An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach

Michael P. Clements
Fred Joutz
Herman O. Stekler

July 26, 2006

**Note: A revised version of this paper is now published in the
Journal of Applied Econometrics 22(1), 121-136.**

An Evaluation of the Forecasts of the Federal Reserve:

A Pooled Approach

Michael P. Clements[*]

Department of Economics

University of Warwick

Coventry, CV4 7AL

Fred Joutz

Department of Economics

George Washington University

Washington DC

Herman O. Stekler

Department of Economics

George Washington University

Washington DC

July 26, 2006

Abstract

The Federal Reserve Greenbook forecasts of real GDP, inflation and unemployment are analysed for the period 1974 to 1997. We consider whether these forecasts exhibit systematic bias, and whether efficient use is made of information, that is, whether revisions to these forecasts over time are predictable. Rather than analyse the forecasts separately for each horizon of interest, we discuss and implement methods that pool information over horizons. We conclude that there is evidence of systematic bias and of forecast smoothing of the inflation forecasts.

JEL classification numbers: C12, C53.

Keywords: Federal reserve forecasts, bias, smoothing, pooling.

1 Introduction

In setting monetary policy, decision makers pay close attention to forecasts about the future state of the economy. These forecasts are one of the many inputs that decision makers use in determining whether changes in monetary policy are required. This is one of the reasons why many studies have analyzed the forecasts made by the staff of the Board of Governors of the Federal Reserve in preparation for the meetings of the Open Market Committee. These forecasts are published in the Greenbook, the data and analytic notebook that is prepared for the Open Market Committee prior to each meeting. These meetings occur several times each quarter.

The staff predict GDP, its components, various price indices, unemployment, etc. for one to eight quarters ahead. A number of studies have examined the accuracy and rationality of these forecasts, focusing primarily on the predictions of real and nominal GDP growth rates, inflation rates and unemployment. For example, Karamouzis and Lombra (1989) concluded that the 1973-82 forecasts displayed large errors and contained unexploited information. Scotese (1994) found no significant biases in either the real GNP or inflation forecasts, but she noted that the forecast revisions had been smoothed. The studies by Jansen and Kishnan (1996, 1999) and Joutz and Stekler (2000) all showed that the inflation forecasts were unbiased but that there was a question whether the real GDP predictions were also unbiased. The Romer and Romer (2000) and Sims (2002) studies were primarily concerned with the accuracy of the Fed's inflation forecasts. While the rationality of those forecasts was not the primary concern of those studies, both papers found no significant biases in those forecasts. While there may be slight differences in some of the results, the overall conclusion of these studies is that these forecasts are unbiased. Moreover, all these papers were based on the Mincer-Zarnowitz (1969) procedure.

The approaches used in these studies compared the forecasts and outcomes *separately for each horizon.* A new procedure developed by Davies and Lahiri (1995, 1999), following on from Keane and Runkle (1990), permits us to pool the forecasts for each variable across all horizons.[1] This has a number of advantages. It is natural to consider all the forecasts together and ask whether they are biased, rather than asking the question separately of each individual horizon.

---

[1] See the review article by Pesaran and Weale (2006) for background on recent developments in testing for rationality of expectations using survey data.

A finding that j-step forecasts are biased, while the $j + 1$ step forecasts are not biased, for example, is difficult to interpret. Moreover, if we test a number of sets of forecasts then the chances of rejecting the null for at least one set (horizon) will exceed the nominal test level. Utilising all the forecasts in a single test should also make for a more powerful test of the null of unbiasedness, where the forecasts of rational agents (with squared-error loss functions[2]) should be unbiased at all horizons.

Techniques for testing for the efficient use of information were discussed in the context of fixed-event forecasts by Nordhaus (1987) and Clements (1995, 1997). Nordhaus (1987) examined the forecasts for a single target date using a large number of forecasts of that target made at varying times (and so with different horizons). Clements (1997) considered pooling over a small number of targets. We adapt these ideas to our set up, where we have a large number of targets but relatively few forecasts of each, and also calculate pooled versions of these tests. We show that if we assume the forecast-error decomposition of Davies and Lahiri (1995), the test based on the lack of correlation on adjacent forecast revisions of Nordhaus is invalid. Instead, we present versions which are valid under the Davies-Lahiri setup.

In section 2 we briefly review the standard approaches for testing for bias, where each horizon is analyzed separately, and then outline the pooling procedure. Section 3 sets out the approach we adopt for analyzing forecast revisions. This enables us to examine the forecast smoothing issue raised by Scotese. Namely that forecasters are averse to making large and/or frequent changes to their subsequent forecasts for the same target. Section 4 explains the structure of the forecasts at our disposal, that is, the nature of the FED forecast data set, as well as the results of our assessment. Section 5 presents our conclusions.

2 Testing for forecast bias

Before we present the approach to pooling forecasts made with various leads (i.e., over different horizons), we discuss the tests that have customarily been used in testing for bias and rationality.

---

[2]See Zellner (1986) for a discussion of this caveat, and Artis and Marcellino (2001) for an illustration. For comparability with the earlier studies, we assume symmetric loss functions.

2.1 Traditional procedure

The conventional test for bias is based on the regression:

$$A_t = \alpha + \beta F_{th} + \epsilon_t \tag{1}$$

where A_t and F_{th} are the actual and predicted values for time t. The prediction is made at $t - h$, and the error term is allowed to be an MA($h - 1$) process. The test involves the joint null hypothesis that $\alpha = 0$ and $\beta = 1$. A rejection of this null would indicate that the forecasts are biased: see also Mincer and Zarnowitz (1969). Whilst the joint null entails unbiasedness, $E(A_t - F_{th}) = 0$, Holden and Peel (1990) note that the null is a sufficient, but not a necessary, condition for unbiasedness, since unbiasedness holds for:

$$\alpha = (1 - \beta) F_{th}. \tag{2}$$

A more satisfactory test of unbiasedness is via a test of $\tau = 0$ in the regression:

$$A_t - F_{th} = \tau + \nu_t. \tag{3}$$

The important point is that either (1) or (3) would typically be run, and the relevant null test statistic computed, separately for each h, $h = 1, 2, 3 \ldots$. More powerful tests may be obtained by pooling, especially when the available sample of forecasts for any given h is relatively small.

2.2 Pooling procedure

In this section we outline the procedures we adopt to test for bias based on pooling the forecasts. Pooling across horizons requires that we posit a model of the forecast errors that enables us to infer the correlation structure across errors of different targets and lengths that is consistent with rationality. To this end, we will adopt the decomposition of forecast errors of Davies and Lahiri (1995), except that their forecasts vary across individuals as well as targets and horizons. For a single individual, with targets $t = 1, \ldots, T$, and horizons, $h = 1, \ldots, H$, we have:

$$A_t - F_{th} = \alpha + \lambda_{th} + \varepsilon_{th}. \tag{4}$$

4

The ε_{th} are the 'idiosyncratic shocks', and the λ_{th} are the aggregate or common macroeconomic shocks. λ_{th} is the sum of all shocks that occurred between $t - h$ and t, labelled as u_{th} to u_{t1}:

$$\lambda_{th} = \sum_{j=1}^{h} u_{tj}. \tag{5}$$

As we discuss below, the distinguishing feature of these two shocks (or surprises) is that $var\,(\varepsilon_{th})$ is constant over h (and t), because it is an error specific to the forecast of t being made at time $t - h$. By way of contrast, $var\,(\lambda_{th})$ decreases as the horizon h shortens, as λ_{th} cumulates the shocks between $t - h$ and t.

The forecast error is defined as:

$$
\begin{aligned}
e_{th} &= A_t - F_{th} = \alpha + \lambda_{th} + \varepsilon_{th} \\
&= \alpha + v_{th}.
\end{aligned}
\tag{6}
$$

Stacking the forecasts in the TH dimension vector $\mathbf{F}$ as

$$\mathbf{F}' = (F_{1H}, \ldots, F_{11}, F_{2H}, \ldots, F_{21}, \ldots, F_{TH}, \ldots, F_{T1})$$

and letting $\mathbf{A}^* = (A_1, A_2, \ldots, A_T)'$, then defining $\mathbf{A} = \mathbf{A}^* \otimes \mathbf{i}_H$ we can write the TH vector of forecast errors as $\mathbf{e} = \mathbf{A} - \mathbf{F}$.[3] Then in matrix notation (6) can be written as:

$$\mathbf{e} = \mathbf{i}_{TH}\alpha + \mathbf{v} \tag{7}$$

where $\mathbf{v}$ is the TH vector from stacking v_{th} conformably with $\mathbf{A}$ and $\mathbf{F}$, and $\mathbf{i}_{TH}$ is the unit vector of dimension $T \times H$. From (7) it is clear that the bias α is restricted to be the same across all horizons. This assumption can be relaxed by instead specifying the regression equation as:

$$\mathbf{e} = (\mathbf{i}_T \otimes \mathbf{I}_H)\boldsymbol{\alpha}_H + \mathbf{v} \tag{8}$$

---

[3] Note the Kronecker product is defined such that a typical block $\mathbf{C} \otimes \mathbf{D}$ is given by $C_{ij}\mathbf{D}$, and $\mathbf{i}_s$ is an s dimensional vector of 1's.

where the parameter vector $\boldsymbol{\alpha}_H = (\alpha_H, \ldots, \alpha_1)'$ allows a separate bias for each horizon, and the matrix of explanatory variables $(\mathbf{i}_T \otimes \mathbf{I}_H)$ is TH by H, where $\mathbf{I}_H$ is the identity matrix of order H. Tests of bias based on (8) are F-tests that $\boldsymbol{\alpha}_H = \mathbf{0}$. The model with the null hypothesis imposed is the same as in (7) but the alternative hypotheses differ. If there were systematic differences in the ways in which the forecasts of different horizons were calculated, this could result in them having different biases. Consequently, the assumption of a common bias may erroneously lead to a failure to reject the null $H_0 : \alpha = 0$ in (7). Note that testing based on (8) is not equivalent to the standard approach of testing the forecast errors for each horizon using separate regressions. Davies and Lahiri (1995) adopt a common bias over horizons but allow individual-specific biases. We begin with horizon-specific biases, and test the assumption of a common bias over horizons.

The distinction between macro shocks (λ_{th}) and idiosyncratic errors (ε_{th}) is perhaps questionable when there is a single forecaster, as in our analysis of the FED. Clearly, when there is 'private information' (see Davies and Lahiri (1995, footnote 13)), the distinction is lost. Private information leads to to the idiosyncratic error term ε_{th} in (4) being replaced by $\eta_{th} = \sum_{j=1}^{h} \varepsilon_{tj}$, with $\lambda_{th} = \sum_{j=1}^{h} u_{tj}$ unchanged, such that as h gets smaller the variance of the private component, $Var(\eta_{th})$, declines. In our case, if we think of the FED as possessing confidential ('private') information, then although conceptually distinct from the macro shocks, it cannot be separately distinguished when it takes the form $\eta_{th} = \sum_{j=1}^{h} \varepsilon_{tj}$ (for a horizon h), as

$$A_t - F_{th} \;\; = \;\; \alpha + (\lambda_{th} + \eta_{th}) \tag{9}$$

$$= \;\; \alpha + v_{th} \tag{10}$$

where

$$v_{th} = \sum_{j=1}^{h} (u_{tj} + \varepsilon_{tj}) = \sum_{j=1}^{h} u_{tj}^{*}$$

say. One can motivate the formulation with idiosyncratic terms by thinking of these not so much as shocks as particular or idiosyncratic errors made by the forecaster which can include inefficient use of information. Thus, the ε_{th} represent forecast errors at a particular horizon (or state of the business cycle) that the forecaster is prone to make. These may be due to forecast model misspecifications and/or the forecaster's own interventions.

6

The importance of the $\{\varepsilon_{th}\}$ entering as summations ($\eta_{th} = \sum_{j=1}^{h} \varepsilon_{tj}$) in the model for the forecast errors $A_t - F_{th}$ is that the variance of the (common) shock $\{v_{th}\}$ is directly proportional to h. So the practical upshot is that the stance we take on the idiosyncratic component will affect the correlation structures of the forecast errors and may therefore affect the outcomes of the tests of the FED forecasts. Rather than taking a view on which assumption is appropriate, we will assess the robustness of our results to these alternative assumptions about the model of the forecast error process. As a shorthand we will sometimes use $E\left(\varepsilon_{th}^2\right) \equiv \sigma_\varepsilon^2 = 0$ to refer to the case where the idiosyncratic component is absent, in which case σ_u^2 becomes the variance of the $\left\{u_{tj}^*\right\}$. Otherwise, when $\sigma_\varepsilon^2 \neq 0$, σ_u^2 implicitly refers to $var\left(u_{tj}\right)$.

Based on the above, we can assume that $E\left(\mathbf{v}\right) = \mathbf{0}$, but $\mathbf{\Sigma} = E\left(\mathbf{vv}'\right)$ will not be proportional to the identity matrix. Allowing for idiosyncratic errors, and assuming that the ε_{th} are not correlated, $E\left(\varepsilon_{th}\varepsilon_{sj}\right) = \sigma_\varepsilon^2$ when $s = t$ and $j = h$, and zero otherwise, we obtain:

$$\mathbf{\Sigma} = \sigma_\varepsilon^2 \mathbf{I}_{TH} + \mathbf{\Psi}. \tag{11}$$

$\mathbf{\Psi}$ captures the fact that forecasts which have common (macroeconomic) shocks, be they of the same target, or of different targets, will be correlated. The precise form of $\mathbf{\Psi}$ depends on the maximum and minimum forecast lengths, etc., and will be described below. But the important point is that $\mathbf{\Psi}$ depends only on the single unknown variance parameter, $\sigma_u^2 = E\left(u_t^2\right)$, assuming homoscedastic aggregate shocks. The assumption of homoscedastic macro shocks enables us to construct the relevant covariance matrices in sections 3 and 4 in a fairly straightforward way, but may be questionable if the U.S. economy has become less volatile since the early to mid 1980s, as argued by some (see Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)). We find that the root mean square forecast error (RMSFE) of the current quarter forecast errors before and after 1985 is not markedly different, suggesting that the assumption of homoscedastic shocks may not be unreasonable.

The test of bias is based on the OLS estimator of α in (7),

$$\widehat{\alpha} = (TH)^{-1} \sum_{t=1}^{T} \sum_{h=1}^{H} e_{th} \tag{12}$$

with the consistent covariance matrix estimator

$$Var\left(\widehat{\alpha}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = (TH)^{-2}\,\mathbf{i}'_{TH}\mathbf{\Sigma}\mathbf{i}_{TH} \tag{13}$$

where $\mathbf{X} = \mathbf{i}_{TH}$. In the case of (8), we estimate a vector $\widehat{\boldsymbol{\alpha}}_H$, and the covariance matrix is as (13) but with $\mathbf{X} = \mathbf{i}_T \otimes \mathbf{I}_H$. To obtain a consistent standard error for OLS, $\mathbf{\Sigma}$ needs to be replaced by an estimate $\widehat{\mathbf{\Sigma}}$. To calculate $\widehat{\mathbf{\Sigma}}$ we require estimates of σ_u^2 and σ_ε^2. These can be obtained as follows. $\widehat{\alpha}$ is estimated from (12). Then $\widehat{v}_{th} = e_{th} - \widehat{\alpha}$ is the calculated deviation of each forecast error from the estimate of the bias. The variance of this term is a combination of the idiosyncratic shocks at each forecast date and the common macro shocks over time. Using $E\left(v_{th}^2\right) = \sigma_\varepsilon^2 + h\sigma_u^2$, we can obtain estimates $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\sigma}_u^2$ as the estimated coefficients $\widehat{\varphi}_0$ and $\widehat{\varphi}_1$ in the regression:

$$\widehat{\mathbf{v}} \odot \widehat{\mathbf{v}} = \varphi_0 \mathbf{i}_{TH} + \varphi_1 \boldsymbol{\tau} + \boldsymbol{\omega} \tag{14}$$

(where $\odot$ is the Hadamard product, denoting element-by-element multiplication) $\boldsymbol{\tau} = \mathbf{i}_T \otimes \boldsymbol{\tau}_H$, and $\boldsymbol{\tau}'_H = (H, H-1, \ldots, 1)$. Thus $\widehat{\varphi}_0$ is an estimate of σ_ε^2, the variance of idiosyncratic shocks, and $\widehat{\varphi}_1$ estimates σ_u^2, the variance of the (homoscedastic) macro shocks. When $\sigma_\varepsilon^2 = 0$, an estimate of σ_{u*}^2 is given by the coefficient φ_1^* in the regression:

$$\widehat{\mathbf{v}} \odot \widehat{\mathbf{v}} = \varphi_1^* \boldsymbol{\tau} + \boldsymbol{\omega}^*. \tag{15}$$

3 Econometric analysis of forecast revisions

After considering forecast bias, a natural question is to ask whether forecast errors are predictable from information available at the time the forecast was made. A closely related question is whether forecast revisions are predictable. Forecast errors are predictable if in the regression:

$$e_{th} = \delta x_{t,h+1} + \phi + \left(\lambda_{th} + \varepsilon_{th}\right) \tag{16}$$

we reject the null hypothesis that $\delta = 0$, where $x_{t,h+1}$ is any variable known at the time the forecasts are made (it is the value of x in period $t - h - 1$). An analysis of the predictability of forecast revisions results from taking the difference between adjacent forecasts errors of the

same target, i.e.:

$$e_{t,h+1} - e_{th} = f_{th} - f_{t,h+1} = \delta \left(x_{t,h+2} - x_{t,h+1} \right) + \left(u_{t,h+1} + \varepsilon_{t,h+1} - \varepsilon_{th} \right). \tag{17}$$

This is equivalent to the forecast revision between $h + 1$ and h. If the forecasts are rational, the revisions should not be predictable, i.e., $\delta = 0$. The advantage of testing revisions based on (17) rather than errors (based on (16)) is that we avoid the possible problem of the regressor and disturbance being correlated in (16) (see Davies and Lahiri (1995, p.217), *inter alia*), and we can easily test whether the FED forecast revisions have been "smoothed"; an issue first addressed by Scotese (1994).

Define the revision to the forecast of t between the h and $h - 1$ step forecasts as $r_{t;h-1,h} \equiv f_{t,h-1} - f_{t,h}$. The revision $r_{t;h-1,h}$ should be unpredictable given all the information known at the time $f_{t,h}$ is made. Otherwise it would be known in advance that $f_{t,h-1}$ would be revised relative to $f_{t,h}$ in a way that is in part predictable. Since 'all information' is unbounded, this hypothesis can be made testable in a simple way by considering only the information set consisting of revisions to past targets $\{r_{t-1;h-1,h}, r_{t-2;h-1,h}, \ldots\}$, and of past revisions for that given target $\{r_{t;h-1+i,h+i}\}$, $i = 1, 2, \ldots$. To keep the analysis manageable, we consider tests based on these two information sets separately. To do so, we first derive the correlation structures implied by the model of forecast errors discussed in section 2.2.

3.1 The information set consisting of revisions to past targets

From $r_{t;h-1,h} \equiv f_{t,h-1} - f_{t,h} = e_{t,h} - e_{t,h-1}$, and recalling that $e_{t,h} = \alpha_h + \varepsilon_{t,h} + \sum_{j=1}^{h} u_{tj}$, we have:

$$r_{t;h-1,h} = -\varepsilon_{t,h-1} + \varepsilon_{t,h} + u_{t-h} + \alpha_h - \alpha_{h-1}.$$

It follows that $Cov \left(r_{t;h-1,h}, r_{t-s;h-1,h} \right) = 0$ for all $s > 0$. Thus, rational forecast revisions of adjacent targets will be uncorrelated whether or not there are idiosyncratic shocks. This suggests a test of $\gamma = 0$ in:

$$r_{t;h-1,h} = \gamma r_{t-1;h-1,h} + \delta + \omega_t, \quad t = 2, \ldots, T, \tag{18}$$

9

where δ is an intercept.

Under the null, $\omega_t = r_{t;h-1,h} - \delta$, and the covariance matrix, $Cov\left(\boldsymbol{\omega}\boldsymbol{\omega}'\right) = \left(2\sigma_\varepsilon^2 + \sigma_u^2\right)\mathbf{I}_{T-1}$, where $\boldsymbol{\omega}$ is a vector of the $T-1$ disturbances, ω_t. We have this diagonal form because $Cov\left(\omega_t\omega_{t-s}\right) = 0$ for $s \neq 0$ and $var\left(\omega_t\right) = 2\sigma_\varepsilon^2 + \sigma_u^2$ for all t. Inference can be conducted using OLS. When there are no idiosyncratic shocks, $\sigma_\varepsilon^2 = 0$, the σ_u^2 are re-defined as σ_{u*}^2 to be the variance of the macro and private information shocks, but this has no effect on the analysis of the revisions. The covariance matrix of the disturbances remains proportional to the identity matrix. Our main focus is on revisions between forecasts of adjacent quarters.[4]

As well as estimating (18) separately for $h = 2, 3, \ldots,$ we could also pool. By letting $\mathbf{r}_{1,2} = \left(r_{2;1,2} \ \ldots \ r_{T;1,21}\right)'$, $\mathbf{r}_{2,3} = \left(r_{2;2,3} \ \ldots \ r_{T;2,32}\right)'$ etc. we can stack the adjacent revisions as:

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_{1,2} \\ \mathbf{r}_{2,3} \\ \vdots \\ \mathbf{r}_{H-1,H} \end{bmatrix}$$

where $H = 5$ for our analysis. Then consider the balanced pooled regression:

$$\mathbf{r} = \gamma\mathbf{r}_{-1} + \boldsymbol{\delta}_d + \boldsymbol{\omega}$$

where $\mathbf{r}_{-1} = \left(r_{1;1,2} \ \ldots \ r_{T-1;1,2} \ r_{1;2,3} \ \ldots \ r_{T-1;2,3} \ \ldots\ldots r_{1;H-1,H} \ \ldots \ r_{T-1;H-1,H}\right)$, and both $\mathbf{r}$ and $\mathbf{r}_{-1}$ are $(H-1)\times(T-1)$ vectors. Here, $\boldsymbol{\delta}_d = \left(\mathbf{I}_{H-1} \otimes \mathbf{i}_{T-1}\right)\boldsymbol{\delta}$, and $\boldsymbol{\delta} = \left(\delta_1 \ \delta_2 \ \ldots \ \delta_{H-1}\right)'$, so that $\boldsymbol{\delta}_d$ is a set of $H-1$ horizon dummies with parameter vector $\boldsymbol{\delta}$. This allows for horizon-specific biases, such that the means of the revisions between different horizons may differ.

For $H = 5$, the covariance matrix of $\boldsymbol{\omega}$ has the form under the null:

$$E\left(\mathbf{rr}'\right) = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} \\ \mathbf{B}' & \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B}' & \mathbf{A} & \mathbf{B} \\ \mathbf{D}' & \mathbf{C}' & \mathbf{B}' & \mathbf{A} \end{bmatrix}$$

---

[4]If the revisions are defined between 'distant' forecasts, then one can deduce that revisions between t and $t-s$, $s > 1$ will be correlated because of the macro shocks. For example, $Cov\left(r_{t;j,i}, r_{t-2;j,i}\right) \neq 0$ when $i > j+2$.

where $\mathbf{A} = \left(2\sigma_\varepsilon^2 + \sigma_u^2\right)\mathbf{I}_{T-1}$, and we can deduce that[5]:

$$\mathbf{B} = \begin{bmatrix} -\sigma_\varepsilon^2 & \sigma_u^2 & 0 & \ldots & & 0 \\ 0 & -\sigma_\varepsilon^2 & \sigma_u^2 & 0 & & \vdots \\ \vdots & 0 & -\sigma_\varepsilon^2 & & \ddots & \\ & & 0 & -\sigma_\varepsilon^2 & \ddots & \\ & & & \ddots & \ddots & \sigma_u^2 \\ 0 & \ldots & & & 0 & -\sigma_\varepsilon^2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & \sigma_u^2 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_u^2 & \ddots & \vdots \\ \vdots & 0 & 0 & 0 & \ddots & 0 \\ & & 0 & 0 & \ddots & \sigma_u^2 \\ & & & \ddots & \ddots & 0 \\ 0 & \ldots & & & 0 & 0 \end{bmatrix}.$$

$\mathbf{D}$ is as $\mathbf{C}$ except that the $(1,4)$, $(2,5)$, $(3,6)\ldots$ terms are equal to σ_u^2 (rather than the $(1,3)$, $(2,4)$, $(3,5)\ldots$ elements). Testing is by OLS with corrected standard errors.

3.2 The information set consisting of past revisions

In the previous section we considered whether the revision to the forecast of (say) 1974 Q4 made between 1974 Q4 and 1974 Q3 is predictable from the revision to the forecast of 1974 Q3 made between 1974 Q3 and 1974 Q2. In addition, in this section we consider whether (say) the revision to the forecast of 1974 Q4 between 1974 Q4 and 1974 Q3 is predictable from the revision to the forecast of 1974 Q4 between 1974 Q3 and 1974 Q2. So the null hypothesis is now that the revision is not predictable from the previous revision for the same target.

It is straightforward to establish that $Cov\left(r_{t;h-1,h}, r_{t;h,h+1}\right) = -\sigma_\varepsilon^2 \neq 0$, so we would not expect $\gamma = 0$ in:

$$r_{t;h-1,h} = \gamma r_{t;h,h+1} + \delta + \omega_t, \quad t = 1,\ldots,T, \tag{19}$$

unless $\sigma_\varepsilon^2 = 0$. Since we do not wish to impose this restriction, instead we consider:

$$r_{t;h-1,h} = \gamma r_{t;h+1,h+2} + \delta + \omega_t, \quad t = 1,\ldots,T, \tag{20}$$

Thus, we consider whether the revision to the forecast of (say) 1974 Q4 between 1974 Q4 and 1974 Q3 is predictable from the revision to the forecast of 1974 Q4 between 1974 Q2 and 1974

---

[5]In section 4.1.2 we give an example of how a similar covariance matrix is constructed.

Q1. Tests of revisions to the same target have been stressed by authors such as Nordhaus (1987).

Given $H = 5$, we test for the dependence of $r_{t;1,2}$ on $r_{t;3,4}$ and $r_{t;2,3}$ on $r_{t;2,5}$ in two ways: separately and by pooling. We again focus on first-order dependence. By letting $\mathbf{r}_{1,2} = (r_{1;1,2} \ \ldots \ r_{T;1,2})'$ and $\mathbf{r}_{2,3} = (r_{1;2,3} \ \ldots \ r_{T;2,3})'$, we stack the revisions as $\mathbf{r} = \left(\mathbf{r}'_{1,2} \ \mathbf{r}'_{2,3}\right)'$, where in the pooled regression:

$$\mathbf{r} = \gamma \mathbf{r}_{-1} + \boldsymbol{\delta}_d + \boldsymbol{\omega} \tag{21}$$

we now have $\mathbf{r}_{-1} = (r_{1;2,4} \ \ldots \ r_{T;3,4}, r_{1;4,5} \ \ldots \ r_{T;4,5})$. Both $\mathbf{r}$ and $\mathbf{r}_{-1}$ are $2 \times T$ vectors. In this regression, $\boldsymbol{\delta}_d = \left(\mathbf{I}_2 \otimes \mathbf{i}_{T-1}\right) \boldsymbol{\delta}$, and $\boldsymbol{\delta} = (\delta_1 \ \delta_2)'$, to again allow for horizon specific biases.

The covariance matrix of $\boldsymbol{\omega}$ has the form under the null:

$$E\left(\mathbf{rr'}\right) = \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{A} \end{bmatrix}$$

where $\mathbf{A} = \left(2\sigma_\varepsilon^2 + \sigma_u^2\right) \mathbf{I}_T$, and $\mathbf{C}$ is defined in the previous sub-section. An estimate of σ_u^2 can be obtained from the OLS residuals from the estimation of (21).

4   Data set and results

Our FED forecast set contains a number of forecasts of three variables: real GDP, inflation as measured by the GDP deflator, and unemployment. These forecasts were made for each quarter from 1965.4 to 1997.4, yielding 129 target dates. The maximum lengths of the forecasts differ by target date. For example, for the early target dates, there were only 'current quarter' and one-quarter ahead forecasts. By the 1970's, the FED staff was making forecasts four or more quarters ahead. Consequently, the number of observations is not identical for each horizon. There is more than one forecast made during many of the quarters for a given 'target' – we take the latest forecasts made in the quarter as the forecasts for that quarter. We follow Joutz and Stekler (2000) in treating the forecasts made during the first few days of a quarter as belonging to the previous quarter, because little information would have accrued. Joutz and Stekler (2000) analyse the relationships between the forecasts made at various times during the months. They find that forecasts made later in the quarter are more accurate for forecasts

of the current quarter, but not subsequent quarters. We consider only one forecast a quarter as inter-quarter forecast changes would be expected to be more important than intra-quarter changes. By taking the last forecast in each quarter, we ensure that as far as possible the forecasts are made at approximately the same time in each quarter.

For much of our analysis, we consider forecasts of the quarters 1974:2 through to 1997:4 (so $T = 95$) as we have current ($h = 1$) to 4-quarter ahead forecasts ($h = 5$). Note that a current-quarter forecast is acually a 1-step forecast, as the current-quarter value is unknown when the forecast is made, a next-quarter forecast is a 2-period forecast, and so on up to $h = 5$. Care need is required in distinguishing between step and quarter.

A possible complication is that the FED's forecasts are conditional on a counterfactual path for monetary policy (see, e.g., Reifschneider, Stockton and Wilcox (1997)), whereby a more or less constant path is often assumed. Because policy is generally thought to affect the economy with a lag, even though the FED conditions its forecasts on a counterfactual path it might be reasonable to assume that short horizon forecasts are essentially equivalent to unconditional forecasts, whilst longer-horizon forecasts would have a forecastable error component due to the difference between the actual and assumed policy rates. For example, Romer and Romer (2000, p.437) indicate that monetary policy has little impact on either real output or inflation for at least 3 or 4 quarters. Our maximum horizon is four quarters ahead. To the extent that policy has a significant effect on the macroeconomy within this timeframe, our longer horizon forecasts may exhibit systematic biases for this reason.

The actual values used for the construction of forecast errors are taken to be the 45-day release numbers from the Bureau of Economic Analysis (BEA), and the first published unemployment figures released by the Bureau of Labor Statistics (BLS). It is sometimes argued that forecasters seek to forecast earlier announcements of data rather than later revisions (see e.g., Keane and Runkle (1990)). This position appears to be consistent with the literature on the effects of data vintage on model specification, estimation, and forecast evaluation: see Croushore and Stark (2001, 2003) and Koenig, Dolmas and Piger (2003).

4.1 Forecast bias results

4.1.1 Biases by horizon, tested separately

We first test the 1 through 5-step ahead forecasts separately for bias, by regressing the forecast errors on a constant and testing whether the coefficient on the constant is significantly different from zero, as in standard analyses. In so doing we allow an $\mathrm{MA}(h-1)$ serially-correlated error process. For example, a 2-step forecast of 1998Q1 made in December 1997 will be made before the actual value for 1997Q4 is known, as the actual values are taken to be the 45-day release from the BEA. But 1-step (current quarter) forecasts are not serially correlated. For example, the actual data for 1997Q3 is known – having been published in November 1997 – before the current quarter forecast of 1997Q4 is made in December 1997.

In calculating autocorrelation-consistent standard errors we adopt uniform weights by horizon in the HAC standard errors of Newey and West (1987). The results of doing so are recorded in Table 1, where the calculations are based on the maximum number of forecasts available for each horizon (also recorded in the table). As a precursor to pooling over horizons, we also record the results of testing separately by horizon for a shorter period, 74.2 – 97.4. This matches the sample of forecasts available for the pooling exercise, and so allows a direct comparison of the individual horizon and pooled results on a common sample period.

The results suggest there is no evidence that either the inflation or the unemployment rate forecasts are biased at any horizon, while the 1-step real growth forecasts are clearly biased on the shorter, common sample. We reject the null at the 10% level for the whole sample. The under-prediction (positive bias) for the period as a whole is less pronounced than for the common sample, suggesting less pessimistic predictions of real growth during the 1965-74 period. Notice though that the longer horizon real GDP forecasts show no evidence of bias. Although the estimated biases for the longer horizon forecasts are quite large for unemployment and real growth, the autocorrelation-corrections to the standard errors result in insignificant t-values. The RMSEs are shown for the common sample, 1974-97, but differ little if calculated for the maximum number of available forecasts. The RMSEs are similar to the comparable figures in Joutz and Stekler (2000, Table 2). We also calculated RMSEs separately for the sub-samples 1974:2 to 1984:4 and 1985:1 to 1997:4. For inflation we obtained RMSEs of 1.11

and 0.731, compared to 0.92 for the whole period 1974-97; for unemployment 0.118 and 0.094, compared to 0.105; and for real growth, 1.37 and 1.12, compared to 1.24. The RMSEs for the more recent period are lower, but with the possible exception of inflation, are not markedly so. We conclude from this that the assumption of homoscedastic shocks is not unreasonable.

4.1.2 Pooling over horizons

As well as testing the forecasts separately at each step ahead, we can 'pool over horizons' as discussed in section 2. As noted, the nature of the forecast data is such that the maximum forecast horizons for the earlier period are short. For example, we only have 1 and 2-step ahead forecasts of the quarters 66:1 to 68:4 (i.e., forecasts made in 65:4 and 66:1 of 66:1, and forecasts made in 68:3 and 68:4 of 68:4). For some of the latter periods there are forecasts made 7 or 8 periods earlier. Selecting 74:2 as the first 'target date', we have $h = 1$ to 5-step forecasts ($H = 5$) for each of $T = 95$ quarters (74:2, 74:3 to 97:4). Given the large amount of overlap, for this set of forecasts, $\boldsymbol{\Psi}$ in (11) takes the form:

$$
\boldsymbol{\Psi} = \boldsymbol{\sigma}_u^2
\begin{bmatrix}
\mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} & \mathbf{0} & \cdots & & & \mathbf{0} \\
\mathbf{B}' & \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \ddots & \ddots & & & \\
\mathbf{C}' & \mathbf{B}' & \mathbf{A} & \mathbf{B} & \mathbf{C} & & \ddots & & & \\
\mathbf{D}' & \mathbf{C}' & \mathbf{B}' & \mathbf{A} & \mathbf{B} & & \ddots & & & \\
\mathbf{E}' & \mathbf{D}' & \mathbf{C}' & \mathbf{B}' & & \ddots & & & & \\
\mathbf{0} & & & \ddots & \ddots & & & & & \\
\vdots & \ddots & & & & & & & & \\
 & & & & & & & & & \mathbf{C} \\
 & & & & & & & & \mathbf{A} & \mathbf{B} \\
\mathbf{0} & & & & \mathbf{0} & & & \mathbf{D}' & \mathbf{C}' & \mathbf{B}' & \mathbf{A}
\end{bmatrix}
$$

where the component matrices are given by:

$$
\mathbf{A} = \begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 4 & 3 & 2 & 1 \\ 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad
\mathbf{B} = \begin{bmatrix} 4 & 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 & 0 \\ 3 & 3 & 2 & 1 & 0 \\ 2 & 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}, \quad
\mathbf{C} = \begin{bmatrix} 3 & 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}
$$

and:

$$
\mathbf{D} = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad
\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

The form of $\mathbf{\Psi}$ can be understood by example. Consider the $(1,6)$ element, say. This is the covariance between the macro shocks for a forecast of 74:2 made in 73:2, and a forecast of 74:3 made in 73:3 (i.e., two 5-step ahead forecasts). The relevant macro shocks for the former occur in {74:2,74:1,73:4,73:3 and 73:2}, and for the latter {74:3,74:2,74:1,73:4 and 73:3}. There are 4 shocks in common {74:2,74:1,73:4 and 73:3}. Hence the 4 as the $(1,1)$ element of $\mathbf{B}$. The reason for including the forecast-origin shock, 73:2 for the former, and 73:3 for the latter, is our assumption that the state of the economy is not known when the forecast is made due to lags in data collection etc. This logic leads to a single shock at 74:2 in a 1-step or 'current-quarter' forecast of 74:2 made at 74:2 (the $(5,5)$ element of $\mathbf{A}$).

In Table 1 we showed that the null of unbiasedness was only rejected for the $h = 1$ forecasts of real GDP growth. Table 2 contains the results from running regressions based on equation (7). The pooled forecasts are tested for bias when the maintained model assumes a common bias for all horizons. The table reports $\hat{\alpha}$, and the estimated standard error and p-value of $\alpha = 0$ for idiosyncratic shocks, and when such shocks are absent. The table also records the estimates of σ_ε^2 and σ_u^2 based on (14). For unemployment the estimate of σ_ε^2 was negative,[6] and

---

[6]OLS with non-negativity constraints could be used to enforce a non-negative estimate, but this would presumably be very similar to (14) with the intercept omitted, which is the regression equation (15).

the table records only the estimate of the standard error of $\hat{\alpha}$ obtained from (15). As shown in Table 2, omitting σ_ε^2 has virtually no effect on inference concerning α for either inflation or growth. Whether we assume an indiosyncratic error component is immaterial for inference on bias. The OLS pooled estimates suggest the forecasts of all three variables are unbiased.

Table 3 contains the results of testing when we allow the bias to vary across horizon. The results were obtained using equation (8). The table contains results when idiosyncratic shocks are either present or absent. If there are idiosyncratic shocks, we can accept the null hypothesis of a common bias at all horizons for inflation and real growth, but not for unemployment. However, if the null hypothesis is that the common bias is zero, then we reject the null for all three series. In the absence of idiosyncratic shocks, we reject both hypotheses: the null of a common bias and that it is equal to zero.

The results in Tables 2 and 3 indicate that whether we enforce a common bias across horizons in the maintained model is crucial. The forecasts for all three variables are biased when we jointly test that all the horizon-specific bias terms are zero. Given that we find significant biases, it is of interest to consider these a little more carefully. At the longer horizons, we find over-predictions (negative biases) for both unemployment and output growth, as well as for inflation for the 1974-97 period (except for $h = 5$). Although both unemployment and output growth are over-predicted, further investigation revealed that the forecasts errors are significantly negatively correlated at all steps other than $h = 1$, which is what one might expect to find. The inflation and output growth errors are also significantly negatively correlated. The correlations between the inflation and unemployment forecast errors are insignificant at all horizons.

4.2 Forecast revisions results

Table 4 records the results of testing the revisions separately, as in equation (18), for example. This is the more traditional approach to analysing forecast revisions. The first part of the table reports the results of tests for dependence in four revisions series; $r_{t;h,h+1}$, for $h = 1, \ldots, 4$. We test whether the revision to one target period helps predict the same horizon revision to the subsequent target period. There is reasonable evidence of potential smoothing of revisions for inflation at all horizons except when $h = 1$. There is evidence for inflation and real growth for

$h = 1$ if we consider one-sided alternatives. No further evidence of smoothing is found at the remaining horizons, and none at all for unemployment.

The bottom panel of Table 4 shows the results of testing for dependence in revisions to the same target. We find that the revisions between the 3 and 2-step forecasts of inflation are predictable from the change between the 5 and 4-step forecasts – γ is found to be significantly positive so that changes appear to be implemented gradually. This implies if upward revisions have been made, the revision 2 quarters for the same target is also likely to be positive. Otherwise there is no evidence of inefficiency. We suspect this reflects the systematic errors in forecasting inflation in the later half of the 1970s and first half of the 1980s. There were under-predictions of inflation in the 1970s, and the over-predictions in the 1980s, even as the Federal Reserve was attempting to establish low inflation credibility. In addition, there were persistent fears of inflation accelerating in the 1990s.

The results for the tests of first-order correlation in the stacked revisions (see section 3) are summarised in Table 5, where we again present two sets of results. The first two rows relate to regressing $r_{t;i,j}$ on $r_{t-1;i,j}$, the last two rows look at revisions for the same target t. The estimates of σ_ε^2 and σ_u^2 from testing for bias are used to calculate $E(\mathbf{rr}')$ to test $\gamma = 0$ in rows (1) and (3). Rows (2) and (4) assume $\sigma_\varepsilon^2 = 0$ and form $E(\mathbf{rr}')$ from σ_u^2. The evidence in the first part of the table suggests that revisions are predictable for inflation, and possibly for growth, but not for unemployment. The final rows report the pooled tests of whether revisions are predictable from earlier revisions of the same target. Our results suggest there is no smoothing in the forecast revisions of this sort. These are the more standard tests of forecast revisions akin to Nordhaus (1987), but as noted, we do not consider adjacent revisions.

Using the stacked and separate regression approaches, and the two different information sets, we find some evidence that the revisions in the inflation and real GDP forecasts were predictable. This indicates that the forecasts were not rational. In particular, changes were smoothed over the 1974.2-1997.4 sample period. These results complement those of Scotese (1994) who found some evidence of smoothing in the FED forecasts of real GNP for some periods. She had based her findings on a comparison of the FED forecasts with those of a benchmark, a BVAR. Our analysis does not require any benchmark and is based solely on an examination of the forecasts, their errors and revisions.

5 Conclusions

We have analyzed the FED Greenbook forecasts to determine whether they are unbiased, and rational, as measured by the predictability of the forecast revisions. The forecasts were analyzed in two ways: separately at each horizon, and pooled across horizons. Pooling has a number of advantages in terms of efficiency. Our results differ from those of the traditional approach that tests for forecast bias separately at each horizon. In our case the traditional approach does not reject unbiasedness. The results of our pooling approach (allowing horizon-specific biases) indicated that *the forecasts of all three variables are biased.*

We identify two forms of smoothing of forecast revisions. The first concerns revisions to the same target, as discussed by Nordhaus (1987). We found forecast smoothing of this sort for inflation based on the individual regressions. While the reasons why the forecasts are smoothed is beyond the scope of this paper, a conjecture that would apply to smoothing of revisions of forecasts of the same target is the following. Scotese had suggested that the FED forecasters wanted to maintain their "reputation" as well as be accurate. We can provide one possible reason why "reputation," interpreted to mean credibility, is important to the FED staff. Since FED policy cannot oscillate between stimulative and contractionary actions, the forecasts must be credible to form the basis for policy decisions. If there were large fluctuations in the forecasts from FOMC meeting to meeting, the forecasters would lose some of their credibility with the Board members and Bank Presidents. Consequently, when there is uncertainty and the staff receives data that tend to be volatile (and subject to revision), the staff may only partially revise its previous forecast. This avoids the problem of having to reverse the changes incorporated into the new forecast if later data partially reverse the earlier movements. If the later data corroborate the earlier numbers, the next revision will be in the same direction as the previous revision. This may explain why the forecasts have been smoothed and why the revisions in the FED forecasts have been predictable.[7]

There was stronger evidence of serial correlation in revisions across adjacent targets for inflation and real growth from the pooled approach. To the best of our knowledge, ours is the

---

[7]This explanation is also consistent with the finding in e.g., McNees (1990), that judgmental adjustments to model-based forecasts are found to be valuable, but that such adjustments are also made to make the forecasts appear "plausible" (see e.g., the review by Önkal-Atay, Thomson and Pollock (2002)).

first study to address this issue. Our findings imply that if the revision between the j and $j-1$ horizon forecasts of period t is positive, then the likelihood is that the revision between the j and $j-1$ horizon forecasts of period $t+1$ will also be positive. We leave further exploration of this topic to future research.

References

Artis M, Marcellino M. 2001. Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal*, **4**: S20–S36.

Clements MP. 1995. Rationality and the role of judgement in macroeconomic forecasting. *Economic Journal*, **105**: 410–420.

Clements MP. 1997. Evaluating the rationality of fixed-event forecasts. *Journal of Forecasting*, **16**: 225–239.

Croushore D, Stark T. 2001. A real-time data set for macroeconomists. *Journal of Econometrics*, **105**: 111–130.

Croushore D, Stark T. 2003. A real-time data set for macroeconomists: Does the data vintage matter?. *The Review of Economics and Statistics*, **85(3)**: 605–617.

Davies A, Lahiri K. 1995. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, **68**: 205–227.

Davies A, Lahiri K. 1999. Re-examining the rational expectations hypothesis using panel data on multi-period forecasts. In Hsiao C, Lahiri K, Lee LF., and Pesaran MH. (eds.), *Analysis of Panels and Limited Dependent Variable Models*, pp. 226–254: Cambridge University Press.

Holden K, Peel DA. 1990. On testing for unbiasedness and efficiency of forecasts. *Manchester School*, **58**: 120–127.

Jansen DW, Kishnan RP. 1996. An evaluation of Federal Reserve forecasting. *Journal of Macroeconomics*, **18**: 89–100.

Jansen DW, Kishnan RP. 1999. An evaluation of Federal Reserve forecasting: A reply. *Journal of Macroeconomics*, **21**: 189–203.

Joutz F, Stekler HO. 2000. An evaluation of the predictions of the Federal Reserve. *International Journal of Forecasting*, **16**: 17–38.

Karamouzis N, Lombra N. 1989. Federal Reserve policymaking: An overview and analysis of the policy process. *Carnegie Rochester Conference Series on Public Policy*, **30**, 7–62.

Keane MP, Runkle DE. 1990. Testing the rationality of price forecasts: new evidence from

panel data. *American Economic Review*, **80**: 714–735.

Kim CJ, Nelson CR. 1999. Has the US economy become more stable? a Bayesian approach based on a Markov-Switching model of the Business Cycle. *Review of Economics and Statistics*, **81**: 1–10.

Koenig EF, Dolmas S, Piger J. 2003. The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, **85**: 618–628.

McConnell MM, Perez-Quiros, GP. 2000. Output fluctuations in the United States: What has changed since the early 1980s?. *American Economic Review*, **90**: 1464–1476.

McNees SK. 1990. The accuracy of macroeconomic forecasts. In Klein PA. (ed.), *Analyzing Modern Business Cycles*, pp. 143–173. London: M. E. Sharpe Inc.

Mincer J, Zarnowitz V. 1969. The evaluation of economic forecasts. In Mincer J. (ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.

Newey WK, West KD. 1987. A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica*, **55**: 703–708.

Nordhaus WD. 1987. Forecasting efficiency: Concepts and applications. *Review of Economics and Statistics*, **69**: 667–674.

Önkal-Atay D, Thomson ME, Pollock AC. 2002. Judgmental forecasting. In Clements MP, and Hendry DF (eds.), *A Companion to Economic Forecasting*, pp. 133–151: Oxford: Blackwells.

Pesaran MH, Weale M. 2006. Survey expectations. In Elliott G, Granger CW, and Timmermann A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 715–776: Elsevier, Horth-Holland.

Reifschneider DL, Stockton DJ, Wilcox DW. 1997. Econometric models and the monetary policy process. *Carnegie-Rochester Conference Series on Public Policy*, **47**: 1–37.

Romer CD, Romer DH. 2000. Federal Reserve private information and the behaviour of interest rates. *American Economic Review*, **90**: 429–457.

Scotese CA. 1994. Forecast smoothing and the optimal underutilization of information at the Federal Reserve. *Journal of Macroeconomics*, **16**: 653–670.

Sims CA. 2002. The role of models and probabilities in the monetary policy process. *Brookings*

Papers on Economic Activity: 2, The Brookings Institute, 1–40.

Zellner A. 1986. Biased predictors, rationality and the evaluation of forecasts. *Economics Letters*, **21**: 45–48.

Table 1: Tests of the null of no bias at different horizons

	Using maximum no. of forecasts			74:2 to 97:4		
		Inflation				
h	bias	p-value	no.	bias	p-value	RMSFE
1	-0.017	0.586	129	-0.141	0.933	0.920
2	0.028	0.427	128	-0.189	0.899	1.22
3	0.136	0.291	115	-0.107	0.701	1.49
4	0.164	0.304	112	-0.050	0.573	1.69
5	0.087	0.405	103	0.007	0.492	1.87
		Real growth rates				
h	bias	p-value	no.	bias	p-value	RMSFE
0	0.179	0.045	129	0.295	0.009	1.24
1	-0.030	0.546	128	0.150	0.314	2.89
2	-0.203	0.704	115	-0.054	0.551	3.15
3	-0.359	0.822	112	-0.214	0.712	3.45
4	-0.226	0.743	103	-0.231	0.742	3.58
		Unemployment				
h	bias	p-value	no.	bias	p-value	RMSFE
0	-0.001	0.546	129	0.008	0.238	0.105
1	-0.059	0.974	128	-0.066	0.963	0.316
2	-0.087	0.893	115	-0.099	0.886	0.537
3	-0.096	0.812	112	-0.122	0.837	0.698
4	-0.114	0.788	103	-0.123	0.788	0.817

Notes to table: The test for bias regresses the forecast errors on a constant, and records the p-value (columns 3 and 6) of the t-statistic. The p-value is the probability of observing a value in excess of the test statistic if the null is true. The last column reports RMSFEs.

Table 2: Tests of the null of no bias; Pooled over horizons; Common bias assumed. The sample period is 74.2 – 97.4.

	Idiosyncratic shocks			No idiosyncratic shocks			
	$\hat{\alpha}$	se$(\hat{\alpha})$	p-value	se$(\hat{\alpha})$	p-value	σ_ε^2	σ_u^2
Inflation	-0.096	0.251	0.649	0.258	0.645	0.154	0.672
Growth	-0.012	0.496	0.509	0.521	0.508	1.07	2.61
Unemp.	-0.080	–	–	0.103	0.783	–	0.113

Table 3: Tests of the null of no bias; Pooled over horizons; Horizon-specific bias. The sample period is 74.2 – 97.4.

	Idiosyncratic shocks		No idiosyncratic shocks	
	$H_0 : \boldsymbol{\alpha}_H = \mathbf{i}_H\alpha$	$H_0 : \boldsymbol{\alpha}_H = \mathbf{0}$	$H_0 : \boldsymbol{\alpha}_H = \mathbf{i}_H\alpha$	$H_0 : \boldsymbol{\alpha}_H = \mathbf{0}$
Inflation	0.565	0.000	0.000	0.000
Growth	0.798	0.016	0.000	0.000
Unemp.	0.000	0.000	0.000	0.000

The entries in the the table are p-values of chi-squared statistics of the stated null. The estimates of the horizon-specific biases are identical to the those based on the individual regressions in table 1.

Table 4: Tests of the null that forecast revisions were not predictable from past revisions at same horizon and past revisions from the same target. The sample period is 74.2 – 97.4.

	Inflation			Real growth			Unemployment		
	$\hat{\gamma}$	se$(\hat{\gamma})$	p-value.	$\hat{\gamma}$	se$(\hat{\gamma})$	p-value.	$\hat{\gamma}$	se$(\hat{\gamma})$	p-value.
$r_{t;h,h+1}$ on explanatory variable $r_{t-1;h,h+1}$									
h									
1	0.133	0.103	0.100	0.148	0.103	0.077	-0.045	0.101	0.671
2	0.208	0.100	0.020	0.088	0.103	0.197	0.044	0.104	0.338
3	0.148	0.097	0.066	0.011	0.100	0.455	0.023	0.104	0.413
4	0.285	0.099	0.002	-0.124	0.103	0.883	0.034	0.104	0.374
$r_{t;h,h+1}$ on explanatory variable $r_{t;h+2,h+3}$									
h									
1	-0.101	0.159	0.736	0.024	0.214	0.456	-0.057	0.077	0.772
2	0.196	0.108	0.036	0.101	0.123	0.208	0.009	0.110	0.467

All inference based on OLS.

Table 5: Tests of null that forecast revisions were not predictable, stacked. The sample period is 74.2 – 97.4.

	Inflation			Real growth			Unemployment		
	$\hat{\gamma}$	se($\hat{\gamma}$)	p-value.	$\hat{\gamma}$	se($\hat{\gamma}$)	p-value.	$\hat{\gamma}$	se($\hat{\gamma}$)	p-value.
$r_{t;i,j}$ on explanatory variable $r_{t-1;i,j}$									
(1)	0.178	0.101	0.039	0.092	0.072	0.101	0.021	0.096	0.415
(2)	0.178	0.095	0.031	0.092	0.059	0.058	0.021	0.079	0.397
$r_{t;i,i+1}$ on explanatory variable $r_{t;i+2,i+3}$, for $i = 1, 2$									
(3)	0.039	0.140	0.390	0.056	0.143	0.349	-0.026	0.069	0.649
(4)	0.039	0.123	0.376	0.056	0.111	0.309	-0.026	0.057	0.678

All inference based on OLS. The first and third rows use the estimates of σ_u^2 and σ_ε^2 obtained when testing for bias. Rows 2 and 4 use the estimate for σ_u^2 when σ_ε^2 is set to zero when testing for bias.