The Future of FinTech

Sanjiv R. Das
Santa Clara University*

June 3, 2019

Abstract

In this perspectives article, I describe the growing field of FinTech, and the different technologies that support it. FinTech will be a disintermediation force and I discuss the underlying technological drivers of various disruptive technologies. This framework comprises ten primary areas in FinTech and I briefly survey each one. The pitfalls of FinTech are also analyzed. Overall, the great strides made in computing technology, mathematics, statistics, psychology, econometrics, linguistics, cryptography, big data, and computer interfaces have combined to create an explosion of financial technologies.

JEL classification: C00; G1; G2; G3; G4;
Keywords: FinTech; Blockchains; Trading; Payments; Robots; AI; RegTech.

1 Introduction

Financial firms are rapidly using technology to transform their businesses. In this article, I survey ten areas in which FinTech is poised to deliver high value to firms, markets, and regulators. Much of this value is created through the use of machine learning technology, big data and cloud computing, and cryptographic methods.

FinTech refers to various financial technologies used to automate processes in the financial sector, from routine, manual tasks to non-routine, cognitive decision-making. Various areas of finance are subject to disruption, such as payment systems, contract checking, trading, risk management, quantitative asset management, lending, mobile banking, customer retention, and investment banking. Annual FinTech financing in 2018 was $112 billion, comprised of 2,196 deals, doubling over that of the previous year (2017: $51 billion).\footnote{https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/the-pulse-of-fintech-2018.pdf.} Figure 1 presents one representation of the FinTech landscape.

FinTech may be characterized by technological change in three broad areas of finance: (i) raising capital, (ii) allocating capital, and (iii) transferring capital. FinTech is disrupting

*I am grateful to Kresimir Kalafatic and Seoyoung Kim for discussion and comments, and for commenting on the basic taxonomy of areas in FinTech.

\footnote{https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/the-pulse-of-fintech-2018.pdf.}
Figure 1: The FinTech landscape. These “lumascapes” diagrams show a sampling of a business domain and firms within it. [Link](https://pbs.twimg.com/media/CYOHdfgW8AERjdO.png)
Definition: FinTech is any technology that eliminates or reduces the costs financial intermediation.

Definitions of FinTech vary depending on source, and the definition here is a general approach to capturing the main flavor of what is currently understood as FinTech. For example, the BIS defines FinTech in credit to “broadly include all credit activity facilitated by electronic (online) platforms that are not operated by commercial banks.” (See BIS Quarterly Review, September 2018, page 31.)

What is the long-run driver for FinTech as a disruptor? The main factor is that the cost of financial intermediation is very high and has historically always been so. In an interesting study, Philippon (2016) shows that the average cost of intermediation has held steady at around 2% of transaction amounts. Figure 2 shows that this cost has hovered around this level since 1880 till current times, for an astonishingly extended period of time of over a century. While one may speculate about the reasons for this high cost of financial intermediation, such as lack of competition on the supply side, or ignorance of consumers on the demand side, the fact remains that these rents accruing to large financial institutions are now ripe for the picking by smaller, agile, FinTech players. After all, technology is often a cheaper intermediary, and a driver of competition. Over the last two decades, employment in the financial sector has expanded from 5% to 6.5% of the workforce, as shown in Figure 3. FinTech may very likely reduce this number dramatically.

Figure 2: The cost of financial intermediation has held steady at 2% over time. See Philippon (2016).
FinTech applications range from simple automation to complex decision-making. Many rely on big data, and necessitate investments in cloud infrastructure and analytics. Successful FinTech applications display some common characteristics. First, it is valuable to develop models from a theoretical foundation before bringing in data. This helps both in preparing the data based on extant theory, and makes interpretation of the results facile as they are seen from the backdrop of a theoretical foundation. For example, automated lending models are based on a theoretical foundation of financial concepts such as leverage, and customer behavior, which suggests the econometric specification for the data. Second, sharp definition of the problem statement is critical – the question is primary, data is secondary. To consider the problem of lending again, there are many different firms with varied offerings, as may be seen in Figure 1. What differentiates them from each other is the specific pain point in the lending process that they seek to ameliorate. Any entrant in this congested area of FinTech must sharply define the niche problem they are solving so as to create a unique value proposition. Third, many FinTech offerings confront big data and computational bottlenecks. They face problems with data extraction, integration, curation, quality, and analytics. How well these are handled will determine the success of a FinTech venture. For a detailed description of these issues, see Alexander et al. (2017). Fourth, FinTech tends to be multidisciplinary, calling for teams that cut across many fields. A case in point is robo-advising, where an amalgamation of experience from portfolio management, behavioral finance, user interface design, risk management, legal aspects of consumer finance, etc., are all required. Another example is in cryptocurrencies, where specialized knowledge of cryptography, monetary systems, payment methods, distributed computing, etc., come into play. Applications that are based on natural language processing also entail many different sorts of expertise in linguistics, computer science, AI, and economics. Machine learning and
analytics figure prominently in many FinTech applications. Figure 4 is a sampling of areas in which large banks may benefit from FinTech implemented analytics.

The incredible effectiveness of machine learning (ML) in FinTech relies on advances in computational algorithms, such as matrix factorization, deep learning, classification methods, etc., and a proliferation of special purpose hardware (cloud platforms, GPUs, etc.) that exploit vast quantities of data. The confluence of these scientific developments has led to huge advantages for firms that have invested in ML technologies. Figure 5 shows that the performance of hedge funds that use machine learning outstrips that of funds that do not.

The scope of FinTech is huge. In a recent paper, Srinivasan (2016) states that as of 2015, Ernst & Young reports that there were 1,400 FinTech firms, with more than $33 billion in funding. In comparison, credit card fraud accounted for $31 billion in losses in a year. Therefore, even small improvements in just this area will generate substantial value compared to the investments being made in FinTech. Furthermore, modern hardware, cloud infrastructure, and software tools have made possible the rapid development of sophisticated FinTech platforms by very small teams, enabling entrants with minimal funding to compete in this space. With great power comes great responsibility. Lo (2016) points out that “...the unintended consequences of technology-leveraged finance include firesales, flash crashes, botched initial public offerings, cybersecurity breaches, catastrophic algorithmic trading errors, and a technological arms race that has created new winners, losers, and systemic risk in the financial ecosystem.” He suggests a strong focus on robust technology to manage this problem.

The distribution of FinTech is naturally uneven. Interestingly, one of the countries with high adoption rates for FinTech is Hong Kong, with the US coming second, followed by Singapore. The Chinese government is now one of the biggest spenders on deep learning technologies.

In the following sections, I will discuss ten areas where I expect FinTech to grow rapidly

- Monitoring corporate buzz.
- Analyzing data to detect, analyze, and understand the more profitable customers or products.
- Targeting new clients.
- Customer retention.
- Lending activity (automated)
- Market prediction and trading.
- Risk management.
- Automated financial analysts.
- Financial forensics to prevent rogue employees.
- Credit cards: optimizing use, marketing offers.
- Fraud detection.
- Detecting market manipulation.
- Social network analysis of clients.
- Measuring institutional risk from systemic risk.
Figure 5: Hedge fund performance is driven by machine learning. AI/Machine Learning Hedge Fund Index vs quants and traditional hedge funds. See: http://www.eurekahedge.com/Research/News/1614/Artificial-Intelligence-AI-Hedge-Fund-Index-Strategy-Profile.

and act as both an innovation and disruption. After these discussions, I will explore the use of machine learning, AI, and deep learning through some use case examples. These examples will also highlight the tremendous advances in machine learning technology that has made these disruptions possible. In the ensuing sections, we will examine some of the ideas and underpinnings of various uses of mathematics and computer science, married with economic models in the space of FinTech solutions.

2 Machine Learning, Artificial Intelligence and Deep Learning

As described in Culkin and Das (2017), artificial intelligence has come of age through implementing “deep learning” neural networks\(^2\), because of the confluence of three important factors. (i) The efficacy of mathematical analysis for calibrating neural nets; (ii) improvements in hardware and software that allow very large (deep) neural nets to be computed efficiently; and (iii) the availability of big data with which to train these models. Deep learning models have been proven to uncover subtle nonlinearities in data that are not discoverable using standard, more or less, linear econometric models. These models implement pattern

\(^2\)A neural net is a highly nonlinear fitting model made up of several simple nested nonlinear functions that take inputs and train them to generate a discrete or continuous output, through training on large data sets. Neural nets are used for example in natural language translation, and in driving autonomous vehicles. This is a very simplistic definition of neural nets, and nets with several tens of layers are known as “deep-learning” nets.
recognition to high levels of accuracy. By casting finance problems as pattern recognition problems, we are able to avail of higher predictability than before. We can also use deep learning to train pricing models from data on inputs and prices in the markets, bypassing the need for theoretical models, such as those used for option pricing. In the words of Peter Norvig, “All models are wrong, and increasingly you can succeed without them.” He makes the case that data is the new theory.\(^3\)

In order to explore this idea we revisit work from the 1990s. Hutchinson et al. (1994) explored using neural nets to learn the Black and Scholes (1973) option pricing formula, also developed by Merton (1973). In Culkin and Das (2017) we created data for the assessment of how a deep neural net would learn this equation, by randomly simulating a range of call option prices. The data was divided into two random sets, one for training, and the other for validation. Before passing the prices to the deep learning net, the data was normalized by exploiting a facet of the Black-Scholes call option function, i.e., that the pricing function is linear homogenous in \((S, K)\), i.e., \(C(S, K) = K \cdot C(S/K, 1)\). Therefore,

\[
C(S, K)/K = C(S/K, 1)
\]

The data was normalized by dividing both stock price \(S\) and call price \(C\) by strike price \(K\). This normalized data was applied to the deep learning net to fit the input variables \(S, K, T, q, r, \sigma\) (the feature set) to the output prices \(C\). The out-of-sample root mean-squared error (RMSE) is 0.0112, with an average error of \(\pm 1\%\) of the strike. The percentage pricing error (error divided by option price) is 0.0421, i.e., 4\% on average. The distribution of pricing errors is shown in Figure 6. A regression of the model values on the true values has very high \(R^2 = 0.9982\).

We also found that with the exception of very small option prices, where the percentage error tends to be magnified, moneyness is not correlated with pricing error.

This simple example supports Norvig’s assertion that machine learning has the potential to replace theoretically derived models with data-driven models. This is now being

\(^3\)https://www.wired.com/2008/06/pb-theory/.
implemented widely across the financial industry. Here are some examples. Bridgewater
Associates, the world’s largest hedge fund has a project to automate decision-making to
save time and eliminate human emotion volatility.\footnote{https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-ai-artificial-intelligence-management.} Goldman Sachs now has two out of 600
equity traders left in one part of their business. They found that four traders can be replaced
by one computer engineer.\footnote{http://www.zerohedge.com/news/2017-02-13/goldman-had-600-cash-equity-traders-2000-it-now-has-2.} It is estimated that by 2020 at least 5 percent of all economic
transactions will be handled by autonomous software.\footnote{http://www.gartner.com/smarterwithgartner/gartner-predicts-our-digital-future/.}

AI will process payment functions and learn from customer behaviors, through Intelligent
Payment Management (IPM). The potential savings are huge – AI will help consumers make
daily financial decisions and monitor spending. New Personal Financial Management apps
use contextual awareness, which measures spending habits and online footprints to create
personalized advice. Combining pooled financial data with end-user control to offer tailor-
made services is a classic AI solution that we should expect to see plenty of. Algorithms
will automatically mine customer data and undertake cross-selling of financial products.
Automation of cognitive tasks is now occurring rapidly. For example, J.P. Morgan has a
software called COIN that interprets commercial loan agreements and has resulted in a
saving of 360,000 hours of lawyer time.\footnote{https://futurism.com/an-ai-completed-360000-hours-of-finance-work-in-just-seconds/.}

There are several interesting applications underway. In the retail banking arena, Mizuho
Financial Group sent Pepper, its humanoid robot into its Tokyo branch to handle customer
inquiries.\footnote{https://www.mizuhobank.com/mizuho_fintech/news/pepper/index.html.} They partnered with IBM to enable Pepper to understand human emotions, and
build interaction into apps. Royal Bank of Scotland is trial testing Luvo AI, a customer
service assistant to interact with staff and customers.\footnote{https://www.businessinsider.com/royal-bank-of-scotland-launches-ai-chatbot-luvo-using-ibm-watson-2016-9?r=UK&IR=T&IR=T.} AXA (the insurer) has an app-basedot called Xtra, it engages in bespoke conversations with customers about healthy living.\footnote{https://www.the-digital-insurer.com/dia/xtra-by-axa-ai-driven-personal-wellness-coaching-app/.} And finally, AI is used in peer2peer lending.\footnote{http://www.nanalyze.com/2017/04/ai-fintech-startups-loans-new-credit/.} The role of chatbots in changing the interface
with banking customers is growing rapidly.\footnote{https://www.americanbanker.com/news/this-is-how-financial-services-chatbots-are-going-to-evolve.}

AI is also permeating the operations of hedge funds. BlackRock is replacing human stock
pickers with machine algorithms, using deep learning neural nets, as described earlier.\footnote{https://fortune.com/2017/03/30/blackrock-robots-layoffs-artificial-intelligence-ai-hedge-fund/.} Sentient Technologies is a hedge fund run entirely using AI.\footnote{https://en.wikipedia.org/wiki/Sentient_Technologies.} It is supposed to have a
secret algorithm with adaptive learning that uses thousands of machines. Numerai is a
hedge fund that makes trades by aggregating trading algorithms submitted by anonymous contributors, prizes are awarded in cryptocurrency called Numeraire, which reside on the Ethereum blockchain.\footnote{https://numer.ai/} Numerai open-sources transformed big data, which is not revealed in pure form because it scrambles the data using homomorphic encryption, a form of encryption where the data can be used for pattern recognition and analysis without being able to extract any original information from it. A skeptical viewpoint is that there is very little data about the track record of these hedge funds, as the business remains secretive. It is also argued that such funds will fail because investors will remain reluctant to turn over their money completely to a machine. Yet, this is being disproved by the little data we have. Figure 5 shows that funds using machine learning outperform others quite handily. And money is marching into AI driven funds, as Numerai Fund 1 LP raised over a million dollars in short order.\footnote{https://www.sec.gov/Archives/edgar/data/1667103/000166710316000002/xslFormDX01/primary_doc.xml.}

Does the fact that that hedge funds are successfully beating the market imply an unanticipated failure of market efficiency? In order to explore this proposition, Das et al. (2018) conducted a simple experiment using pattern recognition. Employing deep learning neural nets, they trained an algorithm on data from all stocks in the S&P500 index, and attempted to predict, using a look back number of days (30, 60, 90 days for example), the direction of movement (up or down) in the index the following day. Undertaking different experiments, the results, using daily data from 1963 show that it is possible to train an algorithm to guess the sign of the index return with marginally better than 50% accuracy. Out-of-sample accuracy levels (58.2%) are higher than average increase days in the markets (52.7%), suggesting that the model does improve portfolio choice to some degree. However, the accuracy levels are low enough to suggest that, even with a greatly expanded information set, markets are still efficient. The availability of data and open-source software packages such as TensorFlow make such tests of market efficiency easy to run. Indeed, this is probably one of the first tests of market efficiency, where the conditioning information set is as large as all stocks in the S&P 500 index.

3 Network Models: FinTech for Systemic Risk Modeling

In this section, we explore how graph theory is being applied to understanding the risks of the financial system, also known as “systemic” risk. Systemic risk has some universally accepted characteristics. It is a risk that has (i) a large impact, (ii) is widespread, and (iii) creates a ripple effect that endangers the viability of the economic system.

Systemic risk is an attribute of the economic system and not that of a single entity. Its measurement should have two important features: (a) Quantifiability: It must be measurable on an on-going basis. (b) Decomposability: Aggregate system-wide risk may be broken down into risk contributions from all financial entities in the system. Financial institutions (FIs)
that have large risk contributions to aggregate systemic risk may be deemed “systemically important.”

The Dodd-Frank Act of 2010, and Basel III regulations characterize a systemically risky FI as one that is (a) large, (b) complex, (c) interconnected, and (d) critical, i.e., provides hard to substitute services to the economy. Failure of such an institution is potentially disruptive to the financial system. The Dodd-Frank Act does not offer quantification specifics.

Recently, Das (2016) proposed a metric for systemic risk that has both the attributes of system risk, captures the features of SIFIs, and is consistent with the three universal characteristics of systemic risk. In order to accurately characterize systemic risk, graph theory is used, and a network of banks is constructed. To see an example of this network construction, see Burdick et al. (2011), where text analytics was used to extract interbank loan transactions from SEC filings, and construct a co-lending network of money flows between banks. Coupling this interconnectedness information with credit quality information from banks leads to a new measure of systemic risk, that has attractive properties, as formalized in Das et al. (2019). There is now a vast literature speaking to this problem, as in papers by Espinosa-Vega and Sole (2010); Billio et al. (2012); Chan-Lau et al. (2016), to cite just a few. It takes a careful combination of graph theory and economics, using large quantities of data to generate a single metric for measuring systemic risk. The requirement to do this in real time makes this a particularly interesting FinTech problem. For example, downloading and text mining all SEC filings relating to interbank loans in order to visualize the interbank lending network is an especially interesting problem, requiring terabytes of data combined with graph theory. See Figure 7 for the networks plotted on an annual basis for 2005, and for 2006-2009.

The mathematics for this model in its stochastic form are created by generalizing the Merton (1974) single-firm credit risk model. This extended model derived in Das et al. (2019) allows us to construct stochastic risk networks in a structural credit risk modeling framework. The model’s data are standard Merton model inputs for each firm:

- Equity price = \( s = \{s_1, s_2, ..., s_n\} \)
- Equity volatility = \( \sigma = \{\sigma_1, \sigma_2, ..., \sigma_n\} \)
- Number of shares = \( m = \{m_1, m_2, ..., m_n\} \)
- Risk free rate = \( r \)

The model variables are all derived from the Merton model):

- \( n \) = number of banks in the system
- \( a = n \)-vector with components \( a_i \) that represent the assets in bank \( i \) (derived from \( s, \sigma, m, r \)).
- \( \lambda = n \)-vector with components \( \lambda_i \) that represent the average yearly chance of bank \( i \) defaulting (from \( s, \sigma, r \)).
Figure 7: Lending networks between banks for five years, 2005-2009. Burdick et al. (2011).
• \( E = n \times n \) matrix with components \( E_{ij} \) that represent the probability that if bank \( j \) defaults, it will cause bank \( i \) to default (from \( s, \sigma, r \)). This matrix represents the risk connectedness of the banks.

The following calculations lead to a single metric for systemic risk that captures both, interconnectedness and credit quality of all banks. Define \( c \) to be an \( n \)-vector with components \( c_i \) that represent bank \( i \)'s credit risk. More specifically, we define

\[
c = a \odot \lambda,
\]

where \( \odot \) represents component multiplication (the Hadamard product); that is, \( c_i = a_i \lambda_i \).

The aggregate systemic risk created by the \( n \) banks in the system is

\[
R = \frac{\sqrt{c^\top \cdot E \cdot c}}{1^\top a},
\]

where \( 1 \) is an \( n \)-vector of ones, so the denominator \( 1^\top a = \sum_{i=1}^{n} a_i \) represents the total assets in the \( n \) banks. We note that \( r \) is linear homogenous in \( \lambda \).

The metric developed has four valuable properties, derived in Das et al. (2019):

1. All other things being equal, \( R \) should be minimized by dividing risk equally among the \( n \) financial institutions, and maximized by putting all the risk into one institution.
2. \( R \) should increase as the financial institutions become more interconnected.
3. If all the assets, \( a_i \), are multiplied by a common factor, \( \alpha > 0 \), it should have no effect on \( R \).
4. Substanceless partitioning of a bank into two banks should have no effect on \( R \).

Implementation of this simple system on a large-scale across all banks is an interesting challenge in the FinTech space.

The metric also leads to other useful metrics, which are easy to compute. The metric \( R \) is linear homogeneous in \( \lambda \). Let \( \alpha \) be any scalar constant. If we replace \( \lambda \) with \( \alpha \lambda \), it immediately follows that \( c \) is replaced by \( \alpha c \), and, by our equation for \( R \), we see that \( R \) is replaced by \( \alpha R \).

We can also calculate the sensitivity of \( R \) to changes in \( \lambda \): Differentiating our equation for \( R \) with respect to \( \lambda \)

\[
\frac{\partial R}{\partial \lambda} = \frac{1}{2} a \odot \left[ (E + E^\top)c \right] \cdot \frac{1}{1^\top a \sqrt{c^\top E c}}
\]

whose components represent the sensitivity of \( R \) to changes in each bank’s value of \( \lambda \). This is the basis of Risk Decomposition, equal to \( D = \left( \frac{\partial R}{\partial \lambda} \odot \lambda \right) \), a vector containing each bank’s contribution to \( R \).

A version of this model was implemented in India, with support from the Reserve Bank of India. Metrics may be produced daily. More details are presented in Das (2016), and for
Figure 8: Network of banks in India and Risk Decomposition, December 3, 2015. See Das (2016).

Illustration Figure 8 in this proposal shows the Indian network, systemic score $S$, and risk decomposition $D$. The metric is therefore easy to implement and offers a real-time systemic risk management tool for regulators.

Using publicly derived data for the United States banking system, Das et al. (2019) also calculated the top risky banks and top risky links. See Figure 9.

4 Personal and Consumer Finance

Households are consumers and allocators of capital. On the consumption side, they borrow money to finance consumption and investments in capital assets. They also earn incomes, generate savings, and allocate their wealth into various asset classes.

In recent times, households have faced several hurdles in the consumption-investment cycle. First, asset returns have become exceedingly low, and this has made retirement targets more elusive. For older investors, who rely on safe income streams after completing their work lives, decent risk free returns are no longer available, as risk-free interest rates have dropped to near zero. Further, risk premiums on speculative assets have seemingly shrunk, though there are differences of opinion on the expected equity risk premium. Second, medical advances have made longevity risk severe. There are heightened concerns about senior citizens outliving their savings. Third, there is substantial volatility risk when reaching for yield. Seeking higher returns in alternative asset classes comes with substantial risk, not just in the second moment of returns, but from the presence of negative skewness and excess kurtosis. Fourth, whereas risk-adjusted returns keep falling, high cost providers of financial services have retained their pricing schedules. As shown earlier in Figure 2, the average unit cost of financial intermediation is 2%. Fifth, inequality is also rising, and the financial system operates adversely against the poor, in a growing cycle of bias (Piketty (2014)). The share of income of the top decile of the population has grown from 35% in the 1940s to around 50% today. These factors scream out for a FinTech solution to raise risk-adjusted
Figure 9: Top systemically risky banks in US and top risky links (banking relationships). The figures show the ticker symbols of the financial institutions. The period covered for each row is the last six months leading up to the date shown in the leftmost column. (Dates are shown in YYYYMMDD format.) See Das et al. (2019).
returns so that investors may seek a less anxious retirement, with lower costs and inequality.

Various FinTech initiatives are addressing these issues. Robo-advising now enables investors to use technology to place their money in well-diversified asset pools at much lower cost, while offering solutions to the retirement problem. Firms such as Wealthfront (www.wealthfront.com) and Betterment (www.betterment.com) are cutting out high cost retirement providers using technology, while also educating naive, small investors. On the lending side, we have better credit models that enable firms to segment customers whose FICO scores otherwise excluded them from accessing financing. Not all subpar FICO scores come from low quality borrowers, and the ability to tease out the subset that might be of better quality offers FinTech lenders an opportunity to open up a new lending sector. Interesting approaches are explored in this space, for example social media interactions are used to identify good customers, as in Wei et al. (2016). Digital footprints may be used to assess individual default risk, as shown in Berg et al. (2018). Friendship networks may be exploited in peer-lending schemes, see Lin et al. (2013). Firms such as PayActiv17 are disintermediating the payday lending market, dropping the costs of borrowing by around 90%. Big data helps remove biases that often arise with small data, and these may be eliminated as discussed in Chowdhry et al. (2016). However, big data needs to be handled carefully, as biases within may end up permeating FinTech algorithms, see O’Neill (2016). Overall, consumer finance is ripe for new, refreshing improvements, driven by FinTech innovation.

5 Nowcasting

Forecasting has traditionally been a punctuated process with annual or at best quarterly forecasts. Macroeconomic forecasts are rarely able to use current data, as statistics for GDP, inflation, unemployment, etc., are usually available only with a lag. Moreover, these reported statistics are often updated and corrected over time, so they are being revised just as they are being used to make forecasts. This is a shifting target, comprised of very low frequency data.

The availability of new sources of almost streaming time series that are correlated with macroeconomic data has opened up the possibility that forecasting may be undertaken at high frequency in real time. This approach is known as “nowcasting” – and has been gaining popularity as a new area in FinTech.

The Federal Reserve Bank of Atlanta has implemented a nowcasting system known as GDPNow, see Higgins (2014). This model forecasts GDP growth by aggregating 13 subcomponents of the GDP with a chain-weighting methodology used by the U.S. Bureau of Economic Analysis. The model beats other forecasting approaches and delivers more timely forecasts. Figure 10 shows that the forecasts become more and more accurate as the date of the GDP number release nears. For a survey of nowcasting methods, see Banbura et al. (2014).

Nowcasting generally involves combining data of different frequencies to create as high-frequency a forecast as possible, see Giannone et al. (2008). In this sense, it is akin to ensemble modeling, where different data is combined to improve prediction accuracy. In

Aastveita et al. (2014), U.S. real-time data is used to produce combined density nowcasts of quarterly Gross Domestic Product (GDP) growth, based on a system of three commonly used model classes. Nowcasts are updated whenever new information is released. Combined density nowcasts always perform well relative to the model classes in terms of both logarithmic scores and calibration tests. The density combination approach performs more effectively than standard approaches.

Another facet of nowcasting is the creation of real-time indices that are non-traded. Of course, for traded indices, the process of trading itself generates the indicators of value in real time. But for non-traded indices, nowcasting using other traded variables and streaming data is feasible. For example, Chacko et al. (2016) provide a theoretical model in which it is possible to generate an illiquidity index from any market sector for which an ETF exists. In this model, liquidity in any market sector is modeled as set of options related to immediacy of trading. When immediacy is hard to come by in illiquid markets, the option to trade perforce has higher value. Converting this option value into basis points delivers the price of liquidity as a trading spread. The model is easy to implement and only requires two inputs at any point in time, the price of the ETF and the NAV of its underlying holdings, both of which are reported daily by exchange traded funds. The formula for the illiquidity index is as follows.

\[
BILLIQ = -10,000 \times \ln \left[ \frac{NAV}{NAV + |ETF - NAV|} \right]
\]  

(2)

where \(NAV\) is the net asset value of the exchange traded fund and \(ETF\) is its price. In liquid markets both \(ETF\) and \(NAV\) trade very close to each other, but in illiquid markets it is easier to trade the exchange-traded fund rather than the underlying components (i.e., the net asset value), so the two diverge in price, and the formula above picks up this difference.
and converts it into a basis points spread, shown in the formula above.

Inflation forecasting has leapfrogged traditional CPI indicators that have been criticized for two reasons. One, they are delayed and infrequent, and two, they are often based on outdated baskets of goods and therefore, do not represent current inflation levels. Using easily scraped prices from internet shopping sites, various efforts are underway to create real-time inflation indices. The Billion Prices project at MIT is a fascinating example of this new phenomenon.\footnote{http://www.thebillionpricesproject.com/} These examples suggest that the future of Nowcasting is a direct consequence of the growing availability of streaming data, and the technologies that enable its use. These platforms will eventually provide feature sets that will form the kernel of deep learning trading systems considered in Section 2.

6 Cybersecurity and Clandestine FinTech

People trust financial institutions to keep their wealth and information secure. (They also trust banks to operate responsibly, but that impression of trust erodes with every financial crisis.) The increasingly digital manifestation of banks has opened them up to hacking, and the nature of theft has changed dramatically. No longer is information security a matter of simply maintaining a good firewall. Rogue agents, in both digital and human form, reside deep inside bank systems, and Chief Information Security Officers (CISOs) at banks no longer assume that their firewalls are failsafe. PwC (2014) reports that 45% of financial sector companies experience crime versus an average of 34% in other industries.

Cyberthreats to financial institutions may be categorized into three major forms. First, threats come from state-organized actors. The news is replete with fact and conjecture about the role of state-sponsored hacking. Second, organized crime has discovered that hacking is a source of easy profits, at much less risk. When you can break and enter and loot assets or information without leaving the comfort of your home, it’s easy to see why criminals see this as a far better (often transnational) enterprise than old-style smash and grab.\footnote{Despite this, physical bank robberies remain numerous; the FBI reported 3,033 in 2018, down from 4,251 in 2016. See https://www.fbi.gov/investigate/violent-crime/bank-robbery} Cybercrime has resulted in stealing information as in the well-known scandal involving Target.\footnote{The cyber breach of Target resulted in the theft of information of 40 million customers. See: http://www.reuters.com/article/us-target-breach-idUSBRE9BH1GX20131219.} The “Bangladesh Bank cyber heist” in February 2016 entailed five fake money transfers through the SWIFT network, totaling $101 million, of which only $38 million was recovered.\footnote{https://en.wikipedia.org/wiki/Bangladesh_Bank_robbery.} The bank’s account at the New York Fed was hacked, and the thefts were traced to Sri Lanka and the Philippines, evidence that cybercrime in the financial sector operates easily on a global scale. The third actor in financial cybercrime is rogue bank employees. These are sleeper agents who gain access to servers as employees and wait for an opportune moment to operationalize their malicious intent. Internal bank security systems are being geared up to detect suspicious activity inside firewalls.
The recent dissemination of cybersecurity protocols by the SANS Institute has laid the groundwork for a formalized approach. The essential idea is that there are 20 Critical Security Controls (CSCs) that may be measured, and then aggregated into a cybersecurity score. This score may now be integrated into a bank’s risk management practices. Figure 11 shows the different CSCs and the overall framework. It also shows the various tools that are widely used to manage compliance with the CSCs.

Financial cybercrime has been detailed in popular books. Poulsen (2011) recounts how a hacker in the US, Max Butler, stole 1.8 million credit card accounts from two Russian crime syndicates, and was eventually caught by the FBI. This was lucky for him, as he may easily have fallen into the hands of the Russians, who might have meted out a much worse sentence than the 13 years Butler received. Menn (2010) tells the tale of a dyslexic young Californian who worked with the Feds to counteract hacking and financial ransom by the Russian mob and the mafia. It also describes how UK’s cybercrime units proved to be effective against Russian hackers. In the area of cybercrime in finance, fact reads like thriller fiction.

Finally, a discussion of financial cybercrime would be incomplete without mentioning the role of the “Dark Web”. A large amount of clandestine FinTech occurs on the dark web, that resides on the dark net, i.e., networks on the web that require special handshaking software to connect servers (closed versus open protocols). Most of our normal daily experience involves the complement of the dark net, i.e., the clear net. The “dark net” is a subnet of the internet where people connect servers to each other via private networks. The “deep web” is a subset of the internet that is not indexable by search engines. While different from the dark web, there is some overlap of dark and deep webs. The role of the dark web is to make agents who interact safe from detection, i.e., it assures anonymity and non-traceability of the source of activities on the web. As one might imagine, money laundering relies heavily on the dark net. For a detailed treatment of the dark net, see Bartlett (2014).

An example of the kind of FinTech evolving on the dark net is the creation of “tumblers” that are services to hide the source of transactions in bitcoin or other cryptocurrencies. By sending transactions through a tumbler, one can mask origination. Tumblers mix up transactions by acting as intermediaries and take a fee for doing so. These are algorithmic services, and may also be bundled automatically with some cryptocurrencies such as Cloakcoin (appropriately named!). Because the blockchain is a public, distributed, and immutable transaction ledger, it makes tracing a transaction possible, rather than otherwise. It is the fact that it is decentralized and does not need a single recording agency who controls transactions and can change rules at whim that make it attractive. Contrary to popular belief, it is because the blockchain does not guarantee anonymity, that the role of tumblers has become widespread.

Figure 11: SANS Institute Critical Security Controls (CSCs), shown in the top diagram. The lower graphic shows the various tools used for compliance with the CSCs. see: https://www.sans.org/media/critical-security-controls/critical-controls-poster-2016.pdf. See also https://www.sans.org/security-resources/posters/20-critical-security-controls/55/download.
7 Fraud Detection and Prevention

An interesting aspect of financial crime is that it mostly allows the perpetrators to be removed from the scene of the crime. Online fraud is an excellent example of this phenomenon. Much of credit card fraud now occurs on shopping sites, and the old business of making copies of charge cards with stolen credit card numbers is quickly turning irrelevant. Much of the fraud occurs in three ways: account takeovers, synthetic id use, and business email compromise, and the number of successful attempts has risen 34% from 2013 to 2016 (see Hasham et al. (2018)).

Online financial fraud management begins with record keeping, and it is essential that all online activity be logged so that traceback is possible. Strict authentication is important too. Many financial platforms are designed around multifactor authentication, as in the popular two-step verification implemented by many banks and asset management firms. End to end encryption is necessary to prevent man in the middle style attacks. Authentication is implemented in many ways, using tokens, passwords, pin codes, digital keys, biometrics, etc. Many digital wallets today embed multiple digital passwords, often three, and can be unlocked with at least two of these, which offers a high level of security.

However, authentication is a futile process if someone has stolen information through a data breach, because the authentication system will not detect a malicious use of validly accessed data. Activity analysis is needed and additional data might be brought to bear. For example, stolen credit card numbers offer an online thief full access to purchasing power so they need to be detected through recognizing that an attempted purchase does not follow the user’s standard behavior patterns. This is the realm in which machine learning has proven to be extremely effective, especially when alternative data is used. Solutions in this space come from using social media to detect anomalous behavior, noticing different devices, unexpected patterns in email use, unusual locations of use of the credit card from standard ones, etc. This is also known as the field of adaptive behavioral analytics, epitomized by firms such as Bionym, EyeVerify, and BioCatch.

As we have seen, the first layer in fraud prevention is authentication, which is related to cybersecurity. As is well known, this is proving to be a hard problem, as firewalls and accounts are breached in large numbers. A second layer of fraud detection lies in using varied data and conducting behavioral analysis, which is not easy to do, and requires extensive data curation. A third aspect of the problem lies in the difficulty of training machine learning algorithms to detect fraud. Algorithms that use big data in fraud detection are trained on highly imbalanced data sets. For example, detecting credit card fraud is fraught with

---

23 Despite the increase in cybercrime, bank robberies continue to occur at high rates. In 2016, there were more than 4,000 bank robberies, see: https://www.fbi.gov/file-repository/bank-crime-statistics-2016.pdf/view.

24 Sometimes these levels of security may be breached when hackers find holes in the code for the wallet application itself, as occurred in July 2017 in the case of the Parity wallet, see https://www.coindesk.com/30-million-ether-reported-stolen-parity-wallet-breach/. About $30 million of Ether (ETH) was stolen.


20
imbalanced categories. The rate of fraudulent transactions in this realm is \( \sim 0.10\% \), i.e., 1 in 1000 transactions is fraudulent, which means the data has 999 zeros for every 1 binary outcome. Precision and recall are very hard to control in such data settings, and false positives are rampant. These “anomaly detection” problems have been addressed through new techniques like topological data analysis (TDA), popularized by the firm Ayasdi.\(^{26}\) Other techniques have been developed by firms like Simility\(^{27}\). Various machine learning techniques are used from oversampling to boosting to deep learning to random forests. Overall, fraud detection is addressed with an interesting cocktail of authentication, behavioral modeling, device monitoring, and anomaly detection using machine learning on big data.

### 8 Payment and Funding Systems

Digital payments have changed the way in which we move money around, but they have also changed our concept of money. Starting with Paypal we now exchange money on several platforms, and bypass the banking system. Since then we have many such services, for example: Venmo, Apple Pay, Samsung Pay, Google Pay, Dwolla, etc.

This disintermediation takes many forms, but is leaving an electronic payment trail that is making black markets nervous. See for example the recent demonetization of the 500 and 1000 rupee notes in India. Transacting in cash was hampered, and millions were lost in concealed currency hoards, mostly used for property transactions. But it also gave rise to new transacting systems, with digital payments, such as the Paytm platform. In 2015, the Indian central bank gave Paytm authorization to open a payments bank, known as “Paytm Payments Bank Limited”. As of writing, Paytm now has over 250 million wallets in their transaction stream.

Another form of payments-driven disintermediation comes from disruption in the payday lending space from firms like PayActiv.\(^{28}\) This firm works with employers to issue advances against earned wages to low income employees, part of a package of offerings that reduces financial stress for people who live from paycheck to paycheck. It is estimated that of a US workforce of approximately 130 million people who receive a paycheck, about two-thirds live from one pay cycle to the next, and are unable to withstand temporary cashflow shocks. These people are victims of payday lenders who charge them annualized funding rates of 400-500%. Disintermediation by firms like PayActiv results in driving those borrowing costs down to one-tenth or less.

The lending space is also being disintermediated by peer-to-peer (p2p) markets. These electronic markets connect individual lenders and borrowers directly, and because they bypass large institutions who have high overheads, they have the opportunity to offer cheaper loans. Interest rates are set by the p2p platform through analysis of borrower data or through competition amongst lenders using reverse auctions. Prosper\(^{29}\) is one of the oldest such platforms, operating for more than a decade. They support small loans, mostly un-

\(^{26}\)https://www.ayasdi.com/.
\(^{27}\)https://simility.com/.
\(^{28}\)https://www.payactiv.com/.
\(^{29}\)https://www.prosper.com/.
nder $5,000, and make extensive use of borrower data. Most p2p lending platforms include borrower/lender anonymity, provide loan pricing support, offer lenders autonomy in picking their borrowers, and allow for both, secured and unsecured lending. Administrative services such as recording, credit checking, and online automation are standard. Targeted lending using big data and machine learning brings greater benefits to lenders (Crespo et al. (2018)).

FinTech also enables new venture financing, through crowdfunding. Top sites for this are GoFundMe30, Kickstarter31, indiegogo32, Kiva33, etc. Crowdfunding results in over $35 billion of fund-raising per year. The platforms usually take 5-10% of the money raised as their fee, so in terms of profit, this is at least a $1-2BN business. There may also be a processing fee added, so this parallels underwriting fees charged by investment banks. Payment system technologies represent one of the fastest growing areas of financial disintermediation, but their success will eventually depend on whether they are truly able to offer payment services at cost levels much below that of traditional players. These costs will also be adapted through AI-driven pricing, see Rizzi et al. (2018).

9 Automated and High-Frequency Trading

Whereas FinTech is new terminology, high-frequency trading (HFT) has been around for a long time, marking some of the first high-tech advances in finance. TradeWorx34 and Automated Trading Desk (ATD, bought by Citibank for $680M in 2007), were pioneers in the field. Algorithmic trading now accounts for 50% of executed trades in the equity markets, down from around 2/3 of stock trades in late 2000s. Estimates vary of course, and Aldridge and Krawciw (2017) estimate the share of market trading at closer to 40%. The profits from algorithmic trading are under competitive pressure and regulatory oversight.

There is a vast academic literature on HFT, and several debates surround this somewhat under-the-radar field. Some of the highlights are as follows. First, Since 2013, 2/3 of the top 30 cited papers on HFTs show positive market effects from high-frequency traders. Second, automated firms reduced trading costs, and contrary to popular opinion, improved market depth and stability. Third, new research is possible because various forms of high-frequency, streaming FinTech data have become available. Tick by tick data sets are now much more prevalent, and firms are working with academics much more closely than in the past. Fourth, there is now clear evidence that HFTs stabilize markets, Hendershott and Riordan (2013); HFTs improve market quality and reduce bid-ask spreads, Hasbrouck and Saar (2013); and HFTs reduce trading costs, Menkveld (2013). Fifth, trading in Dark Pools has been prevalent for quite some time, but has changed form many times, as risks and technologies evolve. For a theoretical analysis of this activity, see Buti et al. (2017).

This latest manifestation of the older field of “market microstructure” is here to stay. It is being reinvented through technology. It is also blamed for market disruptions such as

31 https://www.kickstarter.com/.
33 http://www.kiva.org/.
34 http://www.tradeworx.com/.
the “flash crash” of May 6, 2010, which lasted all of a half hour in afternoon trading, when the indexes crashed and rebounded, hitting several market triggers, and transferring vast amounts of profit/loss between trading accounts. Markets are also being transformed as a better understanding of these trading models on new trading platforms is creating (hopefully) marketplaces that are treating small players more fairly than in the past. A good example is the creation of the Investor’s Exchange (IEX), immortalized in the book “Flash Boys” by Lewis (2014).

Looking ahead the big changes anticipated in HFT are (i) an increase in regulation, (ii) a reduction in profits from competition, (iii) lower relevance on sheer speed of execution, (iv) a greater role for the use of myriad sources of information, (v) the entry of deep learning and AI, (vi) a reduced human role in favor of greater automation, and (vii) an emphasis on hardware over software.

10 Blockchains and Cryptocurrencies

On August 13, 2017, the price of cryptocurrency Bitcoin (BTC) surged past $4,000. This was a 20% increase over the previous week, after a plan to speed up trade execution was agreed upon. The new solution, denoted SegWit2x\(^35\), has been a bone of contention in the BTC community. However, trade execution under older protocols was slow, and this innovation is going to be a game-changer. Nevertheless, trading BTC is still fraught with risk. The daily volatility of BTC is around 5%, much higher than gold (1.2%), major currencies (0.5% to 1%), or tech stocks, 1-2%. And, a few months later, on November 29, 2017, the price of bitcoin broke the $10,000 price barrier, an astounding price rise for any currency. It ramped up to close to $20,000 in mid-December 2017, but then dropped drastically to below $7,000 by February 2018, and it has remained around that level ever since, after dropping even lower towards the end of 2018. Of course, there is no clear sense as to where its eventual trading range lies. We do not have a good model for pricing bitcoin using fundamentals.

Blockchains and cryptocurrencies are distinct things. A blockchain is a distributed ledger that has four properties. It has decentralized validation; it is immutable, i.e., the record may not be changed; it is secure, i.e., tamper proof; and trusted\(^36\), such that only valid transactions will enter the ledger and it will prevent double spending. It is also a ledger on a peer-to-peer network.

A cryptocurrency is a medium of exchange and a store of value, just like fiat currencies, though legally it is a security, and may be thought of as an asset class. Transactions in this currency are recorded on the blockchain. Bitcoin’s inventor is as yet unknown, though the original paper on which it is based is attributed to Nakamoto (2009). Bitcoins are traceable, but generally offer anonymity through masking methods. Whether bitcoin is a currency or a speculative asset has been debated in Yermack (2015).

Transactions in cryptocurrencies on the blockchain are secured by encryption methods,

\(^35\)https://segwit2x.github.io/

\(^36\)Sometimes, this is called the opposite, a “trustless” system, because there is no need for a trusted player in the middle of all other transacting entities.
and transactions are recorded in a decentralized set of nodes after approval. Approval of a transaction occurs when a “miner” processes a transaction block by adding a number (the “nonce”) to a variable length transaction, and then solves a hashing problem to generate a fixed-length (256 bits) hash with a set number of leading zeros (required to be 17 at the current time). This random guessing takes computational power (and electricity) and is rewarded with a set number of BTC, which is BTC 12.5 as of writing. This mining process produces a “proof of work” that validates the transaction block. With today’s technology, a block is solved every ten minutes.

Transaction volumes in BTC are currently quite low, about 1.5-2.0 transactions per second (tps), or roughly 300,000 transactions per day, versus PayPal at 125 tps, and Visa at 4000 tps. However, the technology is now widely accepted and way beyond mere proof of concept. It forms a core part of modern day FinTech.

A broader theme for the future is that the blockchain offers a networked platform on which decentralized and automatic contracting is made possible. This has been implemented most famously on the Ethereum platform, which is a programmable blockchain. Anyone can create a smart contract on the Ethereum chain, using the Ethereum Virtual Machine (EVM), and proof of work is rewarded with the cryptocurrency Ether (ETH), used for payments and fees on the platform. For example, one may establish a trading exchange on Ethereum, and contracts can be automatically settled on the platform once the program is so designed. It allows anonymous trading while still permitting a regulator to get an overall picture of risk concentration in the market, thereby allowing for systemic risk management. Another example of the use of the Ethereum platform is in real estate contracts. Several benefits exist such as title verification, settlement, shared equity in properties, and liquid trading of real estate assets.

Ethereum has also supported the growth of Decentralized Autonomous Organizations (DAOs), where group arrangements are explicitly contracted on the Ethereum blockchain as smart contracts. One in particular, known as the DAO (same name) was instituted to be a $150 million venture fund, invested in by thousands of people in a crowd-sourced success. It was however hacked in 2016 and several millions of Ether were stolen but later restored by canceling the stolen Ether on the chain, and resetting the blockchain, in direct violation of the principle of immutability. This resulted in a hard fork in the Ethereum blockchain. However, the hack was a result of weak security on the DAO side, not on that of the Ethereum blockchain. The Parity wallet was hacked on Ethereum for $31 million (in July 2017). Another $150 million was vulnerable, but white hat hackers stepped in and hacked those accounts out to save them, in a strange situation where the disease itself provided the remedy.37 These issues have not been addressed by regulators who are struggling to keep pace. Recently the SEC ruled that tokens issued by DAO were to be treated as securities and are now regulated by Federal law. A cynical viewpoint is that after regulation, cryptocurrencies will end up following the old rules of money in a new digital setting.

Another finance application that has grown rapidly are initial coin offerings (ICOs). These are analogous to a pre-product sale, where coins are issued on a blockchain as a store

of future value. These will also attract regulation by the SEC, and we have seen 46 ICOs in July 2017 alone. The SEC recently declared that ICOs are securities. Volumes in ICOs are doubling annually. In 2018, by June 30, 537 ICOs ($13.7 billion) were registered. In 2017, there were a total of 552 ICOs ($7.0 billion). The average size of an ICO has doubled from $12.8 million in 2017 to over $25.5 million in 2018.\footnote{https://cointelegraph.com/news/pwc-report-finds-that-2018-ico-volume-is-already-double-that-of-pre} 

Companies like Numerai\footnote{https://medium.com/numerai/an-ai-hedge-fund-goes-live-on-ethereum-a80470c6b681.} have created their own cryptocurrency (called a “Numeraire”), which investors in their crowd-sourced hedged fund may use for investments, withdrawals, transaction fees, and payments to developers of trading algorithms they may invest in. This cryptocurrency was made possible by hosting it on the Ethereum blockchain.

We are in the nascent stages of the blockchain revolution, but expect to see large-scale innovation in this space. It is likely that there will be vast changes in financial contracting, trading, risk management, and corporate finance, all implemented on blockchain infrastructure. For a comprehensive outlook on the disruptive (and beneficial) potential of blockchains, see Harvey (2014); Yermack (2017).

11 Text Analytics

Textual data greatly expands the universe of available data from the merely numerical. The starting point of any text analysis is the quantification of textual data, i.e., a mapping from words and documents to mathematical abstractions such as vectors, matrices, and tensors. The goal is to elucidate qualitative inferences and predictions after suitable mathematical transformations.

There are several areas in which text analytics is now being applied in finance. For a recent survey, see Das (2014). I will discuss some of these applications in this section. Sources of text fall into three categories: (i) blogs, forums, wikis; (ii) news; and (iii) content generated by firms. Early work tended to focus on sentiment extraction using postings on stock message boards, such as Yahoo!, Raging Bull, Motley Fool, and Silicon Investor. Sentiment is also extracted from news sections in the Wall Street Journal, and the Dow Jones and Reuters news services. These sources were at least at daily frequency. More recently, text streams have become available in real time, and Twitter has become a happy hunting ground for sentiment analysts. Bollen et al. (2011) state that they can predict the direction of the Dow Jones index with 87% accuracy using tweets. Since then there have been a slew of papers using tweets, each with mixed success.

Predicting market direction is one of the obvious first uses of text mining, and considerable energy has been devoted to this goal. Early research by Antweiler and Frank (2004) and Das and Chen (2007) constructed a bullishness index.

\[
B = \frac{n_B - n_S}{n_B + n_S}
\]

where \(n_B, n_S\) is the number of messages categorized as “buy” and “sell”, respectively. This measure is also often called “polarity”. This research shows that it is hard to predict market
direction using message board postings, but there is a somewhat weak prediction of volatility. Since then, much more attention has been directed towards using tweets instead. All these initial forays into sentiment analysis of stocks firmly placed textual analysis in finance onto firm footing.

Extensive use is made of corporate textual data for asset management. Data is extracted from the vast universe of regulatory corporate filings, such as 10-K and 10-Q forms. See papers by Loughran and McDonald (2011); Burdick et al. (2011); Bodnaruk et al. (2015); Jegadeesh and Wu (2013); and Loughran and McDonald (2014). These papers found that special words lists, known as lexicons, were indeed useful in determining the sentiment embedded in annual reports. For example, a count of risk related words in annual reports is a good predictor of poor performance in subsequent quarters. Papers by Calomiris and Mamaysky (2017) and Froot et al. (2017) show that textual information from large-scale media sources coupled with market information provides strong predictability of market direction.

Another interesting development emanates from quantifying “readability” of annual reports. It turns out that metrics for readability of text are useful in predicting corporate performance, and the less readable an annual report is, the worse a firm performs. Readability was originally quantified by researchers in linguistics, and measures such as the Gunning-Fog index (Gunning (1952)) have become popular as they are robust measures of readability. The formula for the Fog index is

$$0.4 \times \frac{\#Words}{\#Sentences} + 100 \cdot \frac{\#ComplexWords}{\#Words}$$

(4)

The formula returns a number that indicates the number of years of schooling needed to be able to comprehend the text. In other words, it also renders the grade level at which the text is written. The intuition for why low readability of annual reports correlates with poor performance is that when a firm has bad news to convey, it tends to obfuscate as much as possible. Subsequent research also found that longer annual report discussion (in number of characters) was also an indicator of poor subsequent performance, in fact, it was even better than poor readability. Once again, clearly, obfuscation comes more easily with long-winded writing than with short, pithy text. Finally, research shows that counting characters is unnecessary. Simply looking at the file size of the annual report on the SEC server is a good sorting characteristic. Bigger file sizes predict poor performance!

News analytics is also widely used to enhance asset management. This domain deals with the measurement of qualitative and quantitative attributes of unstructured news articles. Leinweber (2009) and Leinweber and Sisk (2011) offers extensive discussion on news analytics based trading strategies. Tetlock et al. (2008) show that the number of negative words forecasts negative earnings, and news stories that focus on fundamentals are more informative than other articles. For a comprehensive treatment of news analytics, see Mitra and Mitra (2011).

Prediction of corporate or banking failures has the potential to save vast sums of money. In recent work, Das et al. (2017) show how early warning signals may be extracted from text analyzing emails of senior management in a firm. Using Enron as a test case, they propose
a “zero-revelation” technology where a software program can analyze emails for their qualitative characteristics such as sentiment, quantitative characteristics like size and frequency, spatial aspects like networks, or aggregate focus through extracting topics. Because the program reads and provides an aggregate analysis, without revealing the contents of the emails, this is a non-invasive approach that may be used by corporate management or a regulator to predict financial malaise early, thereby adding value to a firm’s risk management process.

The main features of such an approach are as follows: (i) Financials are often delayed indicators of corporate quality. (ii) Internal discussion may be used as an early warning system for upcoming corporate malaise. (iii) Emails have the potential to predict such events. (iv) Software can analyze vast quantities of textual data not amenable to human processing. (v) Corporate senior management may also use these analyses to better predict and manage impending crisis for their firms. (vi) And most important, the approach requires zero revelation of emails.

The empirical analysis of Enron’s emails primarily covers two years, 2000 and 2001. Enron failed in Q4 2001, so this is the period before and during the crisis the firm faced. Some clear results emanate. (i) As Enron’s condition worsens, emails are shorter in length. (ii) Likewise, Enron’s sentiment extracted from emails drops a few months before the crisis, and continues into it. (iii) It is interesting that the length of senior management emails is a better predictor of demise than is sentiment, though both are useful. (iv) It is also interesting that plotting the frequency of usage of words over time also tells a story. The word “losses” sees more frequent usage than average in the crisis period, and the inverse is true for the word “profits”. For example, see the word “credit” plotted against sentiment for the two years in the data set, these track very closely, as shown in Figure 12. (v) A topic analysis also shows that negative topics become more prevalent as we approach the crisis versus positive topics. (vi) We also see that the email network for 2000 looks very different from that in 2001. See the paper for graphs and details.
12 Concluding Comments: the Good and the Bad

In this final section, I consider two important issues that arise around FinTech, labor market changes and implementation pitfalls along the FinTech path.

12.1 The Finance Labor Market

FinTech is a potential disruptor of the financial services labor market, which accounts for 6-7% of US employment. AI and machine learning will erode many of these jobs. However, the debate on whether we will see the process evolve through IA (intelligence augmentation) versus AI (artificial intelligence) is still open. The former will lead to less job loss. It may be easier to replace high-end trading jobs than low-end customer service roles! Firms are recording traders’ keystrokes and screens and using the data to train algorithms to mimic traders. These artificial agents are able to improve through techniques such as reinforcement learning. These algorithms can analyze a certain type of data far more deeply than a human can, but then they are also poorer at considering a breadth of data, though they may eventually be engineered to do so. In contrast, chatbots to replace customer service agents have a long way to go be good at replacing humans. Middle-level jobs such as paralegal work in securities contract checking is being eliminated with algorithms that have lower error rates. This corresponds to Autor (2015)’s hypothesis that while very low-end and very high-end jobs will remain, several middle-level jobs will be lost, such as financial analysts, routine trading functions, paralegals, HR roles, loan officers, financial advisors, etc.

Ray Kurzweil has predicted the supremacy of machines, the “Singularity”, which he suggested will occur in 2045. Norbert Wiener famously said 41 - “We can be humble and live a good life with the aid of machines, or we can be arrogant and die.” Setting these ominous fears aside, we may dig deeper into the kinds of jobs that seem to be getting winnowed away. A cynical hypothesis has been put forward by Graeber (2018) that there are legions of “bullshit” jobs that are finally being eliminated such as corporate lawyers, financial advisors, etc. These are jobs that would make no difference were they to disappear and in fact people doing these jobs know this and bear a psychological stigma as a consequence. 42 I have a simple hypothesis for which jobs will be lost: if the job generates data, then an AI agent can be trained on that data to replace the human. This is what we see happening with trading jobs, paralegal work, etc.

Hopefully, there will be jobs where FinTech will keep the “human-in-the-loop” (HITL). These are jobs where consideration of non-standard cases is required or where a human is needed in the loop to avoid legal liability, e.g., risk managers. Such jobs may escape “technological unemployment.”

41https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html
42Graeber first espoused these ideas in an essay titled “On the Phenomenon of Bullshit Jobs”, https://strikemag.org/bullshit-jobs/. The arrival of AI coincides now with the book on the same idea.
12.2 Pitfalls for FinTech

While FinTech heralds the new age of finance, it is by no means widely prevalent as yet, and is certainly not the panacea for all open issues. While it is both an improving force and a disruptive one, it needs to be implemented thoughtfully. Here, I briefly survey seven pitfalls to avoid when implementing FinTech.

First, beware of the garbage-in, garbage-out (GIGO) problem. FinTech is consuming more and more data, but it needs to be good data. Without “data curation”, financial analyses will produce poor quality output. This issue is itself being addressed with new technologies, where machine learning is being used to clean data, generating better quality inputs as well as reducing the data engineering cycle drastically. This is discussed extensively in Alexander et al. (2017) and technically, companies such as Tamr\(^{43}\) and Paxata\(^{44}\) offer excellent solutions.

Second, we have information overload (IO). This comprises collecting too much data and not using it properly. An antidote for this is good theoretical modeling, which suggests exactly what data should be provided to the algorithm. This approach transcribes the scope of the data needed. It may also save money as a firm may be able to move from distributed computing to using a single, large, and fast machine. Testing trading models and proper backtesting will help in improving signal to noise ratios. Tools for this are becoming widely available and business such as Quantopian\(^{45}\) and Numerai\(^{46}\) are making these widely available so that trading ideas may be crowdsourced, another effective antidote to processing inhuman quantities of market information.

Third, Big is Not Better (BiNB). More data does not mean better results, quality matters. However, AI and Deep Learning are making it possible to let machines make sense of large-scale data. Simple applications in the consumer finance space are seeing positive results from the use of deep learning on large data, especially in the area of anomaly detection, as we have seen with firms like Ayasdi and Simility.

Fourth, financial firms must not confuse correlation with causality. In the search for predictive analytics, the use of big data is turning up better predictive models, but these are untraced to the cause of predictability, and are largely a-theoretic. Given this, firms need to revisit their models on much tighter review cycles in order to ensure that they are still viable for trading. This issue becomes especially acute when considering streaming data in HFT.

Fifth, the infrastructure for FinTech is expensive, and once a firm dips its toes in the water, it should be ready to go all in, else it becomes impossible to get good results. HFT for example, requires an increasingly expensive investment in hardware and communications technology. Deep learning platforms require outlays for expensive chips such as arrays of GPUs. Blockchain infrastructure requires electricity-greedy special purpose hardware. And so on. It’s an arms race!

\(^{43}\)https://www.tamr.com/.
\(^{44}\)https://www.paxata.com/.
\(^{45}\)https://www.quantopian.com/.
\(^{46}\)https://numer.ai/.
Sixth, trust is paramount. FinTech offerings tend to hand off previously human functions to technology, for example, replacing tellers with ATMs and online banking. Cryptocurrencies are completely based on trust, where trust is transferred from a centralized and regulated repository to trust in technology and decentralization. Without trusted algorithms and data, FinTech will fail. Any firm developing a new FinTech business must consider how they will implement trust through technology.

Seventh, while technology can be used to improve experience, more often than not, the opposite is the result. Excessive misdirected automation can lead to a terrible customer experience. A good case in point is the widespread use of chatbots for customer service interactions, where the quality range is huge. Robo-advising is another area in which careful design of the interface becomes necessary. Firms are using Design Thinking to ensure that customers are well-served in FinTech driven businesses. Finance houses need to become much more consumer-centric and use design thinking as actively as tech companies like Apple do.

I end with the following prediction: Finance companies will eventually become quasi-technology firms.

References


