Monitoring Financial Stability in a Complex World

Mark D. Flood
Office of Financial Research

Allan I. Mendelowitz
Committee to Establish the National Institute of Finance

William Nichols
Office of Financial Research

Version 10
January 19, 2012

ALL COMMENTS ARE WELCOME

The authors are grateful to Dick Berner, Mark Flannery, Jonathan Sokobin, John Bottega, and Irina Leonova for numerous helpful comments. Any remaining errors pertain to the authors alone.

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Financial Research or the U.S. Treasury.

Monitoring Financial Stability in a Complex World

Abstract

We offer a *tour d'horizon* of the data management issues facing macroprudential supervisors. Traditional financial oversight has been very firm-centric, with strong respect for the boundaries of the firm. Even in this firm-oriented context, financial information has been expanding much faster than traditional technologies can track. As we broaden to a macroprudential perspective, the problem becomes both quantitatively and qualitatively different. Supervisors should prepare for new ways of thinking, and larger volumes of data.

MONITORING FINANCIAL STABILITY IN A COMPLEX WORLD

Mark Flood, Allan Mendelowitz, and Bill Nichols

1 *Introduction*

This paper outlines a network approach to monitoring threats to financial stability and some of the strategic data management challenges that will confront regulators and market participants as they implement Dodd-Frank Act (DFA).[1] Because of the need to monitor the large and growing data volumes from disparate sources across the financial system, a financial stability supervisor will require specialized techniques for risk measurement and data capture, and expansive capacity for risk analysis.

We identify three strategic forces affecting data management for financial supervisors. First, financial market data volumes are growing exponentially. One should thus expect traditional data management technologies to fail, and they have. In particular, back offices of trading firms have not kept up with their own front office processes (in terms of volume and complexity of data created), nor with evolving practices in other industries to manage growing data volumes and changes in source types and persistence mechanisms. Second, systemic monitoring requires a new focus on the relationships among firms and markets across the financial system. The most important of these are the contractual relationships created by financial transactions. To assess threats to financial stability one must quantify the bilateral and multilateral relationships – and

[1] The DFA is officially the Dodd-Frank Wall Street Reform and Consumer Protection Act; see U.S. Congress (2010). The OFR provisions of the DFA were based on an earlier bill introduced by Sen. Jack Reed; see U.S. Senate (2010). Among many other things, the DFA created the Financial Stability Oversight Council (FSOC) and Office of Financial Research (OFR) to monitor threats to financial stability in the U.S. The Federal Reserve Board established a new Office of Financial Stability Policy and Research. The Federal Deposit Insurance Corporation (FDIC) established a new Office of Complex Financial Institutions. Similar significant initiatives exist at other central banks, regulatory agencies and multilateral institutions worldwide.

the chains and networks of relationships – between/among entities (including investors and issuers). Third, the possibility for diverse contract types to create very similar economic exposures, and the large volume of data needed to monitor the entire system, require a supervisor to build cognitive capacity. All these are especially important in a macroprudential context, where the data may help inform regulatory decisions that affect the whole system. However, our goal in this paper is simply to call attention to the scope of data management issues for the macroprudential supervisor to a nonfinancial audience—issues that are too often ignored.

2 Legacy Financial Supervision

Before turning to the problems of complexity and supervision at the systemic level, we first consider the issues for data management at a microprudential scale.

2.1 Firm-level supervision and disintermediation

Traditional accounting still is the data management framework most widely used for monitoring risks in financial institutions, especially for regulatory purposes. In general, a firm's risk exposures enter through its assets and liabilities, which appear on the balance sheet, a straightforward, well defined reporting format that has been refined over centuries. There are off-balance-sheet exceptions, of course, such as exposures through unconsolidated subsidiaries or assets under management, but these only reinforce the issues we describe here. Financial standards are quite explicit and intentional in their focus on the reporting entity and the "boundaries of the firm" (see, for example, FASB, 2008), as this is the managerial locus of decision-making and control, and the scope of legal obligation. This intense distinction between intra-organization activities versus inter-organization transactions has a long history, covered most famously in Coase's (1937) essay on transaction costs, and surveyed more recently by Holmström and Roberts (1998).

Firm-level accounting measures are central to prudential supervision. Banks, broker-dealers, investment companies and other market participants are all supervised at the level of the corporate entity. Capital requirements apply at the corporate level for both individual entities as well as their parent holding companies. The observation frequency for generally accepted accounting principles (GAAP) reporting is typically quarterly or annual: the state of a typical *non*-financial firm changes only gradually over time as sales are completed and expenses incurred, so a quarterly reporting cycle is usually adequate. Indeed, for firms practicing just-in-time (JIT) inventory or manufacturing, quarterly filing schedules are trivial to meet. Most financial regulatory reporting has adopted this same frequency.[2] In contrast, large financial firms with significant trading operations are able to modify their valuation and risk profiles much more rapidly, and large banks are therefore generally subject to continuous on-site supervision.

The firm-centric conception of risk inherited from accounting also appears in many of the modeling abstractions that are commonplace in applied risk management. Value at risk (VaR), economic value of equity (EVE), risk-weighted assets (RWA), and other familiar metrics are good examples. As we argue below, there are important risk-management and data-management implications of an exclusive focus on firm-level exposures. An individual financial firm exists in a volatile marketplace with only limited visibility into the views, preferences and constraints that

[2] Workhorse regulatory data collections in the U.S. include the SEC's 10-K (annual) and 10-Q (quarterly) reports, bank Call Reports (quarterly), and Thrift Financial Reports (quarterly). While it is difficult to generalize, reporting abroad tends to be less frequent than in the U.S.

guide the behavior of its counterparties. From the perspective of a manager or regulator, it can often be a plausible and convenient simplification to regard the firm as an “island,” treating the market values of assets and liabilities myopically as the outputs of some random process. Assuming some stability in the data-generating process, one can estimate the parameters of the price distribution, calculate confidence intervals, and use them to set position limits and allocate capital. This is the essence of the VaR methodology. In part because it links firm-level managerial objectives to microprudential regulatory objectives, this procedure is well suited to programmatic risk management within the firm as well as standardized capital regulation from without.[3] Unfortunately, the firm-as-island conceptualization ignores important system-level phenomena. For example, the recent crisis demonstrates amply that systemic effects matter. Danielsson and Shin (2003) highlight the fallacy of composition inherent in a strictly microprudential supervisory regime. The whole is not the sum of the parts: firm-level risk management and supervision alone are inadequate.[4] An important practical manifestation is the so-called “volatility paradox,” (see, e.g., Brunnermeier, Eisenbach and Sannikov, 2011) whereby an episode of low volatility such as the 2003-2007 period combines low firm-level risk measures (e.g., VaR) with growing systemic risk, as aggregate imbalances are accumulated.

Accounting has other limitations as a source of risk information. Traditionally, valuations were recorded at historical cost, with the advantage – extremely useful for contract enforcement – of being unambiguous. However, historical cost is a backward-looking measure, and therefore a

[3] Blundell-Wignall and Atkinson (2010) describe many of the issues that arise in this approach. Adrian and Shin (2010) discuss ways in which myopic VaR implementations can exacerbate boom-bust leverage cycles in financial markets.

[4] The literature on systemic risk measurement is large and growing. Bisias, Flood, Lo and Valavanis (2011), IMF (2009), and ECB (2010b) provide overviews.

very poor choice for risk accounting. Relatively recent changes to "fair value" standards are more forward-looking, but ironically make GAAP financial statements more difficult to interpret.[5] There are intricate rules for distinguishing "held to maturity" versus "trading" or "available for sale" securities. The former are recorded at amortized cost, while the latter are typically recorded at fair value. This has the potential to mislead by confounding two measurement frameworks in a single balance sheet. Because up-to-date market prices are not always available, determination of fair value introduces significant discretion and ambiguity into the measured valuations. For example, Benston (2006) recounts the use (and abuse) of discretionary "Level 3" mark-to-model fair valuations by Enron, contributing to its demise.

For a static regulatory regime to provide consistent system-level supervision over time implicitly requires some stability in the institutional structure of the regulated sector. However, as documented by Kaufman and Mote (1994) or Boyd and Gertler (1994) updated by Feldman and Lueck (2007), financial activity has systematically disintermediated away from traditional banking institutions over many decades. The fact that many of the new markets are relatively lightly regulated compared to banks has been a significant enticement to disintermediate. While securities regulators have typically focused on disclosure and transparency over more the intrusive regulation faced by banks, many firms in the so-called "shadow" banking system, such as hedge funds, now operate with little scrutiny at all.[6] At the same time, the shadow banking

[5] The final rule on fair value measurement was adopted by the Financial Accounting Standards Board (FASB) and the International Accounting Standards Board (IASB) in 2011; see IASB (2011). This harmonizes the FASB and IASB approaches, and replaces earlier, very similar guidance under FASB Topic 820, formerly the Statement of Financial Accounting Standards (SFAS) 157.

[6] Poszar, et al., describe the shadow banking sector in greater detail. Lo (2011b) asserts that outsiders know almost nothing about the nature and magnitude of the risk exposures of the hedge fund industry, and are forced to resort instead to plausible "conjectures." Based on data from *Institutional Investor*, he emphasizes that the size of the now defunct LTCM is an order of magnitude smaller (in 1998 dollars) than a number of current hedge funds.

system maintains close ties and interactions with traditional intermediaries, so that their activities cannot be isolated.

2.2 Financial innovation and the complexity of data management

The trend toward disintermediation has also been facilitated by the opportunities created by financial innovation. Especially noteworthy are the enormous growth in derivatives markets since the late 1970s, the expansion of trading systems and securitization markets since the late 1980s, and advances in the modeling and management of portfolio credit risk since the late 1990s.[7] Innovating firms typically view new contract types favorably. Because they face limited competition, innovative contracts earn larger economic rents for the seller, typically in the form of higher spreads. Some securities conceal embedded short puts or other contingent losses to entice investors to overpay. For example, Ingersoll, Spiegel, Goetzmann and Welch (2007) document ways in which contingent exposures can be used to manipulate standard investment performance metrics published to investors. In banking firms, a similar problem emerges in the “opaqueness” of assets. For example, Flannery, Kwan and Nimalendran (2004, 2010) show that, prior to the crisis, there was little evidence from equity trading characteristics that investors in large banks were significantly deprived of information about the credit quality of bank assets. During the crisis, however, many of these same large institutions were cut off from funding altogether due to uncertainties about their solvency, suggesting that they held large contingent exposures that came into the money in the context of the crisis.

[7] The general literature on technological innovation is largely beyond our scope. See Antonelli (2009) for a general overview. See Tufano (2003) for an economist’s overview of the literature on financial innovation.

While innovation is widely encouraged on trading desks, the ramifications for data management within the innovating firm are typically neglected. Anecdotally, sell-side firms are reported to systematically underinvest in the back-office infrastructure needed to support their front-office innovations properly. Gottfredson and Aspinall (2005) demonstrate that this pathology is not limited to the financial sector. Incentive schemes, such as the alternative investment management industry's standard "2-and-20" rule, that reward increases in gross revenues incentivize innovations that boost measured performance. Gottfredson and Aspinall (2005) argue that it is commonplace for firms of all types to fail to account for the managerial and operational complexity implied by their product innovations, resulting in a phenomenon of excessive complexity and over-innovation. In short, the introduction of new products necessitates the costly development of specialized data-management infrastructure to track transactions. Failing that, the burden of data integrity falls upon back-office personnel, with the inevitable incidence of operational errors. Many of these costs are inframarginal, since the operational sluggishness engendered by an innovation tends to affect existing product lines as well. Most of these complexity costs also fall on the back office. Notably, Gottfredson and Aspinall (2005) propose the count of a firm's distinct SKUs (stock-keeping units) as a basic operational complexity metric. Unlike most manufacturing and retail sectors, there is as yet no comprehensive, shared SKU system – i.e., a globally standard set of instrument type identifiers – in finance.

Securitization innovations have helped to supplant traditional portfolio lending with an originate-to-distribute business model, fundamentally altering the lender's production of data and information in subtle and not-so-subtle ways. So-called "soft" information about creditworthiness, derived in part from a loan officer's subjective experience of working directly with the borrower, is discarded when the loan is sold into a securitization. Instead, all information on loans intended for securitization is reduced to a set of "hard" information defined by the inputs to an automated underwriting calculator. A series of recent papers explores how lending, including the types of credits underwritten, differs systematically between small and large banks as a result of the distillation by larger institutions of the underwriting information to a set of strictly hard criteria.[8] At the extreme, some mortgage securitization underwriters have submitted predefined pool characteristics to mortgage bundling operations which then accumulated newly originated loans to fulfill the specifications. In such cases, no information other than the predefined attributes was ever collected about the loans. In short, information loss along the supply chain is a function of provenance and lineage practices; the fact that there are no requirements to guard against information loss means that no firm willingly incurs the costs to maintain this information. Indeed, under pressure to fill the pools, tolerance for occasional missing values expanded naturally via "ignorance creep" to create a large specialized market for low-doc and no-doc loan pools.[9]

Paradoxically, the rise of the originate-to-distribute model has increased the overall volume of data in the financial system by fracturing the traditional origination process into a sequence of highly specialized transactions. Many parties along the securitization pipeline have an active interest in the performance of each loan. As financial products such as mortgages have been systematically securitized – and those securitizations structured and repackaged – loan details

---

[8] See, for example, Berger, Miller, Petersen, Rajan, and Stein (2005), Agarwal and Hauswald (2010), and Petersen and Rajan (2002). In the mortgage industry, the two most common automated underwriting systems are Freddie Mac's *Loan Prospector* and Fannie Mae's *Desktop Underwriter*.

[9] For example, Cordell, Huang and Williams (2011, p. 25, emphasis in the original), citing IOSCO (2008, p. 2), note that, "clearly data quality was a problem, fueled as it was by declining underwriting standards. One very valid point on the data is that the quality of the data being provided deteriorated significantly in the buildup to the crisis because of declining underwriting standards, by the IOSCO's reckoning, 'beginning in late 2004 and extending into early 2007.'"

that once might have been recorded only by the borrower and the originating lender (holding the whole loan in portfolio), are shared by the borrower, originating bank (for recourse commitments), loan servicer, securitization trust, securitization bondholders, and the buyers and sellers of credit protection derivatives. On the one hand, the data reconciliation process necessitated by inter-firm contracting should improve data quality by putting many sets of eyes on the same set of facts – assuming that individual participants have enough information in one place to support reconciliation at all. On the other, having so many consumers of the information multiplies the validation burden more than proportionally.[10]

2.3 Scalability of data management

The steady expansion of new market segments has moved significant portions of the financial system to the fringes of regular supervision. We lack accurate, up-to-date estimates of the total size of many over-the-counter markets. As a result, we know surprisingly little about the simple scale of certain market segments, limiting our understanding of the overall data management problem in the financial system. Thanks to technological innovation, the problem is growing in size. Individual innovations tend to be disruptive events, but they also accumulate over longer periods into a smoother high-level growth trajectory. A look at the basic orders of magnitude is helpful (see Figure 1). Similar to Moore's law for transistor densities, data volumes – for example, proxied by aggregate digital storage space or Internet throughput – have grown

---

[10] See Flood (2009) on the costs of data reconciliation across multiple schemas and systems. The discussion here implicitly assumes that the information collected at origination comprises a stable and well defined set of attributes. Because relational databases are costly to change in production systems, back-office practices typically require static data models describing instruments that traded repeatedly. Front-office dealers, on the other hand, frequently prefer customized deal structures built from one-off or semi-standardized components with idiosyncratic data representations. This can overwhelm back-office data modeling.

globally at an exponential rate.[11] Hilbert and López (2011a, 2011b), for example, estimate 1986-2007 average annual growth rates for storage capacity (23%/year). Koh and Magee (2006) estimate the long-range exponential growth rates in data storage and transport-layer bandwidth, and Nagy, Farmer, Trancik, and Gonzales (2011) reprise that study, arguing that growth has in fact been super-exponential. Financial activity is data- and information-intensive, and exemplifies this growth experience. Data validity is critical for financial activity; that is far different from much of the generic traffic on the Internet. At 25 frames per second, a downloaded video (even heavily compressed) incorporates a great deal of signal redundancy. A few corrupted bits or even the loss of an entire frame would seldom be noticed, let alone provoke misinterpretation. In part because contractual ambiguity is potentially very costly, signal redundancy is much less common in a financial context. As a result, corrupting a few digits in a transaction confirmation or payment instruction could easily be cause for significant concern. Flipping a single bit might mean the difference between paying and receiving millions of dollars: Nick Leeson's billion-dollar rogue trading loss in the Barings Bank scandal began with an innocent clerical error of this sort (see Bookstaber, 2007, 38-39).

FIGURE 1 APPROXIMATELY HERE

Figure 1 suggests the nature of the problem. Starting with double-entry bookkeeping, participants deploy a range of technologies and processes to scrub and validate their data. Traditionally, these techniques have relied heavily on human diligence and attentiveness. Even

[11] The numbers provided here are intended to be suggestive of the situation in financial markets, rather than conclusive. The growth in processing power represented by Moore's Law is particularly relevant as a benchmark for the growth in storage requirements in finance, since advances processor power help enable the development of new market segments. Valuation of structured securitizations, for example, makes frequent use of CPU-intensive Monte Carlo analyses; see, for example, Berthold, et al., (2011). Similarly, while high-frequency trading is typically latency-dependent, it nonetheless benefits from high-performance processing power; see, for example, Intel (2010).

in processes involving a large degree of automation, a “human in the loop” will typically be a binding constraint, so we use global population as a rough proxy measure of aggregate capacity for processes that depend significantly on manual input. Population has more than doubled in the last half-century, while stock market trading volume has increased almost a thousand fold.[12] In turn, the trend in trading volume is broadly consistent with Hilbert and López’s (2011b) estimates of the growth in aggregate storage volumes.[13]

These measures are also consistent with recent evidence that the trade settlement process is increasingly staggering under the activity load. Exception management accounts for a large fraction of the total cost of trade processing. For example, Bradley, et al. (2011, Figure 2) note that overall settlement fails have been generally increasing since at least 1996. The failure rate series is volatile, with occasional severe spikes. Trimbath (2008) finds that, prior to the financial crisis, settlement failures in U.S. bond markets rose over the last decade, with the trend interrupted by regulatory and market actions. In some cases, back-office behavior has been

[12] For at least two reasons, the S&P 500 trading volume depicted here represents a lower bound on the growth in data generated by the financial system. First, it does not encompass the vast increase in derivative markets that has occurred since 1980. Comprehensive data on outstanding balances (not trading volumes) for OTC derivatives are available only since 1998; see BIS (2010). These have shown a roughly order-of-magnitude increase over the past decade, with approximately $600 trillion notional outstanding in June 2010 (ca. $25 trillion in market value), dominated by interest-rate swaps. The growth in trading is also reflected in and compounded by the growing “financialization” of the economy: the share of GDP represented by the U.S. financial sector (including insurance) has tripled since World War II, and nearly doubled since 1980 (see Philippon, 2008, Figure 1, p. 36). Second, each transaction generates a number of internal and external versions of the trade information for financial reporting, regulatory compliance, risk management, etc. These ancillary data sets should all be kept consistent, but the number of reconciliations required does not typically scale linearly with the number of positions or transactions (see Flood, 2009). Note that time scales in financial markets have also been shrinking, evidenced by the growth of algorithmic trading; see Castura, et al. (2010) or Hendershott, et al. (2011). Because more must happen faster, the consequences of process failure are correspondingly larger.

[13] Extrapolating from their 23% approximate annual growth rate over the 1986-2007 period – and assuming it applies at least equally to the financial services sector – we see that data storage requirements are on the order of 10,000 times greater in 2005 compared to 1980. For comparison, they estimate annual growth rate for worldwide computing capacity at 58%/year, and telecommunications traffic at 28%/year. At the same time, advances in processing power are also creating engineering challenges as applications impose heavier demands on legacy database technologies; see, for example, Stonebraker, et al. (2007) and Pandis, et al. (2010).

chastened by losses and invigorated regulation in the wake of the 2008 crisis, rendering pre-crisis evidence obsolete or suspect. For example, Bradley, et al. (2011, Figure 3) show that settlement fails in the market for U.S. Treasuries dropped sharply after imposition by the Federal Reserve of a penalty for fails in this market in May 2009. However, the same chart indicates that the (unpenalized) fails in the mortgage-backed securities (MBS) market have continued to grow steadily over the same time period.[14] To be effective, regulation must be applied and enforced; it doesn't occur automatically.

FIGURE 2 APPROXIMATELY HERE

The practical implications of pushing too much data through a process can be painful to watch. For example, mortgage foreclosure rates have skyrocketed since the collapse of the market in 2007-08. Figure 2 shows delinquency and charge-off rates for residential mortgage loans jumping abruptly above historical precedent starting in 2007. The foreclosure rate (total foreclosed loans as a percent of all mortgage loans; not shown) tracks very closely with the delinquency rate. While the delinquency rate roughly quintupled during this episode, the charge-off rate at the peak was roughly 20 times higher than its 1990s-era average. Foreclosure of mortgage loans has historically been handled on a case-by-case basis, with much manual processing.[15] A natural consequence of an unanticipated increase in the foreclosure-processing

[14] TPMG (2011) offers main mechanisms for settlement fails: miscommunication, operational problems (e.g., the September 11, 2001 disruption), "daisy chain" fails in which failure to receive collateral on one deal leads to failure to deliver on another (this is an example of "tight coupling" as described by Perrow, 1999, and Bookstaber, 2007), and "strategic" fails in which the "short" counterparty intentionally reneges, typically because the cost of borrowing securities to fulfill his commitment approaches or exceeds the time-value opportunity cost of postponing delivery. Strategic fails are thus exacerbated by episodes of low interest rates.

[15] Note that this is the delinquency rate for mortgages overall, including both prime and subprime loans. The delinquency rate for subprime loans in isolation was much worse, peaking at over 15.5% in the final quarter of 2009. Prime mortgage borrowers are easy for mortgage servicers to handle: until the crisis, defaults and foreclosures were rare, and loans typically had very standard structures. As a result, the mortgage servicing business became

throughput was an acceleration of the legacy (largely manual) processes to accommodate the new volume. One of the practical manifestations of this has been "robo-signing" of foreclosure documents.[16] As Kaufman, et al. (2010), Holland (2011), Wallace (2011) and Hunt, et al. (2011) all make clear, this is not an isolated problem, but emblematic of a range of institutionalized and partially manual processes throughout the mortgage industry. As other parts of the securitization plumbing have increased their throughput, the narrower pipes are often overwhelmed, provoking process failures. The de facto inability to perform proper diligence at this scale results in dilemma between costly type I (excessive foreclosure) and type II (excessive forbearance) errors. In principle, the information for accurate decisions is available, but the processing power is not.

Poor incentives and externalities also plague data management. The cost of remediating backlogs and building new processes are borne directly by the firms involved, while many of the risks involved are by nature systemic, and therefore not internalized. This creates a natural role for a supervisory intervention. For example, in 2005 the largest New York dealers in the market for credit default swaps (CDSs) were admonished by regulators for their enormous paperwork backlog, and agreed to clean it up (see Senior Managements, 2005). As with settlement fails in the Treasuries market, mentioned above, this was less a question of inadequate technology, and

---

concentrated in a handful of specialized banks that invested in the relevant information technology infrastructure. In contrast, subprime mortgages employed a variety of innovative terms ostensibly intended to constrain the monthly mortgage payment to a level expected to be sustainable for the borrower. In addition to a more complex servicing process, subprime loans exhibit very different default rates. In hindsight, it is apparent that both underwriting standards and credit pricing were too lax for an enormous number of subprime mortgages, especially those originated after 2005. Dungey (2007a) provides a good introduction to the mechanics of the mortgage servicing process. Dungey (2007b) is a similar overview of the foreclosure process on the eve of the crisis.

[16] Robo-signing is the practice of attaching signatures to affidavits and other foreclosure documents so quickly that it is inconceivable that a reasonable review occurred. This is a data-validation issue on two levels: First, the signature is an attestation, based on (supposedly) diligent human review, of the validity of the information in the affidavit. Second, because it seems in many cases that the task was delegated to unauthorized and unqualified shills as an expedient, the signatures themselves become data requiring subsequent validation.

more a question industry leadership and new regulatory incentives.[17] In a more famous episode, the Paperwork Crisis of 1968-1970, increases in stock market trading volume overwhelmed the industry's back-office capacity. Settlement fails and "DK" (don't know) transactions proliferated. The New York Stock Exchange (NYSE) began closing early to help clear the backlog, and the markets closed entirely on Wednesdays for a period of several weeks in July-August 1968. The episode culminated in the Securities Investor Protection Act of 1970, which created the Securities Investor Protection Corporation (SIPC).[18]

Because manual processes are the least efficient, they are the most vulnerable to input surges, and straight-through-processing (STP) architectures are gradually replacing them.[19] Unfortunately, Figure 1 depicts steady exponential growth in data throughput volumes. Few processes – even automated ones – scale well when they are pushed orders of magnitude beyond their designed capacity. In this context, the transition to STP simply moves the automation boundary. That is, after the shift to automated application of data-validation rules (i.e., the shift to STP) has extracted its efficiency gains, additional efficiency will again be demanded. Perhaps

---

[17] Fleming and Garbade (2005) provide a contemporary analysis of settlement fails in the Treasuries market. The Counterparty Risk Management Policy Group Report (CRMPG, 2005), a statement by participants of industry best practices, was a catalyst for change at the time. When the operational costs are small and/or not internalized, unilateral remediation is difficult to justify.

[18] See Markham (2002, 362-367) and SEC (1972, 3-6).

[19] On the need for straight-through processing, see CPSS-IOSCO (2001) and CPSS (2008). For an example of an implementation perspective, see Ciulla, Bloom and Justin (2010). CPSS (2011) identifies five main categories of financial market infrastructure, each of which encompasses a multitude of processes, and each of which might benefit from STP:

- payments systems
- securities and other settlement systems (SSSs)
- central securities depositories (CSDs)
- central counterparties (CCPs)
- trade data repositories (TRs)

Ironically, technological advances may also encourage novel practices – such as transacting via text messages from wireless devices – that place further demands on data management and validation.

this requirement will be satisfied by techniques for the automated generation of the data-validation rules themselves, or thereafter by the automated generation of domain-specific languages for specifying data-validation rules.[20] However, because risk is a central concern of the supervisory process, seemingly straightforward outlier-detection rules that are useful for low-intelligence bulk validation in other domains are likely inappropriate in this context; see, for example, Ghoting, et al. (2008) or Han, et al. (2011, ch. 12). For risk applications, the most interesting facts are the very often ones that appear as "outliers" in the data.

3 Systemic Supervision and the Network of Counterparty Claims

Information management is more challenging still at a systemic level. In addition to a proliferation of institutions and institution types, there is a separate set of inter-firm considerations that only apply to macroprudential supervision.

3.1 Networks and information

While the implementation of the DFA is fundamentally redefining the supervisory process, some implications for systemic risk monitoring are already becoming clear. First, because it is systemic, data-validation challenges are likely to be severe. The broad scope – all financial sectors and institutions – implies very large data volumes. Systemic supervision also implies more kinds of data (e.g., accounting, macroeconomic, contractual terms and conditions, etc.)

---

[20] Madnick and Zhu (2006) offer some concrete examples of the role of semantic context in defining the quality of a data set, as well as suggestions for effective management of that semantic context to improve data quality. Fueber, Hepp and Wischnewski (2011) indicate a similar path forward, defining a data quality constraints language targeted at the Semantic Web.

from multiple markets sectors (e.g., equities, over-the-counter (OTC) derivatives, commercial loans, etc.). Tools and techniques for reducing the data-reporting burden and streamlining the data-validation process will be especially welcome.

Moreover, the notion of risk changes as the context broadens to the systemic level. As described above, it is commonplace for both firm-level risk managers and microprudential supervisors to regard the firm (expediently) as an island buffeted by unpredictable random shocks. Individual firms typically try to evaluate their immediate counterparties, but they cannot peer more deeply into the network than that. Portfolio positions and risk exposures are closely held business secrets. For example, Andrew Lo highlights this problem of myopia in his own attempts to understand the behavior of hedge funds:[21]

> "… you know for a fact that there are people out there that know what actually happened, but they're not talking. So, in fact, this entire paper could be science fiction or it could be dead on, we have no idea. To this day, we don't know, because nobody's talking. They're not allowed to talk, because that would disadvantage their shareholders."

In contrast, a view of the full financial network provides additional conditioning information relative to what is available to the individual firm. Price events that appear to the myopic participant to be deep in the tail of the unconditional distribution – the so-called "black swans" – might be much less surprising with knowledge of the connections and imbalances observable in the web of counterparty claims. Macroprudential supervision could well focus on the network of contractual relationships. This is conditioning information with high marginal value.

As noted above, disintermediation has been one important influence in the growing significance overall of macroprudential factors – especially the network of claims – with important

[21] See Lo (2011a, at 13:18). The study he refers to is Khandani and Lo (2011).

implications for information and data management. Securitization in particular moves a lending relationship away from the originating lender, which traditionally maintained extensive hard and soft information about the borrower, and distributes responsibility for it across a variety of specialized agents, including the loan servicers, bond investors, CDS protection sellers, etc. To support this web of interests and relationships, data about the loans is compartmentalized and replicated across a range of participants. The issues are particularly acute for tranched or structured products, such as collateralized debt obligations (CDOs). Judge (2011) refers to this process as fragmentation, and coins the term "fragmentation node" to describe a counterparty where cash flows and associated data are parceled into subsets and repackaged for sharing with others in the network. As discussed above, such data fragmentation is a "lossy" conversion, in the sense that most of the soft information from the origination process is lost as a loan enters a securitization. In other words, the pre-fragmentation information set is typically greater than the sum of the post-fragmentation parts. Securitization distills it all down to a narrow subset of hard information, with the responsibility for collecting and maintaining the information distributed across a range of participants.

Moreover, fragmentation per se is an obstacle to the comparison and aggregation of information. Cordell, Huang and Williams (2011), for example, do the difficult work of comparing subprime MBSs to "structured finance asset-backed securities collateralized debt obligations (SF ABS CDOs)" based on those same subprime MBSs. Ordinary MBSs have a relatively simple senior/subordinated structure, while CDOs, because they typically combine multiple tranches from each of many MBSs, have a much more intricate subordination scheme mapping the individual loans through the MBSs to the particular tranches of the higher-level CDO structure. After examining write-downs on the universe of publicly traded ABS/MBS securities and SF

ABS CDOs issued between 1999 and 2007, Cordell, Huang and Williams (2011, p. 24) highlight an extraordinary difference between subprime MBS and the more structured ABS CDOs: "only 4% of the AAA-rated subprime MBS securities issued from 2005 to 2007 were impaired or are expected to become impaired. By our calculations, 98% of the AAA-rated SF ABS CDOs issued between 2005 and 2007 will suffer write-downs." For investors and rating agencies to take seriously the AAA rating on these SF ABS CDOs required either highly implausible assumptions for loss experience, or – more likely – failure to perform the analysis at all.

In some cases, contractual complexity can render diligence impossible. It is easier to create certain pricing problems – for example, constructing an intricately structured derivative security – than to solve those problems. A recent paper by Arora, Barak, Brunnermeier, and Ge (2011) illustrates the difficulties.[22] The standard argument, presented by DeMarzo (2005), is that issuers can reliably signal the quality of newly issued security by taking a first-loss position (junior tranche). In contrast, Arora, Barak, Brunnermeier, and Ge (2011) show how a CDO issuer can "boobytrap" a subset of its CDOs by strategically hiding the worst-performing assets in them. This creates a natural information asymmetry, in which the creator of the contract inherently knows more about its value than prospective buyers. In extreme cases, it is literally impossible, due to computational bounds, for the seller to prove that the offering price is reasonable, and likewise impossible for the buyer to verify the seller's claims. Because information asymmetries in financial markets are typically profit opportunities, complex securities tend to arise endogenously; they are not accidents of nature. While the boobytrap example demonstrates the impossibility of full diligence in certain cases, it also suggests that

[22] Flood, Kyle and Raschid (2010) also discuss some of the implications of financial complexity for information management.

issuers strategically deceive investors, begging the question of investor naïveté: in a repeated game, why are deceptive issuers not ultimately driven from the market? Even if the impossibility of diligence defeats the usefulness of signaling via a first-loss position, reputation should discourage manipulative behavior (see, for example, Hartman-Glaser, 2011). However, the dynamics of selection do not require intentional deception in order for the market to prefer complex securities: any product for which diligence and reasoning are imperfect, and for which errors in analysis tend to favor the issuer – the so-called "winner's curse" (see Thaler, 1988) – will have an "evolutionary advantage."

At the network level, the web of claims helps to obfuscate, because important system-level patterns are not visible to individual, myopic participants. Indeed, this is an important justification for government supervision of the system. Moreover, shocks can propagate in surprising ways. For example, Bookstaber (2007) offers the example of the LTCM failure in 1998, in which Russia's sovereign bond default ricocheted through the network of claims to hit the market for Danish mortgage bonds. The latter had no immediate connection to Russian debt, but simply happened to be among the more liquid assets in large portfolios that were otherwise exposed to Russia. Although this connection is surprising – certainly it was for LTCM – in principle, such indirect linkages may be foreseeable, since portfolio holdings are a matter of fact, while the behavior of portfolio managers in a panic will likely be more tightly constrained and predictable than otherwise.

3.2 An example – rehypothecation of repo collateral

We offer the example of rehypothecation of repo collateral to illustrate the importance for monitoring threats to financial stability of investor myopia amid the network of contractual relationships. Rapid deleveraging in the repo markets was an important crisis propagation channel in the wake of the Lehman Brothers failure in the fall of 2008. As discussed below, feedback and contagion among leveraged institutions can produce non-linear responses to exogenous shocks at the system level.

A "repo" is a sale of securities (i.e., collateral) combined with a simultaneous commitment to repurchase them at a later date, usually in the near term.[23] A relatively simple example is a hedge fund that wants the risk and return profile of a particular security (e.g., corporate bonds) for its portfolio, but wants to boost returns by leveraging its capital. In this example, the hedge fund buys the bonds on the open market and immediately sells them into a repo transaction with its prime broker.[24] The hedge fund gets the desired bonds for its portfolio, but is effectively using borrowed money to pay for them. Of course, the hedge fund does not receive the full value of the bonds in the front leg of the repo; a haircut is assessed to protect the prime broker against fluctuations in the value of the collateral. The net effect is one of leveraging, as the hedge fund can use the cash proceeds from the repo sale to purchase additional bonds. It is common for the prime broker in a repo transaction to take absolute title to the collateral. This facilitates the sale

[23] Taub (2008), IMF (2001), and Copeland, et al. (2010) describe the mechanics of the repo markets in greater detail. The repo markets are very large, and there are naturally numerous variations.

[24] A prime broker is a specialized firm that provides a range of related services to hedge funds and other investment managers. Typical services include custody, securities settlement, tax accounting, and account-level reporting. Lehman Brothers acted as prime broker for a number of large hedge funds at the time of its demise. In the example here, the hedge fund is the "collateral pledger," and the prime broker is the "collateral pledgee."

of collateral by the prime broker in the event the collateral pledger fails to repurchase it as promised at the maturity of the repo.

However, depending on the jurisdiction and the details of the prime brokerage agreement, the collateral pledgee will have a "right to use" the collateral.[25] Among other things, a prime broker with a right to use may rehypothecate (re-lend) the pledger's collateral to third parties for other purposes. For example, another hedge fund might pay to borrow the collateral to use in a short sale transaction. Gorton and Metrick (2009, p. 8) note that collateral is a scarce resource in securitization markets, so that there are strong incentives to leverage it through rehypothecation. Deryugina (2009, p. 257) observes that both the pledger and pledgee can benefit from the additional revenues generated by this reuse.

FIGURE 3 APPROXIMATELY HERE

These relationships are depicted in Figure 3, which shows both a simple repo transaction on the left, and a repo involving rehypothecated collateral on the right. Note that rehypothecation has the effect of connecting two subgraphs, significantly complicating the topology in the counterparty network graph.[26] We emphasize that the rehypothecation occurs invisibly to the original pledger of collateral ("Hedge Fund #1" in the figure); although pledgers are aware that rehypothecation goes on, they do not in general observe when specifically their own collateral is rehypothecated or to whom. This lack of transparency about the network of relationships played an important role in the recent crisis. Deryugina (2009, pp. 274-75) notes that, when Lehman

[25] Deryugina (2009) describes the structure of rehypothecation transactions and related legal considerations in detail. She emphasizes the importance of the relatively lenient U.K. rules on rehypothecation in attracting prime brokerage business to London.

[26] Pozsar and Singh (2011) further explore the complexities introduced by rehypothecation of collateral.

Brothers International Europe (LBIE) failed in London in September 2008, it had rehypothecated or commingled over $20 billion worth of client collateral, much of which LBIE could not identify immediately. Most of those pledgers became general creditors in the subsequent bankruptcy; Deryugina (2009, pp. 274-75, note 111) quotes from the court's response to pledgers' petition for information about the whereabouts of their collateral:

> "[I]t would be necessary to investigate particular records held by LBIE and to obtain data and records from relevant third party custodians, depositaries and other parties. ... [T]he difficulties that this process faces, not least the refusal of a number of custodians and others to comply with demands for information and that, in the meantime, the administrators are only able to call upon limited LBIE resources."

The flip side of the financial boom sustained by increasing leverage of collateral is the self-reinforcing deleveraging cycle that ensues when the value of the collateral is called into question. In such a cycle, redemption of collateral at fire-sale prices depresses the value of remaining collateral, forcing additional margin calls and subsequent redemptions. Gorton and Metrick (2009) and Singh and Aitken (2010) describe this process in detail in the context of the Lehman failure. If it were simply a question of investor disclosure, an obvious fix would be to impose tighter restrictions on pledgees' ability to reuse collateral without explicit permission of the pledgers.

However, this would not remove the incentives for pledgees to reuse scarce collateral. Their gains from leveraging collateral are internalized, but the risks of a contagious deleveraging are externalized, suggesting a possible role for prudential supervision. Because of the intrinsic myopia of individual participants, supervisory transparency into the full network of relationships is especially valuable. Kashyap, Berner, and Goodhart (2010) survey the economic literature on fire-sale contagion during the crisis, and argue that the fire-sale problem fits naturally into the broader framework of macroprudential policy.[27] They draw a straightforward but powerful conclusion from a sketch of a simple three-sector (households, financial institutions and a central bank) model of the economy. Just as an airplane pilot has three sets of control surfaces to manage roll, pitch and yaw, a regulator charged with managing defaults, credit crunches and fire-sale contagion in financial markets requires three policy tools to do the job effectively. Capital requirements and liquidity requirements are two such instruments (supplemented with backstop capital and liquidity facilities during the emergency phase of the crisis). Evidence is strong that fire-sale contagion is a third significant threat, and minimum collateral margin (or "haircut") requirements are a plausible tool to address it. From a data management perspective, tools such as regulatory haircut requirements demand that policy makers be able to observe and measure emerging patterns amid the contractual network. From an accounting perspective, this will mean tracking financial relationships as objects each with its own explicit identity in the system, rather than simply as attributes that describe the legal entities. In other words, a graph consists of both nodes and edges, and both are important.[28]

3.3 Implications for supervisory implementation

The foregoing paints a daunting picture of the data requirements facing macroprudential supervisors. Summarizing, there are (at least) three major technical challenges. First, there is the exponential growth in data volumes. Second, there is the need to monitor financial relationships, especially contractual relationships and ownership hierarchies. Collecting

---

[27] Shleifer and Vishny (2010) survey the issues surrounding fire sales and contagion.

[28] A "graph" is an abstract mathematical formalism of a set of elements, called "nodes" (or vertices, or points), and a corresponding set of "edges" (or lines) that connect the nodes. Graph theory has developed a large body of proved propositions describing the nature of graphs. See, for example, Diestel (2006) for further details.

contractual terms and conditions is a prerequisite to forward-looking cash-flow and risk analysis; terms and conditions are not systematically collected by supervisors today. Contracts are also a key ingredient for mapping the network of contractual relationships for systemic modeling. Measuring the edges – i.e., financial contracts – in the counterparty network graph will require the capture of much more detail about those contracts than is the case under traditional firm-centric accounting systems. Supervisors need to know who is connected to whom. As a first step, this requires a reliable system of legal entity identifiers (LEIs) to unambiguously identify the parties to any contract. Third, there is the issue of complexity, which can occur both at the level of the individual contract as well as in the network of contractual relationships. We propose that collecting – *intelligently* – contract-level terms and conditions can balance these challenges.

Financial contracts have several characteristics that make them desirable digital records of the financial system. First, by definition, contracts connect the individual entities in the system, creating something beyond a simple aggregation of its constituent parts. In particular, the potential for feedback effects and spillovers explain the inadequacy of strictly microprudential (i.e., firm-centric) supervision.[29] Second, there are strong incentives to make the contracts valid, complete and unambiguous statements of the promises and commitments being made. Parties to the contract benefit directly from this transparency, while basic legal principles like the parole evidence rule and contractual "integration" clauses encourage clarity to be built into the contract from the start, since it cannot be added after the fact.[30] This helps in defining foundational truths

[29] The literature on network models of systemic risk is large and growing. For recent overviews, see Haldane (2009), ECB (2010a), or Moussa (2011).

[30] See, for example, Gooch and Klein (1997), especially pp. 63-64.

to support datum-level validation, as well as the internal consistency needed for contract-wide data integrity rules. Third, many (but not all) financial contracts already exist in well understood digital representations; in these cases the data representation problem is largely solved. To facilitate large-scale processes for trade confirmation, settlement, corporate actions, etc., firms encode most contracts in highly structured and well documented public messaging schemas, such as ISO20022 (2011), FIX (2011) or FpML (2011). Lastly, and most importantly, contracts define the contingent cash flows that constitute the financial essence of the relationship. The details of who pays whom how much when and under what circumstances are the key to calculating valuations and understanding risk exposures. A fundamental capability is to capture and understand each contract's cash flow commitments – often contingent on other factors – between the counterparties. Understanding the cash flows is crucial, because it is possible for two contracts or portfolios to generate substantially identical cash flow patterns, even when their legal or accounting representations differ widely. Much of financial engineering is devoted to repackaging a fixed set of cash flow commitments into a different contractual configuration, perhaps to manage or lay off risk, avoid taxable events, reduce the market impact of a trade, or simply to obfuscate the activity.

Monitoring risks from across the financial spectrum implies comparing and aggregating seemingly disparate exposures, such as a structured mortgage-backed security and a subordinated corporate debenture. Doing it in individual portfolios is one thing. However, to do it at the scale and scope of the full financial system will require additional automation and analytics, even if the monitoring frequency is not continuous. The upshot is a need for robust instrument type identification, including standardized, structured, machine-readable representations of financial contracts, and data integration technologies that build on top of them. Those technologies should include the capability to project any contract into the financial space of state-contingent cash flows, abstracting from other details that do not affect the contractual cash flows. Brammertz, et al. (2009) suggest a solution along these lines that collapses the seemingly disparate universe of financial contracts into a manageable number of cash flow patterns. Hence, two contracts with the same state-contingent cash flows appear as identical contracts for the purposes of this approach, irrespective of whether they are called loans, bonds, or derivatives, etc. A limited number of cash flow patterns can be used as building blocks to assemble more complicated patterns, so that the state-contingent cash flow obligations from the vast majority of financial contracts can be handled in a standardized and manageable way. Projections of this sort would create a set of equivalence classes that implicitly define instrument types based on financial considerations (i.e., cash-flows) rather than legal, accounting or regulatory distinctions.

While collecting contract-level details for the full system is a powerful supervisory approach, it is a major challenge that will take a long-term sustained effort to execute. It will also take careful design and structuring to avoid overwhelming the macroprudential supervisor with data storage, security, and validation burdens. Other industries have been innovative in this area where finance has not: for example, retail merchandising has deployed "eventually correct" architectures with distributed processing.[31] Techniques for resolution reduction are another obvious response, which should also support systemic risk monitoring in the nearer term. While resolution reduction originated in the visualization community as a set of techniques to compress images while still retaining important patterns and features, it has broader applicability to other domains where data compression is useful. For example, in defining the "optimal granularity" of

[31] See Gilbert and Lynch (2002) on eventually consistent architectures and the so-called "CAP theorem. See Srivastava (2006) on other recent advances in data archtiectures.

supervisory reporting for counterparty credit risk on OTC derivatives, Mutnikas and Zerbs (2011) propose that supervisors collecting contingent exposures only from the 50 largest firms, for five to ten future value dates, and under chosen set (ca. 200) contingent scenarios. Moreover, this reporting would collect aggregated gross and net bilateral exposures. Duffie (2011) suggests a similar subset-and-aggregate approach to resolution reduction. Unfortunately, surveillance requirements depend intensely on the state of the world. During a crisis, or in the aftermath of a firm's failure, the supervisor's need for information will be much more extensive and urgent than in the ordinary course of events. For example, state-contingent data collection is a central motivation for the "living-will" requirements of the DFA. FDIC (2011) describes the role of its new Office of Complex Financial Institutions (OCFI) thus:

> "A critical component of successfully addressing a distressed SIFI [systemically important financial institution] is having sufficient information and clear strategic options at the time of failure to enable decision makers to reasonably foresee the outcomes of alternative scenarios. One of the FDIC's biggest challenges during the fall of 2008 was not having the information necessary to make informed decisions. Robust pre-planning – which entails understanding how and where these enterprises operate, as well as the structure of their business lines, counterparties, business risks, their role in the financial system, and their place in financial intermediation – is essential in giving regulators viable resolution options other than a bailout in the midst of a crisis. OCFI's monitoring activity of these systemic enterprises will be the principal mechanism for validating the entities' resolution plans and informing the FDIC on the development of Title II resolution plans."

"Robust pre-planning" should include the technical ability to ingest fully granular terms and conditions on financial contracts held by the relevant firms. The capacity for state-contingent resolution enhancement should be available for the supervision of the counterparty network as well, with the important extension that the network graph also has a role in the early warning toolkit. IMF (2009), for example, highlights an empirical model of the financial network with some ability to foreshadow systemic events. Alternatively, supervisors might simulate shock on the network to learn how different topologies propagate disruptions.

Even a very low-resolution instance of the network graph could prove to be a powerful supervisory tool. Consider a graph that identifies all of the contracts in the system (or some subsystem of interest), but with only a very minimal set of attributes for each contractual edge in the network – for example, counterparties, notional amount, and some instrument type classification. Such a "thin graph" would reveal the contractual network topology, exposing accumulating imbalances and suggesting crisis propagation channels. By presenting limited information at the contract level, it would avoid the issues of aggregation (loss of information, programming effort/bugs, reconciliation, etc.) while nonetheless limiting the burdens of data validation, security, and confidentiality. At the same time, the thin graph would provide the basic scaffolding to support resolution enhancement in a crisis, by attaching a fuller set of terms and conditions as attributes of the edges in the network. As noted above, a basic requirement for building such a graph is consistent and reliable counterparty identification.[32] Large complex financial institutions may comprise hundreds or thousands of distinct legal entities. Because of this, building a network graph to monitor threats to financial stability will require data on such corporate ownership families. While not the primary focus of such an effort, an additional benefit of systematic issuance of counterparty identifiers is that it should yield significant operational cost savings for financial firms by materially reducing the number of failed trades caused by the inconsistent designation of counterparties. Finally, the thin graph would provide a

[32] The DFA, at §154(b)(2)(A)(i), also requires the OFR to build a "financial company reference data base." This will not be trivial because many individual obligors exist in parent-subsidiary hierarchies with de facto cross-guaranties. In some cases, these are de jure cross-guaranties: the DFA (at §616) reiterates and extends the "source of strength" doctrine that requires bank and thrift holding companies provide financial support to their subsidiary depository institutions.

baseline scoping of the size and coordinate dimensions of the financial system: how many firms and instruments exist, and of what types. Such a perspective is crucial for prioritizing the various options for research and supervision. To avoid looking “only under the streetlights” requires new sources of insight and light for the broader market.

Addressing these challenges will depend on the overall *cognitive capacity* of the organization, which includes:

- situational awareness of the financial system,
- decision support for policymakers, and
- crisis response capability.

In addition, there must be a research function to augment and refine each of the foregoing, and publication channels to inform external stakeholders.[33]

A core task for situational awareness is data collection and ingestion. Data collections will typically revolve around regularly repeated inflows of structured, machine-readable numeric series, such as market prices or transaction reports.[34] Data ingestion is an important step in the process, since this is where a number of important structuring activities occur, including data validation, basic classification, application of retention and filtering rules, cataloging, initial

[33] Situational awareness is a concept that originated in a military context to describe the outcome of a tactical process of perception, comprehension and projection onto a near-term decision space; see, for example, Leedom (2001). The issues of organizational capacity for systemic surveillance are better developed and understood in certain other disciplines. See, for example, Wagner, Moore and Aryel (2006).

[34] There are important exceptions, of course. Unstructured data, for example, articles from newspapers and the trade press or interviews with regulators or industry participants, will be an important source of information. The information on settlement fails – which by definition do not result in contracts – presented by Bradley, et al. (2011) might provide the basis for a systemic key risk indicator. Bisais, et al. (2011) identify a class of early warning models that are based solely on macroeconomic aggregates.

versioning, and establishment of provenance.[35] The resulting metadata will provide the core informational support for subsequent "functional accessibility" to the data – the ability to navigate, query, link, and define the data. For example, the machine representations of the contracts might be mapped to a semantic context (e.g., a semantics repository), to provide additional interpretive specificity; in this case, both the contract schemas and associated semantics should be explicitly versioned over time. [36] Metadata also matters especially for data dissemination: financial exchanges, regulators, and other participants share a wide range of information – including both raw data inputs and calculated outputs – with each other and with third parties. Standardization of term definitions, classification schemes, etc., and methods to evolve them across the regulatory and industry communities will be critical; absent this, the ability to aggregate information sensibly will not occur. Because of the large volumes of data involved, it will likely not be possible to achieve perfection in data validation at a fully granular level. Resource constraints will imply a trade-off between quantity and accuracy.[37] This trade-off should be managed to avoid mistakes and to prioritize access to the most important data. For example, incoming data might be staged in its raw, pre-ingested state until demanded by some

---

[35] Provenance is a technical term for the metadata to support repeatable collection or derivation of the data. In many cases where issues regarding chain of custody or data lineage apply, establishing accurate data provenance can be crucial. Data source tagging – i.e., citation of the source – is a basic technique. There are standard markup languages, such as the Data Documentation Initiative (see DDI, 2009) for capturing provenance metadata in a structured format.

[36] Similarly, efforts to build a "semantic repository" for finance – a comprehensive set of standard, structured, and interrelated definitions to augment the data model and help specify the attributes of contractual relationships; for example, see Bennett (2011) or Madnick and Zhu (2005) – are extremely useful, but not sufficient. A semantics repository is also only one input into the process of understanding, and not a full solution or a methodology. Other important techniques include object definition, unique entity symbology, information standardization and business process flow; these are beyond the scope of the present paper.

[37] For example, Vogels (2009), in a discussion of the "eventual consistency" model of distributed and replicated data, cites Brewer's (2000) "CAP (consistency, availability, partition-tolerance)" proposition that, "of three properties of shared-data systems – data consistency, system availability, and tolerance to network partition – only two can be achieved at any given time." A formal proof is given by Gilbert and Lynch (2002).

downstream process, effectively creating a just-in-time inventory system. The prioritization analysis might itself be assisted by techniques for automated discovery, inference and pattern recognition. Based on accumulated experience, perhaps supported by machine learning, newly arriving data might contain easily detected features or anomalies of special interest.

Because statistical analysis and data visualization are powerful tools for data aggregation, pattern extraction, and dimensionality reduction, both should play an important role in decision support in this data-rich environment. Decision support is one of the most important applications for the assembled information resources. Given the vast amounts of data involved and the complexity of relationships, there must be techniques for systematizing, streamlining and rationalizing the raw data into presentations tailored to the needs of policymakers and other stakeholders. Regarding statistical analysis, Bisias, et al. (2011) survey a diverse range of economic models of threats to financial stability, which they classify into five broad categories based on the modeling techniques employed and financial phenomena considered: macroeconomic measures, illiquidity measures, probability distribution measures, network analysis measures, and contingent-claims/default measures. In addition, they organize the paper around an alternative breakdown into broad categories based on data inputs and outputs, and analytical methods applied: macroeconomic measures, granular foundations and financial networks, forward-looking risk assessment, stress tests, cross-sectional measures, and liquidity/insolvency and crisis behavior. Finally, they identify the particular data inputs required by the individual models examined (see Bisias, et al, 2011, Table 1). Beyond traditional econometrics, well designed dashboard graphics and animations can condense important information for rapid assimilation for decision support. Data exploration is another area where visualization tools can make a major contribution. Certain basic rules for data classification, analysis and triage can be automated, but many others will require a human in the loop. For example, analysis of anomalous market activity is an example of something that it may be difficult to train a machine to do well. Graphics are a useful technique for aggregating data for broader publication, as important decisions are taken not only by regulators and policymakers, but also by investors and other market participants.

Finally, rapid response is a required capacity for what are perhaps the most interesting facts of all, namely news of major unanticipated events. The costs of poor decisions and deferred decisions can be large, and the benefits of good decision support correspondingly large. By their nature, the data delivery mechanism in such cases is unpredictable: news of a large price jump could arrive through a regular data-ingestion process; alternatively, news or direct experience of a terrorist attack might arrive outside of normal channels. The ability to react appropriately will depend on having the right skills, computational capacity, and functional accessibility to information in place when the news arrives. For example, the CFTC's ability to provide timely and effective support for the preliminary report on the "flash crash" (see CFTC-SEC, 2010a) was significantly enhanced by the data-ingestion infrastructure that was already in place when the event occurred. Rapid response capability implies a need for a very broad range of specialized expertise, some of which might be outsourced through networks of on-call analysts and researchers outside the agency. Like a triage in an emergency room, the first task will be to assess the nature of the event, so it can be handed off to the proper expert or team for classification (diagnosis) and finally response (treatment). An example of a possible response is a "flash report," defined as a decision memo offering preliminary findings and policy options within 24 hours of a significant market event. In a rapid-response context, even short-horizon early warning indicators from a risk dashboard can serve a useful function by escalating situational awareness and alerting the on-call network.

4 Summary

The preceding sections highlight important forces that shape the landscape for monitoring threats to financial stability. First, data volumes are growing at an exponential rate far exceeding the growth rate in human population. While this is a general phenomenon, it also appears to apply with even greater force to financial data flows. Traditional data management processes are unsustainable in this environment. Second, monitoring the financial system will require much greater attention to the edges in the network – financial contracts – than is available with traditional accounting or supervisory techniques. Individual participants in the system will always have limited visibility beyond their own immediate relationships. This creates a natural role for a macroprudential supervisor to monitor the evolution of the counterparty network as a whole. Third, the complexity of the problem domain, combined with the volume of data involved and the pace of decisions and activity will create a very challenging information environment for a financial stability monitor. Significant attention and resources should be devoted to building cognitive capacity in the organization.

5 References

Adrian, Tobias and Hyun Song Shin, 2010, "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), July 2010, 418-437.

Agarwal, Sumit, and Robert Hauswald, 2010, "Distance and Private Information in Lending," *Review of Financial Studies*, 23(7), 2757-2788.

Andolfatto, David, Aleksander Berentsen, and Christopher Waller, 2011, "Undue Diligence," working paper, Federal Reserve Bank of St. Louis, June. http://www.iiw.uni-bonn.de/konstanz/2011/Paper_Andolfatto.pdf

Archard, Noel, 2011, "Statement Of Noel Archard, CFA, Blackrock, Inc., Managing Director, Head U.S. IShares Product before the United States Senate Committee On Banking, Housing And Urban Affairs, Subcommittee On Securities, Insurance And Investment, October 19, 2011." http://banking.senate.gov/public/index.cfm?FuseAction=Hearings.Hearing&Hearing_id=ad4fdfb9-d589-4ac9-8829-0edf1ad8dc8d

Allen, Franklin, and Ana Babus, "Networks in Finance," ch. 21 in: P. Kleindorfer and J. Wind, eds., *Network-based Strategies and Competencies*, University of Pennsylvania, Wharton School Publishing.

Antonelli, Cristiano, 2009, "The economics of innovation: from the classical legacies to the economics of complexity," *Economics of Innovation and New Technology*, 18(7), 611-646. http://www.tandfonline.com/doi/abs/10.1080/10438590802564543

Arora, Sanjeev, Boaz Barak, Markus Brunnermeier, and Rong Ge, 2011, "Computational Complexity and Information Asymmetry in Financial Products," *Communications of the ACM*, 54(5), May, 101-107.

Bank for International Settlements (BIS), 2010, "Positions in global over-the-counter (OTC) derivatives markets at end-June 2010," technical report, November. http://www.bis.org/publ/otc_hy1011.htm

Bennett, M., 2010. "Enterprise Data Management Council Semantics Repository," Internet resource. http://www.hypercube.co.uk/edmcouncil/.

Benston, George J., 2006, "Fair-value accounting: A cautionary tale from Enron," *Journal of Accounting and Public Policy*, 25(4), July-August, 465-484. http://www.sciencedirect.com/science/article/pii/S0278425406000469

Berger, Allen N., Nathan H. Miller, Mitchell A. Petersen, Raghuram G. Rajan, Jeremy C. Stein, 2005, "Does function follow organizational form? Evidence from the lending practices of large and small banks," *Journal of Financial Economics*, 76, 237-269. http://112.78.41.54/ejurnal/JFE%202005%2076%202/JFE%2005%2076%202-2.pdf

Berthold, Jost, Andrzej Filinski, Fritz Henglein, Ken Friis Larsen, Mogens Steffensen, and Brian Vinter, 2011, "Functional High Performance Financial IT: The Hiperfit Research Center in Copenhagen - Project Paper," in: R. Peña-Marí and R. Page, eds., *Trends in Functional Programming (TFP'11) - 12th International Symposium, Madrid, Spain, May 16-18, 2011, Revised Selected Papers*, Springer. http://hiperfit.dk/pdf/TFP2011HIPERFIT.pdf

Bisias, Dimitrios, Mark Flood, Andrew Lo, and Stavros Valavanis, 2011, "A Survey of Systemic Risk Analytics," working paper, Office of Financial Research, December.

Blundell-Wignall, Adrian, and Paul Atkinson, 2010, "Thinking Beyond Basel III: Necessary Solutions For Capital And Liquidity," *OECD Journal: Financial Market Trends*, 2010(1), 1-23. http://www.oecd.org/dataoecd/42/58/45314422.pdf

Bookstaber, Rick, 2007, *A Demon of Our Own Design*, Wiley.

Bottega, John, 2010, "Information Quality and the Financial Crisis," presentation slides, *MIT 2010 Information Quality Industry Symposium*, Cambridge, MA, July 15, 2010. http://mitiq.mit.edu/IQIS/2010/BottegaKeynote_MIT2010IQIS.pdf

Boyd, J., and Mark Gertler, 1994, "Are Banks Dead? Or Are the Reports Greatly Exaggerated?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 18(3), Summer, 2-23. http://www.minneapolisfed.org/research/QR/QR1831.pdf

Bradley, Harold, Robert Fawls, Robert E. Litan, and Fred Sommers, 2011, "Canaries in the Coal Mine: How the Rise in Settlement 'Fails' Creates Systemic Risk for Financial Firms and Investors," technical report, Ewing Marion Kauffman Foundation, March. http://www.kauffman.org/uploadedfiles/canaries-in-the-coal-mine-report.pdf

Brammertz, Willi, Ioannis Akkizidis, Wolfgang Breymann, Rami Entin, and Marco Rustmann, 2009, *Unified Financial Analysis : the missing links of finance*, Wiley. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470697156.html.

Brewer, Eric, 2000, "Towards Robust Distributed Systems," *Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing (PODC '00).* http://dl.acm.org/citation.cfm?id=343502

Brunnermeier, Markus K., Thomas M. Eisenbach and Yuliy Sannikov, 2011, "Macroeconomics with Financial Frictions: A Survey," working paper, Princeton University, July. https://www.princeton.edu/~markus/research/papers/survey_macroeconomics_frictions.pdf

Buchan, Iain, John Winn, and Chris Bishop, 2009, "A Unified Modeling Approach to Data-Intensive Healthcare," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 91-97. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part2_buchan.pdf

Caballero, Ricardo, and Alp Simsek, 2011, "Fire Sales in a Model of Complexity," working paper, MIT, April. http://econ-www.mit.edu/files/5580

Castura, Jeff, Robert Litzenberger, Richard Gorelick, Yogesh Dwivedi, 2010, "Market Efficiency and Microstructure Evolution in U.S. Equity Markets: A High-Frequency Perspective," working paper, RGM Advisors, August. http://fnce.wharton.upenn.edu/news/Litzenberger_transient_vol5_2010.pdf

Ciulla, Thomas, Daniel Bloom, and Justin Justin, 2010, "Automating the OTC derivative collateral management function," *Journal of Securities Operations & Custody*, 3(2), August, 103-110. http://henrystewart.metapress.com/openurl.asp?genre=article&issn=1753-1802&volume=3&issue=2&spage=103

Coase, R. H., 1937, "The Nature of the Firm," *Economica*, New Series, 4(16), November, 386-405. http://www.sonoma.edu/users/e/eyler/426/coase1.pdf

Committee on Payment and Settlement Systems (CPSS), 2008, "The interdependencies of payment and settlement systems," technical report, Bank for International Settlements, June. http://www.bis.org/publ/cpss84.pdf

______, 2011, "Principles for financial market infrastructures," consultative report, Bank for International Settlements, March. http://www.bis.org/publ/cpss94.pdf

Committee on Payment and Settlement Systems (CPSS), and Technical Committee of the International Organization of Securities Commissions (IOSCO), 2001, "Recommendations for Securities Settlement Systems: Report of the CPSS-IOSCO Joint Task Force on Securities Settlement Systems," Consultative Report, January. http://www.bis.org/publ/cpss42.pdf

Commodity Futures Trading Commission and Securities and Exchange Commission (CFTC-SEC), 2010a, "Preliminary Findings Regarding the Market Events of May 6, 2010," Report of the Staffs of the CFTC And SEC to the Joint Advisory Committee on Emerging Regulatory Issues, May. http://www.sec.gov/sec-cftc-prelimreport.pdf

______,, 2010b, "Findings Regarding the Market Events of May 6, 2010," Report of the Staffs of the CFTC And SEC to the Joint Advisory Committee on Emerging Regulatory Issues, September. www.sec.gov/news/studies/2010/marketevents-report.pdf

Counterparty Risk Management Policy Group (CRMPG), 2005, "Toward Greater Financial Stability: A Private Sector Perspective: The Report of the Counterparty Risk Management Policy Group II," July 27. http://www.crmpolicygroup.org/crmpg2/docs/CRMPG-II.pdf

Copeland, Adam, Antoine Martin, and Michael Walker, 2010, "The Tri-Party Repo Market before the 2010 Reforms," *Federal Reserve Bank of New York Staff Reports*, no. 477, November. http://www.copeland.marginalq.com/res_doc/sr477.pdf

Cordell, Larry, Yilin Huang, and Meredith Williams, 2011, "Collateral Damage: Sizing and Assessing the Subprime CDO Crisis," Working Paper 11-30, Federal Reserve Bank of Philadelphia, August. http://www.philadelphiafed.org/research-and-data/publications/working-papers/2011/wp11-30.pdf

Danielsson, Jon and Hyun Song Shin, 2003, "Endogenous Risk," Ch. 20 in: *Modern Risk Management: A History*, Risk Books, 297-314.

Data Documentation Initiative (DDI), 2009, "Data Documentation Initiative (DDI) Technical Specification Part I: Overview, Version 3.1," technical report, October. http://downloads.sourceforge.net/project/ddi-alliance/Data%20Documentation%20Initiative/DDI%203.1%20%282009-10-18%29/DDI_3_1_2009-10-18_Documentation_XMLSchema.zip?use_mirror=superb-dca2

DellaVigna, Stefano, and Joshua Pollet, 2009, "Investor Inattention and Friday Earnings Announcements," *Journal of Finance*, 64, April, 709-749.

DeMarzo, Peter M., 2005, "The Pooling and Tranching of Securities: A Model of Informed Intermediation," *Review of Financial Studies*, 18(1), Spring, 1-35.

Derygina, Mariya, 2009, "Standardization of Securities Regulation: Rehypothecation and Securities Commingling in the United States and the United Kingdom," *Review of Banking and Financial Law*, Vol. 29, Fall 2010 issue, 253-288. http://www.bu.edu/law/central/jd/organizations/journals/banking/archives/documents/volume29/Deryugina.pdf

Diestel, Reinhard, 2006, *Graph Theory*, *3rd edition*, Springer Verlag.

Duffie, Darrell, 2011, "Systemic Risk Exposures A 10-by-10-by-10 Approach," working paper, Stanford University, July. http://www.darrellduffie.com/uploads/policy/Duffie10By10By10July2011.pdf

Dungey, Doris, 2007a, "Tanta: Mortgage Servicing for UberNerds," Internet posting, *Calculated Risk*, 20 February 2007, downloaded 27 July 2011 from: http://www.calculatedriskblog.com/2007/02/tanta-mortgage-servicing-for-ubernerds.html

Dungey, Doris, 2007b, "Foreclosure Sales and REO for UberNerds," Internet posting, *Calculated Risk*, 7 April 2007, downloaded 27 July 2011 from: http://www.calculatedriskblog.com/2007/04/foreclosure-sales-and-reo-for-ubernerds.html

Engle, Robert F. and Weidman, Scott, 2010, Technical Capabilities Necessary for Regulation of Systemic Financial Risk: Summary of a Workshop, National Research Council of the National Academies, National Academies Press , Washington, DC, http://www.nap.edu/catalog.php?record_id=12841.

Enterprise Data Management Council (EDMC), 2011, Semantics Repository, Internet resource, downloaded 29 July 2011. http://www.edmcouncil.org/sec_semantics.aspx

European Central Bank (ECB), 2010a, "Financial Networks and Financial Stability," *Financial Stability Review*, June, 155-160. http://www.ecb.int/pub/pdf/other/financialstabilityreview201006en.pdf

______, 2010b, "New quantitative measures of systemic risk," *Financial Stability Review*, December, 147-153. http://www.ecb.int/pub/fsr/shared/pdf/ivefinancialstabilityreview201012en.pdf?abea0fb869314b1aca4c72f2d826b897

Financial Accounting Standards Board (FASB), 2008, "Conceptual Framework for Financial Reporting: The Reporting Entity," *Preliminary Views* 1580-100, May 29. http://www.fasb.org/pv_conceptual_framework_reporting_entity.pdf

Fitzgerald, Anne, Brian Fitzgerald, and Kylie Pappalardo, 2009, "The Future of Data Policy," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 201-208. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_fitzgerald.pdf

FIX Protocol (FIX), 2011, "What is FIX?" Internet resource, downloaded 29 July 2011. http://www.fixprotocol.org/what-is-fix.shtml

Federal Deposit Insurance Corporation (FDIC), 2011, "Statement of Martin J. Gruenberg, Acting Chairman, Federal Deposit Insurance Corporation on Enhanced Oversight after the Financial Crisis: Wall Street Reform at One Year before the Committee on Banking, Housing and Urban Affairs, United States Senate; 538 Dirksen Senate Office Building July 21, 2011," U.S. Senate Committee testimony. http://www.fdic.gov/news/news/speeches/chairman/spjul2111.html

Federal Reserve Bank of St. Louis, 2011a, "Delinquency Rate On Single-Family Residential Mortgages, Booked In Domestic Offices, All Commercial Banks (DRSFRMACBS)," *FRED Economic Data*, Internet resource, downloaded 28 July 2011. http://research.stlouisfed.org/fred2/series/DRSFRMACBS/downloaddata?cid=32440

______, 2011b, "Charge-Off Rate On Single Family Residential Mortgages, Booked In Domestic Offices, All Commercial Banks (CORSFRMACBS)," *FRED Economic Data*, Internet resource, downloaded 28 July 2011. http://research.stlouisfed.org/fred2/series/CORSFRMACBS/downloaddata?cid=32439

Feldman, Ron, and Mark Lueck, 2007, "Are Banks Really Dying This Time? An update of Boyd and Gertler." *The Region*, Federal Reserve Bank of Minneapolis, September, 6-51. http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=1139

Financial Products Markup Language (FpML), 2011, "What is FpML®?," Internet resource, downloaded 29 July 2011. http://www.fpml.org/about/factsheet.html

Flannery, Mark J., Simon H. Kwan, and M. Nimalendran, 2004, "Market evidence on the opaqueness of banking firms' assets," *Journal of Financial Economics*, 71(3), March, 419-460. http://www.sciencedirect.com/science/article/pii/S0304405X03001855

______, 2010, "The 2007-09 Financial Crisis and Bank Opaqueness," Working Paper 2010-27, Federal Reserve Bank of San Francisco, September. http://www.frbsf.org/publications/economics/papers/2010/wp10-27bk.pdf

Fleming, Michael, and Kenneth Garbade, 2005, "Explaining Settlement Fails," Current Issues in Economics and Finance, Federal Reserve Bank of New York, 11(9), September. http://www.ny.frb.org/research/current_issues/ci11-9.pdf

Flood, Mark, 2009, "Embracing change: financial informatics and risk analytics," *Quantitative Finance*, 9(3), 243-256. http://www.tandfonline.com/doi/abs/10.1080/14697680802366037

Flood, Mark, A. Pete Kyle, and Louiqa Raschid, 2010, *Knowledge Representation and Information Management for Financial Risk Management: Report of a Workshop, Arlington, VA*, National Science Foundation and Pew Charitable Trusts, July. http://irix.umiacs.umd.edu/docs/FIWreport-FINAL.pdf.

Fuerber, Christian, Martin Hepp, and Markus Wischnewski, 2011, "The Data Quality Constraints Library: Primer, Draft 2010-11-24," Internet site (downloaded 26 November 2011). http://semwebquality.org/documentation/primer/20101124/index.html

Ghoting, Amol, Srinivasan Parthasarathy and Matthew Otey, 2008, "Fast mining of distance-based outliers in high-dimensional datasets," *Data Mining and Knowledge Discovery*, 16(3), 349-364. http://www.springerlink.com/content/x58067325kv41n26/

Gilbert, R. Alton, 1992, "Implications of netting arrangements for bank risk in foreign exchange transactions," *Federal Reserve Bank of St. Louis Review*, 74(1), 3-16.

Gilbert, Seth, and Nancy Lynch, 2002, Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services, *ACM SIGACT News*, 33(2), June, 51-59. http://dl.acm.org/citation.cfm?doid=564585.564601

Gooch, Anthony, and Linda Klein, 1997, "A Review of Case Law Affecting Swaps and Derivative Instruments: United States Case Law," in: R. Schwartz and C. Smith, eds., *Derivatives Handbook: Risk Management and Control*, John Wiley & Sons, 57-134.

Gorton, Gary B., 2009, "Slapped in the Face by the Invisible Hand: Banking and the Panic of 2007," working paper, Yale University, May. http://www.frbatlanta.org/news/CONFEREN/09fmc/gorton.pdf

Gorton, Gary B. and Andrew Metrick, 2009, "Haircuts," Working Paper 15273, National Bureau of Economic Research (NBER), August. http://nber.org/papers/w15273.pdf

Gottfredson, Mark, and Keith Aspinall, 2005, "Innovation Versus Complexity: What Is Too Much of a Good Thing?" *Harvard Business Review*, November. http://hbr.org/2005/11/innovation-versus-complexity-what-is-too-much-of-a-good-thing/ar/1

Gray, Jim, David Liu, Maria Nieto-Santisteban, Alexander Szalay, Gerd Heber, and David DeWitt, 2005, "Scientific Data Management in the Coming Decade," *ACM SIGMOD Record*, 34(4), 35-41. http://arxiv.org/ftp/cs/papers/0502/0502008.pdf

Haldane, Andrew G., 2009, "Rethinking the Financial Network," Speech delivered at the Financial Student Association, Amsterdam, April, Bank of England. http://www.bankofengland.co.uk/publications/speeches/2009/speech386.pdf

Han, Jiawei, Micheline Kamber, and Jian Pei, 2011, *Data Mining: Concepts and Techniques, 3rd Edition*, Morgan Kaufmann.

Hansen, Charles, Chris Johnson, Valerio Pascucci, and Claudio Silva, 2009, "Visualization for Data-Intensive Science," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 153-163. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part3_hansen_johnson.pdf

Hartman-Glaser, Barney, 2011, "Reputation vs. Signaling in a Security Issuance Game," working paper, University of California at Berkeley, January. http://energy.uh.edu/departments/finance/documents/Barney%20main.pdf

He, Bin, Mitesh Patel, Zhen Zhang, and Kevin Chen-Chuan Chang, 2007, "Accessing the deep web," *Communications of the ACM*, 50(5), May, 94-101. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.9091&rep=rep1&type=pdf

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2011, "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, 66(1), February. http://faculty.haas.berkeley.edu/hender/Algo.pdf

Hilbert, Martin, and Priscila López, 2011a, "The World's Technological Capacity to Store, Communicate, and Compute Information," *International Journal of Communication*, forthcoming. http://www.martinhilbert.net/HilbertLopez%20InfoCapacityMethod.pdf

Hilbert, Martin, and Priscila López, 2011b, "How to measure the world's technological capacity to communicate, store and compute information?," *Science*, 332(6025), April, 60-65. http://www.sciencemag.org/content/332/6025/60

Holland, Peter, 2011, "The One Hundred Billion Dollar Problem in Small Claims Court: Robo-Signing and Lack of Proof in Debt Buyer Cases," *Journal of Business & Technology Law*, 6(2), 101-129. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1875727

Holmström, Bengt, and John Roberts, 1998, "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*, 12(4), Fall, 73–94.

Hunt, John, Richard Stanton and Nancy Wallace, 2011, The End of Mortgage Securitization? Electronic Registration as a Threat to Bankruptcy Remoteness, working paper, U. of California at Berkeley, August. http://faculty.haas.berkeley.edu/stanton/papers/pdf/mers.pdf

Ingersoll, Jonathan, Matthew Spiegel, William Goetzmann, and Ivo Welch, 2007, "Portfolio Performance Manipulation and Manipulation-proof Performance Measures," *Review of*

Financial Studies, 20(5), 1503-1546. http://research.ivo-welch.info/journalcopy/2007-rfs.pdf

Intel, 2010, "Extreme Speed, Extreme Throughput," marketing brochure. http://www.thesystech.com/docs/3853-8_IN_ThesysTechCaseStudy-hi.pdf

International Accounting Standards Board (IASB), 2011, "IFRS 13 Fair Value Measurement," guidance document, May 12. http://www.ifrs.org/Current+Projects/IASB+Projects/Fair+Value+Measurement/IFRS+13+Fair+Value+Measurement/IFRS+13+Fair+Value+Measurement.htm

International Monetary Fund (IMF), 2009, "Assessing the Systemic Implications of Financial Linkages," Global Financial Stability Review, Apr09, 73–110. http://www.imf.org/External/Pubs/FT/GFSR/2009/01/pdf/chap2.pdf

International Organization for Standardization (ISO20022), 2011, "ISO 20022 Universal financial industry message scheme," Internet resource, downloaded 29 July 2011. http://www.iso20022.org/

International Organization of Securities Commissions (IOSCO), 2008, "Report on the Subprime Crisis: Final Report," technical report, May. http://www.iosco.org/library/pubdocs/pdf/IOSCOPD273.pdf

______, 2010, "Disclosure Principles for Public Offerings and Listings of Asset-Backed Securities," technical report, April. http://www.iosco.org/library/pubdocs/pdf/IOSCOPD318.pdf

International Monetary Fund (IMF), 2001, "The Macroeconomic Statistical Treatment of Reverse Transactions, Fourteenth Meeting of the IMF Committee on Balance of Payments Statistics Tokyo, Japan, October 24-26, BOPCOM-01-16," technical report, Statistics Department. http://www.imf.org/external/pubs/ft/bop/2001/01-16.pdf

______, 2009, "Assessing the Systemic Implications of Financial Linkages," Global Financial Stability Review, Apr09, 73-110. http://www.imf.org/External/Pubs/FT/GFSR/2009/01/pdf/chap2.pdf

International Swaps and Derivatives Association (ISDA), 2010a, "Market Review of OTC Derivative Bilateral Collateralization Practices," guidance document, March. http://www.isda.org/c_and_a/pdf/Collateral-Market-Review.pdf

______, 2010b, "Best Practices for the OTC Derivatives Collateral Process," guidance document, June. http://www.isda.org/c_and_a/pdf/ISDA-Best-Practices-for-the-OTC-Derivatives-Collateral-Process.pdf

Johnson, C., R. Ross, S. Ahern, J. Ahrens, W. Bethel, K. L. Ma, M. Papka, J. van Rosendale, H. W. Shen, and J. Thomas, 2007, "Visualization and Knowledge Discovery: Report from the DOE/ASCR Workshop on Visual Analysis and Data Exploration at Extreme Scale," technical report, October. http://www.sci.utah.edu/vaw2007/DOE-Visualization-Report-2007.pdf

Kahn, Charles M, James McAndrews, William Roberds, 2003, "Settlement Risk under Gross and Net Settlement," *Journal of Money, Credit and Banking*, 35(4), August, 591-608.

Kashyap, Anil, Richard Berner, and Charles Goodhart, 2011, "The Macroprudential Toolkit," *IMF Economic Review*, International Monetary Fund, 59(2), June, 145-161. http://www.palgrave-journals.com/imfer/journal/v59/n2/full/imfer20114a.html

Kaufman, George G., and Larry R. Mote, 1994, "Is Banking a Declining Industry: A Historical Perspective," *Economic Perspectives*, Federal Reserve Bank of Chicago, 18(2), 2-21. http://www.chicagofed.org/digital_assets/publications/economic_perspectives/1994/ep_may_jun1994_part1_kaufman.pdf

Kaufman, Ted, Richard Neiman, Damon Silvers, J. Mark McWatters, and Kenneth Troske, 2010, "Examining the Consequences of Mortgage Irregularities for Financial Stability and Foreclosure Mitigation," *Congressional Oversight Panel: November Oversight Report*, U.S. Government Printing Office, November. http://www.gpo.gov/fdsys/pkg/CPRT-111JPRT61835/pdf/CPRT-111JPRT61835.pdf

Khandani, Amir E., and Andrew W. Lo, 2011, "What Happened To The Quants In August 2007?" Evidence from Factors and Transactions Data," *Journal of Financial Markets*, 14(1), 1-46. http://web.mit.edu/alo/www/Papers/august07.pdf

Kirilenko, Andrei, Mehrdad Samadi, Albert S. Kyle, Tugkan Tuzun, 2011, "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," working paper, Commodity Futures Trading Commission, January. http://business.nd.edu/uploadedFiles/Academic_Centers/Study_of_Financial_Regulation/pdf_and_documents/2011_conf_Andrei_Kirilenko.pdf

Koh, Heebyung, and Christopher L. Magee, 2006, "A functional approach for studying technological progress: Application to information technology?" *Technological Forecasting and Social Change*, 73, 1061–1083. http://web.mit.edu/iandeseminar/TFSsciencedirect2006_technologicalprogress_magee_koh.pdf

Larus, James, and Dennis Gannon, 2009, "Multicore Computing and Scientific Discovery," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 125-129. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part3_larus_gannon.pdf

Laux, Christian, and Christian Leuz, 2010, "Did Fair-Value Accounting Contribute to the Financial Crisis?" *Journal of Economic Perspectives*, 2010, 24, 93-118. http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.24.1.93

Leedom, D.K., 2001, Sensemaking Symposium, final report to the Command and Control Research Program, Office of the Assistant Secretary of Defense for Command, Control, Communications and Intelligence, U.S. Department of Defense. http://www.au.af.mil/au/awc/awcgate/ccrp/sensemaking_final_report.pdf

Levy, Dan and John Gittelsohn, 2011, "Foreclosure Filings Hit Three-Year Low as U.S. Servicers in 'Dysfunction'," *Bloomberg News*, 10 March 2011, downloaded 28 July 2011. http://www.bloomberg.com/news/print/2011-03-10/foreclosure-filings-drop-to-3-year-low-as-u-s-servicers-in-dysfunction-.html

Lo, Andrew K., 2011a, "Andrew Lo discusses systemic risk," presentation video, *UMD/CMU/NYU/UCB Conference on Systemic Risk and Data Issues,* Washington, DC, October 5, 2011. http://www.youtube.com/watch?v=nuDIoBeNwD0

_______, 2011b, "In the Valley of the Shadow of Hedge Funds," presentation slides, *UMD/CMU/NYU/UCB Conference on Systemic Risk and Data Issues*, Washington, DC, October 5, 2011. http://www.rhsmith.umd.edu/cfp/events/2011/confSystemicRisk1011/

Madnick, Stuart, and Hongwei Zhu, 2006, Improving data quality through effective use of data semantics, *Data & Knowledge Engineering*, 59(2), November, 460-475. http://www.sciencedirect.com/science/article/pii/S0169023X05001497

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, and Angela Hung Byers, 2011, "Big data: The next frontier for innovation, competition, and productivity," technical report, McKinsey Global Institute, May. http://msbapm.business.uconn.edu/files/2011/02/McKinsey_big_data_full_report.pdf

Markham, Jerry, 2002, *A Financial History of the United States, Volume 2: From J.P. Morgan to the Institutional Investor (1900-1970)*, M.E. Sharpe.

Mondria, Jordi, and Climent Quintana-Domeque, 2010, "Financial Contagion and Attention Allocation," working paper, University of Toronto, July. http://individual.utoronto.ca/jmondria/MQ_Contagion.pdf

Moore, Reagan, 2001, "Data Management Services," working paper, San Diego Supercomputing Center. http://sbc.omg.org/workshop/proceedings/Moore.doc

Moussa, Amal, 2011, "Contagion and Systemic Risk in Financial Networks," Ph.D. Thesis, Columbia University. http://academiccommons.columbia.edu/catalog/ac:131474

Mutnikas, Yaacov, and Michael Zerbs, 2011, "Optimal Granularity: A Supervisory Deep Dive on Counterparty Credit Risk," *Think*, November, 32-37. http://www.algorithmics.com/think/Nov11/Algo-THINK1111-OG-Mutnikas-Zerbs.pdf

Nagy, Béla, J. Doyne Farmer, Jessika E. Trancik, and John Paul Gonzales, 2011, "Superexponential Long-Term Trends in Information Technology," *Technological Forecasting and Social Change*, 73, 1061-1083. http://www.santafe.edu/media/workingpapers/10-11-030.pdf

Ng, Jeffrey, 2011, "The effect of information quality on liquidity risk," *Journal of Accounting and Economics*, 52(2-3), November, 126-143. http://www.sciencedirect.com/science/article/pii/S016541011100036X

Office of Financial Research (OFR), 2011, "Office of Financial Research Issues Statement on Progress to Date and Next Steps Forward in the Global Initiative to Establish a Legal Entity Identifier (LEI)," press release, August 12th. http://www.treasury.gov/press-center/press-releases/Pages/tg1275.aspx

Pandis, Ippokratis, Ryan Johnson, Nikos Hardavellas, and Anastasia Ailamaki, 2010, "Data-Oriented Transaction Execution," *Proceedings of the VLDB Endowment*, 3(1), 928-939. http://www.vldb.org/pvldb/vldb2010/pvldb_vol3/R83.pdf

Parastatidis, Savas, 2009, "A Platform for All That We Know: Creating a Knowledge-Driven Research Infrastructure," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 165-172. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part3_parastatidis.pdf

Perrow, Charles, 1999, *Normal Accidents: Living with High-Risk Technologies*, Princeton University Press.

Petersen, Mitchell A. and Raghuram G. Rajan, 2002, "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, 57(6), December.

Pozsar, Zoltan, Tobias Adrian, Adam Ashcraft, and Haley Boesky, 2010, "Shadow Banking," Staff Report no. 458, Federal Reserve Bank of New York, July. http://www.newyorkfed.org/research/staff_reports/sr458.pdf

Pozsar, Zoltan, and Manmohan Singh, 2011, "The Non-Bank-Bank Nexus and the Shadow Banking System," IMF Working Paper WP/11/289, December. http://www.imf.org/external/pubs/ft/wp/2011/wp11289.pdf

Philippon, Thomas, 2008, "Why Has the Financial Sector Grown So Much? The Role of Corporate Finance," NBER Working Paper No. 13405, National Bureau of Economic Research, March. http://w4.stern.nyu.edu/emplibrary/philippon.pdf

Securities and Exchange Commission (SEC), 1972, *Thirty-Eighth Annual Report of the Securities and Exchange Commission, for the Fiscal Year Ended June 30th, 1972*. http://www.sec.gov/about/annual_report/1972.pdf

Senior Managements of: Bank of America, N.A., Barclays Capital, Bear, Stearns & Co., Citigroup, Credit Suisse First Boston, Deutsche Bank AG, Goldman, Sachs & Co., HSBC Group, JP Morgan Chase, Lehman Brothers, Merrill Lynch & Co., Morgan Stanley, UBS AG, Wachovia Bank, N.A. (Senior Managements), 2005, "Letter to Mr. Timothy Geithner, President, Federal Reserve Bank of New York," 4 October 2005. http://www.newyorkfed.org/newsevents/news_archive/markets/2005/industryletter.pdf

Shilling, Gary, 2011, *The Age of Deleveraging: Investment Strategies for a Decade of Slow Growth and Inflation*, Wiley.

Shleifer, Andrei and Robert Vishny, 2010, "Fire Sales in Finance and Macroeconomics," *Journal of Economic Perspectives*, forthcoming.

Singh, Manmohan and James Aitken, 2010, The (sizable) Role of Rehypothecation in the Shadow Banking System, IMF Working Paper WP/10/172, July. http://www.imf.org/external/pubs/ft/wp/2010/wp10172.pdf

Srivastava, Utkarsh Hriday, 2006, "Efficient Query Processing for Modern Data Management," Doctoral Thesis, Stanford University, September. http://ilpubs.stanford.edu:8090/786/1/2006-20.pdf

Stonebraker, Michael, Jacek Becla, David Dewitt, Kian-Tat Lim, David Maier, Oliver Ratzesberger, and Stan Zdonik, 2009, "Requirements for Science Data Bases and SciDB," *CIDR Perspectives*, Conference on Innovative Data Systems Research (CIDR). http://www.scidb.org/Documents/SciDB-CIDR2009.pdf

Stonebraker, Michael, Samuel Madden, Daniel J. Abadi, Stavros Harizopoulos, Nabil Hachem, Pat Helland, 2007, "The End of an Architectural Era (It's Time for a Complete Rewrite)," In: *VLDB '07: Proceedings of the 33rd international conference on Very Large Data Bases*, VLDB Endowment, Vienna, 1150-1160. http://hstore.cs.brown.edu/papers/hstore-endofera.pdf

Szalay, Alexander, and José Blakeley, 2009, "Gray's Laws: Database-centric Computing in Science," in: T. Hey, S. Tansley, and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 5-11. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part1_szalay.pdf

Taub, Leon, 2008, "Borrowed securities: implications for measuring cross-border portfolio investment," in: *IFC Bulletin No. 28, The IFC's contribution to the 56th ISI Session, Lisbon, August 2007*, Irving Fisher Committee on Central Bank Statistics (IFC), August, 169-193. http://www.bis.org/ifc/publ/ifcb28.pdf

Thaler, Richard H., 1988, "Anomalies: The Winner's Curse," *Journal of Economic Perspectives*, 2(1), Winter, 191-202.

Treasury Market Practices Group (TMPG), 2011, "Understanding Settleme nt Fails in Agency Mortgage-Backed Securities," technical report, Federal Reserve Bank of New York, April. http://www.newyorkfed.org/tmpg/tmpg_04292011.pdf

Trimbath, Susanne, 2008, "Trade Settlement Failures in U.S. Bond Markets," Working Paper STP2007_01 (v.2), STP Advisory Services, September. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1016873

Tufano, Peter, 2003, "Financial Innovation," In: G. Constantinides, M. Harris, and R. Stulz, Eds., *Handbook of the Economics of Finance, Volume 1a: Corporate Finance*. Elsevier, 307-336.

U.S. Census Bureau, 2011, "Total Midyear Population for the World: 1950-2050," *International Data Base (IDB)*, Internet resource, downloaded 26 July 2011. http://www.census.gov/population/international/data/idb/worldpoptotal.php.

U.S. Congress, 2010, "Dodd-Frank Wall Street Reform and Consumer Protection Act, Bill Summary & Status 111th Congress (2009-2010), H.R.4173," Enrolled Bill. http://thomas.loc.gov/cgi-bin/query/z?c111:h4173:

U.S. Senate, 2010, "The National Institute of Finance Act of 2010, Bill Summary & Status 111th Congress (2009-2010), S.3005," introduced 4 February 2010. http://thomas.loc.gov/cgi-bin/bdquery/z?d111:s3005:

Vastag, Brian, 2011, "Exabytes: Documenting the 'digital age' and huge growth in computing capacity," *Washington Post*, February 10. http://www.washingtonpost.com/wp-dyn/content/article/2011/02/10/AR2011021004916.html

Vogels, Werner, 2009, "Eventually Consistent," *Communications of the ACM*, 52(1), January, 40-44. http://www.cs.brown.edu/courses/csci2950-u/papers/p40-vogels.pdf

Wagner, Michael, Andrew Moore, and Ron Aryel, 2006, *Handbook of Biosurveillance*, Academic Press.

Wallace, Nancy, 2011, "Private-label Residential Mortgage Securitization: Recording Innovations and Bankruptcy Remoteness," presentation slides, U. of California at Berkeley, March. http://www.law.berkeley.edu/files/bclbe/Wallace_MERS.pdf

Wikipedia, 2011, "Transistor count," Internet resource, downloaded 22 July 2011. http://en.wikipedia.org/wiki/Transistor_count

Yahoo Finance, 2011, "S&P 500 INDEX,RTH (^GSPC)," Internet resource, downloaded 22 July 2011. http://finance.yahoo.com/q/hp?s=%5EGSPC+Historical+Prices

Zhang, Guo-Qing, Guo-Qiang Zhang, Qing-Feng Yang, Su-Qi Cheng, and Tao Zhou, 2008, "Evolution of the Internet and its Cores," *New Journal of Physics*, 10, December, 123027. http://iopscience.iop.org/1367-2630/10/12/123027

6 Figures

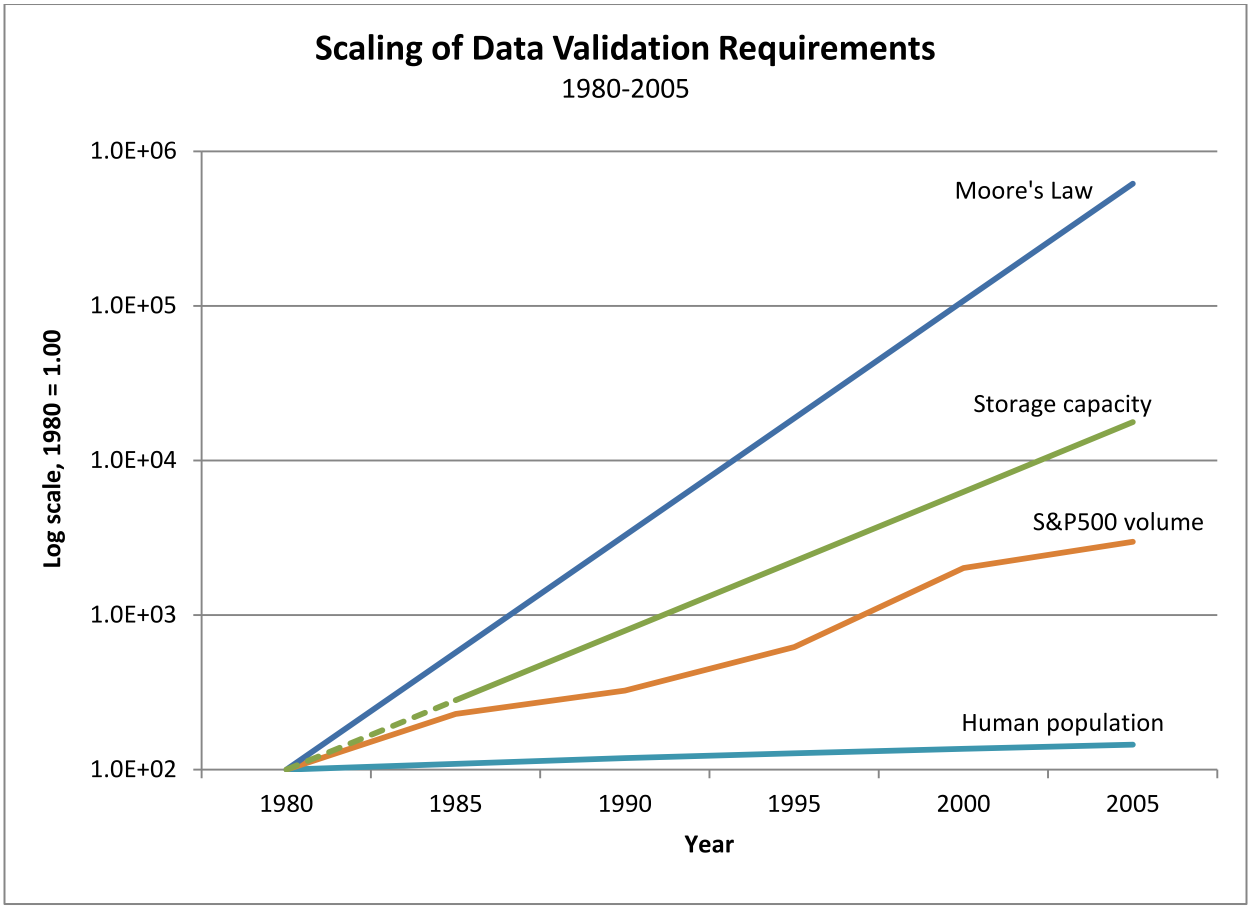


Figure 1: Differential exponential growth rates for data validation requirements

Moore's Law *is estimated as a linear regression of transistor densities (in logarithms) against year of introduction over the 1971-2011 period; data were downloaded from Wikipedia (2011).* ***Storage capacity*** *is based on the average annual growth estimate (23% per year) of Hilbert and López (2011a) for the 1986-2007 period, extrapolated back to cover the 1980-1985 interval.* ***S&P500 trading volume*** *was downloaded from Yahoo Finance (2011).* ***Human population*** *is based on total midyear world population, from the U.S. Census Bureau (2011).*

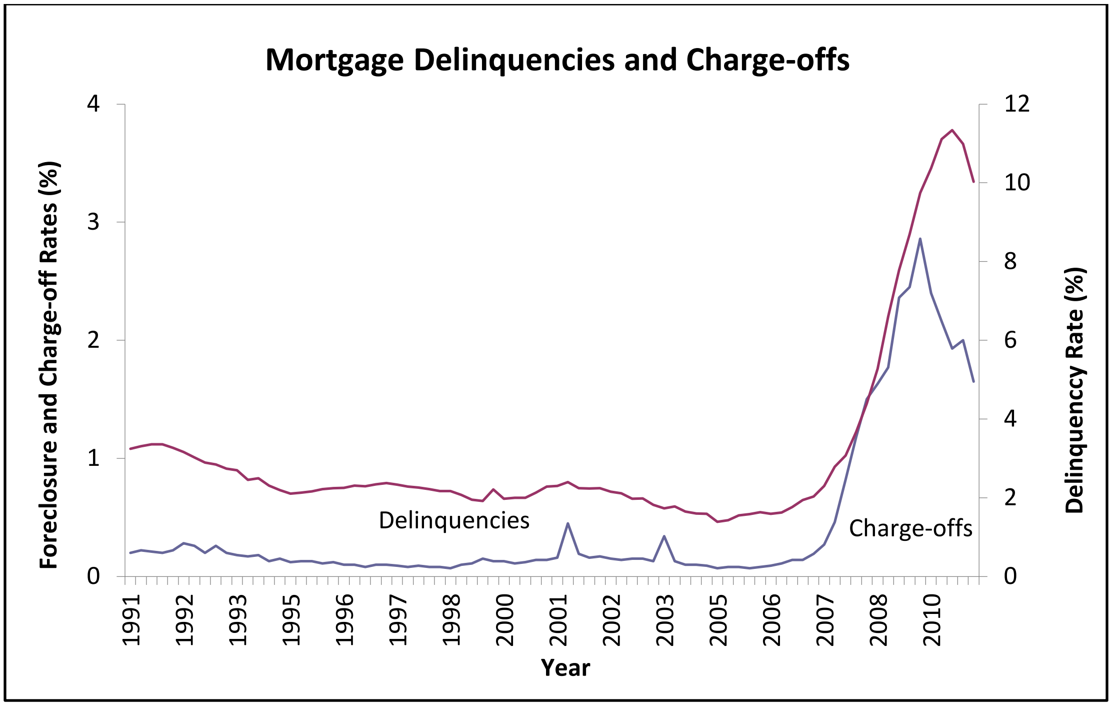
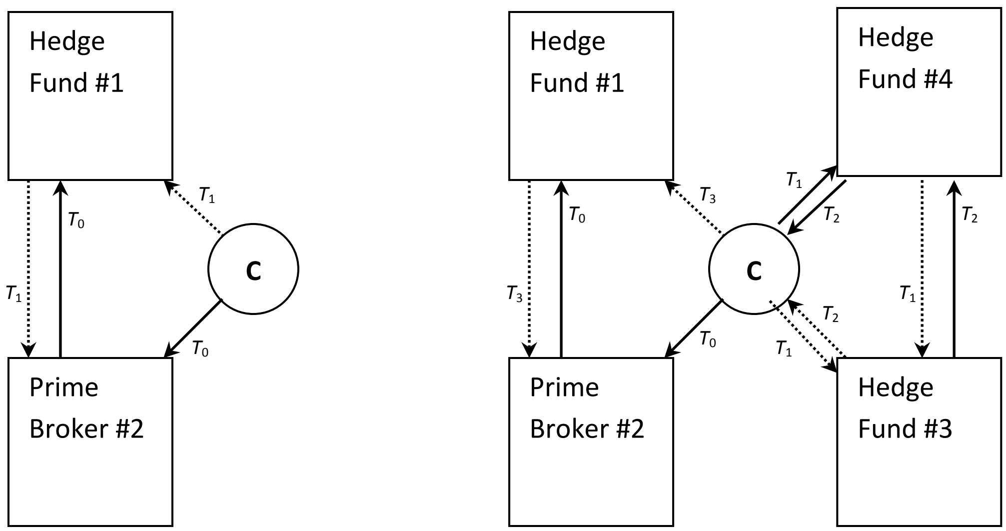


Figure 2: Overwhelming the foreclosure-processing infrastructure

Delinquencies *represent balances on delinquent loans for single-family residential mortgages held by all U.S. commercial banks as a percent of all such loans; data are taken from Federal Reserve Bank of St. Louis (2011a), series DRSFRMACBS.* ***Charge-offs*** *represent balances on charged-off loans for single-family residential mortgages held by all U.S. commercial banks, as a percent of all such loans; data are taken from Federal Reserve Bank of St. Louis (2011b), series CORSFRMACBS.*



Simple repo   Repo with rehypothecation for a short sale

Figure 3: The impact of rehypothecation on interconnectedness

Collateral is indicated by the circle containing a "C" in both examples. For a ***simple repo****, the prime broker pays cash at T_0 and receives the collateral; at T_1, the collateral is returned to the hedge fund, which repays the cash with interest. In* ***repo with rehypothecation for a short sale****, the prime broker lends the collateral at time T_1 to hedge fund #3, who promptly sells it to hedge fund #4. At time T_2, the short sale is reversed, and the collateral returned to the prime broker. At time T_3, the original repo is unwound.*