

Comparison of forecasting methods with an application to predicting excess equity premium

Hsiao, Cheng; Wan, Shui Ki

Published in:
Mathematics and Computers in Simulation

DOI:
[10.1016/j.matcom.2010.03.010](https://doi.org/10.1016/j.matcom.2010.03.010)

Published: 01/03/2011

Document Version:
Peer reviewed version

[Link to publication](#)

Citation for published version (APA):
Hsiao, C., & Wan, S. K. (2011). Comparison of forecasting methods with an application to predicting excess equity premium. *Mathematics and Computers in Simulation*, 81(7), 1235-1246.
<https://doi.org/10.1016/j.matcom.2010.03.010>

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent publication URLs

Comparison of Forecasting Methods with an Application to Predicting Excess Equity Premium

Cheng Hsiao*and Shui Ki Wan†

November 27, 2009

Abstract

This paper reviews various forecast methods including combination using theoretically optimal weights and those under model selection approaches. In addition, we suggest two modified simple averaging forecast combination methods – a mean corrected and a mean and scale corrected method. We conclude that due to the fact that real data is usually subject to structural breaks, rolling forecasting scheme has a better performance than fixed window and continuously updating scheme. In addition, methods that use less information appear to perform better than methods using all the sample information about the covariance structure of the available forecasts. The mean and scale corrected simple average approach yield smaller mean squared forecast error than the three widely used regression approaches suggested by Granger and Ramanathan (1984).

1 Introduction

Forecasting is important to decision makers in both private and public sectors. Much effort has been devoted to the development of modelling and estimation issues to improve forecast accuracy. For instance, models ranging from linear specifications, like simple macro-models and autoregressive models with order p or random walk models to non-linear models like artificial neural network and Markov switching models are proposed to predict exchange rates. Since the seminal work of

*University of Southern California, City University of Hong Kong, and WISE, Xiamen University. Correspondence: chsiao@usc.edu

†University of Southern California. Correspondence: shuiwan@usc.edu

Bates and Granger [2], forecasts combination has gained favor as another way to improve forecasts. Over the past four decades, hundreds of papers are spawned by their ideas. Various methods are proposed to combine forecasts. These include but not limited to Granger and Ramanathan [11], Diebold and Pauly [8] and Chan, Stock and Watson [4]. Clemen [6] provided a nice summary of the applications of forecast combination.

However, there are researchers holding opposing views towards forecast combination. One strand thinks that if we have the information underlying those forecasts, then combining information must be superior to combining forecasts. Diebold once described the role of forecast combination in [7] that "in a world in which information sets can be instantaneously and costlessly combined, there is no role. It is always optimal to combine information sets rather than forecasts". Starting from this point of view, encompassing tests are developed to facilitate the combination of information sets as in Chong and Hendry [5]. The idea is based on hypothesis testing. If a particular model, say model m , encompasses all other models among a set of M models generating forecasts $f_t^{(j)}$, $j = 1, \dots, M$ for our variable of interest y_t , then the regression

$$y_t = \beta_0 + \sum_{j=1}^M \beta_j f_t^{(j)} + u_t, t = 1, \dots, T.$$

leads to the null that $H_0 : \beta_m = 1$ and $\beta_{j \neq m} = 0$ is not rejected. However, as pointed out in Swanson and Zeng [17], traditional hypothesis testing approach works only when the model is correctly specified. In addition, Granger, King and White [10] also stated that it is more difficult to justify using standard hypothesis tests for choosing between two competing models.

Even the underlying information can be obtained costlessly, combination of information doesn't necessarily yield more accurate out-of-sample forecasts. As summarized in Timmermann [18], individual forecasts may be affected by structural breaks in different ways caused by institutional change or technological developments. Especially when the parameters are not known and needed to be estimated, the adjustment to the breaks for different models will be different. Besides, individual forecasting models may be subject to misspecification bias of unknown form. Forecast combination is a way to diversify those unnecessary risk incurred in making forecasts. Theory of combination was established in the seminal work by [2]. Since then, various methods of combination in both classical and Bayesian contexts have been proposed, and empirical studies lend support

to combination of forecasts. For example, Wright [20] showed that Bayesian Model Averaging improves the accuracy in forecasting US inflation. On the other hand, numerous literature have shown that simple averaging performs much better than theoretically optimal weights. Examples include Huang and Lee [12]. Stock and Watson [16] has even termed this as forecast combination puzzle. In this paper, we shall suggest two modified simple averaging methods – a mean corrected and a mean and scale corrected simple averaging method. We shall evaluate their performance vis-a-vis some well-known information approaches – AIC proposed by [1], AICC by [13] and BIC by [15], and forecasts combination approaches – the model averaging using posterior odds ratio [3], Bates and Granger [2], the three approaches in Granger and Ramanathan [11] and simple averaging through both Monte Carlo simulation and an application to predicting excess equity premium.

In Section 2, we shall provide the model and loss function on which comparison are based. We also review various forecast combinations and model selection methods. In Section 3, we shall compare all those methods under different forecasting schemes (continuously updating forecasts, rolling forecasts with fixed window and fixed forecasts as defined in West and McCracken [19]) using a small scale Monte Carlos simulation. In Section 4, an empirical study about predictability of excess equity premium is carried out. Section 5 concludes.

2 Background

Suppose that (y_t, \mathbf{X}_t') are observed, where \mathbf{X}_t is an $N \times 1$ column vector. We would like to construct point forecasts for y_t based on some subsets of \mathbf{X}_t , $\mathbf{X}_t(m)$. Obviously, the optimal forecast model depends on the risk function. In this paper we only consider quadratic loss function $e_t^2 = (y_t - \hat{y}_t)^2$, where \hat{y}_t is a predictor of y_t . For simplicity, only linear regression models are generated. If we are considering all permutation of \mathbf{X}_t except the simple mean model, then we have $M = 2^N - 1$ models.

$$y_t = \tilde{\mathbf{X}}_t(m)' \boldsymbol{\beta}(m) + e_t(m), m = 1, \dots, M, \quad (1)$$

where $\tilde{\mathbf{X}}_t(m)$ includes a constant term and a subset of \mathbf{X}_t , $e_t(m)$ is the forecast residual of model m . We shall assume the first forecast starts at time $P + 1$.

2.1 Three Forecasting Schemes

We shall consider sequences of forecasts of y_t , $t = P + 1, \dots, T$ generated according to the 3 schemes as defined in [19]. The first method is a fixed window forecasting scheme. The first P time series observations are used to estimate the parameter vector $\beta_P(m)$. The whole sequence of forecasts is obtained $f_t^{(m)} = \tilde{\mathbf{X}}_t(m)' \hat{\beta}_P(m)$, $t = P + 1, \dots, T$; $m = 1, \dots, M$. The second one is the continuously updating forecasting procedure. The first regression under this scheme uses $\left\{y_t, \tilde{\mathbf{X}}_t(m)\right\}_{t=1}^P$ to estimate $\beta_P(m)$ and predict y_{P+1} by $f_{P+1}^{(m)} = \tilde{\mathbf{X}}_{P+1}(m)' \hat{\beta}_P(m)$. For the next prediction $f_{P+2}^{(m)}$, the additional time series observations $\left\{y_{P+1}, \tilde{\mathbf{X}}_{P+1}(m)\right\}$ are incorporated to update the estimate $\hat{\beta}_{P+1}(m)$. This procedure continues until $f_T^{(m)}$ is obtained. The third one is the rolling forecast with fixed window size R , $\left\{y_s, \tilde{\mathbf{X}}_s(m)\right\}_{s=t}^{t+R-1}$, $t = 1, \dots, T - R - 1$; $m = 1, \dots, M$ to estimate $\beta_t(m)$, $t = 1, \dots, T - R - 1$ and make forecasts $f_{t+1}^{(m)} = \tilde{\mathbf{X}}_t(m)' \hat{\beta}_t(m)$, $t = P, \dots, T - 1$; $m = 1, \dots, M$.

2.2 Information Combination

We consider two approaches to combine information.

2.2.1 Model Selection Approaches

Given that complete information at time t , $\{y_s, \mathbf{X}_s\}_{s=1}^t$ is available, model selection approaches – *AIC*, *AICC* or *BIC* can be used to select the best predictive model by choosing m of the \mathbf{X}_t variables to predict y_t . Under fixed window scheme, the OLS estimate $\hat{\beta}_P(m)$ is derived based on $\left\{y_s, \tilde{\mathbf{X}}_s(m)\right\}_{s=1}^P$. With a fixed window, the model selection approach will pick a particular model m that minimizes the respective selection criterion and form the forecast sequence $\left(f_{P+1}^{(m)}, \dots, f_T^{(m)}\right)$. The model selection criterion may pick different models at each time point t for the continuously updating or the rolling window scheme. Denote the selected model at time t as m_t , $t = P + 1, \dots, T$. Thus the forecast sequence becomes $\left(f_{P+1}^{(m_{P+1})}, \dots, f_T^{(m_T)}\right)$. The sample estimated mean squared forecast error, *MSFE* is computed as

$$MSFE = \frac{1}{T - P} \sum_{t=P+1}^T (y_t - f_t)^2. \quad (2)$$

Remark: This model selection is different from the one in [17] in that they use *AIC* to select

among competing forecasts. In our paper, given N explanatory variables, there are $M = 2^N - 1$ possible models. However, given finite time series observations T with $T < M$, it is impossible to use AIC to select competing forecasts among M models due to lack of degree of freedom. Therefore, we use AIC to select X s from the information set to form forecasts. We shall use another trimming procedures as stated in the next sub-section.

2.2.2 Simple Average

Simple Average of the $(2^N - 1)$ possible predictive models is used as the predictor of y_t .

2.3 Forecasts Combination

Most of the times, we only observe a set of forecasts instead of the underlying information. In this section, we consider a variety of methods of combining the available forecasts. However, if all the available predictive models are included, methods that require estimation of prediction covariance matrix or regressing y on the available forecasts may break down due to the shortages of degrees of freedom. Therefore, two different trimming procedures are proposed. One for fixed window scheme and continuously updating scheme, another for rolling window scheme.

Definition: Given N variables, we categorize the M models into N class of models M^j where j of the regressors from the information set \mathbf{X}_t are included in the forecasting model. For example, those models with only 1 regressor and a constant are categorized as M^1 , models with 3 regressors of \mathbf{X}_t and a constant are categorized as M^3 . So, we have N classes of models, M^1, \dots, M^N . Within each class of model M^j , the model with highest (average) likelihood is chosen.

Under the fixed window scheme and continuously updating scheme, the likelihood of the first regression is compared. The likelihood of the first regression is defined as

$$L^1(m) = -\ln \left(\frac{\hat{\mathbf{e}}(m)' \hat{\mathbf{e}}(m)}{P} \right) - \frac{1}{2} (P - j), \quad m \in M^j, \quad (3)$$

where $\hat{\mathbf{e}}(m)$ is the $P \times 1$ OLS residuals associated with the first regression. Within each class M^j , the model with the highest likelihood is chosen. So, there are N models remained after trimming.

Under rolling forecasting procedure, average likelihood is applied to select models within each class. Take model 1 for instance, the first regression using $\left\{ y_t, \tilde{\mathbf{X}}_t(1) \right\}_{t=1}^P$ gives likelihood $L^1(1)$,

the second regression using $\{y_t, \tilde{\mathbf{X}}_t(1)\}_{t=2}^{P+1}$ gives likelihood $L^2(1)$, ..., the P^{th} regression gives $L^P(1)$. Suppose we use the first P likelihoods to select one particular model from each class, the average likelihood is defined as

$$\bar{L}(m) = \frac{1}{P} \sum_{j=1}^P L^j(m). \quad (4)$$

The reason for using average likelihood instead of the first likelihood as in fixed window scheme or recursive forecasting scheme is that different windows are used to estimate $\beta_t(m)$. Since the initial time series observations are included in estimating $\beta_t(m)$ under continuously updating forecasting framework, the model selected by comparing the first likelihood will be the same as the model selected by comparing the average likelihood to start with. The average likelihood trimming procedure also gives N models remained, one from each class M^j . Stacking the N forecast sequences $(f_{P+1}^{(m)}, \dots, f_T^{(m)})$ gives matrix \mathbf{F} , which is a $(T - P) \times N$ matrix.

2.3.1 Bates and Granger (1969)

Bates and Granger [2] proposed to use the weights that are inversely proportional to the out-of-sample forecast error variances.

$$w_{BG,t}(m) = \frac{\sigma_t^2(m)^{-1}}{\sum_{j=1}^M \sigma_t^2(j)^{-1}}. \quad (5)$$

However, the population prediction error variance $\sigma_t^2(m)$ is not observed. $\sigma_t^2(m)$ has to be replaced by the maximum likelihood estimate $\hat{\sigma}_t^2(m)$. Since this method requires information about out-of-sample forecast error variances, we have to select, say L , forecasts as an estimation of the prediction covariance matrix. Out of the N models, the prediction error covariance matrix S is formed as follows. Under the continuously updating forecasting scheme, given a sequence of forecasts $(\mathbf{f}_{P+1}, \dots, \mathbf{f}_T)$, we use the first L forecasts $(\mathbf{f}_{P+1}, \dots, \mathbf{f}_{P+L})$ to estimate S_{P+L} . Based on this covariance matrix, the Bates and Granger weight, $w_{BG,P+L+1}(m)$, can be obtained using its diagonal terms as in (5), where $m = 1, \dots, N$. The combined forecast will be $y_{P+L+1}^c = \mathbf{F}(P+L+1)\mathbf{w}_{BG,P+L+1}$, where $\mathbf{F}(s)$ is the row of \mathbf{F} corresponds to forecasts at time s . Then in the next step, $(\mathbf{f}_{P+1}, \dots, \mathbf{f}_{P+L+1})$ is used to estimate S_{P+L+1} . An updated weight $w_{BG,P+L+2}(m)$ and thus the combined forecast $y_{P+L+2}^c = \mathbf{F}(P+L+2)\mathbf{w}_{BG,P+L+2}$ are obtained. Thus, the prediction error covariance matrix

under recursive forecasting scheme is defined as

$$S_t = \frac{1}{P+t} [1'_N \otimes \mathbf{y}(P+1:t) - \mathbf{F}(P+1:t)] [1'_N \otimes \mathbf{y}(P+1:t) - \mathbf{F}(P+1:t)]', t = P+L, \dots, T-1, \quad (6)$$

where $\mathbf{y}(P+1:t) = (y_{P+1}, \dots, y_t)'$, $\mathbf{F}(P+1:t)$ is the rows corresponding to the forecasts (f_{P+1}, \dots, f_t) , $1'_N$ is the row vector of 1s of N dimension and \otimes is Kronecker product.

In the rolling forecast framework, the prediction covariance matrix is defined as

$$S_t = \frac{1}{P} [1'_N \otimes \mathbf{y}(t-P+1:t) - \mathbf{F}(t-P+1:t)] [1'_N \otimes \mathbf{y}(t-P+1:t) - \mathbf{F}(t-P+1:t)]', t = P+L, \dots, T-1, \quad (7)$$

That is, the first combined forecast is obtained by using the first L forecasts $(\mathbf{f}_{P+1}, \dots, \mathbf{f}_{P+L})$ for estimating S_{P+L} , then the second combined forecast is based on $(\mathbf{f}_{P+2}, \dots, \mathbf{f}_{P+L+1})$ to get S_{P+L+1} and so on.

2.3.2 Three Approaches of Granger and Ramanathan

Granger and Ramanathan [11] compared three approaches to select the weights by minimizing the sum of squared forecast errors from the combination forecast. The first method is an unrestricted regression of y_t on all available forecasts without constant term

$$y_t = \mathbf{F}(t) \boldsymbol{\alpha} + u_{1,t}, t = P+1, \dots, P+L. \quad (8)$$

The second method aiming at generating unbiased forecast restricts the sum of weights attached to each forecast to be one

$$y_t = \mathbf{F}(t) \boldsymbol{\alpha} + u_{2,t}, t = P+1, \dots, P+L \text{ and } \boldsymbol{\alpha}' \mathbf{1}_N = 1. \quad (9)$$

The third method of combining has no restrictions on the weights, but a constant term is added

$$y_t = \tilde{\mathbf{F}}(t) \tilde{\boldsymbol{\alpha}} + u_{3,t}, t = P+1, \dots, P+L, \quad (10)$$

where $\tilde{\mathbf{F}}$ includes the columns of 1s on top of the N forecasts. Similar to the estimation of prediction error covariance matrix, the continuously updating forecasting scheme augments the regression by the new forecasts while rolling scheme keeps dropping initial forecasts and adding new forecasts to maintain fixed window size. Both the *BG* and *GR* methods are based on fixed number of predictive models.

2.3.3 Simple Averaging

Equal weight, $\frac{1}{N}$, is assigned to the N predictive models.

2.3.4 Mean Corrected Simple Averaging

Some or all N predictive models could be biased. To correct for possible bias, we suggest a mean corrected simple averaging,

$$\hat{y}_t = \mu + \bar{y}_t \quad , \quad (11)$$

where \bar{y}_t denotes the simple averaging predictor for y_t . The mean μ is obtained as the average of $(y_t - \bar{y}_t)$ for $t = P + 1, \dots, P + L$ for fixed window predictor. For continuously updating or rolling window, μ is updated with the sample under consideration.

2.3.5 Mean and Scale Corrected Simple Averaging

In addition to consider correcting the bias by adding an intercept to the simple averaging predictor, we can also make a scale correction by considering the predictive model

$$\hat{y}_t = \mu + c\bar{y}_t \quad , \quad (12)$$

where the mean μ and scale c are obtained by regressing y_t on a constant and \bar{y}_t .

Remark: The simple averaging can be viewed as a special case of (9) where $\alpha = (\frac{1}{N}, \dots, \frac{1}{N})'$. The mean corrected simple averaging (11) or mean and scale corrected simple averaging can be considered as special cases of (10) with $\tilde{\alpha} = (\mu, \frac{1}{N}, \dots, \frac{1}{N})'$ or $\tilde{\alpha} = (\mu, \frac{c}{N}, \dots, \frac{c}{N})'$ respectively. All three averaging methods may be viewed as constrained regression with slope constraints. In other words, they use less sample information but more prior information.

2.3.6 Bayesian Averaging

To remove the error arising from model uncertainty associated with selecting one particular model out of N models using the model selection criterion, all the models are given a weight generated by the posterior odds as in Buckland, Burnham and Augustin [3]. For example, the weight for model m under AIC is

$$w(m) = \frac{\exp\left(-\frac{1}{2}\Delta AIC_m\right)}{\sum_{j=1}^N \exp\left(-\frac{1}{2}\Delta AIC_j\right)}, \quad m = 1, \dots, N, \quad (13)$$

where $\Delta AIC_i = AIC_i - \min_j (AIC_j)$. Similar weights are assigned to each model m under AICC and BIC with ΔAIC_i replaced by $\Delta AICC_i$ and ΔBIC_i respectively.

3 Monte Carlo Studies

In this section we evaluate the performance of various approaches by considering a small scale Monte Carlo simulation with 500 replications. The variables (y_t, \mathbf{X}'_t) are generated by a common factor structure model with number of factors $K = 1$.

$$\begin{pmatrix} y_t \\ \mathbf{X}_t \end{pmatrix} = \begin{matrix} B \\ (q+1) \times K \end{matrix} F_t + U_t, \quad t = 1, \dots, T, \quad (14)$$

where the idiosyncratic error vector \mathbf{U}_t is drawn from *iid* $N(0, \mathbf{I}_{q+1})$. We let $q = 5$ and $T = 500$. The first forecast starts in 61, i.e. $P = 60$. The rolling window size is $R = P = 60$. We use *AIC*, *AICC* and *BIC* to denote the *MSFE* of the model selection approaches as in (2); *MAAIC*, *MAAICC* and *MABIC* to denote the outcome based on the weight of (13); *FSA* and *HLSA* to denote the outcome based on simple averaging over all possible models and simple averaging over the remaining models after trimming; *BG*, *GR1*, *GR2* and *GR3* are the four forecast combination methods using (5), (8) to (10); *MCSA* and *MSCSA* to denote regression approach of (11) and (12). Table 1 shows the simulation results. Since the data generating process is stationary with zero mean, all predictive models are unbiased. Therefore, it is expected that the fixed and continuously updating scheme should perform better than rolling forecast. The simulation results confirms this. Averaging over information generated forecasts using posterior odds under fixed and continuously

updating scheme help to reduce the $MSFE$ over the method of forecasting using a single model. Because all the predictive models are unbiased, $GR3$ actually performs worse than $GR1$ or $GR2$, so are $MCSA$ and $MSCSA$ compared to simple averaging.

4 Empirical Application

In this section, we use 9 variables chosen by Pesaran and Timmermann [14] to predict excess equity premium on the S&P 500 index. These are one-month lagged dividend yield DY_{t-1} and earning price ratio EP_{t-1} , one-month lagged and two-month lagged Treasury Bill rate TB_{t-1}, TB_{t-2} , one-month lagged and two-month lagged long-term government bond yield GB_{t-1}, GB_{t-2} , two-month lagged inflation Π_{t-2} , two-month lagged annual growth rate of Industrial Production ΔIP_{t-2} and annual growth rate of $M2$ with two months lag ΔM_{t-2} . Monthly data run from 1960M3 to 2008M12. The variable of interest y_t is the excess equity premium on the S&P 500 index.

4.1 Data Sources

Index Prices: This is the Standard & Poor's 500 Index at the closing price on the last trading day of each month obtained in CRSP. Since some of the macro variables have data only after 1959M1, this variable also starts at the same period. The y variable, excess return of S&P, is calculated by $(P_t + D_t - P_{t-12}/P_{t-12}) - TB_{t-1}$.

Dividend Yield ($DY_t = \log D_t - \log P_{t-1}$): This is the difference between the log of dividends and log of lagged prices. Dividends are 12-month moving sums obtained from Welch and Goyal [9].

1M T-bill rate (TB_{t-1}, TB_{t-2}): This is the risk-free rate and will be subtracted from the stock return for calculating the excess return. Data from International Financial Statistics (IFS) is percent per annum.

M2 Seasonally Adjusted ($\Delta M_t = (M_t - M_{t-12})/M_{t-12}$): This month-end monetary aggregate variable as one of the business component is also from IFS. It starts from 1959M1.

Inflation (Π_{t-2}): The inflation rate is computed by producer price index also obtained from IFS.

Rate of Industrial Production ($\Delta IP_t = (IP_t - IP_{t-12})/IP_{t-12}$): This is based on seasonally adjusted index for industrial production downloaded from IFS.

Earnings Price Ratio ($EP_t = \log E_t - \log P_t$): This is the difference between the log of earnings and log of prices. Earnings are 12-month moving sums of earnings on the S&P 500 index. Data are from Welch and Goyal (2008).

Government Bond (GB_{t-1}, GB_{t-2}): This is obtained from IFS.

4.2 Parameters

This empirical studies cover data from 1960Q3 to 2008Q12. Throughout the studies, the forecasts start at time point $P + 1 = 121$. Under the rolling forecasting scheme, the window size is $R = 120$, i.e. 10 years. Since we have 9 variables, the total number of models is $2^9 - 1 = 511$ models. However after the proposed trimming procedures, only $N = 9$ models remains. For those model selection approaches AIC , $AICC$, BIC and the posterior-odd counterparts $MAAIC$, $MAAICC$, $MABIC$ and simple averaging FSA , all models M are considered. However for those Granger methods, trimming procedures are applied to better estimate the prediction covariance matrix.

The results are shown in Table 2 and forecasting paths are plotted against the truth in Figures 1 to 3 under fixed, rolling and continuously updating forecasting framework. S&P500 index during the period from 1960M3 to 2008M12 was subject to several big hits and structural changes, for instances, Arab Oil Embargo in 1973, Gulf War in 1991, the dot.com bubble burst in 2000, the subprime crisis in 2007, the collapse of Lehmann Brothers in 2008. Existing forecasting models may at most predict accurately in a particular period. Therefore, forecast combination can play an important role here. Furthermore, because of the instability of the excess equity premium on S&P Index, the predictions made under fixed window scheme performs worse than predictions under continuously updating forecasting method and rolling forecasting scheme is expected have the best performance.

Since only the first 120 time series observations of $\{y_t, \mathbf{X}_t\}$ are used to form forecasts for the remaining path upto 2008M12 under the fixed window scheme, the forecast performance is unsatisfactory due to the highly volatile return. No particular optimal methods work under this scheme. Although $GR3$ is expected to be better than $GR1$ and $GR2$ as proved in [11], under fixed window scheme, this result doesn't hold anymore. Another striking result is that if we eliminate the possible bias in the simple averaging forecasts by including a constant term in the regression, the methods of $MCSA$ and $MSCSA$ substantially reduce the $MSFE$ and become the best methods.

When the forecasts are updated continuously, it also leads to largely improved performance over the fixed window scheme. Rolling scheme further lowers the $MSFE$ over the continuously updating scheme. Methods using less information of the forecasts, like BG and $MSCSA$ outperform other forecast combination methods. Graphically, those simple averaging methods produce less volatile forecasts. The rolling scheme traces the actual path more closely than the other two schemes.

5 Concluding Remarks

In this paper we have reviewed some popular forecast combination methods and suggested two additional forecast combination methods – a mean corrected simple averaging and a mean and scale corrected simple averaging methods. We have evaluated the predictive performance of these methods through a small scale Monte Carlo and an empirical application to the prediction of excess premium.

When there is no structural change, there is no room for forecasting averaging. Model selection criterion should be used to select the best predictive model. If forecast combination methods are used, the Bayesian averaging method appears to dominate classical sampling method unless the sample is very large. Averaging methods using complete sample information (Granger and Ramanathan ($GR1$)) appear dominate methods using less sample information (BG , $HLSA$, $MCSA$ and $MSCSA$). On the other hand, if there are possibilities of structural change, the optimal weighting methods based on a given structure are no longer optimal. Therefore, those methods that appear to use more sample information actually perform worse than the approaches of using less sample information and rolling windows appear to dominate fixed window as in the case of predicting excess premiums. Moreover, the mean corrected or the mean and scale corrected simple averaging methods perform much better than the simple averaging method because they are more robust against bias and possible misspecifications of predictive models.¹

¹In our empirical application, the simple average actually performs well. This is because all the predictive models contain an intercept. So in principle, they are all unbiased predictors. Therefore, there is no need to consider a mean corrected simple averaging predictor. However, if some predictive models are biased, then a mean corrected or mean and scale corrected simple averaging could do better than simple averaging.

References

- [1] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: B.N. Petroc and F. Csake (Eds.), Second International Symposium on Information Theory, Budapest, 1973, pp. 267-281.
- [2] J.M. Bates, C.W.J. Granger, The combination of forecasts, *Operations Research Quarterly* 20 (1969) 451-468.
- [3] S.T. Buckland, K.P. Burnham, N.H. Augustin, Model Selection: An integral part of inference, *Biometrics* 53 (1997) 603-618.
- [4] Y.L. Chan, James H. Stock and Mark W. Watson, A Dynamic Factor Model Framework for Forecast Combination, *Spanish Economic Review*, 1 (1999) 91-121.
- [5] Y.Y. Chong and David F. Hendry, Econometric Evaluation of Linear Macro-Economic Models, *Review of Economic Studies*, 53 (1986) 671-690.
- [6] Robert T. Clemen, Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting*, 5 (1989) 559-583.
- [7] Diebold F.X. (1989), Forecast Combination and Encompassing: Reconciling two divergent literatures, *International Journal of Forecasting*, 5, 589-592.
- [8] F.X. Diebold and Peter Pauly, The use of prior information in forecast combination, *International Journal of Forecasting*, 6 (1990) 503-508.
- [9] Goyal, A., and I. Welch (2008), A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *The Review of Financial Studies* 21 (4): 1455-1508.
- [10] Granger, C.W.J., King M.L. and White H. (1995), Comments on testing economic theories and the use of model selection criteria, *Journal of Econometrics*, 67, 173-187.
- [11] C.W.J. Granger, R. Ramanathan, Improved methods of combining forecast accuracy, *Journal of Forecasting* 19 (1984) 197-204.

- [12] H.Y. Huang and T.H. Lee, To Combine Forecasts or to Combine Information? Working Paper (2009).
- [13] C.M. Hurvich, C.L. Tsai, Regression and Time Series Model Selection in Small Samples, *Biometrika*, 76 (1989) 297-307.
- [14] M.H. Pesaran and Allan Timmermann, Predictability of Stock Returns Robustness and Economic Significance, *The Journal of Finance*, 1 (4) (1995) 1201-1228.
- [15] G. Schwartz, Estimating the dimension of a model, *Annals of Statistics* 6 (1978) 461-464
- [16] James H. Stock and Mark W. Watson, Combination Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, 23 (2004) 405-430.
- [17] Norman R. Swanson and Tian Zeng, Choosing Among Competing Econometric Forecasts: Regression-based Forecast Combination Using Model Selection, *Journal of Forecasting*, 20 (2001) 425-440.
- [18] Allan Timmermann, Forecast combinations, in: G. Elliott, C.W.J. Granger, A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Elsevier, Amsterdam, 2006, pp. 135-196.
- [19] K. West and Michael W. McCracken , Regression-Based Tests of Predictive Ability, *International Economic Review*, 39 (4) (1998) 817-840.
- [20] Jonathan Wright, Forecasting US Inflation by Bayesian Model Averaging, *Journal of Forecasting*, 28 (2009) 131-144.

Table 1 MSFE of Prediction Error (N=5, T=500, Window Size = 60, Replications = 500)

	Fixed	Continuous	Rolling
AIC	1.3104	1.0983	1.1087
AICC	1.3096	1.0987	1.1094
BIC	1.3117	1.1075	1.1207
MAAIC	1.2944	1.0890 (2)	1.0782 (2)
MAAICC	1.2930	1.0891 (3)	1.0782 (3)
MABIC	1.2903 (5)	1.0957 (5)	1.0847 (4)
FSA	1.3068	1.0884 (1)	1.0780 (1)
HLSA	1.2856 (4)	1.0943 (4)	1.0802 (5)
BG	1.2841 (3)	1.0890 (2)	1.0867
GR1	1.3177	1.1066	1.1453
GR2	1.3400	1.1207	1.1468
GR3	1.3209	1.1129	1.1621
MSCSA	1.2812 (1)	1.1015	1.1159
MCSA	1.2826 (2)	1.0958	1.3695

Number in brackets is the ranking.

Table 2 MSFE of Prediction Error for Excess Equity Premium on S&P 500 (Window Size = 120)

	Fixed	Continuous	Rolling
AIC	2020.949	269.7009	199.5294
AICC	2020.949	269.3486	200.3953
BIC	1799.673	269.4487	202.485
MAAIC	1964.731	266.0164	187.2275 (4)
MAAICC	1958.068	265.9405	186.3489 (3)
MABIC	1999.251	260.0280 (5)	179.4222 (2)
FSA	1530.865 (5)	246.8493 (1)	192.4176
HLSA	2139.737	251.8969 (2)	178.8275 (1)
BG	1866.473	274.4895	191.7456
GR1	650.2812 (3)	264.0966	206.8819
GR2	439.5278 (2)	297.0089	215.8956
GR3	994.7089 (4)	255.6030 (3)	214.0374
MSCSA	310.2087 (1)	258.6374 (4)	187.9163 (5)
MCSA	3592.579	297.9962	200.9472

Number in brackets is the ranking.

Figure 1 Actual versus Predicted Path under Fixed Window Prediction Scheme (1960M3 – 2008M12)

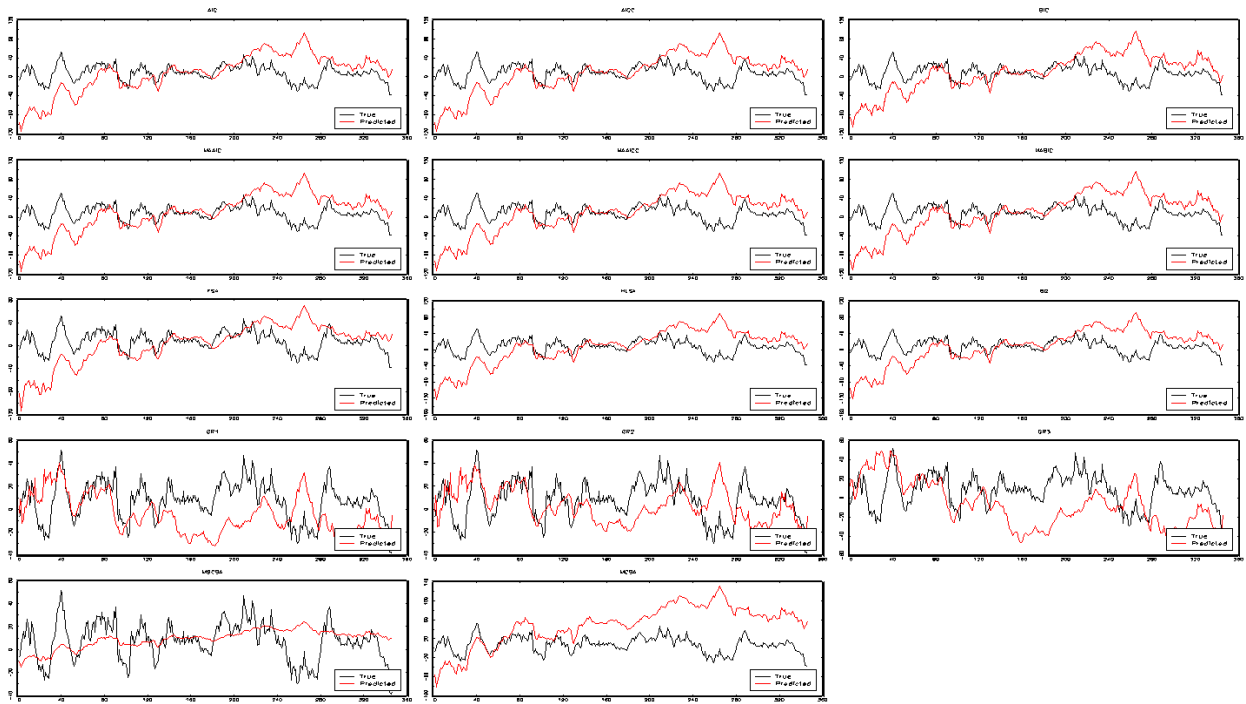


Figure 2 Actual versus Predicted Path under Continuously Updating Scheme (1960M3 – 2008M12)

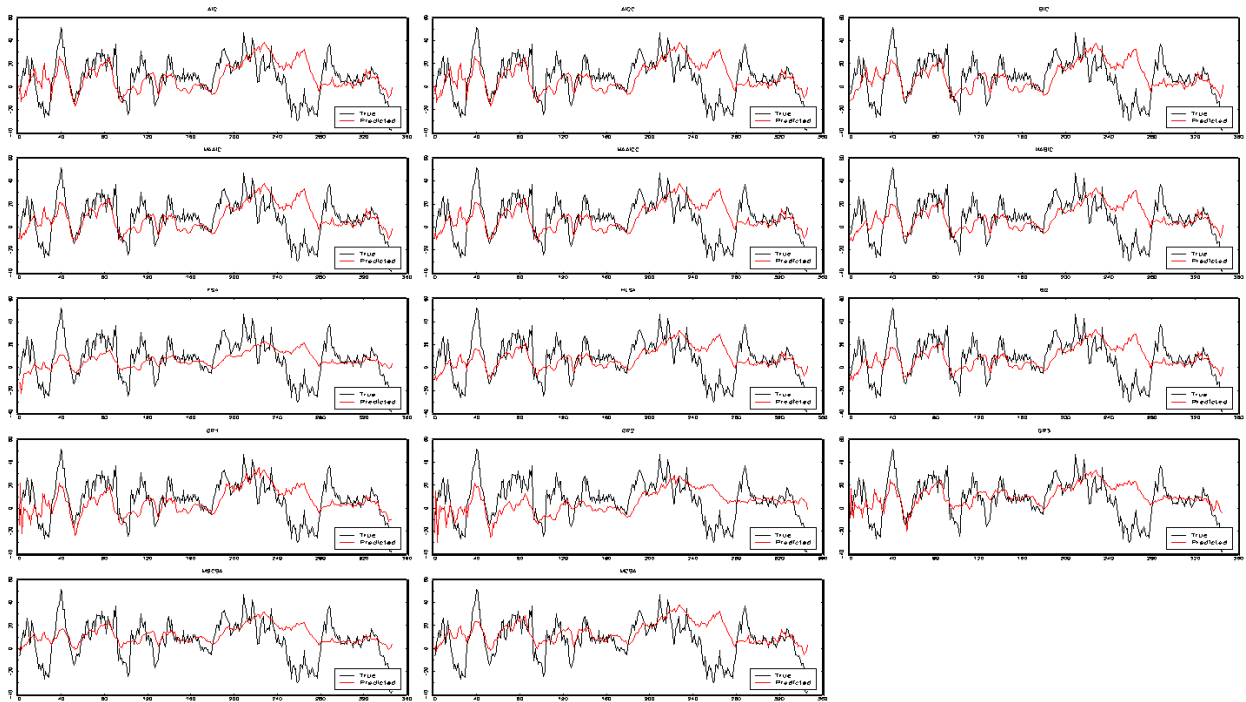


Figure 3 Actual versus Predicted Path under Rolling Prediction Scheme (1960M3 – 2008M12)

