

Reducing Model Bias Bayesian Model Selection and Averaging



ABSTRACT

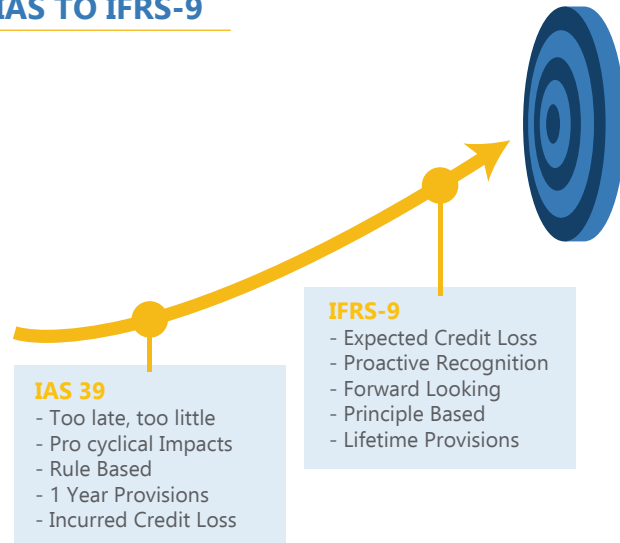
The advent of IFRS 9 across the Banking industry has led to the emergence of various challenges for risk managers. Given the current practices with respect to incorporation of macroeconomic impact in risk parameters, it is imperative to have a robust framework for the same. The current blog enumerates the current industry practices and introduces the methods of statistical model averaging. The method of Bayesian model averaging, which has become an important tool in empirical settings with large numbers of potential regressors and relatively limited numbers of observations is described along with an illustration to provide the reader a perspective of its application as well.

1. Introduction

With the onset of the 2008 financial crisis, followed by the European financial crisis, Brexit and various other macroeconomic changes, the trajectory of the world economy has undergone significant changes, leading to rapid changes in the global financial systems. During the same time period, the IAS39's incurred loss approach had come into heavy criticism for amplifying the pro-cyclical effects on the bank capital regulation and is also partly to be blamed for the 2008 crisis. To overcome these limitations, the IFRS 9 standards were finalized in July 2014 and came into force since January 2018.

The advent of IFRS9's expected credit loss approach in response to IAS39's incurred loss approach has led to the requirement of incorporating forward looking loss estimates for estimating loss provisions. This requires risk managers to build a framework that can take into account the macroeconomic conditions and provide point in time based risk parameters.

IAS TO IFRS-9

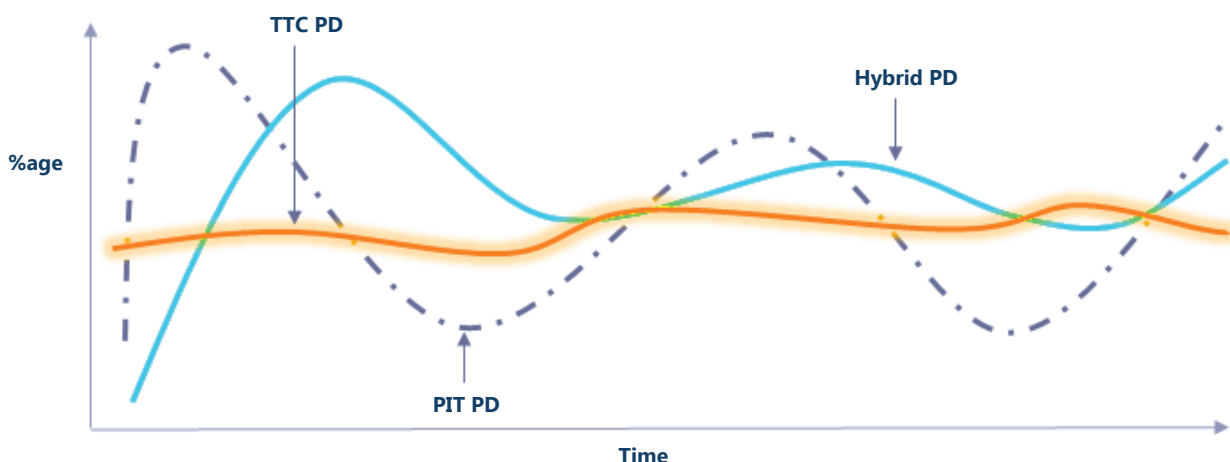


Currently the industry practices are dominated by regression based methods wherein a single equation is used. But given the research in the field of predictive analysis and the advancement in the computing power, particularly the method of Bayesian model averaging (BMA), these contemporary methods have emerged as a means to overcome the limitations associated with a simple regression based framework. BMA has emerged as an important tool to deal with model uncertainty and is particularly useful in empirical settings with a large number of potential models but with limited number of observations. The current blog provides an overview of the model averaging and selection method and discusses how the same can be used to provide more reliable point in time parameters.

2. Characteristics of an Ideal Framework

The risk parameters for IFRS9 namely PD, LGD and EAD are also required to be estimated in accordance with the future looking requirements. One of the key features of the IFRS9 framework is the usage of point in time based risk parameters. For instance, the PIT PD is the PD which is reflective of the current economic conditions and its usage ensures that the risk estimates are reflective of the economic conditions as perceived by the risk manager. Also it must be observed that the TTC PD which takes into account multiple economic cycles is much smoother than the PIT PD. The same can be observed from the graph below wherein the PIT PD exhibits more volatility than the TTC PD.

PIT PD vs. TTC PD vs. Hybrid PD



Hence, given the criticality of the task above wherein different methodologies can be used for adjusting the TTC PD for the impact of the economic cycle to derive a PIT PD, it is important that the framework adopted to perform this task should have the following characteristics:

— Well Specified:

The adopted framework should not suffer from any bias. All the important factors that may impact the expected loss (EL) should be included in the framework.

— Flexibility:

Given the innovations and the rapid changes in the global financial system, it is imperative that the framework adopted should be capable of being modified to incorporate these changes.

— Occam's razor or Simple Model:

A scientific principle according to which the best explanation of an event is the one that is the simplest, using the fewest assumptions or hypotheses; hence suggesting that the framework should not be overly complex.

— Realistic:

The adopted framework should be accurate and have practical assumptions.

— Conceptual Insightfulness:

Conceptually insightful models reveal fundamental properties of economic behaviour or economic systems hence making the system of equations relatable to the observed economic phenomenon.

— Parsimony:

These are simple models since they rely on relatively few special assumptions and they leave the researcher with lower degrees of freedom. Parsimonious models are desirable because they prevent the researcher from consciously or subconsciously manipulating the models so that it over-fits the available facts.

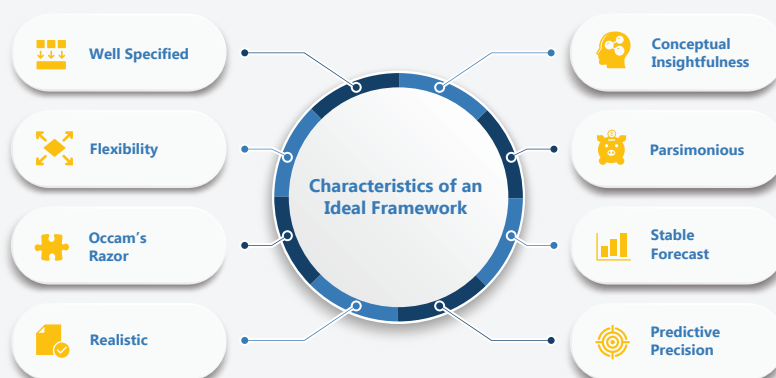
— Stability of Forecasts:

The adopted framework should yield results that are not too volatile; since higher the volatility in the forecasts, the higher would be the quantum of the change in the provisioning numbers that the bank has to undertake. Hence from the implementation and business stability point of view as well, it is desired that the adopted framework should yield reliable and stable forecasts in accordance with the macroeconomic conditions.

— Predictive Precision:

Models have predictive precision when they make precise or strong predictions. Strong predictions are desirable because they facilitate model evaluation and model testing. A model with predictive precision has greater potential to be practically useful if it survives empirical testing. Models with predictive precision are useful tools for decision makers who are trying to forecast future events or the consequences of new policies.

Characteristics of an Ideal Framework



3. Current Practices and Limitations

One of the most practiced method used for deriving the PIT PD from TTC PD is to build a single equation regression model with easily interpretable and significant macroeconomic variables to derive the PIT PD forecasts. Herein a linear regression framework is used with contemporaneous and lagged values of one or more regressors with the primary aim of establishing a satisfactory model which can exhibit the relationship between the bank specific default rates and the macroeconomic factors.

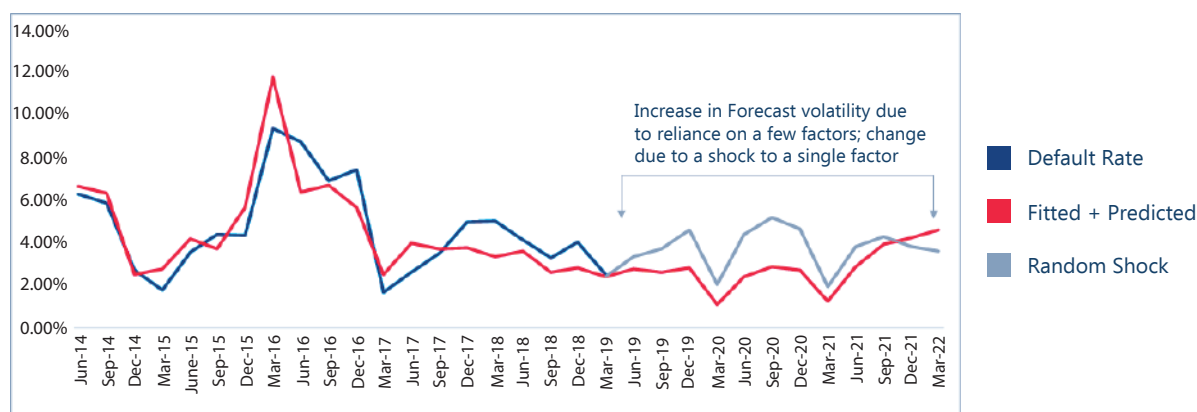
It must be noted that in the emerging markets, the historical default rates data is typically available for only 5-7 years or even lesser; thus even with quarterly default rate computation, a macroeconomic model has 20-28 data points to work with but in situations wherein the number of prospective regressors are close to the number of data points or much more than that, the primary problem then is to select a subset of

variables which can be used to predict future behavior of default rates. Given that the macro variables (herein, the independent variable) are related to the underlying country's economy, there is a high possibility of observing a high degree of multicollinearity amongst them.

Further, the linear regression models use a single equation for making the predictions, this increases the likelihood of excluding key variables as the modeler is forced to choose a few variables in the final model. This also leads to an increased uncertainty pertaining to the range of scenarios the underlying model can incorporate, the stability of the estimates given the sparse number of data points and a low-out-of-sample predictive power as well.

The exhibit below displays the best possible model derived along with out of time forecasts for a series of randomly derived default rates. As can be observed that in spite being a good fit, the current model is still reliant on only a few select macroeconomic variables due to the limited number of data points. Since this model is dependent only on a few macroeconomic variables, shocking one of the variables leads to a significant change in the predicted default rates hence indicating the vulnerability of single equation model; this also leads to erratic forecasts, which can then lead to an erratic change in the provisions as well.

Simple Linear Regression Framework (Default Rates, Predicted Values and Random Shocks)



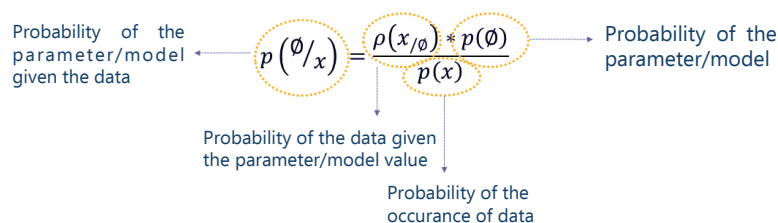
As seen above, under the current approach the variable selection is a combination of statistical methods and judgement of the risk manager. But even if they are deemed satisfactory from a supervisory/ audit point of view, these equations might lead to skewed estimates pertaining to the severity of the macroeconomic situation and hence can lead to incorrect estimation of the loss absorption capacity of the Bank. Moreover, there is no certainty that these mis-specified relationships would hold in the future; hence model uncertainty is a critical issue in the current approach.

4. Model Averaging and Bayesian Model Selection

To overcome the above cited problems and take into account the characteristics which an ideal model framework should possess, the current situation warrants the usage of the statistical methods of selection and averaging. These methods have gained prominence in the recent times since they have emerged as an important tool to deal with model uncertainty, given the current empirical scenarios wherein a large number of potential input variables are there but the number of observations for the dependent variables are limited.

These averaging methods are primarily used to avoid the reliance on a single equation model and to construct the final model such that it takes into account various variables at the same time. This is made possible by estimating all the possible combinations of models and then averaging across the various models to arrive at the final estimates/ forecasts. These methods of model averaging can be further branched into Bayesian Model Averaging (BMA) and Frequentist Model Averaging (FMA).

Bayesian Model Estimation

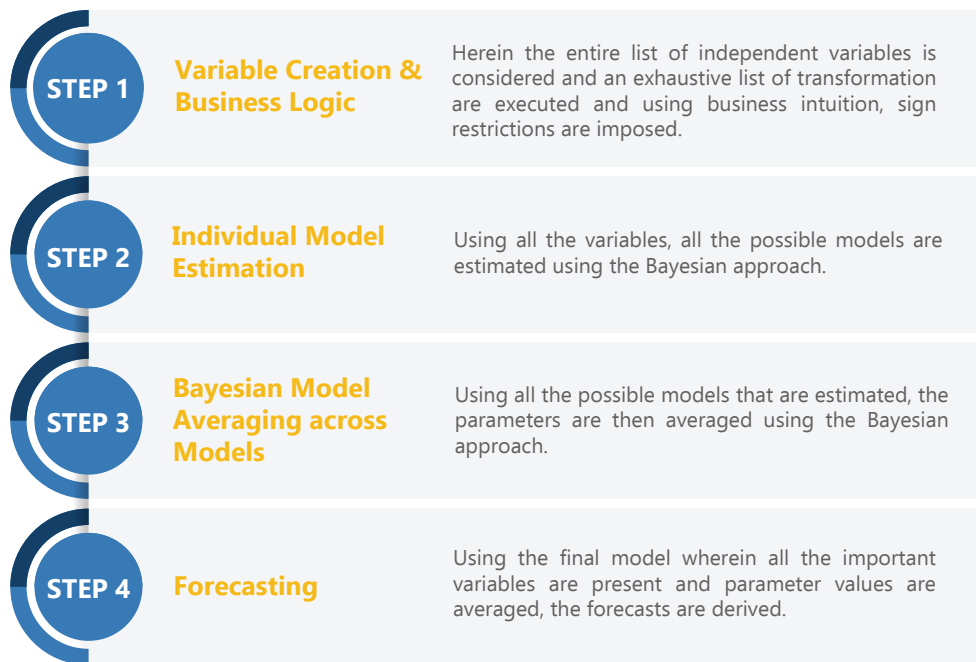


The frequentist methods of model averaging (FMA) tend to focus on deriving the model parameters exclusively from the data. But given the data availability and advancements in computing ability, BMA has emerged as a superior method of obtaining the predictive inference.

The Bayesian Model Averaging overcomes the model uncertainty inherent in the variable selection problem by adopting a two-pronged approach. In the first stage, all the possible models are estimated (if there are n variables then 2^n models are estimated), while the next stage involves averaging over these models according to approximate posterior model probability.

BMA is able to overcome the problem of model uncertainty by taking into account the plethora of independent variables that can be included in the final model instead of relying on a limited set of variables. Herein the weights for the parameter/ model are decided in accordance with the posterior probability of the parameter/ model.

Hence, the probability distribution of the final parameters (posterior) in the model is derived via a combination of the beliefs/ expectation of the modeler or theory (the prior) and the trends exhibited by the data (likelihood function).

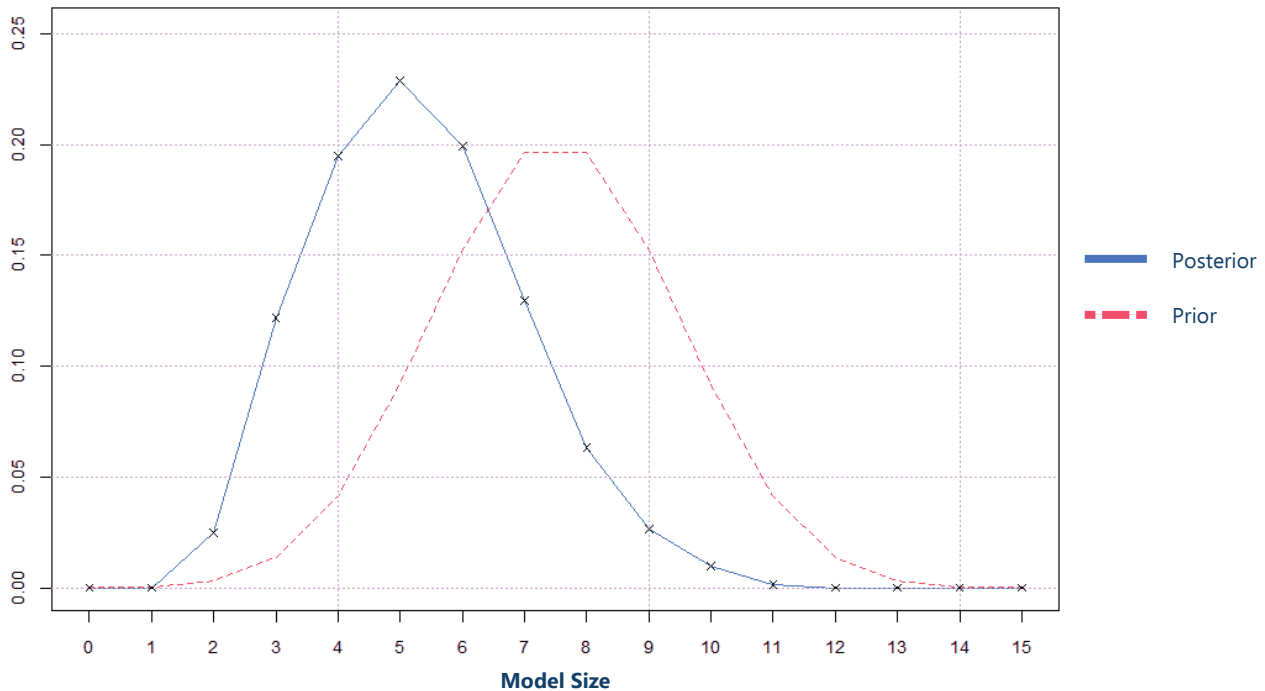


The following illustrates the use of BMA on the same set of data as used in the earlier illustration. As can be observed, the following table ranks all the variables as per their posterior inclusion probability, which indicates the chance that a variable would be included in plethora of models that are estimated. The column Post.Mean provides an estimate of the coefficient of the variable that is averaged across various models for the top 15 variables.

Variable	Posterior Inclusion Probability	Post.Mean
Variable 1	27%	0.01235814
Variable 2	22%	0.00106347
Variable 3	15%	0.00062842
Variable 4	15%	0.01105218
Variable 5	14%	0.00045150
Variable 6	13%	0.00245779
Variable 7	13%	0.00008134
Variable 8	13%	0.00120814
Variable 9	11%	0.01426793
Variable 10	11%	0.00042194
Variable 11	11%	0.00000173
Variable 12	11%	0.03059667
Variable 13	11%	0.00506155
Variable 14	10%	0.00153302
Variable 15	10%	0.00012416

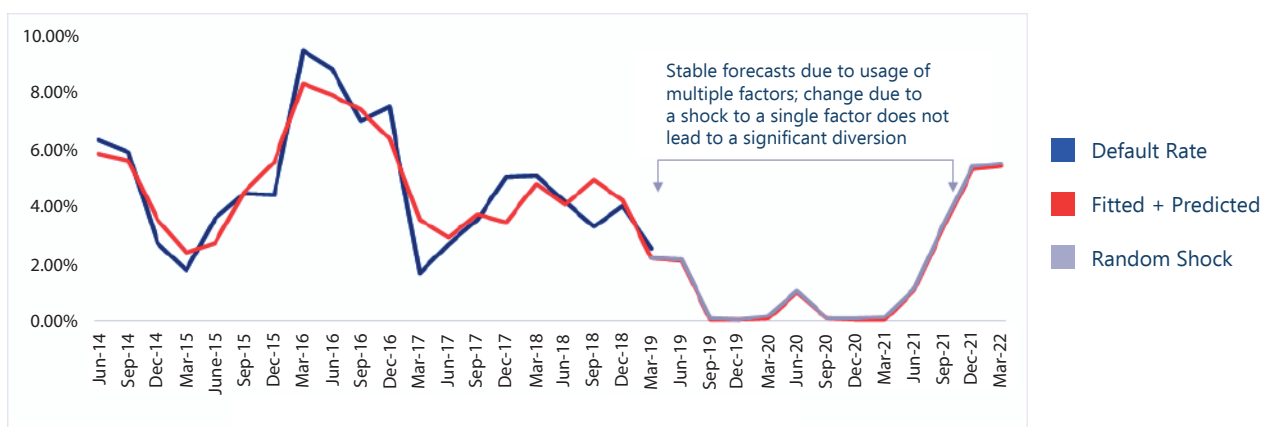
Using the Bayesian approach the user can then also determine the average numbers of variables used per model as well. The exhibit below indicates the same. As can be observed, the average number of variables per model is approximately five. Hence the number of variables considered per model is significantly higher than the number of variables used in a single equation model.

Posterior Model Size Distribution (Mean: 5.297)



Lastly, the exhibit below highlights the fitted and the predicted values in case of BMA. As can be seen that the model herein not only has a better fit via the application of BMA but the forecasts are much more stable. Hence this approach leads to more reliable and stable forecasts.

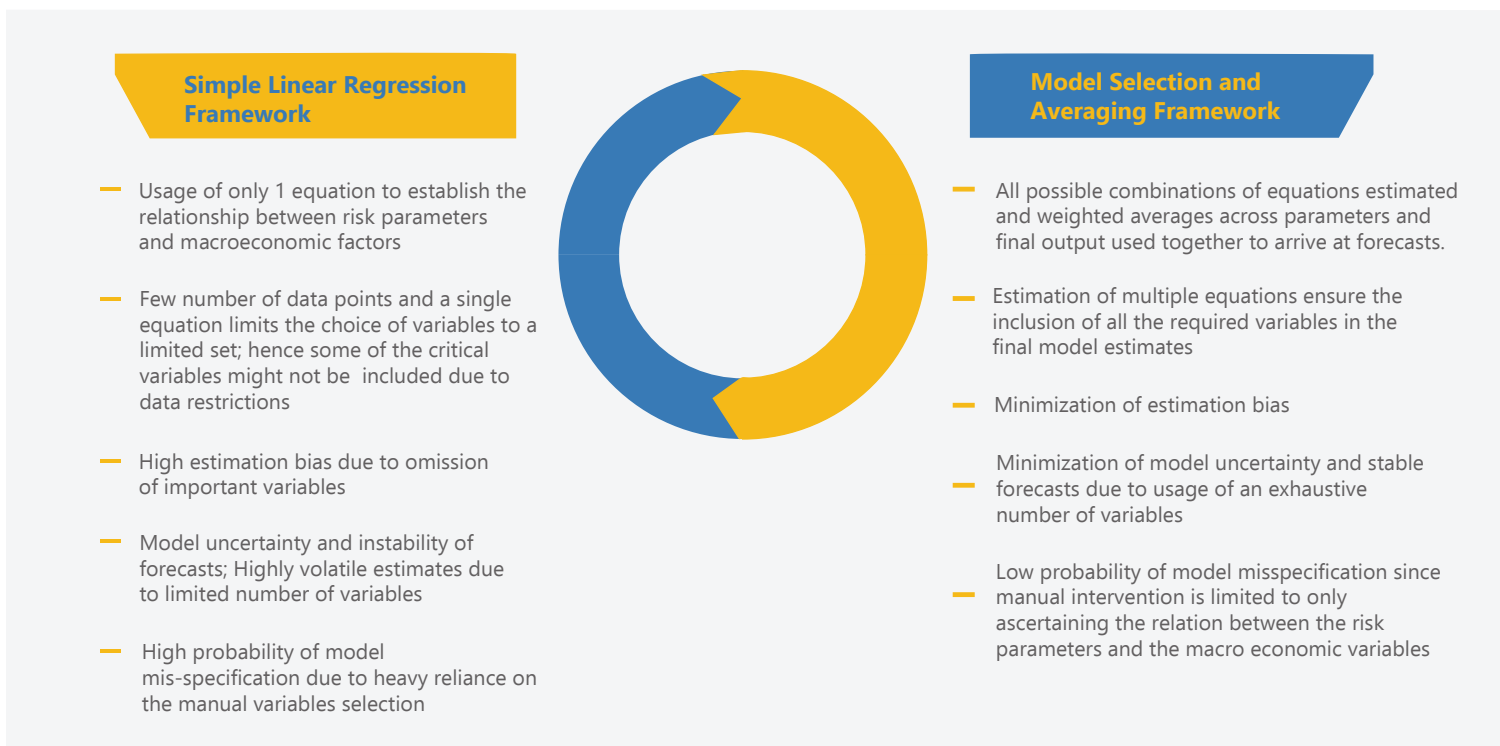
Bayesian Model Averaging (Default Rates, Predicted Values and Random Shocks)



The methods of model averaging tend to use all the variables as potential inputs for estimating all possible models. Further, these several models are then averaged across to arrive at the final model which provides more reliable and stable forecasts

But it must be noted that as per the latest advancements, trimming of a model space is associated with an improvement in the model's performance. Hence for performing the same, various steps can be taken. For instance, considering only those models wherein the sign of the variables is in line with the expectations/ theory, only the top performing models are considered rather than all the models, thus creating a cut-off for the maximum number of variables that can be included in a model etc.

The application of the above would then lead to the estimation of a model that would not only satisfy the requirements of an ideal framework but also would be a robust model as well. The table below provides a point to point comparison between a simple linear regression framework and the model averaging and selection methods that have been discussed above.



To conclude, given the contemporary practices wherein there is an over-reliance on simple linear regression based models and are expected to suffer from not only the model selection bias but also are susceptible to volatile forecasts, the method of Bayesian model selection and averaging provide the perfect solution. These methods not only overcome all the shortcomings of the current practices but also provides the user the ability to incorporate maximum possible iterations so that the prediction results are exhaustive when taking into account the macroeconomic variables.

Contacts



Feel free to send your queries to:

Sandip Mukherjee

Co-Founder | Principal Consultant

Email : sandip.mukherjee@aptivaa.com

Phone : +91- 98210- 41373 | +91- 22- 40838600

Author



Jagan Wadhwa

Senior Consultant

Email : jagan.wadhwa@aptivaa.com

Phone : +91- 22- 40838600

About Aptivaa

Aptivaa is a vertically focused finance and risk management consulting and analytics firm with world-class competencies in Credit Risk, Market Risk, Operational Risk, Basel III, IFRS-9, Risk Analytics, COSO, ALM, Model Risk Management, ICAAP, Stress Testing, Risk Data and Reporting. Aptivaa has emerged as a leading risk management solutions firm having served over 100 clients across 22 countries comprising of highly respected names in the financial services industry.

Aptivaa

USA | UAE | INDIA

Disclaimer:

For our analysis as represented in this document, we have used commercially available market data obtained from sources we generally believe to be reliable. We are not giving an opinion or any other form of assurance on information from these sources. Unless otherwise noted, the values calculated by us are derived using applicable market data parameters and generally accepted valuation methodologies.

Contact: insights@aptivaa.com | [Website: www.aptivaa.com](http://www.aptivaa.com) | www.linkedin.com/company/aptivaa