CHAPTER II

LITERATURE REVIEW

A. Literature Review

1. Indicators of Forecasting Stock Price Index

Prediction stock price index is gaining attention in various fields of forecasting. The idea is to set up a contract whose payoff depends on the outcome of an uncertain future event. This contract can be interpreted as a bet on the outcome of the underlying future event. Participants can trade it. As soon as the outcome is known, participants are paid off in exchange for the contracts they hold. Based on their individual performance, participants can win money. If one thinks the current group estimate is too low (high), one will buy (sell) stocks. Thus, through the prospect of gaining money, participants have an incentive to become active in the group process whenever they expect the group estimate to be inaccurate. There are two important indicators for forecasting stock price (Hassan, Beitollah, and Saeed, 2011). They are fundamental analysis and technical analysis.

a. Fundamental analysis

Fundamental analysis is function on historical and present data, but with the goal of making financial forecasts. It is the process of looking at a business at the basic or fundamental financial level. This type of analysis examines key ratios of a business to determine its financial health and gives us an idea of the value its stock. It uses the information in companies' financial statement (Chih Fong and Wang, 2009). There are two basic approaches one can use; bottom up analysis and top down analysis when they want to examine on stock, futures contract, or currency using fundamental analysis. Also known as quantitative analysis, this involves looking at revenue, expenses, assets, liabilities and all the other financial aspects of a company.

Mahdi, Vali, and Hakim (2011) have stated that each country has its special form related to structure of capital market and introduced models are not generalized. Researches exhibit that the market is not strong performance and price depend on each other in short term and they are independent in long term. They focus on earnings, growth, and value in the market. Nevertheless, in terms of fundamental analysis, the stock market does not have memory and prices are changed randomly. In addition, there are several possible objectives:

a) To behave a company stock valuation and forecast its probable price evolution,
b) To manufacture an estimate on its business performance,
c) To put value on its management and make internal business decisions,
d) To calculate its credit risk.
(http://en.wikipedia.org/wiki/Fundamental_analysis)

Many investors use fundamental analysis alone or in combination with other tools to evaluate stocks for investment purposes. The goal is to determine the current worth and, more importantly, how the market valucs thc stock. Moreover, Foundation analysis focuses on basic value of stock and the stock exchange value. Then scicntific tools like statistics, economy evaluation, financial management and others can find out stock exchange value. This point can determine on financial statement, companies' dividend background, management policy, sales growth, companies' profitability trend and the institution's ability of increasing the profit and net value of assets and many other factors to show the stock value. They have compare the inherent value with current stock price and introduce guidance for financial decision-making (Mahdi, Vali, and Hakim, 2011).

b. Technical analysis

Technical analysis is the forecasting of market prices by means of analysis of data generated by the process of trading. Technical analysis is the study of the movements of a stock/index or any other financial commodity for predicting the future trend direction. Chih-Fong and Wang (2009) had written that technical analysis is researching the trend in stock market will acquire the change rules of stock. Technical analysis is security analysis discipline for forecasting the direction of prices through the study of past market data, primarily price and volume. This is a study, which is based on the analysis of the current prices and volumes, comparing them to historic prices and volumes and looking for similarities that may help to define possible future trend development. Technical analysis relies on the assumption that markets discount everything except information generated by market action, ergo, all you need is data generated by market action. In other words, technical analysis attempts to understand the emotions in the market by studying the market itself, as opposed to its components. It can calculate the intrinsic value of stock (Hassan, Beitollah, and Saeed, 2011). He has noted that the main purpose of technical analysis is forecasting trends of stock price. By the way, forecasts are often not correct and have some errors that the rate decreases with increasing of information.

Sewell (2008) had public that people often estimate future uncertain events by taking a short history of data and asking what broader picture this history is representative of independent of other information about its actual likelihood. Technical analysis is representativeness. He also quotes some experiments from some psychological explanations of why a large number of people have a strong belief in technical analysis.

a) Communal reinforcement is a social construction in which a strong belief is formed when a claim is repeatedly asserted by members of a

community, rather than due to the existence of empirical evidence for the validity of the claim.

b) Selective thinking is the process by which on focuses on favorable evidence in order to justify a belief, ignoring unfavorable evidence.
c) Confirmation bias is a cognitive bias whereby one tends to notice and look for information that confirms one's existing beliefs, whilst ignoring anything that contradicts those beliefs. It is a type of selective thinking.
d) Self-deception is the process of misleading ourselves to accept as true or valid what we believe to be false or invalid by ignoring evidence of the contrary position.

Technical analysis takes a completely different approach; it does not care one bit about the "value" of a company or a commodity. Technicians (sometimes called chartists) are only interested in the price movements in the market. Mahdi, Vali, and Hakim (2011) had explained that drawing the stock price behavior graph, investigating its fluctuations, identifying its long-term behavior sensitivity and forecasting the future of stock were the main goals of this doctrine. Chartists believe that it is not possible to evaluate the inherent value of stock and the history is always repeating. The movement of a stock/index or other commodity is always described by the change in price, trading volume during this price change and change in volatility. In addition, an index movement is described by the number of advance/decline stocks and by advancing and declining volumes. Rahnamay, Falah, and Kordlouie (2011) had sought that manipulating profitability price is able eve in case of having no temporarily change of price and impossibility of cornering market. He had also found model consisting of three traders, such as mass of logical investors, a great informed trader and a great manipulation as great trader by having confidential information. This model is insufficient investor's information and asymmetric information in market is regarded as principal factor of price manipulation. It would be wrong to

state that technical analysis is an art of price analysis only and be focused only on price. By looking solely at price, a trader's vision is limited and he or she will be unable to see the basics processes that describe the cause of the price movement. That is why price, volume, volatility and advance/decline data must be analyzed in harmony (Huong, -). Only a complete analysis of the movement of a stock/index can provides the complete picture of this movement and can be considered an appropriate source for making a trading decision. Stocks with low possibility of liquidity have higher possibility of being exposed to price result in enlargement price fluctuation.

As the result, we can forecast the future trend of stock price by estimating its past trend. So this studying price history can help us to exam the price behavior in future. They can use some diagrams and curves and believe that we can never identify factors that influence supply and demand, because they are too many (Mahdi, Vali, and Hakim, 2011).

2. Forecasting Stock Price Index

We want to say some points about capital market before we can reach out to forecasting stock price index. At the basic, a capital market is efficient if it fully and correctly reflects all relevant information in determining security prices (Timmermann and Granger, 2004). The efficient capital markets hypothesis (sometimes just called the Efficient Markets Hypothesis) states that liquid markets quickly absorb information, so that it is essentially impossible for an average investor to make excess profits trading on public information. Share prices are thus the best indication of the value of a company, because they reflect the consensus view of all available information. Since there are surely positive information and trading costs, the extreme version of the market efficiency hypothesis is surely false. Its advantage, however, was that it was a clean benchmark that allows his to sidestep the messy problem of deciding what were reasonable information and trading costs (Fama, 1991). In an informationally efficient market, price changes must be unforecastable if

they properly anticipated, that is, if they full incorporate the information and expectations of all market participants. By the way, Grossman and Stiglitz (1980) went even farther – they argue that perfectly informationally efficient markets were an impossibility for, if markets were perfectly efficient, there is no profit to gathering information, in which case there would be little reason to trade and markets would eventually collapse. Alternatively, the degree of market inefficiency determines the effort investors were willing to expend to gather and trade on information. Hence, no degenerate market equilibrium will arise only when there were sufficient profit opportunities, that is, inefficiencies, to compensate investors for the costs of trading and information gathering. The profits earned by these attentive investors may be viewed as "economic rents" that accrued to those willing to engage in such activities. Black (1986) gave us a provocative answer: "noise traders," individuals who trade on what they consider to be information but which is, in fact, merely noise.

As the reason above, Timmermann and Granger (2004) had found that everyone with a new prediction method desires to try it out on return from a speculative asset, such as stock market prices, rather that series that are known to be forecastable. Forecasting experiments have to specify at least five factors, namely

a. The set of forecasting models available at any given point in time, including estimation methods;
b. The search technology used to select the best (or a combination of best) forecasting models;
c. The available real time information and ideally the cost of acquiring such information;
d. An economic model for the risk premium reflecting economic agents' trade-off between current and future payoffs;
e. The size of transaction costs and the available trading technologies and any restrictions on holdings of the asset in question.

As the following research from Fortune (1998) had said the most widely quoted stock price index has been supplemented by other popular indices that are

constructed in a different way and pose fewer problems as a measure of stock prices. Stock price indices differ according to the number and characteristics of the stocks included in the index, as well the weights given to each stock. While a stock price index is measuring the level of stock prices, its practical application is to compare values at different points in time, that is, to measure the rate of appreciation (excluding cash dividends) on common stocks. Rahnamay, Falah and Kordlouie (2011) have mentioned that stocks with highest possibility are regarded as price manipulators of stocks. Stocks with low possibility of liquidity have higher possibility of being exposed to price manipulation and price manipulation results in increasing price fluctuation. Instead of treating investor psychology as noise, we should recognize that it is actually the signal. It drives much of the day-to-day price fluctuations. Fama's (1991) seminal papers were based on his interest in measuring the statistical properties of stock prices, and in resolving the debate between technical analysis and fundamental analysis.

Therefore, we can get result from our research; we use Bayesian method to predict stock price index. This method will represent our combination schemes in terms of conditional densities and improves statistical accuracy of forecasts for the Indonesia Stock Exchange index, in particular in terms of density forecasting.

3. Empirical Application in Stock Price Index

Based on paper in Billio et al. (2011a), they had used two different models, such as a White Noise model (WN) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. Firstly, WN model had assumed and predicted log returns which normally distributed with mean and standard deviation equal to the unconditional (up to time T for forecasting at time T+1) mean and standard deviation. WN was a standard benchmark to predict stock returns because it had pointed toward a random walk assumption for price, which was difficult to defeat Welch and Goyal (2008) in contest. The last, GARCH models were often made model use financial time series that exhibited volatility clustering, i.e. periods of swings followed by

periods of relative tranquility. GARCH models were variance conditional on the past. In the classical GARCH models, the conditional variance was expressed as a linear function of the squared past values of the series. Expect to the conditional property of GARCH models, the instrument required to the observations of instant past, so it was including past variances into explanation of future variances (Mina, Mehdi and Mahendran, 2011).

In this research, we just want to forecast stock price index in IDX index. As the reason, we want to focus on GARCH models. Because we do not need to comparison between WN model and GARCH models, we try to analyze aspect of GARCH models, namely, their ability to deliver volatility forecasts. In other words, these models are useful not only for modeling the historical process of volatility but also in giving us multi-period ahead forecasts.

This observation has inspired a wide range of helpful specifications using realized volatility, power variation of several orders, bi-power variation, a jump and an asymmetric term. We focus on the benefits of Bayesian model averaging (BMA) for forecasts of daily average realize volatility. BMA combines individual model forecasts based on their predictive record. Thus, models with good predictions had received large weights in the BMA (Popava and Edward, 2008). The empirical results were showed BMA to be consistently ranked at the top among all benchmark models, including a simple equally weighted model average. Considering all data series and forecast horizons, the BMA was the dominate model. Even thought there were substantial gains in BMA base on density forecasts, point forecasts using the predictive mean had showed smaller improvements. Bollerslev et al. (2007) had documented the importance of GARCH dynamics in time series models of log-volatility. Bayesian model averaging had provided further improvements to density forecasts when we moved away from linear models and average over specifications that allowed for GARCH effects in the innovations to log-volatility (Liu and Maheu, 2008). BMA had provided an optimal way to combine this information that based on

a logarithmic scoring rule, averaging over all the models had given superior predictive ability (Raftery et al., 1997). The established volatility-volume relation motivated the use of volume as the trigger variable in our threshold GARCH model. Since volume and volatility were highly correlated, volume must be treated as an endogenous threshold variable.

a. The Theoretical Framework

We start from the linear regression model. The model does not have to be linear regression model, however; as a start, we must henceforth assume it is. Because we do not have any good reason to accept any other special model structure instead of this most general and widely used linear model.

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \tag{1}$$

where y is the vector of observations.

X is the matrix of regresses including 1 in the first column.

β is a vector of unknown parameters.

ε is an independent, identically distributed (iid).

Following research paper of Magnus, Powell and Prufer (2008), one such treatment is model averaging, where the aim of the investigator is not to find the best possible model, but rather to find the best possible estimates. Each model contributes information about the parameters of interest, and all these pieces of information are combined taking into account the trust we have in each model, based on our prior believes and on the data.

In a sense, all estimation procedures are model averaging algorithms, although possibly extreme or limiting cases. Our framework is the linear regression model

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

where y ($n \times 1$) is the vector of observations, X_1 ($n \times k_1$) and X_2 ($n \times k_2$) are matrices of non random regressors, ε is a random vector of unobservable disturbances, and β_1 and β_2 are unknown parameter vectors. We assume that $k_1 \geq 1$, $k_2 \geq 0$, $k := k_1 + k_2 \leq n-1$, that $X := (X_1 : X_2)$ has full column- rank, and that the disturbances $(\varepsilon_1, \ldots, \varepsilon_n)$ are i.i.d. $N(0 , \sigma^2)$.

The reason for distinguishing between X_1 and X_2 is that X_1 contains explanatory variables that we want in the model on theoretical or other grounds (irrespective of the found t-ratios of the β_1-parameters), while X_2 contains additional explanatory variables of which we are less certain. The columns of X_1 are called 'focus' regressors, and the columns of X_2 'auxiliary' regressors.

There are k_2 components of β_2, and a different model arises whenever a different subset of the β_2's is set equal to zero. If $k_2 = 0$, then no model selection takes place. If $k_2 = 1$, then there are two models to consider: the unrestricted and the restricted model. If $k_2 = 2$, there are four models: the unrestricted, two partially restricted (where one of the two β_2's is zero), and the restricted model. In general, there are 2^{k_2} models to consider. We denote the i-th model by $\mathcal{M}_i$, which we write as

$$y = X_1\beta_1 + X_{2i}\beta_{2i} + \varepsilon \qquad (2)$$

where X_{2i} denotes an $n \times k_{2i}$ matrix containing a subset of k_{2i} columns of X_2, and β_{2i} denotes the corresponding $k_{2i} \times 1$ sub-vector of β_2. We have $0 \leq k_{2i} \leq k_2$.

Model averaging estimation proceeds in two steps. In the first step, we ask how to estimate the parameters, conditional upon a selected model. In the second step, we compute the estimator as a weighted average of these conditional estimators. There exist both Bayesian and non- Bayesian ideas about how to estimate and how to find the weights.

b. Bayesian Methods for Combining Forecasts

When many prediction models are available, one of the challenging issues is to summarize the information on the future values of the variable. The combination of predictions represents a solution to this problem. Following Billio et al. (2011a), we will optimal combination based on the distributional representation of the predictive models. Predictive models remain underutilized, yet an increasing number of scholars have developed forecasting models for specific research domain. As the number of forecasting efforts proliferates, however, there is a growing benefit from developing methods to pool across models and methodologies to generate forecasts are more accurate. Very often, specific predictive models prove to be correct only for certain subsets of observations. Moreover, specific model tend to be more sensitive to unusual events or particular data issues than ensemble methods.

To aid the newfound emphasis on prediction political science, we are advancing recent statistical research aimed at integrating multiple predictions into a single improved forecast. In particular, we are adapting an ensemble method first developed for application to the most mature prediction models in existence – weather models. To generate predictive distribution of outcomes (e.g., temperature), weather researchers apply ensemble methods to forecasts generated from multiple models (Reftery et al., 2004). Thus, state-of-the-art ensemble forecasts aggregate multiple runs of (often multiple) weather prediction models into a single unified forecast.

1) Bayesian Model Averaging (BMA)

A more comprehensive approach to addressing model uncertainty is Bayesian model averaging (BMA), which allows us to assess the robustness of results to alternative specifications by calculating posterior distributions over coefficients and models (Montgomery and Nyhan, 2010). They had explained what BMA is particularly useful in three specific contexts that we illustrate in our empirical

example as in following. First, BMA can be helpful when a researcher wishes to assess the evidence in favor of two or more competing measures of the same theoretical concept, particularly when three is also significant uncertainty over control variables. Second, when there is uncertainty over control variables, researchers can use BMA to test the robustness of their estimates more systematically than is possible under a frequent approach. Finally, BMA may also be valuable for researchers who wish to estimate the effects of large numbers of possible predictors of a substantively important dependent variable. Even though, there are important reasons to be cautious about the conclusions one can draw from such an approach.

Let t be the time index, with $t = 1, 2, \ldots, T$, then given a sequence of vectors $\mathbf{x}_u$ with $u = s, \ldots, t$ and $s \leq t$ we denote with $\mathbf{x}_{s:t} = (\mathbf{x}_s, \ldots, \mathbf{x}_t)$ the collection of these vectors. Let $\mathbf{y}_t$ be an observable variable at time t, we are interested in predicting the future values of the variable $\mathbf{y}_t$. We denote with $\mathbf{y}_t \in \mathcal{Y} \subset \mathbb{R}^L$ the vector of observable variables, $\tilde{\mathbf{y}}_{k,t} \in \mathcal{Y} \subset \mathbb{R}^L$ the typical k-th one-step ahead predictor for $\mathbf{y}_t$, where $k = 1, 2, \ldots, K$. In particular, in a density forecasting exercise, we are interested $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$, where is the distribution of the $\mathbf{y}_t$ conditional on its past values and which is called one-step-ahead prediction density of $\mathbf{y}_t$. In many situations, there are different prediction models available for the variable $\mathbf{y}_t$. In what follows we will assume that at time t a set of K one one-step-ahead predictors $\tilde{\mathbf{y}}_{k,t}$, with $k = 1, 2, \ldots, K$, is available from different models or sources. Moreover, we assume that for each prediction model its conditional density $p(\tilde{\mathbf{y}}_{k,t}|\mathbf{y}_{1:t-1})$ is available analytically or in a approximated form (e.g. through Sequential Monte Carlo, in Appendix 1).

Assume we have some quantity of interest in the future to forecast, $\mathbf{y}_t$, based on previously collected training data $\mathbf{y}^T$ that is fit to K statistical models, $M_1, M_2, \ldots, M_K$. Each model, M_K, is assumed to come from the prior probability distribution $M_k \sim \pi(M_k)$, and the probability distribution function (PDF) for the training data is $p(y^T|M_k)$. The outcome of interest is distributed $p(\mathbf{y}_t|M_k)$. Applying Bayes Rule, we get that:

$$p(M_k|y^T) = \frac{p(y^T|M_k)\pi(M_k)}{\sum_{k=1}^{K} p(y^T|M_k)\pi(M_k)} \quad (3)$$

and the marginal predictive PDF y* is

$$p(y_t) = \sum_{k=1}^{K} p(y_t|M_k)p(M_k|y^T) \quad (4)$$

this BMA PDF (2) can be viewed as the weighted average of the component PDFs where the weight are determined each model's performance within the training data (Montgomery and Hollenbach, 2011) . Likewise, we can simply make a deterministic estimate using the weighted predictions of the components, denoted that

$$E(y_t) = \sum_{k=1}^{K} E(y_t|M_k)p(M_k|y^T) \quad (5)$$

a) Dynamic settings of BMA

We will prescnt operational Bayesian methods for combining individual forecasts, which extend results originally, put forward by bringing in bias parameters cxplicitly and treating the unknown covariance case.

Let t be the time index, with $t = 1, 2, \ldots, T$, then given a sequence of vectors $\mathbf{x}_u$ with $u = s, \ldots, t$ and $s \leq t$ we denote with $\mathbf{x}_{s:t} = (\mathbf{x}_s, \ldots, \mathbf{x}_t)$ the collection of these vectors. Let $\mathbf{y}_t$ be an observable variable at time t, we are interested in prcdicting the future values of the variable $\mathbf{y}_t$. We denote with $\mathbf{y}_t \in \mathcal{Y} \subset \mathbb{R}^L$ the vector of observable variables, $\tilde{\mathbf{y}}_{k,t} \in \mathcal{Y} \subset \mathbb{R}^L$ the typical k-th one-step ahead predictor for $\mathbf{y}_t$, where $k = 1, 2, \ldots, K$. For the sake of simplicity, we present the new combination method for the one-step ahead forecasting horizon (Billio et al., 2010). The methodology easily extends to multi-step ahead forecasting horizons.

They had assumed that the observable vector was generated from a distribution with conditional density $p(\mathbf{y}_{k,t}|\mathbf{y}_{1:t-1})$ and that for each predictor $\tilde{\mathbf{y}}_{k,t}$ there existed a predictive density $p(\tilde{\mathbf{y}}_{k,t}|\mathbf{y}_{1:t-1})$. In order to simplify the exposition, in what follows they had defined $\tilde{\mathbf{y}}_t = \text{vec}(\tilde{\mathbf{Y}}_t')$, where $\tilde{\mathbf{Y}}_t = (\tilde{\mathbf{y}}_{1,t}, \tilde{\mathbf{y}}_{2,t}, \ldots, \tilde{\mathbf{y}}_{K,t})$ is the matrix with predictors in the columns and vec is an operator that stacks the columns of a

matrix into a vector (Jordan and Jacobs, 1993, and Huerta et al., 2003). They denote with $p(\tilde{\mathbf{y}}_{k,t}|\mathbf{y}_{1:t-1})$ the joint predictive density of the set of predictors at time t and let

$$p(\tilde{y}_{1:t}|\mathbf{y}_{1:t-1}) = \prod_{s=1}^{t} p(\tilde{y}_s|\mathbf{y}_{1:s-1})$$

be the joint predictive density of the predictors up to time t.

A combination scheme of a set of predictive densities is a probabilistic relation between the density of the observable variable and a set of predictive densities. They had assume that relationship between the density of $\mathbf{y}_t$ conditionally on $\mathbf{y}_{1:t-1}$ and the set of predictive densities from the K different sources is

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int_{y^{Kt}} p(\mathbf{y}_t|\tilde{\mathbf{y}}_{1:t-1}, \mathbf{y}_{1:t-1}) p(\tilde{\mathbf{y}}_{1:t}|\mathbf{y}_{1:t-1}) d\tilde{\mathbf{y}}_{1:t} \qquad (6)$$

where the dependence structure between the observable and the predictive is not define yet. This relation might be miss-specified because the model set is incomplete or the true data generating process (DGP) is a combination of unknown and unobserved models that statistical and econometric tools can only partially approximate. In the following, in order to model the possibly miss-specified dependencc between forecasting models, they consider a parametric latent variable model. They also assume that the model is dynamic to capture the time variations in the dependence structure.

In order to define the latent variable model and the combination scheme they introduce first the latent space. Let $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$ and $\mathbf{0}_n = (0, \ldots, 0)' \in \Delta_{[0,1]^n} \subset \mathbb{R}^n$ the set of all vector $\mathbf{w} \in \mathbb{R}^n$ such that $\boldsymbol{w}'\mathbf{1}_n = 1$ and $\omega_k \geq 0$, $k = 1, 2, \ldots, n$. $\Delta_{[0,1]^n}$ is called the standard n-dimensional simplex and is the latent space used in all our combination schemes.

Secondly, they introduce the latent model that is a matrix-valued stochastic process $W_t \in \mathcal{W} \subset \mathbb{R}^L \times \mathbb{R}^{KL}$, which represents the time-varying weights of the

combination scheme. Denote with $\omega_{k,t}^{l}$ the k-th column and l-th row elements of W_t, then they had assumed that the vectors $\boldsymbol{w}_t^l = (\omega_{1,t}^l, \omega_{2,t}^l, \ldots, \omega_{KL,t}^l)'$ in the rows of W satisfy $\boldsymbol{w}_t^l \in \Delta_{[0,1]}{}^K$.

The definition of the latent space as the standard simplex and the consequent restrictions on the dynamics of the weight process allow to estimate a time series of [0, 1] weights at time $t-1$ when a forecast is made for $\mathbf{y}_t$. This latent variable modeling framework generalizes previous literature on model combination with exponential weights (see for example Hoogerheide et al., 2010) by inferring dynamics of positivc weights which belong to the simplex $\Delta_{[0,1]}{}^{LK}$. In such a way, one can interpret the weights as a discrete probability density over the set of predictors.

They assume that at time t, the time-varying weight process W_t has a distribution with density $p(\mathrm{W}_t|\mathbf{y}_{1:t-1},\tilde{\mathbf{y}}_{1:t-1})$. Then they can write from (6) as

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) = \int_{y^{Kt}} \left(\int_w p(\boldsymbol{y}_t|W_t,\tilde{\boldsymbol{y}}_t) p(W_t|\boldsymbol{y}_{1:t-1},\tilde{\boldsymbol{y}}_{1:t-1}) dW_t \right) p(\tilde{\boldsymbol{y}}_{1:t}|\boldsymbol{y}_{1:t-1}) d\tilde{\boldsymbol{y}}_{1:t} \qquad (7)$$

In the following, they had assumed that the time-varying weights have a first-order Markovian dynamics and that they may depend on the past values $\tilde{\mathbf{y}}_{1:t-1}$ of the predictors. Thus the weights at time t have $p(W_t|W_{t-1},\tilde{\mathbf{y}}_{1:t-1})$ as conditional transition density. They had usually assumed that the weight dynamics depend on the recent values of the predictors, i.e.

$$p(W_t|W_{t-1},\tilde{\boldsymbol{y}}_{1:t-1}) = p(W_t|\, W_{t-1},\tilde{\boldsymbol{y}}_{t-r:t-1}) \qquad (8)$$

with $r > 0$.

Under these assumptions, the first integral in (7) is now defined on the set $y^{K(r+1)}$ and is taken with respect to a probability measure that has $p(\tilde{\mathbf{y}}_{t-r:t}|\mathbf{y}_{1:t-1})$ as joint predictive density. Moreover, the conditional predictive density of W_t in (7) can be further decomposed as follows:

$$p(W_t|\mathbf{y}_{1:t-1},\tilde{y}_{1:t-1}) = \int_{\mathcal{W}} p(W_t|W_{t-1},\tilde{y}_{t-r:t-1})\, p(W_t|\mathbf{y}_{1:t-2},\tilde{y}_{1:t-2})dW_{t-1}$$

The above assumptions do not alter the general validity of the proposed approach for the combination of the predictive densities. In fact, the proposed combination method extends previous model pooling by assuming possibly non-Gaussian predictive densities as well as nonlinear weights dynamics that maximize general utility functions.

As a conclusion of this section we present some possible specifications of the conditional predictive density $p(\mathbf{y}_t|W_t,\tilde{\mathbf{y}}_t)$. In the next section they will consider different specifications for the weights transition density $p(W_t|W_{t-1},\tilde{\mathbf{y}}_{1:t-1})$.

Example 1: *Gaussian Combination Scheme*

The Gaussian combination model is defined by the probability density function

$$p(\mathbf{y}_t|W_t,\tilde{y}_t) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}_t - W_t\tilde{y}_t)'\,\Sigma^{-1}(\boldsymbol{y}_t - W_t\tilde{y}_t)\right\}$$

where $W_t \in \Delta_{[0,1]^L}$ is the weight matrix defined above and $\sum$ is the covariance matrix.

A special case of the previous model is given by the following specification of the weight density

$$p(\mathbf{y}_t|W_t,\tilde{y}_t) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_t - \sum_{k=1}^{K}\boldsymbol{w}_{k,t}\odot\tilde{y}_{k,t}\right)'\Sigma^{-1}\left(\boldsymbol{y}_t - \sum_{k=1}^{K}\mathbf{w}_{k,t}\odot\tilde{y}_{k,t}\right)\right\}$$

where $\boldsymbol{w}_{k,t} = (\omega_{k,t}^1, \omega_{k,t}^2, \ldots, \omega_{k,t}^L)'$ is a weights vector and $\odot$ is the Hadamard's product. The system of weights is given as $\boldsymbol{w}_t^l = (\omega_{k,t}^1, \omega_{k,t}^2, \ldots, \omega_{k,t}^l)' \in \Delta_{[0,1]^L}$, for $l = 1, 2, \ldots, L$. In this model the weights may vary over the elements of $\mathbf{y}_t$ and only the i-th element of $\mathbf{y}_t$.

A more parsimonious model than the previous one is given by:

$$p(\mathbf{y}_t|W_t,\tilde{\mathbf{y}}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}_t - \sum_{k=1}^{K}\omega_{k,t}\tilde{\mathbf{y}}_{k,t}\right)'\Sigma^{-1}\left(\mathbf{y}_t - \sum_{k=1}^{K}\omega_{k,t}\tilde{\mathbf{y}}_{k,t}\right)\right\} \quad (9)$$

where $\boldsymbol{w}_t = (\omega_{1,t}, \omega_{2,t}, \ldots, \omega_{K,t})' \in \Delta_{[0,1]}{}^{K}$. In this model all the elements of the prediction $y_{k,t}$ given by k-th model have the same weight, while the weights may vary across the models.

As an alternative to the Gaussian distribution, heavy-tailed distributions could be used to account for extreme values, which are not captured, by the pool of predictive densities.

Example 2: *Student-t combination scheme*

In this scheme the conditional density of the observable is

$$p(\mathbf{y}_t|W_t,\tilde{\mathbf{y}}_t) \propto \left(1 + \frac{1}{v}(\mathbf{y}_t - W_t\tilde{\mathbf{y}}_t)'\Sigma^{-1}(\mathbf{y}_t - W_t\tilde{\mathbf{y}}_t)\right)^{-\frac{v+L}{2}} \quad (10)$$

Where $\sum$ is the precision matrix and $v > 2$ is the degrees-of-freedom parameter. The scheme could be extended to asymmetric Student-t as in Li et al. (2010).

Example 3: *Mixture of experts*

Similarly to Jordan and Jacobs (1993) and Huerta et al. (2003), the density of the observable is

$$p(\mathbf{y}_t|\tilde{\mathbf{y}}_t) = \sum_{k=1}^{K} p(W_{k,t}|\mathbf{y}_{1:t-1},\tilde{\mathbf{y}}_{1:t-1})p(\tilde{\mathbf{y}}_{k,t}) \quad (11)$$

where $p(W_t|\mathbf{y}_{1:t-1},\tilde{\mathbf{y}}_{1:t-1})$ is the mixture weight associated to model k, which might be specified similarly to forms.

Following Billio et al. (2011a) had suggested to summarize the information from the different predictive densities in one prediction density for $\mathbf{y}_t$ by conditioning on $\mathbf{y}_t = (\tilde{\mathbf{y}}_{1,t}, \tilde{\mathbf{y}}_{2,t}, \ldots, \tilde{\mathbf{y}}_{K,t})$ and on a combination scheme (10) $\boldsymbol{w}_t = (w_{1,t}, w_{2,t}, \ldots, w_{K,t})$

$$p(\mathbf{y}_t|\boldsymbol{w}_t,\tilde{\boldsymbol{y}}_t) \propto \exp\left\{-(\tfrac{1}{2}\sigma^2)(\mathbf{y}_t - \mathrm{w}_t'\tilde{\boldsymbol{y}}_t)^2\right\} \quad (12)$$

which corresponds to a Gaussian combination where $\mathbf{w}_t$ are the weights:

$$\omega_{k,t} = \frac{\exp\{x_{k,t}\}}{1+\sum_{j=1}^{K-1}\exp\{x_{k,t}\}}, \text{ where } k = 1, 2, \ldots, K-1 \quad (13)$$

$$\omega_{K,t} = 1 - \sum_{j=1}^{K}\exp\{x_{k,t}\} \quad (14)$$

The weights are thus multivariate logistic transformations of a latent process $\mathbf{x}_t$. The transformation allows for positive weights that sum to one and accordingly can be interpreted as the probability associated to a specific prediction model. In this work we assume that the latent factor has the following Gaussian dynamics

$$p(\mathbf{x}_t|\boldsymbol{x}_{t-1},\tilde{\boldsymbol{y}}_{1:t-1}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}_t - \mathbf{x}_{t-1} + \Delta\mathrm{e}_t)'\Lambda^{-1}(\mathbf{x}_t - \mathbf{x}_t + \Delta\mathrm{e}_t\right\} \quad (15)$$

with exogenous variable $\Delta\mathrm{e}_t = \mathrm{e}_t - \mathrm{e}_{t-1}$, where $\mathrm{e}_t = (\mathrm{e}_{1,t}, \mathrm{e}_{2,t}, \ldots, \mathrm{e}_{K,t}$ is a vector of exponentially weighted average errors

$$\mathrm{e}_{k,t} = (1-\lambda)\sum_{i=1}^{r}\lambda^{i-1}(\boldsymbol{y}_{t-i} - \hat{\mathrm{y}}_{k,t-i})^2 \quad (16)$$

with $\lambda \in (0,1)$ being a smoothing parameter and r the size of the window of evaluation of past errors. The past forecasting performance of the predictors is thus included in the weights dynamics. A deterioration of the forecasting performance of the *k*-th prediction model (i.e. $\Delta\mathrm{e}_{k,t} > 0$) reduces its weight in the combination (i.e. $\omega_{k,t}$ decreases). As opposite, an improvement in the prediction performance (i.e. $\Delta\mathrm{e}_{k,t} < 0$) increases the value of the *k*-th weight. This simple method has been innovatively maintained by Aiolfi and Timmermann (2004), and Eklud, Kapetanios and Price (2010) without pattern components.

This combination scheme explains a general relationship between observable, model-specific predictive densities, combination weights and the predictive density for $\boldsymbol{y}_t$

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) = \iint p(\mathbf{y}_t|\boldsymbol{w}_t,\tilde{\boldsymbol{y}}_t)p(\boldsymbol{w}_t|\boldsymbol{y}_{1:t-1},\tilde{\boldsymbol{y}}_{1:t-1})p(\tilde{\boldsymbol{y}}_{1:t}|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{w}_t d\tilde{\boldsymbol{y}}_{1:t}$$

This relationship for the prediction of the observable variable y_t is part of a general filtering and prediction problem which can be explained conditionally on $\tilde{\boldsymbol{y}}_{1:t}$ through the following set of recursions

$$p(\boldsymbol{w}_t|\boldsymbol{y}_{1:t}\tilde{\boldsymbol{y}}_{1:t}) \propto p(\mathbf{y}_t|\boldsymbol{w}_t,\tilde{\boldsymbol{y}}_t)p(\boldsymbol{w}_t|\boldsymbol{w}_{t-1},\tilde{\boldsymbol{y}}_{t-r:t-1})p(\boldsymbol{w}_{t-1}|\boldsymbol{y}_{1:t-2},\tilde{\boldsymbol{y}}_{1:t-2})$$

$$p(\boldsymbol{y}_t|\tilde{\boldsymbol{y}}_{1:t},\boldsymbol{y}_{1:t-1}) = \int p(\mathbf{y}_t|\boldsymbol{w}_t,\tilde{\boldsymbol{y}}_t)p(\boldsymbol{w}_t|\boldsymbol{y}_{1:t-1},\tilde{\boldsymbol{y}}_{1:t-1})\, d\boldsymbol{w}_t$$

$$p(\boldsymbol{w}_t|\boldsymbol{y}_{1:t-1},\tilde{\boldsymbol{y}}_{1:t-1}) = \int p(\mathbf{w}_t|\boldsymbol{w}_{t-1},\tilde{\boldsymbol{y}}_{t-r:t-1})p(\boldsymbol{w}_{t-1}|\boldsymbol{y}_{1:t-2},\tilde{\boldsymbol{y}}_{1:t-2})\, d\boldsymbol{w}_{t-1}$$

b) Non-linear Filtering and Prediction

The density of the observable variable conditional on the combination scheme and on the predictions, and the density of the weights of the scheme conditional on the prediction errors represent a nonlinear and possibly non-Gaussian state-space model. As the following Billio et al. (2010) they had consider a general state space explanation and show how Sequential Monte Carlo methods can be used to approximate the filtering and predictive densities.

Let $\Psi_t = \sigma(\{\boldsymbol{y}_s\}_{s\leq t})$ be the σ–algebra generated by the observable process and assume that the predictors, $\tilde{\boldsymbol{y}}_t = (\tilde{\boldsymbol{y}}'_{1,t},\tilde{\boldsymbol{y}}'_{2,t},\dots,\tilde{\boldsymbol{y}}'_{K,t})' \in Y \in \mathbb{R}^{KL}$ stand from a Ψ_{t-1} measurable stochastic process associated with the predictive densities of the K different models in the pool. Let $\boldsymbol{w}_t = (\boldsymbol{w}'_{1,t},\boldsymbol{w}'_{2,t},\dots,\boldsymbol{w}'_{K,t})' \in X \in \mathbb{R}^{KL}$ be the vector of latent variables associated with $\tilde{\mathbf{y}}_t$ and $\theta \in \Theta$ the parameter vector of the optimal predictive model. Let us include the parameter vector into the state vector and thus define the augmented state vector $\mathbf{z}_t = (\mathbf{w}_t, \theta) \in Y\times\Theta$. The distributional state space form of the optimal forecast model is

$$\boldsymbol{y}_t|\boldsymbol{z}_t,\tilde{\boldsymbol{y}}_t \sim p(\boldsymbol{y}_t|\boldsymbol{z}_t,\tilde{\boldsymbol{y}}_t)$$

$$\boldsymbol{z}_t|\boldsymbol{z}_{t-1} \sim p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},\tilde{\boldsymbol{y}}_{1:t-1})$$

$$\boldsymbol{z}_0 \sim p(\boldsymbol{z}_0)$$

The hidden state predictive and filtering densities conditional on the predictive variables $\tilde{\mathbf{y}}_{1:t}$ are

$$p(\mathbf{z}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{1:t}) = \int_X p(\mathbf{z}_{t+1}|\boldsymbol{z}_t,\tilde{\boldsymbol{y}}_{1:t})p(\boldsymbol{z}_t|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{1:t})\, d\boldsymbol{z}_t$$

$$p(\mathbf{z}_{t+1}|\boldsymbol{y}_{1:t+1},\tilde{\boldsymbol{y}}_{1:t+1}) \propto p(\boldsymbol{y}_{t+1}|\boldsymbol{z}_{t+1},\tilde{\boldsymbol{y}}_{t+1})p(\boldsymbol{z}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{1:t})$$

A major element of interest is the marginal predictive density of the observable variable

$$p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t}) = \int_{X\times Y^{t+1}} p(\mathbf{y}_{t+1}|\boldsymbol{z}_{t+1},\tilde{\boldsymbol{y}}_{t+1})p(\boldsymbol{z}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{1:t})p(\tilde{\boldsymbol{y}}_{1:t+1}|\boldsymbol{y}_{1:t})\, d\boldsymbol{z}_{t+1}d\tilde{\boldsymbol{y}}_{1:t+1}$$

$$= \int_Y p(\mathbf{y}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{t+1})p(\tilde{\boldsymbol{y}}_{t+1}|\boldsymbol{y}_{1:t})\, d\tilde{\boldsymbol{y}}_{1:t+1}$$

where

$$p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{t+1}) = \int_{X\times Y^t} p(\mathbf{y}_{t+1}|\boldsymbol{z}_{t+1},\tilde{\boldsymbol{y}}_{t+1})p(\boldsymbol{z}_{t+1}|\boldsymbol{y}_{1:t},\tilde{\boldsymbol{y}}_{1:t})p(\tilde{\boldsymbol{y}}_{1:t}|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{z}_{t+1}d\tilde{\boldsymbol{y}}_{1:t}$$

is the conditional predictive density of the observable given the predicted variables.

An analytical solution of the previous filtering and prediction problems is not known for the non-linear models presented in the previous sections, thus we apply a numerical approximation method. More specifically, we consider a Sequential Monte Carlo (SMC) approach to filtering. Let $\Xi_t = \{\mathbf{z}_t^i, \omega_t^i\}_{i=1}^N$ be a set of particles, then the basic SMC algorithm uses the particles set to approximate the prediction and filtering densities with empirical prediction and filtering densities, which are defined as

$$p_N(\mathbf{z}_{t+1}|\mathbf{y}_{1:t},\tilde{\mathbf{y}}_{1:t}) = \sum_{i=1}^N p(\mathbf{z}_{t+1}|\mathbf{z}_t,\tilde{\mathbf{y}}_{1:t})\omega_t^i\delta_{\mathbf{z}_t^i}(\mathbf{z}_t)$$

$$p_N(\mathbf{z}_{t+1}|\mathbf{y}_{1:t+1}, \tilde{\mathbf{y}}_{1:t+1}) = \sum_{i=1}^{N} \omega_{t+1}^{i} \delta_{\mathbf{z}_{t+1}^{i}}(\mathbf{z}_{t+1})$$

respectively, where $\omega_{t+1}^{i} \propto \omega_t^i \, p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}, \tilde{\mathbf{y}}_{t+1})$. The hidden state predictive density can be used to approximate the observable prediction density as follows

$$p_N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t+1}) = \sum_{i=0}^{N} \omega_t^i \delta_{\mathbf{y}_{t+1}^{i}}(\mathbf{y}_{t+1})$$

where $\mathbf{y}_{t+1}^{i}$ has been simulated from the measurement density $p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}^{i}, \tilde{\mathbf{y}}_{t+1}, \boldsymbol{\theta})$.

In our applications, we assume that the densities $p(\tilde{\mathbf{y}}_s|\boldsymbol{y}_{1:s-1})$ are discrete

$$p(\tilde{\mathbf{y}}_s|\mathbf{y}_{1:s-1}) = \sum_{i=0}^{N} \delta_{\{\tilde{\mathbf{y}}_t^i\}}(\mathbf{y}_s)$$

where $\delta_{\{x\}}(y)$ denotes the Dirac mass centered at x.

This assumption does not alter the validity of our approach and is mainly motivated by the forecasting practice, see literature on model pooling. In fact, the prediction usually comes from different models or sources. In some cases, the discrete prediction density is the result of a collection of point forecasts from many subjects, such as surveys forecasts. In other cases the discrete predictive is a result of a Monte Carlo approximation of the predictive density (for example, Importance sampling or Markov-Chain Monte Carlo approximations).

Under this assumption, it is possible to approximate the marginal predictive density by the following steps. First, draw j independent values $\mathbf{z}_{1:t+1}^{j}$, with j = 1, 2, …, M form the sequence of predictive densities $p(\tilde{\mathbf{y}}_{s+1}|\mathbf{y}_{1:s})$, with s = 1, 2, …, t. Secondly, apply the SMC algorithm, conditionally on $\tilde{\mathbf{y}}_{1:t+1}^{j}$, in order to generate the particle set $\Xi_t^{i,j} = \{\mathbf{z}_{1:t}^{i,j}, \omega_t^{i,j}\}_{i=1}^{N}$, with j = 1, 2, …, M. At the last step, simulate $\boldsymbol{y}_{t+1}^{i,j}$ from $p(\boldsymbol{y}_{t+1}|\mathbf{z}_{t+1}^{i,j}, \tilde{\mathbf{y}}_{t+1}^{j})$ and obtain the following empirical predictive density

$$p_{N,M}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{i=1}^{N}\omega_t^{i,j}\delta_{\mathbf{y}_{t+1}^{i,j}}(\mathbf{y}_{t+1})$$

2) Application of BMA

The main issue of BMA is how to choose the correct model given the model uncertainty. Contrast to traditional econometrics, which select only one single model for forecasting, BMA stands one step back to consider the tremendous uncertainty the researcher has about the correct model. Equation (4) has assumed relationship between the density of y_t conditionally on $y_{1:t-1}$. Then, the BMA forecast is just the weighted (according to the model posterior probability) average of forecasts from each possible model. More specifically, the procedure of calculating is as follows:

a) Function of Likelihood

Assume that ε has a multivariate Normal distribution with mean θ_n and the covariance matrix $\sigma^2 I_n$. For any realized data, using the definition of the Normal density, we obtain the likelihood of the data under model $y_{1:t-1}$:

$$p(\mathbf{y}_t|\beta_1, \beta_{2t}, \sigma^2, \mathbf{y}_{1:t-1}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n/2} exp\left(-\frac{(\mathbf{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})'(\mathbf{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})}{2\sigma^2}\right)$$

where $p(\mathbf{y}_t|\beta_1, \beta_{2t}, \sigma^2, \mathbf{y}_{1:t-1})$ is the likelihood function of data conditional on the parameters and model $y_{1:t-1}$.

(1) The Prior

Priors are meant to reflect any information the researcher has before seeing the data which she wishes to include and therefore can take any form. By the way, in order to interpret and make computation easier, natural conjugate prior Normal-Gamma distribution is widely chosen by researchers because when combined with the likelihood, it yields a posterior that falls in the same class of distributions. We can get

$$p(\sigma^2|\boldsymbol{y}_{1:t-1}) \propto \frac{1}{\sigma^2}$$

$$p(\beta_1|\sigma^2, \boldsymbol{y}_{1:t-1}) \propto 1$$

$$\beta_{2t}|\beta_1, \sigma^2, \boldsymbol{y}_{1:t-1} \sim N(0, \sigma^2 \boldsymbol{w}_t)$$

It remains to choose w_t. We use the so-called g-prior, which was show in Feernandez, Ley and Steel (2001), as below:

$$g = \begin{cases} 1/k_2^2 \ if \ n \leq k_2^2 \\ 1/n \ if \ n > k_2^2 \end{cases}$$

and $$\boldsymbol{w}_t = (gX_t'\boldsymbol{X}_t)^{-1}$$

Then, the joint prior distribution is:

$$p(\beta_1, \beta_{2t}, \sigma^2|\boldsymbol{y}_{1:t-1}) \propto (\sigma^2)^{-(k_{2t}-2)/2} exp\left(-\frac{\beta_{2t}'\boldsymbol{w}_t\beta_{2t}}{2\sigma^2}\right)$$

(2) The Prior

Combining the prior with the likelihood gives the posterior

$$p(\beta_1, \beta_{2t}, \sigma^2|\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1})$$

$$\propto (\sigma^2)^{-(n+k_{2t}-2)/2} exp\left(-\frac{\beta_{2t}'\boldsymbol{w}_t\beta_{2t} + (\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})'(\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})}{2\sigma^2}\right)$$

It can be established that

$$(\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_2)'(\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_2)$$
$$= [(\boldsymbol{y}_t - X_1b_1 - X_{2t}b_{2t}) - (X_1(\beta_1 - b_1) - X_{2t}(\beta_{2i} - b_{2i}))]'[(\boldsymbol{y}_t$$
$$- X_1b_1 - X_{2t}b_{2t}) - (X_1(\beta_1 - b_1) - X_{2t}(\beta_{2t} - b_{2t}))]$$

b_1 and b_{2t} are the estimator of β_1 and β_{2t} respectively.

In addition, if we let

$$M^* = I_n - X_1(X_1'X_1)^{-1}X_1'$$

and $V_{2t}^{-1} = \boldsymbol{w}_t^{-1} + X_{2t}' M^* X_{2t}$

A little algebra gives

$$V_t = \begin{pmatrix} (X_1'X_1)^{-1} + (X_1'X_1)^{-1}X_1'X_{2t} & -(X_1'X_1)^{-1}X_1'X_{2t}V_{2t} \\ -V_{2t}X_{2t}'X_1(X_1'X_1)^{-1} & V_{2t} \end{pmatrix}$$

We can make as follow:

$$\boldsymbol{p}(\beta_1, \beta_{2t}, \sigma^2 | \boldsymbol{y}_t, \boldsymbol{y}_{1:t-1}) \propto (\sigma^2)^{-(n+k_{2t}-2)/2} exp\left(-\frac{\varphi_t + \tau_t}{2\sigma^2}\right)$$

where $$\varphi_t = \begin{pmatrix} \beta_1 - b_1 \\ \beta_{2t} - b_{2t} \end{pmatrix}' V_t^{-1} \begin{pmatrix} \beta_1 - b_1 \\ \beta_{2t} - b_{2t} \end{pmatrix}$$

$$V_t^{-1} = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_{2t}'X_1 & X_{2t}'X_{2t} + \boldsymbol{w}_t^{-1} \end{pmatrix}$$

$$V_{2t}^{-1} = \boldsymbol{y}_{1:t-1}' + X_{2t}' M^* X_{2t}$$

$$\tau_t = \boldsymbol{y}_t'\boldsymbol{y}_t - \boldsymbol{y}_t'(X_1 X_{2t}) V_t (X_1 X_{2t})' \boldsymbol{y}_t$$

$$= \boldsymbol{y}_t'(M^* - M^* X_{2t} V_{2t} X_{2t}' M^*) \boldsymbol{y}_t$$

Hence, the posterior density of the parameters is the familiar normal-inverse-gamma distribution.

(3) The Marginal Likelihood of Model $y_{1:t-1}$

From the marginal density of $\boldsymbol{y}_t$ in model $\boldsymbol{y}_{1:t-1}$ as

$$p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}) = \iiint p(\boldsymbol{y}_t | \beta_1, \beta_{2t}, \sigma^2, \boldsymbol{y}_{1:t-1}) p(\beta_1, \beta_{2t}, \sigma^2 | \boldsymbol{y}_{1:t-1}) d\beta_1 d\beta_{2t} d\sigma^2$$

$$= \iiint \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\frac{n}{2}} exp\left(\frac{(\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})'(\boldsymbol{y}_t - X_1\beta_1 - X_{2t}\beta_{2t})}{2\sigma^2}\right)$$

$$* (\sigma^2)^{-(n+k_{2t}+2)/2} exp\left(-\frac{\varphi_i + \tau_t}{2\sigma^2}\right) d\beta_1 d\beta_{2t} d\sigma^2$$

$$= c\frac{|w_t^{-1}|^{1/2}}{|V_{2t}^{-1}|^{1/2}}\tau_t^{-(n-k_1/2}$$

where $$c = \frac{\pi^{\frac{n}{2}}\Gamma(\frac{n-k_1}{2})}{|X_1'X_1|\Gamma(-\frac{k_1}{2})}$$

Then, with the expression of τ_t , V_{2t} and some simplification, the marginal density can be written as:

$$p(y_t|y_{1:t-1}) = c\,\frac{|w_t'|^{1/2}}{|w_t^{-1} + X_{2t}'M^*X_{2t}|^{1/2}}[y_t'M^*(M^* - M^*X_{2t}(w_t^{-1} + X_{2t}'M^*X_{2t})^{-1}X_{2t}'M^*)M^*y_t]^{-(n-k_1)/2}$$

To simplify, let $S^* = (M^* - M^*X_{2t}(w_t^{-1} + X_{2t}'M^*X_{2t})^{-1}X_{2t}'M^*)$

The marginal density can be written as:

$$p(y_t|y_{1:t-1}) = c\,\frac{|w_t^{-1}|^{1/2}}{|w_t^{-1}+X_{2t}'M^*X_{2t}|^{1/2}}(y_t'M^*S^*M^*y_t)^{-(n-k_1)/2}$$

In we let $p(y_{1:t-1})$ denote the prior probability that $y_{1:t-1}$ is the true model, and $p(y_{1:t-1}|y_t)$ is the posterior probability for model $y_{1:t-1}$, then

$$p(y_{1:t-1}|y_t) = \frac{p(y_t|y_{1:t-1})p(y_{1:t-1})}{\sum_{j=1}^{k_2} p(y_t|y_{1:j-1})p(y_{1:j-1})} \qquad (17)$$

We now extend BMA from statistical models to dynamical models. The basic idea is that for any given forecast ensemble there is a "best" model, or member, but we do not know what it is, and our uncertainty about the best member is quantified by BMA. One thing left here is the prior probability for model $y_{1:t-1}$. Many researchers feel that simpler models should be preferred to more ones that are complex. A hierarchical structure is for the model proxies for the model prior. The agreement on kind of prior to choose is usually not within reach (Eicher, Papageorgiou and Roehn, 2007). Following Cremers (2002), we can use a flexible prior probability for an individual model as below:

$$p(\boldsymbol{y}_{1:t-1}) = \rho^{k_{2t}}(1-\rho)^{k_2-k_{2t}}$$

With certain value of the parameter ρ, the formula may represent any case of the prior probability. For example, the choice of ρ = 0.5 would assign equal prior probability to all models considered, while if ρ is less than 0.5, model including less explanatory variables is more likely than a model including more variables.

Each model also implies a forecast. In the presence of model uncertainty, the BMA forecast weights each of the individual forecasts by their respective posterior probabilities. We get equation (17) to denote that $\boldsymbol{y}_{1:t-1}$ is parameterized by θ_t. This research has prior belief about the probability that the t-th model is true, denote by $p(\boldsymbol{y}_{1:t-1})$ observes data y_t. It updates her belief to compute the posterior probability that the t-th model is the true model according to

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\theta_t, \boldsymbol{y}_{1:t-1})p(\theta_t|\boldsymbol{y}_{1:t-1})d\theta_t$$

is the marginal likelihood of the t-th model; $p(\theta_t|\boldsymbol{y}_{1:t-1})$ is the prior density of the parameter vector θ_t associated with the t-th model; and $p(\boldsymbol{y}_t|\theta_t, \boldsymbol{y}_{1:t-1})$ is the likelihood function.

To operationalize a BMA forecasting scheme, this research needs only to specify the set of models, the model priors $p(\boldsymbol{y}_{1:t-1})$, and the parameter prior $p(\theta_t|\boldsymbol{y}_{1:t-1})$. Following Faust et al. (2011), we go through a growing literature that considers a large set of very simple models. The models are especially all linear regression models, with each model adding a single regressor to the baseline specification. More formally, the t-th model is given by

$$y_{i+h} = \beta_t X_{ti} + \gamma' Z_i + \varepsilon_{t+h} \qquad (18)$$

where y_i is the variable that this research wishes to forecast at a horizon of h periods, X_{ti} is the predictor specific to model t; Z_i is a ($\rho \times 1$)-vector of predictors that are common to all models; and $\epsilon_{t+h} \sim N(0, \sigma^2)$ is the forecast error (iid). Without loss of

generality, the model-specific predictor X_{ti} is assumed orthogonal to the common predictors Z_i. In our setup, the vector of parameters characterizing the t-th model is thus given by $\theta_t = (\beta_t \gamma' \sigma^2)'$.

In setting the model priors, we suppose that all models are equally likely to involve $p(\boldsymbol{y}_{1:t-1}) = 1/n$. For the parameter prior, we follow general trend to the BMA literature of Fernandez, Ley and Steel (2001) in specifying that the prior for γ and σ^2, denoted by $p(\gamma,\sigma)$, is uninformative and is proportional to $1/\sigma$, while using the g-prior specification of Zhang, Jordan and Yeung (2008)for β_t conditional on σ^2. The g-prior is given by $N(0, \phi\sigma^2(X_i'X_i)^{-1})$, where the shrinkage hyperparameter $\phi > 0$ measures the strength of the prior – a smaller value of ϕ corresponds to a more dogmatic prior.

Letting $\hat{\beta}_t$ and $\hat{\gamma}$ denote the ordinary least square (OLS) estimates of the corresponding parameters in equation (17), the Bayesian h-period-ahead forecast made from model $\boldsymbol{y}_{1:t-1}$ at time T is given by

$$\tilde{y}_{T+h|T}^{t} = \tilde{\beta}_t X_{ti} + \hat{\gamma}' Z_i \,, \tag{18}$$

where $\tilde{\beta}_t = \left(\frac{\phi}{\phi+1}\right)\hat{\beta}_t$ denotes the posterior mean of β_t. In our framework, the marginal likelihood of the t-th model reduces to

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) \propto \left[\frac{1}{1+\phi}\right]^{-\frac{1}{2}} \times \left[\frac{\phi}{1+\phi} SSR_t + \frac{\phi}{1+\phi} SSE_t\right]^{-\frac{(T-\rho)}{2}} \tag{19}$$

where SSR_t is the sum of squares from the t-th the regression and SSE_t is the associated sum of squared errors. The posterior probabilities of the models can then be worked out from equation (17), and the final BMA forecast that takes into account model uncertainty is given by

$$\tilde{y}_{T+h|T} = \sum_{t=1}^{n} p(\mathbf{y}_{1:t-1}|\boldsymbol{y}_t)\, \tilde{y}_{T+h|T}^{t} \tag{20}$$

Clearly, the BMA forecast in equation (20) will depend on the value of the shrinkage hyperparameter ϕ. a small value of ϕ implies that model averaging (Faust et al., 2011). In contrast, a high value of ϕ amounts to weighting the models by their in-sample R^2 values.

We apply BMA to forecasting various indictors of stock price index using in IDX index as predictor. The common predictor Z_i in the predictive regression (18) is a constant and lags of the dependent variable. It is worth emphasizing that we view the forecasting scheme proposed above as a pragmatic approach to database weighting of models and make no claim to its Bayesian optimality properties. Several of the conditions for strict optimality are not met in typical macro time-series applications. First, the regressors are assumed to be strictly exogenous. The second, the forecasts are overlapping h-step ahead forecasts, so the forecast errors less than h periods apart are bound to be serially correlated even though it is assumed that they are independent, identically distributed (iid) normal. Nevertheless, BMA is like other methods that combine a large number of predictors to generate a forecast. It may still have good forecasting properties, even if the premises underlying their theoretical justification are false (e.g. Stock and Watson, 2005). This can viewed as a deterministic forecast in its own right and can be compared with the individual forecasts in the ensemble, or with ensemble mean.

3) BMA estimation of GARCH models

BMA had made additional to concentrate forecasts when we changed from linear models and average over specifications that had agreed for GARCH effects in the originations to log-volatility (Liu and Maheu, 2008). BMA had given an optimal way to combine this information that based on logarithmic scoring rule, averaging over all the models had presented superior predictive ability (Raftery et al., 1997).

Specifically, let y_T denote the forecast stock price index in the variable at time t. The average value of y_T over forecast horizon h is denoted by $y_{T+h}^c = \frac{1}{h+1}\sum_{i=0}^{h} y_{T+i}$. the t-th forecasting model in our step is given by:

$$y_{T+h}^c = \alpha + \beta_t x_{ti} + \sum_{j=0}^{p} \gamma_j y_{T+j} + \varepsilon_{T+h} \quad (21)$$

where x_{ti} is one predictor, and p is the number of lags, is determined by the Bayes Information Criterion (BIC).

The timing, convention in the forecasting regression (21) is as following. We think of forecasts as being made in daily. All stock prices are occurred during the close of trading and daily trading. With these fully real-time data in hand, we use BMA to construct forecasts of the values of the dependent variable for current and period t. Thus, we are considering prediction at horizons up to one year ahead. An important issue in this type of real-time forecasting exercise is the definition of what constitutes the actual values with comparing the BMA forecast. The accuracy of the BMA forecasts is evaluated by comparing the mean-square prediction error (MSPE) of the BMA forecast to that obtained from a univariate autoregression (AR):

$$y_{T+h}^c = \alpha_0 + \sum_{j=0}^{p} \gamma_j y_{T+j} + \varepsilon_{T+h} \quad (22)$$

Where y_{T+h}^c is stock price index at time t; α_0 is intercept. $\sum_{j=0}^{p} y_{T+j}$ is stock price index at time $j = 0, 1, \ldots, p$; $\sum_{j=0}^{p} \gamma_j$ is coefficient of stock price index at time $j = 0, 1, \ldots, p$, and ε_{T+h} is stochastic error term. This is direction autoregression that projects y_{T+h}^c onto p lags of y_T. An alternative would be to estimate an AR(p) model for y_T and then iterate it forward to construct the forecasts. This approach yielded very similar results.

The volatility σ_t of a stock is a measure of uncertainty the returns provided by stock. It is often referred to standard deviation σ_t or variance σ_t^2 in financial market. In financial modeling, volatility is a forward-looking concept (Rachev et al.,

2008). Let Γ_{t-1} is the set of information available up to time $t-1$. The volatility at time t is given by

$$\sigma^2_{t|t-1} = \text{var}(x_t|\Gamma_{t-1}) = \text{E}\left(\left(x_t - \mu_{t|t-1}\right)^2 \middle| \Gamma_{t-1}\right) = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \mu)^2 \quad (23)$$

Where x_t is continuous returns and $\mu_{t|t-1}$ is the mean return, conditional expected return at time t.

Basing on paper of Engle (1982), AutoRegressive Conditional Heteroscedasticity (ARCH) models were accustomed to describe and model detected time series. They are used whenever there is reason to believe that, at any point in a series, the terms will have a characteristic size, or variance. In particular, ARCH models assume the variance of the current error term or innovation to be a function of the actual sizes of the previous time periods' error terms: often the variance is related to the squares of the previous innovations. Following research paper of Bollerslev, Engle and Nelson (1994), ARCH models were employed commonly in modeling financial time series that exhibit time-varying volatility clustering, i.e. periods of swings followed by periods of relative calm.

Let ε_t denote the error terms (return residuals, with respect to a mean process) i.e. the series terms. These ε_t are split into a stochastic piece z_t and a time-dependent standard deviation σ_t characterizing the typical size of the terms so that

$$\varepsilon_t = \sigma_t z_t \,;\; z_t \sim \mathcal{N}(0,1)$$

where z_t is a random variable drawn from a Gaussian distribution centered at 0 with standard deviation equal to 1. (i.e. $z_t \sim \mathcal{N}(0,1)$, idd) and where the series σ_t^2 are modeled by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad (24)$$

Where $\alpha_0 > 0$, $\alpha_i > 0$ and $i > 0$.

An ARCH(q) model can be estimated using ordinary least squares. A methodology to test for the lag length of ARCH errors using the Lagrange multiplier test was proposed by Engle (1982) and Rachev et al. (2008).

This procedure is as follows:

a) Estimate the best fitting autoregressive model AR(q) in equation (22)

b) Obtain the squares of the error $\tilde{\varepsilon}_t^2$ and regress them on a constant and q lagged values:

$$\tilde{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \tilde{\varepsilon}_{t-i}^2 \qquad (25)$$

where q is the length of ARCH lags.

We will be required to generate a forecast daily and on a continuous basis. Although there has been a huge number of studies focusing on modeling stock price volatility by GARCH models, the emerging capital markets has been paid little attention, comparable to mature capital counterparts. We are not in a position to select each day the most appropriate GARCH configuration, we, therefore, propose to adopt a "one size fits all" and will select one GARCH configuration to make all of the forecasts. Liu and Maheu (2008) had sought evidence of Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) dynamics in time-series models of log-volatility. We set all prior in the regression equation as before, they are independent normal $\mathcal{N}(0,1)$. These priors are uninformative.

Extending the framework of Engle (1982) and Rache et al. (2008) generalized the ARCH(q) model to GARCH(p,q) in which they added the q lags of past conditional variance into the equation. The GARCH(p,q) as parameters have formulation as below:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 , \qquad (26)$$

where σ_t^2 is The time-conditional variance at time t, α_0 is represents the argument. α_i is represented the argument and a parameter indicates the contributions to conditional variance of the most recent news at $i = 1, 2, 3, \ldots, p$. $\sum_{i=1}^{q} \varepsilon_t^2$ is a vector of residuals or innovations of an econometric model error at $t = 1$ until $t = p$. β_j is imposed to ensure that the conditional variance is strictly positive and parameter corresponds to the moving average part in the conditional variance at $j = 1, 2, 3, \ldots, q$, and $\sum_{j=1}^{p} \sigma_{t-j}^2$ is sigma lag conditional variance at $t = 1$ until $t = q$.

GARCH model has been the most popular volatility model (Bollerslev, 1986). It has three main problems. Firstly, non-negativity constraint may violate to estimated models. Secondly, GARCH model does not take into account the leverage effect and not allow for feedback between the conditional variance and conditional mean (Vrontos, Dellaportas and Politis, 2003). Since the GARCH model was developed, a huge number of extension models have been proposed as the awareness of GARCH model's weaknesses. The differences among these models are the manner under which evolves overtime. The recent GARCH models try to correct the disadvantages of the previous models for their inefficiency to capture the volatility behaviors. One of the weaknesses of GARCH models is that they enforce symmetric effect of positive and negative shocks on volatility (Naser et al., 2011). However, the negative asset return changes in financial data are argued to impact more significantly on volatility than positive shocks of the same size. In other words, large stock price decreases predict greater volatility than similarly large price increases.

4) Stochastic Volatility Models

The alternative approach to describe the volatility process of a financial time series is called stochastic volatility (SV) models, which we want to concentrate on ARCH(q) models. An innovation term is introduced to conditional variance equation of σ_t^2 in the SV model framework. The equation of σ_t^2 including a second error term

is the main difference between ARCH(q) models and SV models. The conditional variance is completely determined given all the available information, whereas that of SV models contains a second innovation term in ARCH(q) models framework. SV models are extensively discussed by Hoston (1993) and Ait-Sahalia and Kimmel (2006).

SV models use σ_t^2 in most GARCH(p,q) models. The formulations of SV models are flexible due to the volatility error terms. The second noise term increases the difficulty in estimating SV models which cannot be estimated directly by maximum likelihood as GARCH models. The alternative methods to estimate SV models are quasi-maximum likelihood estimation approach (Feernandez, Ley and Steel, 2001) with assumption of Gaussiann volatility proxies and Markov Chain Monte Carlo methods (Billio et al., 2011a and 2011b). The SV model is further extended to allow for long memory in volatility based on the fractional difference. The fact that the aoutocorrelation function of the squared or absolute values of asset return series often slowly decays motivates the development of long-memory SV model (Tsay, 2005). Estimating the long memory SV is considerably complicated.

SV models are widely used in the option pricing studies because it relaxes the assumption of fixed volatility in Black-Scholes formula. By the way, few of scientists used SV models in other financial applications partially due to the complication in estimating the model parameters (Poon, 2008).

5) Forecasting Volatility

The introduction of GARCH(p,q) models give the alternative volatility forecasting models which involve the constant updating of parameter estimates. The forecasting ability of various GARCH(p,q) models is discussed in paper of Liu and Maheu (2008). Given the available information at time t, the researcher can produce the one or more step-ahead forecasts for the conditional variance of as long as the researcher had estimated the GARCH models for conditional variance, error terms

and theirs lagged values (Bollerslev, 1986). The one-step-forecast is available at the current time and the more step-ahead forecasts can be derived based on iterative procedure. Hence, the point is that it is possible to obtain the *s*-step-ahead forecast recursively. The forecasting evaluation of GARCH(p,q) models are considered in a studied paper of Goyal (2000) and Billio et al. (2011a). They showed that well-specified volatility models provide accurate volatility forecast. Moreover, the high-frequency data allow for more accurate ex-post inter-daily volatility measurements.

As mentioned on the above section, SV model including innovation terms is more flexible than GARCH models, so it was found to be better fit to the financial market returns. However, there is no consensus of the SV model superiority over GARCH models in modeling and forecasting volatility. Ray and Panda (2011) carried out the study of volatility forecast for both stock indices and exchange rates. They found that the forecast errors of SV model are larger than GARCH models in exchange rates, but SV model outperforms GARCH models for forecasting stock indices volatility (Lopez, 1999). In contrast, Liu and Lee (2009) had found the equal performance among SV model and other forecast approaches. Evaluating the performance of different forecasting models plays a very important role in choosing the most accurate models (Bera and Higgins, 1993). In particular, the researchers and investors need to decide the evaluating criteria on which to base. Although the vast number of papers has studied the construction of modeling and forecasting volatility, a few of them focus on the volatility forecasting evaluation (Bollerslev, Engle and Nelson, 1994).

According to Bollerslev, Engle and Nelson (1994), the economic loss function, which determines the cost incurred by the investors are the most meaningful measure of forecast evaluation. It is ideal to employ the economic loss function. However, the economic loss function is normally unavailable because it requires the specific details of the investors' decision process and the cost or benefits that result from using these forecasts. Therefore, the statistical loss function is utilized in

practice instead of economic loss function. The most widely used evaluation measures are Mean Error (ME), Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). Lopez (1999) has introduced other evaluation frameworks with probability scoring rules.

B. Review on Related Studies

Mahdi, Vali and Hakim (2011) had said that a proper decision based on scientific principles is a good opportunity for an investor to exploit their capital and by a good knowledge of scientific methods; they can allocate their resources properly. So, they should be acquainted with scientific methods of investment like artificial intelligence. There had always been two investment theories in stock exchange. Firstly, basic analysis focused on market that was formed from reactions between investments; these reactions are based on real information from the company's existing at the market. Company's real information is gathered by the analysis of the existing situation and the perspectives follow it. Market is formed based on rational viewpoints. The last, golden dream focused on market was a result of investors who consider they should buy with the price of 20 units today and sell with the price of 30 units tomorrow. Rahnamay, Falah and Kordlouie (2011) had mention that forecast manipulation in stock price of companies should base on findings variable such as size of company, ratio of price to earnings, clarity of information, stock liquidity and shareholder structure of company. Indeed, Stock and Watson (2003a) had gave detail on their research that contemporaneous forecasts has lower average loss than any of the more sophisticated combination forecasts, a finding consistent with other empirical investigations of combination forecasting. Armstrong (2001) and Timmermann (2005) had stated that combining is useful to the extent that each forecast contains different yet valid information. Combining forecasts had some key principles such as:

a. Different methods or data or both,

b. Forecasts from at least five methods when possible,
c. Formal procedures for combining, equal weights when facing high uncertainty,
d. Trimmed means,
e. Weights based on evidence of prior accuracy,
f. Weights based on evidence of prior accuracy,
g. Weights based on track records if the evidence is strong, and
h. Weights based on good domain knowledge.

Combining is most useful when there were

a. Uncertainty as to the selection of the most accurate forecasting method,
b. Uncertainty associated with the forecasting situation, and
c. A high cost for large forecast errors.

Billio et al. (2011a) had applied their examining in forecast financial stock index showed that their improving statistical accuracy, in particular in terms of density forecasting. Montgomery, Hollenbach and Ward (2011) had discovered model weights that based solely on the components goodness-of-fit with no effort to adjust for their generalize ability. Hoogerheide et al. (2010) had indicated empirical applications that averaging strategies can give higher predictive quality than selecting the best model, and properly specified time varying model weights yield higher forecast accuracy and substantial economic gains compared with other averaging schemes. Reftery et al. (2004) had described that such methods could be combined with the present proposal to produce multimodal and/or multi-analysis ensembles that reproduce spatial correlation of error fields by creating ensembles of field corresponding to each ensemble member and simulating a number of fields from each of these ensembles that is proportional to the corresponding BMA weight.

BMA had given to concentrate forecasts when we changed from linear models and average over specifications that had agreed for GARCH effects in the originations to log-volatility (Liu and Maheu, 2008). The ARCH and GARCH model perform very well but these models only consider the magnitude of the random shocks and do not utilize the direction of the shocks. The non-negativity constraints on the parameters create difficulties in estimation, and in some studies they are found negative. The GARCH models lead to the appearance of some instruments advanced enough to model the financial series (Bollerslev, 1986). The appearance of the GARCH models lead to a better understanding and a modeling of the evolution of the financial series, these models developing both in un-multivariate and multivariate models. In addition to incorporating the nonlinearity in the threshold GARCH model, the threshold or trigger variable takes into account the effect of correlation between conditional variance and other observed variables that represent trading activities. The use of the threshold model is particularly motivated by the volatility volume relationship (Bollerslev, Engle and Nelson, 1994). The established volatility-volume relation motivates the use of volume as the trigger variable in our threshold GARCH model. Since volume and volatility are highly correlated, volume must be treated as an endogenous threshold variable. In such models the time-varying covariance matrix is generated by a small number of orthogonal univariate GARCH models, identified using principal componcnts analysis. In contrast, in this paper we propose to estimate the dynamics of time-varying co-variances and correlations using full multivariate GARCH models (Bera and Higgins, 1993). This avoids the use of variance reduction or similar techniques which can yield very poor forecasts in some practical applications. In such models univariate GARCH processes are estimated for each financial instrument.