IDSS

SUMMARY REPORT

Workshop on Data, Analytics, and Risk in Finance

April 8, 2016

On April 8, 2016, MIT's new **Institute for Data, Systems, and Society** convened the Workshop on Data, Analytics, and Risk in Finance. The workshop's nearly 200 participants included representatives from more than 50 companies and regulatory agencies, as well as top academics in economics, finance, business, math, and engineering. They came together with the common goal of creating a more robust, transparent, and resilient global financial system.

Technological advancements have transformed financial markets into complex and dynamic systems. Organizations are increasingly interconnected and are processing transactions at sub-millisecond time scales while generating, storing, and managing masses of data. Big data, machine learning, predictive analytics, and encryption offer new opportunities for understanding and managing risk in financial systems. Consequently, it has become imperative to develop better models, deeper insights, and clearer understandings of the mechanisms that drive the financial system. There is much work to be done to rigorously test data-driven models that help us to understand what is happening and why, address the underlying theory, and translate these models into better policy and business decisions.

Key to this endeavor is a multidisciplinary approach: combining the efforts of theorists, empiricists, practitioners, and academics to obtain a broad view of the system and to see the less-obvious interactions brought to light by the massive amounts of data that were previously unavailable to researchers.

An exciting range of perspectives, ideas, and opportunities for collaboration came from the workshop. At IDSS, we continue to promote a research agenda focusing on systemic risk through the development of data-driven and decision-oriented models that allow for early detection and mitigation of such risk. Although this type of research program faces challenges, we intend to examine classes of interconnected financial systems for which concrete analysis can be performed and metrics can be derived.

IDSS would like to thank all of the Workshop speakers and participants for their contribution. We invite you to join us in taking the next step in advancing the research and collaborations that will drive progress toward these goals and enable all stakeholders to benefit from a more resilient and stable financial system.

— **MIT IDSS WORKSHOP ORGANIZERS**

Professor Munther Dahleh
Director, IDSS

Professor Asu Ozdaglar
Director, LIDS

Professor Andrew W. Lo
Director, Lab for Financial Engineering

Ms. Elizabeth Bruce
Executive Director, IDSS

Ms. Jackie Wilbur
External Relations Officer, IDSS

INTRODUCTION

This report provides expert commentary in two key areas: first, systemic risk; second, big data and analytics in finance. These insights continue the conversation from the recent MIT **IDSS Workshop on Data, Analytics, and Risk in Finance**. The event's morning session featured both industry representatives and regulators, including panelists from the Federal Reserve Bank of Boston and the Office of Financial Research of the U.S. Department of the Treasury. They joined a host of MIT professors in discussing the need to better understand risk and resiliency in financial systems. The afternoon sessions put the spotlight on opportunities for innovation, big data, and analytics related to assessing and managing risk, as well as developing new products and services.

FINANCIAL STABILITY, MONITORING, AND SYSTEMIC RISK

Systemic risk in the world's financial systems is a complex phenomenon. A major factor in the 2008 global financial crisis, systemic risk occurs when a triggering event, such as the collapse of a major financial institution, creates the potential for a domino effect of further failures in connected institutions and sectors. The trigger can be a large event or a small, idiosyncratic shock—a factor that complicates the picture of systemic risk and how it is understood, modeled, and mitigated.

IDSS workshop participants discussed systemic risk from a variety of angles, drawing out the common challenges across sectors and generating new ideas for collaboration.

Overarching Challenges in Systemic Risk

Economist Simon Johnson, professor of entrepreneurship at the MIT Sloan School of Management and former chief economist of the International Monetary Fund (IMF), framed a portion of the systemic risk conversation by articulating a series of key questions:

- How can we better understand why financial systems fail? How do we identify systemic risk and its sources?

- How do we know big financial institutions have enough equity to absorb shocks in the system? How do we define and measure equity?

- Given financial institutions are becoming increasingly complex and interconnected, can regulators address the failure of major institutions—those designated as systemically important financial institutions (SIFIs)—in a way that prevents system-wide failure?

- Is bigger better? Does greater centralization bring more stability? Are big banks actually more resilient, and how do we measure this?

- Will increased regulation of financial institutions bring more stability or less?

- Can institutions other than banks, such as insurance companies and those institutions in the so-called "shadow banking" system, pose significant systemic risk? If so, what should be done to monitor this?

Identifying Systemic Risk

As Johnson noted, a key issue in mitigating financial systemic risk is to identify the threat and its sources. For example, at an institutional level, systemic risk may be hidden when bankers and regulators fail to understand potential correlations among different assets or sectors.

Revisiting the 2008 financial crisis and resulting recession, Paul Willen, senior economist and policy advisor at the Federal Reserve Bank of Boston, remarked that banks had turned $64 billion of subprime mortgages into $160 billion of collateralized debt obligations (CDOs). Regulators didn't view the CDOs as housing-market exposure on bank balance sheets. Instead, the CDOs were seen as part of the banks' debt portfolios.

As a result, regulators and banks made incorrect assumptions about the magnitude of their exposure to the volatility in the subprime housing markets. "The reason banks made these decisions is that they were very optimistic about housing prices. Nobody believed housing prices could fall," Willen noted. He emphasized the need to better understand the expectations of the housing-market players. One way to accomplish this is collecting and studying different kinds of data not currently captured by regulators. Better understanding large-scale patterns in behavior might have allowed the overly optimistic view of what was happening industry-wide to be recognized before it was too late.

In contrast, "our definition of systemic risk is all correlations going to one in a portfolio," said Stephen Malinak, global head of persistent analytics and data science at Thomson Reuters, which develops and

> **Is bigger better? Does greater centralization bring more stability? Are big banks actually more resilient, and how do we measure this?**

markets hundreds of data sets for use by investors. "Many factors can drive securities prices and cause exposure to multiple varieties of shocks," he noted. Investors who seek to reduce volatility in their portfolios diversify. If, however, a crisis causes all assets to drop sharply in value simultaneously, the benefit of diversification is lost.

Since the 2008 crisis, institutions have changed the ways they model potentially risky situations. Citibank, for example, has designed financial scenarios that it runs regularly to determine whether its positions in any area pose a systemic risk, according to Amitabh Arora, head of Citigroup's systemic risk group. Citibank tests for "one-in-50 events" that would cause the bank to breach capital-ratio requirements. It also tests for "one-in-10 events" that would wipe out the bank's quarterly earnings. The goal is to make sure the bank has sufficient buffers so that no one scenario would create such internal risks.

One lesson from the 2008 financial crisis is the need to "focus on quantities and not just ratios," said Arora, who, prior to Citigroup, worked for Lehman Brothers. For example, if banks' corporate customers start parking huge deposits in their short-term accounts, the banks need to understand when those clients might remove the deposits, since the amounts could have a significant impact on their risk scenarios. Citibank has also created a system for identifying off-balance-sheet liabilities that might magnify exposures and includes them in its modeling scenarios.

Network Effects

Modeling network interactions is critical to gaining a better understanding of how our financial systems work. For the decade from 1998 to 2008, financial institutions had maintained that the way to reduce systemic risk was to share risk widely, according to Daron Acemoğlu, the Elizabeth and James Killian Professor of Economics at

MIT. But after September 15, 2008—the date Lehman Brothers Holdings filed for bankruptcy, intensifying the financial crisis—it became clear that shared risk had not mitigated the effects of the crisis but had amplified them instead.

In studying this unexpected outcome, Acemoğlu and his colleagues discovered that in densely connected networks, widely shared risk flips from being a source of stability to a source of fragility. At times of maximum stress, there is a kind of phase change, Acemoğlu said. "A sharp threshold occurs, and everything gets turned on its head. Then you want weakly connected networks —islands of institutions that trade with each other" to help prevent failures from propagating throughout the system.

To be able to predict network effects like these, Acemoğlu said, "you need a mathematical model that corresponds to the underlying economic mechanism." Although difficult to achieve, these network models are clearly useful for understanding interconnections that can lead to cascading events in financial systems— whether bubbles or collapses—and help us design systems that are inherently more resilient.

Modeling Systemic Risk

Mark Flood, a financial economist and research principal for the OFR, advocated the use of big data for monitoring financial stability and emphasized the need for selective modeling to do this well. "You can't monitor and model everything all the time," he said. "You need mechanisms for selective modeling with the ability to drill down when necessary." Developing better models that flag areas of high risk could provide regulators with better tools for identifying potential financial crises at an earlier stage.

A major challenge involves the implementation issues associated with managing large, diverse data sets.

Because financial systems are so large, regulators need to consider:

1. Coverage (how much data to collect).

2. Frequency (how often to sample data).

3. Granularity (what level of detail is needed).

Lack of coordination is a major data challenge for regulators in both the United States and Europe. Each regulator collects massive amounts of data related to a specific area, but data is typically managed in isolated silos. Because the data isn't coordinated, or identified consistently, risks can remain invisible until they cause significant problems.

Another challenge is timeliness. Trading algorithms execute orders in less than one second—far faster than data is collected by any regulator. An example from macroprudential supervision is real-time monitoring of high-frequency data streams during a flash crash, as cited in a recent OFR working paper, "Big Data Challenges and Opportunities in Financial Stability Monitoring."

The financial industry is also developing new data standards, such as the Global Legal Entity Identifier, which provides unique identifiers for any entity participating in financial transactions worldwide. However, there is a need for additional industry standards that could improve data management and sharing, ultimately reducing costs and inefficiencies— benefits to both industry and government stakeholders.

It is openly acknowledged that the current relationship between regulators and banks is strained, due in part to a perceived lack of trust on both sides and the banks' fear that regulators will place undue burdens on their industry. However, MIT's Johnson hinted that improved system-risk models, increased transparency, and new kinds of aggregate risk metrics could lead to a future with much less onerous regulatory requirements and might provide the tools needed to rebuild trust. Participants strongly agreed that industry players could collectively do a much better job at risk monitoring and management, that better risk models could inform better policies, and that everyone in the room—and well beyond—should have a vested interest in getting it right.

It is unlikely that researchers will be able to accurately predict rare, or "black-swan," events, such as an economic collapse, in the near future. Big data and predictive analytics can't provide a foolproof way to identify all critical problems before they happen. However, developing better early-warning indicators and ways to monitor patterns at scale, as well as quickly resolving problems once they occur, are approaches more likely to head off disaster and are worthwhile endeavors for this community.

Looking Ahead

Managing systemic financial risk presents plenty of hard challenges. "The systemic risk agenda has to be at the top of research topics from a social perspective," said Darryll Hendricks, chief operating officer of UBS Investment Bank's Non-Core and Legacy Division. Clearly, we need to use more advanced analytics techniques to understand domino effects and cascading risk.

By improving transparency and making more data available, market makers and regulators can improve

> " We're trying to use new techniques to tackle problems... We're trying to get new data and new ways of looking at that data. "
>
> – Andrew W. Lo, Director, MIT Laboratory for Financial Engineering and the Charles E. and Susan T. Harris Professor of Finance at Sloan

their ability to avoid financial meltdowns, participants said. The hitch: most financial firms are, for good reason, wary of sharing more data. Banks and hedge funds closely guard their proprietary trading strategies and hedging techniques as trade secrets. Issues involving access to data for research must be addressed to develop new ideas, models, tools, and techniques.

Regulatory officials and academic researchers indicated that they would like to see more transparency around the flows of funds among market players and internationally. For example, the U.S. Consumer Financial Protection Bureau would appreciate ways to securely share data with housing-market regulators.

Participants agreed that academic centers such as MIT could serve as trusted data brokers for research, aggregating critical data from many different sources. New public/private partnerships are clearly needed to share the type of data required to conduct research on vast and fast-moving financial markets and create new publicly available risk models. Finally, researchers need efficient methods for discovering, accessing, and integrating data at scale while managing proprietary information and privacy risks.

INNOVATION, BIG DATA, AND ANALYTICS

Many financial institutions are finding innovative ways to use data and analytics to improve their internal systems for assessing risk, making operations more auditable, and providing actionable insights to the market.

"We're trying to use new techniques to tackle problems," said Andrew W. Lo, director of the MIT Laboratory for Financial Engineering and the Charles E. and Susan T. Harris Professor of Finance at Sloan. "We're trying to get new data and new ways of looking at that data."

UBS's Hendricks called the criticality of data "hard to overstate." For bankers, he said, "the regulatory side has turned into data manipulation, analysis, and submission."

Both industry players and regulators are looking at ways to use non-traditional data—from sources such as text mining, social media, satellite imagery, and drone data—to derive accurate, real-time insights into what is happening in different markets.

Automating Compliance and Auditability

Compliance managers have an interest in using data more effectively to monitor risk within institutions. But there's a natural tension between the need to watch data and the temptation to micromanage it. Corporate boards of directors struggle with the conflicts between wanting to delegate operational decision making to managers while remaining aware of risks that could threaten the firm's viability.

The challenge for compliance officers is knowing how to use all the data collected. Regulated activities are both data-rich and messy. Because different regulators

> **Banking and insurance industries are undergoing a period of transformation as new regulations and technologies rapidly change the ways consumers manage loans, invest, save money, and protect their assets.**

require institutions to collect different kinds of data for different purposes, it can be difficult to combine data effectively in ways that yield timely, accurate insights.

Analyzing communications among employees and customers is crucial, said Peter Ferns, a technology fellow and senior engineer at Goldman Sachs. Almost every compliance problem results from personal interactions, he explained. By identifying communications from insiders involved in a merger, for example, compliance monitors can look for stock trading by people with whom they are in contact.

Similarly to other investment banks, Goldman Sachs has created a "data lake" that contains much of the information the company generates and makes it available for analytics. It collects hundreds of millions of transactions and six million electronic communications daily, each registered to its producer. While the text-mining technology was initially developed to deal with U.S. Senate hearings and regulatory issues, Ferns said it's now being expanded for use in customer relationship management and human relations functions.

Digital Currencies and Blockchain

Blockchain technology intrigues many in the industry. It holds the potential for creating a transparent view of collateral and payments. This distributed-ledger technology eliminates the need for central clearinghouses and trusted third parties to verify trading partners. In theory, it could increase security and reduce fraud. Bitcoin is currently the best-known example of

blockchain technology in use, but researchers anticipate alternative designs to guarantee secure transactions in the future.

Financial institutions are experimenting with a number of different blockchain designs. Cryptography expert Silvio Micali, the Ford Professor of Engineering in MIT's Electrical Engineering and Computer Science Department, helped develop a blockchain called Algorand, a new platform for digital currency and fast settlement of stock trades. Algorand requires much less computation than Bitcoin's exchange, making it more computationally efficient.

Paul Wojcik, chief risk officer at fund manager T. Rowe Price, summed up blockchain's potential impact: "It will revolutionize the whole structure of counterparty settlement."

Computing on Encrypted Data

Alex "Sandy" Pentland, the Toshiba Professor of Media Arts and Science at MIT, described Enigma, a blockchain solution recently developed by MIT researchers. Enigma encrypts data by breaking it up into chunks that are distributed to computer nodes in the Enigma network, said Pentland, who is also director of MIT's Connection Science and Human Dynamics labs. Each node processes its chunk of data, which is then reassembled. The ownership of each piece of data is defined by metadata stored on multiple computers in a Bitcoin-like design that ensures they can't be forged or counterfeited.

If financial institutions are confident they could share data without revealing proprietary secrets, platforms like Enigma have the potential to become a trusted means for industry stakeholders to develop new aggregate risk metrics and better understand systemic risk.

Predicting Real-Time Markets

For any investor, the seemingly unattainable Holy Grail would be the ability to predict the future. Toward that end, Devavrat Shah, an associate professor in MIT's Department of Electrical Engineering and Computer Science, has devised a way to accurately predict movements in the dollar price of highly volatile Bitcoins. After testing various models, Shah said, he and a colleague determined that by using Bayesian regression analysis, they could predict Bitcoin price movements so well they achieved an 89 percent return over 50 days and more than 2,000 trades.

Innovative Applications for Big Data

Banking and insurance industries are undergoing a period of transformation as new regulations and technologies rapidly change the ways consumers manage loans, invest, save money, and protect their assets. Anxious to avoid being bypassed, investment managers and bankers are racing to explore and exploit data and novel sources of information.

Financial market makers are intrigued by the potential of using new, unstructured data to improve returns and reduce risk. "Potential for self-reinforcing feedback is an area where there's a lot of value in doing more research," said UBS's Hendricks. Arora, of Citigroup, encourages more research that uses big data analytics to spot changes in investor attitudes and behavior, especially behaviors that might translate into rapid changes and/or higher risk in the market.

Bankers are eager to take advantage of technology that provides advantages for their clients. "We search to provide insights to clients," Hendricks said. For example, companies are deploying drones in regions around the world to collect real-time data on new construction and general trends for clients in the real-estate sector.

Banks are using technology to better understand investors' needs, interests, and trends. Similar to how Amazon tracks what its customers are reading on their Kindle devices, banks "can track how much time users spend on each sentence in research reports," Hendricks noted. That capability better informs the bank about what information investors value and helps analysts improve and customize their reports.
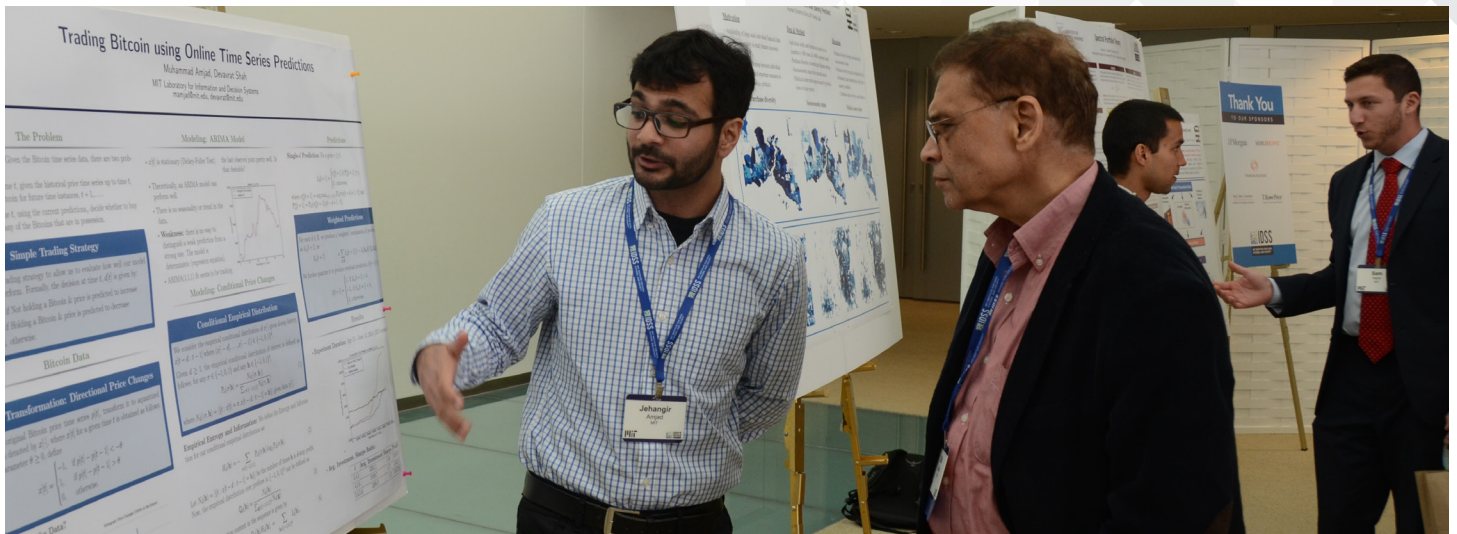
James Darby Nielson, managing director of quantitative research at Fidelity Management and Research, described his mission as enhancing stock picking with previously unexplored data sources including social media data. For instance, when a fund analyst wanted to analyze consumer reaction to the illnesses customers experienced recently after eating at Chipotle Mexican Grill restaurants, the analyst used Twitter to assess the strength of the negative reaction.

Retail analysts are using satellite imagery of shopping-mall parking lots to assess retail trends, Nielson said. Similarly, AIG uses satellite imagery to assess damage from hurricanes, even before adjusters can reach an affected site. Satellite and sensor data can also be used to assess agricultural markets—providing better data on how much acreage was planted by crop—leading to better predictions of commodity futures pricing.

Meanwhile, the insurance industry is looking at new ways to use natural-language processing. A recent project, for example, attempts to better assess potential chemical hazards. With new chemicals

entering the market faster than ever, insurers want better early-warning systems for possible risks. Mining the 20,000+ related academic articles published each year can help identify and assess risk by flagging important and harmful toxins before governments even start regulating them.

Looking Ahead

In addition to enabling sophisticated risk-management strategies, big data, machine learning, and predictive analytics hold great promise in creating new financial insights, products, and services. MIT's Shah asked: "How do we build good recommendation algorithms for investing?" Today, we use Yelp to help us decide where—or where not—to dine out, in the future we will have recommendation algorithms that provide personalized advice on how to invest.

In just a few years, a trillion sensors will provide us with real-time data streams worldwide, leading Siddhartha Dalal, AIG's chief data scientist, to ask one of the workshop's most provocative questions: "How do you use unconventional data to get ahead of high-severity, long-tail events?"

New kinds of data, including data from social media, along with natural-language processing tools will provide us with much greater insight into human and institutional behavior. We can use these tools to better understand investor behavior, including irrational behavior and "herding" (in which investors copy the behavior of other investors), which can trigger bubbles and collapses in markets.

> **Whether predictive analytics and big data are facilitating innovative uses of data or enabling more effective management of systemic risk, they're clearly critical to the future of our financial ecosystem.**

A better understanding of behavior in financial markets will better inform policies and decision making as well as aid risk-management strategies. So far, social media analytics has provided only short-term signals that are of limited use to fundamental analysis. Those signals are most useful after an event, "to explain rather than predict," as Fidelity's Nielson put it. But it is an area that's certainly ripe for future exploration and research.

One activity that would benefit all stakeholders would be finding new ways to manage and share data across industries as well as with academic researchers and government regulators. Sharing data across agencies could greatly improve efficiencies if, for example, company filings with the U.S. Securities and Exchange Commission could also be used by other regulators.

Another area of interest is developing new practical encryption technologies for securely sharing data with regulators, researchers, and other trusted parties. The research challenge here lies in developing practical implementations that can guarantee market players that their financial information and proprietary strategies will remain secure, and do not add significant overheads in terms of cost, management, and computational burden.

Whether predictive analytics and big data are facilitating innovative uses of data or enabling more effective management of systemic risk, they're clearly critical to the future of our financial ecosystem. Organizations such as banks, asset-management firms, technology companies, and regulatory agencies are all looking at how new data and better models will play an increasingly important role in ensuring financial stability as well as opening the door to new products and markets.

The agenda and full roster of speakers for the workshop can be found at **idssfinance2016.mit.edu.**

Produced by MIT Technology Review Custom

MIT Technology Review Custom produces world-class print, online, and live-event solutions that align clients with a trusted 117-year-old brand.

www.technologyreview.com/media

All event photographs were taken by Erica Bates.

ABOUT THE MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

The MIT Institute for Data, Systems, and Society (IDSS) is committed to addressing complex societal challenges by advancing education and research at the intersection of statistics, data science, information and decision systems, and social science.

Our lives are shaped by complex, interconnected systems, from the energy grids that power our homes to our transportation networks. In the 21st century, these systems are generating vast amounts of data and becoming increasingly interconnected, a proliferation that is redefining fields from economics to engineering.

IDSS is at the forefront of these developments. Bringing together researchers from all five MIT schools, IDSS uses rigorous analytical methods to better model, understand, and improve society's most complex systems. Our work drives important discoveries about resilience, risk, and sustainability globally, with specific application to areas like finance, transportation, energy, and health.

For details on the IDSS Industry Program or the IDSS Finance and Risk Initiative, please e-mail **idss-partners@mit.edu**.

idss.mit.edu   facebook.com/mitidss   twitter.com/mitidss