Yip Chee-YIN, PhD
University Utara, Malaysia
E-mail: yip@uum.edu.my
Ng Kok-HAUR, PhD
Institute of Mathematical Sciences
University of Malaya, Kuala Lumpur
Malaysia
E-mail: kokhaur@um.edu.my
Lim Hock-EAM, PhD
University Utara Malaysia
E-mail:lheam@uum.edu.my

THE RELATIVE PREDICTIVE ABILITY OF FORECAST WEIGHT AVERAGING AND MODEL AVERAGING PROCEDURE

Abstract. *This paper proposes a procedure called forecast weight averaging which is a specific combination of forecast weights obtained from different methods of constructing forecast weights for the purpose of improving the accuracy of pseudo out of sample forecasting. It is found that under certain specified conditions, forecast weight averaging can lower the mean squared forecast error obtained from model averaging. In addition, we show that in a linear and homoskedastic environment, this superior predictive ability of forecast weight averaging holds true irrespective whether the coefficients are tested by t statistic or z statistic provided the significant level is within the 10% range. By theoretical proofs and simulation study, we have shown that model averaging like, variance model averaging, simple model averaging and standard error model averaging, each produces mean squared forecast error larger than that of forecast weight averaging. Finally, this result also holds true marginally when applied to business and economic empirical data sets, Gross Domestic Product (GDP growth rate), Consumer Price Index (CPI) and Average Lending Rate (ALR) of Malaysia.*

Key Words*: model averaging, forecast weight averaging, mean squared forecast error.*

JEL Classification: C 530

1. INTRODUCTION

Forecasting has a long history. Accurate forecasting plays a very important role in almost every sector of life especially in decision making. In the financial sector, accurate forecasts can result in big jump in profit. Standard forecasting is usually based on well specified models like the linear regression models. But one problem arises from this type of models is how to identify the well specified model out of so many possible models which can be constructed from the same set of regressors. We usually use model selection procedure to overcome this problem. Model selection has been around for a long time. Many different methods of model selection has been proposed and advocated and each one of them is based on distinctive estimation criteria. Some of the well known criteria are Akaike information criterion (AIC) (Akaike, 1973), Mallows criterion (MC) (Mallows, 1973), Bayesian information criterion (BIC) (Schwarz, 1978), the focused information criterion (Claeskens and Hjort, 2003) and many others.

It has long been suspected that model selection procedure has flaws which will influence the final inference. This has motivated many researchers to look into how model selection can affect inferences. Potscher (1991) showed that AIC model selection method results in distorted inference. Buhlmann (1999) examined and presented conditions under which post-model-selection (PMS) estimators are mostly adaptive. Then Leeb and Potscher (2003, 2005, 2006) studied the unconditional and conditional distribution of PMS estimators and found that they cannot be uniformly estimated. This literature suggests that model selection may not the best procedure for constructing the best model for forecasting. The main reason is the existence of model selection uncertainty.

Model averaging is an alternative method to model selection and its advantages are that it can reduce estimation variance and at the same time control omitted variable bias. There is a large literature on model averaging notably Bayesian model averaging (BMA) and an ever growing frequentist literature. Draper (1995) and Raftery, et al. (1997) made seminal contributions to BMA. In the frequentist literature, Buckland, et al. (1997), and Burnham and Anderson (2002) suggested exponential AIC weights for model averaging. Hansen (2007) introduced the idea of model averaging by using weights which minimize the Mallows criterion.

Each of the model averaging methods mentioned above has its own shortcoming. No one of them can claim to be the best. This fact motivates us to think of averaging the forecast weights from different methods, see Yip, et al. (2011). We continue with Yip's work and investigate whether their forecast weights averaging procedure can be applied more accurately to variance model averaging (VMA) or standard error model averaging (EMA) other than simple model averaging (SMA) and BMA. We would also like to find out whether the behavior of forecast weight averaging (FWA) estimator remain unchanged if t statistical test is conducted instead of z test for the case of linear combination of two linear models in a homoskedastic environment.

The rest of the paper is organized as follows: Section 2 introduces model averaging procedure notably model averaging of two simplest linear models. Section 3 reviews existing methods of selection of forecast weights. Section 4 presents forecast weight averaging (FWA) while Section 5 presents and discusses the results of a simulation study. Section 6 discusses three empirical examples of how model selection, averaging and forecast weights averaging are applied in an economic environment. Section 7 discusses the theoretical framework of the FWA procedure. Section 8 concludes this paper.

2. THE MODEL AVERAGING PROCEDURE

Suppose that we have two models, M_1 and M_2. Let f_1 and f_2 be the forecasts obtained by using M_1 and M_2 respectively.

If f_c represents the combined or averaging forecasts, w_1 and w_2 are the forecast weights respectively for f_1 and f_2, then we have model averaging forecasts or combined forecasts be given by

$$f_c = w_1 f_1 + w_2 f_2 \tag{1}$$

and

$$w_1 + w_2 = 1 .$$

The values of w_1 and w_2 are constrained so that its sum is equal to unity in order that each forecast contributes the correct amount of share to the forecast combination. If f_1 and f_2 are unbiased, then it can be shown that f_c is indeed unbiased. With f_c unbiased, then mean squared forecast error (MSFE) of f_c is equivalent to $Var(e_c)$, where e_c, e_1 and e_2 are respectively the combined forecast error, forecast errors of models 1 and 2. We are mainly interested in MSFE. Following Equation (1), we have an equivalent equation relating e_c, e_1 and e_2, that is

$$e_c = w_1 e_1 + w_2 e_2 . \tag{2}$$

Its variance is given by

$$Var(e_c) = w_1^2 Var(e_1) + w_2^2 Var(e_2) + w_1 w_2 Cov(e_1, e_2) + w_2 w_1 Cov(e_2, e_1) . \tag{3}$$

There are two covariance term connecting the two forecast errors in Equation (3). This suggests that a matrix approach is necessary to show that the forecast combination variance is no greater than either each one of the individual forecast variances.

Let covariance matrix, $\mathbf{V}$ is given by

$$\mathbf{V} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \tag{4}$$

where σ_{11} and σ_{22} are respectively denote the variance for the forecast errors e_1 and e_2, σ_{12} and σ_{21} are covariance of forecast errors e_1 and e_2.

Consider the combined forecast error defined in Equation (2), we can rewrite the combined forecast error variance in Equation (3) as follows:

$$\Sigma_c = \sum_{i=1}^{2} \sum_{j=1}^{2} w_i w_j \sigma_{ij} . \tag{5}$$

We would have to search for a set of weights which minimizes Equation (5) subjected to the condition that $w_1 + w_2 = 1$. By using Lagrange undetermined multiplier technique, it can show that

$$[\Sigma_c]_{\min} = \mathbf{w}_{\min}^T \mathbf{V} \mathbf{w}_{\min}$$

$$= (\mathbf{I}_2^T \mathbf{V}^{-1} \mathbf{I}_2)^{-1} \tag{6}$$

where $\mathbf{I}_2^T = (1 \quad 1)$.

From Equation (6), we obtain

$$\frac{1}{[\Sigma_c]_{\min}} = \frac{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}. \tag{7}$$

After rearranging, Equation (7) becomes

$$\frac{1}{[\Sigma_c]_{\min}} - \frac{1}{\sigma_{11}} = \frac{(\sigma_{11} - \sigma_{12})^2}{\sigma_{11}(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} \geq 0 , \tag{8}$$

and

$$\frac{1}{[\Sigma_c]_{\min}} - \frac{1}{\sigma_{22}} = \frac{(\sigma_{22} - \sigma_{12})^2}{\sigma_{22}(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} \geq 0 . \tag{9}$$

Equations (8) and (9) imply that the combined forecast variance is less than the minimum of the two individual forecast variances, that is

$$[\Sigma_c]_{\min} < \min(\sigma_{11}, \sigma_{22}) . \tag{10}$$

Equation (10) shows that the forecast error variance is always smaller than each of the individual forecast error. This verification is based on fact that σ_{ii} is constant throughout the data series, but this is rarely the case in practice because σ_{ii} will change slightly at least from one part of the series to the next. Moreover Equation (10) is designed for population setting and we deal with samples most of the time. Thus Equation (10) may not hold all the times in practice.

2.1 MODEL AVERAGING OF TWO SIMPLEST LINEAR REGRESSION MODELS

Our approach is similar to that of Buckland (1997) and Yip et al. (2011).We focus our discussion and analysis on two simplest linear regression models, M_1 and M_2. However, it must be noted that the method discussed in this section can be applied equally well for averaging multiple models. We combine these two simplest linear regression models for forecasting the conditional means by using model averaging technique. Then we replace the forecast weights by forecast weights averaging from three different methods of assigning forecast weights. There are two risk measures, mean squared error (MSE) and MSFE to measure forecasting ability of any model. However, since out of sample forecasting focus primarily on MSFE. We use mainly MSFE to verify that forecast weights averaging technique is indeed can improve forecast accuracy. Our two simplest Models M_1 and M_2 are nested strongly nested. This is because they belong to the same regression family and it is necessary for application of the methods discussed in this paper. Extending the methods to allow for non-nested models is highly desirable.

METHODOLOGY:

Let the two models be denoted by M_1 and M_2. These two models are shown in Equations (11) and (12).

$$M_1: \qquad y = \beta_0 + u_1 \tag{11}$$

$$M_2: \qquad y = \beta_0 + \beta_1 x + u_2 \tag{12}$$

We make the following assumptions:

a) The two models are linear with constant variance.

b) The error terms u_1 and u_2 follow normal distribution with zero mean, $u_1 \sim N(0, \sigma_1^2)$ and $u_2 \sim N(0, \sigma_2^2)$.

c) x is a realization from an independent and identically distributed (iid) with mean zero. This assumption will ensure that $\sum_{i=1}^{n} x_i = 0$ and that x can be taken as a constant. We let the conditional means for M_1 and M_2 be represented by θ_1 and θ_2 respectively. With that, we have the followings:

$$M_1: \quad \theta_1 = \beta_0 \; ; \qquad E(y) = \hat{\theta}_1 = \hat{\beta}_0 \tag{13}$$

$$M_2: \quad \theta_2 = \beta_0 + \beta_1 x \; ; \quad E(y) = \hat{\theta}_2 = \hat{\beta}_0 + \hat{\beta}_1 x \tag{14}$$

We would like to forecast $\theta = \beta_0 + \beta_1 x$ he future conditional mean by assigning weights for θ_1 and θ_2. Thus we have the following equation

$$\hat{\theta} = w_1 \hat{\theta}_1 + w_2 \hat{\theta}_2 \tag{15}$$

where w_1 and w_2 are the forecast weights under a constrained condition that

$$w_1 + w_2 = 1 . \tag{16}$$

By substituting Equations (13) and (14) into Equation (15), we obtain the following:

$$\hat{\theta} = \hat{\beta}_0 + w_2 \hat{\beta}_1 x . \tag{17}$$

The above derivation for two models can always be generalized to accommodate combination of multiple models.

2.1.1 T STATISTIC VERSUS Z STATISTIC

Yip, et al. (2011) has used a similar approach for deriving Equation (17). However, they use z statistic to test for the validity of the coefficient β. Z statistic is not a pivotal statistic because it depends on the nuisance parameter x. We need to use t statistic for the test, but t statistic has a degree of freedom which makes it not so independent after all. We propose to study the behavior of the t statistical test in the vicinity that it can be approximated well by z statistic. This is our first contribution in this paper. This section deals mainly the theoretical aspect of this asymptotic behavior of the estimator.

Equation (17) is a linear regression, so we have to test whether the parameters $\beta = 0$ against $\beta \neq 0$ by using t test where $\beta^T = [\beta_0 \; \beta_1]$. This is usually done by using the estimated $\hat{\beta}_1$ from the sample for the test. Thus we have: $\left[I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) < t_{\alpha,v}\right) + I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) > t_{1-\alpha,v}\right)\right]$ is the condition for $\hat{\beta}_1$ to be valid so that β_1 is nonzero and $t_{\alpha,v}$ is the $(1-\alpha)100\%$ point of the $t-$distribution with v degrees of freedom . With that, we have our estimated linear equation to be given by

$$\hat{\theta} = \hat{\beta}_0 + w_2 \hat{\beta}_1 x \left[I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) < t_{\alpha,v}\right) + I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) > t_{1-\alpha,v}\right)\right]. \tag{18}$$

We let $J = \left[I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) < t_{\alpha,v}\right) + I\left((\hat{\beta}_1 / \sqrt{v_{\beta_1}}) > t_{1-\alpha,v}\right)\right]$ so that Equation (18) can be written as

$$\hat{\theta} = \hat{\beta}_0 + w_2 \hat{\beta}_1 x J. \tag{19}$$

We need to make sure that the two estimated beta, (β_0 and β_1) are independent of each other. This is because uncorrelated/independent regressors is the basic condition for linear regression to be valid. To satisfy this requirement, we transform the t statistic into z statistic by defining a local to zero framework which has a parameter c which we name as conversion parameter. We investigate the asymptotic behavior of the approximated t statistical test and see how it can affect our final inference. We usually choose c to be small and n large so that t statistic can be approximated accurately by z statistic. However, in our experiment, we vary the values of c from -50 to 50 and fix the value of n as 100 for the purpose of finding at what significant level that z and t test will produce the same result in our analysis, that is

$$\hat{\beta}_1 / \sqrt{v_{\beta_1}} = c/n + z \tag{20}$$

where $z = (\hat{\beta}_1 - \beta_1)/\sqrt{v_{\beta_1}}$ and the value of c is allowed to vary from -50 to 50, -40 to 40, -30 to 30, -20 to 20, -10 to 10 and -5 to 5 as follows:

$$c = (-a, a) \text{ for a} = 50,40,30,20,10,5 \quad \text{and } n \to 100 \tag{21}$$

The ratio of c to n is included in equation for the purpose of adjusting the t statistic to z statistic. In addition we need $\sqrt{v_{\beta_1}}$ to be constant, but it depends on the values of x. We transform $\sqrt{v_{\beta_1}}$ as follows:

$$\sqrt{v_{\beta_1}} = \alpha_2 \sigma_2 \tag{22}$$

where σ_2^2 is the homoskedastic variance of the regression error and α_2 is the second diagonal element of the inverse matrix of $(\mathbf{X}^T\mathbf{X})^{-1}$ which is given as follows:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} \alpha_1^2 & a \\ b & \alpha_2^2 \end{bmatrix}. \tag{23}$$

By using Equations (20) and (22), we obtain:

$$\hat{\beta}_1 = \alpha_2\sigma_2[c/n + z]. \tag{24}$$

After substituting Equation (24) into Equation (19), we obtain

$$\hat{\theta} = \hat{\beta}_0 + w_2 x\alpha_2\sigma_2[c/n + z]J\,. \tag{25}$$

By substituting Equation (19) into Equation (25), we obtain Equation (26) linking β_1 to c, n and σ_2, that is

$$\beta_1 = (c\alpha_2\sigma_2)/n\,. \tag{26}$$

Using $\theta = \beta_0 + \beta_1 x$, Equations (25) and (26), we obtain the following mean bias, variance bias and mean squared forecast error or bias. We omit the condition J for the validity of β_1, for clarity. These bias, variance and forecast error are shown as follows:

$$E(\hat{\theta} - \theta) = x\alpha_2\sigma_2 c(w_2 - 1)/n\,, \tag{27}$$

$$Var(\hat{\theta} - \theta) = Var(\hat{\theta}) = \sigma_1^2/n + x^2 w_2^2\alpha_2^2\sigma_2^2\,, \tag{28}$$

and

$$\text{MSFE}(\hat{\theta}) = \sigma_1^2/n + x^2\alpha_2^2\sigma_2^2\left[w_2^2 + c^2(w_2 - 1)^2/n^2\right]. \tag{29}$$

where σ_1^2/n is the variance for the intercept term $\hat{\beta}_0$ Note that in the derivation of Equation (17), we have assumed that $\hat{\beta}_0$ is the same for Equations (13) and (14) which is true since we have assumed that x follows an iid with mean zero and that it is a realization of the distribution.

3 FORECAST WEIGHTS SELECTION REVIEW

Model averaging is used to construct forecasting model combination for performing out of sample forecasting. Forecast combination has been introduced in the literature long time ago. The idea was introduced by Bates and Granger (1969) and then it was extended by Granger and Ramanathan (1984). The literature is large and some excellent reviews include Granger (1989), Diebold and Lopez (1996), Hendry and Clements (2002), Timmermann (2006) and Stock and Watson (2005). As a whole, there is a broad consensus that forecast combination can improve forecast accuracy. However, there is no consensus as to how to construct forecast weights for the forecast model combination. This motivates us to follow Yip, et al. (2011) to experiment on averaging different forecast weights, which is our second contribution in this paper. We shall now review some forecasting combination literature, notably seminal paper by Bates and Granger (1969) and the recent one, simple model averaging by Stock and Watson (2004).

Bates and Granger (1969) have elaborately identified the conditions for combination of forecasts. Among the five methods for forecasting combination presented in the paper, the use of forecasts variance of errors as forecast weights is good as variance of errors is a simple way to illustrate the mechanism of Akaike Information Criterion.

We shall use this proportion of variance of errors as our first method to generate forecast weights. We name this method as variance error model averaging (VMA). Besides VMA, we also use the ratio of standard errors of the two models M_1 and M_2 to construct forecast weights. We name this method as EMA as defined in the abstract. This EMA is our third method of generate forecast weights.

The recent forecasting literature focused on forecasts based on Bayesian model averaging (BMA) and simple model averaging (SMA). Wright (2003a,b), Stock and Watson (2004) have demonstrated that using SMA forecast combination, forecast accuracy not only improve tremendously but also change little over time. In particular simple model averaging (equal weights) and Bayesian model averaging have demonstrated big success in forecasting. We also have Hansen (2008) forecast combination model based on using forecast weights selected by the Mallows model averaging criterion. Our second method of selecting forecast weights is the SMA using equal weights. This amounts to saying that we give a simple average of the total number of models M, involved as weight to each forecast from different models. That is each forecast weight is effectively $1/M$. This method of giving weights to each forecast works intriguingly well not only in the real term but also it is stable over time. However, Stock and Watson (2004) do not offer a definitive explanation for the excellent performance of SMA.

We choose SMA as our second method for selecting weights mainly because it performs well in forecasting as illustrated by Stock and Watson (2004) 's experiment. Moreover we think that all data especially economic data are generated by human activities. In general term, most of us prefer moderation in most aspects of life and this drives the data generating process towards the central mean. Thus we have three different methods, one of which is old but reliable, second is the most recent successful one in the forecasting literature and the third is ratio of the standard errors of both models. Since we have advocated considerable attention to the method of averaging, it would be only logical to assume that forecast weights from different methods can also be estimated by some sort of averaging. It is this idea of averaging the forecast weights from different methods that motivates us to continue work done by Yip, et al. (2011) who has done forecast weight averaging for SMA and BMA (Bayesian weight averaging). Instead we consider FWA with weights as weighted averaging of, VMA, SMA and EMA.

3.1 VMA MODEL AVERAGING

We impose the following conditions on our two model averaging models. They are:

a. The individual sets of forecasts are made unbiased. This is because an unbiased set of forecast if combine with a biased set of forecast will most likely produce mean squared error larger than the unbiased forecasts.

b. The forecast combination produces the lowest mean squared forecast error.
c. The performance of the forecast combination is stationary that is stable over time.
d. As the number of forecasts increases, the average forecast weight should approach the optimum value.
e. The weights should vary slightly about the optimum value.

For VMA model averaging, we select weights so that the forecast error variance for the combined models is minimized.

3.1.1 SELECTION OF WEIGHTS FOR VMA

We select the weights under the assumption of homoskedastic variance that is the variance is constant for a particular model. Suppose that the forecast error variances of the models M_1 and M_2 are denoted by σ_1^2 and σ_2^2 respectively. Then the two VMA weights for the combined forecast should be:

$$w_{V1} = \sigma_1^2 / \left(\sigma_1^2 + \sigma_2^2\right),$$

and

$$w_{V2} = \sigma_2^2 / \left(\sigma_1^2 + \sigma_2^2\right). \tag{30}$$

We have chosen that the forecast weight is directly proportional to the respective forecast error variance because of two reasons. One of them is error variance of regression is the main building block for Akaike Information Criterion and the Bayesian Information Criterion. This implies that error variance plays a very important role on the suitability of the model. The other reason is that we have assumed that the forecast error variance is constant for a particular model. Thus in effect, Equation (30) is almost equivalent to a scalar weighted average. Since different model has different forecast error variance and we expect them to behave randomly. We treat this forecast weights w_V as a random variable in general.

The MSFE for VMA method of constructing forecast weights is given by

$$\mathrm{MSFE}(\hat{\theta})_{\mathrm{V}} = \sigma_1^2 / n + x^2 \alpha_2^2 \sigma_2^2 \left[1 + (cu^2 / n)^2\right] / (1 + u^2)^2 \tag{31}$$

where $u = \sigma_1 / \sigma_2$.

3.2 SELECTION OF WEIGHTS FOR SMA

For SMA model averaging, we have a number of choices notably medians, trimmed mean and simple mean. We choose simple mean. This means that if we have M models for averaging, the chosen simple weights are equal to $1/M$. This simple mean or central mean is chosen because of two reasons. One of which is that we believe most data sets mirror the way human being behave since the data are mostly generated by human activity. Most human being are moderate in thinking and they have central tendency. Most of the data carries this piece of information. So when we model data, we should look into its central tendency behavior first that is to say, we put focus on its mean. The second reason is that Stock and Watson (2004) have shown that selecting equal weights as forecast weights can prove to be very successful in real data out of sample forecasting. Thus, we have for two models, $M = 2$:

$$w_{S1} = w_{S2} = 1/2 . \tag{32}$$

By substituting $w_{S2} = 1/2$ in Equation (29), The MSFE for SMA method is given by

$$\mathrm{MSFE}(\hat{\theta})_{\mathrm{S}} = \sigma_1^2 / n + x^2 \alpha_2^2 \sigma_2^2 \left[1 + (c/n)^2\right] / 4 . \tag{33}$$

3.3 SELECTION WEIGHTS FOR EMA

The forecast weights are the ratio of the respective standard deviation. The two forecast weights for $M = 2$ are given below:

$$w_{E1} = \sigma_1 / (\sigma_1 + \sigma_2), \tag{34}$$

and

$$w_{E2} = \sigma_2 / (\sigma_1 + \sigma_2). \tag{35}$$

The MSFE for EMA method is then given by

$$\mathrm{MSFE}(\hat{\theta})_{\mathrm{E}} = \sigma_1^2 / n + x^2 \alpha_2^2 \sigma_2^2 \left[1 + (cu/n)^2\right] / (1+u)^2 . \tag{36}$$

4. FORECAST WEIGHT AVERAGING (FWA)

The formulation of forecast weight averaging would be defined in the same way as model averaging except that we are considering forecast weights from different methods. The rationale underlying this formulation is that forecast weights from each method would contain certain important information about the forecast not available in the other method of constructing forecast weight. We discuss the general formulation initially and only then, we discuss in detail how to combine forecast weights from two and three specific methods of constructing forecast weights. It must be noted here that FWA depends on the suitability of the model averaging procedure. Let say we have three methods, A, B and C of assigning forecast weights, denoted by w_A, w_B and w_C. Each method is used to combine forecasts from two models only. Since we are combining weights from three methods only, we would have the following information.

$$w_{A1} + w_{A2} = 1 , \tag{37}$$

$$w_{B1} + w_{B2} = 1 , \tag{38}$$

and

$$w_{C1} + w_{C2} = 1 . \tag{39}$$

We select equal weight method to assign weight to each forecast weight w_A, w_B and w_C. We shall explain the reason for this selection latter. Thus we have the following equations

$$w_{Ca1} = (w_{A1} + w_{B1} + w_{C1}) / 3 \tag{40}$$

and

$$w_{Ca2} = (w_{A2} + w_{B2} + w_{C2}) / 3 \quad . \tag{41}$$

Note that $w_{Ca1} + w_{Ca2} = 1$, as required by our assumption. As w_{Ca1} and w_{Ca2} are difficult to calculate if BMA and other criterions are used, most of the time we calculate one of them only for forecasting. This can be done by using the restricted sum of w_{Ca1} and w_{Ca2} and our final formula would be Equation (43).

$$f_{Ca} = w_{Ca1} f_1 + w_{Ca2} f_2 \tag{42}$$

$$f_{Ca} = f_1 + w_{Ca2}(f_2 - f_1) \tag{43}$$

We propose to combine the three weights w_V, w_S and w_E to form forecast weight averaging (FWA). It is logical to think that averaging out the averages would give us a much better measure to smoothen the total forecast fluctuation. Thus we would be able to obtain more accuracy in forecasting if we average out forecast weights from different models combination. We select only two models for FWA. However, the method describes in this paper can be generalized to accommodate M models combination.

We have to choose weights w_{Ca} by combination averaging of the VMA weights w_V , SMA weights w_S and EMA weights w_E. A linear combination of all these three variables is given in Equation (44), that is

$$w_{Ca2} = \left[\sigma_2^2/(\sigma_1^2 + \sigma_2^2) + 1/2 + \sigma_2/(\sigma_1 + \sigma_2)\right]/3 \tag{44}$$

Substituting Equation (44) into Equation (43) will give us the following combination forecast.

$$f_{Ca} = f_1 + \left[\sigma_2^2/(\sigma_1^2 + \sigma_2^2) + 1/2 + \sigma_2/(\sigma_1 + \sigma_2)\right](f_2 - f_1)/3 \tag{45}$$

By substituting Equation (44) into Equation (29), we would obtain Equation (46) which is the MSFE for FWA forecast weight averaging.

$$\text{MSFE}(\hat{\theta})_{\text{F}} = \frac{\sigma_1^2}{n} + \frac{1}{9}x^2\alpha_2^2\sigma_2^2\left[\left(\frac{1}{1+u^2} + \frac{1}{2} + \frac{1}{1+u}\right)^2 + \frac{c^2}{n^2}\left\{\left(\frac{1}{1+u^2} + \frac{1}{2} + \frac{1}{1+u}\right) - 2\right\}^2\right] \tag{46}$$

Thus by comparing Equations (31), (33), (36) and (46), we will be able to ascertain whether FWA is a better forecasting model than VMA, SMA and EMA. We use two different methods for this comparison. First, we use simulation and then we use mathematical verification of the simulation result.

5. A SIMULATION STUDY

A simulation study is carried out to assess the performance of all forecasting methods by computing their MSFE as illustrated by Equations (31), (33), (36) and (46). First, we generate data with sample size of $n = 99$ from model 1 and model 2 as given below:

Model 1 $\quad y = 1 + u_1, \qquad u_1 \sim N(0,2) \qquad (47)$

Model 2 $\quad y = 1 + 3x + u_2, \qquad u_2 \sim N(0,2) \qquad (48)$

where u_1 and u_2 are independent. We next calculate the values of MSFE for each method for given values of $x = 20$ and c. Finally, the above steps repeat for

$N = 2000$ replications and obtain the average values of MSFE for all methods. Figures 1 illustrates the relative performance of forecasting ability of all the four models as the values of c vary.

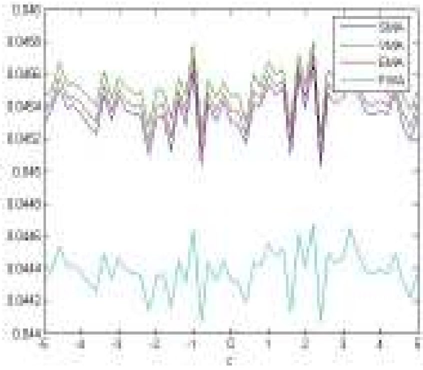

Figure 1: **Comparison of MSFE when $x = 20$ for range of c from -5 to 5**

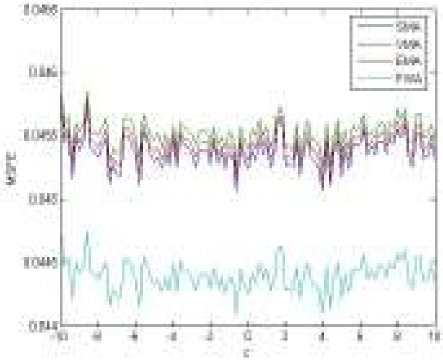

Figure 2: Comparison of MSFE when $x = 20$ for range of c from -10 to 10

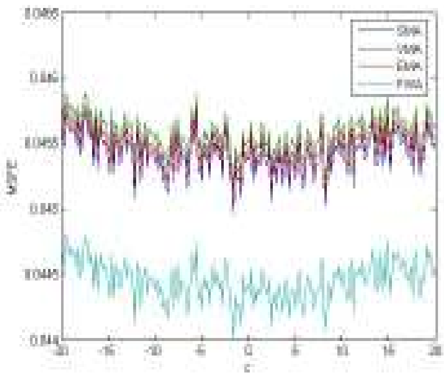

Figure 3: Comparison of MSFE when $x = 20$ for range of c from -20 to 20

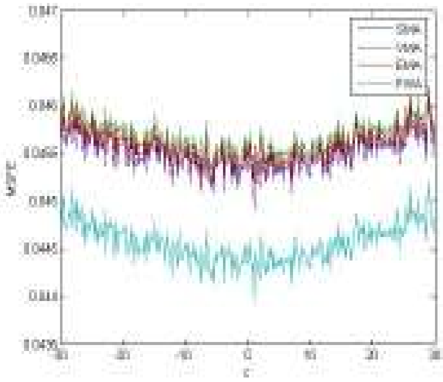

Figure 4: Comparison of MSFE when $x = 20$ for range of c from -30 to 30

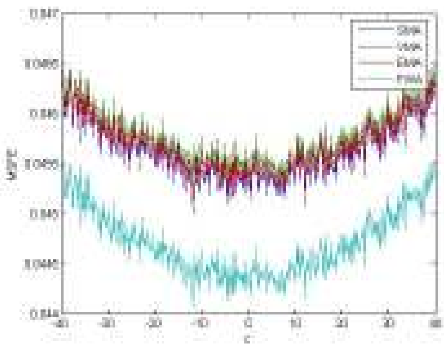

Figure 5: Comparison of MSFE when $x = 20$ for range of c from -40 to 40

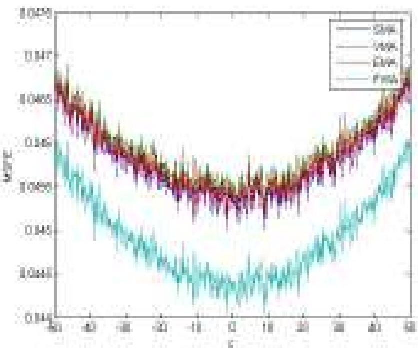

Figure 6: Comparison of MSFE when $x = 20$ for range of c from -50 to 50

Figures 1 to 6 show the simulation results with different range of values of c such that Figure 1 has the smallest range value of c (-5 to +5) and Figure 6 has the largest range value of c (-50 to +50). Figures 1 to 6 clearly depict the values of MSFE for all the estimators FWA, VMA, EMA and SMA and that they are independent of the values of c only for a small range of the values of c ranging from -5 to +5 (Figure 1) and -10 to +10 (Figure 2). Nevertheless, from the values of c more than ± 10, the MSFE appears to be positively correlated with the value of c (see Figure 6). These results imply that it is only within the range of values of c from -10 to +10, that t and z tests can produce similar results within some specific range of the significant level.

Relating to the relative predictive performance of the FWA, VMA, EMA and SMA, from Figures 1 to 6, it is found that the FWA has clearly the lowest value of MSFE than VMA, EMA and SMA, across the various range value of c. Whereas, the VMA, EMA and SMA have almost similar values of MSFE. Quantitatively, from Figure 6, it appears that for the range value of c from -10 to +10, FWA is able to reduce the MSFE of VMA, EMA or SMA by more than 3%. This percentage is obtained by visually examining the graph in Figure 6 whereby it is clear that MSFE decreases from 0.0455 to 0.0445. However, the better predictive performance of FWA is reducing with the increasing value of c (in absolute term). Figure 6 indicates that as the value of c increases, for instance at value of c at 50 (or -50), FWA is able to reduce the MSFE of VMA, EMA or SMA by only around 1 %.

In short, the simulation results suggest that FWA has better predictive performance than VMA, EMA and SMA. However, this better predictive performance is deteriorating with increase in the value of c, the local to zero parameter. In the following section, we will present the empirical results.

6. EMPIRICAL RESULTS

We have conducted three real life business and economic examples of Malaysian macroeconomic series. Firstly, the Gross Domestic Product (GDP) with net export (NEX) as regressor. Secondly, the Consumer Price Index (CPI) with employment (EMP) as regressor. Lastly, the Average Lending Rate (ALR) with money supply (MS) as regressor. The data are obtained from the database of IFS (International Financial Statistics) in quarterly from Q1 1988 to Q3 2006 (real GDP), Q1 1999 to Q3 2006 (NEX), Q1 1957 to Q3 2006 (CPI), Q1 1998 to Q3 2006 (EMP), Q4 1986 to Q3 2006 (ALR), and Q4 1969 to Q3 2006 (MS). The last ten observations (Q1 2004 to Q3 2006) are used for calculation of MSFE.

For each example, we fit the data with the models given by Equations (11) and (12). For instance, the first example of GDP, we estimate the following two models:

$$M_1 : y_t = \beta_0 + u_t$$
$$M_2 : y_t = \beta_0 + \beta_1 x_t + u_t$$

where y_t denotes respectively GDP and x_t stands for NEX. We use these models, M_1 and M_2 again for the 2nd and 3rd real life examples.

Before we conduct model averaging and forecast weight averaging, we check the BIC values for M_1 and M_2. By applying the BIC information criterion, model with the smallest BIC value is the more stable than the other models. With this consideration, we conclude that M_2 is more stable than M_1 because it returns a smallest BIC value. To confirm that M_2 is the more stable model, we perform the perturbation instability in estimation (PIE) test. Model with the smallest PIE is always more stable than model with larger PIE. We obtained PIE values for models M_1 and M_2 as shown in Table 1. Table 1 shows clearly that GDP and CPI have lower PIE values for Model 2 than Model 1. As for ALR, PIE of Model 2 is almost similar to Model 1. As a whole, PIE in Table 1 confirms that Model 2 is more stable for GDP, CPI and ALR because the smaller the PIE, the more stable the model.

Table 1: **PIE values for M_1 and M_2**

Data	Model	
	M_1	M_2
GDP	1.4305	1.2355
CPI	0.4906	0.3093
ALR	0.1196	0.1330

Table 2: **Comparison of MSFE with different methods of model averaging**

Method	MSFE		
	GDP	CPI	ALR
M_2	210	6	0.28
VMA	307	7	1.00
SMA	937	680	1.02
EMA	499	24	1.40
FWA	551	118	1.41
M_1	2194	2569	6.48

Table 2 compares the results of MSFE for these three real life examples based on five different models, that are M_2, VMA, SMA EMA, FWA and M_1. Table 2 clearly illustrates that the model M_2 has is the lowest MSFE and model M_1 has the highest MSFE across the three real life examples. Since the real GDP, CPI and ALR series of Malaysia are clearly not a mean-reverted series; the M_1 which used unconditional mean for forecasting is expected to have the highest possible MSFE suggesting that it is the worst model. On the contrary, due to the open economy of Malaysia, the net export is a significant predictor of real GDP.

By economic theories, the employment is a significant predictor of CPI (Phillip curve); the money supply is a significant predictor of ALR (money market equilibrium). Thus, the model M_2 is expected to have significantly lower value of MSFE than model M_1 which means that model M_2 is the best model relatively speaking. As such, the MSFE of the combined forecast of models M_1 and M_2 should fall between the MSFE of models M_1 and M_2. The one with MSFE closest to M_2 is expected to have the best predictive ability.

Figures 7, 8 and 9 depict the MSFEs of three real life examples for all models discussed. Across the three real life examples, it is found that SMA has the lowest predictive performance (its MSFE is the closest to MSFE of model M_1). This is due to its combined forecast weight which split equally between M_1 and M_2, regardless the fact that the forecast of M_2 should be assigned higher weight than M_1.

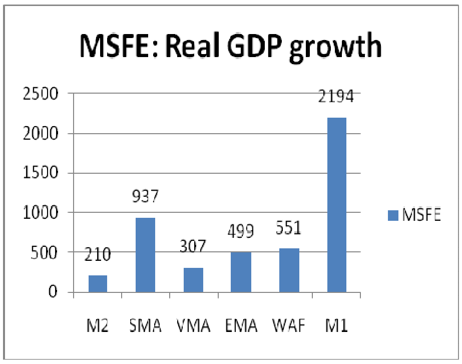

Figure 7: MSFE for Real GDP growth

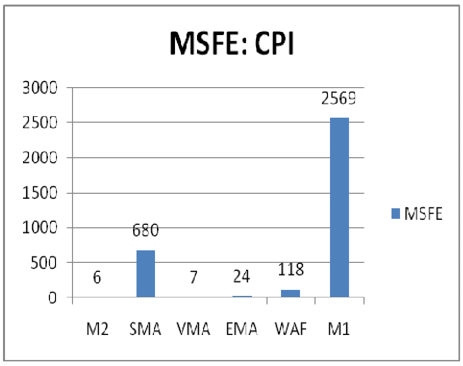

Figure 8: MSFE for Consumer Price Index

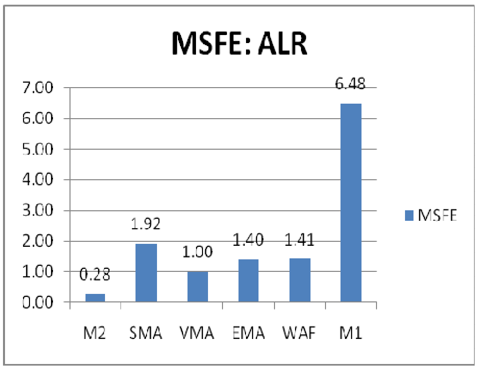

Figure 9: MSFE for Average Lending Rate

The VMA has the highest predictive performance as its MSFE is the nearest to M_2. This is due to its combined forecast weight which assign the heaviest weight to M_2 (the best model) and the lowest weight to M_1 (the worst model). The FWA is found to have not the lowest MSFE. This finding appears inconsistent with the simulation study in Section 5. However, as explained early about PIE stability test, M_2 is the preferred model because its PIE value is 0.3093 and 0.1330 for CPI and ALR, which is well below the mark of 0.5 where model averaging is preferable (see Yuan and Yang (2005)). Intuitively if model averaging is not the best procedure to choose for forecasting purpose as in this case, we expect FWA cannot be the optimal forecasting model as well because FWA is based primarily on model averaging. It must also be noted that these real life data do not satisfy the condition that the sum of the regressor values must be zero. This is usually the case in most practical situation. Therefore β_0 in model M_1 and M_2 may not be the same.

7. MATHEMATICAL VERIFICATION OF THE SUPERIORITY OF FWA

Since the term σ_1^2/n and $x^2\alpha_2^2\sigma_2^2$ are the same for VMA, SMA, EMA and FWA, we compare the rest of the terms in Equations (31), (33), (36) and (46). We compare FWA with each of VMA, SMA and EMA.

7.1 FWA VERSUS VMA

VMA $$\frac{1}{(1+u^2)^2}\left[1+\frac{c^2}{n^2}u^4\right]=\frac{1}{(1+u^2)^2}+\frac{c^2}{n^2}\frac{u^4}{(1+u^2)^2}$$

When the value of u ranges from 1 to 2 by following the definition $u=\sigma_1/\sigma_2$, which is more than 1 most of the time, FWA will returns a lower MSFE than VMA.

7.2 FWA VERSUS SMA

The comparing factor for SMA and FWA are:

SMA $$\frac{1}{4}\left[1+\frac{c^2}{n^2}\right]$$

FWA $$\frac{1}{9}\left[\left(\frac{1}{1+u^2}+\frac{1}{2}+\frac{1}{1+u}\right)^2+\frac{c^2}{n^2}\left\{\left(\frac{1}{1+u^2}+\frac{1}{2}+\frac{1}{1+u}\right)-2\right\}^2\right]$$

It is obvious that $1/9<1/4$, $\left(1/(1+u^2)+1/2+1/(1+u)\right)^2<1$ and that $\left\{\left(1/(1+u^2)+1/2+1/(1+u)\right)-2\right\}^2\approx 1$. Thus FWA returns a lower MSFE than SMA.

By using a similar type of proof, we can also show that FWA can return a lower MSFE than EMA.

8. CONCLUSION

From the simulation result it is obvious that FWA will return a smaller value of MSFE than VMA, EMA or SMA. FWA is effective in forecasting mainly because it averages out information from different models of constructing forecast weights. This result is confirmed by mathematical verification. However, it is assumed that each variance is constant which is difficult to obtain in practice even in the homoskedastic environment. This practical aspect of the variance reveals clearly in our empirical experiment results. Empirical results suggest that the averaging methods are superior if and only if we have imperfect information about the worst and the best models. This is consistent with the theory of portfolio investment where the diversification of investment is preferable due to imperfect market information.

ACKNOWLEDGEMENT

The authors are very grateful to the referee for making various suggestions which have considerably improved the article. We would also like to thank God for giving us an opportunity to send this article for publication.

REFERENCES

[1] **Akaike, H. (1973),** ***Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models***. *Biometrika*, 60, 255 -265;

[2] **Bates, J.M. and Granger, C.W.J. (1969),** ***The Combination of Forecasts.*** *Operations Research Quarterly*, 20, 451-468;

[3] **Buckland,S.T., K.P. Burnham, and N.H. Augustin (1997),** ***Model Selection: An Integral Part of Inference***. *Biometrics*, 53, 603-618;

[4] **Buhlmann, P. (1999*), Efficient and Adaptive Post Model Selection Estimators***. *Journal of Statistical Planning and Inference*, 79, 1-9;

[5] **Burnham, K.P., and D.R. Anderson (2002),** ***Model Selection and Multimodel Inference: A Practical Information – Theoretic Approach***. Berlin: Springer-Verlag.;

[6] **Claeskens, G. and Hjort, N. L.(2003),** ***Frequentist Model Average Estimators***. *Journal of the American Statistical Association*, 98, 900-916;

[7] **Draper, D. (1995),** ***Assessment and Propagation of Model Uncertainty***. *Journal of the Royal Statistical Society*, Ser.B, 57, 45-70;

[8] **Granger, C.W. (1989), Invited Review:** ***Combining Forecasts–twenty Years Later***. *Journal of Forecasting*, 8, 167-173;

[9] **Granger, C.W.J. and Ramanathan, R. (1984),** ***Improved Methods of Combining Forecasts***. *Journal of Forecasting*, 3, 197-204;

[10] **Hansen, Bruce E. (2007),** ***Least Squares Model Averaging***. *Econometrica*, 75, 1175-1189;

[11] **Hansen, Bruce E. (2008),** ***Least Squares Forecast Averaging***. *Journal of Econometrics*, 146, 342-350;

[12] **Hendry, D.F. and Clements, M.P.(2002),** ***Pooling of Forecasts***. *Econometrics Journal*, 5, 1-26;
[13] **Leeb, H. and B.M. Potscher (2003),** ***The Finite-sample Distribution of Post-model -selection Estimators and Uniform versus Non-uniform Approximations***. *Econometric Theory*, 19, 100-142;
[14] **Leeb, H. and B.M. Potscher (2005),** ***Model Selection and Inference: Facts and Fiction***. *Econometric Theory*, 21, 21-59;
[15] **Leeb, H. and B.M. Potscher (2006),** ***Can One Estimate the Conditional Distribution of Post-model-selection Estimators?".*** *The Annals of Statistics*, 34(5), 2554-2591;
[16] **Mallows, C.I.(1973),** ***Some Comments on C_p***. *Technometrics*, 15, 661-675;
[17] **Potscher, B. M. (1991),** ***Effects of Model Selection on Inference***. *Econometric Theory*, 7, 163-185;
[18] **Raftery, A.E. and Y. Zheng (2003),** ***Long Run Performance of Bayesian Model Averaging***. Working Paper, University of Washington;
[19] **Schwarz, G. (1978),** ***Estimating the Dimension of a Model***. *Annals of Statistics*, 6, 461-464;
[20] **Stock. J. H., Watson, M.W. (2004),** ***Combination Forecasts of Output Growth in a Seven-country Data Set***. *Journal of Forecasting*, 23, 405-430;
[21] **Stock. J. H., Watson, M.W. (2005),** ***An Empirical Comparison of Methods for Forecasting Using Many Predictors***. Working Paper, NBER;
[22] **Timmermann, A. (2006),** ***Forecast Combinations***. In: *Elliott, G., Granger, C.W.J., Timmermann. A*. (Eds.*), Handbook of Economic Forecasting*". 1, 135-196;
[23] **Wright, J. H. (2003a),** ***Bayesian Model Averaging and Exchange Rate Forecasts.*** Board of Governors of the Federal Reserve Board International Finance Discussion Paper, No. 779;
[24] **Wright, J.H. (2003b*), Forecasting U.S. Inflation by Bayesian Model Averaging***. Board of Governors of the Federal Reserve Board International Finance Discussion Paper, No. 780;
[25] **Yip, Chee-Yin, Ng, Kok-Haur, and Lim, Hock-Eam (2011*), Forecast Combination Modeling Using a Pooled Forecast Weights in a Linear and Homoskedastic Environment***. *International Journal on GSTF Business Review*, 1, 122-127;
[26] **Yuan, Z., and Yang, Y. (2005*), Combining Linear Regression Models: When and How?.*** *Journal of the American Statistical Association*, 100 (472), 1202-1214.