HONG KONG MONETARY AUTHORITY

Working Paper 01/2008

10 March 2008

PREDICTING STOCK MARKET RETURNS BY COMBINING FORECASTS

Prepared by Laurence Fung and Ip-wing Yu[1]

Market Research Division

Research Department

Abstract

The predictability of stock market returns has been a challenge to market practitioners and financial economists. This is also important to central banks responsible for monitoring financial market stability. A number of variables have been found as predictors of future stock market returns with impressive in-sample results. Nonetheless, the predictive power of these variables has often performed poorly for out-of-sample forecasts. This study utilises a new method known as "Aggregate Forecasting Through Exponential Re-weighting (AFTER)" to combine forecasts from different models and achieve better out-of-sample forecast performance from these variables. Empirical results suggest that, for longer forecast horizons, combining forecasts based on AFTER provides better out-of-sample predictions than the historical average return and also forecasts from models based on commonly used model selection criteria.

JEL Classification Numbers: G11; G12; C13

Keywords: Forecasting; Model combination; Model uncertainty

Author's E-Mail Address:

Laurence_KP_Fung@hkma.gov.hk; Ip-wing_Yu@hkma.gov.hk

[1] The authors acknowledge the comments from Hans Genberg and Cho-hoi Hui.

Executive Summary:

- *The predictability of stock market returns has been a challenge to market practitioners and financial economists. While a large number of financial, macroeconomic and valuation variables have been found to have good in-sample predictive power, they often perform poorly for out-of-sample forecast.*

- *This study applies a new model combination method known as "**A**ggregate **F**orecasting **T**hrough **E**xponential **R**e-weighting (AFTER)" to examine its out-of-sample forecast performance. The method is used to generate out-of-sample return forecasts for eight stock market indices in six economies – Hong Kong, Japan, the US, the UK, France and Germany.*

- *The empirical results based on the forecast evaluation tests suggest that for a short (one month) horizon, forecasts derived from AFTER are "as good as" those from the historical average return, a commonly used benchmark. For longer forecast horizons (six months and longer), forecasts from AFTER are in general able to outperform the benchmark and sometimes those based on model selection criteria.*

- *In view of its relatively better performance, the AFTER forecast could be used with caution as a reference if there is a need to know the long-term performance of the stock market in policy makers' monitoring work.*

I. INTRODUCTION

The predictability of stock market returns has been a challenge to market practitioners and financial economists. This is also important to central banks responsible for monitoring financial stability. On the one hand, numerous studies have proposed various financial, macroeconomic and stock valuation variables and found them able to predict subsequent stock market returns for long horizons.[2] On the other hand, several studies express concerns about spurious results based on these variables (Nelson and Kim (1993) and Stambaugh (1999)) and cast doubts on the evidence of return predictability, given that the out-of-sample predictions have often performed poorly (Bossaerts and Hillian (1999) and Goyal and Welch (2006)).[3]

The performance of out-of-sample forecasts also varies with the models and variables used over time. Given the large number of potential variables that are relevant to predict stock market return, there is little consensus on what kinds of models have the best predictive power. The use of different statistical model selection criteria does not help in choosing the single "best" model from these candidates.[4] Facing the model uncertainty and/or model instability, taking an average over a number of forecasts from different models might improve the out-of-sample forecast performance.[5]

This study addresses the model uncertainty issue and examines the out-of-sample predictability of stock market returns by using a new model combination method known as "**A**ggregate **F**orecasting **T**hrough **E**xponential **R**e-weighting (AFTER)" introduced by Yang (2004). The weighting scheme in AFTER takes into account prominent features of financial data such as the time-varying volatility with simple implementation procedure (Hagmann and Loebb (2006)). We apply the AFTER weighting algorithm to examine the out-of-sample return forecasts of eight stock market indices from six economies. Empirical results suggest that, for long forecast horizons (six months and longer), combining forecasts based on AFTER provides better out-of-sample predictions than the historical average return and sometimes those provided by commonly used model selection criteria.

---

2 Fama and French (1988) and Campbell and Shiller (1988a, 1988b), among many others, find that variable such as the dividend yield, the price-earnings ratio, the default premium, the term premium and the short-term interest rate forecast excess stock market returns. Several studies suggest other variables like corporate payout and financing activity (Lamont 1998, Baker and Wurgler 2000), and the consumption to wealth ratio (Lettau and Ludvigson 2001, Guo 2006) as predictors of stock market returns.

3 In some recent studies, however, it is shown that the previous in-sample results are not spurious and the negligible out-of-sample predictive power may be due to small sample sizes. For details, see Inoue and Kilian (2004), Campbell and Thompson (2006) and Hjalmarsson (2006).

4 See Bossaerts and Hillion (1999) and Pesaran and Timmermann (1995, 2000).

5 Hendry and Clements (2004) provide formal evidence on the value-added of averaging over several models. Aiolfi and Favero (2005) find that averaging over a subgroup of models significantly improves the quality of the stock return forecast.

This paper contributes to the financial return forecasting literature using the model averaging approach in two aspects: (i) instead of focusing on one particular stock market (i.e. Hagmann and Loebb (2006) on S&P 500 Index), this study applies AFTER to eight stock market indices from six different economies; and (ii) while other studies only conduct one-period ahead forecast, we apply forecast combination to various forecast horizons up to 24 months.

The remainder of this paper is organised as follows. In Section II we provide an overview of model combination and a description of Yang's AFTER algorithm. Section III discusses the stock markets covered in the study, the empirical implementation of the AFTER algorithm and the various tests for forecast accuracy comparison. Section IV presents the empirical evidence on the predictability of stock market returns arising from model combination. Section V summarises the results and concludes.

II. MODEL COMBINATION AND AFTER

2.1 Model Combination

Despite the vast interest in studying stock market predictability, there is little consensus on the specification of the "best" predictive model. In picking a model, many studies base on some standard statistical model selection criteria (e.g. the adjusted R^2 and Akaike's information criterion) and/or hypothesis testing. However, a model that meets the selection criteria and has good predictive power relative to others in one subperiod does not usually have the same predictive ability in another subperiod. To capture the strengths of other competing models, it is desirable to average over a number of forecasts from different models. The idea that forecast combination outperforms any individual forecast has attracted attention since it was introduced by Bates and Granger (1969), when they combined two separate sets of forecasts of airline passenger data and found that the composite set of forecasts yielded lower mean-square error than either of the original forecasts. Stock and Watson (2003, 2004) considered inflation and output growth prediction in each of the G7 economies using a large number of possible models. They showed that the best prediction performance is obtained by simply averaging the forecasts from the large number of models.[6]

[6] Hibon and Evgeniou (2005) argued that while forecast combination does not always outperform the best single model, it is less risky in practice to combine forecasts than to select an individual forecasting method.

2.2 The AFTER Algorithm

In this study, we follow Hagmann and Loebb (2006) and use a model combination algorithm called "**A**ggregate **F**orecasting **T**hrough **E**xponential **R**e-weighting (AFTER)" introduced by Yang (2004) to investigate out-of-sample stock market predictability. Hagmann and Loebb find AFTER a better way of averaging forecasts as AFTER takes into account prominent features of financial time series such as time-varying volatility, while in the meantime, it also shares some of the Bayesian interpretation.[7] A key feature in AFTER, as pointed out by Hagmann and Loebb, is that the time-varying weights associated with a member of the models are directly linked to the past out-of-sample performance of that model. This property is attractive in the evaluation of return predictability as it is the out-of-sample prediction performance that matters. In addition, the AFTER algorithm is simple and easy to implement. Hangmann and Loebb (2006) implement AFTER for predicting S&P 500 index returns and find that AFTER significantly beats the constant, unconditional benchmark model.

The following section describes Yang's AFTER algorithm. Suppose that future returns on stock market r_{t+1} are forecastable using a simple linear regression specification by a set of k financial and macroeconomic indicators available at time t. With these k variables, there are 2^k different competing linear models. Each of these models is given by:

$$r_{t+1} = \alpha_j + \beta_j^{'} x_{j,t} + \varepsilon_{t+1} \tag{1}$$

where j is a model-specified indicator, $x_{j,t}$ is a model-unique subset of n variables, ε_{t+1} is normally distributed with mean zero and standard deviation $\sigma_{j,t+1}$. The parameter n ranges between zero and k. When $n = 0$, future returns are assumed to be constant. When $n = k$, all k variables are included in the model. We assume that the forecast of future return is obtained by averaging over the resulting forecasts from the 2^k competing models:

$$\hat{r}_{t+1}^{*} = \sum_{j=1}^{2^k} w_{j,t} \hat{r}_{j,t+1} = \sum_{j=1}^{2^k} w_{j,t} (\hat{\alpha}_j + \hat{\beta}_j^{'} x_{j,t}) \tag{2}$$

---

[7] As to be shown later, the posterior probability assigned to the model in the AFTER framework depends on the out-of-sample forecast performance of the model. This is updated over time when more data are available.

where $\hat{\alpha}_j$ and $\hat{\beta}_j$ are coefficient estimates for model *j* obtained from ordinary least squares estimation. Yang (2004) proposes to choose the model weight ($w_{j,t}$) in Equation (2) at time *t* as:

$$w_{j,\,t} = \frac{w_{j,\,t-1}\hat{\sigma}_{j,\,t}^{-1}\exp(-(r_t - \hat{r}_{j,\,t})^2 / 2\hat{\sigma}_{j,\,t}^2)}{\sum_{j'=1}^{2^k} w_{j',\,t-1}\hat{\sigma}_{j',\,t}^{-1}\exp(-(r_t - \hat{r}_{j',\,t})^2 / 2\hat{\sigma}_{j',\,t}^2)} \qquad (3)$$

where $\hat{r}_{j,t}$ and $\hat{\sigma}_{j,\,t}^2$ are forecast return of model *j* and its variance respectively, which are available at time *t*-1. The weight of each competing model is updated on the basis of past out-of-sample performance. As mentioned in Yang (2004), the weighting scheme in Equation (3) has a Bayesian interpretation. If we view $w_{j,\,t-1}$ as the prior probability of model *j* before observing the actual r_t, then $w_{j,\,t}$ can be treated as the posterior probability of this model after r_t is known. Hence, this posterior probability depends on the out-of-sample forecast performance of $\hat{r}_{j,\,t}$ which was available at time *t*-1. The idea of AFTER is to assign more weight for those models that have been predicting more accurately in the past.[8]

III. DATA AND EMPIRICAL ESTIMATION

3.1 The Data

The empirical examination uses monthly observations on eight benchmark stock market returns from six economies. Table 1 highlights these benchmark stock market indices.

Table 1. Benchmark Stock Market Indices

Stock Market	Benchmark Index
Hong Kong	Hang Seng Index (HSI), Hang Seng China Enterprises Index (H-share)
Japan	TOPIX Index (TOPIX)
US	Dow Jones Industrial Average (DJIA), S&P 500 (SP500)
UK	FTSE 100 Index (FT100)
Germany	DAX 30 Index (DAX)
France	CAC 40 Index (CAC)

Source: Bloomberg

[8] For more on the theoretical properties of AFTER, see Yang (2004).

In order to avoid the data-snooping bias of exhaustively including as many potential variables into different models as possible, we include only those financial and macroeconomic variables, that are found to be important in previous studies on return predictability (see Cremers (2002) for an overview of variables used in past research studies), in the linear regression specification (as in Equation (1)).[9] In particular, at time t and for a forecast horizon of h-period ahead, the following eight variables are used:[10]

- The stock market return at time t-h
- The dividend yield of the respective stock market index at time t-h
- The price-earnings ratio of the respective stock market index at time t-h
- The inflation rate (year-on-year change in the consumer price index) at time t-h
- The month-on-month change in the yield of 3-month government instrument at time t-h
- The term premium measured by the difference between the yields of 10-year and 3-month government instruments at time t-h
- The interest differential between the overnight and 3-month interbank rates at time t-h
- The 'Fed Model':[11] it is the spread between the earnings yield of the stock market index and the yield of 10-year government instrument at time t-h

3.2 Empirical Estimation

3.2.1. Sample period, setting of initial weight and model specification

Monthly data from January 1970 to July 2007 are covered in the empirical analysis.[12] In this study, we consider four different forecasting horizons, namely one month, six months, one year and two years. The setup of the linear regression specification in Equation (1) becomes:

$$r_{t+h} = \alpha_j + \beta_j^{'} x_{j,t} + \varepsilon_{t+h} \qquad (4)$$

where r_{t+h} is the h-month ahead log return of the stock market index, h is the forecast horizon and $x_{j,t}$ is a model-unique subset of n variables available at time t. Given that

9 There is no consensus on which predictive variables should be included as each study focuses on a particular set of predictive variables to forecast index returns.
10 Data of these variables are from Bloomberg, CEIC and Datastream.
11 The 'Fed Model' is discussed in the Board of Governors of the Federal Reserve System (1997).
12 Note that due to data availability, not all the data series of the predictive variables start from January 1970.

there are eight variables, a combination of these variable results in a total of 256 individual models in the form of Equation (4).

The forecast comparison is based on out-of-sample forecasts from January 1999 to July 2007. The period from January 1970 to December 1998 is used for calculating initial parameter estimates of the 256 regression models. For each forecasting month, we re-estimate all 256 regression models with an expanding window resulting in a total of 103 out-of-sample forecasts as well as their variances. The time series of forecasts and variances are then used to generate a sequence of model weights according to Equation (3) of the AFTER algorithm. With a sequence of model weights for each of the 256 models, a time series of combined out-of-sample forecasts can be derived using Equation (2).

To implement Equation (3) of the AFTER algorithm, some starting weights are chosen at the very beginning of the forecasting period. Since the "true" set of parameters is seldom known and we do not have "a prior" knowledge on which variable (or model) is more informative in stock market forecasting, a simple way is to assign the same weight to each of the 256 models at the beginning.[13]

Finally, a measure of the error variance for each model is needed to implement the AFTER algorithm of Equation (3). As it is well known that financial time series are most likely to have time-varying volatility, each of the 256 models in the form of Equation (4) is estimated under a simple GARCH(1,1) specification with the conditional variance (σ^2) specified as:

$$\sigma_{t+h}^2 = \omega + \alpha\varepsilon_{t+h-1}^2 + \beta\sigma_{t+h-1}^2 \tag{5}$$

3.2.2. Competing methods

For out-of-sample forecast accuracy comparison, the benchmark is the historical average return (Average) obtained recursively from an unconditional constant model.[14] In addition to the historical average, the forecasts obtained from AFTER are also compared with those obtained from the following methods:

---

[13] We also assign alternative initial weights following Hagmann and Loebb (2006) and we note that their impacts on the forecast comparison results are insignificant. Thus, we only present the empirical results with equal initial weights.

[14] The use of historical average as a benchmark in the forecast comparison is also found in Goyal and Welch (2006) and Campbell and Thompson (2006).

a. model selection based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC): the out-of sample forecast is taken from the single "best" model selected recursively at each out-of-sample forecast period based on the respective information criterion;[15]

b. equal-weight moving averages (EW5): a 5-year moving average of actual returns; and

c. decaying-weight moving averages (DW5): a 5-year moving average with decaying weights based on actual returns[16]

Altogether, the forecasts derived from the AFTER algorithm are compared with forecasts from the benchmark model as well as those from four alternative methods.

3.3 Evaluation of out-of-sample forecast

Out-of-sample forecast is evaluated by several methods and tests, including:

1. Root-mean-square error (RMSE): defined as:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{r}_{i,t} - r_t)^2}$$

where r_t is the actual stock market return at time t, $\hat{r}_{i,t}$ is the out-of-sample forecast of stock market return at time t based on the forecast method i at time $t - h$, and T is the number of effective out-of-sample forecasts. The smaller the RMSE of a model is, the better the model performs.

15 Among all estimated models, the one with the smallest value of AIC (or BIC) is preferred. Please refer to Akaike (1974) and Schwarz (1978) for details of the respective information criterion.

16 The weights are assigned in a decaying function of time. More weight is given to the more recent stock return than to returns in the distant past. The decay rate is 10% per month, i.e. if a weight of 100% is given to the return in the recent month, then a weight of 90% is given to the return in the previous month.

2. The Diebold and Mariano (1995) test statistic (*DM*-test): to statistically compare the forecasts of two models, we perform the *DM*-test using the squared error (SE) loss function with the constant, unconditional historical average as the benchmark model. The squared error loss function is defined as:

$$d_t = (\varepsilon_{i,t})^2 - (\varepsilon_{c,t})^2$$

where $\varepsilon_{i,t}$ is the out-of-sample forecast error from forecast method *i* at time *t*, and ε_c is the out-of-sample forecast error from the historical average model at time *t*. In general, the *DM* statistic is given as

$$DM = \frac{\bar{d}}{SE(\bar{d})}$$

where $\bar{d}$ is the sample mean of the loss function d_t and $SE(\bar{d})$ is the standard error of $\bar{d}$.[17] The *DM* statistic is asymptotically *t*-distributed under the null hypothesis of equal forecast accuracy. A *DM* statistic of 1.96 or larger implies that the difference between the two squared errors is statistically significant.

3. The Theil (1961) U statistic (*U*-statistic): a measure of the degree to which the forecast differs from the actual. The statistics is computed as:

$$U\text{-statistic} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{r}_{i,t} - r_t)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{r}_{i,t})^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t)^2}}$$

The *U*-statistic is bounded between zero and one. The more accurate the forecasts are, the smaller is the value of the *U*-statistic. The model with its *U*-statistic closer to zero is preferred.

---

[17] As mentioned in Diebold and Mariano (1995), $SE(\bar{d})$ has to be a consistent estimate. Therefore, $SE(\bar{d})$ is corrected for the presence of heteroskedasticity and autocorrelation by Newey and West (1987).

4. Direction-of-change measure: defined as the percentage of correct direction forecasts. Two variables are defined as:

$$DC^{*}_{\hat{r}_{i,t}} = \begin{cases} 1 \ \ if \ \ \hat{r}_{i,t} > 0 \\ 0 \ \ otherwise \end{cases}, \text{ and}$$

$$DC_{r,t} = \begin{cases} 1 \ \ if \ \ r_t > 0 \\ 0 \ \ otherwise \end{cases}$$

The performance of forecast method *i* in terms of the percentage of correct direction change (*CDC*) is:

$$CDC = \frac{1}{T}\sum_{t=1}^{T}\Phi\{DC^{*}_{\hat{r}_i,t} = DC_{r,t}\}$$

where Φ takes the value one when $DC^{*}_{\hat{r}_{i,t}} = DC_{r,t}$ and zero otherwise.

5. The Pesaran and Timmermann (1992) non-parametric test (*PT*-test): similar to the *CDC*, the *PT*-test examines whether the directional movements of the actual and forecast returns are in line with one another. The larger the *PT* statistic, the better is the match. To calculate the *PT* statistic, we first define the following variables:

$$X_t = \begin{cases} 1 \ \ if \ \ r_t > 0 \\ 0 \ \ otherwise \end{cases}$$

$$Y_t = \begin{cases} 1 \ \ if \ \ \hat{r}_{i,t} > 0 \\ 0 \ \ otherwise \end{cases}$$

$$Z_t = \begin{cases} 1 \ \ if \ \ r_t \hat{r}_{i,t} > 0 \\ 0 \ \ otherwise \end{cases}$$

Now, the *PT* statistic (*PT*) is defined as:

$$PT = \frac{P - P^{*}}{\sqrt{V(P) - V(P^{*})}} \sim N(0,1)$$

where $P = \sum_{t=1}^{T} Z_t / T$,

$P^* = P_X P_Y + (1 - P_X)(1 - P_Y)$,

$P_X = \sum_{t=1}^{T} X_t / T$,

$P_Y = \sum_{t=1}^{T} Y_t / T$,

$V(P) = T^{-1} P^* (1 - P^*)$, and

$$V(P^*) = T^{-1}(2P_Y - 1)^2 P_X (1 - P_X) + T^{-1}(2P_X - 1)^2 P_Y (1 - P_Y) + 4T^{-2} P_Y P_X (1 - P_Y)(1 - P_X)$$

If the *PT* statistic of a forecast method is greater than 1.96, then the null hypothesis of no predictive performance at the 5% significance level is rejected. For this test, the larger the *PT* statistic is, the more the method is preferred.

IV. RESULTS

4.1 Out-of-sample forecast comparison

In this section, we present the out-of-sample forecast comparison results of different stock markets based on AFTER and other forecasting methods. Tables 2 to 9 report the summary statistics of the six out-of-sample forecast evaluation tests for individual stock markets under different forecast horizons.

4.1.1. Hong Kong Hang Seng Index and H-share Index

Table 2. Forecast Comparison Statistics: Hang Seng Index

	Forecast Horizons			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	6.4	12.0	19.8	24.4
Average	6.1	16.0	26.1	38.9
EW5	6..1	15.0	23.9	34.9
DW5	6.2	13.8	20.7	26.1
AIC	7.2	23.0	20.0	13.9
BIC	6.1	21.6	20.6	14.0
***DM*-statistic** [2]				
AFTER	0.89	-2.33*	-2.67*	-3.06*
EW5	-0.18	-0.56	-0.68	-0.82
DW5	0.48	-0.98	-1.32	-1.96*
AIC	1.00	0.96	-1.40	-3.00*
BIC	-0.48	0.92	-1.20	-3.00*
***U*-statistic** [3]				
AFTER	0.82	0.50	0.46	0.37
Average	0.80	0.61	0.58	0.59
EW5	0.91	0.80	0.78	0.79
DW5	0.81	0.57	0.51	0.44
AIC	0.74	0.66	0.41	0.20
BIC	0.79	0.65	0.42	0.20
CDC [4]				
AFTER	0.61	0.72	0.81	0.87
Average	0.61	0.60	0.67	0.71
EW5	0.53	0.52	0.67	0.59
DW5	0.63	0.72	0.83	0.78
AIC	0.56	0.61	0.78	0.93
BIC	0.58	0.62	0.77	0.95
***PT*-statistic** [5]				
AFTER	0.47	4.17^+	5.58^+	5.19^+
Average	1.00	1.00	1.00	1.00
EW5	0.29	0.12	2.59^+	0.11
DW5	2.01^+	4.07^+	6.23^+	3.76^+
AIC	-0.08	1.87	4.83^+	6.58^+
BIC	-0.18	2.04^+	4.62^+	6.78^+

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. $^+$ indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

Table 3. Forecast Comparison Statistics: H-share Index

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	10.2	24.1	34.9	40.3
Average	9.7	24.4	33.0	53.9
EW5	9.7	24.8	33.2	49.1
DW5	9.8	23.9	28.4	31.2
AIC	11.2	31.9	39.0	30.7
BIC	10.9	30.3	39.5	30.0
***DM*-statistic** [2]				
AFTER	1.31	-0.12	0.55	-2.81*
EW5	0.10	0.36	0.12	-2.23*
DW5	0.94	-0.29	-1.24	-4.39*
AIC	1.84	1.73	1.42	-3.54*
BIC	1.46	1.55	1.49	-3.62*
***U*-statistic** [3]				
AFTER	0.81	0.60	0.53	0.50
Average	0.89	0.83	0.70	0.68
EW5	0.85	0.70	0.60	0.57
DW5	0.81	0.61	0.45	0.32
AIC	0.73	0.60	0.58	0.32
BIC	0.73	0.59	0.57	0.31
CDC [4]				
AFTER	0.56	0.66	0.74	0.78
Average	0.57	0.68	0.77	0.62
EW5	0.57	0.65	0.66	0.64
DW5	0.57	0.66	0.77	0.81
AIC	0.54	0.51	0.68	0.86
BIC	0.57	0.53	0.64	0.86
***PT*-statistic** [5]				
AFTER	0.62	2.64^{+}	3.88^{+}	3.99^{+}
Average	1.46	2.94^{+}	5.79^{+}	4.30^{+}
EW5	1.52	3.66^{+}	4.58^{+}	4.49^{+}
DW5	0.96	1.89	4.52^{+}	5.59^{+}
AIC	1.13	0.28	3.55^{+}	5.71^{+}
BIC	0.87	0.54	2.13^{+}	5.71^{+}

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. $^{+}$ indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

Table 2 shows that for the predictability of the HSI return, the forecast under AFTER does not appear to perform well for the one-month horizon when compared to other methods, even though the *DM*-statistics indicate equal accuracy between AFTER and the historical average benchmark model. For longer horizons (six months and longer), AFTER consistently beats the unconditional historical average benchmark model, and sometimes the four other alternative methods, in the forecast evaluation tests. Only for a 24-month horizon, forecasts based on the two model selection methods (AIC and BIC) outperform those from AFTER.

For the H-share index return, the results in Table 3 indicate that AFTER does not outperform other alternative methods significantly under all forecast horizons. In fact, based on these results, the historical average benchmark model performs as well as (sometimes outperforms) other alternative methods. AFTER beats the benchmark model only for the 24-month horizon. While the RMSEs from AFTER are, in general, larger than those from the benchmark model for other horizons, the *DM*-statistics suggest that the forecasts from AFTER and the benchmark model are equally accurate. Similar to the HSI case, the forecasts from the two model selection methods (based on AIC and BIC) for the 24-month horizon are more accurate than those from AFTER.

4.1.2. Japan TOPIX Index

Table 4. Forecast Comparison Statistics: TOPIX

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	4.6	14.1	20.0	29.3
Average	4.6	16.1	25.8	32.6
EW5	4.6	15.3	23.3	33.6
DW5	4.5	13.5	18.4	20.5
AIC	4.8	16.8	21.5	25.1
BIC	4.7	16.3	22.0	24.5
***DM*-statistic** [2]				
AFTER	0.49	-1.54	-1.97*	-0.77
EW5	0.54	-1.35	-2.32*	0.89
DW5	-0.17	-1.83	-2.66*	-4.65*
AIC	1.32	0.32	-1.96*	-2.26*
BIC	1.01	0.08	-1.76	-2.55*
***U*-statistic** [3]				
AFTER	0.90	0.65	0.56	0.49
Average	0.97	0.81	0.83	0.83
EW5	0.91	0.85	0.83	0.80
DW5	0.76	0.58	0.49	0.35
AIC	0.78	0.65	0.63	0.41
BIC	0.81	0.64	0.62	0.40
CDC [4]				
AFTER	0.50	0.69	0.71	0.73
Average	0.44	0.43	0.41	0.51
EW5	0.45	0.45	0.47	0.43
DW5	0.60	0.71	0.76	0.81
AIC	0.52	0.61	0.62	0.76
BIC	0.50	0.68	0.64	0.76
***PT*-statistic** [5]				
AFTER	-0.25	3.86^{+}	4.15^{+}	4.54^{+}
Average	-1.20	-0.30	-0.82	2.95^{+}
EW5	-0.83	-0.52	0.28	2.37^{+}
DW5	1.92	4.10^{+}	5.14^{+}	6.51^{+}
AIC	0.51	2.21^{+}	2.00^{+}	5.64^{+}
BIC	0.03	3.69^{+}	3.00^{+}	5.64^{+}

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. $^{+}$ indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

For the stock market in Japan, Table 4 indicates that for the forecast horizons of six months and longer, AFTER produces better results under all tests than the unconditional historical average benchmark model.[18] In these cases, the out-of-sample forecasts from AFTER yield smaller RMSEs (even though the *DM*-statistics sometimes suggest insignificant differences between forecasts from AFTER and the benchmark model), *U*-statistics closer to zero, higher percentage of correct directional change, and a significant *PT*-statistic. Even though the forecasts from AFTER have an edge over those from the historical average benchmark model for long horizons, it is noted that their better performance is not obvious when compared to forecasts from AIC, BIC and the 5-year decaying-weight moving average (DW5), especially for the 24-month horizon.[19]

[18] For the 1-month horizon, however, the forecast from AFTER is only marginally better than that of the benchmark model.

[19] For a 24-month horizon, forecasts from AIC, BIC and DW5 have smaller RMSEs than AFTER forecasts, *U*-statistics much closer to zero, higher *CDC*s and more significant *PT*-statistics.

4.1.3. US Dow Jones Industrial Average (DJIA) and S&P 500 Index (SP500)

Table 5. Forecast Comparison Statistics: DJIA

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	4.1	10.3	13.9	19.9
Average	4.1	9.1	13.1	19.1
EW5	4.2	9.8	15.1	25.3
DW5	4.2	8.1	10.2	13.2
AIC	4.2	11.9	15.8	24.8
BIC	4.1	11.9	15.9	24.4
***DM*-statistic** [2]				
AFTER	1.08	1.29	0.49	1.28
EW5	1.89	1.10	1.87	3.69#
DW5	1.34	-1.05	-1.48	-1.91
AIC	1.68	1.69	1.38	1.79
BIC	-1.02	1.72	1.41	1.62
***U*-statistic** [3]				
AFTER	0.83	0.60	0.58	0.58
Average	0.86	0.68	0.61	0.58
EW5	0.84	0.65	0.62	0.67
DW5	0.84	0.56	0.44	0.40
AIC	0.82	0.61	0.58	0.60
BIC	0.86	0.62	0.58	0.59
CDC [4]				
AFTER	0.55	0.66	0.67	0.67
Average	0.55	0.66	0.68	0.67
EW5	0.48	0.55	0.59	0.56
DW5	0.54	0.71	0.81	0.76
AIC	0.52	0.66	0.68	0.69
BIC	0.55	0.66	0.68	0.68
***PT*-statistic** [5]				
AFTER	1.00	1.00	1.46	1.00
Average	1.00	1.00	1.00	1.00
EW5	-1.28	-0.91	-1.09	-2.51
DW5	0.35	3.17^{+}	5.48^{+}	4.53^{+}
AIC	-1.59	1.00	1.00	1.94
BIC	1.00	1.00	1.00	1.51

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. # indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly larger than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. $^{+}$ indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

Table 6. Forecast Comparison Statistics: SP500

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	4.1	12.3	18.1	25.5
Average	4.1	10.6	17.3	27.8
EW5	4.2	11.6	19.6	35.1
DW5	4.1	8.6	12.0	17.8
AIC	4.3	13.9	23.5	39.5
BIC	4.1	13.6	22.7	39.4
***DM*-statistic** [2]				
AFTER	0.34	1.60	0.92	-1.50
EW5	2.32#	1.56	1.75	3.60#
DW5	0.40	-1.37	-1.71	-2.04*
AIC	1.73	1.91	2.01#	1.96#
BIC	-1.97*	2.00#	2.24#	1.96#
***U*-statistic** [3]				
AFTER	0.85	0.70	0.63	0.66
Average	0.85	0.72	0.66	0.68
EW5	0.85	0.73	0.71	0.76
DW5	0.80	0.51	0.41	0.39
AIC	0.86	0.70	0.68	0.74
BIC	0.85	0.70	0.68	0.76
CDC [4]				
AFTER	0.55	0.66	0.68	0.63
Average	0.57	0.66	0.69	0.62
EW5	0.43	0.40	0.34	0.23
DW5	0.57	0.79	0.86	0.80
AIC	0.51	0.65	0.70	0.60
BIC	0.57	0.66	0.71	0.60
***PT*-statistic** [5]				
AFTER	-0.48	1.00	-0.68	1.46
Average	1.00	1.00	1.00	1.00
EW5	-1.85	-2.96	-4.37	-6.21
DW5	0.96	5.23+	6.96+	5.90+
AIC	-1.33	-0.02	1.36	-0.20
BIC	1.00	0.49	1.95	-0.20

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model. # indicates the average squared error under the respective forecast method is significantly larger than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. + indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

The forecast comparison results in Tables 5 and 6 illustrate the performance of the forecasts from AFTER in the US stock market. From the forecast results on DJIA in Table 5, we conclude that AFTER is unable to beat the historical average benchmark model in several tests, including the RMSE (even though the *DM*-statistics are insignificant, suggesting that the forecasts from AFTER and that of the benchmark model are equally accurate) and *CDC*. Nonetheless, AFTER yields a smaller *U*-statistic compared to the benchmark model, suggesting that the degree to which the forecasted return from AFTER differs from the actual return is smaller than the forecast based on the benchmark model. Compared to the alternative methods, forecasts from AFTER only outperform marginally. Generally, forecasts from AFTER have smaller RMSEs than those from AIC, BIC and the equal-weight moving average (EW5). However, in terms of the *U*-statistics and *CDC*, their forecast results are very similar. Similar findings are observed for SP500 in Table 6.[20] The forecast results in this study of the S&P 500 Index using AFTER are not as good as the findings suggested by Hagmann and Loebb (2006). The different results in the two studies may be due to the differences in the number of variables included (and thus the number of models being combined), the setup of the dependent variables and the sample period.

[20] It is noted that for both DJIA and SP500, the method that outperforms the benchmark model and other methods is the decaying-weight moving average (DW5), of which forecasts yield the smallest RMSE, higher *CDC*, and smaller *U*-statistics among all forecast methods and across all forecast horizons (except the 1-month case).

4.1.4. UK FTSE100 Index (FT100)

Table 7. Forecast Comparison Statistics: FT100

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	4.0	9.9	10.0	11.5
Average	3.9	10.6	19.8	33.2
EW5	4.0	11.1	18.5	34.0
DW5	3.9	7.9	10.7	16.5
AIC	4.0	9.9	9.6	9.8
BIC	3.9	9.6	10.0	9.7
***DM*-statistic** [2]				
AFTER	1.03	-1.17	-2.65*	-3.80*
EW5	0.62	0.48	-0.57	0.28
DW5	-0.25	-1.44	-2.19*	-2.84*
AIC	0.90	-0.68	-2.82*	-3.44*
BIC	0.13	-1.00	-2.80*	-3.52*
***U*-statistic** [3]				
AFTER	0.84	0.60	0.40	0.25
Average	0.82	0.70	0.69	0.70
EW5	0.88	0.80	0.79	0.83
DW5	0.79	0.50	0.40	0.36
AIC	0.77	0.57	0.37	0.21
BIC	0.85	0.55	0.39	0.20
CDC [4]				
AFTER	0.54	0.69	0.89	1.00
Average	0.58	0.63	0.66	0.55
EW5	0.42	0.38	0.37	0.32
DW5	0.62	0.79	0.85	0.82
AIC	0.54	0.76	0.86	0.98
BIC	0.54	0.76	0.86	0.99
***PT*-statistic** [5]				
AFTER	-0.80	2.93^{+}	7.90^{+}	10.20^{+}
Average	1.00	1.00	1.00	1.00
EW5	-1.94	-2.88	-3.21	-3.69
DW5	2.05^{+}	5.44^{+}	6.91^{+}	6.58^{+}
AIC	-0.01	4.68^{+}	7.22^{+}	9.76^{+}
BIC	-0.50	4.70^{+}	7.22^{+}	9.98^{+}

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. $^{+}$ indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

For the UK stock market, Table 7 shows that for 1-month horizon, AFTER is unable to beat the historical average benchmark model in most tests. Similarly, forecasts from other alternative methods also underperform against those based on the simple historical average. However, for forecast horizons of six months and longer, AFTER performs better against the benchmark model with significantly smaller RMSEs, smaller *U*-statistics, larger *CDC* and significant *PT*-statistics. In most cases, AFTER also outperforms the alternative methods.[21]

[21] Forecasts from the two model selection methods (AIC or BIC) are sometimes more accurate than that from AFTER.

4.1.5. France CAC 40 Index (CAC)

Table 8. Forecast Comparison Statistics: CAC

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	5.2	12.6	20.0	29.9
Average	5.3	16.0	25.7	39.7
EW5	5.5	16.9	28.9	49.9
DW5	5.3	12.2	18.1	25.8
AIC	5.3	14.7	17.5	35.7
BIC	5.5	14.8	19.5	36.9
***DM*-statistic** [2]				
AFTER	-0.28	-2.21*	-1.98*	-1.98*
EW5	2.73#	0.98	2.13#	2.87#
DW5	-0.07	-1.50	-1.75	-2.62*
AIC	0.17	-1.01	-2.24*	-0.96
BIC	0.91	-0.94	-1.65	-0.57
***U*-statistic** [3]				
AFTER	0.74	0.49	0.46	0.40
Average	0.87	0.67	0.68	0.78
EW5	0.86	0.79	0.77	0.77
DW5	0.75	0.48	0.41	0.35
AIC	0.69	0.54	0.38	0.46
BIC	0.72	0.55	0.39	0.48
CDC [4]				
AFTER	0.56	0.82	0.78	0.87
Average	0.62	0.69	0.68	0.63
EW5	0.42	0.35	0.33	0.27
DW5	0.65	0.81	0.83	0.81
AIC	0.60	0.74	0.89	0.82
BIC	0.59	0.77	0.90	0.81
***PT*-statistic** [5]				
AFTER	1.00	5.57+	4.45+	7.46+
Average	1.00	1.00	1.00	1.00
EW5	-2.19	-4.06	-4.50	-5.11
DW5	2.34+	5.58+	6.04+	6.15+
AIC	1.98+	3.36+	7.61+	6.45+
BIC	1.63	4.17+	7.84+	6.08+

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model. # indicates the average squared error under the respective forecast method is significantly larger than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. + indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

Similar to the UK case, the forecast results of the French stock market return in Table 8 show that for the 1-month horizon, AFTER marginally outperforms the historical average benchmark model with smaller RMSEs and *U*-statistics.[22] However, for longer horizons, AFTER is able to beat the benchmark model under all tests. Among the alternative methods, the forecasts from the model selection method based on AIC and the decaying-weight moving average method (DW5) sometimes perform slightly better than those from AFTER under some particular horizons. For instance, the forecasts from AIC beat AFTER's for the 12-month horizon, while those from DW5 outperform AFTER's for the 24-month horizon.

[22] Nonetheless, AFTER forecasts have a lower *CDC* than the benchmark model and sometimes insignificant *PT*-statistics.

4.1.6. German DAX 30 Index (DAX)

Table 9. Forecast Comparison Statistics: DAX

	Forecast Horizon			
	1-month	6-month	12-month	24-month
RMSE [1]				
AFTER	6.9	17.7	29.8	46.7
Average	6.9	18.7	28.2	43.9
EW5	7.1	20.5	33.0	54.6
DW5	7.0	16.1	22.2	27.9
AIC	7.3	18.7	34.0	55.0
BIC	7.0	18.0	32.0	53.4
***DM*-statistic** [2]				
AFTER	0.43	-1.08	0.47	1.00
EW5	1.96#	1.67	2.52#	4.38#
DW5	0.36	-0.97	-1.39	-2.58*
AIC	1.33	-0.04	0.87	2.63#
BIC	1.00	-0.59	0.66	2.26#
***U*-statistic** [3]				
AFTER	0.82	0.60	0.63	0.64
Average	0.92	0.78	0.86	0.81
EW5	0.89	0.81	0.79	0.79
DW5	0.79	0.55	0.46	0.36
AIC	0.81	0.62	0.61	0.72
BIC	0.89	0.66	0.61	0.71
CDC [4]				
AFTER	0.54	0.79	0.69	0.58
Average	0.54	0.63	0.64	0.62
EW5	0.47	0.37	0.33	0.35
DW5	0.57	0.77	0.81	0.84
AIC	0.51	0.75	0.75	0.58
BIC	0.53	0.69	0.72	0.59
***PT*-statistic** [5]				
AFTER	1.00	5.60+	2.75+	0.14
Average	1.00	1.00	1.00	1.00
EW5	-0.76	-3.02	-3.95	-3.08
DW5	1.24	4.96+	5.79+	6.67+
AIC	-1.19	4.44+	4.35+	-0.36
BIC	-0.93	3.19+	3.57+	-0.06

Source: HKMA staff estimates.

Notes:
1. The method with a smaller RMSE is preferred.
2. * indicates rejection of the null hypothesis of equal forecast accuracy (comparing with the forecast from the historical average model) based on the *DM*-statistic at the 5% significance level, and the average squared error under the respective forecast method is significantly less than that under the historical average model. # indicates the average squared error under the respective forecast method is significantly larger than that under the historical average model.
3. The method with its *U*-statistic closer to zero is preferred.
4. The method with a higher *CDC* is preferred.
5. + indicates rejection of the null hypothesis of no predictive performance at the 5% significance level. The method with a *PT*-statistics larger than 1.96 is preferred.

For the German stock market, the results in Table 9 indicate that the forecast performance of AFTER is slightly better than that of the historical average benchmark model. While the forecasts based on AFTER have larger RMSEs (in 12-month and 24-month cases) than that from the benchmark model, *DM*-statistics indicate that their SEs are statistically indifferent. Nevertheless, the AFTER forecast has smaller *U*-statistics, higher percentage of *CDC* and significant *PT*-statistics (in 6-month and 12-month cases) compared to the benchmark model. Among the other alternative methods, only the results from the decay-weight moving average method (DW5) have the advantage over AFTER, especially for the 24-month forecast horizon.

In summary, a common observation in Tables 2 to 9 is that as the forecast horizon lengthens, the forecast performance of AFTER actually improves.[23] With longer horizons, it is shown that the *U*-statistic from AFTER forecasts decreases and approaches closer to zero, its *CDC* increases to over 70% or higher, and its *PT*-statistic becomes much larger than the 5% significance level of 1.96. Such an improvement is not generally present with the forecasts based on the historical average benchmark model.

V. SUMMARY AND CONCLUSION

This paper studies the issue of stock market return predictability by applying a new model combination methodology known as "**A**ggregate **F**orecasting **T**hrough **E**xponential **R**e-weighting (AFTER)" introduced by Yang (2004). Out-of-sample return forecasts of eight stock market indices from six economies are examined and their results are compared with those of the historical average benchmark model as well as four other alternative methods.

The empirical results suggest that the performance of the forecasts derived from AFTER is relatively better, in terms of smaller root-mean-square forecast errors, smaller Theil's U statistics, and more accurate in predicting the direction of change (as shown by the indicator for the direction of change and the statistics in the Pesaran-Timmerman test). For 1-month horizon, the forecasts derived from AFTER are "as good as" those from the simple historical average return. For longer horizons, forecasts from AFTER are able to outperform those of the historical average return. This is the case for the Hang Seng Index (forecast horizons at six months and longer), the H-share Index (24-month horizon), the S&P500 Index (24-month horizon), the TOPIX Index (six months and longer), the FTSE100 Index (six months and longer), the CAC 40 Index (six months and longer) and the DAX 30 Index (six months and 12 months).

---

[23] With the exception of the forecast performance for the stock markets in the US and Germany.

Despite the relatively better performance of using the AFTER algorithm for long horizons, one should bear in mind that it is still a daunting task to forecast stock market movement. Therefore, the application of AFTER for stock market forecast should be used with caution. In practice, the AFTER forecast could be used as a reference if there is a need to know the long-term performance of the stock market in policy makers' monitoring work.[24]

[24] For example, the outlook of stock market performance is likely to be a factor in the macroeconomic model developed by central banks.

REFERENCES

Aiolfi, M. and C. A. Favero (2005): “Model Uncertainty, Thick Modelling and the Predictability of Stock Returns”, *Journal of Forecasting* 24(4), 233-254.

Akaike, H. (1974): “A New Look at the Statistical Model Identification”, *IEEE Transactions on Automatic Control* 19(6), 716-723.

Baker, M and J. Wurgler (2000): “The Equity Share in New Issues and Aggregate Stock Returns”, *Journal of Finance* 55, 2219-2257.

Bates, J. and C. Granger (1969): “The Combination of Forecasts”, *Operational Research Quarterly* 20(4), 451-468.

Board of Governors of the Federal Reserve System (1997): “Monetary Policy Report to the Congress Pursuant to the Full Employment and Balanced Growth Act of 1978”, Federal Reserve Board.

Bossaerts, P. and P. Hillion (1999): “Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?”, Review of Financial Studies 12(2), 405-428.

Campbell, J. Y. and R. J. Shiller (1988a): “The Dividend-price Ratio and Expectations of Future Dividends and discount Factors”, *Review of Financial Studies* 1, 195-228.

——————————— (1988b): “Stock Prices, Earnings, and Expected Dividends”, *Journal of Finance* 43, 661-676.

Campbell, J. Y. and S. B. Thompson (2006): “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?”, forthcoming *Review of Financial Studies*.

Cremers, K. J. M. (2002): “Stock Return Predictability: A Bayesian Model Selection Perspective”, *Review of Financial Studies* 15(4), 1223-1249.

Diebold, F. X. and R. S. Mariano (1995): “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics* 13(3), 253-265.

Fama, E. F. and K. R. French (1988): “Dividend Yields and Expected Stock Returns”, *Journal of Financial Economics* 22, 3-25.

Goyal, A. and I. Welch (2006): “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction”, forthcoming *Review of Financial Studies*.

Guo, H. (2006): “On the Out-of-Sample Predictability of Stock Market Returns”, *Journal of Business* 79(2), 645-670.

Hagmann, M. and J. Loebb (2006): "Model Combination and Stock Return Predictability", Swiss Finance Institute *Research Paper Series* 06-5.

Hendry, D. F. and M. P. Clements (2004): "Pooling of Forecasts", *Econometrics Journal* 7(1), 1-31.

Hibon, M. and T. Evgeniou (2005): "To Combine or Not to Combine: Selecting Among Forecasts and Their Combinations", *International Journal of Forecasting* 21(1), 15-24.

Hjalmarsson, E. (2006): "Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns?", Board of Governors of the Federal Reserve System *International Finance Discussion Papers* #855.

Inoue, A. and L. Kilian (2004): "In-sample or Out-sample Tests of Predictability: Which One Should We Use?", *Econometric Reviews* 23, 371-402.

Lamont, O. (1998): "Earnings and Expected Returns", *Journal of Finance* 53, 1563-1587.

Lettau, M. and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns", *Journal of Finance* 56, 815-849.

Nelson, C. and M. Kim (1993): "Predictable Stock Returns: The Role of Small Sample Bias", *Journal of Finance* 48, 641-661.

Newey, W. and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708.

Pesaran, M. H. and A. Timmermann (1992): "A Simple Nonparametric Test of Predictive Performance", *Journal of Business and Economic Statistics* 10(4), 461-465.

______________________ (1995): "Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance* 50(4), 1201-1228.

______________________ (2000): "A Recursive Modelling Approach to Predicting UK Stock Return", *The Economic Journal* 110(460), 159-191.

Schwarz, G. (1978): "Estimating the Dimension of a Model", *The Annals of Statistics* 6(2), 461-464.

Stambaugh, R. F. (1999): "Predictive Regressions", *Journal of Financial Economics* 54, 375-421.

Stock, J. H. and M. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices", *Journal of Economic Literature* 41, 788-829.

__________________ (2004): "Combination Forecasts of Output Growth in a Seven-Country Data Set", *Journal of Forecasting* 23, 405-430.

Theil, H. (1961), *Economic Forecasts and Policy*, 2nd Edition, Amsterdam: (North-Holland).

Yang, Y. (2004): “Combining Forecasting Procedures: Some Theoretical Results”, *Econometric Theory* 20(1), 176-222.