Forecast Combination With Entry and Exit of Experts*

Carlos Capistrán† Allan Timmermann‡

Banco de México UCSD

June 29, 2007

Abstract

Combination of forecasts from survey data is complicated by the frequent entry and exit of individual forecasters which renders conventional least squares regression approaches infeasible. We explore the consequences of this issue for various combination methods in common use and propose a new method that projects actual outcomes on the equal-weighted forecast to adjust for biases and noise in the underlying forecasts. Through simulations and an application to inflation forecasts we show that the entry and exit of individual forecasters can have a large effect on the real time performance of conventional combination methods. The proposed projection works well in practice.

KEYWORDS: Survey Data, Survey of Professional Forecasters, Composite Forecasts, Inflation.

*We thank Mark Watson (the discussant) and seminar participants at the Real-Time Data Analysis and Methods in Economics conference at the Federal Reserve Bank of Philadelphia. We also thank seminar participants at the 2nd Workshop on Macroeconomic Forecasting, Analysis and Policy with Data Revision at CIRANO, and Banco de México for many helpful comments and suggestions. Andrea San Martín and Gabriel López-Moctezuma provided excellent research assistance. The opinions in this paper are those of the authors and do not necessarily reflect the point of view of Banco de México.

†Banco de México, Dirección General de Investigación Económica. Av. 5 de Mayo No. 18, 4o. Piso, Col. Centro, 06059, México DF, México. Email: ccapistran@banxico.org.mx.

‡Department of Economics and Rady School of Management, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508. Email: agtimmermann@ucsd.edu.

1 Introduction

Evidence of in-sample predictability−established on the same sample used to estimate and select the forecasting model−is widely regarded as being insufficient to demonstrate the value of the resulting forecasts to a decision maker. By ignoring problems associated with parameter estimation errors, model uncertainty and data availability, in-sample forecasts can grossly overstate evidence of genuine predictability. Real-time forecasting experiments are designed to deal with such problems. Establishing that a variable could have been predicted in real time requires using the original data vintages (Amato and Swanson 2001; Croushore and Stark 2001), but also accounting for recursive parameter estimation and even the uncertainty surrounding model selection in real time (Pesaran and Timmermann 2005).

Many of these issues are not a concern when it comes to evaluating forecasts from survey data. By construction, such forecasts were computed in real time. However, when interest lies in real-time combination of survey forecasts, new problems arise. An important issue that has largely been ignored in the literature on forecast combinations is that most expert surveys take the form of unbalanced panels as individual forecasters frequently enter and exit from the surveys. This issue is pervasive and affects the Livingston survey, the Survey of Professional Forecasters (both maintained by the Federal Reserve Bank of Philadelphia), the Michigan surveys (Survey Research Center, University of Michigan), the survey of the Confederation of British Industry (CBI), Consensus Forecasts (Consensus Economics) and surveys of financial analysts' forecasts of corporate earnings (Institutional Brokers' Estimate System, IBES).

As an illustration of this problem, Figure 1 shows how participation in the Survey of Professional Forecasters, available from the Federal Reserve Bank of Philadelphia, evolved over the 5-year period from 1995 to 1999.[1] Each quarter, participants are asked to predict the implicit price deflator for Gross Domestic Product. Forecasters constantly enter, exit and re-enter following a period of absence, creating problems for standard combination ap-

[1] For more on the Survey of Professional Forecasters see Croushore (1993).

proaches that rely on estimating the full covariance matrix for the individual forecasts. Such approaches are not feasible since many forecasters may not have overlapping data and so the covariance matrix cannot be estimated.

This paper considers ways to combine expert opinions that work even in the presence of forecast data that is incomplete with many missing observations. We consider methods such as the equal-weighted average, odds ratio or the previous best forecast in addition to least squares and shrinkage methods modified by trimming forecasts from participants who do not report a minimum number of data points. We also propose a new and very simple approach that first computes the equal-weighted forecast and then projects the realized value on a constant and this forecast. This affine transformation of the equal-weighted forecast does not require each of the underlying forecasts to be unbiased. Furthermore, it can be shown to be optimal for a wider set of parameterizations of the covariance matrix of forecasts than the simple equal-weighted forecast. The method only requires estimating an intercept and a slope parameter through linear projection. Finally, since the method nests the standard equal-weighted forecast (obtained with a zero intercept and a slope of unity), it is easy to test if it improves upon the standard forecast.

We compare the real-time forecasting performance of these methods through Monte Carlo simulations in the context of a common factor model that allows for bias in the individual forecasts, dynamic dependencies in the common factors, and heterogeneity in individual forecasters' ability. In situations with a balanced panel of forecasts the least squares combination methods perform quite well except for when the cross-section of forecasts (N) is large relative to the length of the time-series (T). If the parameters in the Monte Carlo simulations are chosen so that equal-weights are sufficiently suboptimal in population, least-squares combination methods dominate the equal-weighted forecast. Interestingly, however, the simple modification of projecting the outcome on an intercept and the equal weighted forecast continues to outperform regression-based and shrinkage combination forecasts even in many of these experiments.

In the simulations that use an unbalanced panel of forecasts calibrated to match actual survey data, the simulated ("pseudo") real-time forecasting performance of the least squares combination methods deteriorates relative to that of the equal-weighted combination. This happens because the panel of forecasters must be trimmed to get a balanced subset of forecasters from which the combination weights can be estimated by least squares methods. This step entails a loss of information relative to using the equal-weighted forecast which is based on the complete set of individual forecasts. The out-of-sample forecasting performance of the projection on the equal-weighted forecast continues to be very good in the unbalanced panel since this approach makes use of the full set of forecasts in the first stage and then adjusts for any biases remaining in the equal-weighted forecasts in the second stage.

These conclusions are confirmed in an empirical application to inflation forecasts based on the Survey of Professional Forecasters (the data shown in Figure 1). Our analysis estimates combination weights recursively and uses these to compute out-of-sample forecasts. We evaluate these forecasts using both real-time and revised data for actual values. The method that bases the forecast on a projection of the actual value on the equal-weighted forecast is found to do very well out-of-sample compared to a range of alternatives in common use.

The plan of the paper is as follows. Section 2 presents a theoretical framework that allows us to establish conditions under which equal-weights are optimal in population. Section 3 describes commonly used estimation methods from the forecast combination literature and introduces the projection on the mean forecast. Section 4 conducts a Monte Carlo simulation experiment based on the common factor model from Section 2, while Section 5 provides an empirical application to inflation forecasting. Section 6 concludes.

2 Theoretical Results

2.1 Equal-Weighted Forecast Combination

Forecast combinations such as simple averages have proven surprisingly difficult to outperform. This seems to be a robust finding and has been reported in large forecasting experiments involving different types of modeling approaches and a variety of variables in economics, finance and other fields (see, e.g., Clemen 1989; Makridakis and Hibon 2000; Stock and Watson 2001, 2004).

The robustness of the simple average forecast across different data types, time periods, and forecasting methods remains a puzzle. One would expect to find considerable heterogeneity in experts' forecasting ability and this ought to be exploitable by differentiating the weights applied to different forecasts. In practice, however, individual forecasters' true ability−and consequently the combination weights−are unknown and improving upon the equally weighted average requires having a procedure for estimating the combination weights which ensures that the sample estimates do not get too far removed from their true but unknown values. Least-squares procedures (e.g., Granger and Ramanathan 1984) require estimating the covariance matrix of the forecast errors. Achieving a precise estimate of this is often either very difficult or simply not feasible due to (i) the availability of short and incomplete data samples for individual forecasters; (ii) the dimensionality of the problem at hand with a large number of forecasters relative to the length of the time-series; or (iii) instability of the covariance matrix (Kang 1986; Elliott and Timmermann 2005) reflecting structural breaks, time-varying coefficients or other changes in the underlying data generating process.

In this section we establish conditions under which it is optimal (in a population sense) to use equal-weighted forecast combinations and when it is not. This sets a benchmark that proves helpful in understanding the finite-sample forecasting performance in simulations and experiments with actual data. It also points towards directions for improving on the simple average forecast.

Let the variable we are interested in forecasting one step ahead given information at time t be denoted by Y_{t+1} and assume that an $N \times 1$ vector of forecasts computed at time t, $\hat{\mathbf{Y}}_{t+1|t}$, is available. We follow common practice and minimize mean squared error (MSE) loss so only the first two moments of the joint distribution of the predicted variable and forecasts matter:[2]

$$\begin{pmatrix} Y_{t+1} \\ \hat{\mathbf{Y}}_{t+1|t} \end{pmatrix} \sim \left(\begin{pmatrix} \mu_y \\ \boldsymbol{\mu} \end{pmatrix} \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{y\hat{\mathbf{y}}} \\ \boldsymbol{\sigma}_{y\hat{\mathbf{y}}} & \boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} \end{pmatrix} \right). \tag{1}$$

Define the forecast errors associated with the N forecasts as $\mathbf{e}_{t+1,t} = Y_{t+1}\boldsymbol{\iota} - \hat{\mathbf{Y}}_{t+1|t}$, where $\boldsymbol{\iota}$ is an $N \times 1$ vector of ones. From (1) the covariance matrix of the forecast errors, $\boldsymbol{\Sigma}_e = E[\mathbf{e}_{t+1,t}\mathbf{e}'_{t+1,t}]$, is given by:

$$\boldsymbol{\Sigma}_e = (\sigma_y^2 + \mu_y^2)\boldsymbol{\iota}\boldsymbol{\iota}' + \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}_{\widehat{\mathbf{y}}\widehat{\mathbf{y}}} - \boldsymbol{\iota}\boldsymbol{\sigma}'_{y\widehat{\mathbf{y}}} - \boldsymbol{\sigma}_{y\widehat{\mathbf{y}}}\boldsymbol{\iota}' - \mu_y\boldsymbol{\iota}\boldsymbol{\mu}' - \mu_y\boldsymbol{\mu}\boldsymbol{\iota}'. \tag{2}$$

Suppose that the individual forecasts are unbiased so that $\boldsymbol{\mu} = \mu_y\boldsymbol{\iota}$ and consider minimizing the expected forecast error variance subject to the constraint that the weights add up to one (so the combined forecast remains unbiased):

$$\begin{aligned} &\min \boldsymbol{\omega}'\boldsymbol{\Sigma}_e\boldsymbol{\omega} \\ &s.t.\ \boldsymbol{\omega}'\boldsymbol{\iota} = 1. \end{aligned} \tag{3}$$

Assuming that $\boldsymbol{\Sigma}_e$ is invertible and solving the associated Lagrangian optimization, we get the standard solution for the optimal weights:

$$\boldsymbol{\omega}^* = (\boldsymbol{\iota}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\iota})^{-1}\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\iota}. \tag{4}$$

In general the optimal weights depend on the full covariance matrix, $\boldsymbol{\Sigma}_e$. Only in very

[2] To simplify the notation we drop t subscripts on these moments, but it is implicit that all moments can depend on conditioning information available at the time the forecast is computed. The results can easily be generalized to an arbitrary $h-$step forecast horizon.

special cases does (4) reduce to equal weights—the most prominent special case being when the forecast errors have identical variance, σ^2, and identical pair-wise correlations, ρ, $(-1 < \rho < 1)$. In this case we get:

$$\begin{aligned}
\boldsymbol{\Sigma}_e^{-1} &= \frac{1}{\sigma^2(1-\rho)}\left(\mathbf{I} - \frac{\rho}{1+(N-1)\rho}\boldsymbol{\iota}\boldsymbol{\iota}'\right) \\
&= \frac{1}{\sigma^2(1-\rho)(1+(N-1)\rho)}\left((1+(N-1)\rho)\mathbf{I} - \rho\boldsymbol{\iota}\boldsymbol{\iota}'\right),
\end{aligned}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. Inserting this expression in (4), we have:

$$\begin{aligned}
\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\iota} &= \frac{\boldsymbol{\iota}}{\sigma^2(1+(N-1)\rho)} \\
(\boldsymbol{\iota}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\iota})^{-1} &= \frac{N}{\sigma^2(1+(N-1)\rho)},
\end{aligned}$$

and it follows that the optimal weights are given by:

$$\boldsymbol{\omega}^* = \left(\frac{1}{N}\right)\boldsymbol{\iota}. \tag{5}$$

Hence equal-weights are optimal in situations with an arbitrary number of forecasts when the individual forecast errors have mean zero, identical variance and (arbitrary) identical pair-wise correlations. The weights add up to unity only as a result of imposing these constraints and will not otherwise hold in general. In the absence of some notion of the underlying data generating process, it is difficult to tell how plausible such constraints are so we next turn to a model that allows us to better interpret the constraints.

2.2 A Common Factor Model

Realistic and empirically plausible covariance structures can be obtained from common factor models which are widely used empirically to forecast macroeconomic and financial time series (see, e.g., Chan, Stock and Watson 1999 and Stock and Watson 2006). Moreover, intuition can be gained in terms of the structure of the factor loadings, idiosyncratic noise and variability of the individual factors. Accordingly, let the target variable, Y_{t+1}, and the individual forecasts,$\hat{Y}^i_{t+1|t}$, be driven by the following common factor model:

$$\begin{aligned} Y_{t+1} &= \mu_y + \boldsymbol{\beta}'_{yF}\mathbf{F}_{t+1} + \varepsilon_{yt+1}, \quad \varepsilon_{yt+1} \sim N(0, \sigma^2_{\varepsilon_Y}) \\ \hat{Y}^i_{t+1|t} &= \mu_i + \boldsymbol{\beta}'_{iF}\mathbf{F}_{t+1} + \varepsilon_{it+1}, \quad \varepsilon_{it+1} \sim N(0, \sigma^2_{\varepsilon_i}), \quad i = 1, ..., N, \end{aligned} \tag{6}$$

where we assume that $E\left[\varepsilon_{it+1}\varepsilon_{jt+1}\right] = 0$ if $i \neq j$, and $E[\varepsilon_{it+1}\varepsilon_{yt+1}] = 0$ for $i = 1, ..., N$. As we shall see, an advantage of this model is that it is sufficiently rich to cover a variety of empirically relevant scenarios.

Dynamics in the n_f common factors can also be introduced:

$$\mathbf{F}_t = \mathbf{B}_F\mathbf{F}_{t-1} + \boldsymbol{\varepsilon}_{F_t}, \quad \boldsymbol{\varepsilon}_{F_t} \sim N(\mathbf{0}, \mathbf{D}_{\varepsilon_F}), \tag{7}$$

where $\mathbf{D}_{\varepsilon_F}$ is an $n_f \times n_f$ diagonal matrix with entries:

$$\mathbf{D}_{\varepsilon_F} = \begin{pmatrix} \sigma^2_{F_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{F_2} & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & & \sigma^2_{F_{n_f}} \end{pmatrix},$$

and $E[\varepsilon_{yt+1}\boldsymbol{\varepsilon}_{Ft}] = E[\varepsilon_{it+1}\boldsymbol{\varepsilon}_{Ft}] = \mathbf{0}$, for $i = 1, ..., N$. We assume that the eigenvalues of $\mathbf{B}_F$ all lie outside the unit circle so $(\mathbf{I} - \mathbf{B}_F)^{-1}$ exists and the initial value $\mathbf{F}_0$ can be drawn from the unconditional distribution of the factors. This gives the following convenient form of the (unconditional) covariance-matrix of the joint distribution of Y and $\widehat{\mathbf{Y}}$:

$$\begin{aligned}
\sigma_y^2 &= \boldsymbol{\beta}_{yF}'(\mathbf{I}-\mathbf{B}_F^2)^{-1}\mathbf{D}_{\varepsilon_F}\boldsymbol{\beta}_{yF}+\sigma_{\varepsilon_Y}^2, \\
\boldsymbol{\sigma}_{y\hat{\mathbf{y}}}[i] &= \boldsymbol{\beta}_{iF}'(\mathbf{I}-\mathbf{B}_F^2)^{-1}\mathbf{D}_{\varepsilon_F}\boldsymbol{\beta}_{yF} \\
\boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}[i,j] &= \boldsymbol{\beta}_{iF}'(\mathbf{I}-\mathbf{B}_F^2)^{-1}\mathbf{D}_{\varepsilon_F}\boldsymbol{\beta}_{jF}+\mathcal{I}_{\{i=j\}}\sigma_{\varepsilon_i}^2,
\end{aligned} \tag{8}$$

where $\mathcal{I}_{\{i=j\}}$ is an indicator function that equals unity if $i=j$ and otherwise is zero.

This model nests several cases of common interest. A particular forecast is conditionally unbiased when $\mu_i=\mu_y$, $\boldsymbol{\beta}_{iF}=\boldsymbol{\beta}_{yF}$ and $\sigma_{\varepsilon_i}^2=0$. As $\sigma_{\varepsilon_i}^2$ increases, an increasingly important noise component is added to the forecasts which become less valuable and it becomes more attractive to combine forecasts rather than using a single prediction. Such noise may be due to model misspecification such as when the wrong predictor variables or predictor variables subject to measurement errors are used in the forecaster's model. When $\sigma_{\varepsilon_Y}^2>0$, the target variable, Y_{t+1}, comprises an unpredictable component and as $\sigma_{\varepsilon_Y}^2$ goes up, the predictive R^2 of each individual forecast declines. Cross-sectional heterogeneity in the individual forecasters' performance can be introduced by letting any one of the parameters $(\mu_i,\beta_{iF},\sigma_{\varepsilon_i}^2)$ differ across forecasters.

Since the factor model is quite general, we impose additional structure to ensure that the individual forecasts are sensible. In particular, notice that the best linear projection of Y_t on $\hat{Y}_{t|t-1}^i$ is given by:

$$\frac{\boldsymbol{\beta}_{iF}'(\mathbf{I}-\mathbf{B}_F^2)^{-1}\mathbf{D}_{\varepsilon_F}\boldsymbol{\beta}_{yF}}{\boldsymbol{\beta}_{iF}'(\mathbf{I}-\mathbf{B}_F^2)^{-1}\mathbf{D}_{\varepsilon_F}\boldsymbol{\beta}_{iF}'+\sigma_{\varepsilon_i}^2}. \tag{9}$$

We can choose parameter values such that this is equal to unity, ensuring that the individual forecasts are unbiased, a restriction often deemed sensible.

Furthermore, letting $\sigma_{F_i}^2=\sigma_F^2$ for $i=1,..,n_f$, $\sigma_{\varepsilon_i}^2=\sigma_\varepsilon^2$ for $i=1,..,N$, assuming (purely for simplicity) that there is no factor dynamics ($\mathbf{B}_F=\mathbf{O}$) and letting $\boldsymbol{\beta}_{iF}=\beta\boldsymbol{\iota}$, $\boldsymbol{\beta}_{yF}=\beta_y\boldsymbol{\iota}$, we have:

$$\boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(\mathbf{I} - \frac{Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2 + Nn_f\beta^2\sigma_F^2}\boldsymbol{\iota\iota}'\right)$$

and so the optimal combination weights are:

$$\begin{aligned}
\omega^* &= \boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\boldsymbol{\sigma}_{y\hat{\mathbf{y}}} \\
&= \frac{1}{\sigma_\varepsilon^2}\begin{pmatrix}
1-\frac{Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} & \frac{-Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} & \cdots & \frac{-Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} \\
\frac{-Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} & 1-\frac{Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} & & \frac{-Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} \\
\vdots & & \ddots & \vdots \\
\frac{-Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2} & \cdots & & 1-\frac{Nn_f\beta^2\sigma_F^2}{\sigma_\varepsilon^2+Nn_f\beta^2\sigma_F^2}
\end{pmatrix}\begin{pmatrix} n_f\beta\beta_y\sigma_F^2 \\ \vdots \\ \\ n_f\beta\beta_y\sigma_F^2 \end{pmatrix}.
\end{aligned}$$

Equal weights that sum to unity are optimal in this setting provided that:

$$\frac{n_f\beta\beta_y\sigma_\varepsilon^2\sigma_F^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + Nn_f\beta^2\sigma_F^2)} = \frac{1}{N}.$$

This constraint only holds in the special case where:

$$\sigma_\varepsilon^2 = n_f N\beta\sigma_F^2(\beta_y - \beta). \tag{10}$$

Using identical combination weights that sum to unity is clearly only optimal as a very special case and restricting the individual forecasts to be unbiased through (9) does not, in general, ensure that it is optimal to use equal weights. Furthermore, variations in the variance-covariance parameters introduce heterogeneity in forecasting performance and generally have the effect of moving the optimal weights even further away from $1/N$.

Conversely, if $\boldsymbol{\beta}_{iF} = \boldsymbol{\beta}_{jF}$ and $\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon j}^2$ for all i, j, then all diagonal elements of $\boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}$ are identical as are the off-diagonal elements. This means that equal weights are optimal although the weights will not necessarily sum to one. In such situations the optimal forecast is formed as a scalar (not necessarily equal to one) times the equal-weighted average of the individual forecasts. While still a special case, this covers many more situations than the case where the simple equal-weighted forecast is optimal.

3 Methods for Estimating Combination Weights

The theoretical analysis in the previous section suggested that equal-weighted combinations are only optimal under a set of highly restrictive conditions on the joint distribution of the forecasts and target variable. That this forecasting method generally performs so well empirically can therefore in all likelihood be attributed to the fact that it does not require the estimation of any combination weights. In practice, parameter estimation error is an important determinant of relative forecasting performance. This also explains why least squares methods which require estimating the covariance matrix of forecast errors tend to perform poorly empirically and why shrinkage towards equal-weights—a practice that introduces bias but reduces the effect of parameter estimation errors—often is found to improve on least squares methods. Clearly the effect of estimation error on forecasting performance can be very significant.

To address this point we next describe a variety of methods for forecast combination in common use and propose a new and simple method that modifies the equal-weighted forecast,

$$\bar{Y}_{t+1|t} = N_t^{-1} \sum_{i=1}^{N_t} \hat{Y}_{t+1|t}^{(i)} \qquad (11)$$

which serves as a natural benchmark. Another option is simply to use the previous best model based on past performance. This approach places all the weight on the single forecast with the lowest historical MSE-value

$$\begin{aligned} \hat{Y}^*_{t+1|t} &= \hat{Y}^{i_t^*}_{t+1|t}, \quad \text{where} \\ i_t^* &= \arg \min_{i=1,\ldots,N_t} t^{-1} \sum_{\tau=1}^{t} (Y_\tau - \hat{Y}^i_{\tau|\tau-1})^2 \end{aligned} \qquad (12)$$

While it may seem that this method does not require any estimation, this is not quite true since the ranking of the various models itself follows a stochastic process that may lead to shifts in the selected model as new data emerges.

3.1 Least Squares Estimation of Combination Weights

It is common to estimate combination weights by ordinary least squares, regressing realizations of the target variable, Y_τ on the N-vector of forecasts, $\hat{\mathbf{Y}}_{\tau|\tau-1}$ using data over the period $\tau = 1, ..., T$:

$$\hat{\boldsymbol{\omega}}_T = (\sum_{\tau=1}^{T-1} \hat{\mathbf{Y}}_{\tau+1|\tau} \hat{\mathbf{Y}}'_{\tau+1|\tau})^{-1} \sum_{\tau=1}^{T-1} \hat{\mathbf{Y}}_{\tau+1|\tau} Y_{\tau+1}. \tag{13}$$

Different versions of this least squares projection have been proposed. Granger and Ramanathan (1984) consider three regressions:

$$\begin{aligned}
(i)\ Y_{t+1} &= \omega_t^0 + \boldsymbol{\omega}_t' \hat{\mathbf{Y}}_{t+1|t} + \varepsilon_{t+1} \\
(ii)\ Y_{t+1} &= \boldsymbol{\omega}_t' \hat{\mathbf{Y}}_{t+1|t} + \varepsilon_{t+1} \\
(iii)\ Y_{t+1} &= \boldsymbol{\omega}_t' \hat{\mathbf{Y}}_{t+1|t} + \varepsilon_{t+1}, \text{ s.t. } \boldsymbol{\omega}_t' \boldsymbol{\iota} = 1.
\end{aligned} \tag{14}$$

The first and second of these regressions can be estimated by standard OLS, the only difference being that the second equation omits an intercept term. The third regression omits an intercept and can be estimated through constrained least squares. The first and more general regression does not require the individual forecasts to be unbiased since any bias can be adjusted through the intercept term, ω_t^0. In contrast, the third specification is motivated by an assumption of unbiasedness of the individual forecasts. Imposing that the weights sum to one then guarantees that the combined forecast is also unbiased.[3] One could further impose convexity constraints $0 \leq \omega_t^i \leq 1$, $i = 1, .., N_t$ to rule out that the combined forecast lies outside the range of the individual forecasts.

An obvious problem with this approach is that it is very poor at handling unbalanced

[3]This specification may not be efficient, however, as the latter constraint can lead to efficiency losses as $E[\hat{\mathbf{Y}}_{t+1|t} \varepsilon_{t+1}] \neq \mathbf{0}$.

data sets such as those from the Survey of Professional Forecasters shown in Figure 1. It is simply not feasible to estimate the complete covariance matrix for this type of data. In such cases, minimum data requirements must be imposed and the set of forecasts trimmed. For example, one can require that forecasts from a certain minimum number of (not necessarily contiguous) common periods be available.

To partially address this issue, we also apply a weighting scheme which, for forecasters with a sufficiently long track record, uses weights that are inversely proportional to their historical MSE-values, while using equal-weights for the remaining forecasters (normalized so the weights sum to one).

3.2 Shrinkage

Shrinkage methods have been widely used in forecasting. For example, Stock and Watson (2004) propose shrinkage towards the arithmetic average of forecasts. Let $\hat{\omega}_t^i$ be the least-squares estimator of the weight on the ith model in the forecast combination obtained, e.g., from one of the regressions in (14). The combination weights considered by Stock and Watson take the form:

$$\begin{aligned} \omega_t^i &= \psi\hat{\omega}_t^i + (1-\psi)(1/N_t), \\ \psi_t &= \max(0, 1 - \kappa N_t/(T-1-N_t-1)), \end{aligned}$$

where κ regulates the strength of the shrinkage with larger values of κ implying a lower ψ_t and thus a greater degree of shrinkage towards equal weights. As the sample size, T, rises relative to the number of forecasts, N, the least squares estimate gets a larger weight. While this approach can be assumed to work well in some situations, because it is based on the least squares estimator it is also likely to suffer from the deficiencies of that approach. Conversely, for fixed values of T and N, larger values of κ correspond to more shrinkage towards equal-weighting (smaller ψ_t) and hence present some of the problems associated with using equal weights.

3.3 Odds Matrix Approach

The odds matrix approach (Gupta and Wilton 1987) computes the combination of forecasts as a weighted average of the individual forecasts where the weights are derived from a matrix of pair-wise odds ratios. Each entry in the matrix is interpreted as the odds that forecast i will outperform forecast j. If the odds matrix is denoted $\mathbf{O}$, then the weight vector, $\boldsymbol{\omega}$, is obtained from the solution to $(\mathbf{O} - N\mathbf{I})\,\boldsymbol{\omega} = \mathbf{0}$, where $\mathbf{I}$ is the identity matrix.[4] Estimation of the matrix $\mathbf{O}$ is accomplished by estimating the pair-wise probabilities π_{ij} that represent the probability that the ith forecast will outperform the jth forecast in the next realization. The entries of the $\mathbf{O}$ matrix are then $o_{ij} = \frac{\pi_{ij}}{\pi_{ji}}$. There are several ways to estimate the pair-wise probabilities. Following the empirical application of Gupta and Wilton (1987), we use $\pi_{ij} = \frac{a_{ij}}{(a_{ij}+a_{ji})}$, where a_{ij} is the number of times forecast i had a smaller absolute error than forecast j in the historical sample.[5]

3.4 Projection on the (Equal-Weighted) Mean

We next propose a new method for forecast combination that exploits some of the advantages of using the equal-weighted average but uses information in this average in a more flexible manner. To motivate the approach, consider the following insights from the literature on forecast combination (for a survey, see Timmermann 2006): (i) estimation of additional parameters used to combine the forecasts quickly leads to deteriorating forecasting performance; (ii) individual expert forecasts are often biased and the slope coefficient in

[4] Gupta and Wilton (1987, 1988) propose to use the normalized eigenvector associated with the eigenvalue that solves $\mathbf{O}\boldsymbol{\omega} = \tau_{\max}\boldsymbol{\omega}$ (the largest positive eigenvalue) as an estimate of the weight vector. This is the approach we follow.

[5] Shrinkage and the odds matrix approach can both be viewed as Bayesian combination methods. More formal Bayesian methods could also be employed, for example, Bayesian Model Averaging (see Jacobson and Karlsson (2004)) or one could use Bagging (see Inoue and Kilian (Forthcoming)).

a regression of the realized value on individual forecasts often differs from unity;[6] (iii) bias correction is best done at the level of the combined forecast by including a single intercept and more refined adjustments generally do not lead to large improvements; (iv) forecasts from data sources such as surveys are generally highly unbalanced which makes standard covariance-based approaches difficult to apply.

Based on these considerations, we propose a simple affine transformation of the equal-weighted forecast, $\bar{Y}_{t+1|t} = N_t^{-1} \sum_{i=1}^{N_t} \hat{Y}^i_{t+1|t}$:

$$\tilde{Y}_{t+1|t} = \hat{\alpha}_t + \hat{\beta}_t \bar{Y}_{t+1|t}. \tag{15}$$

This extension of the equal-weighted combination only requires estimating two parameters, α and β, which can be done through least squares regression. As in the case with the simple equal-weighted average, information from forecasters with no more than a single data point can be used. By including a constant, the forecast combination method adjusts for biases that may be present in the individual forecasts as well as in the aggregate. By allowing for a slope coefficient different from unity, as shown in Section 2, the method is likely to work well under a much broader set of scenarios than the simple equal-weighted forecast.

4 Monte Carlo Simulations

To analyze the determinants of the performance of the various forecast combination methods, we next conduct a series of Monte Carlo experiments in the context of the factor model described in Section 2. In all experiments we use two factors, $n_f = 2$, so that $F = 1, 2$. We let the sample size, T, vary from 100 to 500 and 1000 and let the number of forecasts (N) assume values of 4, 10 and 20. This covers situations with large N relative to the sample size t (e.g., $N = 20, T = 100$) as well as situations with plenty of data points relative to the

[6] Zarnowitz (1985) finds evidence against efficiency in individual forecasters' predictions. Davies and Lahiri (1995) report evidence that informational efficiency is rejected for up to half of the survey participants in their data.

number of estimated parameters (e.g., $N = 4$, $T = 1,000$). All forecasts are one-step-ahead, simulated out-of-sample, and are computed based on recursive parameter estimates using only information available at the time of the forecast. To forecast Y_{t+1} we therefore use information only up to period t, including for the estimates of the combination weights, $\hat{\boldsymbol{\omega}}_t$.

The first set of experiments assumes that the individual forecasts are unbiased and set $\mu_y = \mu_i = 0$ $(i = 1, ..., N)$. In the base experiment (experiment 1) we further assume that β_{i1} solves (10) for all i so the true optimal weights are identical and sum to unity. Furthermore, we set:

$$\begin{aligned}
\boldsymbol{\beta}_y &= (1 \;\; 1)' \\
\sigma_{\varepsilon_Y} &= \sigma_{\varepsilon_{F_1}} = \sigma_{\varepsilon_{F_2}} = 1; \quad \sigma_{\varepsilon_i} = 1 \quad i = 1, .., N \\
\mathbf{B}_F &= \mathbf{0}.
\end{aligned}$$

In experiments 2-7 we assume that $\beta_{i1} = 0.5$, $(i = 1, ..., N)$ while β_{i2} solves (9) for $i = 1, ..., N$ ensuring that the regression coefficient of Y_{t+1} on the individual forecasts $\hat{Y}^i_{t+1|t}$ is unity. Factor dynamics is introduced in experiment 3 by letting $\mathbf{B}_F = 0.9 \times \mathbf{I}$. Heterogeneity in the individual forecasters' ability is introduced by drawing the factor loadings, β_{if}, from a Beta distribution centered on 0.5 with either low dispersion ($Beta(5, 5)$ in experiment 4) or high dispersion ($Beta(1, 1)$ in experiment 5). To allow for the possibility that different experts capture different predictable components (thus enhancing the role of forecast combinations over the individual models), experiment 6 considers a scenario where different groups of forecasts load on different factors. Finally, forecast biases are introduced by allowing for a non-zero intercept in experiment 7. To summarize, we alter the base scenario as follows:

	Scenarios	Change in Parameters
1	Base scenario ($\boldsymbol{\omega}'\boldsymbol{\iota} = 1$)	–
2	Identical weights ($\boldsymbol{\omega}'\boldsymbol{\iota} \neq 1$)	β_i solves (9)
3	Factor Dynamics	$\mathbf{B}_F = 0.9 \times \mathbf{I}$
4	Weak heterogeneity	$\beta_{if} \sim Beta(5,5)$
5	Strong heterogeneity	$\beta_{if} \sim Beta(1,1)$
6	Factor-loadings in blocks	$\boldsymbol{\beta}'_{i1} = \begin{cases} 1 & \text{if } 1 \leq i \leq N/2 \\ 0 & \text{if } N/2 < i \leq N \end{cases}$ $\boldsymbol{\beta}'_{i2} = \begin{cases} 0 & \text{if } 1 \leq i \leq N/2 \\ 1 & \text{if } N/2 < i \leq N \end{cases}$.
7	Biased forecast	$\mu_i = \begin{cases} 1/2 & \text{if } 1 \leq i \leq N/2 \\ 0 & \text{if } N/2 < i \leq N \end{cases}$.

Following the analysis in Section 3, we compare the following *ten* combination methods:

Label	Combination Method
EW	Equal-weighted forecast
PEW	Projection on constant and equal-weighted forecast
GR1	Unconstrained OLS: (14, (i))
GR2	OLS w/o constant: (14, (ii))
GR3	Constrained OLS w/o constant: (14, (iii))
Shrink1	Shrinkage with $\kappa = 0.25$
Shrink2	Shrinkage with $\kappa = 1$
Odds	Odds ratio
Previous Best	Forecast from previous best model
Inverse MSE	Weight is the inverse of the historical MSE

4.1 Balanced Panel of Forecasters

Results for the case with a balanced panel of forecasts are reported in Table 1 in the form of out-of-sample MSE-values computed relative to the MSE-value associated with the equal-weighted forecast (which is thus always equal to unity). In the base scenario the simple equal-weighted forecast performs best since it imposes a true constraint on the combination weights and hence achieves efficiency gains.[7] However, the simple equal-weighted forecast is not producing particularly precise forecasts in the other scenarios (experiments 2-7) even though the parameters of the Monte Carlo experiments are chosen such that the population value of a regression of Y_{t+1} on the individual forecasts $\widehat{Y}^{i}_{t+1|t}$ is unity. The reason is that although using equal weights is optimal in settings without cross-sectional heterogeneity (see our discussion in Section 2), the optimal weights need not add up to unity.

In the base scenario (experiment 1) the best combination scheme among those proposed by Granger and Ramanathan (1984) is to exclude an intercept and impose that the weights sum to unity. This holds across all sample sizes and cross-sections—imposing a true constraint ensures efficiency gains. The improvement over the most general least squares regression (GR1) is, however, quite marginal—about 1-2%. Conversely, when the true weights do not sum to unity, as in the second experiment, the most constrained least squares combination (GR3) produces MSE-values that are far worse than the less constrained models (GR1 and GR2). Constraining the intercept to be zero (GR2) leads to marginally better performance than under the unconstrained least squares model (GR1) when this constraint holds as in experiments 2-6, although it leads to inferior performance when the underlying forecasts are in fact biased (experiment 7).

Turning to the shrinkage forecasts, these generally improve on the benchmark equal-weighted combination's performance. In most cases the shrinkage approach does as well as or slightly better than the best least squares approach. When the sample size is small, the

[7] The odds ratio and the inverse MSE approaches seem to marginally outperform the equal-weighted average but the numbers are very close to one and are more likely the result of sampling variation.

model with the largest degree of shrinkage does best. However, using a smaller degree of shrinkage becomes superior as the sample size, T, is raised (for fixed N). The benefit from shrinkage is particularly sizeable when the number of models is large as when $N = 20$.

Although the differences in MSE-values are small, the odds matrix and inverse MSE approaches generally dominate using equal-weights. In contrast, choosing the single best model does not lead to good forecasting performance in the experiments without heterogeneity where (*ex ante*) the forecasting models are equally good. Combining forecasts therefore works well in such settings as it allows the user to dilute the noise in the individual forecasts. As expected, the out-of-sample forecasting performance of the previous best model improves as the degree of heterogeneity across models gets stronger and a clearer picture of the single best model emerges.

Factor dynamics—introduced in the third Monte Carlo experiment—leads to deteriorating forecasting performance across all combination schemes. This is not surprising since the effective sample size is smaller in the presence of persistent factors. Interestingly, factor dynamics also has the effect of improving the relative performance of the most general least squares methods (GR1 and GR2), shrinkage and projection on the equal-weighted mean (PEW).

Heterogeneity in the factor loadings of the various forecasts—introduced by drawing these from a beta distribution—has two effects. First, it means that the true performance now differs across forecasting models. Models with larger factor loadings have a higher R^2 than models with small factor loadings. Secondly, the combination schemes that are based on equal weights now perform worse. This follows from our discussion in Section 2 which showed that (generically) equal weights are optimal only when the forecasts errors have identical variances with the same pair-wise correlations. The effect of weak heterogeneity on the performance of the various combination schemes ($\beta_{if} \sim Beta(5,5)$ in experiment 4) is quite minor. However, as stronger heterogeneity is introduced in the distribution of factor loadings ($\beta_{if} \sim Beta(1,1)$ in experiment 5), the simple equal-weighted forecasts perform worse and using the previous best forecasting model becomes more attractive−although this strategy is still dominated by many of the other combination methods.

When half of the forecasts track factor one while the remaining half of the forecasts track factor two (experiment 6), the benefits from combining over using the single best model (which can only track one factor at a time) tend to be particularly large. Moreover, the projection on equal weights (PEW) performs very well and the least constrained OLS and shrinkage forecasts also continue to perform well relative to the benchmark.

When we let half of the forecasts be biased with a bias equal to one-half of the standard deviation parameters, the efficiency gain due to omitting an intercept in the least squares combination regression is now more than out-done by the resulting bias. This explains why the general Granger-Ramanathan scheme (GR1) which includes an intercept term now produces better results than the constrained Granger-Ramanathan regressions (GR2 and GR3). The previous best model produces worse forecasting performance than in the case without bias, as there is always the risk of selecting a biased model. Since the shrinkage methods pull the least squares forecast towards the biased equal-weighted forecast, this also explains why the shrinkage schemes perform worse than in the case without a bias and generally produce worse results than the least squares methods. In contrast, the performance of the forecast that uses a projection on an intercept and the equal-weighted forecast is unchanged compared with the results in the second experiment since only the intercept is changed.

Overall, the best forecasting performance is produced by the simple combination method that regresses Y_{t+1} on an intercept and the equal-weighted forecast, $\overline{Y}_{t+1|t} = N^{-1} \sum_{i=1}^{N} \widehat{Y}_{t+1|t}^{i}$. This approach produces better results than the equal weighted forecast in all experiments except the first one (for which a small under-performance of up to two percent is observed). Furthermore, it generally does best among all combination schemes in experiments 2-7, with slightly better results observed for the least squares and shrinkage methods in the presence of strong heterogeneity (experiment 5).[8]

4.2 Unbalanced Panel of Forecasters

We next perform the same set of experiments on data generated from the two-factor model filtered so as to mimic the unbalanced panel structure of the Survey of Professional Forecasters data shown in Figure 1. To this end we first group the experts into frequent and infrequent forecasters defined according to whether a forecaster participated in the survey a minimum of 75 percent of the time. Next, we pool observations within each of the two groups of forecasters and estimate two-state Markov transition matrices for each group, where state one represents participation in the survey while state 2 is absence from the survey. The estimated transition matrices were:

$$Frequent\ Participation : \begin{pmatrix} 0.84 & 0.16 \\ 0.41 & 0.59 \end{pmatrix}, \tag{16}$$

$$Infrequent\ Participation : \begin{pmatrix} 0.69 & 0.31 \\ 0.03 & 0.97 \end{pmatrix}. \tag{17}$$

Among frequent forecasters there is an 84 percent chance of observing a forecast next period if a forecast was reported in the current period. This probability declines to 40 percent if no forecast was reported in the current period. Conversely, the extremely high probability (0.97) of repeated non-participation among the infrequent forecasters shows that this category covers forecasters who rarely participate in the survey.

We use these transition matrices to generate a matrix of zeros and ones that indicates when a forecaster participated in the survey. We then multiply, element-by-element, the zero-one participation matrix with the matrix of one-step forecasts generated from the two-factor

[8] Under strong heterogeneity among the forecasters, the equal-weighted forecast is sub-optimal. This explains why in experiment 5 regression-based approaches which assign different weights to the individual forecasters perform relatively better than methods based on the equal-weighted forecast.

model and apply the combination methods to the resulting (unbalanced) set of forecasts.

To apply least-squares combination methods we trimmed those forecasters with fewer than 20 contiguous forecasts or no prediction for the following period. Among the remaining forecasters we next used the largest common data sample to estimate the combination weights. If there were no forecasters with at least 20 contiguous observations or if there are fewer remaining forecasters than parameters to be estimated, we simply use the average forecast for next period. Our simulations assume that the proportion of frequent forecasters is set at 40 percent. This means that we only have to resort to using equal-weights one-third of the time−a number similar to that found in the empirical application in the next section.

Results are reported in Table 2 in the form of out-of-sample MSE-values again measured relative to the values generated by the equal-weighted forecast combination. Since unbalanced panels are more likely to occur in settings with a relatively large number of forecasters, we only report results for $N = 20$.

Compared with the earlier results in Table 1, the previous best, odds matrix and inverse MSE approaches perform more like the equal-weighted approach in the unbalanced panel. A similar finding holds for the least squares combination and shrinkage approaches. The performance of such methods (relative to the simple equal-weighted approach) is therefore relatively worse in the unbalanced panel. Two reasons explain this finding. First, in about one-third of the periods the regression methods revert to using the equal-weighted forecast because a balanced subset of forecasters with a sufficiently long track record cannot be found. Second, since the regression methods trim the set of forecasters to obtain a balanced subset of forecasts, they discard potentially valuable information. Consequently, these methods perform worse than the equal-weighted average when conditions are in place for the latter to work well and only outperform by a small margin otherwise.

Overall, the PEW method continues to perform better than the other approaches even with an unbalanced panel of forecasts. Moreover, the relative performance of this approach generally improves over the case with a balanced panel. For example, in the second experiment the relative MSE-value associated with this approach goes from 0.72 in the balanced panel to 0.56 in the unbalanced panel. Similar improvements are observed in experiments 4 and 5. The only experiment where a worse (relative) performance is observed is in the sixth experiment with a block diagonal factor structure. However, even in this case the PEW approach remains the best overall.

We conclude that, across the board, the proposed equal-weighted projection method is better than the other methods that can be used when estimation of the full covariance matrix of the forecast errors is not feasible (equal weights, odds matrix or previous best forecast). It also performs better than the regression and shrinkage approaches modified so they can be used on a balanced subset of forecasters. Although the PEW approach does better in most experiments irrespective of whether a balanced or an unbalanced panel of forecasts is available, the extent to which the PEW approach outperforms tends to be greater in the unbalanced panel.

5 Application to Inflation Forecasts

To illustrate the performance of the combination methods on actual data we use one- and four-step-ahead inflation forecasts from the Survey of Professional Forecasters. Inflation is measured as the annualized quarterly change in the output deflator using either fully revised data or real-time data from the Federal Reserve of Philadelphia's web site. Fully revised data is the last revision as of January 2007, whereas real-time data corresponds to the first revision. We restrict the data sample to start in the fourth quarter of 1979 to take into account the change in monetary policy that occurred when Paul Volcker took office as Chairman of the Federal Reserve. This change is widely regarded as having affected the behavior of inflation (see Clarida, Galí, and Gertler 2000, among others). Our sample ends with the forecasts made for the third quarter of 2006. At each period in time there are between 9 and 49 forecasters, with a median value of 32.

We calculate the mean forecast at each point in time (equal-weighted combination, $\overline{Y}_{t+h|t} = (1/N_t)\sum_{i=1}^{N_t} \widehat{Y}^i_{t+h|t}$, where h, the forecast horizon, is either 1 or 4) and use the first R forecasts to estimate the parameters of the regression of Y_{t+h} on an intercept and the equal-weighted forecasts. This projection is then used to generate out-of-sample forecasts for observation $R+h$. Denoting the full sample size as $T+h$, up to $P = T+1-R$ out-of-sample forecasts can be generated in this way. We use either recursive estimation—where the estimation window expands so the first window has R observations and the last window $R+P-1$ observations—or rolling window estimation where the length of the estimation window remains fixed at R observations. In both cases we set $R = 30$, so that we end up with 77 out of sample forecasts for $h = 1$ and 74 for $h = 4$.

For the other combination methods we keep forecasters with a minimum of 10 contiguous observations. Finally, we estimate the combination weights on the largest common sample.[9] We also calculated an equal-weighted combination and the projection on the prediction of this group of forecasters.

Before presenting the combination results it is useful to consider the performance of the best individual forecasters. Using one and four-step-ahead forecasts, Figure 2 presents time-series and histograms of the identification numbers of the survey participants with the best historical forecasting record at a given point in time. While the plots reveal some persistence in the identity of the best forecaster, there is considerable turnover among the forecasters at the top, indicating the potential for successfully combining forecast from this data set.

Figure 3 shows time-series plots of the four-step-ahead forecast errors, calculated using real-time data and based on the equal-weighted, rolling window projection, and unrestricted least squares (GR1) approaches. Due to their use of a common target (actual) value, there is a substantial common component in the series. Even so, it is clear that the differences

[9] We also did the application allowing at least 20 contiguous observations, with the rule that if there are no forecasters with at least 20 contiguous observations or if there are fewer observations than parameters (arising in about a third of the cases), we simply use the equal-weighted average forecast (this rule is not necessary when we restrict the application to at least 10 contiguous observations). The results are qualitatively the same as in the case reported here.

between the three sets of forecasts are economically large, suggesting that it makes a material difference which of the approaches is used for forecasting. One of the features that also comes across from these figures is the advantage of bias-adjusting the simple average.

Empirical results in the form of (pseudo real time) root mean squared error values for the various combination methods and estimation procedures (recursive or rolling) for one-step-ahead and four-steps-ahead forecasts are presented in Table 3. The table presents separate results for revised and real-time data.

The PEW combinations generate the smallest RMSE-values in the great majority of cases. However, this cannot be conclusive evidence in favor of PEW combinations since sample variability in forecasting performance needs to be taken into account, as indicated by Granger and Newbold (1986), Meese and Rogoff (1988), and Diebold and Mariano (1995). To find out if the differential performance of the various forecasts is significantly different from zero, we calculate three sets of $p-$values. First, we apply the Diebold and Mariano (1995) test to the MSE differences (PEW vis-a-vis alternative models) using:

$$\widehat{v}_{t+h|t} = \left(Y_{t+h} - \widehat{PEW\_roll}_{t+h|t}\right)^2 - \left(Y_{t+h} - \left(\widehat{AM}_{t+h|t}\right)\right)^2. \tag{18}$$

Here $\widehat{PEW\_roll}$ represents the forecasts calculated using the PEW method under a rolling estimation window and $\widehat{AM}$ stands for the alternative model under consideration. Under the null that the expected difference in MSE-values equals zero, $E\left[v_{t+h|t}\right] = 0$, a standard application of the central limit theorem yields $P^{\frac{1}{2}}\overline{v} \longrightarrow N\left(0, S\right)$, where $\overline{v} = P^{-1}\sum_{t=R}^{T+1}\widehat{v}_{t+h|t}$, and S is the spectral density of $\overline{v}$ at frequency zero scaled by 2π.[10] This approach does not account for recursive parameter updating. Because of this−and because we do not know how the individual forecasts were generated by the survey participants−we also adopt the forecast evaluation approach recently proposed by Giacomini and White (2006). Their null hypothesis, $E\left[\widehat{v}_{t+h|t}|I_t\right] = 0$, is different from the one used by Diebold and Mariano in two

[10] We estimate S using Newey and West's (1987) autocorrelation and heteroskedasticity consistent variance estimator.

aspects. First, losses depend on parameter estimates, rather than on their probability limits. And second, the expectation is conditional on the information set I_t. Giacomini and White show that under their null, $P\overline{Z}'\left(S^g\right)^{-1}\overline{Z} \longrightarrow \chi_q^2$, where $\overline{Z} = P^{-1}\sum_{t=R}^{T+1} h_t\widehat{v}_{t+h|t}$, and h_t is a $q \times 1$ test function, with $h_t \subset I_t$. To implement this test, we use a constant and the lagged difference as instruments, and Newey and West's procedure to estimate S^g.[11]

The second and third panels of Table 3 report the associated p-values of these tests applied to our data. First consider the results with the revised data. At the one-quarter forecast horizon, the Diebold-Mariano $p-$values suggest that the performance of the PEW method is significantly better than the Granger-Ramanathan approaches, one of the shrinkage methods, and the odds ratio approach. It is also significantly better than the previous best forecast. However, it is not possible to distinguish statistically between the PEW forecasts and those produced by the equal-weighted, one of the shrinkage approaches as well as the inverse MSE method. This could be due to the shortness of our evaluation sample. Similar results are produced by the Giacomini-White $p-$values and when four-step-ahead forecasts rather than one-step-ahead forecasts are considered.

However, the results change substantially when real-time data is used for the actual values. In this case, despite continuing to produce the smallest RMSE-value, the PEW forecasts are only significantly better than the GR1 and GR2 forecasts when four-steps-ahead forecasts are considered. These results suggest that while the PEW forecasts produce better forecasts independently of whether revised or real-time data is used to measure the 'actual' value, the evidence in support of this approach is strongest when revised data is used as the forecast target.

The third and final set of $p-$values is based on the test statistic recently proposed by Clark and West (2006). In effect this approach allows us to decompose the previous test statistic and obtain a sharper comparison between the equal-weighted forecast and the PEW

[11] All the combination methods that we use satisfy Giacomini and White's limited memory requirement except PEW_rec. The requirement is satisfied because for EW we only use the forecasts available that period, for PEW_roll we use a fixed estimation window, and for the other methods we use, each period, the largest common sample among the forecasters that have a minimum of 10 contiguous forecasts.

method. When nested models are compared, uncertainty introduced by the estimation of additional parameters under the alternative and more general model (the projection method in our case compared to the simple null that $\alpha = 0, \beta = 1$) must be accounted for. Under the null of identical forecasting performance of the two approaches, the sample difference of the mean MSE-values is therefore not zero but negative. To account for this, Clark and West suggest the following adjustment to the test statistic:

$$\widehat{v}^a_{t+h|t} = \left(Y_{t+h} - \overline{Y}_{t+h|t}\right)^2 - \left(Y_{t+h} - \left(\widehat{\alpha}_\tau + \widehat{\beta}_\tau \overline{Y}_{t+h|t}\right)\right)^2 + \left(\overline{Y}_{t+h} - \left(\widehat{\alpha}_\tau + \widehat{\beta}_\tau \overline{Y}_{t+h|t}\right)\right)^2. \quad (19)$$

Clark and West establish conditions under which the distribution of $\overline{v}^a = P^{-1} \sum_{t=R}^{T+1} \widehat{v}^a_{t+h|t}$ is given by $P^{\frac{1}{2}} \overline{v}^a \longrightarrow N(0, S^a)$. Again we use Newey and West's procedure to estimate S^a. P-values for the null that the difference in MSE-values (equal–weighted vis-a-vis PEW) equals zero are shown in the final panel of Table 3. The first two p-values compare PEW recursive and PEW rolling with equal weights, respectively, whereas the third p-value compares PEW using only forecasters with at least 10 contiguous forecasts with equal weights applied to the same subset of forecasters. In all cases the null gets rejected at the 5% level. The projection on the mean forecast is therefore better than the equal-weighted forecast irrespective of the method used to estimate the parameters of the projection and independently of how inflation is measured.

To better understand the performance of the PEW method, it is helpful to study the evolution over time in the estimates of the projection parameters α and β. To this end, Figure 4 provides a plot of these estimates using rolling windows of 30 observations, revised data and one and four-steps-ahead forecasts. The estimates of α and β display persistent deviations from zero and one, respectively. Such persistent deviations are consistent with and help explain the gains from using the projection approach and from using rolling window methods to estimate α and β.

6 Conclusion

Successful schemes for real-time combination of expert forecasts achieve a favorable trade-off between the bias induced by using sub-optimal forecast combination weights and the effect of parameter estimation error arising from the use of estimated combination weights. This trade-off is key to the real-time performance of different combination approaches and is the explanation for our finding that the entry and exit of experts from surveys of professional forecasters has such a large effect on the merit of the different approaches. Essentially, the unbalanced panel structure of survey data means that the real-time performance of combination methods that require estimating the full covariance between the experts' forecasts deteriorates relative to that of more robust methods such as equal-weighting or using posterior odds. It also explains the good overall performance of the new approach proposed here of projecting the outcome variable on a constant and the equal-weighted forecast. This approach uses information in the full set of individual forecasts (incorporated into the equal-weighted average) but then adjusts for possible bias and noise in this aggregate forecast.

Rather than using a simple linear projection of the outcome on a constant and the equal-weighted forecast, $\bar{Y}_{t+1,t}$, more flexible functions can of course be adopted if the sample size permits their estimation. For example, one could use $\hat{Y}_{t+1|t} = f(\bar{Y}_{t+1|t};\theta)$, where $f(\cdot)$ is given by a neural net or sieve estimator with parameters θ. The median or a trimmed mean could also be used instead of the mean forecast in situations where "extreme" forecasts would otherwise lead to poor performance.

References

[1] Amato, J.D. and Swanson, N.R. (2001), "The Real Time Predictive Content of Money for Output," *Journal of Monetary Economics,* 48, 3-24.

[2] Chan, Y.L., Stock, J.H. and Watson, M.W. (1999), "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review*, 1, 91-121.

[3] Clarida, R., Galí, J. and Gertler, M. (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics,* 115, 147-180.

[4] Clark, T.E. and West, K.D. (2006), "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics,* 135, 155-186.

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

[6] Croushore, D. (1993), "Introducing: The Survey of Professional Forecasters," *Federal Reserve Bank of Philadelphia Business Review,* November/December, 3-13.

[7] Croushore, D. and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics,* 105, 111-130.

[8] Davies, A. and Lahiri, K. (1995), "A New Framework for Analyzing Three-Dimensional Panel Data," *Journal of Econometrics,* 68, 205-227.

[9] Diebold, F.X. and Mariano, R. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* 13, 253-265.

[10] Elliott, G. and Timmermann, A. (2005), "Optimal Forecast Combination Weights Under Regime Switching," *International Economic Review,* 46, 4, 1081-1102.

[11] Giacomini, R. and White, H. (2006). "Tests of Conditional Predictive Ability," *Econometrica,* 74, 6, 1545-1578.

[12] Granger, C.W.J. and Newbold, P. (1986), *Forecasting Economic Time Series* (2nd ed.), New York:Academic Press.

[13] Granger, C.W.J. and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts," *Journal of Forecasting,* 3, 197-204.

[14] Gupta, S. and Wilton, P.C. (1987), "Combination of Forecasts: An Extension," *Management Science,* 33, 356-372.

[15] Gupta, S. and Wilton, P.C. (1988), "Combination of Economic Forecasts: An Odds-Matrix Approach," *Journal of Business and Economic Statistics,* 6, 373-379.

[16] Inoue, A. and Kilian, L. (forthcoming), "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation," *Journal of the American Statistical Association,* forthcoming.

[17] Jacobson, T. and Karlsson, S. (2004), "Finding Good Predictors of Inflation: A Bayesian Model Averaging Approach," *Journal of Forecasting,* 23, 479-496.

[18] Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science,* 32, 683-695.

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

[20] Meese, R.A. and Rogoff, K. (1988), "Was it Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating-Rate Period," *Journal of Finance,* 43, 933-948.

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

[22] Pesaran, M.H. and Timmermann, A. (2005), "Real Time Econometrics," *Econometric Theory*, 21, 212-231.

[23] Stock, J.H. and Watson, M.W. (2001), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in eds. R.F. Engle and H. White, *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive Granger*, USA:Oxford University Press.

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

[25] Stock, J.H. and Watson, M.W. (2006), "Forecasting with Many Predictors," in eds. G. Elliott, C.W.J. Granger and A. Timmermann, *Handbook of Economic Forecasting*, Elsevier Press.

[26] Timmermann, A. (2006), "Forecast Combinations," in eds. G. Elliott, C.W.J. Granger and A. Timmermann, *Handbook of Economic Forecasting*, Elsevier Press.

[27] Zarnowitz, V. (1985), "Rational Expectations and Macroeconomic Forecasts," *Journal of Business and Economic Statistics,* 3, 293-311.

Table 1: Simulation results from forecast combinations under factor structure

# of Forecasts	Sample Size	EW	PEW	GR1	GR2	GR3	Shrink 1	Shrink 2	Odds	Previous Best	Inverse MSE
Experiment 1 : Equal weights summing to one											
4	100	1.000	1.015	1.052	1.046	1.037	1.045	1.043	0.993	1.538	0.988
4	500	1.000	1.004	1.010	1.008	1.006	1.008	1.008	0.998	1.653	0.997
4	1000	1.000	1.002	1.002	1.001	1.000	1.001	1.001	0.999	1.664	0.998
10	100	1.000	1.021	1.133	1.123	1.110	1.116	1.098	0.986	2.620	0.979
10	500	1.000	1.004	1.020	1.017	1.015	1.017	1.016	0.997	2.943	0.994
10	1000	1.000	1.002	1.012	1.011	1.010	1.011	1.011	0.999	3.027	0.997
20	100	1.000	1.020	1.253	1.236	1.222	1.206	1.129	0.981	4.357	0.971
20	500	1.000	1.006	1.040	1.037	1.034	1.036	1.034	0.996	4.970	0.993
20	1000	1.000	1.002	1.021	1.019	1.019	1.019	1.019	0.998	5.293	0.997
Experiment 2: Equal weights											
4	100	1.000	0.873	0.905	0.899	1.036	0.899	0.897	0.995	1.126	0.994
4	500	1.000	0.862	0.867	0.866	1.006	0.866	0.866	0.999	1.179	0.999
4	1000	1.000	0.864	0.864	0.864	1.000	0.864	0.864	0.999	1.184	0.999
10	100	1.000	0.785	0.868	0.859	1.108	0.854	0.843	0.993	1.111	0.992
10	500	1.000	0.769	0.780	0.779	1.016	0.779	0.779	0.998	1.204	0.998
10	1000	1.000	0.773	0.781	0.781	1.009	0.781	0.781	0.999	1.235	0.999
20	100	1.000	0.735	0.907	0.893	1.229	0.872	0.832	0.991	1.082	0.990
20	500	1.000	0.715	0.740	0.737	1.034	0.737	0.736	0.998	1.196	0.998
20	1000	1.000	0.721	0.735	0.735	1.022	0.735	0.734	0.999	1.240	0.999
Experiment 3: Factor Dynamics											
4	100	1.000	0.730	0.757	0.752	1.032	0.752	0.751	0.993	1.279	0.989
4	500	1.000	0.722	0.726	0.724	1.004	0.724	0.724	0.998	1.369	0.998
4	1000	1.000	0.719	0.720	0.721	1.003	0.721	0.721	0.999	1.431	0.999
10	100	1.000	0.513	0.567	0.563	1.103	0.561	0.557	0.990	1.340	0.986
10	500	1.000	0.503	0.511	0.510	1.021	0.510	0.510	0.998	1.488	0.997
10	1000	1.000	0.494	0.500	0.500	1.010	0.500	0.500	0.999	1.512	0.999
20	100	1.000	0.406	0.498	0.491	1.221	0.481	0.486	0.988	1.300	0.984
20	500	1.000	0.405	0.418	0.417	1.032	0.417	0.417	0.997	1.494	0.997
20	1000	1.000	0.403	0.410	0.410	1.019	0.410	0.409	0.999	1.572	0.998
Experiment 4: Weak heterogeneity											
4	100	1.000	0.870	0.886	0.877	0.980	0.876	0.874	0.983	1.029	0.982
4	500	1.000	0.877	0.876	0.874	0.976	0.873	0.873	0.990	1.105	0.990
4	1000	1.000	0.853	0.845	0.845	0.958	0.845	0.845	0.989	1.087	0.989
10	100	1.000	0.766	0.830	0.821	0.980	0.817	0.808	0.979	1.013	0.978
10	500	1.000	0.765	0.775	0.773	0.920	0.773	0.772	0.986	1.076	0.986
10	1000	1.000	0.755	0.757	0.755	0.906	0.755	0.755	0.987	1.070	0.986
20	100	1.000	0.739	0.916	0.903	1.041	0.881	0.838	0.978	1.002	0.977
20	500	1.000	0.725	0.745	0.743	0.870	0.743	0.742	0.985	1.038	0.984
20	1000	1.000	0.714	0.722	0.722	0.853	0.722	0.721	0.986	1.055	0.985

Table 1: Simulation results (continuation)

# of Forecasts	Sample Size	EW	PEW	GR1	GR2	GR3	Shrink 1	Shrink 2	Odds	Previous Best	Inverse MSE
Experiment 5: Strong heterogeneity											
4	100	1.000	0.861	0.851	0.843	0.898	0.842	0.841	0.959	0.939	0.955
4	500	1.000	0.860	0.828	0.826	0.887	0.826	0.826	0.966	0.962	0.963
4	1000	1.000	0.840	0.815	0.814	0.882	0.814	0.814	0.966	0.972	0.963
10	100	1.000	0.762	0.816	0.809	0.858	0.804	0.794	0.949	0.881	0.945
10	500	1.000	0.743	0.721	0.719	0.767	0.719	0.719	0.953	0.877	0.949
10	1000	1.000	0.748	0.724	0.724	0.762	0.724	0.723	0.953	0.882	0.949
20	100	1.000	0.734	0.866	0.851	0.871	0.832	0.799	0.943	0.809	0.939
20	500	1.000	0.706	0.710	0.708	0.733	0.708	0.707	0.950	0.813	0.945
20	1000	1.000	0.705	0.700	0.700	0.725	0.700	0.700	0.951	0.828	0.946
Experiment 6: Block-diagonal factor structure											
4	100	1.000	0.778	0.807	0.802	1.033	0.802	0.800	0.995	1.060	0.995
4	500	1.000	0.765	0.769	0.768	1.007	0.768	0.768	0.999	1.104	0.999
4	1000	1.000	0.760	0.761	0.760	1.002	0.760	0.760	0.999	1.114	1.000
10	100	1.000	0.642	0.708	0.701	1.105	0.697	0.690	0.994	1.032	0.995
10	500	1.000	0.629	0.639	0.638	1.018	0.637	0.637	0.999	1.108	0.999
10	1000	1.000	0.631	0.637	0.637	1.008	0.637	0.637	1.000	1.122	0.999
20	100	1.000	0.566	0.697	0.687	1.233	0.671	0.654	0.993	1.008	0.995
20	500	1.000	0.550	0.567	0.565	1.032	0.565	0.565	0.998	1.082	0.999
20	1000	1.000	0.554	0.566	0.565	1.023	0.565	0.565	1.000	1.143	1.000
Experiment 7: Bias in individual forecasts											
4	100	1.000	0.841	0.872	0.924	1.019	0.923	0.922	0.991	1.123	0.989
4	500	1.000	0.830	0.835	0.901	0.996	0.901	0.901	0.995	1.161	0.995
4	1000	1.000	0.830	0.830	0.890	0.987	0.890	0.890	0.995	1.159	0.995
10	100	1.000	0.756	0.836	0.892	1.087	0.887	0.875	0.988	1.095	0.987
10	500	1.000	0.737	0.749	0.799	0.991	0.799	0.799	0.994	1.172	0.993
10	1000	1.000	0.741	0.749	0.807	0.987	0.807	0.807	0.995	1.192	0.995
20	100	1.000	0.706	0.871	0.909	1.191	0.887	0.844	0.986	1.084	0.985
20	500	1.000	0.690	0.714	0.750	1.008	0.749	0.749	0.994	1.185	0.994
20	1000	1.000	0.691	0.705	0.742	0.990	0.742	0.742	0.995	1.210	0.994

Notes: Results are based on 10,000 simulations. EW: equal-weighted forecast, PEW: projection of actual value on an intercept and EW forecast, GR1: unconstrained OLS, GR2: OLS w/o constant, GR3: OLS w/o constant and weights constrained to add to unity, Shrink1: shrinkage with $\kappa=0.25$, Shrink2: shrinkage with $\kappa=1$, Odds: Odds ratio approach, Previous Best: forecast from previous best model, Inverse MSE: weight equals the inverse of the historical MSE.

Table 2: Simulation results from forecast combinations under factor structure with survey-like data

# of Forecasts	Sample Size	EW	PEW	GR1	GR2	GR3	Shrink 1	Shrink 2	Odds	Previous Best	Inverse MSE
Experiment 1 : Equal weights summing to one											
20	100	1.000	1.000	1.040	1.030	1.520	1.030	1.030	1.520	1.540	1.000
20	500	1.000	0.986	1.040	1.030	1.510	1.030	1.030	1.510	1.520	1.000
20	1000	1.000	0.990	1.030	1.020	1.530	1.020	1.020	1.530	1.550	1.000
Experiment 2: Equal weights											
20	100	1.000	0.583	0.988	0.982	0.979	0.982	0.981	0.978	0.979	1.000
20	500	1.000	0.555	0.994	0.988	0.984	0.988	0.987	0.983	0.984	1.000
20	1000	1.000	0.575	0.995	0.989	0.984	0.989	0.988	0.984	0.985	1.000
Experiment 3: Factor Dynamics											
20	100	1.000	0.355	0.973	0.976	0.975	0.976	0.976	0.975	0.976	1.000
20	500	1.000	0.337	0.966	0.973	0.974	0.973	0.973	0.974	0.975	1.000
20	1000	1.000	0.338	0.962	0.971	0.969	0.971	0.970	0.969	0.969	1.000
Experiment 4: Weak heterogeneity											
20	100	1.000	0.577	0.990	0.985	0.980	0.984	0.984	0.980	0.980	1.000
20	500	1.000	0.551	0.984	0.980	0.976	0.980	0.980	0.977	0.977	1.000
20	1000	1.000	0.561	0.999	0.991	0.985	0.991	0.991	0.985	0.985	0.998
Experiment 5: Strong heterogeneity											
20	100	1.000	0.563	0.987	0.981	0.976	0.981	0.981	0.976	0.976	1.000
20	500	1.000	0.552	0.988	0.983	0.980	0.983	0.983	0.981	0.981	1.000
20	1000	1.000	0.557	0.988	0.983	0.977	0.983	0.982	0.977	0.977	1.000
Experiment 6: Block-diagonal factor structure											
20	100	1.000	0.761	1.000	0.998	0.993	0.997	0.997	0.992	0.992	1.000
20	500	1.000	0.739	1.010	0.999	0.993	0.999	0.998	0.993	0.993	1.000
20	1000	1.000	0.754	1.000	0.998	0.994	0.998	0.998	0.994	0.994	0.998
Experiment 7: Bias in individual forecasts											
20	100	1.000	0.586	0.987	0.992	0.990	0.992	0.992	0.990	0.991	1.000
20	500	1.000	0.564	0.994	0.998	0.997	0.998	0.998	0.996	0.997	0.999
20	1000	1.000	0.579	0.995	0.997	0.995	0.997	0.997	0.995	0.995	0.998

Notes: Results are based on 10,000 simulations. The minimum number of contiguous observations used by the least squares and shrinkage combinations is 10. EW: equal-weighted forecast, PEW: projection of actual value on an intercept and EW forecast, GR1: unconstrained OLS, GR2: OLS w/o constant, GR3: OLS w/o constant and weights constrained to add to unity, Shrink1: shrinkage with $\kappa=0.25$, Shrink2: shrinkage with $\kappa=1$, Odds: Odds ratio approach, Previous Best: forecast from previous best model, Inverse MSE: inverse of historical MSE when available or equal weights.

Table 3: Empirical application to inflation forecasts from the Survey of Professional Forecasters. [1/]

	1-Step-Ahead		4-Steps-Ahead	
	Revised Data	Real-Time Data	Revised Data	Real-Time Data
RMSE				
EW	0.88	0.90	1.01	1.00
EWc	1.00	0.98	1.15	1.15
PEW, Recursive	0.90	0.90	1.00	0.98
PEW, Rolling	0.79	0.88	0.83	0.93
PEW c	0.90	0.97	0.78	1.00
GR1	2.28	8.00	1.08	1.38
GR2	1.21	1.76	1.12	1.09
GR3	1.12	1.04	1.02	0.98
Shrink 1	1.24	2.84	1.85	1.03
Shrink 2	2.60	6.91	5.50	1.50
Odds	0.89	0.89	1.07	1.09
Previous Best	0.94	0.88	1.05	1.09
Inverse of MSE	0.87	0.90	1.01	0.98
P-Values Diebold-Mariano Test [2/]				
EW	0.16	0.67	0.11	0.50
EWc	0.01 **	0.15	0.01 **	0.10 *
PEW, Recursive	0.02 **	0.43	0.06 *	0.29
PEWc	0.22	0.23	0.40	0.49
GR1	0.09 *	0.27	0.11	0.06 *
GR2	0.00 ***	0.15	0.04 **	0.10 *
GR3	0.03 **	0.16	0.05 **	0.58
Shrink 1	0.02 **	0.29	0.19	0.28
Shrink 2	0.16	0.29	0.22	0.23
Odds	0.10 *	0.81	0.05 **	0.18
Previous Best	0.06 *	0.92	0.03 **	0.10 *
Inverse of MSE	0.21	0.75	0.09 *	0.66
P-Values Giacomini-White Test [3/]				
EW	0.15	0.28	0.00 ***	0.01 ***
EWc	0.02 **	0.08 *	0.00 ***	0.03 **
PEWc	0.01 **	0.26	0.24	0.13
GR1	0.10 *	0.22	0.25	0.04 **
GR2	0.00 ***	0.28	0.16	0.30
GR3	0.05 **	0.10 *	0.00 ***	0.56
Shrink 1	0.07 *	0.39	0.31	0.59
Shrink 2	0.25	0.40	0.28	0.17
Odds	0.06 *	0.22	0.00 ***	0.11
Previous Best	0.04 **	0.24	0.00 ***	0.25
Inverse of MSE	0.17	0.32	0.00 ***	0.45
P-Values Clark-West Test [4/]				
PEW Recursive	0.01 ***	0.05 **	0.00 ***	0.00 ***
PEW Rolling	0.00 ***	0.02 **	0.00 ***	0.00 ***
PEWc	0.00 ***	0.03 **	0.00 ***	0.03 **

* p<0.10. ** p<0.05. *** p<0.01.

1/ The minimum number of contiguous observations required is 10, except for EW, PEW Recursive, and PEW Rolling, where no restriction was imposed. For PEW Rolling a fixed window with 30 observations was used. The number of out-of-sample forecasts equals 77 for 1-step-ahead and 74 for 4-steps-ahead.

2/ Computed with respect to PEW Rolling.

3/ Computed with respect to PEW Rolling. Test is conditional on the (first/fourth) lag of the difference of the losses.

4/ The first two comparisons are with respect to EW, the third one is with respect to EWc.

Figure 1: Participants in the Survey of Professional Forecasters (inflation forecasts)

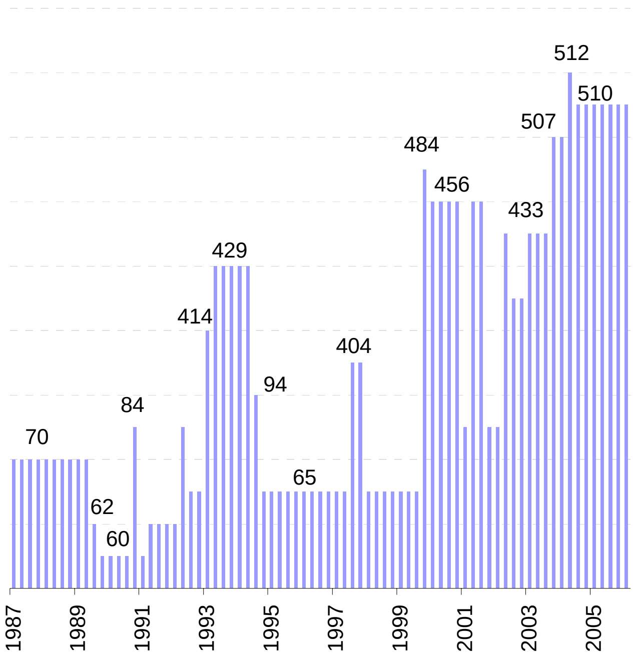
ID	95.1	95.2	95.3	95.4	96.1	96.2	96.3	96.4	97.1	97.2	97.3	97.4	98.1	98.2	98.3	98.4	99.1	99.2	99.3	99.4
20	x	x	x	x	x		x		x	x		x	x		x	x		x	x	x
40	x	x	x	x	x					x						x	x	x	x	x
65	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
84	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
94		x	x	x	x										x		x	x	x	
99	x	x		x		x		x		x		x		x		x	x			x
404	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x		x	
405	x	x		x		x	x										x			x
407		x	x	x	x	x	x	x	x		x		x	x		x	x	x	x	
409	x	x	x	x	x															
410																				
411	x	x	x	x			x	x	x	x	x	x	x	x	x	x			x	
414	x	x	x			x	x	x	x	x	x	x	x	x	x		x	x	x	
416	x	x																		
421	x	x	x	x		x	x	x	x	x	x	x		x	x	x	x	x		x
423		x	x	x		x					x								x	
426	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	
428	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
429	x	x	x	x											x	x	x	x		x
431	x		x	x	x	x	x	x	x			x	x	x		x	x	x		x
439	x		x	x	x	x	x	x	x	x	x	x		x	x		x	x		x
442	x	x	x	x	x	x			x			x	x							
443			x				x	x												
446	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x
448	x	x	x	x																
450	x	x	x					x	x											
451	x	x	x	x		x	x				x									
452	x	x	x	x	x			x		x	x		x	x	x	x				
455	x	x	x	x	x	x														
456	x	x	x	x	x	x	x		x	x		x	x	x	x	x	x	x	x	x
457		x	x	x																
458		x	x		x		x	x	x	x	x	x	x		x	x	x	x		x
460							x													
462		x	x	x		x	x	x		x	x	x	x	x		x	x			
463		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
464		x	x	x		x	x		x		x	x	x			x	x	x	x	
465				x	x	x	x	x	x	x	x	x	x		x		x	x	x	x
466		x	x			x			x	x		x			x	x	x		x	x
467					x															
468			x																	
469		x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
470						x				x	x	x	x		x	x		x	x	x
472		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
473		x																		
474			x					x	x		x									
475		x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
476		x				x														
479			x																	
481		x	x	x	x				x	x	x	x		x	x	x	x			x
483		x	x				x			x	x	x		x	x		x			x
485		x	x	x		x	x	x	x	x	x	x	x	x		x				
486			x	x			x	x		x		x		x						
487		x	x	x																
488		x	x	x		x	x	x	x	x	x	x	x		x		x		x	x
489		x																		
490				x			x	x	x	x	x									
491							x	x												
494												x								
495														x	x					
496															x	x				
497															x	x	x		x	
498																x	x	x	x	x
499																		x	x	x
500																			x	x
501																		x	x	x
502																		x	x	x
504																		x	x	x
505																		x	x	x

Notes: The ID corresponds to the identification number assigned to each forecaster in the survey. The columns represent the quarter when the survey was taken. The Xs show when a particular forecaster responded to the inflation part of the survey and provided a one-step-ahead forecast.

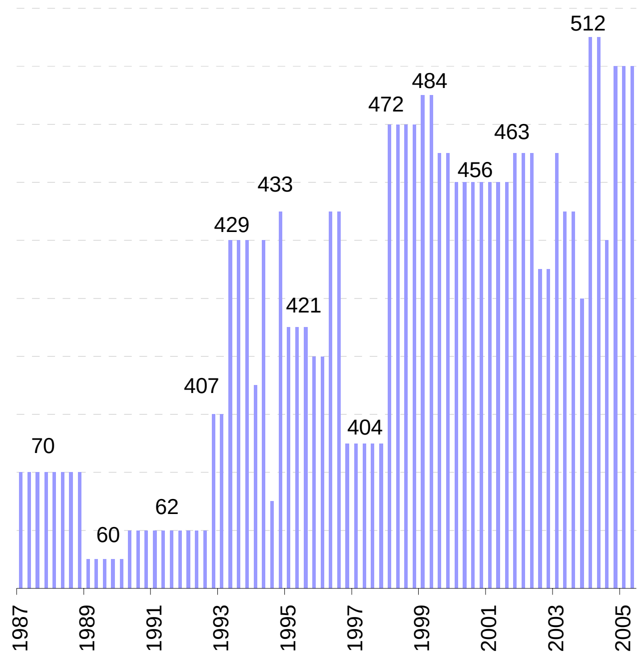
Figure 2: Previous Best Forecaster

a) Time-Series

1 step-ahead Revised Data Forecasts

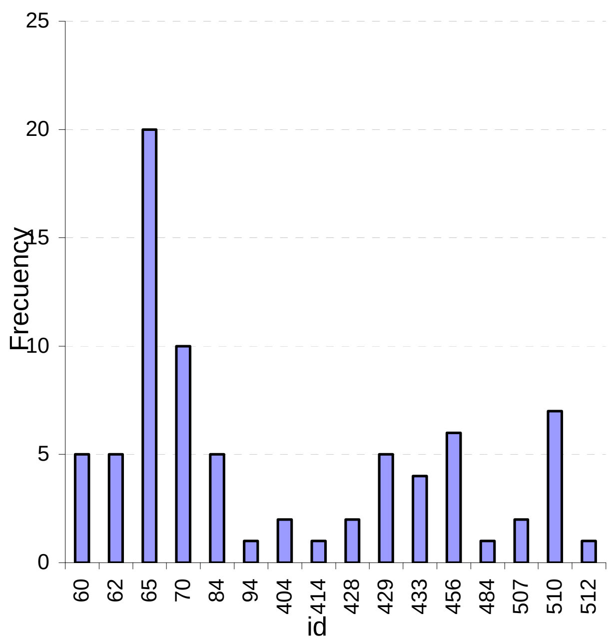


4 step-ahead Revised Data Forecasts



b) Histogram

1 step-ahead Revised Data Forecasts



4 step-ahead Revised Data Forecasts

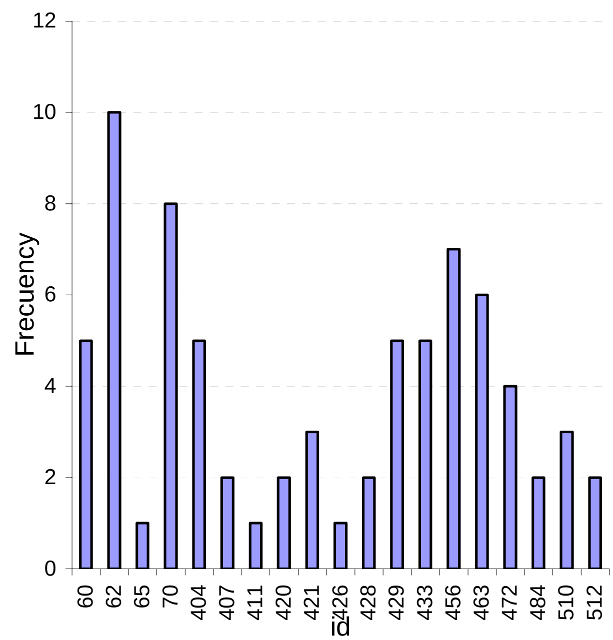


Figure 3: Forecast Errors, Four-Steps-Ahead, Real-Time Data

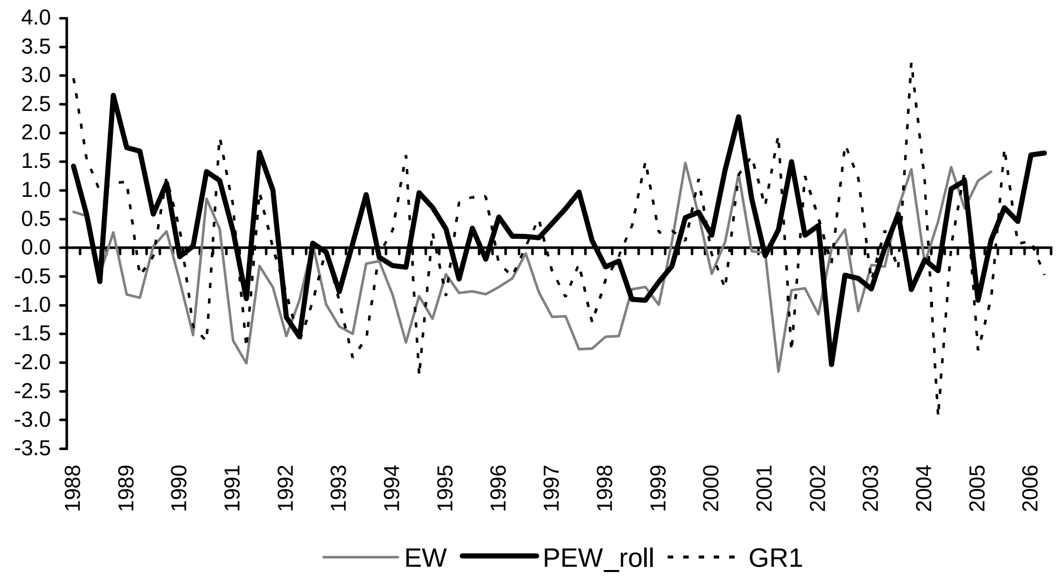


Figure 4: Estimated Parameters of PEW Rolling