Model Risk Assessment Series

Model doesn't begin with a model

Owen Lau, CFA, FRM
August 2013

1. Executive Summary

The recent expansion of model development from origination to exposure valuation benefits the decision making process of the management in financial institutions, however, models also come with costs, as well as risk. During 2008, many models returned incorrect prediction or valuation because of the wrong assumption of correlation during financial distress. Afterwards, model risk came to attention to the industry. The industry then shifted the focus from model implementation to model risk assessment. The understanding of the model risk assessment starts with the understanding of model development. Each step of the model development process has potential risks. This article focuses on the first two steps of the model development process: data collection and data cleansing. During the data collection and data cleansing process, the key risks include:

- **Initial variable selection risk**
- **Data availability risk**
- **Data extraction risk**
- **Data integrity risk**
- **Data transformation risk**

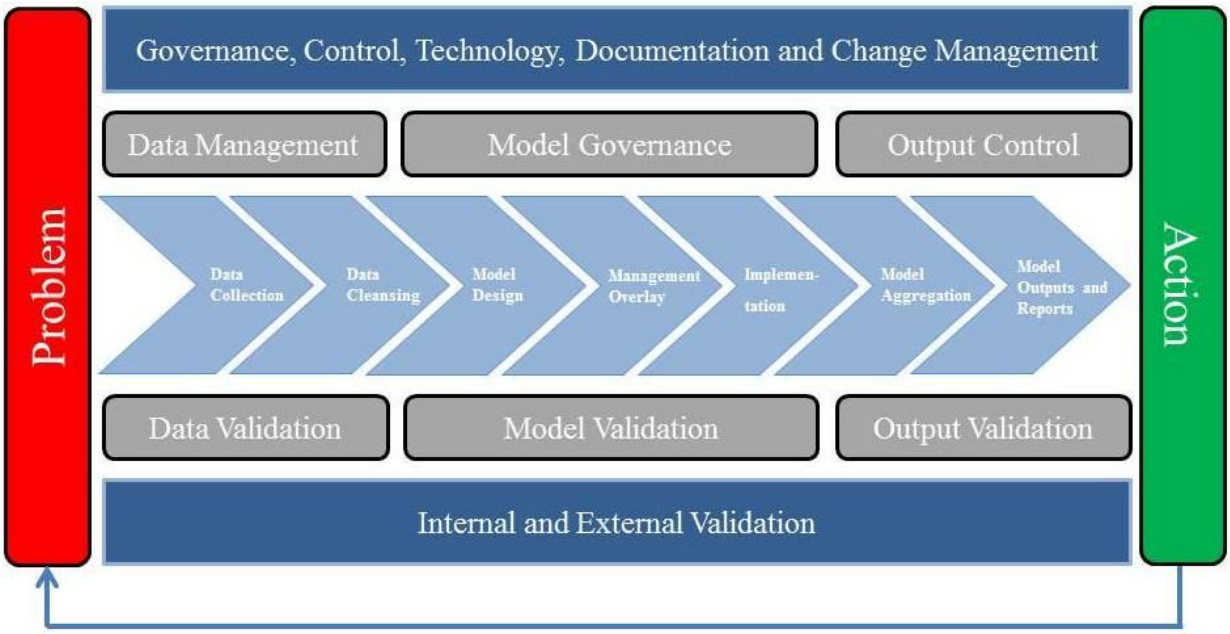
2. Getting Start

Model development doesn't begin with the model, it begins with a problem. The purpose of a model is to provide one, but not unique, solution to solve a business problem. An effective model commences with a clear statement of purpose to ensure that the model built will align with the intended use. For example, probability of default (PD) model is to quantify how likely the counterparty will be unable to pay back, stress testing model is to give management and regulators the sense of loss and capital level under different economic stress scenarios so that they can formulate the action plan accordingly. Once the problem is clearly defined, the project team and each person's role and responsibility have to be determined. The project team usually comprises of model owner, project manager, project sponsor, model developer, model auditor and IT expert. The project manager should compile a project timeline which is reasonable to the development of the model and get the signoff from the project sponsor. The key risk at this stage is the choice of people and technology because they are very sticky to the project, organization should consider the complexity of the model, the knowledge level of the developer, the vulnerability of the technology and the regulatory impact of the model. Typically the organizational or project structure is outside the scope of model risk assessment, but the importance of the project structure can't be under-estimated.

3. Effective Model Risk Assessment

In order to assess the model risk appropriately, we need to understand the steps of the modeling process. **Exhibit 1** shows the model risk assessment framework.

Exhibit 1
Model Risk Assessment Framework



Seven Steps of Model Development

1	***Data Collection***
2	***Data Cleansing***
3	***Model Design***
4	***Management Overlay***
5	***Implementation***
6	***Model Aggregation***
7	***Model Outputs and Reports***

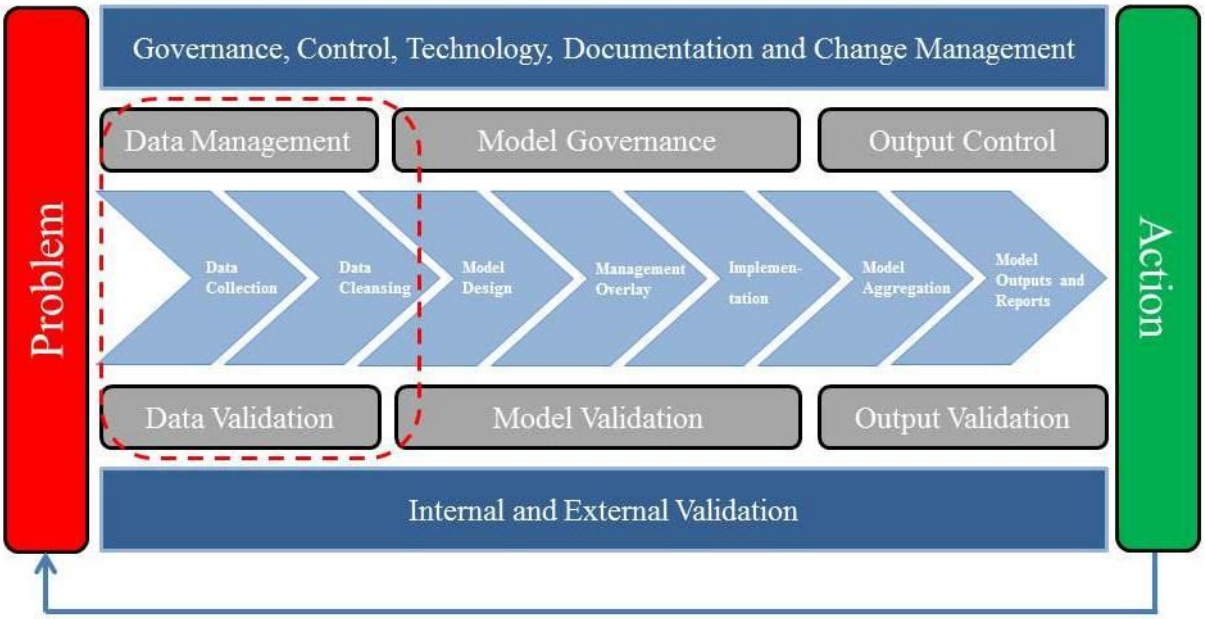
Throughout the process, control and governance should be applied to make sure the model development is in compliance with regulatory requirement and within the project timeline. There should be policies in place so that the project team acts within the boundary of the organization. Documentation is important but is often ignored. Appropriate documentation engages the senior management to get the sign-off, it also helps knowledge transfer because the model developer is usually not the model user. In addition, the model validation team relies heavily on the documentation to undertake the qualitative and quantitative analysis to assess the risks of the model. Another area that cuts across the process is change management. Model, database and process are changed and upgraded continuously, effective monitoring and management of the version is critical for the process continuity. Internal and external validation provides assurance

to the model. In many U.S. financial institutions, the emphasis of validation is on model validation, as illustrated in the guideline from Federal Reserve Board SR11-7 and OCC 2011-12 "Supervisory Guidance on Model Risk Management". Model validation is the set of processes and activities intended to verify that models are performing as expected, in line with their design objectives and business uses. Model validation team should be independent from the model development team in order to provide an objective assessment of the model and identify the potential limitations and assumptions. Ultimately the outputs of the model will provide the model user an idea – typically not a direct outcome – to formulate an actionable plan to answer the business problem.

Each step of the process represents a major milestone of the project. The project team takes a significant effort in each of the above steps for the model development, because of the amount of resources put in each step, it adds a certain level of risk to the process. Below will illustrate more on the risks of the data collection and data cleansing process, as shown in **Exhibit 2**.

Exhibit 2

Model Risk Assessment Framework



4. Data Collection

It is the process to identify the initial set of financial and macro-economic variables. The experience and judgment of the project team including model owner, developers and other experts will greatly influence the appropriate selection of this initial set. They apply their day-to-day experience to make intuitive inference between the predictors and the outcome. For example, unemployment rate is commonly used in predicting the probability of default (PD) of a credit card portfolio because a job reflects the repaying ability of a consumer. House price index (HPI) is another commonly used indicator to predict the loss given default (LGD) of a residential real estate portfolio. Financial variables vary differently for different models (e.g. PD vs LGD) and line of business (e.g. credit card vs commercial mortgage). For example, financial variables of commercial portfolio include total revenue, earnings before interest and tax, interest expense, total asset, total interest-bearing debt, total liability, capital expenditure and etc. Macro-economic variables include unemployment rate, house price index, consumer price index, real GDP growth, 10-year treasury yield and etc.

Once the initial set of variables is defined, the project team has to compile the data dictionary which lists out the field name, data type, field definition and data source of each variable. Sometimes there are multiple databases storing the same financial variables, so the project team has to assess which location is the right place to extract the data point. Another consideration is the length of year of the data extraction. BASEL requires that the minimum data observation period for PD and EAD should be at least five years while the minimum data observation period for LGD should ideally cover at least one complete economic cycle but must in any case be no shorter than a period of seven years for at least one source. If the available observation period spans a longer period for any source, and the data are relevant, this longer period must be used[1]. The last step of the data collection process is to extract the data from the original source to the database for data cleansing.

What are the risks?

► Initial Variable Selection Risk

The initial variable selection is a multidisciplinary activity drawing on economics, finances, statistics, mathematics and other fields. It is critical to select the right set of financial and macro-economic variables as the predictors, the experience and judgment of the project team will determine the predicting power of the model, so there is a risk of not selecting the right set of initial variables for modeling. To minimize this risk, the composition of the group to select the variables has to be examined and the initial set of variables has to be reviewed to make sure that it makes intuitive business sense for those variables to align the intended use.

► Data Availability Risk

When the group of developers selects the initial set of variables, not all variables are always available, it draws on the risk of data availability. For the probability of default

(PD) model for credit card portfolio, it is desirable to get the personal income level of the credit card holder as the predictor, however, usually it is not easy to capture the personal income from the credit card holder even though personal income could have been a strong predictor for probability of default. Sometimes data proxies are used in replacement of the unavailable data, however, they should be carefully identified, justified and documented.

► Data Extraction Risk

It is a highly omitted area. Data extraction sounds like a routine, straightforward process but it actually posts a key risk which is not easily identified. Typically, financial variables are stored in an original database, they are extracted, converted and aggregated to a certain file format and are sent to another database for further processing (e.g. data validation). In one of my client engagements, I saw an error occurred during the data extraction process. It was due to the fact that the original database allowed an input of full stop (.) to indicate a missing value of the long term debt in the financial statement. During the extraction and aggregation process, all the data had to be converted to comma-separated values (CSV) for data aggregation and transmission. When the file was transferred and uploaded to the database for data cleansing, the full stop (.) disappeared, no value was entered into that field. It was important to differentiate between full stop (.) and no value. Full stop (.) told the project team to follow up with the missing value, getting down into the source and filling it back, or deriving it from other fields. While no value indicated a zero (0) value which was consistent with those companies with zero long term debt. So comparing the "new" data with the "old" data before extraction to minimize the conversion error is critical.

5. Data Cleansing

It is the process to clean up the data. Usually the data in the original source is incomplete and contains many irregularities. Missing values make it hard to run the basic statistical analysis like correlation and outliner will skew the result under normal condition, so a thorough data cleansing process is crucial to remove the sand and stone before cooking. There are many ways to handle the missing values, common approaches are using the mean, median or a certain percentile of the data field to backfill the missing value. Outliners can be defined by plus or minus three standard deviations from the mean so that extreme values are ignored. Sometimes transformation is necessary, for example, EBIT and interest expenses are important elements to predict the default probability, having the two elements together with other variables may stretch the model too big, so interest coverage can be used. Such kind of ratio transformation can reduce the number of predictors used in the modeling process. In addition, dummy variables may be used to indicate discrete event and ranking transformation may be included in the selected variables.

What are the risks?

► Data Integrity Risk

Data integrity is often a cause of headache for developers, when many fields are selected and inter-correlated, sometimes they don't agree with each other. For example, balance sheet items like total asset, total liability and owner's equity are often picked in the initial set of variables, there is a risk that the data entry person enter one of the values incorrectly, say total asset equals to 100, owner's equity equals to 20 and total liability is entered incorrectly as 8 instead of 80, the model may play down the effect of total liability even though it reflects the burden of the company. So data consistency check is a critical process to verify the inter-relation of each data field. Data validation rule can be introduced to cross-check the agreement of the basic accounting rule like total asset = total liability + owner's equity, this kind of accounting validation rule can filter out some inaccurate data fields in income statement and balance sheet. Data integrity check is often an area to be passed over, especially for smaller models, however, the importance of data agreement shouldn't be underrated. As noted in the SR Letter 11-7, the data and other information used to develop a model are of critical importance; there should be rigorous assessment of data quality and relevance, and appropriate documentation. Developers should be able to demonstrate that such data and information are suitable for the model and that they are consistent with the theory behind the approach and with the chosen methodology. [2]

► Data Transformation Risk

When developer transforms the data from one domain, say an index, to another domain, say a ranking, the developer may lose some information in the transformation process. Typically ranking has the characteristic of equal scale between any two ranks, it may not be the case for the pre-transformation domain. So the transformation rationale has to be clearly examined and evaluated for relevance, assumptions and limitations.

6. Concluding Remarks

Model risk assessment has become an inseparable part in the model development process, financial institutions are expected to validate the whole model development process in a comprehensive and holistic way in order to meet regulatory requirement. The methodologies used have to be transparent, robust and complete, in the sense to not only cover the model implementation process but also cover the input. A complete suite of framework is necessary to avoid the garbage-in-garbage-out impression.

7. About the Author

Owen Lau has over ten years of experience in investment and risk analysis. He currently works with S&P Capital IQ. In his role, Owen is responsible for credit assessment, default and recovery study, publishing thought leadership articles and managing client engagements across various

asset classes. Prior to S&P, Owen worked as a research analyst with a value oriented hedge fund Pardus Capital Management, a bank examiner with Federal Reserve Bank of New York and a credit analyst with Barclays and HSBC.
Owen holds the designation of Chartered Financial Analyst (CFA) and Financial Risk Manager (FRM). He graduated as an MBA from Columbia Business School where he was awarded the Dean's Graduation Honor (top 25%). He obtained a MSc from Purdue University where he was inducted as Beta Gamma Sigma member. He also holds a BBA from The Chinese University of Hong Kong.

8. References

[1] Basel Committee on Banking Supervision "International Convergence of Capital Measurement and Capital Standards" June 2006

[2] Board of Governors of the Federal Reserve System "SR Letter 11-7 Supervisory Guidance on Model risk Management" April, 2011