State of Model Risk: A Philosophical Discussion of Model Risk in the Age of Machine Learning, AI and Deep Learning

Murphy Choy
Edinburgh Business School
Heriot Watt University
goladin@gmail.com

Ma Nang Laik
School of Business
Singapore University of Social Science
Singapore
nlma@suss.edu.sg

Abstract

Models are abstractions of reality expressed in mathematical terms. Abstraction simplifies the underlying process and phenomenon being investigated. Model risk arises when the abstraction process is inappropriate as a result of data used or model selected. Model risk poses a problem for any organisation that relies on models to perform tasks and making decisions. The presence of model risk has resulted in the development of model validation techniques to detect inappropriate models. Model validation techniques, such as k -fold cross-validation, has been widely adopted for statistical models and machine learning models. With the advances of newer models such as AI, deep learning and esoteric machine learning techniques, there is a need to develop the appropriate model validation techniques to manage this risk. In this paper, we discuss the philosophical issues for model validation and model limitations. Through this discussion, we hope to provide possible solutions to tackle model risk as well as general principles to tackle model risk in the situation where there are no precedents.

Keywords

Artificial Intelligence, Deep Learning, Machine Learning, Model Validation, Model Risk

Introduction

Mathematical Models are found in many organisations across industries such as logistics, banking and even government. These models are built as an aid to the organisation to enable them to measure, predict and guide their expectations and next course of action. Depending on the industry, some organisations may adapt mathematical models in different parts of the business. With the advances in Computer science, Analytics and Data science, coupled with technological innovations and improvements, more organisations are deploying mathematical models to gain competitive advantages.

However, this widespread application of mathematical modeling without careful management and consideration can be a risky proposition. Organisations in pharmaceuticals, material manufacturing and financial institutions are the first few organisations to adopt mathematical modeling and they have experienced the dangers of applying mathematical models. Clinical trials failure, material fatigue failures and credit risk crises are just some of the many dangers that arises as a result of model misspecification. This model misspecification is also otherwise known as model risk.

The recognition of model risk is one of the significant developments in recent history. There are good motivations for recognising models' risk or the blind belief in mathematical models that are poorly defined. Bad models were significant contributors to financial crises, aeronautical disaster and engineering failures. Memorable examples like the structured credit products and their role in the 2008 Global Financial Crisis and misunderstanding of the O-ring failure model in the Challenger disaster reminds modellers of the inherent dangers of over-reliance on a model that are not grounded in reality. The blind faith in the mathematical models used in Wall Street which led to the 2008 Global Financial Crisis is criticised in a special article by Wired Magazine (Salmon, 2009). The disconnection between mathematical model and reality in the Challenger Disaster is also highlighted Richard Feynman in a memorable quote below from the Roger Commission report (Feynman, 1986).

"For a successful technology, reality must take precedence over public relations, for nature cannot be fooled." - Richard Feynman, 1986

Mathematical models used in physical sciences, fortunately, are grounded in reality and when the models do not actually describe the reality, they are quickly superseded and disposed of in favour of those that are more accurate and plausible. However, models used in non-physical science such as psychology and economics are susceptible to deviation from reality. This natural susceptibility lies in the nature of the data. In the case of physical sciences, the data are generally collected from experiments or through observations of physical systems. Experiments are generally pre-defined and well-designed to ensure proper and accurate data collection. Even in the case of observational studies, there are usually proper design to ensure accurate data collection. In psychology, marketing, finance and economics, the data collected are expressed outcome and not entirely controllable or testable. This presents situation where the data collected may not be accurate and not measuring the right outcomes. When the model is built on these data, the abstraction of the model will be inappropriate.

In physical sciences, most mathematical models are pre-conceived or formulated before being verified or validated using the data. The models are usually quite simple and can be expressed in simple hypothetical terms. Even in the case of more complicated models, there are usually pre-defined models which are then tested using more complex methodologies. However, there are situations where it is impossible to pre-conceive a mathematical model. In these cases, mathematical models are built to fit the data in such a way to explain the relationships in the data. Such models are prone to misspecifications as the model is determined by the data and not necessarily validated as the appropriate model. Even when validated by an independent data set, the model may not have captured the underlying relationships accurately and can be invalidated by another data set. This uncertainty in the model structure forms an important aspect of model risk.

Any mathematical model depends on the underlying modelling assumptions. When there are violations of these assumptions or unexpected deviations, the mathematical will be susceptible to model misspecification or model risk. Most mathematical model assumptions relate to distributions and nature of data such as data symmetry and data point correlation. While these assumptions are important, most of the models ignore the information sufficiency assumption. Even in the case of big data, most practitioners building mathematical models do not attempt to determine whether the data used is sufficient for the model developed. The lack of concern for data sufficiency presents a dilemma where it is uncertain how much of the predictive or explanatory power of the model is determined by a small amount of data points. The other issue is the gap between the model assumptions and the reality. Most practitioners apply the models knowing the gap but believes that the assumptions have been proven or is not applicable to the case (Wasserman, 2015).

In the next section, we will discuss about the history of model risk primarily in the banking industry and how various organisations manage model risks. For the third section, we discuss about the inherent limitations of models. In the fourth section, the discussion will be tuned to solutions to the various types of model limitation. The last section will be the conclusion.

History of model risk management

Mathematical models have long been in use in large financial institutions globally (Jorion, 2006) especially for investment banks dabbling in financial instruments. As a general rule, the formality and comprehensiveness of the model risk management process is proportional to the complexity of the model. In the pre-2000s period, there was a proliferation of mathematical model use which led to a gradual recognition of the corresponding need for model risk management as a result of the LTCM debacle (Jorion, 2006). This period saw the establishment of mathematical model use as the decision support tool in various areas of banking such as valuation, credit evaluation, risk measurement. To tackle the various challenges in the applying mathematical models across different areas, there is an increasing need to customise or develop new models that led to an increase in complexity of models.

The first formal article on model risk management appeared in 2000 by OCC as the bulletin OCC 2000-16. The article is perhaps the first attempt at formalising model risk management practices, the need for such management and what are the possible approaches. The article highlighted many key issues with mathematical models such as the impact of errors, eliminating wrongful application of models, importance of appropriate training and the need for independent audits. Many principles listed were already adopted as part of the traditional model management process in other industries. The comprehensive list of principles, steps and processes were then adapted by the Basel Committee as part of the model validation requirements for internal models that forms the core of the "Basel II" risk-based capital requirements.

The formal adoption of model validation in the regulatory requirements forces the financial institutions to implement model validation. This marks the second period of model risk management between the release of the documents in 2000 to the end of the Global Financial Crisis (early 2010s). The period is marked by the proliferation of complex credit instruments through securitization. The eventual failure of these instrument is the result of the lack of understanding of the structure of these products. The underlying valuation models for these products relies on several core assumption which resulted in an assessment of the riskiness relied on key assumptions that will produce dramatically different results under moderate assumption violations. The 'Gaussian copula function', commonly attributed to David Li, is one such model. It's simplicity and ease of application to an otherwise complicated product drives the widespread adoption in the industry despite serious warning from Li and other researches (Duffie, 2007; Salmon, 2009).

The post Global Financial Crisis (post 2010s) period marks the advancement of model risk management into a systematic and comprehensive approach that extends beyond the pure mathematical aspects into governance and practices. The crisis revealed the widespread failures of risk management practices and economic capital models under stressful conditions. To prevent these recurrences, new regulatory acts are put in place with stringent and robust risk management tests required from the banks. The SCAP (Supervisory Capital Assessment Program) and CCAR programs are the two major risk management programs required for US Banks. Other regulators initiated their own version of the program to manage the risk of the banks as well.

With both SCAP and CCAR, capital regulation is focused on the approval of appropriate capital levels determined through the use stress test models. Both programs expanded the amount of data used from traditional sources such as balance sheet exposures to accounts, income statement changes, economic indicators and market information. Due to the complexities of CCAR reporting needs, there are many modeled outputs required. The models are subjected to a comprehensive and robust model risk management regime where model validation is merely a small portion of it. Building models is a resource intensive effort across an extended period of time.

The improvement in the risk management regime was reflected by the issuance of a joint Supervisory Guidance on Model Risk Management by both the Federal Reserve and the OCC (2011). This approach towards model risk management, defined in the guidance, is a vast improvement to the prior practices by recognizing the nature of mathematical models as an abstraction of reality which embodies various types of risks. To tackle the risk involved, there is a need for a sound model risk management framework. The guidance recognises that modeling is a process and should be viewed as a “life-cycle”, from the preliminary business need, model development, testing and validation, system implementation, regular monitoring and maintenance before retiring the model. Beyond the various steps in modeling, the framework also incorporates infrastructure, management review and board governance.

Since the great financial crisis, the industry is experiencing a new renaissance in applying machine learning and artificial intelligence to banking problem. The industry has been aggressively pursuing the active applications of these models to tackle fraud risk, credit risk, market risk amongst others. These models are complicated and generally black box models. This leads to complications when dealing with model validation. In the next section, we will discuss about the inherent limitation of models generally.

General Limitation of Models

Model creation involves the completion of a series of procedures which differs subjected to the kind of models involved. Model creation begins when the modeler first attempts to understand the key components and context of the real-world situation. After the appropriate familiarisation with the components, purpose and context of the situation, the modeler will determine the information requirements and assess the data available for modeling. With the right data available, the modeler will begin testing the underlying hypotheses and sub models of the full model. If the hypotheses and sub models are significant or operates well, they will be combined to develop the actual model. As the sub models are combined with various permutations, data about the model are collected and assessed for unusual anomalies. Once the modeler is satisfied with the model, the model will be tested for its validity which involves testing the model’s ability to reproduce historical results or new data with known results in terms of its accuracy. Once this has been done, the model is put into use. The entire process seems relatively straightforward but it is usually not the case. We will examine the major limitations of the modeling process.

Limitation 1: Every single model is the result of a series of human judgement.

Even though each step in the modeling process seems mechanical and straight forward, every single step requires human judgment. Mathematical abstraction of the real-world situation is fraught with difficulties. This is compounded by data quality issues. Determining the sub models and hypotheses optimality are non-trivial processes which becomes even more complicated when the information are combined to develop the full model. Human bias and implicit assumptions are built into the model at every juncture. Even for fully automated model building processes that involves Artificial Intelligence (AI), human biases and judgements are present in the AI which are inputs to the model as the necessary judgments are made at various steps. Judgments which are necessarily human in origin are inherent limitations. Thus, the use of any model must be accompanied by a detailed examination of the judgments and decisions associated with the creation of the model and how they influence the nature of the model. Explicit and thorough documentations are required to ensure proper understanding of the model for the model users.

Limitation 2. Models are limited representation of the real world.

All models are a form of abstractions of reality. Any real-world system/situation is highly complicated and contain numerous interrelated elements. Any modelers creating a reasonable version of the system will have to capture the critical components and relationships that simulate the real world system as closely as possible. Any attempt to abstract a sufficiently complex system will virtually be guaranteed to leave out some components either due to the perceived lack of importance or lack of measurements. When important components are left out of the model, the model might fail in certain situation as it does not contain the critical components.

Limitation 3. All Models Assume the Future will be Like the Past.

Most models are created using historical data to reflect a system that captures the information that has happened in the past. These models are very good at explaining what has happened in the past and intrapolate results based historical inputs. Unfortunately, most models are used to forecast future results based on future inputs. The models assume that the relations that are found based on historical data will remain the same in the future. This is of course an illusion. Any real-world system is susceptible to change that renders the model useless. The Global Financial Crisis in 2008 is a fine example of this.

Limitation 4. Data Issues

All mathematical models are the expression of the data used. This implies that should the data be compromised in any way, the mathematical model will also be compromised. The assumptions behind the data and how it is being obtained is important to the integrity of the model. There are multiple possible issues with data and in the era of big data, the problem is even more pronounced. Even though factors such as volume, velocity and variety are commonly cited, in modeling, veracity and validity of the data are more important and have far reaching implications for the use of the model.

Limitation 5. The Developmental History of a Model may be Unclear

Unlike the other limitations which are obvious, unclear model development history can be an odd limitation. Modelers frequently change and update their models to improve the accuracy of the model. There are many ways to update the model such as changing the inputs, adjusting the seed among many other techniques. These changes are often not included and does not provide information about the steps and alternatives explored by the modeler. Without these details, it is difficult for any modelers to independently check the process and pin point possible lapses in the development process.

Limitation 6. The Applicability of a Model is Limited by its intended objective

A model is developed for a particular purpose and objective. It is important that the model should only be used for that purpose. If the model is used for other purposes, its effectiveness and validity will be questionable. There are two major reasons why the application of the model outside of its intended use can be problematic. The first reason is the definition of the target variable that is used in the model. Unless the definitions are aligned, the model will not be measuring the same item and the interpretation of the results will be inappropriate. The second reason is that the underlying sub models and relations might be modeled using data from a different source than the one used for this. This implies that future data acting as inputs for this model might be different in nature and the measurement is different. Due to this, the predicted results will likely to be inappropriate for the new purpose.

Limitation 7. Model predictions are generally not deterministic, they are probabilistic.

Most model predictions appear precise and specific. However, the results are usually in terms of probability. It is important for any users to recognise that the model results are probabilities. Very often, the prediction is bounded by confidence bands. The most common measure of accuracy do not incorporate the confidence bound analysis. This makes it difficult to assess the accuracy of the model. Some of the new modeling techniques do not lend well to confidence band calculations. This can lead to wrong assessment of the model accuracy.

In the light of the limitations of mathematical models, it is important to tackle model risk by addressing the limitations and set up the appropriate measures and techniques to determine the model risk. Any model validation techniques must address the issues created by the limitations of modeling. The next section will address each limitation with existing model validation techniques and measurements.

Overcoming model limitations

Over the years, the banking industry has developed many approaches to tackle model risk through validation and audit processes (Sudjianto, 2017; Chang, 2016). Most approaches focused on the quantitative validation of the model which is affected by limitation 6. In this section, we will review each limitation and review what the current and possible approaches is to tackle the risk of the limitation.

Solution to limitation 1: Every model requires inputs and judgement from multiple modelers through collaborative work.

In the programming world, many developers have started doing pair coding/programming where a pair of developers will switch roles between programmer and auditor. This process is very useful as it enables both parties to code and review the program to ensure the quality and functionality of the code. In this way, the code is developed using the judgement of two individual simultaneously for a single project.

A similar concept can be extended to Modeling as well. However, pair modeling is a difficult process as the modelers might have different background and experience in building model. The other major pitfall is the need to come to a decision on the preferred approach and a considered judgement. To alleviate this shortcoming, it is better to have collaborative modeling. Collaborative modeling can be done in a small team where various team members work together to complete the model. The team can split into smaller team of odd numbers to perform the modeling tasks. There are also two approaches to this collaborative modeling approach - competitive approach or consensus approach.

Competitive approach encourages the various team members to tackle the problem individually with a model performance and validity measurement proposed by each member. The proposed measurement ensures that the various team member's core guiding metric is evaluated with others. By restricting to a single measurement, we ensure that only the most important and relevant metric used by the modeler in their decision process is proposed. Each member of the team will develop their own version of the model and then evaluate the models via the metrics. The optimal model will be determined as the one that has majority of the metrics voting it as the top model. The various team members will then review the champion model development process for any flaws. If there are no issues, then the model will be accepted. For cases where there are minor issues, the changes will be recommended by the team members which will then be implemented, and the metrics re-evaluated. Should the model still emerge as the champion, it will then be accepted. Otherwise the new champion model will be evaluated through this process.

Consensus approach is very different from the competitive approach. As opposed to individual members of the team coming up with their own best version of the model, all the team members will work and review the models building process simultaneously. At every single steps, the various team members will cooperate and work with one another to come up with new suggestions and possibilities for the models. The team members will iterate the process until all members accepts the results. This iteration to consensus is then repeated across the modeling process until the model is implemented.

Solution to limitation 2. Through constant updates and experimentations, models can become comprehensive representations of the real world.

Any model is ultimately a snapshot of the real world (Federal reserve, 2011). This is due to the fact that the model is built on a single sample of data taken from a population in a specific geography within a specific time range. The limited information makes it difficult for the model to be comprehensive even with multiple modelers working on it.

To create a comprehensive representation of the real world, there are two approaches that can be used. The first approach is to rebuild the model across multiple time periods with different demographic attributes. As we rebuild the model across multiple time periods with different demographic attributes, the model's core attributes will be reviewed, and the estimates reevaluated. The changes will be considered and used to determine what are the attributes that are changing and how much changes are there to the estimates. Depending on the type of model, there are specific reviews and tests to determine whether the model is different from the previous iteration. This approach is very suitable for a model used in a relatively static environment across multiple time periods.

The second approach is more suitable for the scenario where the environment is very volatile or dynamic. Models generally suffers from dislocation from reality because it contains relationships which do not adhere or explain the reality. In science, the modeling process involves the use of experimentations to collect information which tests the model. For the volatile environment, the model can be tested using a small sample of data that evaluates the ability of the model to explain the outcome. Due to the dynamic environment, the data sets are varied and do well to test the model in various circumstances. Should the model prove to be robust under such conditions, it can be said to be representative of the environment.

Solution to limitation 3. Testing the model with unseen inputs and validate the outcome to estimate the ability of the model to cope with major changes.

Unlike the second approach for limitation 2, the solution to limitation 3 is to test the model with unseen inputs that has never appeared in the data to test the model. This is to actually test the model's ability to produce reasonable results when given data that is totally unexpected under normal circumstances. The model is also tested to ensure that it does not bend the rules of reality under abnormal circumstances.

Solution to limitation 4. Data uncertainty should be a major measurement in the modeling process.

Data quality will be a major issue in any modeling process as it is the main ingredient for any models. While most modeling case study attempt to understand the impact of data in terms of its validity and reliability, both concepts are not practical in real world systems where the measurements are prone to errors and collection issues. Due to this, it would be better to assume that the data quality is compromised and create some form of data uncertainty metrics (Walker et. Al., 2003) that measures the quality of the model with respect to data. After all, a model built based on useless data is ultimately useless.

In Walker et. Al. (2003), the authors proposed two categories of uncertainty for data. The first uncertainty refers to the uncertainty of the system that generates the data. The second uncertainty refers to the uncertainty of data being driven by external factors. Both cases of uncertainty present the problem of veracity and validity and requires proper measures to evaluate their impact.

The first uncertainty is easier to solve as it involves the way the system captures information. To measure the uncertainty, pre-programmed scenarios are inserted into the system with known results. The output of the system will then be tested for its accuracy. By testing the various inputs from various systems, we can develop a metric that measures the deviation from expectation for all the inputs to the model that acts as a proxy to the uncertainty of the input data.

The second uncertainty remains elusive to measure. The inputs to the internal system are derived from systems external to the organisation. Some of these systems are not controllable and represent market forces. Thus, measuring the uncertainty requires use of proxy and measuring the variability of the data that is coming into the system. One possible way is to measure the inherent level of data variability in the inputs from the external system which is compared to other time periods. The comparison enables us to see how much data shift happens in the system and whether the level of variability in the data changes. By measuring these changes, this provides a measure of impact of external environment on the data coming into the internal system.

Solution to limitation 5. Model development history should be recorded automatically.

Recording the model development history in an automated manner is a very simple task with the current level of collaborative technology. By using the appropriate modeling system, the model development history can be captured easily and reviewed when needed.

Solution to limitation 6. Comprehensive model validity measurements under epistemic uncertainty.

The most difficult aspect of model validation is to identify the model risk under epistemic uncertainty where the user has no idea about the structure of the model and the relationships between the inputs. In recent research papers (Deng, Yu and Deng, 2018), there have been moves towards the use of Dempster-Shafer Theory of Evidence to develop quantitative measurement of model uncertainty.

Solution to limitation 7. Appropriate probabilistic interpretation rather than simple classifications results

The solution is difficult in some scenario as not all models are developed with probabilities in mind. To solve this, modelers has to actually compute the probabilities of their forecasts and predictions. This can also be solved through the use of software that compute the values automatically.

Conclusion and Future Direction

Through the discussion about the history of model risk management, we have reviewed all the events leading up to modern day model risk management practices and challenges. We have discussed in depth of the general limitation of mathematical models and how they impact the accuracy and validity of the models used in risk management. With the relevant solution and measurements reviewed, we hope that risk managers can benefit from this discussion and provide new directions to tackle model risk issues.

References

Basel Committee on Banking Supervision, 1996. "Amendment to the Capital Accord to Incorporate Market Risks (No. 24)", January. Available at http://www.bis.org/publ/bcbs24.htm

Basel Committee on Banking Supervision, 2011. "Basel III: A global regulatory framework for more resilient banks and banking systems," June. Available at http://www.bis.org/publ/bcbs189.htm

Deng, W., Lu, X. and Deng, Y., 2018. Evidential model validation under epistemic uncertainty. *Mathematical Problems in Engineering*, *2018*.

Duffie, Darrell, 2007. "Innovations in Credit Risk Transfer: Implications for Financial Stability," pp. 40-41. July. Available at http://www.darrellduffie.com/uploads/working/DuffieInnovationsCreditRiskTransfer2007.p df

Feynman, Richard P. (1986) Appendix F- Personal Observations on the reliability of the Shuttle.

Hsiu-Mei Chang, 2016. Model Risk Management. CAS ERM Seminar

Jorion, P., 2006. Risk management for hedge funds with position information.

Sudjianto, A., 2017. Quantitative risk management and stress test to ensure safety and soundness of financial institutions. *IFC Bulletins chapters*, *44*.

Salmon, F., 2012. The formula that killed Wall Street. *Significance*, *9*(1), pp.16-20.

Reserve, F., 2011. Supervisory guidance on model risk management. *Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, SR Letter*, pp.11-7.

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B., Janssen, P. and Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, *4*(1), pp.5-17.

Wasserman, L., 2015. *The Role of Assumptions in Machine Learning and Statistics: Don't Drink the Koolaid!*. Technical report, Carnegie Mellon University, 2015. 8.

Biographies

Murphy Choy is currently a director of Technology, Operation and Analytics. His research is in the area of data refinery development, application of quantitative models to communication and application of quantitative methods to solve real life challenges. He was a former instructor in Singapore Management University and has won a teaching award for his work. He holds a Doctorate of Professional Studies in Business Analytics from Middlesex University London, Masters of Finance from University College Dublin and Bachelor of Statistics from National University of Singapore. He also holds a diploma of Economics from University of London and Postgraduate Certificate in Business Research from Heriot-Watt University.

Ma Nang Laik is a Senior Lecturer in the School of Business at one of the autonomous universities in Singapore, Singapore University of Social Sciences (SUSS). She teaches quantitative method, business skills and management, business analytics applications and supervises many students' final year projects. Prior to that she has been working as a director of Master of IT in Business-Analytics (MITB-A) programme in School of Information Systems in Singapore Management University (SMU) for seven years. She holds a PhD from Imperial College, London where her research interest is to apply operations research (OR) to solve real-world problems in the area of optimization of resource, capacity planning, and simulation and decision support systems. She has presented her work in various well-known international conferences and is a winner of EFMD case competition in year 2016. She has worked in MNCs for many years before joining academic and is still working as an academics consultant for business organisations. Her research expertise lies in the simulation and modelling of large scale real-world problems and the development of computationally efficient algorithms to enable sound and intelligent decision making in the organization.