Forecast Combinations*

Marco Aiolfi
Goldman Sachs Asset Management

Carlos Capistrán
Banco de México

Allan Timmermann
UCSD

May 6, 2010

Abstract

We consider combinations of subjective survey forecasts and model-based forecasts from linear and non-linear univariate specifications as well as multivariate factor-augmented models. Empirical results suggest that a simple equal-weighted average of survey forecasts outperform the best model-based forecasts for a majority of macroeconomic variables and forecast horizons. Additional improvements can in some cases be gained by using a simple equal-weighted average of survey and model-based forecasts. We also provide an analysis of the importance of model instability for explaining gains from forecast combination. Analytical and simulation results uncover break scenarios where forecast combinations outperform the best individual forecasting model.

1 Introduction

Since the seminal work of Bates and Granger (1969), forecast combinations have come to be viewed as a simple and effective way to improve and robustify the forecasting performance over that offered by individual models. As a result, forecast combinations are now in widespread use in central banks, among private sector forecasters and in academic studies. Challenges still remain, however, to our understanding of what types of forecasts benefit most from combination, which combination schemes are optimal in a given forecast situation and when to expect the greatest advantage from forecast combination.

This article studies two issues in forecast combination. First, we consider ways to combine forecasts from surveys and time-series models. Second, we consider the possibility, advanced by Hendry and Clements (2004), that model instability can help explain the gains in forecasting performance resulting from combination. These two issues may in fact be closely related since survey participants can sometimes rapidly adjust their forecasts to shifts in the underlying data generating process. In contrast, constant parameter models will typically take longer to adjust to a change in the data generating process. Conversely, in a stable environment time-series models

*We thank Mike Clements and an anonymous referee for helpful and constructive comments. Christian Constandse and Gabriel López-Moctezuma provided excellent research assistance. Timmermann acknowledges support from CREATES, funded by the Danish National Research Foundation. The opinions in this paper are those of the authors and do not necessarily reflect the point of view of Banco de México.

may more efficiently summarize all available information than survey forecasts do. Survey forecasts may therefore serve as a hedge against breaks in the underlying data generating process when combined with time-series forecasts. Whether forecasts from time-series models, survey forecasts or some combination of the two performs best will depend on the degree of model instability, i.e., the frequency and magnitude of model and parameter changes, as well as the ability of survey participants to adapt their forecasts to such changes. Ultimately, which type of forecasting method works best is an empirical issue.

Forecast combinations require deciding both which forecasts to include and how to weight them. This is usually treated as a two-step process in which relatively little attention is paid to the first step (design of the model universe) compared with the second step (assigning weights to the included models). To the extent that the first step acquires much attention, this is often restricted to "trimming", i.e., excluding the models with the worst forecasting performance (Granger and Jeon (2004)). However, often little explicit thought goes into designing the universe of forecasts used in the combination in the first place. Even so, conclusions about the performance of forecast combinations necessarily have limited validity without posing them in the context of the universe of models being combined. For example, the common finding that an equal-weighted forecast is surprisingly difficult to beat−commonly known as the forecast combination "puzzle"−will cease to hold if a large number of poor forecasts dominates the universe of models being combined.

A particularly interesting aspect of the design of the universe of forecasts is whether including both subjective survey forecasts and forecasts from time-series models helps improve forecasting performance. Existing work on forecast combinations most often focuses on either combining forecasts from time-series models or combining subjective forecasts from sources such as the Survey of Professional Forecasters.[1] However, it is clearly of interest to see if combining these two types of forecasts leads to additional gains or whether one approach dominates the other. Forecasts from time-series models−whether linear or non-linear−are often closely correlated, but survey forecasts and time-series forecasts may be less so, opening the possibility of gains from combining these two types of forecasts. In particular, subjective forecasts can incorporate forward-looking information in a way that time-series forecasts cannot, e.g., as a result of a pre-announced or expected change in public policy. Conversely, there are issues related to the weighting of individual survey forecasts and changes in the composition of survey data due to entry and exit of survey participants (Engelberg et al. (2009) and Capistrán and Timmermann (2009)) that may put survey forecasts at a disadvantage.

To consider whether survey forecasts and model-based forecasts should be combined, we explore empirically how standard combination schemes applied to different classes of forecasting models and different types of forecasts (model-based versus survey forecasts) perform as the universe of forecasting models is varied.[2] We also study bias-adjustment of forecasts which take the form of

[1] Although see Clements and Harvey (2010) for an application to subjective and model-based probability forecasts. Elliott and Timmermann (2005) also propose a time-varying combination scheme in which the weights on survey forecasts and time-series (ARIMA) forecasts are allowed to follow a regime-switching process driven by a first-order Markov chain. They find evidence of significant time-variations in the optimal combination weights, suggesting that neither the time-series forecasts, nor the survey forecasts encompass the other group at all points in time.

[2] Under squared error loss, the value of an additional forecast can be measured through its correlation with the

regressions that augment the forecasts of interest, e.g., survey forecasts, with information such as current and past values of the predicted variable as well as common factors. In an empirical analysis that considers 14 U.S. macroeconomic variables, four horizons and a 17-year forecast evaluation sample, we find that the simple equal-weighted average of survey forecasts dominates the best forecast from any of the time-series models in around two-thirds of all cases. We also find that, for the most part, equal-weighted combinations of forecasts from time-series models and survey forecasts lead to improvements over using the time-series models alone but fail to systematically improve on using only the survey forecasts.

Turning to the second issue, instability of individual forecasting models offer an empirically plausible explanation for the good performance of forecast combinations. There is mounting evidence that the parameters of autoregressive models fitted to many economic time series change over time. For example, Stock and Watson (1996) undertake a systematic study of a wide variety of economic time series and find that the majority of these are subject to change. Diebold and Pauly (1987), Clements and Hendry (1998, 1999, 2006), Pesaran and Timmermann (2005) and Timmermann (2006) view model instability as an important determinant of forecasting performance and a potential reason for combining models. Little is known about how different forms of model instability affect forecast combinations, however. Thus, it is important to investigate the influence of this particular form of model misspecification on forecasting performance and its ability to explain the superior performance of simple combination schemes.

To address this issue, we evaluate the determinants of the performance of different forecast combination schemes in the presence of occasional shifts to the parameters of the data generating process. In the context of a simple factor model with stochastic breaks we derive analytical results that reveal the determinants of whether forecast combinations can be expected to outperform forecasts from the best model. Our results show that the relative factor variance, as well as the frequency and size of breaks play a role in determining the performance of forecast combinations.

Finally, we use simulations from dynamic factor models estimated on a range of US macroeconomic time series to investigate the extent to which factor models with and without breaks are able to match empirical findings on the performance of standard forecast combination schemes. We consider three separate scenarios reflecting breaks in the factor loadings, breaks in the dynamics of the underlying factors and breaks in the covariance matrix of the factors. Our results suggest that stable factor models without breaks are unable to match the performance of the forecast combination schemes, whereas breaks in the factor model−particularly if they occur in the factor loadings or in the factor dynamics−bring the results modestly closer to what is observed in the data. This suggests that parameter instability in dynamic factor models offer a partial explanation for why simple combinations outperform the best individual forecasting models.

The outline of the chapter is as follows. Section 2 discusses the design of the universe of forecasting models used in combining forecasts from time-series models and subjective survey forecasts.

forecast error from the existing forecasts and so the additional forecast should only be assigned a non-zero weight in the combination provided that it explains some of the errors from forecasts that are already included. This is related to the notion of forecast encompassing (Chong and Hendry (1986)).

Section 3 undertakes an empirical analysis using forecasts from univariate and multivariate linear models, non-linear models and survey forecasts. Section 4 provides analytical results that shed light on the performance of forecast combinations under model instability, while Section 5 presents empirical results on forecast combinations under breaks. Section 6 concludes.

2 Combinations of Survey and Time-series Forecasts

Users of modern forecasting techniques in economics and finance are faced with an abundance of predictor variables and a plethora of methods for generating forecasts. An important issue is therefore, first, how to summarize and implement such information and, second, whether to adopt a forecasting strategy that seeks out a single best forecasting method or, alternatively, attempt to combine forecasts generated by different models.

As always, we have to consider whether to pool forecasts or pool information. For some types of forecasts, e.g., survey forecasts, we can only pool the forecasts since we do not have access to the individual forecasters' information set. In contrast, when it comes to combining forecasts from surveys with information used by an econometric forecasting model, we have the option of either combining the forecasts or augmenting the survey forecast with information underlying the econometric model, and we will consider both strategies.

Combinations of survey and time-series forecasts are particularly interesting because they represent fundamentally different approaches to forecasting. Surveys reflect individual forecasters' subjective judgement and may be able to adjust rapidly to changes in the data generating process to the extent that these can be monitored or even predicted by survey participants. Conversely, forecasts from time-series models can efficiently incorporate past regularities in the data, but may take longer to adapt to changes in the data generating process.

To explore the performance of combinations of these types of forecasts, we consider a broad set of specifications which includes a variety of time-series models−univariate, multivariate, linear and non-linear−in addition to survey forecasts. First we describe the design of the experiments.

Suppose that we are interested in forecasting a generic variable of interest, y, multiple periods ahead. Let t be the time of the forecast, let $h \geq 1$ be the forecast horizon, so the object is to predict y_{t+h} given information known at time t. Finally, let $\hat{y}_{j,t+h|t}$ be the jth forecast which could be either a survey forecast or a forecast from a time-series model.

For combinations of forecasts we shall focus on equal-weighted combinations, $\hat{y}^{ew}_{t+h|t}$, of the type

$$\hat{y}^{ew}_{t+h|t} = N^{-1} \sum_{j=1}^{N} \hat{y}_{j,t+h|t}, \tag{1}$$

where N is the number of forecasts being combined. We focus on equal-weighted combinations because our sample is quite short and so estimating forecast combination weights is unlikely to lead to any improvements in forecasting performance (Smith and Wallis (2009)).

Turning to the second approach, i.e. extending survey forecasts with information used by the time-series models, we base forecasts on least squares estimates from a simple regression that includes information from subjective survey forecasts, $\bar{y}_{t+h|t}$, current and past values of the variable of interest, $\{y_{t-l}\}_{l=0}^{L}$ and economic factors, $\{\hat{f}_{k,t}\}_{k=1}^{K}$:

$$y_{t+h} = \alpha + \beta \bar{y}_{t+h|t} + \sum_{l=0}^{L} \lambda_l y_{t-l} + \sum_{k=1}^{K} \theta_k \hat{f}_{k,t} + \varepsilon_{t+h}. \tag{2}$$

Here $\{\hat{f}_{k,t}\}_{k=1}^{K}$ comprise the first K principal components of a larger set of multivariate information, X_t, estimated recursively, i.e. using information only available at time t. The lag length, L, and the number of principal components, K, are selected by using the Schwartz information criterion (SIC). The $h-$step ahead 'consensus' survey forecast, $\bar{y}_{t+h|t}$, is given by:

$$\bar{y}_{t+h|t} = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} y_{i,t+h|t} & \text{equal-weighted forecast} \\ \sum_{i=1}^{N}\left[y_{i,t+h|t} MSFE_{i,t|t-h}^{-1} / \sum_{i=1}^{N} MSFE_{i,t|t-h}^{-1}\right] & \text{inverse MSFE-weights} \\ \sum_{i=1}^{N}\left[y_{i,t+h|t}\exp(SIC_{i,t|t-h}) / \sum_{i=1}^{N}\exp(SIC_{i,t|t-h})\right] & \text{SIC weights.} \end{cases} \tag{3}$$

Here $MSFE_{i,t|t-h}$ is the mean squared forecast error of forecaster i at time t assuming an $h-$period forecast horizon, while $SIC_{i,t|t-h}$ is the value of the Schwarz information criterion for forecaster i at time t, again assuming an $h-$period horizon. In both cases we base calculations on a common overlapping sample so that $MSFE_{i,t|t-h}$ and $SIC_{i,t|t-h}$ are comparable across forecasters. The SIC weights can be viewed as approximate Bayesian Model Averaging weights (Garratt et al. (2009)).

Little is known about how forecasts generated by different classes of econometric models, e.g., univariate linear, multivariate linear, locally and globally approximating non-linear models, complement or substitute for one another in a particular forecasting experiment. To explore this point, we further consider combinations based on forecasts from two non-linear univariate models:

$$y_{t+h} = \alpha + \beta \hat{y}_{t+h|t}^{nl} + \sum_{l=0}^{L} \lambda_l y_{t-l} + \sum_{k=1}^{K} \theta_k \hat{f}_{k,t} + \varepsilon_{t+h}, \tag{4}$$

Again the non-linear univariate forecasts, $\hat{y}_{t+h|t}^{nl}$, are estimated recursively so that no use is made of information dated after period-t, for purposes of forecasting for period $t+h$. Following Swanson and White (1997) and Terasvirta (2006), we consider two types of non-linear univariate forecasts, a logistic smooth transition autoregressive (LSTAR) model and a neural network (NNET). The LSTAR model is

$$y_{t+h} = \left(\alpha_1 + \sum_{l_1=0}^{L_1} \lambda_{1,l_1} y_{t-l_1}\right) + d_t \left(\alpha_2 + \sum_{l_2=0}^{L_2} \lambda_{2,l_2} y_{t-l_2}\right) + \varepsilon_{t+h}, \tag{5}$$

with $d_t = 1/\left(1 + \exp\left(\gamma y_{t-1} - c\right)\right).$ L_1 and L_2 are chosen using the SIC. A linear regression is used to estimate the α's and λ's and then γ and c are estimated by minimizing the sum of squared residuals, repeating until convergence. The NNET is a single-layer feedforward model with J hidden units:

$$y_{t+h} = \beta_0 + \sum_{j=1}^{J} \beta_j g \left(\gamma_{0j} + \sum_{l=0}^{L} \gamma_{l,j} y_{t-l} \right) + \varepsilon_{t+h}, \tag{6}$$

where J and L are selected using the SIC, and g is the logistic function. Notice that in each case we consider direct forecasting models, i.e., period-t forecasts of y_{t+h} based on information known at time t. Another approach would be to iterate on a one-period forecasting model, but we do not consider this approach here.

An alternative interpretation of the "information pooling regressions" (2) and (4) is that they are bias-adjustment equations for the survey and non-linear forecasts, respectively. By regressing the actual value on the forecast as well as an intercept, current and past values of the variable of interest and economic factors, we orthogonalize the forecast error with respect to this additional information.

Note that the various approaches differ in terms of how many parameters they require estimating. This is an important issue. Elaborate combination schemes that require estimating multiple parameters have often been found to underperform the equal-weighted combination scheme (Clemen (1989), Stock and Watson (2001)). This finding reflects a bias-variance trade-off: equal-weighted combinations can be expected to be more biased but also may have a lower estimation error than data-driven weighting schemes which makes up for the bias (Granger and Ramanathan (1984), Smith and Wallis (2009)). The ability to efficiently explore this bias-variance trade-off depends on how many forecasts are included. To keep the aggregation scheme limited, "blocking" forecasts is a simple procedure, i.e. forecasts from linear models is one block, forecasts from non-linear models another block, factor-based forecasts a separate block and survey forecasts form yet another block (Aiolfi and Timmermann (2006)).[3]

We consider the following restricted universes of models:

1. Individual time-series models (univariate (AR), multivariate factor-based (AR_FAC), LSTAR and NNET);
2. Combinations of linear and non-linear time-series models;
3. Subjective (survey) forecasts only;
4. Combinations of survey forecasts and forecasts from time-series models.

3 Empirical Application

This section provides an empirical analysis of how the universe of models being combined as well as the chosen combination or bias-adjustment method affects the performance of various combination

[3] This option may not be open if survey information is being combined since the underlying models and information sets are typically unknown. Survey forecasts will, in any case, also be affected by forecasters' subjective adjustments and priors and so it is not clear how to construct a 'meta model' that encompasses information contained in the individual forecasts.

schemes relative to the individual forecasting models.

3.1 Data

To illustrate the empirical performance of the estimation and forecast combination methods, we study 14 variables that are covered by the Survey of Professional Forecasters (SPF) and have data from 1981Q3 through 2006Q4. The variables include the Real Gross Domestic Product (RGDP), the Unemployment Rate (UNEMP), the GDP Implicit Price Deflator (PGDP) and the Consumer Price Index (CPI). A brief description of these variables is provided in Table 1 (and in Table 2 of Capistrán and Timmermann (2009)). Actual values (i.e., the values that we forecast) are taken from the Federal Reserve Bank of Philadelphia's real-time database. Following Corradi, Fernandez and Swanson (2009) we use first release data.

Data on the subjective forecasts are taken from the SPF and contain one- through four-quarter-ahead forecasts. The data used to calculate common factors come from Stock and Watson (2009). This data set consists of 144 quarterly time series for the United States, spanning the sample 1959:Q1-2006Q4. The series are transformed as needed to eliminate trends by taking first or second differences, in many cases after taking logarithms. The variables, sources and transformations can be found in the appendix to Stock and Watson (2009).

3.2 Forecasting Performance of Individual models and surveys

Columns one through four of Table 2 report the forecasting performance of individual time-series models. In each case we use real time data to estimate the model parameters. This provides a fairer comparison with the survey forecasts since final release data is not available until well after the date where the survey forecast is formed. The first column of table 2 presents the annualized root mean squared forecast error (RMSFE) of the benchmark autoregressive (AR) model, i.e., the autoregressive part of equation (2). Subsequent columns show RMSFE ratios, computed relative to the AR forecasts, of several forecasting models.

To estimate the parameters of the initial AR models, we use the first 30 data points from 1981Q3 to 1988Q4. This means that the forecast evaluation sample runs from 1989Q1 to 2006Q4. Parameter estimation and forecasting is always done recursively, using an expanding window. Lag lengths, L, are determined using SIC and an exhaustive search, with a maximum of eight autoregressive terms. We produce 68 pseudo-out-of-sample forecasts for each horizon (the first $h-1$ forecasts are dropped in order to have the same number of forecasts for each horizon).[4]

The factors used in the autoregressive factor model (AR_FAC) shown in the second column are extracted using principal components.[5] The number of factors, K, is determined using SIC and an exhaustive search, from a maximum of five factors. For the LSTAR model (column 3), the number

[4]The exception is the TBILL series which only has 66 out-of-sample forecasts because the data in the SPF starts in 1982Q1.

[5]As in Stock and Watson (2009), the factor estimates are computed using a subset of 110 series that excludes higher level aggregates related by identities to the lower level subaggregates, instead of the full 144 series.

of lags, L_1 and L_2, is chosen using SIC, and an exhaustive search from a maximum of 3 lags. For the NNET model (column 4), the number of hidden units, J, is chosen in the same way, from a maximum of 4 units.[6] The remaining columns present the results using data from the SPF. The combinations are: equal weights (EW), inverse MSFE (IMSFE), and approximate BMA weights (ISIC). The last two combinations are based on the subset of forecasters that at the forecasting date have a minimum of 10 common contiguous observations. The ISIC calculates the SIC for each forecaster by projecting the actual values on a constant and the forecasts and then uses the weighting scheme in (3).

For approximately two-thirds of the variables and forecast horizons (36 out of 56 cases), the simple equal-weighted average of survey forecasts−which appears to be the best weighting scheme−performs better than any of the time-series models.[7] At the shortest horizon, h =1 quarter, the LSTAR model is the best performer for 8 of the 14 variables. In contrast, the purely autoregressive model, the factor-augmented multivariate model and the neural net only generate the lowest RMSFE-values in one, five and two cases, respectively. Interestingly, the value of the survey forecasts is not entirely driven by the early 1980s, a period before the "Great Moderation". When we repeat the exercise using a sample that starts in 1986Q1 instead of 1981Q3, we find that survey forecasts are better than time-series forecasts in 25 out of 56 cases. Hence, although not as dominant, survey forecasts are still the best performing group.

3.3 Combinations of model-based and subjective forecasts

Table 3 shows results from the bias-adjusted combination equation (2) and subsets of it. We only present results using the equal-weighted SPF forecasts because it is generally the best way of weighting the survey forecasts and also because the results were very similar when we used the other combinations applied to data from the SPF. Both autoregressive terms and individual forecasts are selected using SIC. The results are reported as RMSFE ratios with respect to the best autoregressive model. Adding the equal-weighted survey forecast to the AR terms reduces the RMSFE for some variables, notably the CPI, PGDP, RGDP and RFEDGOV. By including autoregressive terms, the AR-EW combination accounts for any autocorrelation not picked up by the professional forecasters. However, the simple average of survey forecasts listed in Table 2 generally outperforms combinations of the survey forecasts with the time series forecasts. For instance, AR_EW only outperforms the equal-weighted survey forecast in ten out of 56 cases.[8]

We saw in Table 2 that the simple equal-weighted survey forecasts outperformed the best individual time-series forecasts roughly two-thirds of the time. Table 3 shows that bias-adjusting

[6] For the forecasts produced with non-linear models, an "insanity filter" was used. It replaces a forecast more than six interquartile ranges away from the median with the previous observation. This filter replaces forecasts for less than 1% of the observations.

[7] Ang, Bekaert and Wei (2007) also find that survey forecasts have superior information when it comes to predicting inflation.

[8] Since we include a constant in the combination and allow for a non-zero weight on the SPF forecasts, this is related to the bias adjusted mean procedure proposed by Capistrán and Timmermann (2009). However, in contrast to the latter, here we also include autoregressive terms.

the equal-weighted survey forecast by adding either autoregressive or autoregressive and factor terms as in equation (2) generally does not lead to better forecasting performance. This is likely driven by the effect of estimation error associated with having to estimate the weights on the equal-weighted forecast in addition to any included autoregressive terms and factors. This suspicion is confirmed by the finding that in 46 out of 56 cases the equal-weighted average of forecasts from the AR model, the factor-augmented model and the surveys, i.e. equation (1) shown in the last column of Table 3, outperform the corresponding combination regression, equation (2). Hence, any gains from combining subjective and model-based forecasts seems to come from the included models and not so much from the estimated weights (i.e., the bias-adjustment). This can be explained by the short samples used here to estimate combination weights and the resulting large estimation errors.

Interestingly, forty percent of the time (i.e., 22 out of 56 cases) the equal-weighted average of forecasts from the autoregressive model, the factor-augmented model and the mean survey forecast improves upon the simple equal-weighted survey forecasts. This suggests that the survey forecasts are modestly biased and that in some cases this bias can be removed by augmenting the survey forecast with information from the time-series models. Moreover, the simple equal-weighted combination of survey forecasts and time-series forecasts, equation (1), almost never produces very poor forecasts.

3.4 Combinations of linear and non-linear models

Table 4 presents results for the combination of linear models, either univariate or multivariate, and non-linear models as in equations (1) and (4), once again based on recursive estimation and SIC selection using exhaustive search. The first two columns present the combinations of the non-linear forecasts with information underlying the autoregressive model in equation (4). For the AR_LSTAR specification, the RMSFE-values mostly fall between those obtained from the pure AR model and those generated by the LSTAR model and the combined forecasts only outperform both the AR forecasts and LSTAR forecasts in a few cases (7 out of 56). There is more room for improvement for the NNET forecasts extended to include information on current and past values of the predicted variable, in part because the NNET forecasts are relatively poor and so, ironically, the NNET forecasts may not be selected in (4) and so the forecast falls back on the autoregressive part.

Empirically, gains in forecast precision again seem to be sensitive to the way in which we combine the AR and the non-linear models. For a given variable, forecast horizon, and time period, we select the weights attached to each of the eight lags and to the forecast from the non-linear model. Hence, the non-linear forecast is only selected when it reduces the SIC with respect to the selected AR model. This helps to alleviate the problem that non-linear models sometimes produce extreme predictions and is what explains the high RMSFE ratios for unemployment when the nonlinear models are used alone (see Table 2). This finding is consistent with the observation by Teräsvirta (2006) that "combining nonlinear forecasts with forecasts from a linear model may sometimes lead to a series of forecasts that are more robust (contain fewer extreme predictions) than forecasts from the nonlinear model" (p. 438-439). Our results suggest that this conclusion is particularly true when inclusion of non-linear forecasts is dictated by a model selection criterion and the combination weights are estimated.

The columns labeled AR_FAC_LSTAR and AR_FAC_NNET in Table 4 present the results corresponding to equations (1) and (4). The third and fourth columns use linear regression to estimate the weights, (4), whereas the fifth and sixth columns combine the same models, but using equal weights as in (1). In the vast majority of cases the equal-weighted combinations continue to outperform the combinations based on estimated weights. However, when we use an equal-weighted average to combine the non-linear forecasts with forecasts from the univariate and multivariate linear models, we find that the RMSFE is reduced, relative to the best of the underlying models' RMSFE-values, for less than half of all cases.

3.5 Combination of the full set of forecasts

We finally combine the full set of forecasts under consideration. Again, we consider two combination schemes. One simply uses an equal-weighted average of the forecasts from the AR, AR_FAC, LSTAR and NNET models along with the mean survey forecasts in equation (1). The other scheme combines the forecasts from the two non-linear models and the surveys with up to eight autoregressive terms and five common factors. The results are presented in Table 5 both for cases with estimated weights and with equal weights. The latter is clearly easier to implement, since it only involves choosing the parameters for each block, i.e., the lags for the AR, the number of lags and factors for the AR_FAC, the number of lags in each regime of the LSTAR, etc. In contrast, model selection and parameter estimation is computationally intensive: for each variable, forecast horizon, and time period, $2\hat{}\,(8+5+1+1+1) = 65,536$ models are estimated and one gets selected by the SIC.[9]

The results indicate that extending the information set underlying the forecast combination often does not maintain the gains from combining only the survey forecasts. In fact only in 18 out of 56 cases does the equal-weighted combination of time-series and survey forecasts lead to an improvement over using the survey forecast alone.

Using equal weights dominates using estimated combination weights in roughly two-thirds of all cases (38 out of 56). However, the results vary a lot across the variables. For the unemployment series, estimating the weights dominates using equal weights, but for other variables, such as CPROF, HOUSING, or RSLGOV using equal-weights dominates. Furthermore, for most variables the results vary across horizons. An example of the latter is PGDP, for which estimating the weights dominates at horizons 1 and 3 quarters, but in turn is dominated by equal weights for horizons of 2 and 4 quarters.

We conclude from this analysis that using equal weights leads to better forecast performance than using estimated combination weights in roughly two-thirds of all cases where forecasts from

[9] There are up to eight autoregressive terms, five factors, one LSTAR forecast, one NNET forecast and one equal-weighted survey forecast.

time-series models and surveys are considered. Interestingly, even if they do not always deliver the most precise forecasts, forecast combinations, particularly equal-weighted ones, generally do not deliver poor performance and so from a "risk" perspective represent a relatively safe choice.

4 Model Instability

Whether forecast combinations offer a "safe pair of hands" depends, to some extent, on which sources of risk they help forecasters hedge against. In this section we address whether combinations improve forecasting performance in the presence of one particular type of risk, namely model instability.

Hendry and Clements (2004) argue that forecast combinations can provide insurance against extraneous (deterministic) structural breaks when individual forecasting models are misspecified. Their analysis provides supporting evidence that simple combinations can work well under a single end-of-sample break in the process governing the dynamics of the predictor variables. They consider a wide array of designs for the break and find that combinations work particularly well when the predictors are shifted in opposite directions and are positively correlated.

In support of the interpretation that model instability may explain the good average performance of forecast combination methods, the findings in Stock and Watson (2001) suggest that the performance of combined forecasts tends to be more stable than that of the individual constituent forecasts entering in the combinations. Interestingly, however, gains from combination methods that attempt to build in time-variations in the combination weights (either by discounting past performance or by modeling time-variations in the weights) have generally proved elusive.

We consider instability in the context of a simple common factor model. Factor models have played an important role in recent work on forecasting in the presence of large numbers of predictor variables and are widely used empirically to forecast macroeconomic and financial time series (Stock and Watson (2002), Bai and Ng (2002), Bovin and Ng (2006), Forni et al. (2000, 2005), Artis et al. (2005), and Marcellino (2004)). Furthermore, intuition can be gained in terms of distribution of factor loadings/exposures and variability of the individual factors.

We consider forecasting a single variable, Y_{0t}, by means of an array of state variables, $Y_{1t}, .., Y_{Nt}$ for $t = 1, ..., T$, where T is the sample size. Our starting point is a factor model of the form

$$\mathbf{Y} = \mathbf{F} \cdot \mathbf{L} + \varepsilon, \tag{7}$$

where $\mathbf{Y}$ is a $T \times (N+1)$ data matrix, $\mathbf{F}$ is a $T \times m$ matrix that contains the m factors, $\mathbf{L}$ is an $m \times (N+1)$ matrix of factor loadings, and ε is a $T \times (N+1)$ matrix of innovations with covariance matrix $E[\varepsilon\varepsilon'] = \mathbf{\Omega}$.

For a particular time period, $1 \leq t \leq T$, we can write the model

$$\mathbf{Y}_t = \mathbf{F}_t \cdot \mathbf{L} + \varepsilon_t, \quad t = 1, ..., T, \tag{8}$$

where $\mathbf{Y}_t = (Y_{0t}, \hat{Y}_{1t}, ..., \hat{Y}_{Nt})$ is now a $1 \times (N+1)$ vector, $\mathbf{F}_t$ is a $1 \times m$ vector of factors and ε_t are the innovations at time t. We use a zero subscript on the first element of $\mathbf{Y}_t$ to indicate that this is the variable whose values are being predicted. The remaining terms, $\hat{Y}_{1t}$, $\hat{Y}_{2t}, ..., \hat{Y}_{Nt}$ are forecasts as indicated by the hats. This is a highly stylized setup where individual forecasts have predictive content because of their correlation with future factor realizations.

Factor dynamics is assumed to be driven by an autoregressive process,

$$\mathbf{F}_t' = \mathbf{A}\mathbf{F}_{t-1}' + \varepsilon_{Ft}', \quad E[\varepsilon_{F_t}\varepsilon_{F_t}'] = \mathbf{\Sigma}, \tag{9}$$

where $E[\varepsilon_{F_t}'\varepsilon_t] = \mathbf{0}$. Following common practice, we further assume that the factors are orthogonal. There are many ways of specifying breaks to this process. We will assume that breaks are tracked by a state indicator, S_t, that can take two possible values, namely $S_t = 1$ or $S_t = 2$.[10] Breaks then take the form of a shift in some of the parameters of the model (8)-(9) as governed by a change in S_t. To this end we partition the factors into $\mathbf{F}_t = (\mathbf{F}_{1t}|\mathbf{F}_{2t})$, where $\mathbf{F}_{1t}$ is $1 \times m_1$ and $\mathbf{F}_{2t}$ is $1 \times m_2$ with $m_1 + m_2 = m$ and associated covariance matrices $Var(F_{1t}) = \Sigma_{F_1}$, $Var(F_{2t}) = \Sigma_{F_2}$, $Cov(F_{1t}, F_{2t}) = \Sigma_{F_{12}}$. Further, suppose that the matrix of loading coefficients can be partitioned as follows

$$\mathbf{L} = (\mathbf{l}_0 \ \mathbf{L}_r)\,, \ \mathbf{L}_r = \begin{pmatrix} \mathbf{L}_{1r} \\ \mathbf{L}_{2r} \end{pmatrix}, \ \mathbf{l}_0 = \begin{pmatrix} \mathbf{l}_{01} \\ \mathbf{l}_{02} \end{pmatrix},$$

where the dimension of $\mathbf{l}_0$ is $m \times 1$, $\mathbf{L}_r$ is $m \times N$, $\mathbf{L}_{1r}$ is $m_1 \times N$, $\mathbf{L}_{2r}$ is $m_2 \times N$, $\mathbf{l}_{01}$ is $m_1 \times 1$ and $\mathbf{l}_{02}$ is $m_2 \times 1$.

Suppose that breaks take the form of a shift in the loadings of the target variable, Y_{0t} from the first m_1 factors to the last m_2 factors:

$$Y_{0t} = \mathbf{1}_{\{S_t=1\}}\mathbf{F}_{1t}\mathbf{l}_{01} + \mathbf{1}_{\{S_t=2\}}\mathbf{F}_{2t}\mathbf{l}_{02} + \varepsilon_{0t}, \tag{10}$$

where $\mathbf{1}_{\{S_t=1\}}$ is an indicator variable that equals one at time t if $S_t = 1$, otherwise is zero. Similarly $\mathbf{1}_{\{S_t=2\}}$ is one at time t only if $S_t = 2$. Assuming that the unconditional state probabilities are given by $\Pr(S_t = 1) = p$, $\Pr(S_t = 2) = 1 - p$, the population value of the projection coefficient of Y_{0t} on $\hat{Y}_{it}$ $(i = 1, .., N)$, β_i, is given by

$$\beta_i = (\mathbf{l}_i'\mathbf{\Sigma}_F\mathbf{l}_i + \sigma_{\varepsilon_i}^2)^{-1}(p\mathbf{l}_{01}'\mathbf{\Sigma}_{F_1}\mathbf{l}_{i1} + (1-p)\mathbf{l}_{02}'\mathbf{\Sigma}_{F_2}\mathbf{l}_{i2}), \tag{11}$$

where $\mathbf{\Sigma}_F = Var(\mathbf{F}_t)$, $\mathbf{\Sigma}_{Fj} = Var(\mathbf{F}_{jt})$, $(j = 1, 2)$, $\mathbf{l}_i$ is the ith $m \times 1$ column vector of $\mathbf{L}_r$, while $\mathbf{l}_{i1}$ and $\mathbf{l}_{i2}$ are $m_1 \times 1$ and $m_2 \times 1$ vectors formed as the ith columns of $\mathbf{L}_{1r}$ and $\mathbf{L}_{2r}$, respectively.

Following common practice, we assume that each forecasting model is based on a linear projection of the target variable, Y_{0t}, on one of the forecasts, $\hat{Y}_{it}$, $(i = 1, ..., N)$. To establish the

[10] Our results can easily be generalized to the case where S_t takes an arbitrary number of discrete values.

properties of such forecasts, notice from (8) that the population projection of Y_{0t} on $\hat{Y}_{it}$ is

$$\hat{Y}_{0t|i} = \beta_i \mathbf{F}_t \mathbf{l}_i + \beta_i \varepsilon_{it}.$$

The associated forecast error is $e_{it} = Y_{0t} - \hat{Y}_{0t|i}$. Moments of the joint distribution of these forecast errors are characterized as follows. Conditional on $S_t = 1$, we have

$$\begin{aligned} Var(e_{it}) &= (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \mathbf{\Sigma}_{F_1} (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1}) + \beta_i^2 \mathbf{l}_{i2}' \mathbf{\Sigma}_{F_2} \mathbf{l}_{i2} + 2\beta_i (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \Sigma_{F_{12}} \mathbf{l}_{i2} + \sigma_{\varepsilon_0}^2 + \beta_i^2 \sigma_{\varepsilon_i}^2, \\ Cov(e_{it}, e_{jt}) &= (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \mathbf{\Sigma}_{F_1} (\mathbf{l}_{01} - \beta_j \mathbf{l}_{j1}) + \beta_i \beta_j \mathbf{l}_{i2}' \mathbf{\Sigma}_{F_2} \mathbf{l}_{j2} \\ &\quad + \beta_j (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \mathbf{\Sigma}_{F_{12}} \mathbf{l}_{j2} + \beta_i (\mathbf{l}_{01} - \beta_j \mathbf{l}_{j1})' \mathbf{\Sigma}_{F_{12}} \mathbf{l}_{i2} + \sigma_{\varepsilon_0}^2. \end{aligned} \tag{12}$$

Conditional on $S_t = 2$, we have

$$\begin{aligned} Var(e_{it}) &= \beta_i^2 \mathbf{l}_{i1}' \mathbf{\Sigma}_{F1} \mathbf{l}_{i1} + (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2})' \mathbf{\Sigma}_{F_2} (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2}) + 2\beta_i \mathbf{l}_{i1}' \Sigma_{F_{12}} (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2}) + \sigma_{\varepsilon_0}^2 + \beta_i^2 \sigma_{\varepsilon_i}^2, \\ Cov(e_{it}, e_{jt}) &= \beta_i \beta_j \mathbf{l}_{i2}' \mathbf{\Sigma}_{F_1} \mathbf{l}_{j1} + (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2})' \mathbf{\Sigma}_{F_2} (\mathbf{l}_{02} - \beta_j \mathbf{l}_{j2}) \\ &\quad + \beta_j (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2})' \mathbf{\Sigma}_{F_{12}}' \mathbf{l}_{j1} + \beta_i (\mathbf{l}_{02} - \beta_j \mathbf{l}_{j2})' \mathbf{\Sigma}_{F_{12}}' \mathbf{l}_{i1} + \sigma_{\varepsilon_0}^2. \end{aligned} \tag{13}$$

This means that the MSFE associated with the ith forecast error is

$$\begin{aligned} MSFE(e_{it}) &= p \left\{ (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \mathbf{\Sigma}_{F_1} (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1}) + \beta_i^2 \mathbf{l}_{i2}' \mathbf{\Sigma}_{F_2} \mathbf{l}_{i2} + 2\beta_i (\mathbf{l}_{01} - \beta_i \mathbf{l}_{i1})' \Sigma_{F_{12}} \mathbf{l}_{i2} \right\} \\ &\quad + (1-p) \left\{ \beta_i^2 \mathbf{l}_{i1}' \mathbf{\Sigma}_{F1} \mathbf{l}_{i1} + (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2})' \mathbf{\Sigma}_{F_2} (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2}) + 2\beta_i \mathbf{l}_{i1}' \Sigma_{F_{12}} (\mathbf{l}_{02} - \beta_i \mathbf{l}_{i2})' \right\} \\ &\quad + \sigma_{\varepsilon_0}^2 + \beta_i^2 \sigma_{\varepsilon_i}^2. \end{aligned}$$

Similarly, the MSFE associated with the average forecast $\bar{Y}_t = N^{-1} \sum_{i=1}^N \hat{Y}_{it}$, and the corresponding error, $\bar{e}_t = (Y_{0t} - \bar{Y}_t)$, is

$$\begin{aligned} MSFE(\bar{e}_t) &= p \left\{ (\mathbf{l}_{01} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}}{N})' \mathbf{\Sigma}_{F_1} (\mathbf{l}_{01} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}}{N}) + \left(\sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}'}{N} \right) \mathbf{\Sigma}_{F_2} \left(\sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}}{N} \right) \right. \\ &\quad \left. +2(\mathbf{l}_{01} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}}{N})' \Sigma_{F_{12}} \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}}{N} \right\} \\ &\quad + (1-p) \left\{ \left(\sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}'}{N} \right) \mathbf{\Sigma}_{F_1} \left(\sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}}{N} \right) + (\mathbf{l}_{02} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}}{N})' \mathbf{\Sigma}_{F_2} (\mathbf{l}_{02} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}}{N}) \right. \\ &\quad \left. +2 \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i1}'}{N} \Sigma_{F_{12}} (\mathbf{l}_{02} - \sum_{i=1}^N \frac{\beta_i \mathbf{l}_{i2}}{N}) \right\} \\ &\quad + \sigma_{\varepsilon_0}^2 + \left(\sum_{i=1}^N \frac{\beta_i^2 \sigma_{\varepsilon_i}^2}{N} \right). \end{aligned}$$

These expressions are quite general and difficult to interpret so we simplify the model to the case where $m = N = 2$ and $\mathbf{L} = \mathbf{I}_2$ so each forecast tracks one factor only and is thus misspecified.[11] Assuming only a break in the factor loadings, $\mathbf{l}_0$, and setting $\mathbf{A} = \mathbf{0}$, we have

$$\begin{aligned} Y_{0t} &= \mathbf{1}_{\{S_t=1\}} F_{1t} + \mathbf{1}_{\{S_t=2\}} F_{2t} + \varepsilon_{0t}, \\ \hat{Y}_{1t} &= F_{1t} + \varepsilon_{1t}, \\ \hat{Y}_{2t} &= F_{2t} + \varepsilon_{2t}. \end{aligned} \tag{14}$$

All variables are assumed to be Gaussian with $F_{1t} \sim N(0, \sigma^2_{F_1})$, $F_{2t} \sim N(0, \sigma^2_{F_2})$, $\varepsilon_{0t} \sim N(0, \sigma^2_{\varepsilon 0})$, $\varepsilon_{1t} \sim N(0, \sigma^2_{\varepsilon_1})$, $\varepsilon_{2t} \sim N(0, \sigma^2_{\varepsilon_2})$, while the innovations, ε, are mutually uncorrelated and uncorrelated with the factors and $Cov(F_{1t}, F_{2t}) = \sigma_{F_1 F_2}$.

The linear projection of Y_t on Y_{1t} can be obtained through the least squares estimator $(i = 1, 2)$

$$\hat{\beta}_i = \left(\frac{1}{T} \sum_{t=1}^{T} Y_{it}^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} Y_{0t} Y_{it} \right)$$

with probability limits

$$\begin{aligned} p \lim(\hat{\beta}_1) &= \frac{p\sigma^2_{F_1} + (1-p)\sigma_{F_1F_2}}{\sigma^2_{F_1} + \sigma^2_{\varepsilon_1}} \equiv \beta_1, \\ p \lim(\hat{\beta}_2) &= \frac{(1-p)\sigma^2_{F_2} + p\sigma^2_{F_1F_2}}{\sigma^2_{F_2} + \sigma^2_{\varepsilon_2}} \equiv \beta_2. \end{aligned}$$

This leads to the following joint distribution of the forecast errors $e_{it} = Y_{0t} - \beta_i Y_{it}$ $(i = 1, 2)$:

Conditional on $S_t = 1$:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (1-\beta_1)^2 \sigma^2_{F_1} + \beta_1^2 \sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_0} & (1-\beta_1)\sigma^2_{F_1} - (1-\beta_1)\beta_2 \sigma_{F_1F_2} + \sigma^2_{\varepsilon_0} \\ (1-\beta_1)\sigma^2_{F_1} - (1-\beta_1)\beta_2 \sigma_{F_1F_2} + \sigma^2_{\varepsilon_0} & \sigma^2_{F_1} + \beta_2^2 \sigma^2_{F_2} - 2\beta_2 \sigma_{F_1F_2} + \beta_2^2 \sigma^2_{\varepsilon_2} + \sigma^2_{\varepsilon_0} \end{pmatrix} \right).$$

Conditional on $S_t = 2$:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \beta_1^2 \sigma^2_{F_1} + \sigma^2_{F_2} - 2\beta_1 \sigma_{F_1F_2} + \beta_1^2 \sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_0} & (1-\beta_2)\sigma^2_{F_2} - \beta_1(1-\beta_2)\sigma_{F_1F_2} + \sigma^2_{\varepsilon_0} \\ (1-\beta_2)\sigma^2_{F_2} - \beta_1(1-\beta_2)\sigma_{F_1F_2} + \sigma^2_{\varepsilon_0} & (1-\beta_2)^2 \sigma^2_{F_2} + \beta_2^2 \sigma^2_{\varepsilon_2} + \sigma^2_{\varepsilon_0} \end{pmatrix} \right).$$

[11] We could easily relax these assumptions and allow Y_{1t}, Y_{2t} to depend on both factors. However, these generalizations come with few additional insights at the cost of complicating the arithmetic and interpretation of the results.

Integrating across states, we get the MSFE values

$$\begin{aligned}
E[e_{1t}^2] &= E\left[((S_t - \beta_1)F_{1t} + (1-S_t)F_{2t} + \varepsilon_{0t} - \beta_1\varepsilon_{1t})^2\right] \\
&= \left(p(1-\beta_1)^2 + (1-p)\beta_1^2\right)\sigma_{F_1}^2 + (1-p)\sigma_{F_2}^2 - 2(1-p)\beta_1\sigma_{F_1F_2} + \beta_1^2\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_0}^2; \\
E[e_{2t}^2] &= E\left[(S_tF_{1t} + (1-S_t-\beta_2)F_{2t} + \varepsilon_{0t} - \beta_2\varepsilon_{2t})^2\right] \\
&= p\sigma_{F_1}^2 + \left(p\beta_2^2 + (1-p)(1-\beta_2)^2\right)\sigma_{F_2}^2 - 2p\beta_2\sigma_{F_1F_2} + \beta_2^2\sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_0}^2.
\end{aligned}$$

Similarly, the MSFE of the combined equal-weighted forecast $\hat{Y}_t^c = (\hat{Y}_{1t} + \hat{Y}_{2t})/2$ is

$$\begin{aligned}
E[e_{ct}^2] &= E\left[((S_t - \beta_1/2)F_{1t} + (1-S_t-\beta_2/2)F_{2t} + \varepsilon_{0t} - 0.5\beta_1\varepsilon_{1t} - 0.5\beta_2\varepsilon_{2t})^2\right] \\
&= (p(1-\beta_1/2)^2 + (1-p)\beta_1^2/4)\sigma_{F_1}^2 + (p\beta_2^2/4 + (1-p)(1-\beta_2/2)^2)\sigma_{F_2}^2 + \\
&\beta_1^2\sigma_{\varepsilon_1}^2/4 + \beta_2^2\sigma_{\varepsilon_2}^2/4 + \sigma_{\varepsilon 0}^2 - \left(p(1-\beta_1/2)\beta_2 + (1-p)\beta_1(1-\beta_2/2)\right)\sigma_{F_1F_2}.
\end{aligned}$$

For simplicity suppose that the factors are uncorrelated so $\sigma_{F_1F_2} = 0$. Then $E[e_{1t}^2] > E[e_{ct}^2]$ if

$$\begin{aligned}
&\left(p(1-\beta_1)^2 + (1-p)\beta_1^2 - p(1-\frac{\beta_1}{2})^2 - (1-p)\frac{\beta_1^2}{4}\right)\sigma_{F_1}^2 \\
&+ \left((1-p) - p\frac{\beta_2^2}{4} - (1-p)(1-\frac{\beta_2}{2})^2\right)\sigma_{F_2}^2 + \frac{3}{4}\beta_1^2\sigma_{\varepsilon_1}^2 - \frac{1}{4}\beta_2^2\sigma_{\varepsilon_2}^2 > 0. \qquad (15)
\end{aligned}$$

This condition is satisfied provided that

$$\left(\frac{3}{4}\beta_1^2 - p\beta_1\right)\sigma_{F_1}^2 + \left(\beta_2(1-p) - \frac{1}{4}\beta_2^2\right)\sigma_{F_2}^2 + \frac{3}{4}\beta_1^2\sigma_{\varepsilon_1}^2 - \frac{1}{4}\beta_2^2\sigma_{\varepsilon_2}^2 > 0.$$

Using the definition of β_1, β_2, this can be written

$$\frac{\sigma_{F_2}^2}{\sigma_{F_1}^2} > \frac{1}{3}\left(\frac{p}{1-p}\right)^2 \frac{(1+\sigma_{\varepsilon_2}^2/\sigma_{F_2}^2)}{(1+\sigma_{\varepsilon_1}^2/\sigma_{F_1}^2)}. \qquad (16)$$

Hence it is more likely that the MSFE from model 1 exceeds that of the equal-weighted forecast provided that the second factor explains a large part of the variation in Y relative to the first factor ($\sigma_{F_2}^2/\sigma_{F_1}^2$ is high); the second factor is in effect more often ($p/(1-p)$ is low) and the second forecast has a low noise-to-signal ratio relative to that of model 1 ($\sigma_{\varepsilon_2}^2/\sigma_{F_2}^2$ is low relative to $\sigma_{\varepsilon_1}^2/\sigma_{F_1}^2$).

By symmetry, the conditions for $E[e_{2t}^2] > E[e_{ct}^2]$ are

$$\frac{\sigma_{F_2}^2}{\sigma_{F_1}^2} < 3\left(\frac{p}{1-p}\right)^2 \frac{(1+\sigma_{\varepsilon_2}^2/\sigma_{F_2}^2)}{(1+\sigma_{\varepsilon_1}^2/\sigma_{F_1}^2)}. \qquad (17)$$

If both (16) and (17) hold, the average forecast will have a lower population MSFE than that of the individual models. We summarize this result in the following proposition.

Proposition 1 *Suppose $(Y_t, \hat{Y}_{1t}, \hat{Y}_{2t})$ is generated by the process (14) with mutually uncorrelated, Gaussian factors and innovations and that $S_t = 1$ with constant probability p while the probability* *that $S_t = 2$ is $(1-p)$. Then the population MSFE of the equal-weighted combined forecast will be lower than the population MSFE of the best model provided that the following condition holds:*

$$\frac{1}{3}\left(\frac{p}{1-p}\right)^2 \frac{(1+\sigma^2_{\varepsilon_2}/\sigma^2_{F_2})}{(1+\sigma^2_{\varepsilon_1}/\sigma^2_{F_1})} < \frac{\sigma^2_{F_2}}{\sigma^2_{F_1}} < 3\left(\frac{p}{1-p}\right)^2 \frac{(1+\sigma^2_{\varepsilon_2}/\sigma^2_{F_2})}{(1+\sigma^2_{\varepsilon_1}/\sigma^2_{F_1})}.$$

Figure 1 shows the MSFE values from models 1, 2 and the combined forecast as a function of p under different assumptions about relative factor variances and the variances of the error terms. In Panel 1 the two forecasts are of equal quality so the equal-weighted forecast is close to being optimal and always dominates the individual forecasts. In contrast, in Panels 2 and 3 it is only for low (Panel 2) and high (Panel 3) values of p that the combined forecast is best. Finally in Panel 4 we show an example where the forecast combination dominates the individual forecasts provided that p lies between 0.2 and 0.8. Since the probability of a switch in the factor structure can be measured by $p(1-p)$, this is also the region where 'breaks' are most likely, suggesting that model instability can be one reason for the good performance of forecast combinations.

An equivalent condition for $E[e^2_{1t}] > E[e^2_{ct}]$ is (letting $k_1 = (1/3)\sigma^2_{F_1}(p/(1-p))^2$)

$$(\sigma^2_{F_1} + \sigma^2_{\varepsilon_1})\sigma^4_{F_2} - k_1\sigma^2_{F_1}\sigma^2_{F_2} - k_1\sigma^2_{\varepsilon_2}\sigma^2_{F_1} > 0 \tag{18}$$

with real roots

$$\frac{k_1\sigma^2_{F_1} \pm \sqrt{k_1^2\sigma^4_{F_1} + 4k_1\sigma^2_{\varepsilon_2}\sigma^2_{F_1}(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}}{2(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}. \tag{19}$$

Similarly, $E[e^2_{2t}] > E[e^2_{ct}]$ provided that

$$(\sigma^2_{F_1} + \sigma^2_{\varepsilon_1})\sigma^4_{F_2} - k_2\sigma^2_{F_1}\sigma^2_{F_2} - k_2\sigma^2_{\varepsilon_2}\sigma^2_{F_1} < 0, \tag{20}$$

where $k_2 = 3\sigma^2_{F_1}(p/(1-p))^2$. This has roots

$$\frac{k_2\sigma^2_{F_1} \pm \sqrt{k_2^2\sigma^4_{F_1} + 4k_2\sigma^2_{\varepsilon_2}\sigma^2_{F_1}(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}}{2(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}. \tag{21}$$

In both cases, one root will be negative while the other is positive. This means that the combined, equal-weighted forecast dominates both individual forecasts provided that $\sigma^2_{F_2}$ lies in the interval

$$\left[\frac{k_1\sigma^2_{F_1} + \sqrt{k_1^2\sigma^4_{F_1} + 4k_1\sigma^2_{\varepsilon_2}\sigma^2_{F_1}(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}}{2(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}; \frac{k_2\sigma^2_{F_1} + \sqrt{k_2^2\sigma^4_{F_1} + 4k_2\sigma^2_{\varepsilon_2}\sigma^2_{F_1}(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}}{2(\sigma^2_{F_1}+\sigma^2_{\varepsilon_1})}\right]. \tag{22}$$

Figure 2 plots the lower and upper bands of this range of values for $\sigma^2_{F_2}$. The lower band reflects the performance of the combination against the first model while the upper band reflects performance against the second model. As p gets higher, the likelihood that model 2 is the true model gets lower so that, when this model is in fact valid, it must be better than model 1 by a larger margin, i.e. $\sigma^2_{F_2}$ must rise. Conversely, the second model's performance when it is valid cannot be too good relative to model 1 - otherwise a strategy of only using the forecast from model 2 would be the dominant strategy - which is the intuition for why the upper bound is needed. For values of $\sigma^2_{F_2}$ between the two bounds, the equal-weighted combination dominates the individual forecasts.

The above conclusions are not altered by allowing for non-zero correlations across the two factors. After some algebra, we can modify equations (16) and (17) by adding a term

$$\frac{2p\left(3+2(\sigma^2_{\varepsilon_1}/\sigma^2_{F_1}+\sigma^2_{\varepsilon_2}/\sigma^2_{F_2})\right)\sigma_{F_1F_2}}{3(1-p)(1+\sigma^2_{\varepsilon_2}/\sigma^2_{F_2})\sigma^2_{F_1}}$$

to the right hand side of (16) and subtracting (1/3) times this term from the right hand side of (17). As expected, this tends to narrow the range where the combined forecast works best.

5 Breaks and Empirical Forecasting Performance

To illustrate the performance of a range of forecast combination methods in the presence of breaks, we finally conduct a Monte Carlo experiment in the context of a simple common factor model.

5.1 Setup

The data set that we consider is the same as that used in Stock and Watson (2004). It consists of up to 43 quarterly time series for the US economy over the period 1959Q1 – 1999Q4, although some series are available only for a shorter period. The 43 series comprise a range of asset prices (including returns, interest rates and spreads), measures of real economic activity, wages and prices, and various measures of the money stock.[12] To achieve stationarity, the series are transformed as needed to eliminate trends by taking first differences, in many cases after taking logarithms. We then standardize the transformed data so as they have sample mean equal to zero and unitary sample variance. Let $y_{i,t}$, $i = 1, ..., N$ denote the individual standardized transformed data series and assume the following factor structure for the transformed variables:

$$Y_t = \Lambda F_t + \epsilon_t, \tag{23}$$

where $Y_t = (y_{1,t}, ..., y_{N,t})'$, $E[\epsilon_t'\epsilon_t] = R$ and the factor dynamics is assumed to be governed by a first-order autoregressive process:

$$F_t = AF_{t-1} + u_t, \tag{24}$$

[12] Following Stock and Watson (2004) the variables are subject to the following transformations. First, in a few cases the series contain a large outlier−such as spikes associated with strikes−and these outliers are replaced by interpolated values. Second, series that show significant seasonal variation are seasonally adjusted using a linear approximation to X11 in order to avoid problems with non-linearities, see Ghysels, Granger and Siklos (1996). Third, data series available on a monthly basis are aggregated to get quarterly observations.

where $E[u_t'u_t] = \Omega$. We estimate $\widehat{\Lambda}$ and $\widehat{F}_t$ by principal components using a balanced panel subset of the full data set that includes 36 series. We focus on the first four factors which account for 65% of the variance of the panel. Given $\widehat{\Lambda}$ and $\widehat{F}_t$ we can estimate $\widehat{\epsilon}_t = Y_t - \widehat{\Lambda}\widehat{F}_t$ and $\widehat{R} = \widehat{\epsilon}_t'\widehat{\epsilon}_t/(T-1)$. We then fit a VAR(1) model to the first four factors and estimate $\widehat{A}$ and $\widehat{\Omega}$. Under a stable factor setup, factors and data are simulated as follows:[13]

$$\begin{aligned} \widehat{F}_t^m &= \widehat{A}\widehat{F}_{t-1}^m + u_t^m, \qquad u_t^m \sim N(0, \hat{\Omega}) \\ \widehat{Y}_t^m &= \widehat{\Lambda}\widehat{F}_t^m + \eta_t^m, \qquad \eta_t^m \sim N(0, \hat{R}), \end{aligned} \tag{25}$$

where $m = 1, ..., 100$ refers to the particular Monte Carlo simulation. We allow for instability in the factor model by means of breaks in either the factor loadings, Λ, in the factor dynamics, A, or in the covariance matrix of the factors, Ω. Generalizing the setup from Section 4, breaks are generated through an indicator series S_t^m from a Markov Switching process that can take two possible values $S_t^m = 1$ or $S_t^m = 2$ with transition probabilities $p_{11} = p_{22} = 0.8$, where $p_{ij} = prob(S_t = i | S_{t-1} = j)$.

We consider three breakpoint scenarios:

1. Breaks in the factor loadings:

$$\begin{aligned} &\widehat{F}_t^m = \widehat{A}\widehat{F}_{t-1}^m + u_t^m, \qquad u_t^m \sim N(0, \hat{\Omega}) \\ &\widehat{Y}_t^m = 2 \times \mathbf{1}_{\{S_t^m=1\}}\widehat{\Lambda}_{1:2}\widehat{F}_{1:2,t}^m + 2 \times \mathbf{1}_{\{S_t^m=2\}}\widehat{\Lambda}_{3:4}\widehat{F}_{3:4,t}^m + \eta_t^m, \qquad \eta_t^m \sim N(0, \hat{R}) \end{aligned}$$

2. Breaks in the dynamics of the factors:

$$\begin{aligned} S_t^m &= \begin{bmatrix} \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=2\}} & \mathbf{1}_{\{S_t^m=2\}} \\ \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=2\}} & \mathbf{1}_{\{S_t^m=2\}} \\ \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=2\}} & \mathbf{1}_{\{S_t^m=2\}} \\ \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=1\}} & \mathbf{1}_{\{S_t^m=2\}} & \mathbf{1}_{\{S_t^m=2\}} \end{bmatrix} \\ \widehat{A}_t^m &= S_t^m \odot \widehat{A} \\ \widehat{F}_t^m &= \widehat{A}_t^m \widehat{F}_{t-1}^m + u_t^m, \qquad u_t^m \sim N(0, \hat{\Omega}) \\ \widehat{Y}_t^m &= \widehat{\Lambda}\widehat{F}_t^m + \eta_t^m, \qquad \eta_t^m \sim N(0, \hat{R}) \end{aligned}$$

[13] The recursion is initialized by setting $\widehat{F}_0^m = N(0, \hat{\Omega})$. We simulate time series with 200 observations and discard the first 65 observations to match the average length of the actual data.

3. Breaks in the covariance matrix of the factors:

$$\begin{aligned}
\psi_t^m &= 2 \times \mathbf{1}_{\{S_t^m=1\}} + 0.5 \times \mathbf{1}_{\{S_t^m=2\}} \\
\Psi_t^m &= [\psi_t^m, \psi_t^m, 1/\psi_t^m, 1/\psi_t^m] \\
\widehat{\Omega}_t^m &= \Psi_t^m \odot \widehat{\Omega} \\
\widehat{F}_t^m &= \widehat{A}\widehat{F}_{t-1}^m + u_t^m, \qquad u_t^m \sim N(0, \hat{\Omega}_t^m) \\
\widehat{Y}_t^m &= \widehat{\Lambda}\widehat{F}_t^m + \eta_t^m, \qquad \eta_t^m \sim N(0, \hat{R})
\end{aligned}$$

Following the analysis of Stock and Watson (2001), we focus on linear forecasting models. Specifically, we consider simple autoregressions with lag lengths selected recursively using SIC with up to $L = 4$ lags:

$$y_{i,t+1} = c_i + \sum_{l=0}^{L} \lambda_{i,l} y_{i,t-l} + \epsilon_{i,t+1}. \tag{26}$$

We also consider all possible bivariate autoregressive models that include a single additional regressor, $y_{j,t}$, drawn from the full set of transformed variables:

$$y_{i,t+1} = c_i + \sum_{l_1=0}^{L_1} \lambda_{i,l_1} y_{t-l_1} + \sum_{l_2=0}^{L_2} \lambda_{i,j,l_2} y_{j,t-l_2} + \epsilon_{i,t+1}. \tag{27}$$

Lag lengths are again selected recursively using the SIC with between one and four lags of $y_{j,t}$ ($L_1 = 4$) and between zero and four lags of $y_{i,t}$ ($L_2 = 4$). Parameter estimation and forecasting are also done recursively, using an expanding window. To estimate the parameters of the initial AR models, we use the first 40 data points. For each of the simulated series we re-estimate the parameters of the linear forecasting models and use these to produce out-of-sample forecasts.[14]

We next present the results of the Monte Carlo experiment. We also report the results obtained using actual data as a benchmark for the Monte Carlo experiment.

5.2 Empirical Results

Table 6 presents results for the out-of-sample MSFE performance of a range of alternative forecast combination methods also considered by Aiolfi and Timmermann (2006) such as the mean forecast, triangular kernel weights (TK, which weights the forecasts by the inverse of their historical MSFE rank) and weights that are inversely proportional to the MSFE (IMSFE). Out-of-sample MSFE values are reported relative to the MSFE-values produced by the previous best (PB) forecasting model. Because the previous best model at a given point in time depends on the individual models' track record up to that point, the identity of the previous best model may change through time.

[14] To avoid extreme values, forecasts greater than four recursive standard deviations of the target variable are replaced by the recursive mean of the dependent variable computed at the time of the forecast.

We first note that the performance of the equal-weighted forecast combination−i.e., the average forecast computed across the 37 univariate and bivariate models−generally (across all variables) is much better than the performance of the previous best model. Specifically, the ratio of out-of-sample MSFE-values of the average forecast (mean) over the previous best model is 0.844. The simulations with a stable factor structure (no break) in the second column generate a much higher value of 0.97, suggesting that the improvement offered by the forecast combinations cannot be well explained in the context of a stable factor structure.

Turning to the results for the factor models with breaks, the form of the break process appears to affect the ability of the simulations to match the actual performance of the combinations. If the break occurs in the covariance matrix of the factor innovations, the performance of the mean forecast is slightly worse than under the stable factor structure (0.97). In contrast, if the break affects the factor loadings, the relative performance of the equal-weighted forecast improves to around 0.93 which is more in line with the empirical data. Breaks in the coefficients determining the factor dynamics also lead to some improvement in the performance of the forecast combinations relative to the single best model.

Finally consider the disaggregate results by category of the economic variables. It is clear that breaks do not affect all time series to the same extent. For example, breaks in the factor loadings bring the simulated results more closely in line with the data for the monetary aggregates. Such breaks do a far worse job for returns, interest rates and spreads although, in most cases, the relative MSFE performance under breaks is closer to the results for the data than in the absence of a break. Overall, the rather modest improvements in the performance of forecast combinations due to breaks suggest that model instability in the form considered here can only be part of the reason why forecast combinations perform better than a strategy of selecting the single best model.

6 Conclusion

Forecast combinations are in widespread use, representing a pragmatic approach for dealing with the misspecification biases that affect individual forecasting models. Since individual models may be biased in different directions, it is important to consider which types of forecasts to combine, i.e., forecasts from linear versus nonlinear models, forecasts from univariate versus multivariate models and combinations of time-series forecasts with subjective survey forecasts.

Our empirical results suggest that the simple equal-weighted survey forecast dominates the best forecast from a time-series model around two-thirds of the time. However, there is some room for improvement by using a simple equal-weighted average of survey forecasts and forecasts from various time-series models and this approach rarely generates poor out-of-sample forecasts. Even if forecast combinations do not always deliver the most precise forecasts they generally do not deliver poor performance and so from a "risk" perspective represent a safe choice.

We also present analytical and simulation results on the performance of a range of forecast combination schemes under instability in dynamic factor models characterizing the dependence structure across variables. These results are modestly encouraging in that they suggest that the performance of forecast combinations gets closer to what is observed in actual data under a factor model that is subject to occasional breaks in parameters, particularly if they take the form of a change in the factor loadings or in the coefficients determining the factor dynamics.

References

[1] Aiolfi, M. and A. Timmermann, 2006, Persistence in Forecasting Performance and Conditional Combination Strategies. Journal of Econometrics, 135, 31-53.

[2] Ang, A., G. Bekaert, and M Wei, 2007, Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better? Journal of Monetary Economics, 54, 1163-1212.

[3] Artis, M., A. Banerjee and M. Marcellino, 2005, Factor Forecasts for the U.K. Journal of Forecasting, 24, 279-298.

[4] Bai, J. and S. Ng, 2002, Determining the Number of Factors in Approximate Factor Models. Econometrica 70, 191-221.

[5] Bates, J.M. and C.W.J. Granger, 1969, The Combination of Forecasts. Operations Research Quarterly 20, 451-468.

[6] Bovin, J., and S. Ng, 2006, Are More Data Always Better for Factor Analysis? Journal of Econometrics, 132, 169-194.

[7] Capistrán, C., and A. Timmermann, 2009, Forecast Combination with Entry and Exit of Experts. Journal of Business and Economic Statistics, 27(4), 428-440.

[8] Chong, Y.Y, and D.F. Hendry, 1986, Econometric Evaluation of Linear macro-economic Models. Review of Economic Studies, 53, 671-690.

[9] Clemen, R.T., 1989, Combining Forecasts: A Review and Annotated Bibliography. International Journal of Forecasting, 5, 559-581.

[10] Clements, M.P. and D. I. Harvey, 2010, Forecast Encompassing Tests and Probability Forecasts. Forthcoming in Journal of Applied Econometrics.

[11] Clements, M.P. and D.F. Hendry, 1998, Forecasting Economic Time Series. Cambridge University Press.

[12] Clements, M.P. and D.F. Hendry, 1999, Forecasting Non-stationary Economic Time Series. The MIT Press.

[13] Clements, M.P. and D.F. Hendry, 2006, Forecasting with Breaks. Pages 605-658 in Elliott, G., C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting vol. 1. North-Holland.

[14] Corradi, V., A. Fernandez, and N.R. Swanson, 2009, Information in the Revision Process of Real-time Datasets. Journal of Business and Economic Statistics, 27(4), 455-467.

[15] Diebold, F.X. and P. Pauly, 1987, Structural Change and the Combination of Forecasts. Journal of Forecasting ,6, 21-40.

[16] Elliott, G., and A. Timmermann, 2005, Optimal Forecast Combination under Regime Switching. International Economic Review, 46(4), 1081-1102.

[17] Engelberg, J., C.F. Manski and J. Williams, 2009, Assessing the Temporal Variation of Macroeconomic Forecasts by a Panel of Changing Composition. Mimeo, Northwestern.

[18] Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2000, The Generalized Factor Model: Identification and Estimation. The Review of Economics and Statistics, 82, 540-554.

[19] Forni, M, M. Hallin, M. Lippi, and L. Reichlin, 2005, The Generalized Dynamic Factor Model: Forecasting and One Sided Estimation. Journal of the American Statistical Association, 100, 830-840.

[20] Garratt, A., G. Koop, E. Mise, and S.P. Vahey, 2009, Real-time Prediction with U.K. Monetary Aggregates in the Presence of Model Uncertainty. Journal of Business and Economic Statistics, 27(4), 480-491.

[21] Ghysels, E., C.W.J. Granger and P.L. Siklos, 1996, Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process? Journal of Business and Economic Statistics, 14, 374-86.

[22] Granger, C.W.J. and R. Ramanathan, 1984, Improved Methods of Combining Forecasts. Journal of Forecasting, 3, 197-204.

[23] Granger, C.W.J. and Y. Jeon, 2004, Thick Modeling, Economic Modelling, 21, 323-343.

[24] Hendry, D.F. and M.P. Clements, 2004, Pooling of Forecasts. Econometrics Journal, 7, 1-31.

[25] Marcellino, M., 2004, Forecast Pooling for Short Time Series of Macroeconomic Variables. Oxford Bulletin of Economic and Statistics, 66, 91-112.

[26] Pesaran, M.H. and A. Timmermann, 2005, Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks. Journal of Econometrics, 129, 183-217.

[27] Smith, J. and K.F. Wallis, 2009, A Simple Explanation of the Forecast Combination Puzzle. Oxford Bulletin of Economics and Statistics, 71(3), 331-355.

[28] Stock, J.H. and M.W. Watson, 1996, Evidence on Structural Instability in Macroeconomic Time Series Relations. Journal of Business and Economic Statistics, 14, 11-30.

[29] Stock, J.H. and M. Watson, 2001, A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, in R.F. Engle and H. White, eds., Festschrift in Honour of Clive Granger. Cambridge University Press, 1-44.

[30] Stock, J.H. and M.W. Watson 2002. Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics, 20, 147-162.

[31] Stock, J.H. and M. Watson, 2004, Combination Forecasts of Output Growth in a Seven-Country Data Set. Journal of Forecasting, 23, 405-430.

[32] Stock, J.H., and M.W. Watson, 2009, Forecasting in Dynamic Factor Models Subject to Structural Instability. In J.L. Castle and N. Shephard (eds.), The Methodology and Practice of Econometrics. A Festschrift in Honour of David Hendry. Oxford University Press, 173-205.

[33] Swanson, N.R. and H. White, 1997, Forecasting Economic Time Series Using Flexible Versus Fixed Specification and Linear versus Nonlinear Econometric Models. International Journal of Forecasting 13, 439–462.

[34] Teräsvirta, T., 2006, Forecasting Economic Variables with Nonlinear Models. In Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.), Handbook of Economic Forecasting. Elsevier, 413–457.

[35] Timmermann, A., 2006, Forecast Combinations. In Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.), Handbook of Economic Forecasting. Elsevier, 135–196.

Table 1. Variable description

Name	Description	Transformation	Sample
CPI	Forecasts for the CPI Inflation Rate. SA, annual rate, percentage points. Quarterly forecasts are annualized quarter-over-quarter percent changes.	None	1981q3-2006q4
CPROF	Forecasts for the quarterly level of nominal corporate profits after tax excluding IVA and CCAdj. SA, annual rate, billions of dollars.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(CPROF_{t+h}/CPROF_{t+h-1}\right)\right]400$	1981q3-2006q4
HOUSING	Forecasts for the quarterly average level of housing starts. SA, annual rate, millions.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(HOUSING_{t+h}/HOUSING_{t+h-1}\right)\right]400$	1981q3-2006q4
INDPROD	Forecasts for the quarterly average of the index of industrial production. SA, index, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(INDPROD_{t+h}/INDPROD_{t+h-1}\right)\right]400$	1981q3-2006q4
NGDP	Forecasts for the quarterly level of nominal GDP. SA, annual rate, billions of dollars.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(NGDP_{t+h}/NGDP_{t+h-1}\right)\right]400$	1981q3-2006q4
PGDP	Forecasts for the quarterly level of the GDP price index. SA, index, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(PGDP_{t+h}/PGDP_{t+h-1}\right)\right]400$	1981q3-2006q4
RCBI	Forecasts for the quarterly level of real change in private inventories. SA, annual rate, base year varies.	None	1981q3-2006q4
RCONSUM	Forecasts for the quarterly level of real personal consumption expenditures. SA, annual rate, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(RCONSUM_{t+h}/RCONSUM_{t+h-1}\right)\right]400$	1981q3-2006q4
RFEDGOV	Forecasts for the quarterly level of real federal government consumption and gross investment. SA, annual rate, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(RFEDGOV_{t+h}/RFEDGOV_{t+h-1}\right)\right]400$	1981q3-2006q4
RGDP	Forecasts for the quarterly level of real GDP. SA, annual rate, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(RGDP_{t+h}/RGDP_{t+h-1}\right)\right]400$	1981q3-2006q4
RRESINV	Forecasts for the quarterly level of real residential fixed investment. SA, annual rate, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(RRESINV_{t+h}/RRESINV_{t+h-1}\right)\right]400$	1981q3-2006q4
RSLGOV	Forecasts for the quarterly level of real state and local government consumption and gross investment. SA, annual rate, base year varies.	GR QoQ change, expressed in annualized percentage points: $\left[\ln\left(RSLGOV_{t+h}/RSLGOV_{t+h-1}\right)\right]400$	1981q3-2006q4
TBILL	Forecasts for the quarterly average three-month Treasury bill rate. Percentage points.	None	1982q1-2006q4
UNEMP	Forecasts for the quarterly average unemployment rate. SA, percentage points.	None	1981q3-2006q4

Notes: GR QoQ stands for Growth Rate, Quarter on Quarter.

Table 2. Forecast performance of individual models and survey information

Horizon	RMSFE	RMSFE ratios with respect to AR					
		Time Series Models			Survey Forecasts		
	AR	AR_FAC	LSTAR	NNET	EW	IMSE	ISIC
				CPI			
h = 1	1.4516	0.9996	0.6204	1.0672	0.9574	0.9663	0.9881
h = 2	1.5634	0.9880	1.0014	1.1564	0.8903	0.8986	0.9529
h = 3	1.5368	1.0396	1.0671	1.2740	0.9008	0.9182	0.9507
h = 4	1.5801	1.0497	1.0754	1.3831	0.9030	0.9047	0.9271
				CPROF			
h = 1	20.4380	0.9951	0.9539	0.9599	0.8872	0.9044	0.9091
h = 2	20.3715	1.0909	0.9832	0.9867	0.8595	0.8916	0.8981
h = 3	20.4939	1.0016	1.0271	1.0217	0.8662	0.8787	0.8883
h = 4	20.8992	1.0224	1.0057	0.9346	0.8935	0.8916	0.9036
				HOUSING			
h = 1	27.8404	1.0101	0.8981	0.8631	0.8918	0.8919	0.8905
h = 2	28.1350	0.9823	1.0728	1.0349	0.9459	0.9457	0.9417
h = 3	29.1552	1.0490	0.9222	0.9431	0.8761	0.8771	0.8759
h = 4	28.1420	1.0621	0.9795	0.9791	0.8939	0.8949	0.8996
				INDPROD			
h = 1	3.6548	0.8666	0.6994	1.2522	0.9649	0.9639	0.9623
h = 2	4.2634	1.0219	1.0979	1.0137	0.8996	0.8984	0.8935
h = 3	4.3009	1.0124	0.9127	1.0012	0.9404	0.9338	0.9278
h = 4	4.1792	1.0200	1.1211	1.0969	0.9837	0.9657	0.9619
				NGDP			
h = 1	2.1161	1.1249	0.8199	1.1649	0.9707	1.0214	1.0507
h = 2	2.1398	1.1134	1.2885	1.2668	0.9799	1.0117	1.0258
h = 3	2.3450	0.9432	0.9834	1.1161	0.9392	0.9550	0.9662
h = 4	2.1909	1.0621	1.2115	1.3317	1.0360	1.0255	1.0504
				PGDP			
h = 1	0.9086	1.0228	0.7594	1.4095	0.9260	0.9080	0.9126
h = 2	1.0238	0.9932	0.9148	1.2537	0.8473	0.8002	0.8033
h = 3	1.0053	1.0035	0.9357	1.4941	0.9218	0.8418	0.8411
h = 4	1.0485	1.3620	1.1592	1.3751	0.9521	0.8752	0.8696
				RCBI			
h = 1	28.1468	0.9540	0.4302	0.8633	1.0957	1.0931	1.0902
h = 2	36.6898	0.9593	0.9213	0.9670	0.9339	0.9335	0.9263
h = 3	39.4043	0.8966	1.0433	0.9225	0.9385	0.9266	0.9144
h = 4	39.8177	0.9156	0.9721	0.9972	0.9673	0.9487	0.9406
				RCONSUM			
h = 1	1.9999	1.0293	0.9558	1.1211	0.9549	0.9695	1.0161
h = 2	1.9889	1.2232	1.0656	1.1779	0.9768	0.9877	1.0308
h = 3	2.0115	1.0923	0.9837	1.1271	1.0436	1.0425	1.0840
h = 4	2.0667	1.0000	1.1801	1.1353	0.9773	0.9970	1.0324
				RFEDGOV			
h = 1	7.6616	1.0252	1.2773	1.3405	0.9222	0.9252	0.9304
h = 2	7.6475	1.1370	1.2568	1.2912	0.9363	0.9415	0.9457
h = 3	7.6041	1.0647	1.1681	1.1850	0.9432	0.9488	0.9528
h = 4	7.6318	1.1248	1.0534	1.2538	0.9556	0.9625	0.9646

Table 2. Forecast performance of individual models and survey information (Cont.)

Horizon	RMSFE	RMSFE ratios with respect to AR					
		Time Series Models			Survey Forecasts		
	AR	AR_FAC	LSTAR	NNET	EW	IMSE	ISIC
				RGDP			
h = 1	2.0670	0.9352	0.6811	0.7594	0.8715	0.8753	0.8912
h = 2	2.1139	0.9507	0.9241	1.0509	0.8968	0.8988	0.9093
h = 3	2.1729	0.9289	0.9924	0.9729	0.9208	0.9325	0.9253
h = 4	2.1054	1.0655	1.0153	1.0723	0.9577	0.9656	0.9636
				RRESINV			
h = 1	9.9230	0.9610	0.7109	0.8241	1.0027	1.0045	1.0078
h = 2	11.3842	1.1000	1.0877	0.9914	0.9312	0.9305	0.9309
h = 3	11.7252	1.1484	0.9988	1.0611	0.9206	0.9190	0.9352
h = 4	12.6665	0.9942	0.8957	0.9313	0.8655	0.8677	0.8699
				RSLGOV			
h = 1	2.7149	1.0282	0.9012	0.8349	0.8473	0.8610	0.8707
h = 2	2.6699	1.0127	1.0455	1.1501	0.8757	0.8742	0.8842
h = 3	2.6808	1.0094	1.0232	0.9798	0.8876	0.8927	0.8947
h = 4	2.6475	1.0000	1.0508	1.0174	0.9315	0.9381	0.9384
				TBILL			
h = 1	0.3504	1.0445	0.6069	5.0496	1.1575	1.3457	1.6660
h = 2	0.7697	1.1151	0.9426	2.1520	0.9629	0.9653	1.0713
h = 3	1.1940	1.1243	1.0968	2.0890	0.9102	0.8725	0.8884
h = 4	1.6074	1.0105	1.0824	1.7213	0.8760	0.8007	0.7985
				UNEMP			
h = 1	0.2159	0.8624	1.2367	4.9265	1.1009	1.0412	1.9799
h = 2	0.3771	0.7867	1.0102	2.2240	0.8901	0.8352	1.5993
h = 3	0.5835	0.7581	1.0636	1.8224	0.7877	0.7629	1.3319
h = 4	0.7487	0.8431	1.1129	1.5891	0.7666	0.8509	1.2629

Notes: The root mean squared forecast error (RMSFE) is calculated with 68 out-of-sample forecasts, except for the T-bill rate which only uses 66 forecasts. AR forecasts are based on autoregressive models with up to eight lags and lag length selected by the SIC through an exhaustive search. AR_FAC is the AR model augmented with a maximum of five factors extracted from 110 underlying series. LSTAR is the Logistic Smooth Transition Autoregressive model. NNET is a Neural Network with one hidden layer. EW refers to the consensus of the Survey of Professional Forecasters (SPF) giving equal weights to each forecaster. IMSE and ISIC refer to the combinations of the SPF data with weights estimated by the inverse of the individual forecasters' RMSFE or weights proportional to the SIC of each forecaster with at least 10 continuous forecasts.

Table 3. Forecast performance of combinations of AR, factor and survey forecasts

Horizon	RMSFE ratios with respect to AR		
	Regression		Equal Weights
	AR_EW	AR_FAC_EW	AR_FAC_EW
	CPI		
h = 1	0.9745	1.0166	0.9567
h = 2	0.9295	0.9421	0.9387
h = 3	0.9671	1.0080	0.9595
h = 4	0.9845	1.0408	0.9612
	CPROF		
h = 1	0.9816	1.0315	0.9384
h = 2	1.0327	1.1044	0.9441
h = 3	1.0328	1.0594	0.9278
h = 4	1.0520	1.1093	0.9382
	HOUSING		
h = 1	1.0331	1.1000	0.9370
h = 2	1.0471	1.0314	0.9452
h = 3	1.0341	1.0490	0.9417
h = 4	1.1200	1.2472	0.9591
	INDPROD		
h = 1	1.0449	0.8666	0.8700
h = 2	1.0193	1.0486	0.9224
h = 3	1.0121	1.0574	0.9379
h = 4	1.0066	1.0269	0.9657
	NGDP		
h = 1	1.0405	1.0714	0.9741
h = 2	1.0130	1.1150	0.9757
h = 3	1.0269	1.0122	0.9142
h = 4	1.0223	1.2418	0.9892
	PGDP		
h = 1	0.8275	0.8298	0.8731
h = 2	0.9040	0.9619	0.9281
h = 3	0.9450	0.9507	0.9576
h = 4	0.9787	1.5257	1.0745
	RCBI		
h = 1	1.0113	0.9873	0.9447
h = 2	1.0565	0.9592	0.9172
h = 3	1.0318	0.9204	0.9163
h = 4	1.0276	0.9156	0.9390
	RCONSUM		
h = 1	0.9521	1.0611	0.9209
h = 2	1.0000	1.2563	1.0130
h = 3	1.0000	1.2267	1.0036
h = 4	1.0000	1.0000	0.9715

Table 3. Forecast performance of combinations of AR, factor and survey forecasts (Cont.)

Horizon	RMSFE ratios with respect to AR		
	Regression		Equal Weights
	AR_EW	AR_FAC_EW	AR_FAC_EW
	RFEDGOV		
h = 1	0.8686	0.9268	0.9645
h = 2	0.9883	0.9967	0.9733
h = 3	0.9669	0.9456	0.9658
h = 4	0.9457	0.9457	0.9686
	RGDP		
h = 1	0.9138	0.9304	0.8801
h = 2	0.9986	0.9607	0.8910
h = 3	1.0268	0.9185	0.9075
h = 4	0.9999	1.0655	0.9817
	RRESINV		
h = 1	0.9896	0.9468	0.9322
h = 2	0.9827	1.0783	0.9819
h = 3	1.0992	1.2111	0.9780
h = 4	1.0000	0.9961	0.9167
	RSLGOV		
h = 1	0.9462	1.0118	0.9192
h = 2	1.0000	1.0127	0.9297
h = 3	1.0000	1.0092	0.9464
h = 4	1.0105	1.0105	0.9615
	TBILL		
h = 1	1.0000	1.0445	1.0220
h = 2	1.0255	1.1238	0.9841
h = 3	1.0183	1.1634	0.9599
h = 4	1.0044	1.0196	0.8883
	UNEMP		
h = 1	1.0237	0.8624	0.8985
h = 2	1.0283	0.8161	0.8253
h = 3	1.0005	0.7920	0.7922
h = 4	1.0087	0.8543	0.8143

Notes: RMSFE is calculated with 68 out-of-sample forecasts, except for the TBILL rate which uses only 66 forecasts. AR_EW refers to the AR model augmented with the equally-weighted survey (SPF) forecasts. AR_FAC_EW refers to the AR model augmented by a maximum number of five common factors as well as the equally-weighted survey (SPF) forecasts.

Table 4. Forecast performance of combinations of AR, nonlinear and factor models

Horizon	RMSFE ratios with respect to AR					
	Regression				Equal Weights	
	AR_LSTAR	AR_NNET	AR_FAC_ LSTAR	AR_FAC_ NNET	AR_FAC_ LSTAR	AR_FAC_ NNET
	CPI					
h = 1	0.4000	0.8869	0.4072	1.0149	0.8355	0.9203
h = 2	0.9994	1.0000	0.9892	0.9880	0.9539	1.0046
h = 3	1.0000	1.0000	1.0396	1.0194	1.0042	1.0480
h = 4	1.0059	1.0000	1.0582	1.2052	0.9870	1.0386
	CPROF					
h = 1	1.0000	1.0000	0.9951	0.9951	0.9636	0.9646
h = 2	1.0000	1.0000	1.0909	1.0909	0.9996	0.9997
h = 3	1.0000	1.0000	1.0016	1.0016	0.9774	0.9776
h = 4	1.0147	1.0000	1.0345	1.0224	0.9768	0.9524
	HOUSING					
h = 1	1.0000	0.9340	1.0330	0.9983	0.9366	0.9366
h = 2	0.9946	1.0061	0.9964	0.9848	0.9781	0.9789
h = 3	1.0000	1.0000	1.0490	1.0490	0.9609	0.9689
h = 4	0.9981	0.9920	1.0641	1.1166	0.9782	0.9818
	INDPROD					
h = 1	0.5642	0.8945	0.4813	0.7868	0.6616	0.7962
h = 2	1.0000	1.0025	1.0205	1.0188	0.9415	0.9464
h = 3	1.0000	1.0005	1.0124	1.0124	0.9441	0.9378
h = 4	1.0000	1.0000	1.0200	1.0200	0.9978	0.9892
	NGDP					
h = 1	0.9158	1.0000	1.0220	1.1249	0.9366	0.9754
h = 2	1.0476	1.1034	1.1301	1.1454	1.0538	1.0437
h = 3	1.0000	1.0000	0.9432	0.9432	0.9284	0.9645
h = 4	1.0000	1.0170	1.0627	1.1252	1.0487	1.0713
	PGDP					
h = 1	0.8925	0.9086	0.9104	0.9293	0.8195	0.9613
h = 2	0.9717	1.0238	1.0305	1.0169	0.9405	1.0197
h = 3	0.9760	1.0053	0.9806	1.0259	0.9427	1.0846
h = 4	1.0485	1.0745	1.4506	1.4949	1.0863	1.1750
	RCBI					
h = 1	0.5329	0.8484	0.5449	0.8096	0.7280	0.8622
h = 2	1.0000	1.0000	0.9593	0.9844	0.9219	0.9307
h = 3	1.0052	1.0111	0.8966	0.9242	0.9465	0.9102
h = 4	1.0000	1.0000	0.9156	0.9156	0.9412	0.9512
	RCONSUM					
h = 1	1.0000	1.0000	1.0293	1.0293	0.9215	0.9682
h = 2	1.0000	1.0000	1.2584	1.2507	1.0125	1.0348
h = 3	1.0000	1.0000	1.0923	1.0923	0.9620	0.9707
h = 4	0.9982	1.0000	0.9982	1.0000	1.0187	0.9815

Table 4. Forecast performance of combinations of AR, nonlinear and factor models (Cont.)

Horizon	RMSFE ratios with respect to AR					
	Regression				Equal Weights	
	AR_LSTAR	AR_NNET	AR_FAC_ LSTAR	AR_FAC_ NNET	AR_FAC_ LSTAR	AR_FAC_ NNET
	RFEDGOV					
h = 1	1.1174	1.0000	1.1943	1.0252	1.0591	1.0772
h = 2	1.0000	1.0000	1.3256	1.1370	1.0521	1.0609
h = 3	1.0000	1.0000	1.0647	1.0647	1.0030	1.0148
h = 4	1.0000	1.0000	1.1248	1.1248	0.9944	1.0203
	RGDP					
h = 1	0.9640	0.8202	0.9427	0.8133	0.8281	0.8016
h = 2	0.9731	1.0000	0.8938	0.9324	0.9132	0.9307
h = 3	0.9979	1.0000	0.9401	0.9408	0.9271	0.9211
h = 4	1.0552	1.0064	1.0013	1.1346	0.9911	1.0192
	RRESINV					
h = 1	0.7517	0.8026	0.7522	0.8186	0.7979	0.7742
h = 2	1.0050	1.0000	1.1908	1.0466	1.0351	0.9999
h = 3	0.9979	1.0000	1.1776	1.1398	1.0245	1.0067
h = 4	1.0006	1.0072	1.0013	0.9935	0.9529	0.9406
	RSLGOV					
h = 1	1.0000	0.9553	1.0275	1.0719	0.9286	0.8907
h = 2	1.0000	0.9951	1.0142	1.0094	0.9706	1.0021
h = 3	1.0000	1.0072	1.0094	1.0157	0.9529	0.9591
h = 4	1.0000	1.0000	1.0000	1.0000	0.9828	0.9502
	TBILL					
h = 1	0.1887	1.1402	0.1884	1.1194	0.5246	1.9548
h = 2	0.9390	1.0223	0.9034	1.1241	0.9379	1.2920
h = 3	1.0000	1.0000	1.1181	1.1243	1.0100	1.2633
h = 4	1.0454	1.0027	1.0165	1.0271	0.9476	1.1442
	UNEMP					
h = 1	0.4223	1.0230	0.4324	0.8681	0.4638	1.7827
h = 2	1.0591	1.0000	0.7867	0.7867	0.8151	1.0946
h = 3	1.0229	1.0000	0.7428	0.7942	0.8428	0.9711
h = 4	1.0000	1.0068	0.7827	0.8385	0.8667	0.9409

Notes: RMSFE-values are calculated with 68 out-of-sample forecasts, except for the T-bill rate which only uses 66 forecasts. AR_LSTAR and AR_NNET refer to the AR model augmented with forecasts generated by the Logistic Smooth Transition Autoregressive models and the Neural Network models, respectively. AR_FAC_LSTAR and AR_FAC_NNET refer to the AR_FAC model augmented by the LSTAR and NNET forecasts, respectively.

Table 5. Forecast performance of combinations of all models

Horizon	Regression AR_FAC_EW_LSTAR_NNET	Equal Weights AR_FAC_EW_LSTAR_NNET	Regression AR_FAC_EW_LSTAR_NNET	Equal Weights AR_FAC_EW_LSTAR_NNET
	RMSFE ratios with respect to AR			
	CPI		RCONSUM	
h = 1	0.4229	0.8296	1.0611	0.9350
h = 2	0.9421	0.9420	1.2835	0.9960
h = 3	0.9871	0.9934	1.2795	0.9561
h = 4	1.2098	0.9768	0.9982	0.9849
	CPROF		RFEDGOV	
h = 1	1.0315	0.9341	0.9842	1.0540
h = 2	1.1035	0.9477	1.1972	1.0431
h = 3	1.0594	0.9380	0.9456	0.9924
h = 4	1.1465	0.9221	1.1267	0.9828
	HOUSING		RGDP	
h = 1	1.1149	0.8903	0.8985	0.7654
h = 2	0.9988	0.9600	0.8930	0.8952
h = 3	1.0490	0.9190	0.9418	0.9160
h = 4	1.3654	0.9425	1.2174	0.9825
	INDPROD		RRESINV	
h = 1	0.4875	0.6585	0.6751	0.7488
h = 2	1.0262	0.9202	1.1208	0.9736
h = 3	1.0565	0.9187	1.2364	0.9663
h = 4	1.0269	0.9817	1.0056	0.8997
	NGDP		RSLGOV	
h = 1	1.0197	0.9128	1.0680	0.8589
h = 2	1.1829	1.0148	1.0113	0.9608
h = 3	0.9991	0.9358	1.0094	0.9240
h = 4	1.2962	1.0616	1.0820	0.9462
	PGDP		TBILL	
h = 1	0.9146	0.9401	0.1876	1.2180
h = 2	0.9396	0.9222	0.9134	1.0877
h = 3	0.9593	0.9713	1.1640	1.1035
h = 4	1.4551	1.0179	1.0400	1.0167
	RCBI		UNEMP	
h = 1	0.5541	0.7745	0.4356	1.1121
h = 2	0.9843	0.9057	0.8161	0.9032
h = 3	0.9145	0.9230	0.8158	0.8957
h = 4	0.9156	0.9416	0.8228	0.8937

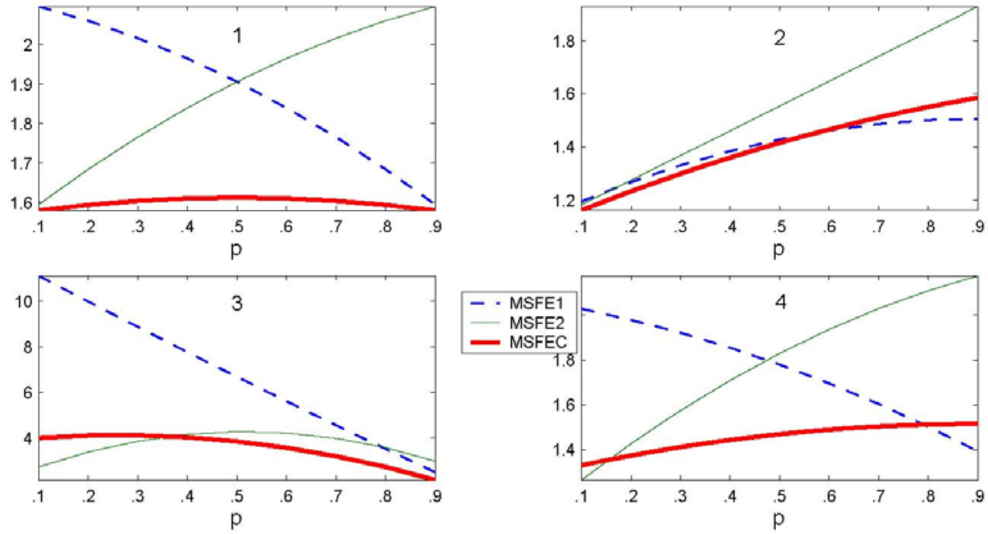
Notes: RMSFE-values are calculated with 68 out-of-sample forecasts, except for the T-bill rate which only uses 66 forecasts. AR_FAC_EW_LSTAR_NNET combines the AR, AR_FAC, EW, LSTAR and NNET forecasts.

Table 6.Out-of-sample MSFE-relative to that of the previous best model using an expanding window – across variables

	Data	Simulation			
		No Breaks	Breaks in the Factor Loadings	Breaks in the Factor Dynamics	Breaks in the Covariance Matrix of Factors
		Panel A: All Variables			
Mean	0.844	0.970	0.935	0.953	0.972
TK	0.837	0.958	0.931	0.951	0.960
PB	1.000	1.000	1.000	1.000	1.000
IMSFE	0.853	0.970	0.935	0.953	0.972
		Panel B: Returns, Interest Rates and Spreads			
Mean	0.777	0.967	0.928	0.954	0.967
TK	0.783	0.957	0.926	0.952	0.958
PB	1.000	1.000	1.000	1.000	1.000
IMSFE	0.777	0.966	0.929	0.954	0.968
		Panel C: Measures of Economic Activity			
Mean	0.990	0.983	0.947	0.949	0.990
TK	0.967	0.960	0.940	0.946	0.966
PB	1.000	1.000	1.000	1.000	1.000
IMSFE	1.004	0.981	0.947	0.949	0.988
		Panel D: Prices and Wages			
Mean	0.814	0.965	0.937	0.953	0.969
TK	0.809	0.955	0.934	0.950	0.959
PB	1.000	1.000	1.000	1.000	1.000
IMSFE	0.823	0.964	0.937	0.952	0.968
		Panel E: Monetary Aggregates			
Mean	0.918	0.975	0.937	0.956	0.975
TK	0.894	0.961	0.932	0.952	0.961
PB	1.000	1.000	1.000	1.000	1.000
IMSFE	0.939	0.975	0.937	0.956	0.975

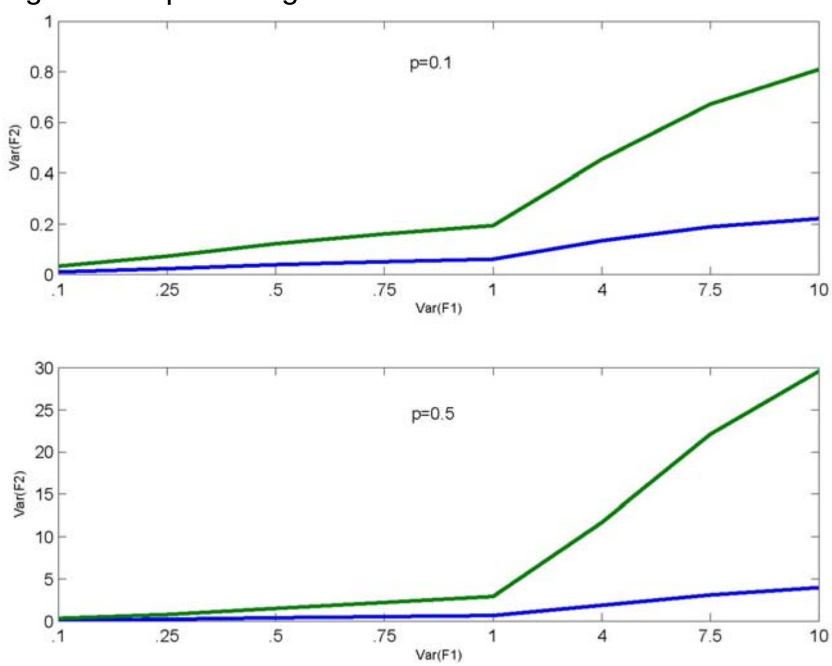
Notes: Mean refers to the mean forecast, TK to the combination using triangular kernel weights (i.e., inverse of historical MSFE rank), PB stands for the previous best forecasts, and IMSFE is a combination with weights that are inversely proportional to the MSFE.

Figure 1. MSFE from model 1, model 2 and equally combined models as a function of p, for different setups



Notes: Panel 1: $\sigma^2_{F2}/\sigma^2_{F1}=1$, $\sigma^2_{\varepsilon 2}/\sigma^2_{\varepsilon 0}=1$, $\sigma^2_{\varepsilon 1}/\sigma^2_{\varepsilon 0}=1$; Panel 2: $\sigma^2_{F2}/\sigma^2_{F1}=0.1$, $\sigma^2_{\varepsilon 2}/\sigma^2_{\varepsilon 0}=1$, $\sigma^2_{\varepsilon 1}/\sigma^2_{\varepsilon 0}=1$; Panel 3: $\sigma^2_{F2}/\sigma^2_{F1}=10$, $\sigma^2_{\varepsilon 2}/\sigma^2_{\varepsilon 0}=1$, $\sigma^2_{\varepsilon 1}/\sigma^2_{\varepsilon 0}=1$; Panel 4: $\sigma^2_{F2}/\sigma^2_{F1}=1$, $\sigma^2_{\varepsilon 2}/\sigma^2_{\varepsilon 0}=0.1$, $\sigma^2_{\varepsilon 1}/\sigma^2_{\varepsilon 0}=1$.

Figure 2. Equal-Weighted Forecast *vs* Best Model Forecast



Notes: The bands describe the regions in which the population MSFE of the equal-weighted forecast is lower than the population MSFE of the best model as a function of the relative variance of the factors, for *p=0.1, 0.5*.