Big Data and Big Money
The Role of Data in the Financial Sector

Jennifer Q. Trelewicz, Deutsche Bank Technology Centre

When we think of industry sectors driven by high tech, for some people, perhaps, banking is not the first that comes to mind. However, when we consider the 3Vs of big data—volume, velocity, and variety—it is hard to think of many sectors whose requirements fit so nicely into the guidelines. For example, in April 2016 alone, the foreign exchange (Forex) markets averaged US$5.1 trillion per day. The Forex markets provide real-time exchange rates between currencies across the world, facilitating global business and settlements.

In this article, I discuss the relevance of big data approaches to the financial sector, outlining challenges to adoption as well as future opportunities for technology development. Because of its transaction and money volumes, I focus on corporate banking (financial markets, corporate credit, trading, and so on), although many application areas are also relevant to consumer finance.

Big Data in the Financial Sector
Let us first examine the relevance of the 3Vs to finance:

- Volume is considered to reach big data levels at many Tbytes or even Pbytes of data. The financial industry produces a huge volume of quotes, market data, and historical trade data. The New York Stock Exchange (NYSE) alone writes more than a Tbyte per day.
- Velocity suits big data when the speed of data storage or processing is on the order of 10^5 transactions per second or more. Generating data at this speed is no challenge for the financial markets. Moreover, the faster systems can process trade data, the faster they can manage trading.
- Variety implies that big data algorithms do well with various formats and data sources. In corporate banking, institutions work with reference data (about legal entities, for example), trade and market data, requests from clients (by electronic and voice means), and many other sources.

What makes the financial sector even more interesting from a big data standpoint is the constant stream of new regulations and reporting standards that bring new data sources and more complex metrics into financial systems. This makes the sector a very interesting place for the data scientist.

The Forex markets, as mentioned earlier, trade 24 hours per day, from morning in Sydney to evening in New York, except for a small window during the weekend. Additionally, algorithmic trading has been used in the financial markets for a long time in one form or another. The NYSE introduced its Designated Order Turnaround (DOT) system in the early 1970s for routing orders to trading desks, where the orders were executed manually. Now, algorithmic trading systems break very large orders into smaller pieces that are executed automatically based on time, price, and volume, optimized for market parameters.

On a continuous basis, the processing of large volumes of data is used for reporting purposes in financial institutions:

- Banking and financial market regulations more and more often require the calculation of various complex metrics, such as XVA (valuation adjustments of derivative instruments, based on counterparty credit risk, cost of funding, margin, and so on). Such metrics are used, for example, to
set the minimal capital reserves of a bank, which directly influences the bank’s profitability.

- Time-sequenced transactional data is analyzed to model market and customer behavior. For example, mapping trade volume with time could help to predict the probability of a default on credit, saving a bank lost resources on a loan.

Some large financial institutions have been slow to adopt big data approaches, but market research from PwC has clarified some of the organizational and cultural inhibitors to adoption in these institutions, many of which are relevant in other industrial sectors as well. First, some financial-sector managers feel that big data algorithms solve technical problems, but not business problems. However, when the data is generated by the business, and the results are used by the business, it is clear that technology is supporting the business. Some do not understand how to gain value from their data streams, while some feel that big data approaches improve technical efficiency but do little for the bottom line. However, the deep analysis that big data approaches can provide can directly support business growth and improved effectiveness. The financial sector has not traditionally been a destination for data scientists, so some institutions have met with difficulty in finding and attracting the needed skills to their organizations. Finally, even when the will for transformation is in place, it might not be clear how and where to start transforming an enterprise to utilize big data approaches.

However, banking is among the top industry sectors investing in big data analytics, according to a recent study from IDC, as Figure 1 shows. Moreover, financial technology, or FinTech, companies are developing solutions and products for a range of banking needs for asset and wealth management; Figure 2 breaks out the proportion of FinTech companies per area according to PwC. Following this trend, there is a growing body of research and algorithm development around other uses of financial data for increasing business effectiveness. We next examine a few of these.

**Market Trading Patterns**

Adaptive models of market trading patterns can provide input to investment strategies for buying and selling certain types of assets. This section explores one example.

Ex-dividend Day (also called X-Day) is the first trading day when the seller of a depository receipt (DR) has the right to receive the last dividend payout. Before X-Day, the buyer of the DR would receive the dividend payout. American DRs (ADRs) are financial instruments traded in the US market by non-American companies. As such, there is a dual tax burden on the dividends: the US taxes and the tax withholding in the country that issued the ADR. As a result, investors are motivated to sell ADRs before X-Day and to buy after X-Day. Naturally, tax policy has a strong influence on the stability of the ADR market.

In recent work, Bi-Huei Tsai examined the ADR market to understand market trading volumes. Analysis of such markets by such a class of algorithms could be used to suggest optimal trading times based on recent market volumes. The author analyzes excess ratios of ADR volume (the daily trading volume minus the “normal” daily trading volume) during the ex-dividend period (X-Day +/- 10 days), positively correlated to dividend taxes, providing a model of tax policy’s
IT TRENDS

influence on the ADR market. Both traders and government tax authorities could use such models to create strategy.

Real-Time Credit Ratings
An application that isn’t specific to financial markets but has relevance to banking for consumers and small and medium businesses is processing data to produce credit scores for applicants in real time. For example, FinTech companies such as Klarna, Lenddo, and Credit Karma provide services related to online credit scoring and verification. In recent work, Ying Wang, Siming Li, and Zhangxi Lin examine the potential for real-time credit scoring for e-commerce.3

Anyone who has applied for a significant amount of credit will be familiar with the timeline of the process. Traditionally, banks collect information about the applicant from both the application form and other sources. Specialists analyze this information to create a credit proposal for the client, which includes the interest rate and terms of repayment. There might be some negotiation between the applicant and the bank, including tradeoffs between various loan parameters for better overall terms. After the credit contract is signed, the client can engage in his or her financial activity and further pay off the loan.

Not only do data collection and terms negotiation take time, but two factors complicate the picture. First, many of the central data stores used for credit ratings are updated only monthly, so the client’s recent financial problems might not be known to the bank from such sources. Moreover, nonfinancial factors could play a role in the risk of default. The authors examined several such factors relating to a large e-commerce platform: frequency of login to the platform, provision of additional contact data (including mobile telephone number), volume of transactions in the last month, number of successful transactions overall, time as a client of the platform, the client’s business sector, and so on. The authors use linear regression analysis over groups of these parameters on historical client data to derive a correlation between the probability of default and the various parameters. The result is a model that could facilitate real-time credit ratings for the e-commerce platform, based on online behavior. Such parameters are not even present in traditional, central credit-rating databases.

Banking Becomes More High-Tech
The two examples described are by no means a limit to the applications of big data algorithms in the financial sector. Although no one can predict the future markets a hundred percent, deep analysis of historical data and current market parameters provide sophisticated, adaptive models of tendencies and behaviors in the markets. In turn, such models facilitate better-informed and faster decisions by traders (including trading systems), financial institutions, and other players.

The opportunity for IT Pro’s readers is to develop new technologies and solutions for a fast-growing sector. Funding for FinTech more than doubled between 2014 and 2015,6 indicating both opportunity and a need for such products and solutions. The broad categories shown in Figure 2 leave a great deal of room for innovation in product, process, and customer experience. It might not be long before banking comes to mind first when we think of industry sectors driven by high tech.

References

Jennifer Q. Trelewicz is the technical risk officer and CTO for Enterprise Risk Technology at Deutsche Bank Technology Centre. Her research interests include big data, high-performance numerical algorithms, and innovation management. Trelewicz has worked in industry-leading technology companies such as IBM, Microsoft, and Google, has numerous granted patents in various countries and publications in international journals and refereed conferences, and volunteers actively with IEEE. She has a PhD in signal processing from Arizona State University, and is a life member of the international honor societies Tau Beta Pi and Phi Kappa Phi. Contact her at jentre@ieee.org.