Modelling the Way
State of the industry overview – learning information from data

Swathes of regulatory change have had a lasting and profound impact on the financial services sector. Regulatory frameworks such as Basel, EMIR, MiFID, and Solvency II have demanded new approaches to capturing, reporting, and deriving value and meaning from data that powers financial institutions.

However, regulation is not the only driver of this change. Massive IT disruption – a blend of new “big data” technologies intersecting with modelling and analytics, simple and sophisticated – is providing new mechanisms to drive business and manage risk. This is facilitating rapid technological change in applications such as innovative portfolio management, more rigorous stress testing of long-term insurance or pension projections, real-time risk monitoring, and fraud detection.

Cultures and internal structures within financial services institutions are also transformative. Large firms with complex proprietary infrastructures and siloed chains of command that were established long before the global financial crisis are having to adapt to new agile, transparent, scalable infrastructures. They face competition from dynamic small firms and challenger shadow-like institutions, often unencumbered by complicated legacy processes and systems.

Amidst all of this are the quants, and developers collaborating with the business to accommodate stringent regulators, serve customers with more choice in fast-moving markets, and ensure the efficacy of increasingly liability-concerned management teams.

Against this backdrop, MathWorks has sought to obtain insights from the financial services industry to better understand the dynamics between data and models and those who:

- Manage data
- Build models and analytics
- Implement models and analytics within data environments and the enterprise
- Use the models and their output (often executives and customers)

Through this lens, we look behind the excitement to examine adoption of machine learning methods, alongside traditional statistics, as tools to make sense of big data. We also identify less fashionable, yet highly impactful trends, such as the software carpentry required to take analytics to the business more quickly and with appropriate quality processes.

The survey of 78 professionals was conducted at the MATLAB Computational Finance Conference in London in June 2014. Respondents span the buy side and sell side, in addition to consultants and representatives from central banks.

We hope you will find it an informative and interesting read.

Steve Wilcockson
Industry Manager – Financial Services, MathWorks
www.mathworks.co.uk/finance-modelling
Key findings at a glance:

- 37% of survey participants consider risk management to be the main driver for their model development.
- 68% identify a “lack of agility to respond to market changes” as the biggest opportunity cost to the business of slow model development.
- Over 60% of respondents report that their institution encourages collaboration between model developers and front-office and middle-office business practitioners.
- 31% report that their institution has implemented a project to react to a specific big data challenge.
- Respondents report that a single software environment facilitates collaboration between model developers and business users (45%), reduces cost of model development (51%) and reduces risk (47%).
- Respondents identify “data quality” (36%) and “creating effective models” (36%) as the greatest challenges associated with big data in financial services.
- More than half of the professionals have concerns about machine learning. For 45% of them, these concerns centre on over-fitting and fear of the implications of a “black box” approach, and 40% want to learn more about machine learning.

Drivers of model development: a balancing act between protection and wealth creation

The industry has arguably been “on the back foot” managing risk since the financial crisis in 2008, with firms working to rebuild their risk oversight and governance frameworks.

Given this renewed focus on risk culture, it is not surprising that 37% of survey respondents identified risk as the main driver for model development within their organisations. However, 27% of respondents are motivated by increased profit potential by rapid response to market movements, more than those noting regulatory requirements specifically (15%). Put simply, greater industry regulation is not a key driver in model production or model use. This data is consistent with results from the 2012 report, Modelling and Analysis: The Financial Services Perspective (www.mathworks.co.uk/financereport).
However, the research highlights an interesting divide between buy-side and sell-side institutions. 38% of those on the buy side and 54% of those on the sell side noted that risk management is the main driver. On the other hand, 45% of buy-side respondents pointed to increasing profit as being key, compared to only 18% of sell-side respondents.

**Collaboration – from concept to tool**

When asked what the opportunity cost was for slow model development, 68% of respondents identify a lack of agility to respond to market changes as their primary concern, followed by ineffective risk management (53%).

This concern around the impact of agility relative to risk is consistent with data from 2012, when 82% identified lack of agility as the biggest opportunity cost. Ineffective risk management was second at 74%.

Another transformative element of financial model development is the throughflow and collaboration between modellers, developers, and business owners such as front office revenue generators — traders, sales teams, portfolio advisers and managers, and middle-office risk managers.
The survey indicates progress, with over 60% of respondents reporting that their institution encourages collaboration between modellers, developers, and front-office and middle-office business practitioners.

The increased collaboration we see reflects two key factors: enabling technologies such as data analytics, and the cultural importance of effective communication between modellers, developers, and users. Model users and decision makers don’t want to be “blinded by science” but they want to interact with the technical teams — whether quants or developers. The survey responses highlight the role of technology, with 45% reporting that using a single software environment throughout the development process from research to production facilitates collaboration, reduces cost of model development (51%), and reduces risk (47%).
Responses support the principle that developing and testing an algorithm or model on the same platform where it will eventually be used is preferable to a disconnected “one language for this and another for that” environment. A single platform can enable modellers and developers to work together and reduce the risk of recoding errors in the final algorithm and providing versatility in deployment. The survey results from 2012 highlight the ambition of both buy-side and sell-side institutions to streamline this process. 75% of buy-side respondents wanted to reduce the time it takes to integrate models into business processes from months to days, while 40% of sell-side respondents wanted to reduce this time to mere hours.

Smaller and medium-size firms can find themselves at an advantage as they maintain simpler managed infrastructures. That said, agile and scrum\(^1\) techniques are accessible to all institutions large and small. Scrum teams can follow a project “holistically” from beginning to end rather than using the “throw software over the wall” approach that is more common in larger institutions with legacy processes. In this scenario, the team on one side of the wall might use a particular tool or environment, while those on the other side might use something different. This fractured approach often requires recoding and introduces model risk.

---

\(^1\) Scrum is an iterative and incremental agile software development framework for managing product development. It defines “a flexible, holistic product development strategy where a development team works as a unit to reach a common goal” and challenges assumptions of the “traditional, sequential approach” to product development. It enables teams to self-organize by encouraging physical colocation or close online collaboration of all team members, as well as daily face-to-face communication among all team members and disciplines in the project.
**Machine learning, with big data**

More demands are being placed on modellers to provide information in a useful, relevant, and robust way. In addition, the model users, the front office, and their risk manager overseers need assurance that the models they use are accurate and suitably applied, can be updated or recalibrated at short notice, and can scale up over the longer term. Financial analytics software should be easy to use and provide useful, actionable information, ideally through visualization.

Modellers also need to be content that their model is trustworthy, robust, and well applied. How much has it been tested? And against what? For example, the model is best tested on new, perhaps live “out-of-sample” data as well as “in-sample” historical test data.

Machine learning techniques are predicated on the notion of learning from data. As such, machine learning algorithms can be used to classify data and determine trends in larger datasets, exposing new insights. There are many machine learning methods, each with benefits and risks, that should be adapted appropriately to the task at hand. “Bagged decision trees” suit different applications than “neural networks,” “support vector machines,” or other machine learning methods.

**What is machine learning?**

Machine learning algorithms use computational methods to “learn” information directly from data without assuming a predetermined equation as a model. They can adaptively improve their performance as you increase the number of samples available for learning. Machine learning can be used for building predictive models from observed data and discovering useful patterns for computational finance, such as credit scoring and algorithmic trading.

While machine learning has been successfully applied in other sectors such as the robotics, medical, and defence industries, financial services firms are exploring capabilities in trading, risk, and portfolio management. Survey responses reveal that machine learning has not yet reached mainstream adoption within financial services, with only 12% reporting that they already use machine learning techniques. The biggest concerns regarding the risks of machine learning are that it is too much of a black box approach (22%) and over-fitting (23%).

![Figure 6. What is your perception of the opportunities and risks of machine learning? (multiple responses allowed; more positive responses left and negative responses right)](image-url)
Interestingly, however, the survey reveals significant interest in learning more: a need to understand before deploying the technologies. 40% report that they want to understand more about machine learning methods. This suggests that while practitioners are aware of the benefits of machine learning, they need to better understand the risks and the nuances of implementation.

Alongside machine learning challenges and opportunities sit big data technologies, such as distributed NoSQL databases, MapReduce protocols, and increasing amounts of simulated data, perhaps generated via GPU or cloud-based parallel processing. Scaling analytics flexibly to larger datasets hosted on server and cloud architectures offers migration challenges, but machine learning provides one approach to automating the process of data discovery and interpretation.

In 2012, 68% of respondents indicated that data quality is the biggest challenge associated with the data deluge, followed by creating effective models (57%) and data variety (38%). This year’s survey also highlights data quality as a challenge (35%), although it is now equal with creating effective models (35%). Only 8% pointed to data variety as a key challenge. Similarly to 2012, the survey reveals that datasets in general are not as large as externally perceived – predominantly measured in gigabytes (46%). That said, 31% report in 2014 that their institution has implemented a project to react to a specific big data challenge.

Critical mass implementation of machine learning, in big data or otherwise, is still some time away in financial services. Education on machine learning, with its benefits and risks, will lead to more effective applications of relevant methods to business-level tasks and successful implementation and scaling in big data infrastructures.
Conclusion

A key driver of change in the financial services IT landscape is increased awareness of risk, driven in part by regulatory and economic concerns after the global financial crisis. In addition, change is increasingly overlaid by “big data” and data analytics technologies.

The survey highlights the need for financial institutions to be quicker, smarter, and more collaborative, as well as maintain a heightened awareness of risk and quality. Despite this, bureaucracy continues to hold many back, particularly on the sell side.

The survey offers interesting insights into how firms are reorganising within this exciting landscape, applying at best some well-constructed, versatile, and implanted models to reduce risk and drive competitive, unique, and better business. Risk managers no longer just report key risk statistics—they suggest alternatives, test against real-world scenarios, and deploy results—often intraday. Risk managers need to be assured, though, that their systems are transparent, quality-assured, and scalable. At worst, layers of spaghetti code, inconsistent, and poorly-maintained applications increase risks—perhaps more so than conforming to the right statistics and stress tests mandated by regulators.

Machine learning is not yet mainstream across financial services – only 12% of survey respondents noted applying machine learning technologies. This is, however, an age of significant evolution, maybe revolution, in the application of machine learning for credit scoring, fraud detection, trading strategy management, and other critical applications. The industry wants to learn more before it implements machine learning to its full potential. However, traditional statistics and equation-based modelling are not going away. To be successful, effective use of big data requires the harnessing of all aspects of statistics, machine learning, and models in general.

Agility, time-to-market, and quality are paramount—arguably the greatest weapons banks have in their bid to both hedge against risk and drive profit. It is only when the right algorithmic models are tested and implemented in the right way that they will deliver true value to a business.

Afterword – the flow of ideas

The financial services industry is not the only sector to have gone through systemic and regulatory change. Other disrupted industries, such as the automotive industry in the 1990s and early 2000s, faced pressures on margins, technology, and resources. The auto industry took the opportunity to transform itself, implementing Model-Based Design with automated development. When research engineers develop new cars, they build models but they often verify, implement, and validate automatically directly onto the vehicle. This results in reduced costs, improved quality, and innovation.

Financial services still has some way to catch up with the automotive industry. Analysing the survey responses, we sense some frustration. Developers and practitioners want to use a single environment to speed up time-to-market and reduce risk but siloes still predominate, particularly among the global, systemically important banks.

There are too many examples of software projects that have caused great financial and reputational damage within the financial services sector. In reaction, MATLAB users report that adopting a quality-assured well-integrated platform can encourage model and analytics knowledge sharing.

http://www.eusprig.org/horror-stories.htm
http://www.nextnewdeal.net/rotybomb/researchers-finally-replicated-reinhart-rogoff-and-there-are-serious-problems
http://www.ft.com/cms/s/0/617d0126-0bd7-11e3-b060-00144feabdc0.html
across multiple domains. This can reduce code siloes, facilitate model reuse and scaling, and speed up development, thereby enhancing analytics innovation and quality, and assuring software quality.

Machine learning is one cog in this fast-paced industry, which can and will deliver benefits to users and businesses. As with any innovation, the underlying framework of model development and implementation must be scrutinised if this exciting technology is to deliver maximum value.

**About MathWorks in Financial Services**

MathWorks is the leading developer of mathematical computing software. MATLAB, the language of technical computing, is a programming environment for algorithm development, data analysis, visualisation, and numeric computation. MathWorks customers include the top 15 asset management companies; 9 of the top 10 U.S. commercial banks; 12 of the top 15 hedge funds; and the reserve banks of all OECD member countries.

Founded in 1984, MathWorks employs more than 3,000 people in 15 countries, with headquarters in Natick, Massachusetts, USA.

For more information, visit [http://www.mathworks.co.uk/financial-services](http://www.mathworks.co.uk/financial-services)