COMBINATION FORECASTS OF OUTPUT GROWTH IN A SEVEN-COUNTRY DATA SET

January, 2003

James H. Stock

Department of Economics, Harvard University
and the National Bureau of Economic Research

and

Mark W. Watson*

Woodrow Wilson School and Department of Economics, Princeton University
and the National Bureau of Economic Research

*We thank Fillipo Altissimo, Frank Diebold, Marcellino Massimilliano, and participants at the Second Workshop on Forecasting Techniques at the European Central Bank for helpful comments and discussions. An earlier version of this paper was circulated under the title, "Combination Forecasts of Output Growth and the 2001 U.S. Recession." This research was funded in part by NSF grant SBR-0214131.

COMBINATION FORECASTS OF OUTPUT GROWTH IN A SEVEN-COUNTRY DATA SET

January, 2003

ABSTRACT

This paper uses forecast combination methods to forecast of output growth in a seven-country quarterly economic data set covering 1959 – 1999, with up to 73 predictors per country. Although the forecasts based on individual predictors are unstable over time and across countries, the combination forecasts often improve upon an autoregressive benchmark. Despite the instability of the constituent forecasts, the most successful combination forecasts are those, like the mean, with the least sensitivity to the recent performance of the individual forecasts. This finding, while consistent with other evidence on the success of simple combination forecasts, is difficult to explain using the theory of combination forecasting in a stationary environment.

Key words: macroeconomic forecasting, high-dimensional forecasting, time-varying parameters

JEL Numbers: C32, E37, E47

1. Introduction

Historically, time series forecasting of economic variables has focused on low-dimensional forecasting models such as autoregressions, low-dimensional vector autoregressions, and single-equation regressions using leading indicators as predictors. These low-dimensional models potentially omit information contained in the thousands of variables available to real-time economic forecasters. To forecast using many predictors, one needs to impose sufficient restrictions that the number of estimated parameters is kept small. One way to impose such restrictions on high-dimensional systems is to suppose that the variables have a dynamic factor structure, and recent research (e.g. Stock and Watson (1999a, 2002), Forni, Hallin, Lippi and Reichlin (2000, 2001)) suggests that there are potential gains from forecasting using high-dimensional dynamic factor models. There are, however, other ways to impose structure on high-dimensional forecasting models, and one such way is to apply the methods of the forecast combining literature.

This paper undertakes an empirical study of high-dimensional forecasting of output growth using forecast combining methods. For introductions to forecast combination methods and surveys of the large literature, see Diebold and Lopez (1996), Newbold and Harvey (2002), and Hendry and Clements (2002). Clemen (1989) provides a comprehensive survey of the literature through the late 1980s, and Makridakis and Hibon (2000) report recent results on combination forecasts. Here, we apply forecast combining methods by first computing a panel of forecasts of output growth, where each constituent forecast is constructed by a (recursively estimated) forecasting regression involving only lagged output and a single predictor, then aggregating this panel of forecasts using forecast combining methods. The data set, which is described in Section 2, covers seven OECD countries from 1961 to 1999 and, for each country, contains up to 73 forecasts based on individual predictors. In previous work with this data set (Stock and Watson (2001)), we found that forecasts based on individual predictors are unstable and that the performance of individual predictors depends on the current economic shocks, as well as on institutional and policy particulars. Surprisingly, however, a preliminary investigation of simple combination forecasts – the median and the trimmed mean of a large panel of forecasts – were stable and reliably outperformed a univariate autoregressive benchmark forecast. Here, we extend that analysis to consider more sophisticated combination forecasts. As discussed in Section 3, the theory of combination forecasting suggests that methods that weight better-performing forecasts more heavily will perform better than simple combination forecasts, and that further gains might be obtained by allowing time variation in the weights or by discounting observations in the distant past.

In Section 4, we use the seven-country data set to explore the empirical performance of various sophisticated forecast combination methods. We find that most of the combination forecasts have lower MSFEs than the benchmark autoregression. The combination methods that have the lowest MSFEs are, intriguingly, the simplest, either with equal weights (the mean) or with weights that are very nearly equal and constant over time. The simple combination forecasts are stable over time and across countries – much more stable than the individual forecasts constituting the panel – and generally outperform forecasts produced using dynamic factor methods. We elaborate on these conclusions in Section 5.

2. The Seven-Country Data Set and Individual Forecasts

This section briefly summarizes the seven-country data set and the panel of forecasts constructed using the individual predictors in that data set.

2.1 The Data

The seven-country data set is the same as is used in Stock and Watson (2001). The data consist of up to 43 time series for each of seven developed economies (Canada, France, Germany, Italy, Japan, the U.K., and the U.S.) over 1959 – 1999 (some series are available only for a shorter period). The 43 series consist of various asset prices (including returns, interest rates, and spreads); selected measures of real economic activity; wages and prices; and measures of the money stock. The list of series is given in Table 1. All the analysis in this paper is done at the quarterly level.

The data were subject to five possible transformations, done in the following order. First, in a few cases the series contained a large outlier, such as spikes associated with strikes, and these outliers were replaced by interpolated values. Second, series that showed significant seasonal variation were seasonally adjusted using a linear approximation to X11. Third, when the data were available on a monthly basis, the data were aggregated to quarterly observations. Fourth, in some cases the data were transformed by taking logarithms. Fifth, the highly persistent or trending variables were differenced, second differenced, or computed as a "gap," that is, a deviation from a stochastic trend. The gaps here were estimated using a one-sided Hodrick-Prescott (1981) filter, which maintains the temporal ordering of the series. For additional details, see Stock and Watson (2001).

In many cases we used more than one version (transformation) of a given series, for example, interest rates were analyzed both in levels and in first differences. The versions of the series analyzed are listed in the final column of Table 1. Counting all the constructed variables (like spreads) and all the variants of the same variable that differ only in the transformation, the maximum number of predictors considered is 73, which is the maximum number of individual-predictor forecasts of output growth in the panel of forecasts for a given country.

2.2 Individual Forecasts

The forecasts based on individual predictors are computed using *h*-step ahead projections. Specifically, let $Y_t = \Delta \ln Q_t$, where Q_t is the level of output (either the level of real GDP or the Index of Industrial Production), and let X_t be a candidate predictor (e.g. the term spread). Let Y_{t+h}^h denote output growth over the next *h* quarters, expressed at an annual rate, that is, let $Y_{t+h}^h = (400/h)\ln(Q_{t+h}/Q_t)$. The forecasts of Y_{t+h}^h are made using the *h*-step ahead regression model,

$$Y_{t+h}^h = \beta_0 + \beta_1(\mathrm{L})X_t + \beta_2(\mathrm{L})Y_t + u_{t+h}^h, \tag{1}$$

where u^h_{t+1} is an error term and β_1(L) and β_2(L) are lag polynomials. Forecasts are computed for h = 2, 4, and 8-quarter horizons.

Model selection and coefficient estimation are done using pseudo out-of-sample methods. Specifically, the coefficients in (1) are estimated recursively using OLS, so that the forecast of Y^h_{t+h} made at date t with estimated coefficients, $\hat{Y}^h_{t+h|t}$, is entirely a function of data for dates 1,…, t. Lag lengths are determined recursively using the AIC with between one and four lags of X_t (we refer to X_t in (1) as the first lag because it is lagged, relative to Y^h_{t+h}) and between zero and four lags of Y_t.

Two univariate benchmark forecasts are used. The first is a multistep autoregressive (AR) forecast, in which (1) is estimated recursively with no X_t predictor and the lag length is chosen recursively by AIC (between zero and four). The second is a recursive random walk forecast, in which $\hat{Y}^h_{t+h|t} = h\hat{\mu}_t$, where $\hat{\mu}_t$ is the sample average of ΔY_s, s = 1,…, t.

All the individual-predictor forecasts considered in this paper are all linear projections. There is evidence that combination forecasts that pool linear and nonlinear forecasts can outperform combination forecasts based solely on linear forecasts (e.g. Stock and Watson (1999b), Blake and Kapetanios (1999)). Incorporating such nonlinear forecasts might improve upon the results reported here, but doing so would go beyond the linear framework of the dynamic factor model forecasts with which we wish to compare the combination forecasts.

3. Combination Forecasts and Forecast Evaluation Methods

Quite a few methods for pooling forecasts have been developed in the large literature on forecast combination. This section describes the combining methods studied in this paper and explains how they will be evaluated by comparing their pseudo out-of-sample forecasts.

3.1 Combination Forecast Methods

Five types of combination forecasts are considered in this paper: simple combination forecasts; discounted MSFE forecasts; shrinkage forecasts; factor model forecasts; and time-varying-parameter (TVP) combination forecasts. These methods differ in the way they use historical information to compute the combination forecast and in the extent to which the weight given an individual forecast is allowed to change over time. These methods, or closely related methods, have appeared previously in the forecast combining literature. Some standard methods for forecast combination, such as Granger-Ramanathan (1984) combining using regression weights, are inappropriate here because of the large number of individual forecasts, relative to the sample size. The methods we use here are variants of linear forecast combinations; although there is evidence that nonlinear combination schemes can produce substantial gains (e.g. Deutsch, Granger and Terasvirta (1994)), the number of forecasts we consider arguably is too large for nonlinear combination methods to be effective.

Notation and estimation periods. Let $\hat{Y}^h_{i,t+h|t}$ denote the i^{th} individual pseudo out-of-sample forecast of Y^h_{t+h}, computed at date t, that is, the i^{th} forecast in the panel of forecasts for a given country. Most of the combination forecasts we consider are weighted averages of the individual forecasts (possibly with time-varying weights) and thus have the form,

$$f_{t+h|t} = \sum_{i=1}^{n} w_{it} \hat{Y}^{h}_{i,t+h|t} \, , \qquad (2)$$

where $f_{t+h|t}$ is the combination forecast, w_{it} is the weight on the i^{th} forecast in period t and n is the number of the forecasts in the panel.

In general, the weights $\{w_{it}\}$ depend on the historical performance of the individual forecast. To evaluate this historical performance, we divide the sample into three periods. The observations prior to date T_0 are only used for estimation of the coefficients in the individual forecasting regression (1). The individual pseudo out-of-sample forecasts are computed starting in period T_0. The recursive MSFE of the i^{th} individual forecast, computed from the start of the forecast period through date t, is

$$MSFE_{it} = \frac{1}{t - T_0 - 1} \sum_{s=T_0}^{t} (Y^{h}_{s+h} - \hat{Y}^{h}_{i,s+h|s})^2 \, . \qquad (3)$$

The pseudo out-of-sample forecasts for the combination forecasts are computed over $t = T_1, \ldots, T_2 - h$. For the empirical work reported in the next section, we used T_0 = 1973:I, T_1 = 1983:I, and T_2 (the end of the sample) is 1999:IV.

Simple combination forecasts. The simple combination forecasts compute the combination forecast without regard to the historical performance of the individual forecasts in the panel. Three simple combination forecasts are used: the mean of the panel of forecasts (so $w_{it} = 1/n$ in (2)); the median; and the trimmed mean. The trimmed mean was computed with 5% symmetric trimming, subject to trimming at least one forecast.

Discounted MSFE forecasts. The discounted MSFE forecasts compute the combination forecast as a weighted average of the individual forecasts, where the weights depend inversely on the historical performance of each individual forecast (cf. Diebold and Pauly (1987)). Specifically, the discounted MSFE combination h-step ahead forecast has the form (2), where the weights are

$$w_{it} = m_{it}^{-1} / \sum_{j=1}^{n} m_{jt}^{-1}, \text{ where } m_{it} = \sum_{s=T_0}^{t-h} \delta^{t-h-s} (Y_{s+h}^h - \hat{Y}_{i,s+h|s}^h)^2, \tag{4}$$

where δ is the discount factor.

The discounted MSFE forecasts are computed for three values of δ, δ = 1.0, 0.95, and 0.9. The case $\delta = 1$ (no discounting) corresponds to the Bates and Granger (1969) optimal weighting scheme when the individual forecasts are uncorrelated.

A related combination forecast is the "most recently best", which as implemented here places all weight on the individual forecast that has the lowest average squared forecast errors over the previous four periods.

Shrinkage forecasts. The shrinkage forecasts compute the weights as an average of the recursive OLS estimator of the weights (the Granger-Ramanathan (1984) estimator, imposing an intercept of zero) and equal weighting. That is, the shrinkage forecasts have the form (2), where

$$w_{it} = \lambda \hat{\beta}_{it} + (1 - \lambda)(1/n), \tag{5}$$

where $\hat{\beta}_{it}$ is the i^{th} estimated coefficient from a recursive OLS regression of Y^h_{s+h} on $\hat{Y}^h_{1,s+h|s},\ldots, \hat{Y}^h_{n,s+h|s}$ for $s = T_0,\ldots, t - h$ (no intercept) and where $\lambda = \max\{0, 1 - \kappa[n/(t - h - T_0 - n)]\}$, where κ is a constant that controls the amount of shrinkage towards equal weighting. The shrinkage forecasts were evaluated for $\kappa = 0.25, 0.5, 1$, with larger values corresponding to more shrinkage towards equal weighting (smaller λ).

The shrinkage forecast based on (5) can be interpreted as a Bayes estimator (see Diebold and Pauly (1990)). In that context, the weight κ could be estimated using empirical Bayes methods, however we do not pursue that here because of difficulties that arise when the number of individual forecasts n is large relative to $t - T_0$.

Factor model forecast combination. Factor model forecast combination entails (i) recursively computing a set of estimated common factors of the panel of forecasts, (ii) estimating a regression of $Y^h_{s+h|s}$ onto the first few common factors, and (iii) forming the forecast based on this regression. Reduction of the many forecasts to a few common factors provides a convenient method for allowing some estimation of factor weights, yet reduces the number of weights that must be estimated. This method has previously been used by Figlewski (1983), Figlewski and Urich (1983), and Chan, Stock, and Watson (1999). One reason to think that this method might work well in this application is that, as described in the introduction, recent work on large forecasting models suggest that large macroeconomic data sets are well described by a few common dynamic factors that are useful for forecasting (e.g. Forni, Lippi, Hallin and Reichlin (2000, 2001), Stock and Watson (1999a, 2002)). The forecast combining application here differs from the usual dynamic factor model approach because the individual series first are used to compute a panel of forecasts, then static common factors are estimated from this panel of forecasts.

Specifically, the factor model forecasts are constructed as follows. Let $\hat{F}^h_{1,s},\ldots, \hat{F}^h_{m,s}$ denote the first m factors of $\hat{Y}^h_{1,s+h|s},\ldots, \hat{Y}^h_{n,s+h|s}$ for $s = T_0,\ldots, t$, computed as the first m principal components of the uncentered second moment matrix of the recursive forecasts over $s = T_0,\ldots, t$.[1] The factor model forecasts are computed using the regression,

$$Y^h_{s+h} = \alpha_1 \hat{F}^h_{1,s} + \ldots + \alpha_m \hat{F}^h_{m,s} + v^h_{s+h}, \tag{6}$$

where the regression coefficients $\alpha_1,\ldots, \alpha_m$ are estimated by OLS over the sample $s = T_0,\ldots, t–h$. The combined forecast is computed using the estimated weights, applied to $\hat{F}^h_{1,t},\ldots, \hat{F}^h_{m,t}$.

Two versions of the factor model forecasts were computed, one with m chosen recursively by AIC, the other by BIC, where $1 \le m \le 4$.

Time-varying parameter forecasts. The TVP combination forecast uses the Kalman filter to estimate time-varying coefficients in the combining regression, where

[1] Because the forecasts are in the same units, the second moment matrix was computed without standardizing the individual forecasts, and the sample mean was not subtracted from the component forecasts.

the coefficients are modeled as evolving according to a random walk. This method is used by Sessions and Chatterjee (1989) and by LeSage and Magura (1992). LeSage and Magura (1992) also extend it to mixture models of the errors, but that extension did not improve upon the simpler Kalman filter approach in their empirical application. Our implementation starts with the Granger–Ramanathan (1984) combining regression, modified to impose a zero intercept and extended to have time varying parameters:

$$Y_{s+h}^{h} = \omega_{1t}\hat{Y}_{1,s+h|s}^{h} + \ldots + \omega_{mt}\hat{Y}_{n,s+h|s}^{h} + \varepsilon_{s+h}^{h}, \quad (7)$$

$\omega_{it} = \omega_{it-1} + \eta_{it}$, where η_{it} are serially uncorrelated, uncorrelated with ε_{s+h}^{h}, and uncorrelated across i. The relative variance var(η_{it})/var(ε_{s+h}^{h}) is in principal estimable but with many forecasts its estimator could be quite unreliable, so instead we set the relative variance to var(η_{it})/var(ε_{s+h}^{h}) = ϕ^2/n^2, where ϕ is a chosen parameter. Larger values of ϕ correspond to more time variation. The initial distribution of ω_{it} sets each weight to $1/n$ with zero variance. Three values of ϕ are investigated: ϕ = 0.1, 0.2, and 0.4. We found that performance of the TVP combination forecasts deteriorated for larger values of ϕ than these.

3.2 Pseudo Out-of-Sample Evaluation Methods

The forecasting performance of a candidate combination forecast is evaluated by comparing its out-of-sample mean squared forecast error (MSFE) to the autoregressive benchmark. Specifically, let $\hat{Y}_{i,t+h|t}^{h}$ denote the pseudo out-of-sample forecast of Y_{t+h}^{h}, computed using data through time *t*, based on the i^{th} combination forecast. Let $\hat{Y}^h_{0,t+h|t}$ denote the corresponding benchmark forecast made using the autoregression. Then the relative mean squared forecast error (MSFE) of the candidate combination forecast, relative to the benchmark forecast, is

$$\text{Relative MSFE} = \frac{\sum_{t=T_1}^{T_2-h}(Y^h_{t+h} - \hat{Y}^h_{i,t+h|t})^2}{\sum_{t=T_1}^{T_2-h}(Y^h_{t+h} - \hat{Y}^h_{0,t+h|t})^2}, \quad (8)$$

where T_1 and $T_2–h$ are respectively the first and last dates over which the pseudo out-of-sample forecast is computed.

When the benchmark model 0 is nested within the candidate forecast *i*, the distribution of the relative MSFE, under the null hypothesis that $\beta_1(L) = 0$ in (1) and the other coefficients are constant, is nonstandard and was obtained by Clark and McCracken (2001). In the analysis here, the null and alternative models are not necessarily nested because of the recursive lag length selection, and the null distribution of the relative MSFE is unknown. Moreover, it is not clear how applicable the Clark and McCracken (2001) distribution theory is when the parameter vector is very large, as is the case for the combination forecasts analyzed here. For these reasons, in this paper we report relative MSFEs but not a measure of their statistical significance, leaving the latter to future work.[2]

---

[2] Clark-McCracken (2001) *p*-values are reported by Stock and Watson (2001) for fixed-lag versions (four lags) of the individual-indicator forecasts that constitute the panel of forecasts analyzed here. The 5% critical value for the relative MSFEs typically range

4. Empirical Results

This section examines the empirical performance of the combination forecasts constructed using the seven-country quarterly data set. We begin by summarizing the performance of the individual forecasts that constitute the panel of forecasts.

4.1 Individual and Simple Combination Forecasts

The individual forecasts for the seven-country data set are discussed and analyzed in detail in Stock and Watson (2001). Consistent with the large literature on forecasting output growth using asset prices, some individual asset prices have predictive content for output in some time periods and in some countries. For example, the term spread (the yield on long term government debt minus a short-term interest rate) was a potent predictor of output growth in the U.S. during the 1970s and early 1980s. There is, however, considerable instability in forecasts based on individual predictors, and good performance in one period and country does not ensure good performance in another. Instead, performance of an individual predictor depends on the configuration of shocks hitting the economy, the current policy regime, and other institutional factors. For

from 0.92 to 0.96 (the critical value depends on nuisance parameters and thus was computed on a series-by-series basis). By this measure, many of the individual-indicator forecasts showed a statistically significant improvement over the AR benchmark, at least in some periods and some countries.

example, the term spread ceased to be a good predictor of output in the late 1980s and 1990s in the U.S.[3]

4.2 Comparison of Alternative Combination Methods

We now turn to an application and evaluation of the various combination schemes to the seven-country data set. All the forecast combination methods were applied to a balanced panel of forecasts, where the individual forecasts entering the balanced panel were those forecasts that commenced in 1973:I, which in turn required data on that predictor starting in 1963:I to ensure sufficiently many observations for the initial estimation of (1). The simple, recent best, and discounted MSFE forecast combination methods do not involve a combining regression, so they are readily computed using an unbalanced panel. Accordingly, these combination forecasting methods were also applied to the full panel of forecasts (all available data). Applying the same combination method to the full data set and the balanced panel subset allows us to ascertain the gain from increasing the number of individual forecasts entering the combination forecast.

The results for forecasts of real GDP growth over two, four, and eight quarters are summarized in Tables 2, 3, and 4, respectively, and the results for IP growth over the three horizons are summarized in Tables 5, 6, and 7. These tables have the same format; the entries for a candidate predictor (the row variable) are its MSFE for the forecast period (indicated in the first row), relative to the MSFE of the benchmark AR forecast. If

---

[3] The instability evident in the individual-predictor forecasts is consistent with other evidence of widespread instability in small econometric and time series models used for macroeconomic forecasting, see for example Stock and Watson (1996), Bernanke and Mihov (1998), Clements and Hendry (2001), Cogley and Sargent (2001, 2002), Sims and Zha (2002), and Marcellino (2002).

the candidate predictor has a relative MSFE less than one, then it outperformed the AR benchmark over the forecast period in that country. The final two rows of relative MSFEs are for forecasts from dynamic factor models, discussion of which is postponed to Section 4.4.

Several results emerge from Tables 2 – 7. First, many of the combination forecasts outperform the AR benchmark reliably across countries, across horizons, and across the variable being forecasted.

Second, combination forecasts based on the full panel generally outperform their counterparts based on the balanced panel subset. Evidently the additional series in the full panel contain information useful for forecasting.

Third, although many if the improvements of the combination forecasts are modest, relative to the AR benchmark (relative MSFEs of .9 or .95), in some cases the gains are large (relative MSFEs of.85 or less).

Fourth, the simple combination forecasts show reliably good performance across series and horizons. Among the simple combination forecasts, there seems to be little difference between the mean and the trimmed mean. The median typically has somewhat higher relative MSFE than either the mean or trimmed mean.

Fifth, the shrinkage forecasts are not robust: for some countries and horizons they perform well, but for others they perform quite poorly. The less shrinkage, the less robust is the resulting combination forecast.

Sixth, the principal component (static factor regression) forecasts have quite variable performance, in some cases far outperforming the AR benchmark but in other cases performing much worse. We return to this below.

Seventh, the results for the methods designed to handle time variation are mixed or poor. The TVP forecasts sometimes work well but sometimes work quite poorly. Similarly, the discounted MSE forecasts with the most discounting ($\delta = .9$) are no better than, and sometimes worse than, their counterparts with less or no discounting ($\delta = .95$ or 1).

4.3 Ranking Combination Forecasts by Average Loss

As another way to compare the combination methods, we also computed an average estimated loss for each combination method, where the average is computed across all countries and across the two different measures of output. There are a total of 13 such cases (7 countries, two measures of output, except not real GDP for France, for which the time series is too short). This average loss of a given combination forecast is computed as the weighted average of the MSFEs for the individual countries (where each MSFE is computed over $t = T_1,\ldots, T_2 - h$), where the country weights are the inverse of the full-sample ($t = 1,\ldots, T_2 - h$) variance of Y_{t+h}^h. Equivalently, the average loss of a given combination forecast is the unweighted average MSFE across countries, where each output measure is standardized to have a unit full-sample variance. One interpretation of this average loss is that it estimates the loss a forecaster would expect to have if she knew she would be forecasting output growth in a developed economy, but is not told which economy or which measure of output growth. The forecast that minimizes the population counterpart of this average loss is the forecast that has the lowest expected loss in the forecasting game in which the forecaster first chooses the combination method, then is told the country and output measure that she must forecast.

The various combination forecasts (along with the random walk forecast), ranked by their average loss, is presented in Table 8 for four-quarter ahead output growth (13 cases averaged) and in Table 9 for all three horizons (39 cases averaged). The results are striking. When the loss is averaged over all countries, dependent variables, and horizons, the best combination forecast is the mean of the full panel of forecasts. The other methods that work well are numerically very close to the mean: the discounted MSE forecast with no discounting, the trimmed mean, and the shrinkage forecast with the greatest shrinkage. In contrast, the combination methods that permit the greatest time variation in weights, or that rely the most on historical evidence to estimate the combination weights, exhibit the poorest performance, in some cases by a wide margin.

4.4 Comparison of Combination and Dynamic Factor Model Forecasts

As discussed in the introduction, an alternative way to forecast using many predictors is to impose a dynamic factor structure, then to compute forecasts based on a small number of estimated dynamic factors. Success with this approach has been reported by Forni, Lippi, Hallin, and Reichlin (2000, 2001) and by Stock and Watson (1999a, 2002). This section compares such dynamic factor model forecasts to combination forecasts.

Construction of dynamic factor model forecasts. The series (with their transformations) used to compute the dynamic factor model forecasts are listed, by country, in Table 10. These series and transformations are a subset of those used for the panel of individual-predictor forecasts, where the subset was the series in the balanced panel dataset, further restricted to avoid singularities and near-singularities of the sample covariance matrix. For the dynamic factor model forecasts, variables were also restricted to appear with only a single transformation.

Following Stock and Watson (1999a, 2002), estimates of the dynamic factors were computed recursively as the first four principal components (ordered by the fraction of the variance explained), computed recursively from the recursive sample correlation matrix of these series. The pseudo out-of-sample dynamic factor forecasts of output growth were then computed by regression *h*-period output growth against the first *k* estimated dynamic factors and *m* lags of ΔY_t, that is, as in (6), except that the factors are estimated using the original series (not the individual forecasts) and the regression also includes a constant and one or more lags of ΔY_t. The number of factors and the number of lags of ΔY_t both were was chosen recursively by AIC (between one and four factors and between zero and four lags of ΔY_t). This procedure (applied to different data) is discussed in detail in Stock and Watson (1999a, 2002).

Empirical results. The MSFEs of the dynamic factor model – principal component (dfm-PC) forecasts, relative to the AR benchmark, are reported in the final section of Tables 2 – 7. Because the dfm-PC forecasts are based on the balanced panel subset, the relevant comparison is between the balanced panel subset combination forecasts and the dfm-PC forecasts. Inspection of Tables 2 – 7 reveals that in many cases the mean combination forecast for the balanced panel and the dfm-PC forecasts have similar relative MSFEs, but for some cases these differ, and if so the mean combination forecast usually outperforms the dfm-PC forecast, sometimes by a wide margin.

The improvement of the combination forecasts over the dfm-PC forecasts is evident in the estimated losses and rankings in Tables 8 and 9. For example, either at the

4-quarter horizon (Table 8) or averaged across all horizons (Table 9), the dfm-PC forecast is comparable to the AR benchmark, and has substantially higher loss than the two best balanced-panel combination forecasts, the discounted MSE(1) and the simple mean. In the overall ranking of Table 9, the dfm-PC forecast performs comparably to the PC combination forecast; in this sense, neither of the methods based on factor models perform nearly as well in this comparison as the simple combination forecasts.[4]

4.5 Forecast Stability

So far the analysis has focused on average performance of the combination forecasts over the full forecast period, 1982 – 1999. Given the instability of the individual forecasts making up the panel of forecasts, however, it is of interest to examine the stability of the high-dimensional forecasts. Accordingly, we divide the pseudo out-of-sample forecast period (the period in the first rows of Tables 2 – 7) in half and compute the MSFEs over the two periods of 1982:I – 1990:II and 1990:III – 1999:IV. A stable and potent forecast would have population MSFEs less than the AR benchmark in both periods, whereas an unstable forecast would have a population relative MSFE less

[4] The dfm-PC forecasts reported in Tables 2 – 7 generally provide substantially smaller improvements upon the AR benchmark than was found in for the U.S. in Stock and Watson (2002). In some cases, such as forecasting Canadian GDP at the 2- and 4-quarter horizon, the dynamic factor model forecasts improve upon the AR benchmark by a considerable margin. In other cases, such IP forecasts for Germany and the U.S. at the 8-quarter horizon, the dfm-PC forecasts are worse than the AR benchmark. But for most countries and horizons the dynamic factor model forecasts have relative MSFEs near one. Additional empirical work (not reported in the tables) suggests that one reason for the difference between these results and the more favorable results in Stock and Watson (2002) is the sample period, which included the 1970s. In addition, considerably fewer series are used here than in the other investigations (for most countries, fewer than 20 series), and the asymptotic theory behind the dfm-PC forecasts relies on the number of series being large.

than one in one period but greater than one in the other period. Because of sampling variability, the sample MSFEs will differ from the population MSFEs, but even without a formal distribution theory for these relative MSFEs (for the reasons discussed in Section 3), examination of the relative MSFEs in the two subsamples can shed some light on the stability of the various forecasting methods.

As a basis for comparison, we first present a scatterplot of the logarithm of the relative MSFEs of the forecasts based on the individual indicators, combined over all horizons, countries, and variables being forecasted. In this scatterplot, given in Figure 1, a point represents the pair of (log) relative MSFEs for a specific predictor, country, horizon, and dependent variable (IP or real GDP); Figure 1 contains 2196 such points. If the forecasting relations were stable, then the points would be scattered around the 45^{o} line, with a cluster around the origin for those predictors that have negligible marginal predictive content for output growth, above and beyond that in lags of output growth. But the points are neither scattered around the 45^{o} line nor clustered around the origin; instead, there are many points far into the northwest and southeast quadrants, which indicate relatively good performance in one period and relatively poor performance in the other. Indeed, there is little structure in this scatterplot, which suggests that performance of a randomly selected individual predictor/country/horizon/dependent variable combination in the first period is largely independent of its performance in the second period (for further discussion, see Stock and Watson (2001)).

The comparable scatterplot for the simple mean combination forecasts (using the full panel) is presented in Figure 2; each of its 41 points represents a pair of (log) relative MSFEs for the simple mean forecast for a particular country, horizon, and dependent variable. Evidently, the simple mean combination forecast shows considerable stability – especially when compared with the instability of the constituent forecasts exhibited in Figure 1. Most of the points for the simple mean forecast are in the southwest quadrant, indicating an improvement over the AR benchmark in both periods.

Additional measures of the stability of the various forecasts are summarized in Table 11, which reports the average relative MSFE for different categories of forecasts (averaged across countries, dependent variables, and horizons), for each subperiod, as well as the average absolute difference between the relative MSFEs between the first and second periods. On average, the forecasts based on the individual predictors do worse than the AR benchmark in both periods. Consistent with Figure 1, those forecasts also have, on average, a large absolute change of 0.40 in the relative MSFE between the two periods. In contrast, the simple mean combination forecast, on average, improves upon the AR benchmark, and has an average change of only .08 between the two periods (the average change of the median forecast is even less, .05). One striking result in Table 11 is that the greater the amount of data adaptivity in the combination forecast, the greater is its instability, with the recent best, PC, shrink(.25), and tvp(.4) combination forecasts all having large average changes in relative MSFEs between the first and second periods.

5. Discussion and Conclusions

The empirical analysis in this paper yields four main conclusions. First, some combination forecasts perform well, regularly having pseudo out-of-sample MSFEs less than the AR benchmark; in some cases, the improvements are quite substantial.

Second, the combination forecasts that perform best generally are those that have the least data adaptivity in their weighting schemes. Aggregated across all horizons, countries, and dependent variables, the forecasting method with the lowest squared error loss was the simple mean combination forecast. In contrast, sophisticated combination forecasts that heavily weight recent performance or allow for substantial time variation in the weights typically performed worse than – sometimes much worse than – the simple combination schemes.

Third, the combination forecasts performed well when compared to forecasts constructed using a dynamic factor model framework. This is interesting in light of recently reported good forecasting results for dynamic factor models. One possibility is that the number of series examined here is relatively small, compared with those examined recently using dynamic factor models. In any event, this finding merits further study.

Fourth, the combination forecasts with the least adaptivity also were found to be the most stable when we divided the pseudo out-of-sample forecast period in half. This result is surprising. After all, the reason for introducing discounting and time varying parameter combining regressions is to allow for instability in the constituent forecasts – which there clearly is – yet doing so worsens the performance of the resulting combination forecast.

Because of the substantial instability of the underlying individual forecasts, we consider it implausible that the classical explanation of the virtues of combination forecasts – the pooling of information in a stationary environment – can explain our results. Indeed, the mean of the contemporaneous forecasts had lower average loss than any of the more sophisticated combination forecasts, a finding consistent with other empirical investigations of combination forecasting. This "forecast combination puzzle" – the repeated finding that simple combination forecasts outperform sophisticated combination methods in empirical applications – is, we think, more likely to be understood in the context of a model in which there is widespread instability in the individual forecasts, but the instability is sufficiently idiosyncratic that the combination of these individually unstable forecasts can itself be stable.

References

Bates, J.M. and Granger, C.W.J. (1969), “The Combination of Forecasts,” *Operations Research Quarterly* 20, 451–468.

Bernanke, B. and I. Mihov (1998), “Measuring Monetary Policy,” *Quarterly Journal of Economics,* 869–902.

Blake, A.P. and G. Kapetanios (1999), “Forecast Combination and Leading Indicators: Combining Artificial Neural Network and Autoregressive Forecasts,” manuscript, National Institute of Economic and Social Research.

Chan, L., J.H. Stock, and M.W. Watson (1999), “A Dynamic Factor Model Framework for Forecast Combination,” *Spanish Economic Review* 1, 91–121.

Clark, T.E. and M.W. McCracken (2001), “Tests of Equal Forecast Accuracy and Encompassing for Nested Models,” *Journal of Econometrics* 105, 85–100.

Clemen, R.T. (1989), “Combining Forecasts: A Review and Annotated Bibliography,” *International Journal of Forecasting* 5, 559 – 583.

Clements, M.P. and D.F. Hendry (1999), *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Cogley, T. and T.J. Sargent (2001), “Evolving Post-World War II U.S. Inflation Dynamics,” *NBER Macroeconomics Annual*, 331 – 372.

Cogley, T. and T.J. Sargent (2002), “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.,” manuscript, Stanford University.

Deutsch, M., C.W.J. Granger, and T. Terasvirta (1994), “The Combination of Forecasts Using Changing Weights,” *International Journal of Forecasting* 10, 47–57.

Diebold, F.X. and J.A. Lopez (1996), “Forecast Evaluation and Combination,” in G.S. Maddal and C.R. Rao (eds.), *Handbook of Statistics*. Amsterdam: North-Holland.

Diebold, F.X. and P. Pauly (1987), “Structural Change and the Combination of Forecasts,” *Journal of Forecasting* 6, 21–40.

Diebold, F.X. and P. Pauly (1990), “The Use of Prior Information in Forecast Combination,” *International Journal of Forecasting* 6, 503-508.

Figlewski, S. (1983), “Optimal Price Forecasting Using Survey Data,” *Review of Economics and Statistics* 65, 813–836.

Figlewski, S. and T. Urich (1983), "Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability and Market Efficiency," *The Journal of Finance* 28, 695–710.

Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000), "The Generalized Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.

Forni, M., M. Hallin, M. Lippi and L. Reichlin (2001), "Do Financial Variables Help Forecasting Inflation and Real Activity in the EURO Area?," CEPR Working Paper

Granger, C.W.J. and R. Ramanathan (1984), "Improved Methods of Combining Forecasting," *Journal of Forecasting* 3, 197–204.

Hendry, D.F. and M.P. Clements (2002), "Pooling of Forecasts," *Econometrics Journal*, 5, 1–26.

LeSage, J.P. and M. Magura (1992), "A Mixture-Model Approach to Combining Forecasts," *Journal of Business and Economic Statistics* 3, 445–452.

Makridakis, S. and M. Hibon (2000), "The M3-Competition: Results, Conclusions and Implications," *International Journal of Forecasting* 16, 451–476.

Marcellino, M. (2002), "Instability and Non-Linearity in the EMU," manuscript, IEP–U. Bocconi.

Newbold, P. and D.I. Harvey (2002), "Forecast Combination and Encompassing," Ch. 12 in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*. Oxford: Blackwell Press.

Sessions, D.N. and S. Chatterjee (1989), "The Combining of Forecasts Using Recursive Techniques with Non-stationary Weights," *Journal of Forecasting* 8, 239–251.

Sims, C.A. and T. Zha (2002), "Macroeconomic Switching," manuscript, Princeton University.

Stock, J.H. and M.W. Watson (1999a), "Forecasting Inflation," *Journal of Monetary Economics* 44, 293–335.

Stock, J.H. and M.W. Watson (1999b), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," ch. 1 in R. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, 1-44.

Stock, J.H. and M.W. Watson (2001), "Forecasting Output and Inflation: The Role of Asset Prices," NBER Working Paper no. 8180 (revised January 2003).

Stock, J.H. and M.W. Watson (2002), " Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, Vol. 20 No. 2, 147-162.

Table 1. Series in the Seven-Country Data Set

Series Label	Sampling Frequency	Description	Transformation(s) for panel of individual forecasts
		Asset Prices	
rovnght	M	Interest Rate: overnight	level, Δ
rtbill	M	Interest Rate: short term Gov. Bills	level, Δ
rbnds	M	Interest Rate: short term Gov. Bonds	level, Δ
rbndm	M	Interest Rate: medium term Gov. Bonds	level, Δ
rbndl	M	Interest Rate: long term Gov. Bonds	level, Δ
rrovnght	Q	Real overnight rate: rovnght – CPI Inflation	level, Δ
rrtbill	Q	Real short term bill rate: rtbill – CPI Inflation	level, Δ
rrbnds	Q	Real short term bond rate: rtbnds – CPI Inflation	level, Δ
rrbndm	Q	Real med. term bond rate: rtbndm – CPI Inflation	level, Δ
rrbndl	Q	Real long term bond rate: rtbndl – CPI Inflation	level, Δ
rspread	M	Term Spread: rbndl – rovnght	level
exrate	M	Nominal Exchange Rate	Δln
rexrate	M	Real Exchange Rate (exrate × relative CPIs)	Δln
stockp	M	Stock Price Index	Δln
rstockp	M	Real Stock Price Index: stockp	Δln
divpr	Q	Dividend Price Index	ln
house	Q	House Price Index	Δln
rhouse	Q	Real House Price Index	ln, Δln
gold	M	Gold Prices	Δln, Δ^2ln
rgold	M	Real Gold Prices	ln, Δln
silver	M	Silver Prices	Δln, Δ^2ln
rsilver	M	Real Silver Prices	ln, Δln
		Activity	
rgdp	M	Real GDP	Δln, gap
ip	M	Index of Industrial Production	Δln, gap
capu	M&Q	Index of Capacity Utilization	level
emp	M&Q	Employment	Δln, gap
unemp	M&Q	Unemployment Rate	Δ, gap
pgdp	Q	GDP Deflator	Δln, Δ^2ln
cpi	M	Consumer Price Index	Δln, Δ^2ln
ppi	M	Producer Price Index	Δln, Δ^2ln
		Wages, Goods and Commodity Prices	
earn	M	Wages	Δln, Δ^2ln
commod	M	Commodity Price Index	Δln, Δ^2ln
oil	M	Oil prices	Δln, Δ^2ln
roil	M	Real Oil Prices	ln, Δln
rcommo d	M	Real Commodity Price Index	ln, Δln
		Money	
m0	M	Money: M0 or Monetary Base	Δln, Δ^2ln
m1	M	Money: M1	Δln, Δ^2ln

m2	M	Money: M2	Δln, Δ^2ln
m3	M	Money: M3	Δln, Δ^2ln
rm0	M	Real Money: M0	Δln
rm1	M	Real Money: M1	Δln
rm2	M	Real Money: M2	Δln
rm3	M	Real Money: M3	Δln

Notes: M indicates that the original data are monthly, Q indicates that they are quarterly, M&Q indicates that monthly data were available for some countries but quarterly data were available for others. All forecasts and regressions use quarterly data, which were aggregated from monthly data by averaging (for CPI and IP) or by using the last monthly value (all other series). The transformations listed in the fourth column are those used to construct the panel of forecasts based in individual predictors; if there are two transformations, that series appears twice in the panel, once with each transformation.

Table 2. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 2-quarter Growth of Real GDP (h = 2)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	81:III – 98:IV	. .	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV
Univariate forecasts							
AR RMSFE	0.016	.	0.013	0.011	0.013	0.010	0.011
random walk	1.03	.	1.03	1.50	2.63	0.98	1.21
Combination forecasts, full panel							
median	0.90	.	0.97	0.92	0.95	0.98	0.99
mean	0.84	.	0.92	0.86	0.92	0.95	0.96
trimmed mean	0.86	.	0.93	0.87	0.93	0.96	0.97
disc. mse(.9)	0.85	.	0.89	0.88	0.96	0.93	0.94
disc. mse(.95)	0.85	.	0.90	0.89	0.96	0.94	0.94
disc. mse(1)	0.85	.	0.90	0.89	0.95	0.94	0.93
recent best	0.53	.	1.31	0.99	0.81	1.16	1.10
Combination forecasts, balanced panel							
median	0.96	.	0.97	1.01	1.00	0.97	0.98
mean	0.90	.	0.92	0.99	0.99	0.95	0.95
trimmed mean	0.92	.	0.93	1.01	1.00	0.95	0.96
disc. mse(.9)	0.88	.	0.90	0.98	0.99	0.94	0.95
disc. mse(.95)	0.88	.	0.91	0.99	0.99	0.94	0.94
disc. mse(1)	0.88	.	0.91	1.00	0.99	0.94	0.93
recent best	0.92	.	1.12	1.34	1.14	1.04	1.47
PC(BIC)	0.85	.	0.83	0.91	0.89	1.46	1.08
PC(AIC)	0.87	.	0.82	0.94	0.90	1.36	1.10
shrink(.25)	0.82	.	1.87	1.39	0.95	1.54	0.96
shrink(.5)	0.82	.	1.22	1.06	0.92	1.19	0.95
shrink(1)	0.89	.	0.93	0.95	0.96	0.95	0.95
tvp(.1)	0.79	.	0.86	0.76	0.80	0.99	0.96
tvp(.2)	0.78	.	0.86	0.70	0.81	1.04	0.99
tvp(.4)	0.76	.	0.87	0.70	0.83	1.05	1.04
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	0.77	.	1.07	1.16	1.07	0.92	1.03
dfm-PC(2,3)	0.84	.	1.08	1.09	1.22	0.90	1.06
n (full panel)	47		44	38	45	39	61
n (bal. panel)	37		37	27	31	29	54

Notes to Table 2: The entry in the first row is the root mean squared forecast error of the benchmark autoregressive forecast (in decimal values of the *h*-period growth, i.e. not at an annual rate). The pseudo out-of-sample forecast period is given in the first row, and the number of forecasts combined in the full panel and in the balanced panel subset are given in the final two rows. The remaining entries are the MSFE of the forecast indicated in the first column, relative to the AR forecast. The forecast mnemonics are:

median	median of individual forecasts at date t
mean	average of individual forecasts at date t
trimmed mean	trimmed mean of individual forecasts at date t, 5% symmetric trimming
disc mse(δ)	combining weights are inversely proportional to discounted forecast errors with discount factor δ
recent best	individual forecast with lowest average squared forecast error over past four quarters
PC(BIC) PC(AIC)	forecasts from regression onto principal components of the panel of forecasts; number of principal components determined by BIC or AIC
shrink(κ)	combining weights are linear combination of equal weighting and recursive OLS, with shrinkage weight $\max\{0, 1 - \kappa[n/(t - T_0 - n)]\}$
TVP(ϕ)	combining weights follow random walk, estimated by Kalman filter, with relative variance $\phi^2(t-T_0)/n^2$
dfm-PC(AIC)	principal components computed using individual leading indicators, and forecasts computed from a dynamic regression including m principal components and p lags of y, where m and p are selected by AIC.
dfm-PC(2,3)	principal components computed using individual leading indicators, and forecasts computed from a dynamic regression including 3 principal components and 2 lags of y.

Table 3. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 4-quarter Growth of Real GDP (h = 4)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	82:I – 98:IV	. .	82:I – 98:IV	82:I – 98:IV	82:I – 98:II	82:I – 98:IV	82:I – 98:IV
Univariate forecasts							
AR RMSE	0.025	.	0.018	0.019	0.023	0.018	0.016
random walk	0.99	.	1.05	1.31	2.97	0.96	1.04
Combination forecasts, full panel							
median	0.92	.	0.99	0.91	0.93	1.00	0.92
mean	0.88	.	1.00	0.82	0.88	0.98	0.90
trimmed mean	0.90	.	0.99	0.82	0.89	0.99	0.91
disc. mse(.9)	0.90	.	0.98	0.85	0.93	0.94	0.90
disc. mse(.95)	0.90	.	1.00	0.89	0.93	0.96	0.89
disc. mse(1)	0.91	.	1.00	0.91	0.92	0.97	0.87
recent best	1.06	.	1.43	0.58	0.87	1.60	1.03
Combination forecasts, balanced panel							
median	0.99	.	1.01	1.03	1.01	0.99	0.92
mean	0.96	.	1.05	1.06	0.98	0.94	0.89
trimmed mean	0.97	.	1.05	1.06	1.00	0.95	0.90
disc. mse(.9)	0.94	.	1.03	1.01	0.97	0.93	0.91
disc. mse(.95)	0.95	.	1.05	1.05	0.98	0.94	0.90
disc. mse(1)	0.95	.	1.05	1.08	0.98	0.94	0.88
recent best	0.90	.	1.36	1.85	1.05	1.26	1.45
PC(BIC)	0.82	.	1.05	0.76	0.68	1.43	0.95
PC(AIC)	0.75	.	1.11	0.77	0.67	1.39	0.98
shrink(.25)	1.32	.	2.26	1.48	1.07	2.14	0.95
shrink(.5)	1.09	.	1.37	1.20	1.02	1.49	0.88
shrink(1)	0.99	.	1.00	1.02	0.97	1.09	0.89
tvp(.1)	0.85	.	0.91	0.55	0.63	1.11	0.98
tvp(.2)	0.97	.	0.98	0.49	0.63	1.27	1.15
tvp(.4)	1.07	.	1.11	0.52	0.65	1.33	1.41
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	0.87	.	1.26	1.12	0.99	1.08	0.88
dfm-PC(2,3)	0.89	.	1.27	1.05	1.14	0.98	0.91
n (full panel)	47		44	38	45	39	61
n (bal. panel)	37		37	27	31	29	54

Notes: See the notes to Table 2.

Table 4. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 8-quarter Growth of Real GDP (h = 8)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	83:I – 97:IV	. .	83:I – 97:IV	83:I – 97:IV	83:I – 97.25	83:I – 97:IV	83:I – 97:IV
Univariate forecasts							
AR RMSE	0.046	.	0.030	0.038	0.046	0.034	0.025
random walk	0.94	.	1.08	1.14	2.74	0.95	0.98
Combination forecasts, full panel							
median	0.95	.	0.98	0.82	0.95	1.04	0.99
mean	0.87	.	0.98	0.68	0.89	1.09	0.96
trimmed mean	0.89	.	0.96	0.71	0.90	1.07	0.98
disc. mse(.9)	0.97	.	1.00	0.79	0.94	1.03	0.97
disc. mse(.95)	0.96	.	1.00	0.81	0.93	1.03	0.96
disc. mse(1)	0.96	.	0.97	0.82	0.93	1.03	0.96
recent best	0.93	.	1.51	0.48	0.95	1.92	1.17
Combination forecasts, balanced panel							
median	1.00	.	1.01	1.02	1.01	1.01	1.00
mean	1.00	.	1.05	0.96	0.99	1.06	0.98
trimmed mean	0.99	.	1.04	1.02	1.00	1.03	0.99
disc. mse(.9)	1.00	.	1.06	0.93	0.98	1.05	0.98
disc. mse(.95)	0.99	.	1.06	0.95	0.99	1.05	0.97
disc. mse(1)	0.99	.	1.05	0.96	0.98	1.06	0.97
recent best	0.86	.	1.18	1.00	1.29	1.85	1.09
PC(BIC)	0.84	.	1.00	0.33	0.43	1.86	1.35
PC(AIC)	0.78	.	1.07	0.36	0.41	2.15	1.30
shrink(.25)	1.68	.	1.58	1.20	0.44	3.65	1.15
shrink(.5)	1.27	.	0.98	0.87	0.53	1.87	0.98
shrink(1)	1.00	.	1.02	0.78	0.78	1.17	0.98
tvp(.1)	0.87	.	1.01	0.32	0.65	1.53	1.15
tvp(.2)	1.08	.	1.24	0.37	0.69	1.94	1.41
tvp(.4)	1.22	.	1.46	0.45	0.69	2.31	1.72
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	0.94	.	1.22	0.98	0.99	1.15	1.03
dfm-PC(2,3)	0.90	.	1.26	0.96	1.18	1.08	1.05
n (full panel)	47		44	38	45	39	61
n (bal. panel)	37		37	27	31	29	54

Notes: See the notes to Table 2.

Table 5. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 2-quarter Growth of IP (h = 2)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV	81:III – 98:IV
Univariate forecasts							
AR RMSE	0.031	0.018	0.026	0.028	0.026	0.018	0.019
random walk	1.17	1.20	1.00	1.07	2.35	1.00	1.30
Combination forecasts, full panel							
median	0.98	0.90	0.94	0.92	0.96	0.97	0.93
mean	0.92	0.89	0.90	0.90	0.94	0.97	0.88
trimmed mean	0.93	0.89	0.90	0.90	0.94	0.97	0.89
disc. mse(.9)	0.92	0.91	0.89	0.93	0.96	0.96	0.88
disc. mse(.95)	0.92	0.92	0.89	0.94	0.96	0.96	0.87
disc. mse(1)	0.92	0.92	0.88	0.93	0.96	0.96	0.85
recent best	0.75	1.31	1.13	0.96	0.83	1.23	0.92
Combination forecasts, balanced panel							
median	0.99	1.02	0.96	0.98	1.01	0.96	0.94
mean	0.93	1.08	0.90	1.00	1.02	0.98	0.89
trimmed mean	0.95	1.05	0.91	1.00	1.03	0.97	0.89
disc. mse(.9)	0.93	1.03	0.90	0.98	1.01	0.97	0.89
disc. mse(.95)	0.92	1.06	0.89	0.99	1.02	0.97	0.88
disc. mse(1)	0.92	1.08	0.89	1.00	1.02	0.98	0.86
recent best	1.01	1.33	1.00	1.07	1.05	1.11	1.18
PC(BIC)	0.99	1.05	0.87	0.83	1.00	1.05	0.85
PC(AIC)	0.92	1.17	0.85	0.84	1.00	1.05	0.82
shrink(.25)	1.27	2.43	1.18	2.35	1.40	2.19	0.89
shrink(.5)	1.04	1.63	1.02	1.16	1.16	1.45	0.89
shrink(1)	0.95	1.29	0.92	0.90	1.02	1.03	0.89
tvp(.1)	0.92	0.97	0.88	0.93	0.92	0.97	0.89
tvp(.2)	0.92	0.93	0.86	0.88	0.88	0.97	0.90
tvp(.4)	0.94	0.96	0.84	0.87	0.90	0.98	0.91
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	0.88	1.17	0.99	1.09	1.13	1.03	1.01
dfm-PC(2,3)	0.88	1.01	1.01	1.07	1.09	1.06	1.00
n (full panel)	47	30	44	38	45	39	61
n (bal. panel)	37	23	37	27	31	29	54

Notes: See the notes to Table 2.

Table 6. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 4-quarter Growth of IP (h = 4)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	82:I – 98:IV	82:I – 98:IV	82:I – 98:III	82:I – 97:IV	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV
Univariate forecasts							
AR RMSE	0.047	0.031	0.037	0.041	0.052	0.026	0.029
random walk	0.96	1.05	1.06	1.11	1.95	1.01	1.09
Combination forecasts, full panel							
median	0.96	0.91	0.97	0.88	0.93	0.96	0.94
mean	0.90	0.91	0.95	0.84	0.88	0.93	0.85
trimmed mean	0.92	0.90	0.95	0.85	0.89	0.95	0.87
disc. mse(.9)	0.95	0.93	0.95	0.90	0.90	0.93	0.86
disc. mse(.95)	0.95	0.94	0.95	0.91	0.89	0.94	0.84
disc. mse(1)	0.95	0.93	0.93	0.91	0.86	0.94	0.83
recent best	1.71	0.93	1.32	0.71	0.78	1.45	1.57
Combination forecasts, balanced panel							
median	0.99	1.05	0.99	1.01	1.01	0.97	0.92
mean	0.96	1.16	0.98	1.03	1.02	0.95	0.86
trimmed mean	0.97	1.12	0.98	1.03	1.03	0.96	0.87
disc. mse(.9)	0.96	1.08	0.97	0.99	1.01	0.94	0.88
disc. mse(.95)	0.96	1.12	0.96	1.02	1.01	0.95	0.85
disc. mse(1)	0.96	1.15	0.96	1.04	1.02	0.95	0.84
recent best	1.69	1.27	1.23	1.83	1.57	1.47	2.10
PC(BIC)	0.89	0.99	0.88	1.01	0.77	1.07	0.77
PC(AIC)	0.89	0.98	0.91	1.00	0.77	1.04	0.80
shrink(.25)	1.13	2.13	1.60	2.01	1.04	2.49	0.85
shrink(.5)	1.01	1.55	1.22	1.32	1.06	1.61	0.86
shrink(1)	0.98	1.32	1.00	0.97	1.06	1.07	0.86
tvp(.1)	0.95	0.92	0.93	0.85	0.83	0.95	0.88
tvp(.2)	1.02	0.89	0.90	0.81	0.80	0.98	0.92
tvp(.4)	1.17	0.95	0.91	0.86	0.85	1.04	1.02
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	0.87	1.30	1.27	1.19	1.19	1.10	1.16
dfm-PC(2,3)	0.88	1.11	1.24	0.92	1.16	1.00	1.15
n (full panel)	47	30	44	38	45	39	61
n (bal. panel)	37	23	37	27	31	29	54

Notes: See the notes to Table 2.

Table 7. MSFES of Combination Forecasts, Relative to Autoregression: Forecasts of 8-quarter Growth of IP (h = 8)

	Canada	**France**	**Germany**	**Italy**	**Japan**	**U.K.**	**U.S.**
Forecast period	83:I – 97:IV	83:I – 97:IV	83:I 97.50	83:I 96.75	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV
Univariate forecasts							
AR RMSE	0.070	0.050	0.054	0.059	0.111	0.041	0.042
random walk	0.99	1.12	1.18	1.23	1.50	1.00	1.00
Combination forecasts, full panel							
median	0.95	0.85	0.94	0.83	0.88	1.02	0.95
mean	0.89	0.83	0.93	0.77	0.79	1.04	0.87
trimmed mean	0.92	0.82	0.91	0.78	0.82	1.03	0.89
disc. mse(.9)	1.08	0.91	0.94	0.88	0.84	0.97	0.90
disc. mse(.95)	1.05	0.95	0.92	0.89	0.84	0.97	0.89
disc. mse(1)	1.04	0.98	0.90	0.90	0.81	0.98	0.89
recent best	1.93	1.21	0.84	0.99	0.55	1.60	1.81
Combination forecasts, balanced panel							
median	1.00	1.02	0.98	0.99	1.02	0.99	0.97
mean	1.00	1.04	0.99	0.98	0.99	1.03	0.89
trimmed mean	1.01	1.05	0.97	1.01	1.01	1.01	0.90
disc. mse(.9)	1.05	0.95	0.96	0.96	0.96	0.99	0.90
disc. mse(.95)	1.04	0.99	0.94	0.98	0.97	1.01	0.89
disc. mse(1)	1.04	1.02	0.93	0.99	0.98	1.02	0.88
recent best	2.28	1.25	0.62	1.62	0.76	1.48	2.33
PC(BIC)	0.92	1.01	0.85	0.71	0.56	1.53	1.06
PC(AIC)	0.92	1.08	0.84	0.71	0.56	1.79	0.98
shrink(.25)	1.30	1.68	1.44	1.98	1.16	2.66	1.16
shrink(.5)	1.07	1.26	0.98	1.11	0.98	1.18	0.90
shrink(1)	1.01	0.96	0.97	1.02	0.95	0.96	0.89
tvp(.1)	1.00	0.82	0.90	0.68	0.59	1.13	0.93
tvp(.2)	1.16	0.95	0.91	0.79	0.56	1.37	0.97
tvp(.4)	1.39	1.10	0.96	1.00	0.62	1.85	1.04
Dynamic factor model forecasts, balanced panel							
dfm-PC(AIC)	1.05	1.18	1.26	0.92	0.95	1.07	1.41
dfm-PC(2,3)	1.08	1.10	1.35	0.89	0.98	1.09	1.37
n (full panel)	47	30	44	38	45	39	61
n (bal. panel)	37	23	37	27	31	29	54

Notes: See the notes to Table 2.

Table 8. Combination Forecasts Ranked by Average Losses: Both Output Measures, 4-quarter Growth

Forecast	Average Loss
disc. mse(1)	0.483
disc. mse(1) - bal. panel	0.489
PC(BIC)	0.489
disc. mse(.95)	0.491
shrink(.5)	0.498
mean	0.498
shrink(1)	0.498
mean - bal. panel	0.498
disc. mse(.95) - bal. panel	0.499
disc. mse(.9)	0.503
trimmed mean - bal. panel	0.504
PC(AIC)	0.504
trimmed mean	0.507
disc. mse(.9) - bal. panel	0.510
shrink(.25)	0.512
median - bal. panel	0.524
tvp(.1)	0.528
median	0.532
AR	0.571
dfm-PC(AIC)	0.586
tvp(.2)	0.588
dfm-PC(2,3)	0.594
random walk	0.609
tvp(.4)	0.687
recent best	0.750
recent best - bal. panel	1.024

Notes: The average losses are weighted averages of the loss of the indicated combination forecast across countries and output measures, where the weighting is by the inverse of the full-sample standard deviation of the variable being forecasted. The average is over 13 sets of forecast (six countries for real GDP, seven countries for IP).

Table 9. Combination Forecasts Ranked by Average Losses:
Both Output Measures, All Horizons (2, 4, 8-quarter growth)

Forecast	Average Loss
mean	0.648
disc. mse(1)	0.653
disc. mse(1) - bal. panel	0.654
disc. mse(.95)	0.655
disc. mse(.95) - bal. panel	0.659
trimmed mean	0.659
disc. mse(.9)	0.659
shrink(1)	0.662
mean - bal. panel	0.662
shrink(.5)	0.663
disc. mse(.9) - bal. panel	0.664
trimmed mean - bal. panel	0.667
median	0.685
median - bal. panel	0.693
random walk	0.698
AR	0.706
tvp(.1)	0.735
PC(AIC)	0.805
shrink(.25)	0.814
tvp(.2)	0.842
dfm-PC(2,3)	0.850
PC(BIC)	0.853
dfm-PC(AIC)	0.858
tvp(.4)	0.976
recent best	1.048
recent best - bal. panel	1.202

Notes: The average losses are averages over both output measures and seven countries (six for real GDP) at three horizons, for a total of 39 cases. See the notes to Table 8.

Table 10. Series Used for Dynamic Factor Model Forecasts

Series Label	Transfor-mation	Canada	France	Germany	Italy	Japan	U.K.	U.S.
rovnght	Δ			X		X		X
rtbill	Δ	X						X
rbnds	Δ							X
rbndm	Δ				X			X
rbndl	Δ	X	X	X	X		X	X
rspread	level			X				X
rexrate	Δln	X	X	X	X	X	X	X
rstockp	Δln	X	X	X	X	X	X	X
rgold	Δln	X	X	X	X	X	X	X
rgdp	Δln	X		X	X	X	X	X
ip	Δln	X	X	X	X	X	X	X
capu	level	X						X
emp	Δln	X		X		X	X	X
unemp	Δ	X		X	X	X	X	X
pgdp	Δ^2ln	X		X	X	X	X	X
cpi	Δ^2ln	X	X	X	X	X	X	X
ppi	Δ^2ln	X		X		X	X	X
earn	Δ^2ln	X	X	X		X		X
roil	Δln	X	X	X	X	X	X	X
rcommod	Δln	X	X	X	X	X	X	X
rm0	Δln							X
rm1	Δln	X		X	X			X
rm2	Δln			X				X
rm3	Δln		X		X			X
n		17	10	18	14	14	13	24

Notes: An "X" indicates that the row series was included, with the indicated transformation, in the construction of the balanced-panel dynamic factor model principal components (dfm-PC) forecasts. The final row gives the total number of series used for the dfm-PC forecasts for each country.

Table 11. Stability of Combination Forecasts: Average Relative MSFEs in Two Subsamples

Forecast	Mean 82:I – 90:II	Mean 90:III – 99:IV	Mean absolute difference, 1st vs. 2nd period	*n*
Univariate forecasts				
Individual Forecasts	1.25	1.13	0.40	2196
random walk	1.23	1.30	0.15	41
Combination forecasts, full panel				
median	0.94	0.94	0.05	41
mean	0.91	0.91	0.08	41
trimmed mean	0.91	0.91	0.08	41
disc. mse(.9)	0.94	0.92	0.09	38
disc. mse(.95)	0.95	0.92	0.09	38
disc. mse(1)	0.95	0.91	0.11	38
recent best	1.31	1.28	0.62	40
Combination forecasts, balanced panel				
median - bal. panel	0.99	0.99	0.03	38
mean - bal. panel	0.99	0.98	0.06	38
trimmed mean - bal. panel	0.99	0.98	0.05	38
disc. mse(.9) - bal. panel	0.97	0.97	0.07	38
disc. mse(.95) - bal. panel	0.98	0.97	0.07	38
disc. mse(1) - bal. panel	0.99	0.97	0.08	38
recent best - bal. panel	1.32	1.56	0.77	38
PC(BIC)	1.26	0.88	0.61	38
PC(AIC)	1.28	0.89	0.64	38
shrink(.25)	1.89	1.66	1.17	38
shrink(.5)	1.15	1.29	0.48	38
shrink(1)	0.99	1.03	0.15	38
tvp(.1)	0.93	0.91	0.19	38
tvp(.2)	1.01	0.99	0.29	38
tvp(.4)	1.13	1.08	0.35	38
Dynamic factor model forecasts, balanced panel				
dfm-PC(AIC)	1.24	1.03	0.34	41
dfm-PC(2,3)	1.17	1.06	0.29	41

Notes: The entries in the second and third columns are the average of the relative MSFEs for the class of forecasts indicated in the first column, over the 1982:I – 1990:II (second column) and the 1990:III – 1999:IV subsample (third column). The fourth column contains the average absolute difference between the relative MSFE in the first and second period, by forecasting method, averaged over the forecasting methods indicated in the first column. The final column reports the number of such methods included in the averages in columns 2, 3, and 4.

Figure 1
Relative MSFE (logarithm) of Pseudo Out-of-Sample Forecasts
Based on Individual Predictors

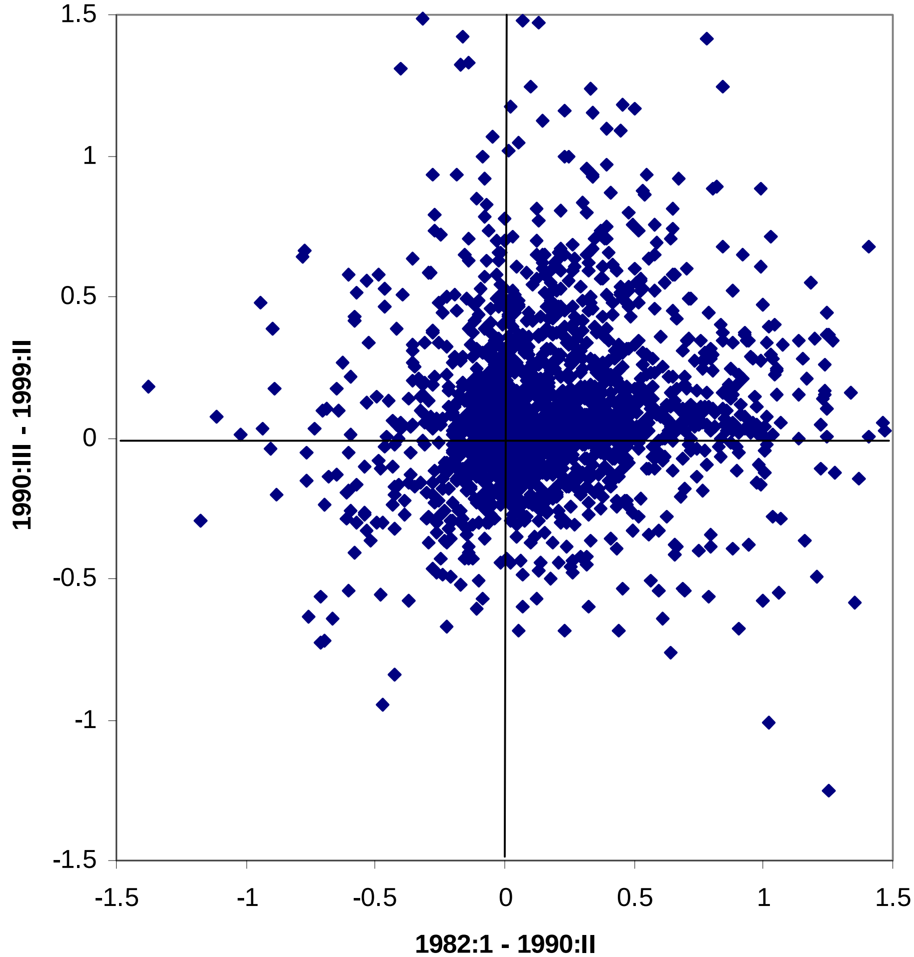

Figure 2
Relative MSFE (logarithm) of Pseudo Out-of-Sample Forecasts
Based on Combined (Mean) Forecast