# Betasmartz Research – Confidential - Business Cycle Classification Algorithm

## Objective

The intention of this document is to introduce and outline all the steps that are involved in determination of the business cycle classification algorithm.

This algorithm provides the probability forecasting model for a particular labeling in time.

## Data Needed

Two types of data are needed, one will be the labeling that wants to be identified for the algorithm and the other will be the potential explanatory data for the label variable.

## Modelling framework definition

There are numerous classification algorithms that could provide a classification regarding the business cycle identification. However, not all the methodologies are robust and most of them are prone to overfitting. In this case the algorithm that was used is called a Bayesian Modeling Average. This technique is very robust and is not sensitive to overfitting so it almost assures from its building blocks that the model reflects the actual state of the population distribution.

### Brief introduction to BMA

The Bayesian model averaging is a technique created to adress model uncertainty in selecting a particular model structure (it is not only dependent on the regression problem).

Let me explain further with a little example:

*There is a researcher that wants to assess the size of a covariate effect on survival time with a view of designing future interventions and additionally he would like to predict the survival time for different patients. So he conducts a data-driven search to select covariates for a specific proportional hazards regression model, M\*, that will provide the framework for inference. So he checks if the model M\* fits to the data and then it proceeds to estimate the models significance, test the covariates size effects and finally make predictions. This is a standard statistical procedure but sadly it is not entirely satisfactory. Suppose there exists an alternative way to fit the proportional hazard model, M\*\*, that also provides a good fit but leads to significantly different estimated effect sizes, different standard errors and different predictions? Sadly there is no standard procedure on how to tackle this situation, proceed with the model M\* will be too risky and could lead to ambiguity. There is where the model selection procedure comes as a solution of this problem and Bayesian model averaging is a way around this problem.*

The BMA adress this problem with the following equation:

This has a practical interpretation of:

***“The probability of that one effects is effectively that size given actual data is equal to the probability of the effect given a particular model and the actual data averaged by the probability that the model is correct given the actual data”***

This will forcedly lead to a use of Bayesian transformation for averaging terms that leads to:

Where the probability depending of the model is calculated as:

This represents the integrated likelihood of model represents the vector parameters in the case of a regression will be: .

* Represents the prior density of under the model .
* is the actual likelihood.
* is the prior probability that the is the true model.

Theoretically this is a very appealing solution to the model uncertainty; sadly the actual implementation of this methodology has several difficulties as:

1. The number of models to be considered in (1) could be enormous, so the exhaustive summation could be infeasible.
2. The integrals that need to be computed are very difficult to solve numerically speaking, gladly a good Markov Chain Montecarlo engine will overcome most of the difficulties but there are still technical challenges to this.
3. The prior distributions for the models and parameters are very challenging and there is not work in there yet.
4. Choosing what class of models to average is a modeling task where there is not a final word on it yet.

### BMA implementation framework for Betasmartz