Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Real-Time Forecast Averaging with ALFRED

Chanont Banternghansa
and
Michael W. McCracken

Working Paper 2010-033A
http://research.stlouisfed.org/wp/2010/2010-033.pdf

October 2010

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Real-Time Forecast Averaging with ALFRED *

Chanont Banternghansa
Federal Reserve Bank of St. Louis

Michael W. McCracken
Federal Reserve Bank of St. Louis

August 2010

Abstract

This paper presents empirical evidence on the efficacy of forecast averaging using the ALFRED real-time database. We consider averages taken over a variety of different bivariate VAR models that are distinguished from one another based upon at least one of the following: which variables are used as predictors, the number of lags, using all available data or data after the Great Moderation, the observation window used to estimate the model parameters and construct averaging weights, and for forecast horizons greater than one, whether or not iterated- or direct-multistep methods are used. A variety of averaging methods are considered. Our results indicate that the benefits to model averaging relative to BIC-based model selection are highly dependent upon the class of models being averaged over. We provide a novel decomposition of the forecast improvements that allows us to determine which types of averaging methods and models were most (and least) useful in the averaging process.

JEL Nos.: E52, E58, C53
Keywords: forecast averaging, real-time data, forecast accuracy

* *Banternghansa*: Research associate, Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; `chanont.banternghansa@stls.frb.org`. *McCracken (corresponding author)*: Research officer, Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; `michael.w.mccracken@stls.frb.org`. We are grateful to ... for valuable comments. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

This paper provides evidence on the ability of various forms of forecast averaging to improve the real–time forecast accuracy of monthly frequency bivariate VAR forecasts of headline and core CPI-based inflation, growth in Industrial Production (IP), and the unemployment rate. We consider a range of approaches to averaging forecasts made by a variety of primitive methods for managing the estimation of each bivariate VAR model. The averaging methods include: equally weighted averages, medians, MSE–weighted averages, Bayesian model averages based upon a BIC approximation, and combinations based on percentile average forecasts. For each of these averaging approaches, we construct forecasts of each variable using real–time data taken from the ALFRED database. We compare our results to bivariate VAR models selected using BIC as a model selection procedure.

Of course, model averaging for forecasting is nothing new. There is plenty of evidence suggesting that model averaging can provide improvements in terms of forecast accuracy relative to model selection. Empirical examples of this evidence include, but are certainly not limited to, Stock and Watson (2004), Kapetanios, Labhard, and Price (2008), and Kascha and Ravazzolo (2010). Theoretical results include Hansen (2008), Elliott and Timmermann (2004), Clark and McCracken (2008), and many others.

In some instances, for example Clark and McCracken (2010) and Faust and Wright (2009), the real-time nature of the data is accounted for when forecasting with model averages. Even so, such examples are the exception and not the norm. Here we use the ALFRED database to mimic the type of data that a forecaster would have access to at each point in time as they construct their monthly forecasts. This is important because it takes account of the fact that economic data is often subject to revision and hence the actual value of a variable may change as we move across forecast origins. In addition, using real time data allows us to account for the fact that most macroeconomic data arrive only after a substantial lag and, moreover, these time lags can vary widely across variables from as little as a week (for employment figures) to as along as two months (for trade data). Finally, by using the ALFRED database as the universe of potential predictors, we allow for the fact that new series become available across time and existing series sometimes become discontinued.

In accordance with the previous literature, our results indicate that model averaging can –but does not always– improve forecast accuracy relative to the more standard BIC-based approach to model selection. Put differently, model averaging per se is not a panacea for improving forecast accuracy. Improvements from model averaging depend critically upon the type of models we average over. There appears to be some advantage to pre-selecting which primitive models should be used in the averaging process. For example, when forecasting core CPI-based inflation there appears to be substantial gains in forecast accuracy at all horizons when we average over only those models estimated using a rolling observation window of fixed size rather than a recursive, expanding, observation window. In contrast, we find that when forecasting indistrial production there are gains in forecast accuracy when we average over only those models estimated using a recursive window rather than a rolling window of observations.

With these two examples in mind, we provide a novel decomposition of the relative root mean-square error improvements for each of our dependent variables, at each forecast horizon, that allows us to determine which types of primitive models and which model averaging techniques are, on average, most (and least) beneficial in the averaging process. In some, though not all, instances we find that our decomposition meshes well with those permutations of types of models and types of averaging procedures that produce the most accurate forecasts.

The remainder of the paper proceeds as follows. Section 2 describes the real-time data used in our analysis. Section 3 provides a synopsis of the primitive models that we average across as well as a description of the types of model averaging we consider. Section 4 presents our results on forecast accuracy and our decomposition. Section 5 concludes.

2 Data

We obtained our data from the ALFRED database maintained by the Federal Reserve Bank of St. Louis. This database consists of collections of vintages of data for each variable – vintages that vary across time as either new data is released or existing data is revised by the relevant statistical agency. By using this database we are able to insure that at each monthly forecast origin we are using only data that was available at the time that the forecast was made. We therefore have defined "real-time" forecasting as using any data available by the end of the month from which we are forecasting.

Choosing the end of a month as the forecast origin is non-trivial. Nearly all monthly macroeconomic data is released after the end of the month it references. A model that uses data associated with January 1996 must therefore be constructed after that month has ended. If we choose the first day of February as our forecast origin, the forecast would be very timely but there would be almost no data associated with January to make use of-thus reducing the accuracy of the forecast. On the other hand, if we choose the first day of May as our forecast origin we would have all the January data available to us but the forecast would be very much out of date. As a middle ground we choose the end of the month following the most recent data vintage as the relevant forecast origin. This implies that our 1-step ahead forecasts, constructed using January 1996 vintage data, made at the end of February, will be a forecast of data associated with February 1996.[1]

From this database, we use a total of 238 unique monthly macroeconomic series in our analysis. Of these 238 series, 67 of them are available for the January 1996 vintage data. As we progress across time, we allow the number of variables used to increase or decrease with data availability. For example, the number of series more than doubles in November 1996. By the end of our forecasting exercise in December of 2008 we have a total of 193 series that are used either as dependent variables or as predictors. This is less than the total number of variables because 45 series were discontinued or did not have enough observations at some point in time to adequately estimate either the model parameters or model averaging weights.[2] There are 29 Output & Production series, 8 Income, Outlays, & Savings series, 40 Labor Market series, 52 Monetary Aggregate and Reserve series, 35 Exchange Rate series, 38 Financial Market and Interest Rate series, 34 Price series, and 2 Survey series. The detailed list available from the authors upon request.

At each forecast origin starting in February of 1996, we construct forecasts of three variables: headline CPI-based inflation, IP growth, and the unemployment rate. We begin forecasting core CPI-based inflation using December 1996 vintage data –the first available vintage for this series. For each of these four variables we construct $h = 1, 3, 6, 12$, and $24-$month ahead forecasts. For unemployment, the target variable being forecasted is y_{t+h}, the unemployment rate at the forecast horizon h. For the inflation and growth forecasts, the target variable being forecasted is the average annualized monthly rate of growth over the forecast horizon and hence interpretation of the target variable varies with

[1] Giannone, Reichlin, and Small (2008) refer to this type of forecast as a "nowcast."

[2] In our analysis we set a few basic rules for which variables to include: (i) we do not use seasonally unadjusted data when the seasonally adjusted version is available, (ii) we do not use regional data for our analysis, (iii) we omit a variable if there is less than 10 years worth of data to use for estimating the model parameters, and (iv) we omit a variable if we do not have at least 24 pseudo out-of-sample forecast errors to calculate the MSE weighted forecasts.

the forecast horizon. More precisely, if we let y_t denote the time t log-difference in, say, headline CPI, the target variable being forecasted at horizon h is $y_{t+h}^{(h)} = (\frac{1200}{h}) \sum_{i=1}^{h} y_{t+i}$.

When constructing our forecast errors, we use the third release (or equivalently, the second revision) of the variable as the realized value of our target variable. In total, since December 2008 is the final vintage we use to evaluate our forecasts, for each model we have roughly 155 1-month ahead forecast errors that we use to measure accuracy. This number shrinks to 151, 145, 133, and 109 for the 3-, 6-, 12-, and 24-month ahead forecasts respectively.

Following Marcellino, Stock and Watson (2006), each variable we use is transformed to ensure stationarity using differences or log-differences. For our dependent variables, we treat the unemployment rate as stationary in levels but treat headline CPI, core CPI, and IP as stationary in log first differences. These transformations are done across all vintages uniformly. We do not allow for the possibility that the type of transformation could be different across vintages. After transforming our variables we then check for outliers defined as observations that are greater than 6 times its interquartile range. These are replaced with the mean of the series (without the outlier) from the relevant vintage. This is done vintage by vintage and hence the outlier detection is not influenced by observations not available at each forecast origin. Note that across the time period for which we forecast, the price and IP indices have been periodically renormalized so that the units of measurement are not the same across all vintages. To avoid mixing and matching, we renormalized each vintage relative to that of the December 2008 vintage.

3 Methods

In this section we describe the primitive models that we average across as well as describe the various model averaging approaches that we consider. All models have one thing in common: they all (see section 3.4 below for a caveat) take the form of an OLS estimated bivariate VAR in the variable to be predicted and one additional predictor. Otherwise all of the primitive models differ by at least one of 6 features: (i) which element of the ALFRED database is used as a predictor, (ii) how many lags of the dependent variable are being used as a predictor, (iii) how many lags of the additional predictor are being used, (iv) whether the model is being estimated using all available data (i.e. the recursive scheme) or is estimated using a moving window of observations (i.e. the rolling scheme), (v) whether the model is estimated only using post Great Moderation data or uses data as far back as the vintage has available, or (vi) for forecast horizons greater than one, whether we use iterated- or direct-multistep methods to create our primitive forecasts.

3.1 Predictors

As noted above, we use the ALFRED database for our real-time forecasting exercise. In particular, we treat it as the universe of potential variables that could be used as a predictor for any one of our four dependent variables. Since the number of variables in ALFRED changes across forecast origins, the number of primitive models that we average across changes across forecast origins. At the beginning of our sample, January 1996, we have a total of only 66 potential predictors for each dependent variable. At the last potential forecast origin (November 2008), for a 1-step ahead forecast we have a total of 192 potential predictors. While the number of predictors typically grows – sometimes dramatically as was the case of November 1996 – there are a few instances in which the number of predictors falls as various variables are discontinued or dropped because of insufficient data.[3]

3.2 Full Sample and Great Moderation Sample

For each of the above models, we estimate the regression parameters using one of two subsets of data. In the first we use all available data in that vintage (Full). While the first observation varies across individual variables, many date back to as early as January 1959. For the second subset of data we restrict attention to only that data which occurs starting in January 1983 (Post), roughly the time frame in which the Great Moderation is considered to have started. Note that this implies that for each vintage we use for estimation, any historical observations pre-1983 are discarded.

We consider both of these separate subsets of data since there is considerable evidence, including that in D'Agostino, Giannone, and Surico (2007), that the predictability of many macroeconomic variables has changed since the onset of the Great Moderation. Even so, there is a trade-off. Using less information to estimate model parameters may generate estimates that are closer to being unbiased because older data comes from a different macroeconomic regime, but it also can decrease the precision of the estimates. In practice, this trade-off may favor using more or less data to estimate parameters due to a bias variance trade-off.

[3] See footnote 2 for more detail.

3.3 Recursive and Rolling Windows

For each of the above models, and conditional on whether we use the Full or Post sample, we estimate the bivariate VAR using one of two observation windows. In the recursive scheme (Rec), we estimate the model by OLS using all available data. Hence as we move forward from one month to the next we use one more observation to estimate the model parameters. In the rolling scheme (Roll), we estimate the model by OLS using only the past 10 years of available data. Hence as we move forward from one month to the next we use the same number of observations to estimate the model parameters.

In some ways, our decision to consider two sample subsets (Full vs. Post) as well as two types of observations windows (Rec vs. Roll) may seem redundant. We view the two choices as distinct but related. In the former, we are essentially assuming a discrete break in the data and seeing how doing so helps forecast accuracy. For the latter, we are assuming something closer to a smooth sequence of breaks. Since we are unsure of which is the proper way to manage forecasting in the presence of uncertain forms of potential structural change, we consider both. See Clark and McCracken (2010) for further discussion on this issue.

3.4 IMS and DMS

For each permutation of predictor, sample, and observation window, we estimate our bivariate VAR-based forecasting model using two different methods: the textbook approach that induces an iterated multi-step (IMS) forecast and the somewhat easier-to-implement method of using a direct multi-step (DMS) forecasting model. Below we give a brief description of each. In the following, let y_t denote either the time t level of the unemployment rate or the time t log-first difference of headline or core CPI or IP. In addition, recall that the target variable to be forecasted at forecast horizon h is $y_{t+h}^{(h)} = (\frac{1200}{h})\sum_{i=1}^{h} y_{t+i}$ for the price and IP indices but is simply y_{t+h} for unemployment.

When using the iterated multi-step (IMS) approach to forecasting, at each forecast origin t we first use OLS to estimate the bivariate VAR model

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha_{y,0} \\ \alpha_{x,0} \end{pmatrix} + A(L)\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix} \tag{1}$$

where $A(L)$ denotes a lag operator of appropriate dimension for the given number of lags used in both the y and x equations. With the regression parameter estimates in hand, the recursive nature of the VAR is used to generate a sequence of 1 through h step ahead forecasts $\hat{y}_{t+i}$ $i = 1, ..., h$. For the unemployment rate, $\hat{y}_{t+h}$ is the resulting forecast of our target variable. For the other dependent variables, we follow Marcellino, Stock, and Watson (2006) and define our $h-$step ahead IMS forecast as $(\frac{1200}{h})\sum_{i=1}^{h}\hat{y}_{t+i}$. Note that for each forecast horizon, the same parameter estimates are used when constructing the forecasts.

When using the direct multi-step (DMS) approach to forecasting, a distinct model is estimated separately for each forecast horizon h. For the unemployment rate and a fixed value of h, this model takes the form

$$y_{t+h} = \alpha_{y,0} + A_y(L)y_t + A_x(L)x_t + \varepsilon_{y,t+h} \tag{2}$$

where $A_y(L)$ and $A_x(L)$ denote lag operators of appropriate dimension for the given number of lags used for y and x respectively. For each separate forecast horizon the forecast is defined as $\hat{y}_{t+h} = \hat{\alpha}_{y,0} + \hat{A}_y(L)y_t + \hat{A}_x(L)x_t$. For the price and production indices, the model takes the slightly different form

$$y_{t+h}^{(h)} = \alpha_{y,0} + A_y(L)y_t + A_x(L)x_t + \varepsilon_{y,t+h}. \tag{3}$$

For each separate forecast horizon the forecast is similarly defined as $\hat{y}_{t+h}^{(h)} = \hat{\alpha}_{y,0} + \hat{A}_y(L)y_t + \hat{A}_x(L)x_t$. Note that in each of the above examples, the parameter estimates resulting from these models will vary with the forecast horizon.

3.5 Lags

Each of the IMS and DMS specifications above require choosing the number of lags of y and x to use as predictors. The textbook choice would be to use a model selection procedure such as BIC. Such a choice, however, contrasts with our goal of providing evidence on the benefits of model averaging relative to model selection techniques. In addition, due to the considerable degree of evidence suggesting that there has been a change in the degree of persistence in inflation (e.g. Levin and Piger 2006), one might consider the possibility that the lag order structure of the model for inflation, in particular has changed over time. We therefore consider all 144 permutations of up to 12 lags of either the y or x variable.

3.6 Averaging Methods

After considering all the permutations of model elements discussed above, for each variable we have 76,128 1-month ahead forecasting models estimated in January 1996 and 221,280 1-

month ahead forecasting models estimated in November of 2008.[4] With this rich collection of individual forecasting models as building blocks, we consider a range of approaches to model averaging with an eye towards determining which types of model averaging are most useful and moreover, which types of primitive models are the most useful to average over.

Our first set of model averages are the simplest. We consider the equally weighted average (Average) as well as the median forecast from among these models (Median). While these methods are not statistically exciting, there is substantial evidence suggesting that simple forms of model averaging can perform quite well (e.g. Smith and Wallis 2009). Note that this form of model averaging implies model weights that are invariant to the forecast horizon.

We then consider two distinct forms of weighted model averaging. In the first, we follow Stock and Watson (2004) as well as others and consider using relative inverse mean square forecast error (MSE)-based weights to combine our models. The intuition is that if one has historical evidence on the accuracy of a collection of models in the form of those model's MSEs, then one might want to use those measures of accuracy to give a particular model more weight if it has been more accurate than the other models. Computationally, let $MSE_{i,t,h}$ denotes the known MSE associated with individual model i at forecast origin t associated with a sequence of past $h-$step ahead forecast errors, the weight given to model i is $\frac{MSE_{i,t,h}^{-1}}{\sum_{j=1}^{N_t} MSE_{j,t,h}^{-1}}$ where $j = 1, ..., N_t$ denotes an index of all the available primitive models at forecast origin t.

In our application, for the relevant vintages of data needed to estimate a particular model at forecast origin t, we conduct a pseudo out-of-sample forecasting exercise to generate these MSEs. The particulars of what we do depend upon whether we are using the Full or Post sample and whether we use the recursive or rolling scheme to construct our forecasts. If the recursive (rolling) scheme is used to construct the model forecast, then the recursive (rolling) scheme is used to construct the pseudo out-of-sample forecasts that are used to construct the model weights. If the Full sample is used, the first pseudo out-of-sample forecast is constructed using data from January 1960 to December 1969 and iterates forward until the availability of real-time data, at time t, is insufficient to calculate a forecast error using the third release of the relevant dependent variable. If the Post sample is used, the first pseudo out-of-sample forecast is constructed using data from January 1984 to December

[4] The number of models not only changes across forecast origins but varies slightly across forecast horizons due to data availability. See footnote 2.

1993 and iterates forward in time as discussed above. Since our forecasting exercise starts in January 1996, this implies that the model weights constructed using the Full sample are estimated based upon an average MSE that uses many more squared forecast errors than those constructed using the Post sample.

We also consider an approximate Bayesian model averaging strategy in which we calculate a posterior probability from prior probabilities and marginal likelihoods for each model, with each model assigned the same prior probability. Following Garrat, Koop, and Vahey (2006), the marginal likelihood of a given model is approximated using it's BIC. In our analysis, for each vintage we estimate each model using the relevant subset of the available data (i.e. the Full or Post sample) and based upon the subsequent residuals, calculate the value of the BIC. Computationally, if we let $BIC_{i,t,h}$ denote the value of the BIC associated with the residuals from individual model i at forecast origin t, the weight given to model i is $\frac{\exp(-.5*BIC_{i,t,h})}{\sum_{j=1}^{N_t}\exp(-.5*BIC_{j,t,h})}$. For the iterated multistep models the BIC is constructed in the typical fashion using equation (1) that implicitly assumes that the residuals are serially uncorrelated. For the direct multi-step models however, we know that when $h > 1$ the residuals from equation (2) are not serially uncorrelated and hence the typical formulation is invalid.[5] For simplicity, we use the standard BIC formula regardless.

In addition to the weighted forecasts described above that averages across all models, we also considered a variant that filters out those models considered "less accurate" by some metric and averages over only those remaining. Specifically, at each forecast origin t we follow Aiolfi and Timmermann (2006) as well as Clark and McCracken (2010) by calculating a Top 10% MSE and a Top 10% BIC weighted average that is constructed using only the top ten percent of the available models. For the Top 10% MSE models this is done by averaging only over the models with the lowest 10% of pseudo out-of-sample MSEs based on the most recent vintage of data. For the Top 10% BIC models this is done by averaging only over the models with the lowest 10% of values of BIC based on the most recent vintage of data.

3.7 Benchmark Forecast

In reporting our results it is useful to get a feel for the magnitude of the benefits of model averaging. To do so we need to choose a baseline for comparison. Since our goal is to observe the benefits of model averaging relative to model selection, using a fixed autoregressive

[5] See Hansen (2010) for a discussion of how this affects the definition of BIC.

model with known lags is insufficient. Not only does that baseline not capture the time varying nature of model selection in a real-time forecast setting, in many cases it doesn't even serve as a particularly difficult benchmark to beat. For example, we could have used the standard random walk benchmark but as we will see below, while this turns out to be a strong benchmark for the unemployment rate, it is a horrible benchmark for industrial production and both of the price indices.

Instead we use the recursively estimated, iterated multi-step, BIC-selected forecast estimated over the Full sample as our benchmark. At each forecast origin t this entails calculating the value of the BIC for each IMS model from equation (1), estimated using the Full sample, separately across all possible lag permutations, and choices of predictor and then choosing the model with the lowest BIC as the model that is used to construct the forecast. The reason behind our selection is that this particular BIC-selected forecast is the conventional model that a textbook in time series econometrics would suggest we use. For completeness, we also report the relative RMSEs associated with the random walk model.

3.8 Summary of Methods

In all, for each variable and each horizon, we consider 6 different forms of model averaging: Average, Median, MSE weight, Top 10% MSE, BIC weight, and Top 10% BIC. Each of these forms of averaging are then applied separately to several distinct classes of models –models indexed by whether they are constructed using (i) the Full and/or Post samples, (ii) the Rec and/or Roll schemes, and (iii) the IMS and/or DMS approaches to forecasting. Note that since we allow for averaging over (for example) models estimated using either the Rec or Roll schemes, there are $3^3 = 27$ model classes that we consider. In all, this gives us $6 \times 3^3 = 162$ distinct permutations of form of model averaging and the types of models that are averaged over.

4 Results

In this section we discuss our results on the value of using forecast averaging as a tool for improving forecast accuracy. For brevity, however, we do not present the tables associated with all 162 model averaging and model class variants. Instead, in Tables 1 and 2 we present results for each of the types of model averaging when we average over *all* models. Table 1 presents results for headline and core CPI-based inflation while Table 2 presents results for growth in Industrial Production and the unemployment rate. The values in the first row of each panel are the RMSEs associated with the benchmark model chosen using BIC at each forecast origin. The remaining values in each panel are relative RMSEs. Values greater than 1 favor the benchmark model while values less than 1 favor the form of model averaging denoted in the first column. For each forecast horizon, the best relative RMSE is highlighted in bold.

4.1 RMSE of Nominal Variables

The first panel of Table 1 provides the results on forecasting headline CPI-based inflation when we average across all models. Here we find that at the shortest three horizons there are little if any advantages to forecast averaging across all models in terms of RMSEs. When averaging over all the primitive models the benchmark is either better than model averaging or only marginally worse. However, as the forecast horizon increases to 12 months, model averaging improves accuracy by roughly five percent and at the longest horizon, forecast averaging improves accuracy by as much as thirty percent. In each of these latter horizons, the Top 10% BIC weighted forecasts gave the lowest RMSEs.

The second panel of Table 1 provides the same results but for core CPI-based inflation. In contrast to the results for headline inflation, here there are consistent improvements at all horizons when model averaging across all models. At the shortest horizons the gains were on the order of a modest five percent but as the horizon increases the improvements rise to something closer to fifteen percent. Across all horizons, no single averaging approach really sticks out as consistently giving the greatest improvements: The Average, MSE weighted, and Top 10% BIC each perform best in at least one horizon.

4.2 RMSE of Real Variables

The first and second panels of Table 2 parallel those in Table 1 in terms of the benefits of model averaging. As was the case for headline CPI, model averaging across all models provides little-to-no improvement relative to model selection when forecasting IP growth at the shortest horizons. In fact, model averaging typically does worse than model selection at the longest horizons with losses of roughly five percent.

But again, in contrast to the results in the first panel, model averaging across all models consistently improves forecast accuracy relative to model selection when forecasting the unemployment rate. Each model averaging procedure improves forecast accuracy at every horizon. Somewhat surprisingly the improvements are U-shaped: the improvements in RMSE are roughly seven percent at the shortest and longest horizon but are closer to twelve percent at the intermediate horizons. Across all but the longest horizon, the Top 10% MSE weighted forecast has the largest improvements relative to the benchmark. At the longest horizon the BIC weighted average performs best.

4.3 Decomposition Regression Analysis

In Tables 1 and 2 we see that while model averaging can improve forecast accuracy, it does not always do so relative to our model selection-based benchmark. Moreover, when model averaging does provide improvements the best form of model averaging varies across dependent variables and across forecast horizons. Finally, though obviously not apparent in Tables 1 and 2 (which present results when averaging over *all* of the primitive models), comparable conclusions can be reached were we to report all of the remaining $162-6=156$ permutations of types of model averaging and model classes for each dependent variable and each horizon.

Even so, it may be that on average across all of these permutations, some simple patterns emerge that could help us identify which types of model averaging are best to use and which classes of models should be averaged over. In order to parse out such effects we estimate a regression in which we use dummy variables for the types of model averaging and model classes as predictors for the corresponding relative RMSEs. Specifically, for each dependent variable and each forecast horizon h, we use OLS to estimate the regression:

$$\begin{aligned} RMSE_i^h - 1 &= \alpha_1 \text{DMS} + \alpha_2 \text{Post} + \alpha_3 \text{Roll} \\ &\quad + \beta_1 \text{IMS/DMS} + \beta_2 \text{Full/Post} + \beta_3 \text{Rec/Roll} \\ &\quad + \gamma_1 \text{Equal} + \gamma_2 \text{Weight} + \gamma_3 \text{Top 10\%} + \gamma_4 \text{MSE} + \epsilon_i^h \end{aligned} \tag{4}$$

where $RMSE_i^h$ is the relative RMSE of permutation $i = 1, ..., 162$. By subtracting 1 the coefficients are more easily interpreted as indicating percent improvement (a negative coefficient) or percent deterioration (a positive coefficient) relative to our benchmark.

The α coefficients are associated with variables that indicate how an individual forecast is made: DMS takes the value 1 if only direct multi-step models are included and zero other wise, $Post$ takes the value 1 if only Great Moderation data is used and zero otherwise, and *Roll* takes the value 1 if only a rolling window of observations is used to estimate the model parameters and zero otherwise. The β's are associated with the different combination of the α's: IMS/DMS takes the value of 1 if the weighted forecast combines both IMS and DMS forecasts and zero otherwise, $Full/Post$ takes the value 1 if the weighted forecast combines both the Full and Post samples, and $Rec/Roll$ takes the value 1 if the weighted forecast combines both recursive and rolling estimation schemes. The γ's are associated with how the weighted forecasts are constructed: *Equal* takes the value 1 if either the Average or Median averaging methods are used and zero otherwise, $Weight$ takes the value 1 if the models are weighted unequally and zero otherwise, Top 10% takes the value 1 if the averaging only uses the Top 10% of forecast and zero otherwise, and MSE takes the value 1 if MSE weights are used and zero otherwise.

4.4 Results for Nominal Variables

Table 3 shows our decomposition results for both headline and core CPI-based inflation. In each panel, the first six rows relate to the selection of models to average over while the latter four rows relate to the type of averaging method. We begin by looking at panel 1, that associated with headline inflation.

In the first two rows of panel 1, those associated with choosing to average over DMS models, IMS models, or both, there appears to be little statistically significant advantage to using any of these particular forecasting methods. The sole exception is that at the 12−month horizon where DMS models appear to be favored. The results are stronger for the choice of data used to estimate the models. Across all horizons there appears to be a significant advantage to using only Great Moderation data to estimate the models: Not only are the coefficients on Post significantly different from zero and negative, they are more negative than the coefficients associated with averaging over both Full and Post samples. The results for the choice of sampling scheme are a bit messier but still instructive. At the shortest and longest horizon, combining both the recursive and rolling schemes - as suggested by Clark and McCracken (2008) - appears to provide the largest average benefits in terms of reducing RMSEs. At the other horizons, simply using the rolling scheme tends to be best.

Moving to the final four rows of panel 1, the results for the type of averaging method clearly indicate that the simple equally weighted averaging methods are significantly worse than using the benchmark. At all horizons the coefficient associated with *Equal* is positive and different from zero. Unfortunately, the remaining 3 rows are less clearly interpretable.

While the MSE, $Weight$, and Top 10% coefficients are typically negative –suggesting that a Top 10% MSE weighted average might be best – the coefficients are only statistically significant in a few of instances at the longer horizons.

The results in panel 2, those associated with core inflation, are similar to the results for headline inflation with a few specific differences. The evidence in favor of using the DMS approach to forecasting is stronger at all horizons and significantly so. And again, for all but the shortest horizon, the evidence favors using only Great Moderation data for estimation of the model parameters. Similarly we find that using the rolling scheme or a combination of the rolling and recursive schemes is the preferred approach.

In the final four rows of panel 2, the results on the type of averaging method are much sharper than those for headline inflation. In all but the shortest horizons, the simple equally weighted averaging methods are significantly worse than using the benchmark. But at the 1-month horizon, it appears that using a simple averaging method does provide significant gains in forecast accuracy and moreover, those gains are larger than were we to use some form of weighting. For horizons longer than 1-month, the coefficients on Top 10% are all significantly negative which, along with the negative MSE and $Weight$ coefficients, suggest that a Top 10% MSE weighted average might be best.

4.5 Results for Real Variables

The results for the real variables, and in particular those for IP, are very different from those we saw for the nominal variables. A quick glance at the first six rows of panel 1 indicates quite clearly that the preferred type of models that should be averaged are now IMS forecasting models estimated recursively using the Full sample–a sharp contrast to the type of models chosen for both headline and core inflation. Moreover, in the final four rows of panel 1, it appears that while there is some evidence favoring MSE weights relative to BIC weights, the bulk of the evidence suggests that one would do even better to use one of the simple equally weighted averages rather than use a weighted or Top 10% weighted average.

The results in panel 2, those associated with the unemployment rate, are less clear cut than those for IP and even those for headline and core inflation. At the 3- and 6-month horizons the DMS approach to forecasting appears to be best but at the longest horizon the IMS appears to be best. Similarly, at the intermediate horizons, using the Great Moderation sample appears to be best but at the longest horizon using the Full sample appears to be best. And while using the rolling scheme or a combination of the recursive and rolling schemes tends to best at the shortest horizons, the recursive scheme clearly tends to dominate at the 6-month and longer horizons Finally, as was the case for the results for IP, it appears that while there is some evidence favoring MSE weights relative to BIC weights, the bulk of the evidence suggests that one would do even better to use one of the equally weighted averages rather than use a weighted or Top 10% weighted average.

4.6 Rankings

The results in Tables 3 and 4 give some indication as to which types of model averaging should be used and which classes of models should be averaged over. However, we should emphasize that those results are indicators of average treatment effects across all 162 permutations of averages and model classes. They do not necessarily indicate which permutations actually turn out to be the best. In Tables 5 and 6, we provide a brief description of those permutations that do turn out to perform best. In particular, we list the 10 best performing permutations of averaging methods and model classes, along with their respective relative RMSEs, for each variable and each of the 1-, 3-, and 12-month horizons.[6] In addition, we provide the 5 worst performing permutations for the sake of comparison.

The first panel of Table 5 provides the rankings for headline CPI-based inflation. There are several striking features. In line with the results from Table 1, at the 1- and 3-month horizons their are few if any gains to model averaging irrelevant of model class. But as the horizon increases to 12-months, gains of roughly ten percent are available when using Top 10% weighted averages; in line with the decomposition results from Table 3. In addition, across all horizons, all of the top ten ranked permutations use either the rolling scheme or a combination of the rolling with the recursive. In contrast all of the lowest ranked permutations exclusively use the recursive scheme. Finally, as suggested in Table 3, all but one of the worst five permutations makes use of the simple equally weighted averaging schemes.

In the second panel of Table 5, that associated with core inflation, we get a slightly different picture on the benefits of model averaging relative to model selection. In particular, as was the case in Table 1, there are consistent benefits to model averaging at all horizons so long as the right permutations of model averages and model classes are used. The top ten

[6] We present these three horizons for brevity. A complete set of results is available from the authors upon request.

permutations outperform the benchmark by roughly seven percent at the shortest horizon and by as much as twenty-five percent at the longest horizon. On the other hand, while the bottom five permutations outperform the benchmark at the 1-month horizon, they fail to do so at the 3- and 12-month horizons.

Interestingly, the types of model averages that do best and worst for core inflation match up nicely with the results in Table 3. At the shortest horizon, equally weighted averages tend to do best but as the horizon increases, the Top 10% weighted averages begin to dominate. And by in large, the class of models to average over match up well too: the top ten ranked permutations are dominated by DMS forecasting models estimated over the Great Moderation sample, or an average of the Great Moderation and Full samples, using the rolling scheme, or a combination of the rolling and recursive. One feature that does not match up is that at the 12-month horizon, the BIC weighted averages appear to be best while the results in Table 3 suggest the MSE weighted would do better.

The first panel of Table 6 provides the rankings for IP growth. As we saw in Table 2, the advantages to model averaging relative to model selection, while feasible, are not particularly large with a maximum of only five percent at the 12-month horizon. As indicated in the decomposition, the equally weighted averages seem to do best at the 1-month horizon but as the horizon increases to 3-months, Top 10% weighted models appear to gain some traction among the best performing averaging methods – a sharp contrast to the decomposition in Table 4. Apparently part of the problem is that many of the worst performing models are also Top 10% averages and hence when we average across all permutations, our decomposition in Table 4 indicates the equally weighted averages should perform better. One point that clearly matches our decomposition is the choice of sampling scheme: Nearly all of the best performing permutations average across models estimated recursively while all the worst performing permutations average across models estimated using the rolling scheme.

In the second panel of Table 6, that associated with forecasts of the unemployment rate, a few things pop out immediately. The first is that model averaging uniformly improves forecast accuracy across all horizons and all permutations. In fact, at the 3-month horizon the worst performing model average provides an improvement of ten percent relative to the benchmark. Also, across all horizons the best ten types of model averaging are of the Top 10% form. This is in sharp contrast with the prediction from our decomposition which predicted that the equally weighted averages tended to perform best. Even so, as was the case for our decomposition in Table 4, it appears that at the shortest horizon the rolling scheme appears to perform best but as the horizon increases the recursive scheme becomes preferred. At the 12-month horizon, all of the bottom five performing permutations use the rolling scheme.

5 Conclusion

Using the ALFRED real-time database, we provide empirical evidence on the real time benefits to model averaging monthly frequency forecasts of headline and core CPI-based inflation, growth in industrial production, and the unemployment rate. Our results support much of the literature on forecasting: Model averaging typically improves forecast accuracy relative to a benchmark chosen using model selection. Even so, we emphasize a different point that is typically glossed over in the literature on forecast averaging. The choice of models that we average across can make a big difference on the efficacy of the averaging methods.

Using a novel decomposition of the benefits to forecast averaging relative to using model selection methods a few rules of thumb seem to be evident. First, when forecasting either headline or core CPI-based inflation, DMS forecasting models estimated over the Great Moderation sample (or an average of the Great Moderation and Full samples) using the rolling scheme (or a combination of the rolling and recursive schemes) seem to perform best. Second when forecasting either IP growth or the unemployment rate, averaging over models estimated using the recursive scheme (or an average of the rolling and recursive) seems to perform best. Third, while the Top 10% averaging approach frequently provides the best improvements in forecast accuracy, it is not immune to performing poorly relative to equally weighted averages because past model performance does not always indicate future model performance.

References

Aiolfi, Marco and Timmermann, Allan. "Persistence in Forecasting Performance and Conditional Combination Strategies." *Journal of Econometrics*, 2006, 135, pp. 31-54.

Clark, Todd E. and McCracken, Michael W. "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts." *International Economic Review*, 2008, 50, pp. 363-395 .

Clark, Todd E. and McCracken, Michael W. "Averaging Forecasts from VARs with Uncertain Instabilities." *Journal of Applied Econometrics*, January/February 2010, 25(1), pp. 5-29

D'Agostino, Antonello; Giannone, Domenico and Surico, Paolo. "(Un)Predictability and Macroeconomic Stability." Working Paper Series 605, European Central Bank, 2006.

Elliott, Graham, and Timmermann, Allan. "Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions." *Journal of Econometrics*, 2004, 122, pp. 47-79.

Faust, Jon and Wright, Jonathan H. "Comparing Greenbook and Reduced Form Forecasts using a Large Real-time Dataset." *Journal of Business and Economic Statistics*, October 2009, 27(4), pp. 468-479.

Garratt, Anthony; Koop, Gary and Vahey, Shaun P. "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty." Reserve Bank of New Zealand Discussion Paper 2006/02, 2006.

Giannone, Domenico; Reichlin, Lucrezia and Small, David. "Nowcasting: The Real Time Informational Content of Macroeconomic Data Releases." *Journal of Monetary Economics*, 2008 55(4), pp. 665-676.

Hansen, Bruce. "Least Squares Forecast Averaging." *Journal of Econometrics*, 2008, 146, pp. 342-350.

Hansen, Bruce. "Multi-Step Forecast Model Selection." Manuscript, University of Wisconsin, April 2010.

Kapetanios, George; Labhard, Vincent; and Price, Simon. "Forecasting using Bayesian and Information Theoretic Model Averaging: An Application to UK inflation." *Journal of Business and Economic Statistics*, January 2008, 26, pp. 33-41.

Kascha, Christian and Ravazzolo, Francesco. "Combining Inflation Density Forecasts." *Journal of Forecasting*, 2010, 29, pp. 231-250.

Levin, Andrew T. and Piger, Jeremy M. "Is Inflation Persistence Intrinsic in Industrial Economies?" Manuscript, University of Oregon, 2006.

Marcellino, Massimiliano; Stock, James H and Watson, Mark W. 2006. "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead." *Journal of Econometrics*, 2006, 135, pp. 499-526.

Smith, Jeremy and Kenneth, Wallis F. "A Simple Explanation of the Forecast Combination Puzzle." *Oxford Bulletin of Economics and Statistics*, 2009, 71(3), pp. 331-355.

Stock, James H. and Watson, Mark. “Combination Forecasts of Output Growth in a Seven–Country Data Set.” *Journal of Forecasting*, 2004, 23, pp. 405-430.

Table 1: **RMSE of Out-of-Sample Forecasts of Nominal Variables**

Headline CPI	1-month	3-month	6-month	12-month	24-month
BIC, Recursive, IMS, Full	3.560	**2.741**	1.622	1.146	0.800
Random Walk	1.151	1.519	2.298	2.851	4.234
Median	0.995	1.018	**0.990**	0.934	0.760
Average, all forecasts	0.995	1.021	1.008	0.952	0.816
MSE Weight, all forecasts	0.995	1.018	0.993	0.934	0.778
MSE Weight, Top 10%	1.000	1.030	1.006	0.943	0.821
BIC Weight, all forecasts	**0.994**	1.024	1.007	0.946	0.781
BIC Weight, Top 10%	**0.994**	1.023	0.996	**0.913**	**0.667**
Core CPI					
BIC, Recursive, IMS, Full	1.233	0.805	0.606	0.586	0.591
Random Walk	1.198	1.580	1.858	1.855	1.954
Median	0.938	0.942	0.899	0.867	0.827
Average, all forecasts	**0.931**	0.962	0.944	0.944	0.967
MSE Weight, all forecasts	0.934	0.938	**0.884**	**0.840**	0.827
MSE Weight, Top 10%	0.958	0.949	0.893	0.848	0.834
BIC Weight, all forecasts	0.936	0.946	0.918	0.918	0.922
BIC Weight, Top 10%	0.946	**0.932**	0.888	0.842	**0.810**

Notes:

(i) Values associated with the first row are RMSEs. The remaining values are ratios of RMSEs relative to that of the first row.

Table 2: **RMSE of Out-of-Sample Forecasts of Real Variables**

Industrial Production	1-month	3-month	6-month	12-month	24-month
BIC, Recursive, IMS, Full	9.951	6.229	5.050	4.136	**2.837**
Random Walk	1.125	1.263	1.313	1.561	2.281
Median	**0.985**	0.972	0.986	**0.994**	1.070
Average, all forecasts	**0.985**	**0.970**	**0.981**	**0.994**	1.055
MSE Weight, all forecasts	0.986	**0.970**	0.982	0.996	1.056
MSE Weight, Top 10%	0.988	0.974	0.982	1.011	1.067
BIC Weight, all forecasts	0.987	0.972	0.983	1.001	1.072
BIC Weight, Top 10%	0.989	0.990	1.005	1.023	1.094
Unemployment Rate					
BIC, Recursive, IMS, Full	0.167	0.301	0.463	0.696	1.023
Random Walk	0.995	0.958	0.947	0.982	1.031
Median	0.936	0.882	0.864	0.922	0.936
Average, all forecasts	0.935	0.877	0.857	0.917	0.922
MSE Weight, all forecasts	0.935	0.877	0.856	0.916	0.926
MSE Weight, Top 10%	**0.922**	**0.865**	**0.836**	**0.910**	0.969
BIC Weight, all forecasts	0.935	0.879	0.862	0.918	**0.919**
BIC Weight, Top 10%	0.938	0.895	0.889	0.953	0.985

Notes:
(i) Values associated with the first row are RMSEs. The remaining values are ratios of RMSEs relative to that of the first row.

Table 3: **Decomposition Regression of Nominal Variables**

CPI	1-month	3-month	6-month	12-month	24-month
DMS	0.000	−0.004	−0.001	−0.016*	0.014
DMS/IMS	0.000	−0.001	0.001	−0.005	−0.000
Post	−0.011***	−0.027***	−0.049***	−0.066***	−0.147***
Post/Full	−0.009***	−0.019***	−0.035***	−0.050***	−0.113***
Roll	−0.004***	−0.045***	−0.075***	−0.086***	−0.141***
Rec/Roll	−0.012***	−0.031***	−0.050***	−0.078***	−0.177***
Equal	0.018***	0.074***	0.088***	0.082***	0.087**
Weighted	−0.001	−0.002	−0.000	0.001	−0.017
Top 10%	0.000	−0.002	−0.011*	−0.020**	−0.047***
MSE	−0.001	0.002	−0.006	−0.022***	−0.011
N	162	162	162	162	162

Core CPI					
DMS	−0.000	−0.010*	−0.034***	−0.085***	−0.152***
DMS/IMS	−0.000	−0.004	−0.011	−0.025	−0.044
Post	0.001	−0.041***	−0.075***	−0.130***	−0.230***
Post/Full	−0.002	−0.034***	−0.060***	−0.102***	−0.176***
Roll	−0.002*	−0.069***	−0.134***	−0.236***	−0.414***
Rec/Roll	−0.013***	−0.056***	−0.105***	−0.182***	−0.317***
Equal	−0.048***	0.046***	0.096***	0.215***	0.439***
Weighted	−0.004***	−0.013**	−0.011	−0.015	−0.027
Top 10%	0.009***	−0.011**	−0.025**	−0.049***	−0.073***
MSE	0.004**	0.000	−0.020**	−0.040**	−0.041
N	162	162	162	162	162

Notes:
(i) ***, **, and * denote statistical signifiance at the 1%, 5%, and 10% levels, respectively.
(ii) Each column in each panel provides the coefficients associated with a distinct OLS estimated version of equation (4).

Table 4: **Decomposition Regression of Real Variables**

IP	1-month	3-month	6-month	12-month	24-month
DMS	−0.000	0.003***	0.013***	0.024***	0.025***
DMS/IMS	−0.000	0.000	0.004**	0.001	−0.003
Post	0.000	0.000	0.006***	0.006**	−0.001
Post/Full	−0.000	−0.000	0.004**	0.004	−0.001
Roll	0.012***	0.013***	0.049***	0.089***	0.147***
Rec/Roll	0.004***	0.006***	0.025***	0.050***	0.085***
Equal	−0.018***	−0.034***	−0.047***	−0.053***	−0.009**
Weighted	0.001***	0.002***	0.000	−0.000	0.003
Top 10%	0.003***	0.007***	0.007***	0.002	0.003
MSE	−0.002***	−0.006***	−0.003*	0.002	−0.006
N	162	162	162	162	162

UR					
DMS	0.000	−0.006***	−0.016***	0.004	0.077***
DMS/IMS	0.000	−0.002	−0.004*	−0.002	0.001
Post	0.001	−0.009***	−0.010***	−0.009***	0.026***
Post/Full	−0.001	−0.007***	−0.008***	−0.009***	0.014
Roll	−0.014***	−0.002	0.013***	0.054***	0.159***
Rec/Roll	−0.009***	−0.004**	0.004	0.023***	0.071***
Equal	−0.054***	−0.107***	−0.127***	−0.088***	−0.151***
Weighted	0.003***	0.003*	0.003	0.002	0.001
Top 10%	−0.007***	−0.005**	−0.006*	0.001	0.034***
MSE	−0.008***	−0.009***	−0.014***	−0.010***	−0.006
N	162	162	162	162	162

Notes:
(i) ***, **, and * denote statistical signifiance at the 1%, 5%, and 10% levels, respectively.
(ii) Each column in each panel provides the coefficients associated with a distinct OLS estimated version of equation (4).

Table 5: **Ranking of Model Averagings for Nominal Variables**

CPI	1-month		3-month		12-month	
1	BIC-DMS/IMS-Post-Rec/Roll	0.993	Top Ten% BIC-DMS/IMS-Post-Roll	1.005	Top Ten% MSE-DMS-Full-Roll	0.894
2	BIC-IMS-Post-Rec/Roll	0.993	Top Ten% BIC-DMS/IMS-Full-Roll	1.005	Top Ten% BIC-IMS-Post-Rec/Roll	0.899
3	BIC-DMS-Post-Rec/Roll	0.993	Top Ten% BIC-DMS/IMS-Full/Post-Roll	1.005	Top Ten% BIC-DMS/IMS-Post-Rec/Roll	0.900
4	Simple Avg-IMS-Post-Rec/Roll	0.994	Top Ten% BIC-IMS-Full-Roll	1.005	Top Ten% MSE-DMS-Full-Rec/Roll	0.902
5	Simple Avg-DMS-Post-Rec/Roll	0.994	Top Ten% BIC-IMS-Post-Roll	1.005	Top Ten% BIC-IMS-Full/Post-Rec/Roll	0.911
6	Simple Avg-DMS/IMS-Post-Rec/Roll	0.994	Top Ten% BIC-IMS-Full/Post-Roll	1.005	Top Ten% BIC-DMS/IMS-Full/Post-Rec/Roll	0.913
7	BIC-DMS/IMS-Full/Post-Rec/Roll	0.994	Top Ten% BIC-DMS-Post-Roll	1.006	Top Ten% MSE-DMS/IMS-Full-Roll	0.915
8	BIC-IMS-Full/Post-Rec/Roll	0.994	Top Ten% BIC-DMS-Full-Roll	1.006	Top Ten% MSE-DMS/IMS-Full-Rec/Roll	0.919
9	BIC-DMS-Full/Post-Rec/Roll	0.994	Top Ten% BIC-DMS-Full/Post-Roll	1.006	MSE-DMS-Full-Roll	0.919
10	Top Ten% BIC-DMS-Post-Rec/Roll	0.994	Median-DMS-Full-Roll	1.006	BIC-DMS-Post-Rec/Roll	0.921
⋮						
158	Simple Avg-IMS-Full-Rec	1.032	BIC-IMS-Full-Rec	1.113	Median-DMS/IMS-Full-Rec	1.195
159	Simple Avg-DMS-Full-Rec	1.032	Median-DMS-Full-Rec	1.113	Median-DMS-Full-Rec	1.196
160	Median-DMS-Full-Rec	1.040	Simple Avg-IMS-Full-Rec	1.114	Simple Avg-DMS/IMS-Full-Rec	1.197
161	Median-DMS/IMS-Full-Rec	1.040	Median-DMS/IMS-Full-Rec	1.117	Median-IMS-Full-Rec	1.201
162	Median-IMS-Full-Rec	1.040	Median-IMS-Full-Rec	1.120	Simple Avg-DMS-Full-Rec	1.212
Core CPI						
1	Simple Avg-DMS-Full-Rec/Roll	0.929	Top Ten% BIC-DMS-Full-Roll	0.928	Top Ten% BIC-DMS-Post-Roll	0.761
2	Simple Avg-DMS/IMS-Full-Rec/Roll	0.929	Top Ten% BIC-DMS-Post-Roll	0.928	Top Ten% BIC-DMS-Full-Roll	0.761
3	Simple Avg-IMS-Full-Rec/Roll	0.929	Top Ten% BIC-DMS-Full/Post-Roll	0.928	Top Ten% BIC-DMS-Full/Post-Roll	0.761
4	MSE-DMS-Full-Rec/Roll	0.930	MSE-IMS-Post-Roll	0.929	Top Ten% BIC-DMS-Full-Rec/Roll	0.769
5	MSE-DMS/IMS-Full-Rec/Roll	0.930	MSE-IMS-Full/Post-Roll	0.929	Top Ten% BIC-DMS-Full/Post-Rec/Roll	0.780
6	MSE-IMS-Full-Rec/Roll	0.930	Top Ten% BIC-DMS/IMS-Full-Rec/Roll	0.930	Top Ten% BIC-DMS-Post-Rec/Roll	0.783
7	Simple Avg-DMS-Full/Post-Rec/Roll	0.931	Top Ten% BIC-DMS-Full-Rec/Roll	0.930	MSE-DMS-Post-Roll	0.797
8	Simple Avg-DMS/IMS-Full/Post-Rec/Roll	0.931	Top Ten% BIC-DMS-Full/Post-Rec/Roll	0.931	MSE-DMS-Full/Post-Roll	0.798
9	Simple Avg-IMS-Full/Post-Rec/Roll	0.931	MSE-DMS/IMS-Post-Roll	0.931	Top Ten% MSE-DMS-Full-Roll	0.799
10	Top Ten% MSE-DMS-Full-Rec/Roll	0.932	MSE-DMS/IMS-Full/Post-Roll	0.931	BIC-DMS-Full/Post-Roll	0.801
⋮						
158	Top Ten% MSE-DMS/IMS-Post-Roll	0.966	BIC-DMS/IMS-Full-Rec	1.105	Median-IMS-Full-Rec	1.397
159	Simple Avg-IMS-Full-Rec	0.967	Median-IMS-Full-Rec	1.108	Top Ten% BIC-DMS/IMS-Full-Rec	1.436
160	Simple Avg-DMS/IMS-Full-Rec	0.967	BIC-IMS-Full-Rec	1.115	BIC-DMS/IMS-Full-Rec	1.478
161	Simple Avg-DMS-Full-Rec	0.967	Simple Avg-DMS/IMS-Full-Rec	1.116	BIC-IMS-Full-Rec	1.484
162	Top Ten% MSE-IMS-Post-Roll	0.967	Simple Avg-IMS-Full-Rec	1.133	Simple Avg-IMS-Full-Rec	1.523

Table 6: **Ranking of Model Averagings for Real Variables**

IP	1-month		3-month		12-month	
1	Median-IMS-Full/Post-Rec	0.978	Top Ten% BIC-DMS-Full-Rec	0.962	BIC-IMS-Full-Rec	0.949
2	Median-DMS-Full/Post-Rec	0.978	Top Ten% MSE-IMS-Full/Post-Rec	0.964	Simple Avg-IMS-Full-Rec	0.949
3	Median-DMS/IMS-Full/Post-Rec	0.978	Top Ten% MSE-DMS-Full-Rec	0.964	BIC-DMS/IMS-Full-Rec	0.949
4	Simple Avg-DMS-Full/Post-Rec	0.981	Top Ten% BIC-IMS-Full-Rec	0.964	Top Ten% BIC-DMS/IMS-Full-Rec	0.949
5	Simple Avg-DMS/IMS-Full/Post-Rec	0.981	Top Ten% MSE-DMS/IMS-Full-Rec	0.965	Median-IMS-Full-Rec	0.950
6	Simple Avg-IMS-Full/Post-Rec	0.981	Simple Avg-IMS-Post-Rec	0.965	MSE-IMS-Full-Rec	0.950
7	Top Ten% BIC-DMS-Full-Rec	0.981	Median-IMS-Post-Rec	0.965	MSE-DMS/IMS-Full-Rec	0.951
8	Top Ten% BIC-IMS-Full-Rec	0.981	Top Ten% MSE-IMS-Full-Rec	0.965	Simple Avg-DMS/IMS-Full-Rec	0.952
9	Top Ten% BIC-DMS/IMS-Full-Rec	0.981	Top Ten% MSE-DMS/IMS-Full/Post-Rec	0.965	Median-DMS/IMS-Full-Rec	0.952
10	BIC-DMS/IMS-Full/Post-Rec	0.981	MSE-IMS-Post-Rec	0.965	Simple Avg-IMS-Full/Post-Rec	0.954
⋮						
158	Top Ten% BIC-IMS-Full-Roll	1.001	Top Ten% BIC-IMS-Full-Roll	0.996	MSE-DMS-Full-Roll	1.088
159	Top Ten% BIC-DMS-Full-Roll	1.001	Top Ten% BIC-IMS-Post-Roll	0.996	Median-DMS-Full-Roll	1.090
160	Top Ten% BIC-DMS-Full/Post-Roll	1.001	Top Ten% MSE-DMS-Post-Roll	0.997	Median-DMS-Full/Post-Roll	1.090
161	Top Ten% BIC-IMS-Post-Roll	1.001	Top Ten% BIC-DMS-Full/Post-Rec/Roll	0.997	Median-DMS-Post-Roll	1.090
162	Top Ten% BIC-IMS-Full/Post-Roll	1.001	Top Ten% BIC-DMS-Post-Rec/Roll	0.998	Top Ten% MSE-DMS-Full-Roll	1.093
UR						
1	Top Ten% MSE-IMS-Post-Roll	0.905	Top Ten% MSE-DMS-Post-Rec/Roll	0.848	Top Ten% MSE-DMS-Post-Rec	0.855
2	Top Ten% MSE-DMS/IMS-Post-Roll	0.905	Top Ten% MSE-DMS-Post-Rec	0.849	Top Ten% MSE-DMS-Post-Rec/Roll	0.860
3	Top Ten% MSE-DMS-Post-Roll	0.905	Top Ten% MSE-DMS-Post-Roll	0.851	Top Ten% BIC-DMS-Full-Rec	0.863
4	Top Ten% MSE-IMS-Full/Post-Roll	0.913	Top Ten% MSE-DMS/IMS-Post-Rec/Roll	0.851	Top Ten% BIC-DMS-Post-Rec	0.867
5	Top Ten% MSE-IMS-Post-Rec/Roll	0.913	Top Ten% MSE-DMS/IMS-Post-Rec	0.851	Top Ten% MSE-DMS-Full/Post-Rec	0.867
6	Top Ten% MSE-DMS/IMS-Full/Post-Roll	0.913	Top Ten% MSE-DMS/IMS-Post-Roll	0.853	Top Ten% BIC-DMS-Full/Post-Rec	0.874
7	Top Ten% MSE-DMS/IMS-Post-Rec/Roll	0.914	Top Ten% MSE-IMS-Post-Rec	0.854	Top Ten% MSE-DMS-Full/Post-Rec/Roll	0.875
8	Top Ten% MSE-DMS-Full/Post-Roll	0.914	Top Ten% MSE-IMS-Post-Rec/Roll	0.854	Top Ten% MSE-DMS/IMS-Post-Rec	0.877
9	Top Ten% MSE-DMS-Post-Rec/Roll	0.915	Top Ten% MSE-IMS-Post-Roll	0.856	Top Ten% MSE-DMS/IMS-Full/Post-Rec	0.892
10	Top Ten% MSE-IMS-Full/Post-Rec/Roll	0.921	Top Ten% MSE-DMS-Full/Post-Roll	0.863	Top Ten% BIC-DMS/IMS-Full-Rec	0.892
⋮						
158	Simple Avg-DMS-Post-Rec	0.951	Top Ten% BIC-IMS-Post-Roll	0.899	MSE-DMS-Full-Roll	0.987
159	Simple Avg-IMS-Post-Rec	0.951	Top Ten% BIC-IMS-Full-Roll	0.899	Median-DMS-Full/Post-Roll	0.989
160	Median-DMS-Post-Rec	0.952	Top Ten% BIC-IMS-Full/Post-Roll	0.899	Median-DMS-Full-Roll	0.989
161	Median-DMS/IMS-Post-Rec	0.952	Median-DMS-Full-Rec	0.899	Median-DMS-Post-Roll	0.989
162	Median-IMS-Post-Rec	0.952	Top Ten% BIC-IMS-Full-Rec/Roll	0.900	Top Ten% MSE-DMS-Full-Roll	0.995