekly return data and risk factors.

⁹ The sample is an unbalanced panel, owing mainly to the impact of mergers. During the sample period, several of the BHCs were acquired, generally by other BHCs in the sample. In addition, some BHCs in the sample acquired large BHCs that were not part of the sample. In estimates, we treat the pre- and post-merger acquiring BHCs as separate entities. Observations for the year in which a given merger was completed are omitted. Finally, some BHCs enter the sample midway through the sample period because their trading assets crossed the U.S. \$500 million threshold or because they converted to BHCs during the 2007-09 financial crisis.

¹⁰ The revenue concentration index is based on the shares of net interest income, fiduciary income, deposit service charges, trading revenue, and other non-interest income in overall operating income. Stiroh (2006) shows that revenue concentration is a significant determinant of BHC equity price volatility.

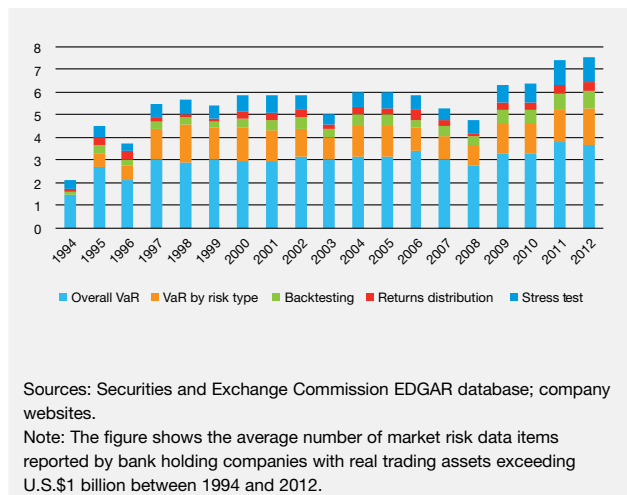


Figure 1 – Average market risk disclosure index, 1994-2012

Basic statistics for all of the risk and performance measures are reported in Table 1.

The market risk disclosure index is the other key variable in the analysis. As explained above, this index captures the amount of information that banks disclose about their forward-looking estimates of market risk exposure in their annual reports or 10-K filings with the SEC.¹¹ The index covers eighteen specific types of information that BHCs could provide in their filings, primarily related to their value-at-risk (VaR) estimates.

VaR is a very commonly used measure of market risk exposure from trading activities. VaR is an estimate of a particular percentile of the trading return distribution, assuming that trading positions are fixed for a specified holding period. VaR estimates made by banks in the sample are typically based on a one-day holding period, generally at the 95th percentile and above.¹² VaR estimates form the basis of banks' regulatory capital requirements for market risk [Hendricks and Hirtle (1997)] and have been the focus of disclosure recommendations made by financial industry supervisors [Multidisciplinary Working Group on Enhanced Disclosure (2001); Basel Committee on Banking Supervision (2015)].

The eighteen items covered in the market risk disclosure index include information about a BHC's VaR estimates for its entire trading portfolio ("overall VaR"), VaR by risk type (for example, risk from interest rate or equity price movements), the historical relationship between VaR estimates and subsequent trading returns ("backtesting"), the distribution of actual trading outcomes ("returns distribution"), and stress-testing.

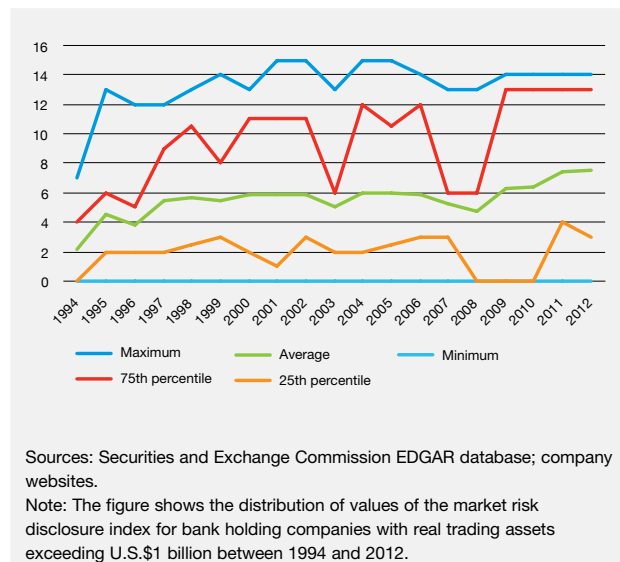


Figure 2 – Distribution of market risk disclosure index, 1994-2012

The specific items included in the index are listed in Table 2. These items were selected based on a review of a sample of BHC disclosures to determine which items were disclosed with enough frequency to be meaningfully included in the index, and also by benchmarking the individual items and the five broader categories against those listed in a rating agency evaluation of banks' disclosure practices [Moody's Investors Service (2006)].

The market risk disclosure index measures the amount of information that BHCs disclose about their market risk exposures, not the content of that information. It is a count of the number of data items disclosed, not an indicator of the amount or nature of market risk exposure undertaken by the BHC. In that sense, it is similar to the disclosure indexes constructed by Nier and Baumann (2006) and Zer (2013), though it is based on different types of data. It is also quite similar to a VaR disclosure index developed independently by Pérignon and Smith (2010).¹³ The Pérignon and Smith (2010) index covers much of the same information as the index in this article, though the

11 We used the SEC's EDGAR database to access the 10-K filings. The EDGAR database is available at: <http://www.sec.gov/edgar.shtml>.

12 See Jorion (2006) for an extensive discussion of VaR modeling, and Moody's Investors Services (2006) for a description of typical VaR parameter choices at banks and securities firms.

13 Fang (2012) uses a disclosure index similar to the one used in this article, in Hirtle (2007), and in Pérignon and Smith (2010).

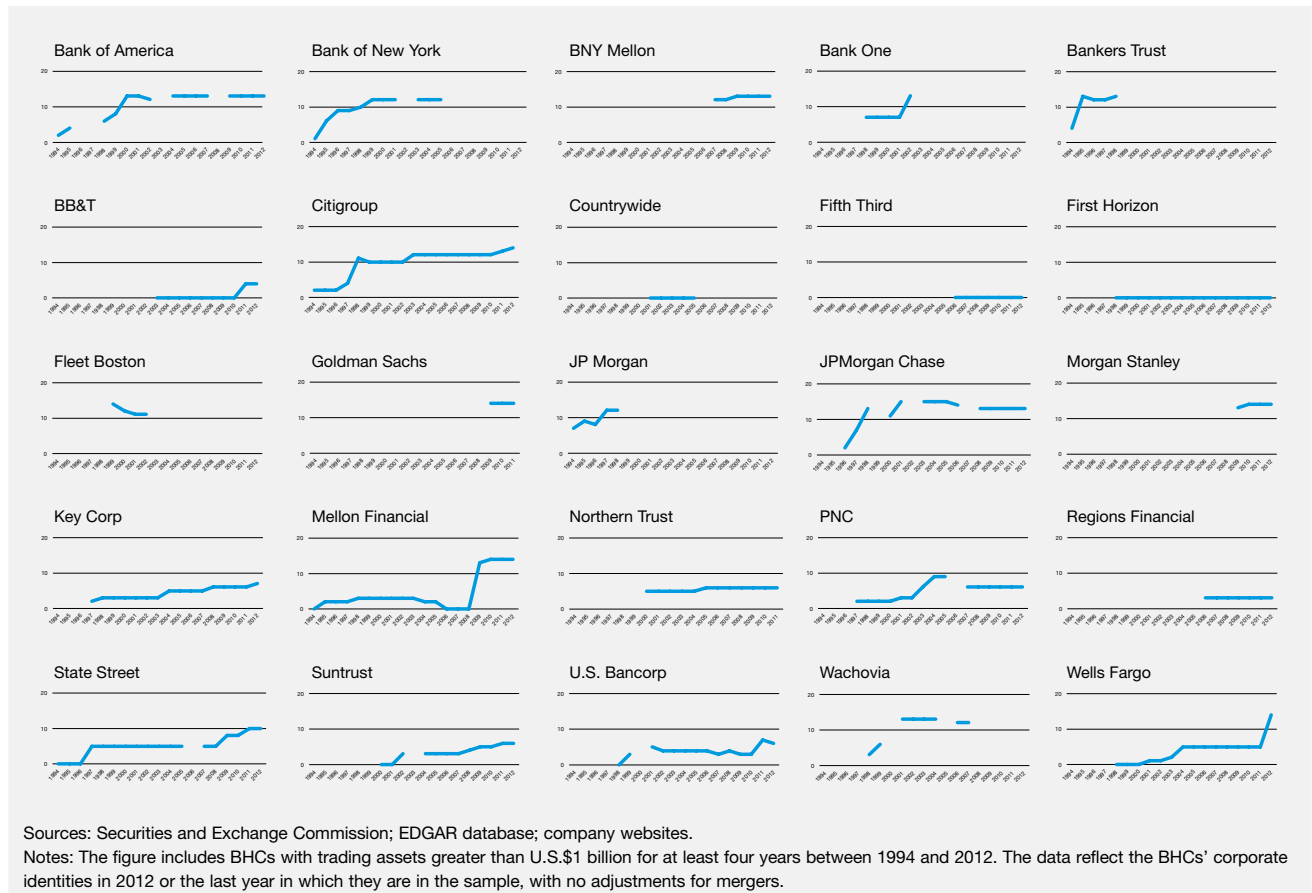


Figure 3 – Disclosure index for large BHCs

authors use their index primarily to make cross-country comparisons of disclosure practices rather than to examine the link between the index and future risk and performance.¹⁴

Figure 1 shows the average value of the market risk disclosure index between 1994 and 2012. The average value of the index increases from just over 2 in 1994 to nearly 8 in 2012. Most of this increase occurs during the early part of the sample, between 1994 and 1998.

The growth through 1998 reflects two significant regulatory developments. First, following the international agreement in Basel, U.S. risk-based capital guidelines were amended in 1998 to incorporate minimum regulatory capital requirements for market risk in trading activities, with the requirements taking full effect in January of that year [Hendricks and Hirtle (1997)]. The market risk capital charge introduced through this amendment is based on the output of banks' internal VaR models, and the need to comply with the new

capital requirements spurred the development of VaR models in the banking industry. On a separate track, SEC Financial Reporting Release (FRR) 48 required all public firms with material market risk exposure to make enhanced quantitative and qualitative disclosures about these risks, starting in 1997 [U.S. Securities and Exchange Commission (1997)]. FRR 48 included three options for forward-looking, quantitative market risk disclosures, one of which was VaR.¹⁵ Together, these two regulatory developments spurred disclosure of VaR estimates and related information.

14 Pérignon and Smith (2010) examine the link between VaR estimates and subsequent trading volatility, a question that is related to, but distinct from, the one we address. They find that VaR estimates contain little information about future trading volatility. This finding is similar to that in Berkowitz and O'Brien (2002) but stands in contrast to the results in Jorion (2002), Hirtle (2003), and Liu et al. (2004), all of which find that VaR measures contain information about future trading income volatility.

15 The Pérignon and Smith (2006) index also grows through 1998, and the authors cite the influence of FRR 48 in this finding for the U.S. banks in their sample. See Roulstone (1999) for an assessment of the impact of FRR 48 on nonfinancial firms.

	Market risk disclosure index	Average real assets	Average real trading assets	Average trading assets divided by total assets
Market risk disclosure index	1.000			
Average real assets	0.627 (0.000)	1.000		
Average real trading assets	0.653 (0.000)	0.881 (0.000)	1.000	
Average trading assets divided by total assets	0.605 (0.000)	0.464 (0.000)	0.705 (0.000)	1.000

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Securities and Exchange Commission EDGAR database; company websites.

Notes: Figures in the table reflect average values for the thirty-six BHCs that have trading assets of more than U.S.\$1 billion at some point between 1994 and 2012. Total assets and trading assets are in 2013 dollars and are averaged across the years that a BHC is in the sample. P-values are shown in parentheses.

Table 3 – Correlation between market risk disclosure index and BHC asset size and trading activity

Figure 1 shows the average value of the market risk disclosure index, but the average masks considerable diversity across BHCs in the sample. Figure 2 illustrates the range of disclosure index values by year. Specifically, the figure shows the minimum and maximum values of the index by year and the 25th and 75th percentiles, along with the averages reported in Figure 1. The maximum value of the index grows from 7 in 1994 to 15 in the mid-2000s, falls back to 13, and then settles at 14 near the end of the sample period. At least one BHC in each year reported no market risk information (in other words, generated an index value of zero). As the average value of the disclosure index increases, the dispersion within the sample BHCs grows. The interquartile range (25th to 75th percentile) more than doubles over the sample period, owing mainly to growing differentiation in the top half of the distribution after 1998. Over the full period, the distance between “top reporting” BHCs and those nearer to the average widened considerably.

Data Item	Share of observations			
	Overall value at risk	All observations	1994	2012
Overall value at risk				
Holding period and confidence interval	0.749	0.538	0.737	
Annual average VaR	0.624	0.308	0.789	
Year-end VaR	0.475	0.154	0.474	
Minimum VaR over the year	0.488	0.154	0.737	
Maximum VaR over the year	0.536	0.231	0.789	
VaR limit (dollar amount)	0.115	0.000	0.053	
Histogram of daily VaR	0.058	0.076	0.105	
VaR by risk type				
Annual average VaR by risk type	0.342	0.000	0.421	
Year-end VaR by risk type	0.217	0.000	0.316	
Minimum VaR by risk type	0.315	0.000	0.421	
Maximum VaR by risk type	0.319	0.000	0.421	
Backtesting				
Chart of daily profit and loss versus daily VaR	0.112	0.077	0.211	
Number of days losses exceeded VaR	0.349	0.077	0.579	
Returns distribution				
Histogram of daily profit and loss	0.220	0.154	0.368	
Largest daily loss	0.075	0.000	0.053	
Stress-testing				
Mention that stress-tests are done	0.420	0.308	0.579	
Describe stress-tests	0.231	0.077	0.473	
Report stress-test results	0.017	0.000	0.000	

Sources: Securities and Exchange Commission EDGAR database; company websites.

Notes: Figures are from 1994 to 2012 10-K reports of the thirty-six bank holding companies in the market risk sample. These companies each have trading assets exceeding U.S.\$1 billion (in 2013 dollars) at some point between 1994 and 2012.

Table 4 – Frequency of individual data items in the market risk disclosure index

Figure 3 shows the market risk disclosure index at the individual BHC level. The BHCs shown in the figure are those that are in the sample for at least four years, traced backward from the BHCs’ corporate identity at the end of the sample period without adjusting for mergers. Not surprisingly given the average results, the index tends to increase over the sample period at the individual BHC level. The typical pattern is for the index to rise in steps over time, though there are certainly cases in which the index declines.

Disclosure Variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.058 ^b (0.029)	-0.057 ^c (0.029)	-0.193 ^c (0.111)	-0.189 (0.114)	1.997 ^c (1.000)	2.050 ^b (0.972)
Aggregate disclosure index	0.010 ^a (0.002)		0.044 ^a (0.013)		0.332 ^b (0.154)	
First principal component		0.018 ^a (0.004)		0.077 ^a (0.023)		0.687 ^b (0.307)
BHC characteristics						
Log (asset size)	-0.061 ^a (0.018)	-0.064 ^a (0.019)	-0.404 ^a (0.111)	-0.412 ^a (0.116)	0.001 (0.964)	-0.165 (0.926)
Risk-weighted assets divided by total assets	-0.085 (0.098)	-0.072 (0.098)	-0.073 (0.716)	-0.014 (0.715)	7.322 ^c (3.789)	7.790 ^b (3.776)
Common equity divided by total assets	-0.011 ^b (0.005)	-0.011 ^b (0.005)	-0.089 ^a (0.033)	-0.090 ^a (0.033)	0.106 (0.198)	0.103 (0.194)
Trading assets divided by total assets	-0.646 ^b (0.243)	-0.652 ^b (0.245)	-2.060 ^c (1.174)	-2.084 ^c (1.175)	17.346 (11.585)	17.102 (11.553)
Non-interest income divided by operating income	-0.060 (0.093)	-0.060 (0.093)	0.168 (0.762)	0.168 (0.763)	5.807 ^b (2.302)	5.771 ^b (2.303)
Revenue source concentration	0.089 (0.146)	0.084 (0.145)	0.141 (0.941)	0.113 (0.937)	14.656 ^b (6.343)	14.733 ^b (6.491)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	293	293	293	293	295	295
R-squared	0.781	0.781	0.314	0.313	0.177	0.186
P-Value: Disclosure Variables = 0?	0.000	0.000	0.000	0.000	0.021	0.017

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: Risk-adjusted market return is the annual average of weekly equity price returns divided by the standard deviation of those returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. BHC characteristics are from the Federal Reserve Y-9C reports. Disclosure information is from the BHCs' annual reports. Stock data are from CRSP. Disclosure leader is a dummy variable indicating that a BHC is the only BHC to disclose a particular data item in a given year. Aggregate disclosure index is the market risk disclosure index. First principal component is based on the eighteen individual data items that comprise the aggregate index. The sample consists of all U.S.-owned BHCs that have trading assets greater than U.S.\$1 billion (in 2013 dollars) at any time between 1994 and 2012, starting with the year that trading assets exceed U.S.\$500 million. The regressions include BHC fixed effects and year dummy variables. Residuals are clustered at the BHC level.

a Significant at the 1% level, b significant at the 5% level, c significant at the 10% level.

Table 5 – Disclosure and risk-adjusted returns

On a cross-sectional basis, the index tends to be higher at larger BHCs and at BHCs with more trading activity, on both an absolute and relative level. Table 3 reports the correlation between the value of the market risk disclosure index and real (2013 dollar) assets, trading assets, and trading asset share, where values are averaged across the years that a BHC is in the sample. Reading down the first column of the table, the correlation coefficients between the disclosure index and the measures of BHC and trading activity scale are large and positive.

Finally, Table 4 reports the frequency with which the individual data items in the market risk disclosure index are reported.

The first column reports the frequency across all observations between 1994 and 2012, while the next two columns report the frequency at the beginning and end of the sample period. The most commonly reported data element is the holding period and confidence interval of the VaR estimate, reported for about 75% of the BHC-year observations. This data item is a close proxy for whether a BHC disclosed any information about VaR at all. About 30% of the observations include some information about VaR by risk type, while information about back-testing and the distribution of returns is reported in 10 to 35% of the observations. About 40% of the observations indicate that the BHC does some kind of stress-testing, but

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.049 (0.033)	-0.047 (0.033)	-0.199 (0.125)	-0.192 (0.128)	1.741 (1.190)	1.823 (1.163)
Aggregate disclosure index	0.010 ^a (0.003)		0.040 ^a (0.014)		0.302 ^c (0.155)	
First principal component		0.018 ^a (0.005)		0.070 ^a (0.026)		0.635 ^b (0.308)
BHC characteristics						
Log (asset size)	-0.058 ^b (0.029)	-0.060 ^c (0.030)	-0.330 ^b (0.156)	-0.337 ^b (0.164)	-0.590 (1.382)	-0.737 (1.341)
Risk-weighted assets divided by total assets	-0.022 (0.116)	-0.009 (0.115)	-0.174 (0.638)	-0.123 (0.636)	7.500 ^b (3.483)	7.852 ^b (3.483)
Common equity divided by total assets	-0.011 ^c (0.006)	-0.011 (0.006)	-0.043 (0.031)	-0.043 (0.032)	0.062 (0.351)	0.071 (0.337)
Trading assets divided by total assets	-0.625 ^b (0.242)	-0.631 ^b (0.246)	-1.401 (1.067)	-1.417 (1.081)	25.188 ^c (13.429)	24.891 ^c (13.262)
Non-interest income divided by operating income	-0.109 (0.109)	-0.109 (0.108)	-0.466 (0.603)	-0.464 (0.603)	8.281 ^a (2.771)	8.164 ^a (2.708)
Revenue source concentration	0.149 (0.193)	0.140 (0.191)	0.273 (0.807)	0.231 (0.802)	13.418 ^b (6.174)	13.467 ^b (6.273)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	247	247	247	247	249	249
R-squared	0.782	0.783	0.424	0.424	0.160	0.170
P-Value: Disclosure variables = 0?	0.000	0.000	0.002	0.002	0.070	0.057

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: Risk-adjusted market return is the annual average of weekly equity price returns divided by the standard deviation of those returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. BHC characteristics are from the Federal Reserve Y-9C reports. Disclosure information is from the BHCs' annual reports. Stock data are from CRSP. Disclosure leader is a dummy variable indicating that a BHC is the only BHC to disclose a particular data item in a given year. Aggregate disclosure index is the market risk disclosure index. First principal component is based on the eighteen individual data items that comprise the aggregate index. The sample consists of all U.S.-owned BHCs that have trading assets greater than U.S.\$1 billion (in 2013 dollars) at any time between 1994 and 2012, starting with the year that trading assets exceed U.S.\$500 million. Observations for the years 2007, 2008, and 2009 are omitted. The regressions include BHC fixed effects and year dummy variables. Residuals are clustered at the BHC level.

a Significant at the 1% level, b significant at the 5% level, c significant at the 10% level.

Table 6 – Disclosure and risk-adjusted returns omitting the financial crisis period

only a tiny share – less than 2% – report the results of these efforts. As a comparison of the columns with data from 1994 and 2012 makes clear, the frequency of reporting increased over the span of the sample period for nearly every data item.

In the regressions, we use the overall market risk disclosure index as the baseline measure of disclosure, but we also construct the first principal component of the cross-sectional variation in reporting of the eighteen individual data items in the index. The basic index is a simple linear weighting (sum) of the individual elements. The first principal component provides an alternate linear combination, with weights that reflect the

common variation across BHC-year observations. It captures about 40% of this variation, suggesting a meaningful common component of reporting across the individual data items. Finally, we create an indicator variable if a given BHC is the only one in the sample to disclose a particular data item in a particular year (“disclosure leader”), to assess the impact of innovations in disclosure practice.¹⁶

¹⁶ The typical pattern is that once one BHC discloses a particular kind of information, others follow in subsequent years. In that sense, BHCs that are the only ones to report an item in a given year are leaders or innovators.

DISCLOSURE AND RISK-ADJUSTED PERFORMANCE

Table 5 presents the basic results of the estimates relating market risk disclosure to subsequent risk-adjusted returns on trading activities and for the firm as a whole. The first set of columns of the table present the results for risk-adjusted market returns, the second set of columns present the results for alpha, and the final set of columns contain the results for trading returns.

The estimates uniformly suggest that increased disclosure is associated with higher risk-adjusted returns, both for trading activities and for the BHC as a whole. The coefficients on the aggregate market risk disclosure index and the first principal component variable are positive and statistically significant in each specification. Aside from being statistically significant, the results are economically important: an increase of one standard deviation in the disclosure index or the first principal components measure is associated with a 0.35 to 0.45 standard deviation increase in risk-adjusted market returns and alpha and a 0.50 to 0.60 standard deviation increase in risk-adjusted trading returns.

The coefficient estimates on the disclosure leader variable (indicating that the BHC is the only company to disclose a particular index item in a given year) are less robust across specifications. The coefficients are negative and weakly statistically significant in the equations using the market-based measures, but positive and statistically significant in the equations for risk-adjusted trading returns. These results suggest that being a first mover in disclosure is associated with better risk-adjusted performance in the trading activities associated with the disclosure but is less strongly associated with market-based returns for the firm as a whole. One potential explanation for these seemingly inconsistent results is that there are learning costs for investors in understanding and putting into context new types of information.

The sample period for the performance data, 1995 to 2013, includes the 2007-09 financial crisis. Since the crisis was a period of extraordinary volatility in financial markets and for the banking sector, one question to ask is how does including this period in the sample affect the results. To explore the impact of the unusual market conditions during the financial crisis, we re-estimated the equations omitting observations from the peak crisis years, 2007 to 2009. These results are reported in Table 6.

On the whole, omitting the financial crisis period does not significantly alter the results concerning the relationship between disclosure and subsequent risk-adjusted performance. The

coefficients on the disclosure variables continue to be positive and statistically significant, with little change in magnitude. The primary difference is that the disclosure leader variable no longer enters the equations with a statistically significant coefficient, though the signs and approximate size of the coefficients are similar to those in the basic results. Thus, the exceptional market and banking sector volatility during the financial crisis does not appear to be driving the overall results.

A related question is whether BHCs that disclosed more risk information experienced higher risk-adjusted returns during the financial crisis. The ideal way to answer this question would be to generate completely separate estimates for the crisis period, but this is not possible owing to limited annual observations. To provide some insight, however, we re-estimate the equations allowing the coefficients on the disclosure index variables to differ between the non-crisis and crisis periods (with the crisis period again defined as 2007 to 2009). Note that the disclosure leader variable is not estimated separately for the two time periods because there is insufficient variation during the crisis period to separately identify the impact. These results are reported in Table 7.

The results differ across the three measures of risk-adjusted performance. For risk-adjusted market returns, the coefficients on the disclosure index and the first principal components variables are positive and statistically significant in both the crisis and non-crisis periods. The hypothesis that the coefficients are the same cannot be rejected (see the last row of the table, which reports p-values for tests of equality of the coefficients). In contrast, for alpha and for risk-adjusted trading returns, the coefficients are positive and statistically significant only during the non-crisis period. These findings suggest that BHCs that disclosed more trading risk information did not have better (or worse) risk-adjusted trading performance during the financial crisis, while the evidence about overall firm performance is mixed.

Overall, the results in Tables 5 to 7 suggest that increased market risk disclosure is associated with higher risk-adjusted returns. If this link is achieved through market discipline on trading activities, then we might expect that the effect would be stronger for BHCs that are more heavily engaged in trading. To explore this question, we examine results where the coefficients on the disclosure variables are allowed to differ between BHCs that are "intense traders" and the rest of the sample. These results are shown in Table 8. "Intense traders" are defined as the ten BHCs in the sample with trading assets greater than or equal to U.S.\$20 billion, where trading assets represent at least 10% of total assets. Note that by construction, all BHCs in the sample have large trading accounts in absolute dollar terms,

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.058 ^a (0.029)	-0.056 ^c (0.029)	-0.283 ^b (0.139)	-0.274 ^c (0.141)	1.719 ^c (0.985)	1.783 ^c (0.965)
Crisis period (2007-09)						
Aggregate disclosure index	0.010 ^a (0.003)		-0.005 (0.023)		0.169 (0.179)	
First principal component		0.019 ^a (0.006)		-0.000 (0.043)		0.428 (0.347)
Non-crisis period						
Aggregate disclosure index	0.010 ^a (0.002)		0.046 ^a (0.013)		0.337 ^b (0.153)	
First principal component		0.018 ^a (0.004)		0.079 ^a (0.024)		0.691 ^b (0.306)
BHC characteristics						
Log (asset size)	-0.061 ^a (0.018)	-0.063 ^a (0.019)	-0.439 ^a (0.115)	-0.435 ^a (0.117)	-0.114 (0.987)	-0.244 (0.950)
Risk-weighted assets divided by total assets	-0.085 (0.098)	-0.071 (0.098)	-0.103 (0.671)	-0.073 (0.665)	7.218 ^c (3.808)	7.590 ^c (3.807)
Common equity divided by total assets	-0.011 ^b (0.004)	-0.011 ^b (0.004)	-0.102 ^a (0.033)	-0.100 ^a (0.033)	0.066 (0.215)	0.069 (0.210)
Trading assets divided by total assets	-0.648 ^b (0.249)	-0.661 ^b (0.250)	-1.449 (1.494)	-1.490 (1.490)	19.438 ^c (11.004)	19.137 ^c (10.955)
Non-interest income divided by operating income	-0.060 (0.093)	-0.059 (0.093)	0.119 (0.686)	0.112 (0.692)	5.636 ^b (2.165)	5.575 ^b (2.199)
Revenue source concentration	0.088 (0.147)	0.078 (0.147)	0.645 (0.933)	0.566 (0.947)	16.251 ^b (6.165)	16.186 ^b (6.321)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	293	293	293	293	295	295
R-squared	0.781	0.781	0.338	0.332	0.185	0.193
P-Value: Disclosure Variables = 0?	0.000	0.000	0.000	0.000	0.010	0.009
P-Value: Crisis = Non-Crisis?	0.947	0.760	0.011	0.027	0.071	0.082

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: Risk-adjusted market return is the annual average of weekly equity price returns divided by the standard deviation of those returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. BHC characteristics are from the Federal Reserve Y-9C reports. Disclosure information is from the BHCs' annual reports. Stock data are from CRSP. Disclosure leader is a dummy variable indicating that a BHC is the only BHC to disclose a particular data item in a given year. Aggregate disclosure index is the market risk disclosure index. First principal component is based on the eighteen individual data items that comprise the aggregate index. The sample consists of all U.S.-owned BHCs that have trading assets greater than U.S.\$1 billion (in 2013 dollars) at any time between 1994 and 2012, starting with the year that trading assets exceed U.S.\$500 million. The regressions include BHC fixed effects and year dummy variables. Residuals are clustered at the BHC level.

a Significant at the 1% level, b significant at the 5% level, c significant at the 10% level.

Table 7 – Disclosure and risk-adjusted returns' separate impact during the financial crisis

so this partition identifies not only BHCs with especially large trading portfolios but also BHCs for which trading represents a particularly large share of firm-wide activity.¹⁷

As the results in Table 8 illustrate, a statistically significant relationship exists between disclosure and risk-adjusted returns for both intense traders and other large traders, but this

relationship is more material for intense trading firms. In every case, the coefficient estimate for the intense traders is larger than that for the other large traders, though these differences

¹⁷ "Intense traders" have trading assets that range between 11 and 42% of total assets (with a median of 18%), as compared to a range of 0.1 to 12.0% (with a median of 1.6%) for the other large traders in the sample.

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Intense traders						
Disclosure leader	-0.061 (0.045)	-0.062 (0.045)	-0.191 (0.148)	-0.201 (0.148)	4.203 ^a (1.021)	4.000 ^a (0.980)
Aggregate disclosure index	0.015 ^a (0.003)		0.070 ^a (0.026)		0.436 ^c (0.224)	
First principal component		0.027 ^a (0.005)		0.123 ^a (0.044)		0.736 ^c (0.399)
Other large traders						
Disclosure leader	-0.035 (0.034)	-0.033 (0.033)	-0.094 (0.115)	-0.087 (0.113)	-0.557 (1.132)	-0.440 (1.138)
Aggregate disclosure index	0.008 ^a (0.002)		0.033 ^a (0.010)		0.308 ^c (0.169)	
First principal component		0.013 ^a (0.004)		0.054 ^a (0.018)		0.685 ^c (0.365)
BHC characteristics						
Log (asset size)	-0.058 ^a (0.019)	-0.059 ^a (0.019)	-0.387 ^a (0.117)	-0.388 ^a (0.120)	0.106 (0.963)	-0.100 (0.953)
Risk-weighted assets divided by total assets	-0.071 (0.101)	-0.065 (0.101)	0.001 (0.746)	0.020 (0.747)	7.146 ^c (3.858)	7.438 ^c (3.801)
Common equity divided by total assets	-0.011 ^b (0.005)	-0.011 ^b (0.005)	-0.088 ^a (0.032)	-0.089 ^a (0.033)	0.098 (0.198)	0.093 (0.194)
Trading assets divided by total assets	-0.580 ^b (0.242)	-0.583 ^b (0.244)	-1.734 (1.166)	-1.751 (1.164)	15.129 (11.727)	14.293 (11.593)
Non-interest income divided by operating income	-0.039 (0.099)	-0.036 (0.100)	0.277 (0.804)	0.288 (0.809)	5.982 ^b (2.293)	5.675 ^b (2.286)
Revenue source concentration	0.115 (0.153)	0.105 (0.152)	0.271 (0.976)	0.212 (0.970)	14.589 ^b (6.432)	14.315 ^b (6.567)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	293	293	293	293	295	295
R-squared	0.783	0.784	0.318	0.318	0.191	0.199
P-Value: Disclosure variables = 0?	0.000	0.000	0.003	0.003	0.002	0.001
P-Value: Intense = other large?	0.048	0.018	0.159	0.119	0.606	0.913

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: Risk-adjusted market return is the annual average of weekly equity price returns divided by the standard deviation of those returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. BHC characteristics are from the Federal Reserve Y-9C reports. Disclosure information is from the BHCs' annual reports. Stock data are from CRSP. Disclosure leader is a dummy variable indicating that a BHC is the only BHC to disclose a particular data item in a given year. Aggregate disclosure index is the market risk disclosure index. First principal component is based on the eighteen individual data items that comprise the aggregate index. The sample consists of all U.S.-owned BHCs that have trading assets greater than U.S.\$1 billion (in 2013 dollars) at any time between 1994 and 2012, starting with the year that trading assets exceed U.S.\$500 million. Intense traders are those with trading account assets greater than 10% of total assets and greater than U.S.\$20 billion in 2013 dollars, while other large traders are the remainder of the sample. The regressions include BHC fixed effects and year dummy variables. Residuals are clustered at the BHC level.

a Significant at the 1% level, b significant at the 5% level, c significant at the 10% level.

Table 8 – Disclosure and risk-adjusted returns by extent of trading activity

are not always significant (see the last row of the table). The coefficient estimates suggest that an increase of one standard deviation in the disclosure index metrics is associated with a 0.40 to 0.65 standard deviation increase in risk-adjusted returns for intense traders, but just a 0.20 to 0.45 standard deviation increase for other large trading BHCs. Further, the impact of being a disclosure leader is evident only for the intense traders: these BHCs have higher risk-adjusted trading returns, whereas there is no significant impact from being a disclosure leader among the other larger traders. Thus, the impact of disclosure on risk-adjusted returns is much stronger for those firms with a concentration in trading activity.

Robustness

One potential criticism of these findings is that the disclosure variables may be capturing unobserved characteristics of the BHCs' trading portfolios. For instance, information on VaR by risk type is clearly more relevant for BHCs with trading positions that span multiple risk factors (such as interest rates, exchange rates, equity prices, or commodities) than for those with simple portfolios. Multi-risk-factor portfolios that span riskier or less widely held risk exposures, such as commodities, could have different risk-return characteristics than portfolios composed of positions exposed primarily to interest rates, which are held in nearly all trading portfolios. Alternatively, BHCs that report more information about stress-testing may do so because they hold portfolios with "tail risk" that would not necessarily be realized in annual risk-adjusted returns (that is, risk-adjusted returns could be overstated because "tail risk" is not captured) but for which stress-testing is an important risk management tool. It could be, therefore, that the disclosure variables are capturing differences in underlying risk and return across BHCs rather than the impact of differential disclosure practices.

We performed a series of robustness checks to assess this concern. First, the specification includes BHC fixed effects, so any differences in risk-adjusted returns across BHCs that are related to permanent differences in disclosure should be absorbed by those controls. As a further check, we repeated the regressions including additional variables to control for the composition of BHCs' trading activity. In particular, BHC regulatory reports contain information on trading revenues derived from different types of risk factors, such as interest rates, exchange rates, equity prices, and commodity prices. Nearly all of the BHCs in the sample (91%) report trading revenue from interest rate and foreign exchange positions, but fewer report revenue from equity- or commodity-based positions (64% and 48%, respectively). We re-estimated the regression including dummy variables to capture the impact of these less common

trading risk factors. Regulatory reports also include information on the different types of securities held in the trading account, and we estimated a second alternative specification with variables that captured the composition of trading positions based on these data.¹⁸ Since this information is available only beginning in 1995, we excluded observations from 1994 from these estimates.

As a final test, we used a measure of the trading portfolio risk: the BHC's market risk capital requirement (scaled by trading account assets). As detailed above, minimum regulatory capital requirements for market risk are based on BHCs' internal VaR estimates. In that sense, they are related to the information disclosed in public financial statements about market risk exposure. Unfortunately, market risk capital data are available only beginning in 1998, when the market risk capital requirements were first imposed, and even in the years since then, some BHCs in our sample were not subject to the requirements in every sample year.¹⁹ Overall, the sample size is reduced by about a third when the market risk capital requirement is included as a control variable.

Results of the estimates including these three sets of additional control variables are reported in Tables 9A, 9B, and 9C, respectively. Including the additional control variables does not change the basic results. There continues to be a positive relationship between disclosure and risk-adjusted returns, though, as before, this relationship is stronger for the market-based measures than it is for accounting-based trading returns. The coefficients on the additional control variables are jointly statistically significant in most of the specifications, especially for the market-based return measures. The most consistent result is that higher market risk exposure, as measured by the ratio of market risk capital to trading assets, is associated with lower risk-adjusted returns (see Table 9C). The variables controlling for trading risk factors (commodity- and equity-based revenue) tend to have the least explanatory power, though the results suggest that equity-based revenue is associated with

¹⁸ The specification included variables reflecting the share of trading account assets composed of U.S. Treasury and agency securities, state and local government securities, mortgage-backed securities, other debt securities, trading positions held in foreign offices, revaluation gains on derivatives positions, and other trading account assets.

¹⁹ Only banks and BHCs with trading account assets exceeding U.S.\$1 billion or 10% of total assets are subject to the market risk capital requirement. In addition, supervisors have the option to exempt a bank or BHC that would otherwise be subject to the requirements if its trading risk is shown to be minimal, or to require a bank or BHC to be subject to the requirements if it has significant trading risk, even if it is below the numerical thresholds [Hendricks and Hirtle (1997)].

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.060 ^b (0.029)	-0.059 ^c (0.030)	-0.194 ^c (0.112)	-0.190 (0.114)	1.982 ^b (0.988)	2.038 ^b (0.957)
Aggregate disclosure index	0.010 ^a (0.003)		0.042 ^a (0.014)		0.363 ^b (0.155)	
First principal component		0.018 ^a (0.004)		0.076 ^a (0.025)		0.720 ^b (0.307)
Additional control variables						
Risk factor dummy variables						
Equity-based revenue	0.039 ^b (0.018)	0.041 ^b (0.017)	0.146 (0.144)	0.155 (0.143)	-1.323 ^c (0.731)	-1.250 ^c (0.714)
Commodity-based revenue	-0.018 (0.023)	-0.017 (0.023)	-0.013 (0.128)	-0.009 (0.129)	-0.397 (0.686)	-0.398 (0.694)
BHC characteristics						
Log (asset size)	-0.065 ^a (0.016)	-0.067 ^a (0.017)	-0.405 ^a (0.108)	-0.413 ^a (0.112)	-0.096 (0.769)	-0.250 (0.752)
Risk-weighted assets divided by total assets	-0.133 (0.098)	-0.122 (0.098)	-0.226 (0.702)	-0.178 (0.701)	8.450 ^b (3.672)	8.879 ^b (3.696)
Common equity divided by total assets	-0.010 ^c (0.005)	-0.010 ^c (0.005)	-0.083 ^b (0.031)	-0.082 ^b (0.032)	0.028 (0.205)	0.030 (0.202)
Trading assets divided by total assets	-0.633 ^a (0.235)	-0.638 ^a (0.237)	-1.956 (1.191)	-1.971 (1.192)	15.779 (11.595)	15.613 (11.582)
Non-interest income divided by operating income	-0.073 (0.091)	-0.074 (0.091)	0.114 (0.765)	0.109 (0.765)	6.330 ^a (2.096)	6.271 ^a (2.082)
Revenue source concentration	0.088 (0.148)	0.086 (0.147)	0.162 (0.915)	0.145 (0.909)	14.181 ^b (6.472)	14.193 ^b (6.579)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	293	293	293	293	295	295
R-squared	0.786	0.787	0.319	0.319	0.192	0.201
P-Value: disclosure variables = 0?	0.000	0.000	0.001	0.001	0.014	0.013

Table 9, Panel A – Robustness check – control for trading risk factors

higher risk-adjusted market returns (but lower risk-adjusted trading returns).

Risk-adjusted performance and market discipline

The finding that increased disclosure is associated with higher future risk-adjusted performance suggests that BHCs that disclose more information face a better risk-return trade-off. This finding is consistent with a broad interpretation of market discipline. Much discussion of market discipline has focused on the idea that market participants are concerned primarily about risk, so that enhanced disclosure serves mainly to discipline bank managers in terms of risk taking. However, it is reasonable to assume that investors, creditors, and other stakeholders might also be concerned with efficient risk taking and the relationship between risk and return. In this broader

interpretation, enhanced disclosure facilitates market discipline not merely by affecting risk but by making risk taking and trading activities more efficient and productive.

A related point is that the link between greater disclosure and better performance may not necessarily stem from the impact of market discipline as traditionally defined. Specifically, the same risk management systems that produce better risk-adjusted performance may also generate the information needed to make more detailed risk disclosures, which may be used by the bank as a public signal of its superior risk management abilities. Fang (2012) finds evidence broadly consistent with this hypothesis, as he documents a contemporaneous correlation between enhanced VaR disclosure and corporate governance characteristics. In this view, enhanced disclosure is a by-product of better performance, rather than a cause.

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.052 (0.031)	-0.051 (0.032)	-0.173 (0.114)	-0.169 (0.117)	1.318 (1.010)	1.320 (0.968)
Aggregate disclosure index	0.009 ^a (0.003)		0.048 ^a (0.015)		0.283 (0.175)	
First principal component		0.016 ^a (0.005)		0.086 ^a (0.028)		0.611 ^c (0.353)
Additional control variables						
Trading portfolio asset shares						
Treasury and agency securities	0.083 (0.059)	0.082 (0.059)	0.253 (0.319)	0.246 (0.318)	-0.178 (2.528)	-0.263 (2.458)
State and local government securities	0.160 ^c (0.087)	0.159 ^c (0.088)	0.769 (0.622)	0.766 (0.628)	-3.250 (3.131)	-3.564 (3.204)
Mortgage-backed securities	0.129 ^a (0.036)	0.127 ^a (0.038)	0.465 ^c (0.259)	0.457 ^c (0.268)	-1.750 (2.479)	-1.834 (2.376)
Other debt securities	0.081 (0.079)	0.085 (0.079)	0.995 (0.926)	1.017 (0.930)	-4.866 (3.011)	-4.643 (2.988)
Derivatives revaluation gains	0.050 ^c (0.027)	0.050 ^c (0.027)	0.066 (0.150)	0.064 (0.149)	-0.429 (1.258)	-0.492 (1.253)
BHC characteristics						
Log (asset size)	-0.070 ^a (0.017)	-0.071 ^a (0.017)	-0.469 ^a (0.111)	-0.476 ^a (0.116)	0.278 (1.013)	0.119 (0.985)
Risk-weighted assets divided by total assets	-0.075 (0.096)	-0.064 (0.095)	0.036 (0.687)	0.091 (0.686)	6.622 (4.097)	6.987 ^c (4.099)
Common equity divided by total assets	-0.012 ^b (0.005)	-0.012 ^b (0.005)	-0.102 ^b (0.040)	-0.102 ^b (0.040)	0.113 (0.246)	0.110 (0.242)
Trading assets divided by total assets	-0.534 ^b (0.254)	-0.543 ^b (0.254)	-2.407 ^c (1.236)	-2.451 ^c (1.225)	18.258 (13.203)	17.550 (13.146)
Non-interest income divided by operating income	-0.044 (0.078)	-0.045 (0.078)	0.344 (0.688)	0.339 (0.690)	4.651 ^c (2.481)	4.608 ^c (2.499)
Revenue source concentration	0.066 (0.140)	0.062 (0.139)	0.393 (0.968)	0.368 (0.967)	9.344 (6.364)	9.559 (6.505)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	280	280	280	280	282	282
R-squared	0.777	0.777	0.340	0.340	0.174	0.182
P-Value: disclosure variables = 0?	0.001	0.000	0.001	0.002	0.123	0.101

Table 9, Panel B – Robustness check – control for trading portfolio composition

That said, enhanced disclosure nonetheless provides market participants with important information about the bank that could influence investor actions, which seems consistent with a broad view of market discipline.

One last interesting finding concerns BHCs that are “first movers” in disclosure, in the sense of being the first to disclose a particular type of information. These firms appear to have lower future risk-adjusted market returns, but higher risk-adjusted trading returns. This finding suggests that there may

be learning costs for investors in assessing and putting into context new types of information about risk. To the extent that this is the case, policymakers advocating new and innovative disclosures should also consider the role that the public sector could play in educating investors and market analysts about these new disclosures. This outreach could reduce any negative market reaction to unfamiliar information and thus better align the incentives of firms and policymakers about enhanced disclosure.

Disclosure variables	Risk-adjusted market return		Alpha		Risk-adjusted trading return	
Disclosure leader	-0.109 ^a (0.024)	-0.104 ^a (0.026)	-0.390 ^a (0.132)	-0.350 ^a (0.125)	0.602 (1.584)	0.675 (1.473)
Aggregate disclosure index	0.010 ^b (0.004)		0.072 ^a (0.020)		0.297 (0.197)	
First principal component		0.018 ^b (0.007)		0.122 ^a (0.035)		0.578 (0.393)
Additional control variables						
Market Risk Exposure						
Market risk capital divided by trading assets	-0.085 ^b (0.035)	-0.080 ^b (0.035)	-0.468 ^b (0.195)	-0.434 ^b (0.197)	-2.554 (1.647)	-2.435 (1.569)
BHC Characteristics						
Log (asset size)	-0.082 ^a (0.029)	-0.082 ^a (0.030)	-0.629 ^a (0.164)	-0.623 ^a (0.169)	-0.206 (1.082)	-0.262 (1.061)
Risk-weighted assets divided by total assets	0.015 (0.099)	0.025 (0.101)	0.849 (0.709)	0.916 (0.720)	8.971 ^b (3.912)	9.337 ^b (3.883)
Common equity divided by total assets	-0.009 ^c (0.005)	-0.009 ^c (0.005)	-0.104 ^a (0.034)	-0.103 ^a (0.035)	0.112 (0.263)	0.110 (0.259)
Trading assets divided by total assets	-0.799 ^b (0.336)	-0.795 ^b (0.337)	-3.038 ^c (1.712)	-3.004 ^c (1.715)	11.608 (17.558)	11.449 (17.517)
Non-interest income divided by operating income	-0.108 (0.101)	-0.106 (0.101)	0.084 (0.791)	0.096 (0.795)	4.455 ^b (1.847)	4.523 ^b (1.888)
Revenue source concentration	0.020 (0.186)	0.010 (0.186)	0.871 (1.213)	0.793 (1.217)	18.829 ^b (7.155)	18.905 ^b (7.264)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BHC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	198	198	198	198	199	199
R-squared	0.779	0.779	0.332	0.329	0.216	0.220
P-Value: Disclosure Variables = 0?	0.000	0.000	0.000	0.000	0.175	0.168

Sources: Federal Reserve Board, Consolidated Financial Statements of Bank Holding Companies (FR Y-9C data); Center for Research in Security Prices (CRSP); Securities and Exchange Commission EDGAR database; company websites.

Notes: Risk-adjusted market return is the annual average of weekly equity price returns divided by the standard deviation of those returns. Alpha is the intercept term from a three-factor market return model using Fama-French factors. Risk-adjusted trading return is annual trading revenue divided by the annual standard deviation of quarterly trading revenue. BHC characteristics are from the Federal Reserve Y-9C reports. Disclosure information is from the BHCs' annual reports. Stock data are from CRSP. Disclosure leader is a dummy variable indicating that a BHC is the only BHC to disclose a particular data item in a given year. Aggregate disclosure index is the market risk disclosure index. First principal component is based on the eighteen individual data items that comprise the aggregate index. The sample consists of all U.S.-owned BHCs that have trading assets greater than U.S.\$1 billion (in 2013 dollars) at any time between 1994 and 2012, starting with the year that trading assets exceed U.S.\$500 million. The regressions include BHC fixed effects and year dummy variables. Residuals are clustered at the BHC level.

a Significant at the 1% level, b significant at the 5% level, c significant at the 10% level.

Table 9, Panel C – Robustness check – control for market risk exposure

SUMMARY AND CONCLUSION

Disclosure plays an important role in market discipline because market participants need to have meaningful information on which to base their judgments of risk and performance. Disclosure is particularly important in the banking industry, given that outsiders generally view banks as being opaque. As a result, banking supervisors and other public sector officials have encouraged banking companies to engage in enhanced disclosure, particularly of forward-looking estimates of risk. This article aims to assess whether these kinds of disclosures provide useful information to market participants that can help foster market discipline.

In particular, the article examines disclosures related to market risk in trading and market-making activities. The key variable is an index of market risk disclosure that captures the amount of market risk information banking companies disclose in their annual reports. The index is constructed for a sample of BHCs with significant trading activities over the years 1994 to 2012. The article estimates the extent to which this disclosure affects future risk-adjusted returns on trading activities and returns for the BHC overall, as proxied by the firm's equity price behavior.

The main findings are that increases in disclosure are associated with higher risk-adjusted returns, both for trading activities and for the firm overall. These results are economically meaningful as well as statistically significant. The findings are robust to alternative specifications that include additional controls for the composition of the BHCs' trading portfolios and the sources of trading revenue, and are stronger for BHCs whose trading activity represents a larger share of firm-wide activity. The results are not driven by the 2007-09 financial crisis and, in fact, the relationship between disclosure and risk-adjusted performance appears to be significantly weaker during the crisis period. Overall, the results suggest that as disclosure increases, BHCs experience an improved risk-return trade-off.

REFERENCES

- Acharya, V. V., D. Anginer, and A. J. Warburton, 2014, "The end of market discipline? Investor expectations of implicit government guarantees," available at <http://ssrn.com/abstract=1961656>
- Balasubramanian, B., and K. B. Cyree, 2011, "Market discipline of banks: why are yield spreads on bank-issued subordinated notes and debentures not sensitive to bank risks?" *Journal of Banking and Finance* 35:1, 21-35
- Barth, M. E., Y. Konchitchki, and W. R. Landsman, 2013, "Cost of capital and earnings transparency," *Journal of Accounting and Economics* 55:2-3, 206-224
- Basel Committee on Banking Supervision, 2004, "International convergence of capital measurement and capital standards: a revised framework," Bank for International Settlements
- Basel Committee on Banking Supervision, 2015, "Standards: revised Pillar 3 disclosure requirements," Bank for International Settlements, available at <http://www.bis.org/bcbs/publ/d309.pdf>.
- Baumann, U., and E. Nier, 2004, "Disclosure, volatility, and transparency: an empirical investigation into the value of bank disclosure," *Federal Reserve Bank of New York Economic Policy Review* 10:2, 31-45
- Berkowitz, J., and J. O'Brien, 2002, "How accurate are Value-at-Risk models at commercial banks?" *Journal of Finance* 57:3, 1093-1111
- Bliss, R. R., and M. J. Flannery, 2002, "Market discipline in the governance of U.S. bank holding companies: monitoring vs. influencing," *European Finance Review* 6:3, 361-395
- Bushee, B. J., and C. P. Noe, 2000, "Corporate disclosure practices, institutional investors, and stock return volatility," *Journal of Accounting Research* 38, supplement, 171-202
- Bushman, R. M., and C. D. Williams, 2012, "Accounting discretion, loan loss provisioning, and discipline of banks' risk-taking," *Journal of Accounting and Economics* 54:1, 1-18
- Covitz, D. M., D. Hancock, and M. L. Kwast, 2004a, "A reconsideration of the risk sensitivity of U.S. banking organization subordinated debt spreads: a sample selection approach," *Federal Reserve Bank of New York Economic Policy Review* 10:2, 73-92
- Covitz, D. M., D. Hancock, and M. L. Kwast, 2004b, "Market discipline in banking reconsidered: the roles of funding manager decisions and deposit insurance reform," *Board of Governors of the Federal Reserve System Finance and Economics Discussion Series*, no. 2004-53, August
- Cumming, C. M., and B. J. Hirtle, 2001, "The challenges of risk management in diversified financial companies," *Federal Reserve Bank of New York Economic Policy Review* 7:1, 1-17
- Ellul, A., and V. Yerramilli, 2013, "Stronger risk controls, lower risks: evidence from U.S. bank holding companies," *Journal of Finance* 68:5, 1757-1803
- Eugster, F., and A. F. Wagner, 2011 (revised 2015), "When and how is voluntary disclosure quality reflected in equity prices?" *Swiss Finance Institute Research Paper Series*, no. 11-25
- Fang, X. 2012, "Informativeness of value-at-risk disclosure in the banking industry," Unpublished paper, available at <http://ssrn.com/abstract=1982936>
- Flannery, M. J., 2001, "The faces of 'market discipline,'" *Journal of Financial Services Research* 20:2-3, 107-119
- Flannery, M. J., S. H. Kwan, and M. Nimalendran, 2004, "Market evidence on the opaqueness of banking firms' assets," *Journal of Financial Economics* 71:3, 419-460
- Goyal, V. K., 2005, "Market discipline of bank risk: evidence from subordinated debt contracts," *Journal of Financial Intermediation* 14:3, 318-350
- Hendricks, D., and B. Hirtle, 1997, "Bank capital requirements for market risk: the internal models approach," *Federal Reserve Bank of New York Economic Policy Review* 3:4, 1-12
- Hirtle, B., 2003, "What market risk capital reporting tells us about bank risk," *Federal Reserve Bank of New York Economic Policy Review* 9:3, 37-54
- Hirtle, B., 2006, "Stock market reaction to financial statement certification by bank holding company CEOs," *Journal of Money, Credit, and Banking* 38:5, 1263-1291
- Hirtle, B., 2007, (revised 2015), "Public disclosure, risk, and performance at bank holding companies," *Federal Reserve Bank of New York Staff Reports*, no. 293
- Iannotta, G., 2006, "Testing for opaqueness in the European banking industry: evidence from bond rating spreads," *Journal of Financial Services Research* 30:3, 287-309
- Jagtiani, J., G. Kaufman, and C. Lemieux, 2002, "The effect of credit risk on bank and bank holding company bond yields: evidence from the post-FDICIA period," *Journal of Financial Research* 25:4, 559-575
- Jorion, P., 2002, "How informative are value-at-risk disclosures?" *Accounting Review* 77:4, 911-931

- Jorion, P., 2006, *Value at risk: the new benchmark for managing financial risk*, third edition, McGraw-Hill
- Kwan, S., 2004, "Testing the strong-form of market discipline: the effects of public market signals on bank risk," Federal Reserve Bank of San Francisco Working Paper, no. 2004-19, May
- Leuz, C., and R. E. Verrecchia, 2000, "The economic consequences of increased disclosure," *Journal of Accounting Research* 38:supplement, 91-124
- Linsmeier, T. J., D. B. Thornton, M. Venkatachalam, and M. Welker, 2002, "The effect of mandated market risk disclosures on trading volume sensitivity to interest rate, exchange rate, and commodity price movements," *The Accounting Review* 77:2, 343-377
- Liu, C., S. G. Ryan, and H. Tan, 2004, "How banks' value-at-risk disclosures predict their total and priced risk: effects of bank technical sophistication and learning over time," *Review of Accounting Studies* 9:2, 265-294
- Meyers, S. C., and R. G. Rajan, 1998, "The paradox of liquidity," *Quarterly Journal of Economics* 113:3, 733-771
- Moody's Investors Service, 2006, "Risk disclosures of banks and securities firms," May 12
- Morgan, D. P., 2002, "Rating banks: risk and uncertainty in an opaque industry," *American Economic Review* 92:4, 874-888
- Morgan, D. P., and K. J. Stiroh, 2001, "Market discipline of banks: the asset test," *Journal of Financial Services Research* 20:2-3, 195-208
- Multidisciplinary Working Group on Enhanced Disclosure, 2001, "Final report to the Basel Committee on Banking Supervision, Committee on the Global Financial System of the G-10 Central Banks, International Association of Insurance Supervisors, and International Organization of Securities Commissions," Bank for International Settlements
- Nier, E., and U. Baumann, 2006, "Market discipline, disclosure, and moral hazard in banking," *Journal of Financial Intermediation* 15:3, 332-361
- Pérignon, C., and D. R. Smith, 2010, "The level and quality of value-at-risk disclosure by commercial banks," *Journal of Banking and Finance* 34:2, 362-377
- Roulstone, D. T., 1999, "Effect of SEC financial reporting release no. 48 on derivative and market risk disclosures," *Accounting Horizons* 13:4, 343-363
- Santos, J., 2014, "Evidence from the bond market on banks' 'too-big-to-fail' subsidy," Federal Reserve Bank of New York Economic Policy Review 20:2, 29-39
- Stiroh, K., 2006, "New evidence on the determinants of bank-specific risk," *Journal of Financial Services Research* 30:3, 237-263
- U.S. Securities and Exchange Commission, 1997, "Disclosure of accounting policies for derivatives financial instruments and derivatives commodity instruments and disclosure of quantitative and qualitative information about market risk inherent in derivative financial instruments, other financial instruments, and derivative commodity instruments," Financial Report Release no. 48. January 31, <http://www.sec.gov/rules/final/33-7386.txt>
- Zer, I., 2013, (revised 2015), "Disclosure practices and option-implied probability of default," unpublished paper, available at <http://ssrn.com/abstract=2335717>

What do New Forms of Finance Mean for Emerging Markets?

M. S. Mohanty – Head of Economics and Financial Markets for Asia and the Pacific, Bank for International Settlements¹

Abstract

The size and the nature of financial intermediation matters not only from the perspective of the risk exposure of financial institutions but also in terms of the cost of credit and the effectiveness with which monetary policy is transmitted to the economy. This paper looks at how the forms of finance have changed in major emerging market economies (EMEs) in recent years and what this means for monetary and financial stability in these economies. It argues that the greater access of households to bank credit and of EME corporations to domestic and external securities debt markets is a double-edged sword. On the one hand, it has helped foster financial development in EMEs, diversifying funding sources, and reducing credit risk concentration. On the other hand, it has contributed to increasing risks and vulnerabilities – as many recent

financial market turbulences illustrated. These developments pose challenges to EM monetary authorities in containing monetary and financial stability risks as well as designing appropriate response.

¹ I am thankful to Claudio Borio, Torsten Ehlers, Masazumi Hattori, Emanuel Kohlscheen, Phurichai Rungcharoenkitkul, Hyun Song Shin, Elod Takáts, Philip Turner, and Agustín Villar for comments and Agne Subelyte for statistical assistance. The paper is based on “What do new forms of financial mean for EM central banks? An Overview,” published in the BIS Papers no 83. It draws heavily on the proceedings, survey information, and the papers prepared for a meeting of senior central bank officials at the Bank for International Settlements in 2015. The views expressed in this paper are those of the author and do not necessarily reflect the views of the BIS or the central banks represented at the meeting.

INTRODUCTION

The 2008 Global Financial Crisis (GFC) has focused attention of academics and policymakers alike on the optimal model of financial intermediation for an economy. There is no consensus on this issue. Models of financial intermediation are likely to vary across countries and regions, depending, among other things, on their economic history and stages of financial development. That said, the crisis challenged the traditional views about the merits of financial intermediation models where either banks or markets play a dominant role. Instead, it highlighted the key role played by financial intermediaries and their funding models in the causation and the propagation of financial collapses. The size and the nature of financial intermediation matter not only for the risk exposure of financial institutions but also the cost of credit and the effectiveness with which monetary policy is transmitted to the economy.

This paper looks at how the forms of finance have changed in major emerging market economies (EMEs) in recent years and what this means for monetary and financial stability in these economies. One important trend emerging over the past decade is that while the share of credit intermediated by the EME banking systems has fallen, banks have been allocating a larger fraction of that credit to households, often increasing their non-core liabilities to finance such lending. On the other hand, the non-financial corporations – the traditional clients of banks – have increasingly moved to the offshore bond markets. At the same time, there have been major changes to EMEs' local currency bond markets, which have become not only deeper but increasingly internationalized because of greater openness and rising foreign participation.

The greater access of households to bank credit and of EME corporations to domestic and external securities debt markets is a double-edged sword. On the one hand, it has helped foster financial development in EMEs, diversifying funding sources and reducing credit risk concentration. On the other hand, it has contributed to increasing risks and vulnerabilities – as many recent financial market turbulences illustrated [BIS (2015, 2016)]. Domestic bond markets now react more strongly to global forces. Larger foreign currency debt has made many companies more vulnerable to exchange rate shocks. Credit cycles have also become more pronounced. These developments pose challenges to EM monetary authorities in containing monetary and financial stability risks, raising questions about the appropriate instruments required to stabilize the economy [Sobrun and Turner (2015)].

THE ROLE OF BANKS AND BOND MARKETS

Historically, banks have played a central role in the financial systems of EMEs by allocating domestic savings, transforming the maturity of financial claims, and intermediating international capital flows. However, a series of banking crises in the 1980s and 1990s raised questions about the merits of bank-based financial intermediation and triggered initiatives aimed at diversifying financial systems. The 2008 crisis was a major turning point in many countries. To capture this shift, Table 1 provides broad patterns of financial intermediation in EMEs just before the recent financial crisis and five years following it, as well as in the mid-2000s [see Ehlers and Villar (2015) for more details].

Recent rapid credit growth in EMEs

As the first three panels of Table 1 show, over the 2004–13 period as a whole, total credit extended to the non-financial private sector of EMEs by banks and bond markets taken together (through domestic and international channels) has risen markedly in many countries as a percentage of GDP. This trend started in mid-2000 but picked up particularly sharply after the onset of the GFC. The growth in total credit has been faster in economies that are more financially open and that have tied their exchange rates to the currencies of advanced economies than those that are less open and allow greater exchange rate flexibility. This is particularly true in the case of Hong Kong SAR, with its linked exchange rate system and highly open capital accounts (as well as its role as an international financial center), but also in China, even with its relatively closed capital markets. It is not surprising, therefore, that total credit as a percentage of GDP has grown at a much faster rate in Asian economies than that in other EME regions.

The last three panels of the Table 1 show how much of this credit is intermediated by the banking system. It is clear that banks continue to be the main source of credit in EMEs. However, there are significant changes to financial intermediation in many countries. For instance, over the past decade, the share of credit intermediated by banks has fallen significantly in China, Chile, Hong Kong SAR, Hungary, India, and Korea. Again, this trend is most visible in emerging Asia but less so prominent in other regions.

In what appears to be a general trend across EMEs, the sharp growth in total bank assets has coincided with a rapid increase in bank lending to households, which has been partially offset by a general decline in banks' debt securities holdings and loans to non-financial corporations. This is in sharp contrast to the experience in the 1980s and 1990s when EME banks

	Total credit to non-financial private sector (as a share of nominal GDP) ¹			Bank credit to non-financial private sector (as a share of total credit to non-financial private sector)		
	2004	2007	2013	2004	2007	2013
Emerging Asia²	97	98	130	91	89	86
China	124	118	181	96	91	75
Hong Kong SAR	164	183	261	90	83	81
India	38	50	59	96	93	92
Indonesia	29	28	41	87	93	89
Korea	139	160	185	74	76	67
Malaysia	131	114	135	96	96	100
Philippines ³	41	34	41			
Singapore	101	97	139	91	84	87
Thailand	109	97	127	97	98	97
Latin America²	34	40	55	64	71	73
Argentina	13	15	16	64	76	94
Brazil	34	45	76	88	93	93
Chile	85	94	123	71	71	63
Colombia ³	26	33	44			
Mexico	23	27	34	34	45	41
Peru ²	20	23	37			
Central and eastern Europe²	64	82	99	49	55	53
Czech Republic	69	78	98	39	52	55
Hungary	81	112	121	48	47	39
Poland	43	57	79	59	65	65
Other EMEs²	49	63	64	81	82	88

¹ BIS calculations of total credit to private non-financial sector.

² Regional aggregates are simple averages.

³ Total credit to the private sector estimate based on domestic bank credit to private sector (IMF, International Financial Statistics, line 22d), plus cross-border loans to the non-bank sector, less government exposure from BIS consolidated banking statistics, plus international debt securities issued by non-financial corporations.

Sources: IMF, International Financial Statistics; national data; BIS international banking statistics; BIS securities statistics.

Table 1 – Private sector credit and domestic bank lending in EMEs

followed what is called “one-way financial intermediation” in which they mobilized household deposits to lend to the private corporate sector or the government [Mohanty and Turner (2008)]. By 2000s, however, this picture had changed substantially. For instance, between 2004 and 2013, the share of credit going to households in total bank credit rose in the range of 10-20 percentage points in Argentina, Brazil, Czech Republic, India, Indonesia, Israel, Poland, Russia, Thailand, and Turkey.

In many of these countries, this share now stands at around 30-50%.

Another important aspect of recent changes in credit intermediation relates to the role of international bank lending in EMEs, which has been a major source of financial stress in many economies, as demonstrated clearly during the Asian financial crisis of 1997-98. In more recent period, however, international bank lending has declined significantly, as such lending has been increasingly replaced by financing through international debt securities. As a result, international bank credit (cross-border claims plus local claims of international banks) as a percentage of total domestic bank credit to the non-bank sector has shrunk in Latin America, where it fell from 50% in 2005 to 30% in 2013, as well as in central and eastern Europe, where it went down from 100% in 2008 to 75% in 2013.

Several forces appear to be at work. In most EMEs, an important factor has been easy domestic monetary conditions, which boosted both the demand for and supply of credit. In many commodity-exporting countries, these domestic conditions interacted with sustained improvements in terms of trade up to 2014. However, the factor that is most common across countries is exceptionally easy global financial conditions. One direct channel appears to work through bank deposits. Since most capital inflows ultimately end up on bank balance sheets, they tend to increase banks’ lending capacity. Second, in several countries banks also funded a significant part of their credit growth by directly accessing international debt markets where terms for borrowing were very easy. Finally, in some countries large capital inflows led banks to lower their lending standards, particularly under the threat of competition as their major corporate clients moved to offshore markets. In other words, banks responded to large non-financial firms’ global search for yield by easing their lending terms.

It is important to note that, despite strong credit growth, in several countries, banking system leverage declined over the past decade (left-hand panel of Figure 1). Interestingly, in many countries the decline in banks measured leverage coincided with an increase in the loan-to-deposit ratio – that is, banks expanded their other liabilities to fund the additional loans. As Adrian and Shin (2010) have shown, monetary policy works through changes in asset prices and the yield curve that affect banks’ profitability and lending behavior (the so-called “risk-taking” channel of monetary policy). To the extent that easy domestic and external financial conditions boost asset prices, they tend to increase the value of equity when bank portfolios are marked-to-market. Banks facing capital or

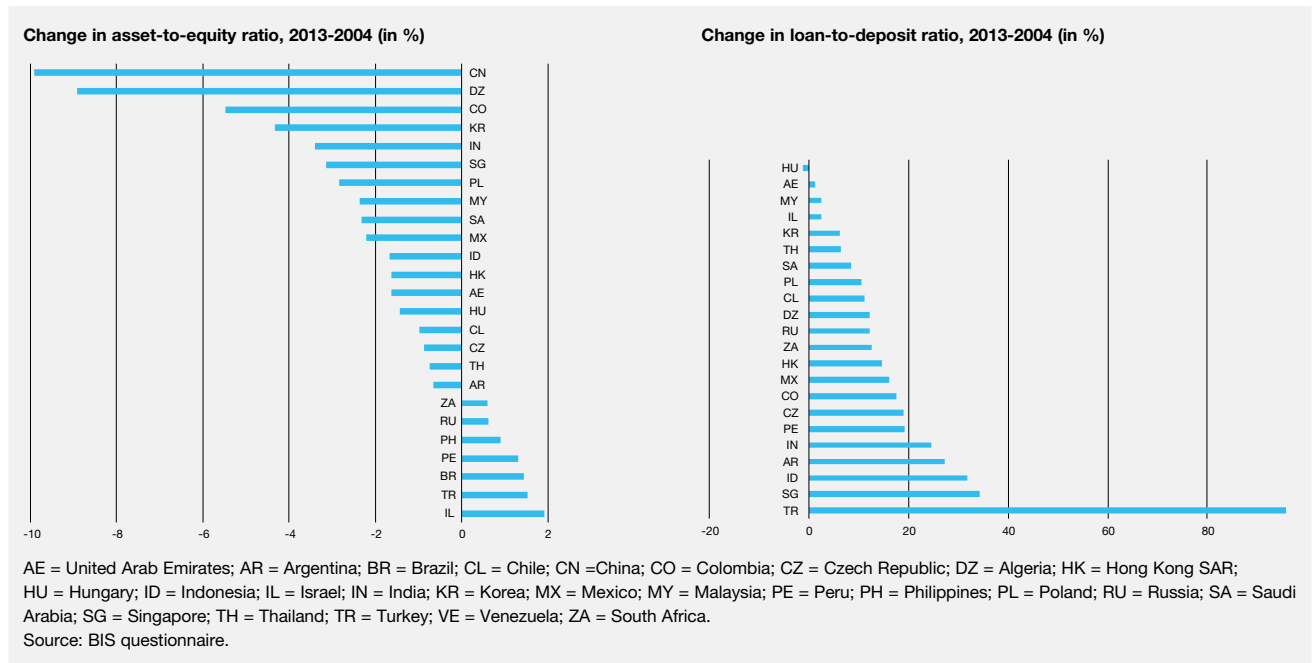


Figure 1 – Banking sector leverage and loan-to-deposit ratio

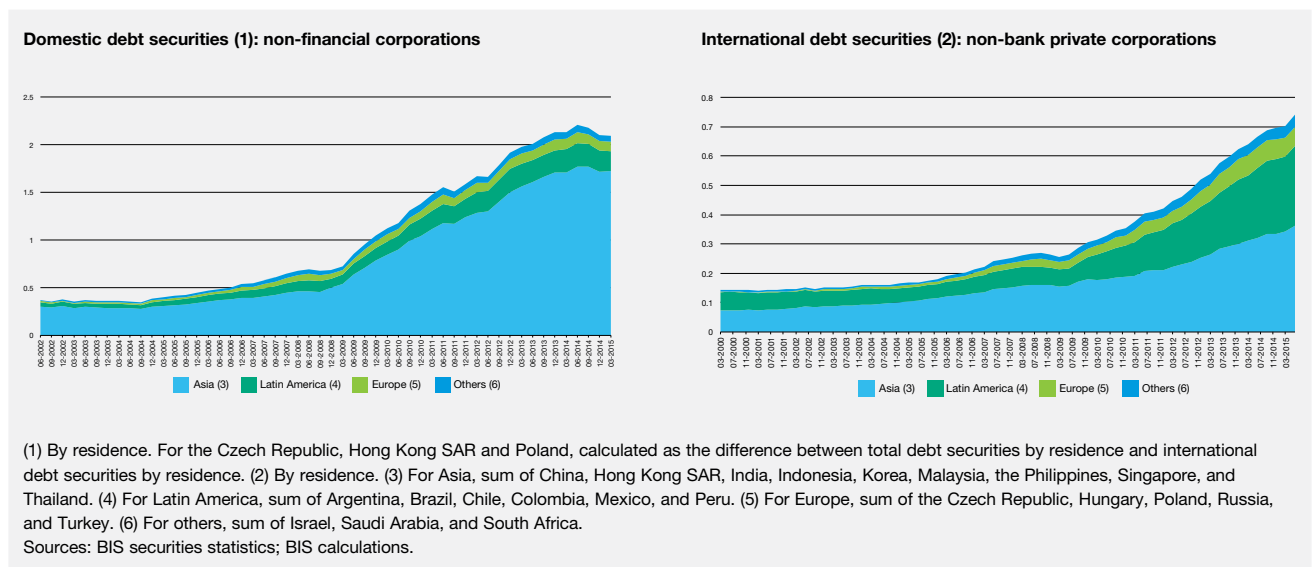


Figure 2 – Domestic and international debt securities (Amounts outstanding, in trillions of U.S.\$)

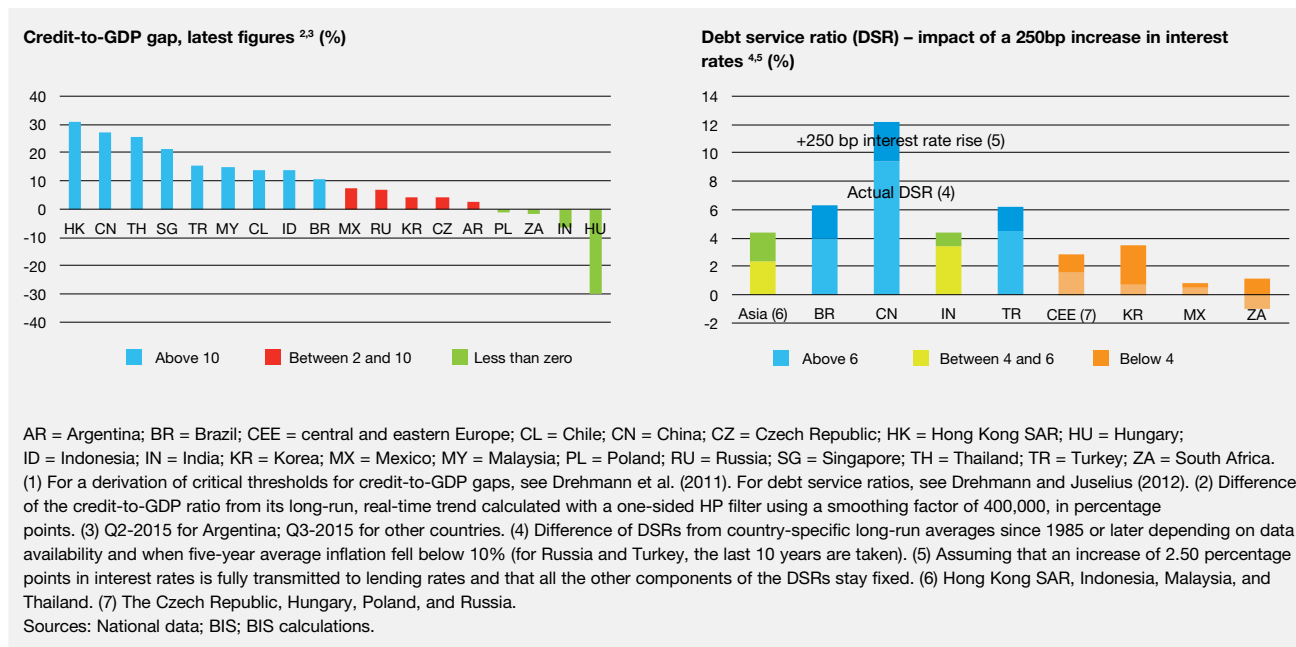


Figure 3 – Early warning indicators for domestic banking crises signal risk ahead¹

value-at-risk (VaR) constraints thus tend to lend more without having to raise additional capital. Asset price booms thus make bank credit highly procyclical even without a rise in the measured leverage ratio.

Rise in debt securities issuance

A clear trend across many EMEs over the past decade is the expansion of debt securities as a funding vehicle. Thus, the combined issuance of debt securities by entities located in EMEs – governments, financial institutions and non-financial corporations together – has grown more than sixfold over the past decade, from U.S.\$2.5 trillion in 2002 to U.S.\$14 trillion in 2014 [Hattori and Takáts (2015)]. Although the issuance of domestic debt securities, which is usually denominated in local currencies, constitutes the largest share of activity (about 80%), the issuance of international debt securities has also risen significantly. Figure 2 provides a snap shot view of debt issuance by EME non-financial corporations in both domestic and international markets. As Figure 2 shows, not only has debt issuance by EME firms risen rapidly after the GFC but a large part of that issuance has moved to offshore markets. Asian and Latin American firms have been very active in issuing debt securities in the international debt markets.

Within this big picture, cross-country differences remain large. In many countries, domestic bond markets still largely consist of government debt securities. With a few exceptions (e.g., Hong Kong SAR and Malaysia), corporate bond markets remain relatively underdeveloped, constraining the supply of long-term finance. At the same time, cyclical factors, such as very low global interest rates, have attracted EME corporations to international debt markets. The preference for issuing debt in international markets may reflect a rational decision by EME corporations to access cheaper funds in deeper international capital markets than in more expensive and less liquid domestic markets.

It is well recognized that larger domestic bond markets offer many benefits to EMEs, not least by fostering financial development. Bond markets help to diversify the sources of funding, avoid credit risk concentration in the banking sector, and enhance opportunities for long-term investment. Moreover, debt issuance by EME borrowers in their own currencies reduces currency mismatches. In many countries, therefore, the recent increase in debt issuance is a direct consequence of EME authorities' concerted efforts to develop local currency bond markets and limit banks' exposure to duration mismatches.

IMPLICATIONS FOR FINANCIAL STABILITY

What do the recent changes in financial intermediation imply for financial stability in EMEs? In assessing the significance of these developments, the underlying factors are likely to be more important than the trends themselves. Clearly, financial deepening plays an important role in the trend increase in credit-to-GDP ratio. Yet, there is increasing concerns that the recent credit growth in EMEs may prove more cyclical than structural. One measure of this vulnerability is the extent to which credit-to-GDP ratio has deviated from its long term trend. Borio and Low (2002) have shown that the credit gap measures are a fairly reliable indicator of financial vulnerability with a lead time of about two years.

That said, accurately estimating the long term value of the credit-to-GDP ratio remains a difficult task in EMEs because of their rapid structural changes. Hence, estimates of credit gaps tend to be less precise for EMEs than those for the mature economies. Keeping this caveat in mind, estimates of credit gaps suggests that bank credit-to-GDP ratio has been generally above its long term trend in many EMEs (left hand panel of Figure 3). For instance, by the end of 2015, the estimated credit gap had widened to over 10 percentage points in several countries – a threshold at which regulators, under the Basel III framework, would ask the banks to hold counter-cyclical capital buffers at the maximum level.

Risks to the banking system

A key question is how these developments affect risks in the EME banking system. There are two sources of risks: one coming from bank borrowers and the other stemming from banks' own balance sheets. Historically, the quality of bank loans and the potential default rates are closely correlated with the debt servicing costs of bank borrowers. For instance, Drehmann and Juselius (2014) show that debt service ratios tend to peak just before strains materialize in the banking system, and, more often than not, rising interest rates prove to be the turning points. The right hand panel of Figure 3 reports Ehlers and Villar's (2015) estimates of the impact of a cumulative 250 points increase in the interest rate that could be triggered by a possible normalization of U.S. monetary policy on debt servicing burdens in a sample of EMEs. The figure shows that, in a number of economies, such an interest rate shock would push debt servicing costs to high levels, exceeding the long-term 6% ratio at which financial stability concerns become important. In many EMEs, therefore, recent rapid growth of credit has prompted authorities to tighten macro-prudential controls as a preventive measure to contain risks.

The exposure of banks to foreign exchange risk through their customer account could also be sizeable. While banks may be hedged against currency risk, their borrowers may not. In some EMEs, borrowers still expect an appreciation of the local currency, increasing incentives for unhedged foreign currency borrowing. One concern in countries with more developed foreign exchange markets seems to be the speculative positions of domestic institutional investors, which can have an influence on the dynamics of exchange rates. Extensive use of hedges against currency appreciation can itself generate appreciation pressures. For instance, buying FX swaps or forwards raises the expected future price of a currency, which feeds back into current market prices. Both investors and borrowers could speculate on currency appreciation, leading to large exposures and potentially disruptive shocks if currency movements were to reverse.

As regards banks' own balance sheets, the direct exposure of banks to interest rate risks is probably limited, as banks manage these risks as part of their routine business. However, bank borrowing from the debt and wholesale deposit markets can still lead to potential funding problems. Aggregate loan-to-deposit ratios in some regions (in Asia and Latin America, in particular) have increased from previously moderate values, even though they are generally below one. In particular, as banks funded a greater part of their incremental lending by mobilizing corporate deposits, they could be vulnerable to tighter external funding conditions, triggering an outflow of deposits from the banking system.

Shin (2013) shows that when banks' assets are growing at a faster rate than their core deposits (such as retail deposits), they tend to increase their non-core liabilities to finance the additional lending growth. In other words, banks shift their funding to more volatile wholesale markets (such as corporate deposits) and international debt markets. Borio et al. (2011) have shown that, historically, EME credit cycles have been synchronized with cross-border credit cycles. In typical boom periods, cross-border credit tends to grow faster than overall credit, with banks accessing wholesale dollar funding markets to finance new asset growth. The process reverses itself, as higher U.S. interest rates cause large-scale unwinding of dollar borrowings and a widespread slowdown of credit in EMEs.

The funding pattern of banks in EMEs have undergone changes. For instance, in Asia, the average share of retail deposits in total assets fell from 37% in 2004 to 31% in 2013. Some countries (e.g., the Philippines and Thailand) have witnessed more rapid declines. In Latin America, the average ratio was generally much smaller (23% in 2004) and fell further (20%).

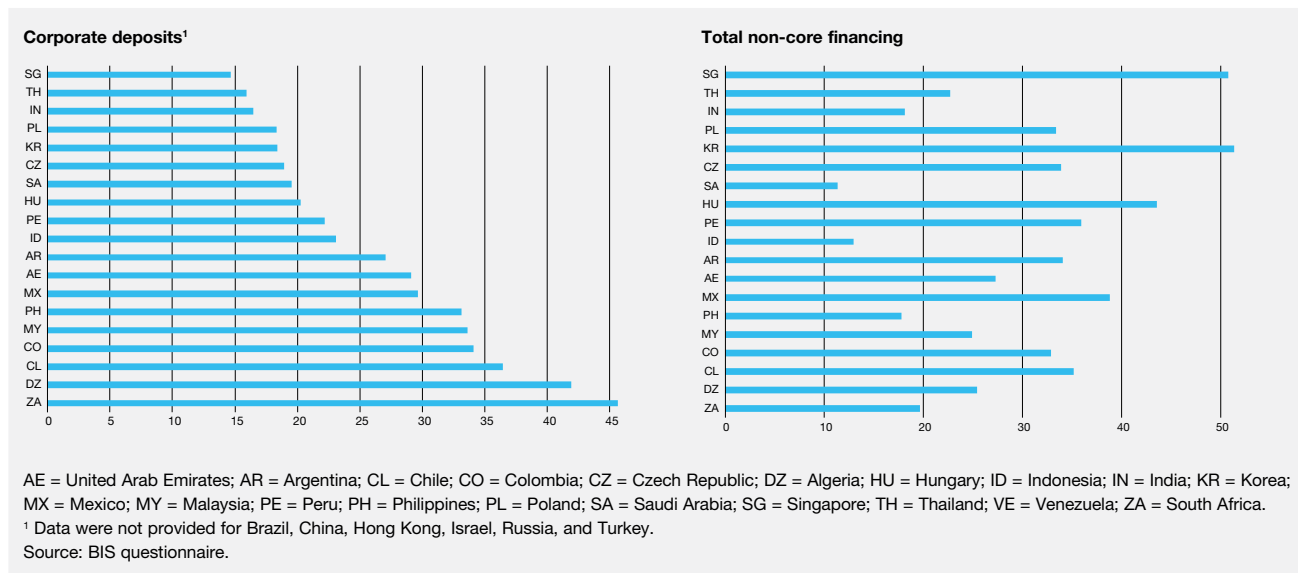


Figure 4 – Sources of funding of banks (end-2013, as a percentage of total assets)

Banks financed a large part of their asset growth by tapping into corporate deposits and increasing debt and other non-equity financing. As the left panel of Figure 4 shows, in a number of countries the share of assets funded by corporate deposits stood at more than 30% in 2013. The median contribution of corporate deposits to total debt liability growth across EMEs as a whole increased from 24% in 2004-2009 (just before the introduction of unconventional monetary policy by advanced economy central banks) to 31% during 2009-13. As the right hand panel of the figure shows, funding of asset growth through non-core liabilities has also been quite strong in EMEs, in some cases exceeding 40% or so.

That said, one notable aspect of EME banking systems has been its increased resilience to external shocks due to the shift by international banks to a subsidiaries-based business model. Because these subsidiaries may tap local deposits for asset growth, they can help reduce currency mismatches in the banking system. If these banks enjoy a measure of protection through access to national deposit insurance schemes, or have a large number of retail customers, subsidiaries would be the preferred model for the host country because their capital could be segregated from the parent bank. In the event of a crisis, however, host-country taxpayers would have to foot the bill – even for foreign banks – although the very anticipation of this risk can prompt the host country regulator to ring-fence assets of subsidiaries. Cross-border banking within a region

(regionalization) heightens the exposure to regional macroeconomic risks. Cooperation between home and host supervisors will be, therefore, essential in any attempt to reduce risks and limit the potentially damaging implications of regulatory arbitrage.

Vulnerability from bond financing

Despite obvious benefits, increased international debt issuance by corporates also creates risks. To the extent that cyclical factors dominate, EME borrowers are exposed to a reversal of easy global financing conditions. Such an eventuality increases the risk that dollar bond issuance may dry up in future; at the same time, the corporate sector would become vulnerable to higher debt repayment and refinancing risks. The interaction between dollar liabilities and large currency depreciation can contribute to magnifying these risks. The January and February 2016 currency market turmoil illustrated this risk quite well, as international bond markets virtually shut down for many EMEs [BIS (2016)].

Figure 5 shows several key parameters of corporate finance and their relationship with debt issuance. For EME corporates as a whole, the stock of debt has continued to rise since 2008. With stagnant earnings before interest and taxes (EBIT), this has meant a steady increase in leverage (upper lefthand panel). Naturally, firms that have issued debt have witnessed a more rapid rise in leverage than those did not (upper

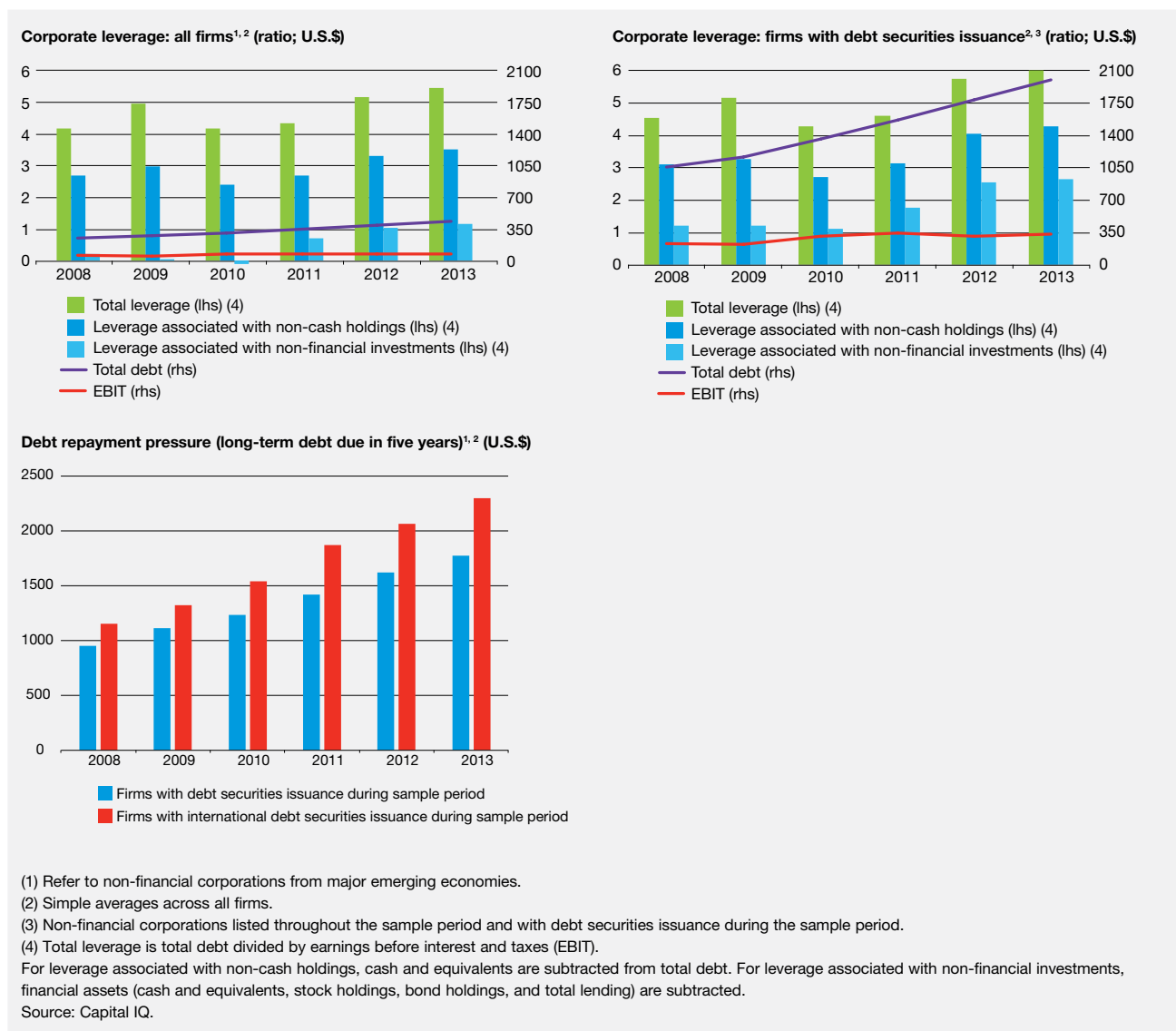


Figure 5 – Corporate leverage and repayment pressure

right-hand panel). However, firms issuing debt have also made larger non-financial investment, implying that they have used these additional resources to build tangible capital stocks.

Risks from higher debt arise from several sources. One is the risk of overinvestment that could lower the rate of return on investment and therefore profitability. There is evidence that the return on assets of EME corporations has fallen recently and the price-to-earnings ratio has risen, suggesting a risk to funding conditions should equity valuations suffer from higher interest rates [Hattori and Takáts (2015) and Chui et al. (2016)].

Second, higher debt repayment burden could reduce future investment prospects and expose highly leveraged firms to potential funding and debt rollover risks (lower left-hand panel of Figure 5). A few oil-exporting countries have been under severe stress because of the recent collapse of oil prices. Given that many oil firms have accumulated substantial dollar debt, they have become vulnerable to large currency depreciations.

Finally, stresses in corporate balance sheets could spread to the banking system. These systemic connections are especially important in countries where banks have obtained a

large part of their funding requirements from corporate deposits, exposing them to withdrawal pressures. By adversely affecting firms' capacity to repay, weaker corporate balance sheets could also feed into the banking system through higher non-performing loan rates.

Another potential source of risk to EMEs relates to the changing investor base in their debt markets. While institutional investors and large asset management companies (AMCs) have a major role to play in deepening EM financial markets, their behavior could also become a source of problem. Hattori and Takáts (2015) discuss several channels through which the portfolio decisions of asset management companies could amplify market volatility. Return and duration mismatches in the portfolios of long-term institutional investors, such as pension funds, could be one source of risk. The search for yield and duration by these investors under conditions of very low long-term interest rates can lead to excessive risk-taking in relatively illiquid markets, causing large price fluctuations. Similarly, AMCs are guided by several investment constraints, such as relative performance targets, risk limits, and minimum credit ratings that have the potential to create procyclical market dynamics in EME bond and equity funds. There is evidence that investment flows into and out of EME funds tend to show greater clustering than flows into and out of advanced economy markets. In addition, discretionary sales by EME bond funds managers tend to amplify investor redemptions [Shek et al. (2015)].

An additional risk could come from carry trade strategies involving EME local currency debt markets. To the extent that foreign investors have not adequately hedged the foreign exchange risk of their bond investments, and have instead intended to profit from expected currency appreciation, their response to unexpected exchange rate movements could aggravate market volatility. EME residents could also make use of dollar debt issuance to undertake similar carry trade strategies. For instance, a recent BIS study of companies from 47 countries outside the U.S. found that EME non-financial companies had used U.S. dollar bond issuance to take on financial exposures that shared the attributes of dollar carry trades [Bruno and Shin (2015)]. The proceeds of such bond issuance were invested in high-yielding bank deposits as well as in shadow banking products and commercial paper.

Yet, how far these risk could actually materialize and affect financial systems depends on specific country situations. For instance, the recent adoption of regulatory and market oversight measures by a number of countries could prove helpful in containing some of those risks. These measures include,

for instance, tighter standards for firms' external funding eligibility and regulations requiring corporations to hedge foreign exchange risk. Moreover, to the extent that bond issuers in EMEs are typically large and of good credit quality, they could be more resilient to negative interest rate shocks than those with weaker balance sheets and credit ratings. Similarly, prudential measures to limit debt build-up in the household sector and building capital buffers in the banking systems could help in containing systemic risks, particularly stemming from external sources.

IMPLICATIONS FOR MONETARY POLICY

Larger and deeper capital markets help to improve the effectiveness of monetary policy, notably by strengthening the transmission of central bank's policy rate to market interest rates. Greater competition from debt markets can also lead to narrower bank intermediation spreads if banks were using their monopoly power to protect high margins, which may affect the equilibrium (or neutral) policy interest rate required to stabilize the economy [Kohlscheen and Rungcharoenkitkul (2015)].

That said, policy challenges in financially open economies can be more complex. Greater global integration of domestic debt markets means that domestic long-term interest rates move more closely with global interest rates than domestic policy rates, which can reduce the effectiveness of monetary policy. Additionally, bank credit may become more volatile as funding conditions of the banking system become more closely linked to the global capital market [Shin and Turner (2015)]. And, greater global debt market integration implies faster transmission of risk aversion shocks, sharper exchange rate movements and, consequently, larger balance sheet movements. All this leads to a stronger "risk-taking channel" of monetary policy [Borio and Zhu (2012)]. Evidence presented by Kohlscheen and Rungcharoenkitkul (2015) indicates that credit flows to EMEs are significantly affected by global risk aversion, such as the VIX index of U.S. stock market volatility and the exchange rate. And the real effects of these variables have increased because of a stronger response of investment to credit flows.

This suggests that the structure of the financial system and the regulatory regime are likely to play a major role in determining the impact of financial intermediation models on monetary policy. For instance, in Malaysia, despite higher foreign ownership of domestic debt markets, the pass-through of the policy rate has increased partly because the share of floating rate loans in total loans has risen [Bank Negara Malaysia (2015)]. In Korea,

recent changes in financial intermediation did not affect the transmission of the policy rate because regulations limiting loan-to-deposit ratio and the loan risk premium reduced the potential divergence between bank lending rates and the policy rate [Jinho (2015)].

How should monetary authorities respond to these changes in financial intermediation? One view is that in times of adverse market conditions, monetary authorities should act as the market-makers-of-last-resort, underpinning liquidity and investor confidence. This would help to reduce the probability of a sharp unwarranted increase in bond yields and tighter monetary conditions. Others have stressed that keeping one's own house in order – e.g., containing macroeconomic and external imbalances – is a necessary (if not sufficient) condition for preventing financial stress from materializing in the first place.

The recent policy focus has generally been to activate measures that help to prevent build-up of imbalances. In this context, macro-prudential policy tools (e.g., loan-to-value ratios in the property sector, limits on currency mismatches, closer monitoring of derivative positions of financial institutions, and minimum holding period for non-resident debt investment) have been seen as a critical component of policy response in many countries. Strong supervision of the banking system is generally seen as an important precondition for the success of micro- and macro-prudential tools. When the non-bank sector outside the supervisory umbrella is a source of systemic risks, the next best response would be to limit funding from the regulated entities to such sectors.

The recent debate is converging to the view that global policy coordination is essential for containing market volatility, particularly during periods of exceptionally low interest rates and large-scale intervention in the foreign exchange markets. Competitive devaluations – what is inherently a non-cooperative game – damage global growth outlook. Even where coordination of policy decisions is judged not to be feasible, there is a scope for coordinating the communication of policy actions.

REFERENCES

- Adrian, T., and H. S. Shin, 2010, "Liquidity and leverage," *Journal of Financial Intermediation* 9:3, 418-437
- Bank for International Settlements, 2015, "EME vulnerabilities take centre stage," *BIS Quarterly Review*, September
- Bank for International settlements, 2015b, "What do new forms of finance mean for EM central banks?" *BIS Papers* no. 83
- Bank for International Settlements, 2016, "Uneasy calm gives way to turbulence," *BIS Quarterly Review*, March
- Bank Negara Malaysia, 2015, "Changing patterns of financial intermediation and implications for central bank policy: the Malaysian perspective", *BIS Papers* no. 83
- Borio, C., and P. Low, 2002, "Assessing the risk of banking crisis," *BIS Quarterly Review*, December
- Borio, C., and H. Zhu, 2012, "Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism?" *Journal of Financial Stability* 8:4, 236-251
- Borio, C., R. McCauley, and P. McGuire, 2011, "Global credit and domestic credit booms," *BIS Quarterly Review*, September
- Bruno, V., and H. S. Shin, 2015, "Global dollar credit and carry trades: a firm-level analysis," *BIS Working Papers* no. 510, August
- Chui, M., E. Kuruc, and P. Turner, 2016, "A new dimension to currency mismatches in the emerging markets: non-financial companies," *BIS Working Papers* no. 550, March
- Drehmann, M., and M. Juselius, 2014, "Evaluating early warning indicators of banking crisis: satisfying policy requirements," *International Journal of Forecasting* 30:3, 759-780
- Ehlers, T., and A. Villar, 2015, "The role of banks," *BIS Papers* no. 83, November
- Hattori, M., and E. Takáts, 2015, "The role of debt securities markets," *BIS Papers* no. 83, November
- Jinho, H., 2015, "Changes in financial intermediation structure: their implications for central bank policies, Korea's experience," *BIS Papers* no. 83
- Kohlscheen, E., and P. Rungcharoenkitkul, 2015, "Changing financial intermediation: implications for monetary policy transmission," *BIS Papers* no. 83, November
- Mohanty, M. S., 2015, "What do new forms of finance mean for EM central banks: an overview," *BIS Papers* no. 83
- Mohanty, M. S., and P. Turner, 2008, "Banks and financial intermediation in emerging Asia: reforms and new risks", *BIS Working Papers* no. 313
- Shek, J., I. Shim, and H. S. Shin, 2015, "Investor redemptions and fund managers' sales of emerging market bonds," *BIS Working Papers* no. 509, August
- Shin, H. S., 2013, "Adapting macro prudential approaches to emerging and developing economies," in Canuto, O., and S. Ghosh (eds.), *Dealing with the challenges of macro financial linkages in emerging markets*, World Bank, Washington D.C.
- Shin, H. S., and P. Turner, 2015, "What does the new face of international financial intermediation mean for emerging markets?" *Bank of France, Financial Stability Review*, 25-30
- Sobrun, J. and P. Turner, 2015, "Bond markets and monetary policy dilemmas for the emerging markets," *BIS Working Papers* no. 508, August

Layout, production and coordination:

Cypres – Daniel Brandt, Kris Van de Vijver and Pieter Vereertbrugghen

© 2017 The Capital Markets Company, N.V.

De Kleetlaan 6, B-1831 Machelen

All rights reserved. All product names, company names and registered trademarks in this document remain the property of their respective owners. The views expressed in The Journal of Financial Transformation are solely those of the authors. This journal may not be duplicated in any way without the express written consent of the publisher except in the form of brief excerpts or quotations for review purposes. Making copies of this journal or any portion thereof for any purpose other than your own is a violation of copyright law.

BANGALORE
BRATISLAVA
BRUSSELS
CHICAGO
DALLAS
DÜSSELDORF
EDINBURGH
FRANKFURT
GENEVA
HONG KONG
HOUSTON
KUALA LUMPUR
LONDON
NEW YORK
ORLANDO
PARIS
SINGAPORE
TORONTO
VIENNA
WASHINGTON D.C.
ZURICH



CAPCO.COM