Efficient forecast tests for conditional policy forecasts

Jon Faust
Johns Hopkins University
faustj@jhu.edu

Jonathan Wright
Federal Reserve Board
jonathan.h.wright@frb.gov

Efficient forecast tests for conditional policy forecasts

March 2006
(revised January 2008)

Abstract

Central Banks regularly make forecasts, such as the Fed's Greenbook forecast, that are conditioned on hypothetical paths for the policy interest rate. While there are good public policy reasons to evaluate the quality of such forecasts, up until now, the most common approach has been to ignore their conditional nature and apply standard forecast efficiency tests. In this paper we derive tests for the efficiency of conditional forecasts. Intuitively, these tests involve implicit estimates of the degree to which the conditioning path is counterfactual and the magnitude of the policy feedback over the forecast horizon. We apply the tests to the Greenbook forecast and the Bank of England's inflation report forecast, finding some evidence of forecast inefficiency. Nonetheless, we argue that the conditional nature of the forecasts made by central banks represents a substantial impediment to the econometric analysis of their quality—stronger assumptions are needed and forecast inefficiency may go undetected for longer than would be the case if central banks were instead to report unconditional forecasts.

Keywords: Forecasting, monetary policy, weak instruments.
JEL Classification: C53, E37, E5.

We thank Steven Durlauf, Chris Sims, Ellis Tallman and an anonymous referee for helpful comments and Jeff Traczinski for valuable research assistance.

NOTE: The view in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

1. Introduction

Forecasts have long played a prominent role in forming policy at central banks. Recently, forecasts have also become an important part of the way many Central Banks communicate with the public about policy. These facts provide many reasons to be interested in the quality of Central Bank (CB) forecasts. Poor forecasts could lead directly to errant policy. Further, from the standpoint of communicating with the public, the quality of the forecast is an important component of the signal-to-noise ratio for the communication scheme. Finally, the relative forecasting precision of the public and CB gives us a measure of the any information advantage of the CB, which is important in some theories of policy (e.g., Canzoneri, 1985).

The nature of common CB forecasts is a significant stumbling block to analysis of these forecasts. The Federal Reserve's Greenbook forecast and the published forecasts of several inflation targeting CBs are conditioned on a particular path for the policy interest rate over the forecast horizon. One standard case is *conditioning* on an unchanged path for the policy interest rate over the forecast horizon. This approach may be adopted because the CB staff do not wish explicitly to forecast the future decisions of policymakers, or because the CB does not wish to publish forecasts of future policy actions. Whatever the reason, the forecasts are conditioned on a path of interest rates that is counterfactual. That is, the path is not meant to be the CBs expectation of the policy rate. We will follow the literature in referring to these forecasts as *conditional* and contrasting them with *unconditional* forecasts, which are taken to be the CB's expectations for the variables in question, conditioned only on the CB's information at the time of the forecast.

There is a long history of evaluating the quality of the Greenbook forecast by simply treating it as an unconditional forecast and applying standard forecast efficiency tests. For example, Romer and Romer (2000) found that the Greenbook inflation forecast is efficient and superior to private sector forecasts. These same exercises are also being conducted on the published forecasts of other central banks (e.g., Andersson, et al. 2005; Bank of England, 2004). The merits of the results of this work are unclear, given that the conclusions rest on an explicit or implicit assumption that the effects of conditioning can be neglected.

The view that the conditional nature of the forecasts can be neglected could be justified by the twin assumptions that the conditioning paths are *not too far* from the CB's unconditional expectation for policy and/or that policy feedbacks are *not too large* over the relevant horizon. While these assumptions may be reasonable at very short forecast horizons, at longer horizons they become more tenuous.

Faust and Leeper (2005) argue that unconditional forecasts of the policy interest rate and goal variables are more effective communication devices for CBs than conditional forecasts of the type presented by central banks. This paper is instead methodological, and develops and applies a framework for testing the efficiency of these conditional forecasts. By positing what we argue to be a reasonable structure relating the conditional and unconditional forecasts, we show that an appropriate forecast efficiency test can be based on an augmented version of standard efficiency regressions. The augmenting terms implicitly take account of the degree to which the conditioning path is counterfactual and the magnitude of policy feedbacks. Thus, we replace the twin assumptions by implicit estimates of the two components.

In the simplest case, we make a CB transparency assumption that private sector and CB interest rate expectations coincide and that private sector expectations can be measured from rates on money market futures contracts. In this case, one can test forecast efficiency by running an OLS regression of the *ex post* conditional forecast error from a forecast at time t on variables known at t, and some augmenting variables. As in the standard case, the efficiency test checks whether the variables known at t are systematically related to the forecast error. The augmenting variables are present to soak up any predictability of the conditional forecast error that is due to the effects of conditioning. Without the transparency assumption, an analogous test can be run, but we must instrument for the augmenting variables.

We present results of these new forecast efficiency tests applied to the Fed's Greenbook forecast and the Bank of England's (BOE's) Inflation Report forecast. The results illustrate the importance of taking conditioning seriously. Under the transparency assumption, we find evidence against the efficiency of both forecasts. Notably, some rejections are stronger than one obtains from a naive test that ignores the conditional nature of the forecasts. Without the transparency assumption there is less evidence against forecast efficiency, especially when we take account of weak instrument issues. We nonetheless find some evidence against efficiency of the Greenbook GDP growth forecast for the current quarter or next quarter, and some weak evidence against the efficiency of the Greenbook inflation forecast at horizons of a few quarters. The practical magnitude of the rejections in these cases may be small.

From the standpoint of one looking to monitor and improve forecasting operations, however, this work has additional implications. The paper makes clear that it takes relatively strong assumptions to form a coherent test of efficiency of the conditional forecasts. Even under these assumptions, the power of the tests may be weak. Thus, rather substantial inefficiency might go unnoticed for a considerable period of time. These facts raise important questions about whether a forecasting operation and/or a communication policy should be focused primarily on conditional forecasts. We discuss these issues in the conclusion.

2. Conceptual Issues and Required Assumptions

The basic setup is that we have a time series of *conditional* forecasts produced by the CBs at various points in time. We would like to deduce whether the CB has used information efficiently in producing the conditional forecast. We do not have the CB's *unconditional* forecast, which the CB may not write down at all, let alone publish. In any case, we refer to an unconditional forecast of the CB, by which we mean its expectation of the variables in question based on its information at the time of the forecast, but not based on any counterfactual values of variables in the forecast period.

Let y_t denote a forecast variable in quarter t; in the application, y_t is either output growth or inflation. Define the unconditional forecast of y_{t+h} based on information available to the CB during quarter t to be $y^u_{t,t+h}$. As the forecast is made during quarter t, y_t will not be, and y_{t-1} may not be, in the current information set. Thus, it may be of interest to consider nonpositive h. We will focus on $h = 0, ..., H$, where the maximum horizon,

H, in our applications is 4. Let i_t denote the policy interest rate in quarter t and let $i^u_{t,t+h}$ denote the unconditional forecast of i_{t+h} at time t.

Our maintained model is that

$$\begin{aligned} y^u_{t,t+h} - y_{t+h} &= \gamma'_{y,h} x_t + \varepsilon^y_{t,t+h} \qquad (1) \\ i^u_{t,t+h} - i_{t+h} &= \gamma'_{i,h} x_t + \varepsilon^i_{t,t+h} \qquad (2) \end{aligned}$$

where $\varepsilon^y_{t,t+h}$ and $\varepsilon^i_{t,t+h}$ are uncorrelated with anything in the central bank's information set at t and x_t is a vector of variables in this information set.[1]

Our goal is to test forecast efficiency, implications of which are:

$$H_y : \gamma_{y,h} = 0 \text{ for all } h$$
$$H_i : \gamma_{i,h} = 0 \text{ for all } h$$

The challenge is that $y^u_{t,t+h}$ and $i^u_{t,t+h}$ are unobserved—instead CBs report conditional forecasts, denoted $y^c_{t,t+h}$ and $i^c_{t,t+h}$. Our proposal for testing these hypotheses is based on a technically simple assumption about the relation between the conditional and unconditional forecasts.

2.1 The key assumption

A1 i) *The unconditional and conditional forecasts are related by the following system of equations:*

$$y^c_{t+h,t} - y^u_{t+h,t} = \sum_{j=0}^{h} \beta_{h,j} (i^c_{t+h-j,t} - i^u_{t+h-j,t}) \qquad (3)$$

$h = 0, \ldots, H$.

ii) The βs satisfy

$$\beta_{h,j} = \beta^*_j \qquad (4)$$

for $j \leq h$, for some β^*_j, $h = 0, ...H$.

Stacking the $H + 1$ equations in (3), we have

$$Y^c - Y^u = B(I^c - I^u) \qquad (5)$$

[1] When y_t is a variable such as GDP, which is continually revised, there is some ambiguity about what to use as the "final" value in defining the forecast error. We discuss this practical issue below.

where $Y^c = (y^c_{t,t}, ..y^c_{t,t+H})'$ and the other superscript c and superscript u variables are defined analogously. Part i of Assumption A1 says that the difference between the conditional and unconditional forecast for y_{t+h} is linear in the $h+1$ terms describing the difference between the unconditional and conditional paths for interest rates from time t to time $t+h$ and so imposes that B in (5) is lower triangular. Part ii requires the effect of a change in the path of interest rates at time $t+h-j$ on the time-t conditional expectation of y_{t+h} to be independent of h and so imposes that

$$B = \begin{pmatrix} \beta_0^* & 0 & \dots & 0 \\ \beta_1^* & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \beta_H^* & \dots & \beta_1^* & \beta_0^* \end{pmatrix} \tag{6}$$

In assumption A1, the $i^c_{t,t+h} - i^u_{t,t+h}$ variables capture the degree to which the conditioning assumption is counterfactual and the βs measure the degree of policy feedback. We now sketch a simple model in which assumption A1 holds. Suppose that each period the CB forms an unconditional forecast. We place no restrictions on how this is formed, but we do make two assumptions on how the conditional forecast is generated. First, assume that the conditional forecasts are formed consistent with Leeper and Zha's (2003) formulation of modest policy interventions—that is, they are treated as a sequence of hypothesized policy shocks of a modest magnitude relative to the unconditional forecast.[2] Second, assume that in a modest neighborhood of the unconditional forecast, the responses of all relevant variables to the policy shock are linear in the shock and well approximated by the conventionally defined impulse responses of the variables to a policy shock.

Under these circumstances, the conditional forecast paths can be written as a deviation from the unconditional forecast using the impulse response functions:

[2] Note that in the Greenbook forecast, the paths for certain variables other than the policy rate (such as exchange rates and some other financial market prices) are specified and taken as given in generating the remainder of the forecast. This might be taken to mean that there is more than one counterfactual path in the conditional forecast and that our tests need to soak up the effects of other conditioning assumptions. We believe that this is not the case; in particular we believe that only the policy rate forecast is deliberately counterfactual.

$$y^c_{t,t+h} = y^u_{t,t+h} + \Sigma^h_{j=0} d_j w_{t-j} \tag{7}$$

$$i^c_{t,t+h} = i^u_{t,t+h} + \Sigma^h_{j=0} g_j w_{t-j} \tag{8}$$

where the impulse responses of y and i to a policy shock at horizon j are given by d_j and g_j, respectively. The second equation implicitly defines the policy shocks, w_s, that are implied by the conditional and unconditional path for rates. In matrix notation, these equations can be written as,

$$Y^c - Y^u = Dw \tag{9}$$

$$I^c - I^u = Gw \tag{10}$$

where the Y and I variables are defined as in (5), and $D = Q(d)$, where $Q(d)$ is the lower triangular Toeplitz matrix with d_j on the j^{th} subdiagonal; $G = Q(g)$ is analogously defined. Thus,

$$Y^c - Y^u = B(I^c - I^u) \tag{11}$$

where $B = DG^{-1}$ is a lower triangular Toeplitz matrix and so A1 is satisfied.[3]

This reasoning gives a coherent recipe for generating forecasts that are conditional in a sense that is probably close to what central bank forecasts are based on. Making more definite statements about the practical relevance of A1 is probably impossible: the CB forecasts are heavily judgmental and involve subjective combination of inputs from many different models and analysts.

We can say a bit more in defense of A1, however. Goodhart (2005) describes the method in use at the Bank of England in terms consistent with A1. Further, conditional forecasts consistent with A1 can arise from a natural procedure. Suppose that the (implicit) model has a representation with two properties. First the representation is structural at least to the extent

[3] Because the inverse of a lower triangular Toeplitz matrix is a lower triangular Toeplitz matrix, and the product of two lower triangular Toeplitz matrices is also a lower triangular Toeplitz matrix.

that it has a policy reaction function. Second, the representation expresses the conditional mean of variables at time t as a function of past values of all variables. The unconditional forecast is given by iterating the model forward given the current information set. A conditional forecast is then generated by suspending the policy reaction function for the forecast horizon, replacing it with the conditional policy path, and iterating the model forward again. So long as the model is locally linear and the conditional policy path implies a modest intervention, A1 will hold.

2.2 Additional possible assumptions

The impulse response interpretation of the βs suggests some possible additional restrictions that might help to sharpen our inferences. In particular, the βs implicitly represent policy feedbacks from monetary policy shocks to the forecast variable. Conventional reasoning about these feedbacks suggests restrictions on these βs that we might want to impose.

For example, under conventional reasoning a positive shock to the policy interest rate has a nonpositive effect on output for at least a few quarters. The sign of the effects of a policy shock on inflation is not so clear. Some evidence and theory supports the view that the positive cost effect of higher interest rates could cause inflation to rise in the short run following a policy tightening. For a more complete discussion of this type of sign restrictions, see Faust (1998); for our purposes, the empirical controversies about these effects is not crucial. What we are interested in are the βs that the CB (implicitly) uses in creating the conditional forecast. Thus, for example, we are not interested in whether or not positive interest rate shocks initially raise inflation; rather, we are interested in whether a rise in the CB conditional policy path leads the CB to raise or lower the conditional forecast for inflation. We suspect that a rise in the policy path leads to nonpositive effects on both inflation and output growth forecasts at short horizons.

We will consider imposing versions of the following assumption:

A 2 *If the conditional policy path is such that*

$$(i^c_{t,t+h} - i^u_{t,t+h}) = c > 0$$

for h = 0, . . . , r then

$$(y^c_{t,t+h} - y^u_{t,t+h}) \leq 0$$

for h = 0, . . . , r.

The assumption says that raising the conditional path by a constant c at horizons 0 through r weakly lowers the conditional forecast for y over the same horizons. We will report results not imposing this assumption, and imposing it only for small values of r: r= 0, 1, 2. Since $y^c_{t,t+h} - y^u_{t,t+h} = \Sigma^h_{j=0}\beta_{h,j}(i^c_{t,t+h} - i^u_{t,t+h})$, imposing this assumption for any given r implies that $\Sigma^r_{j=0}\beta_{h,j} \leq 0$.

In general, we assume that the CB's conditional forecast for the policy rate is observed, but the unconditional forecast is not. However, a particularly simple efficiency test would be available if the CB's unconditional forecast for the policy rate were observed. Accordingly, we define the following transparency assumption:

A 3 *The public's unconditional expectation of the policy interest rate is the same as that of the central bank and can be measured from interest rate futures quotes.*

This assumption has two elements: the central bank and the public share the same policy expectations, and the public's expectations can be measured from financial market proxies.

3. Econometric Methodology

In this section, we turn to the task of constructing efficiency tests for conditional forecasts. The simplest conditional forecast efficiency test imposes assumption A1 (either part (i) alone, or both parts (i) and (ii)) as well as assumption A3. We refer to it as the test under *transparency*. To derive it, substitute $y^u_{t,t+h}$ from the maintained model (1) into (3):

$$y^c_{t,t+h} - y_{t+h} = \gamma'_{y,h}x_t + \sum_{j=0}^{h}\beta_{h,j}(i^c_{t,t+h-j} - i^u_{t,t+h-j}) + \varepsilon^y_{t,t+h} \qquad (12)$$

All the variables on the right-hand side of this equation are observed by the econometrician, under the transparency assumption. This allows us to then estimate the equation by OLS and to test H_y using any test of zero restrictions on the γs. This is a standard regression-based test of forecast efficiency with the augmenting variables in the summation present to *soak up* the consequences of conditioning. The test might reasonably be viewed as a test of the efficiency of $y^u_{t,t+h}$ or of $y^c_{t,t+h}$. More precisely, it is a joint test of the hypothesis that $y^c_{t,t+h}$ can be related to some latent, efficient unconditional forecasts, $y^u_{t,t+h}$ and $i^u_{t,t+h}$ in the manner stated in A1. All this relies on $i^u_{t,t+h-j}$ being observed by the econometrician.

3.1 GMM, IV-based tests

To construct a forecast efficiency test that is valid when the i^us are *not* observed—i.e. under assumption A1 alone—substitute for i^u_{t+h} in (12) using (2):

$$\begin{aligned} y^c_{t,t+h} - y_{t+h} = \gamma'_{y,h} x_t + \textstyle\sum_{j=0}^h \beta_{h,j} (i^c_{t,t+h-j} - i_{t,t+h-j}) \\ - \textstyle\sum_{j=0}^h \beta_{h,j} (\gamma'_{i,h} x_t + \varepsilon^i_{t,t+h}) + \varepsilon^y_{t,t+h} \end{aligned}$$

Collect the terms in x_t and the terms in the εs to write,

$$y^c_{t,t+h} - y_{t+h} = \tilde{\gamma}'_h x_t + \sum_{j=0}^{h} \beta_{h,j} (i^c_{t,t+h-j} - i_{t,t+h-j}) + \tilde{\varepsilon}_{t,t+h} \tag{13}$$

where $\tilde{\gamma}_h = \gamma_{y,h} - \sum_{j=0}^h \beta_{h,j} \gamma_{i,h}$ and $\tilde{\varepsilon}_{t,t+h} = \varepsilon^y_{t,t+h} - \sum_{j=0}^h \beta_{h,j} \varepsilon^i_{t,t+h}$.

These equations have two important properties. First, jointly H_y and H_i imply $\tilde{H}_0$: $\tilde{\gamma}_h = 0$; thus, any valid test of $\tilde{H}_0$ will also be a valid joint test of H_y and H_i.[4] Second, $\tilde{\varepsilon}_{t,t+h}$ is a composite of efficient forecast errors, and, thus, shares with the original εs the property of being orthogonal to any variable in the CB's information set at time t.

Given the second property, any variable in the CB information set at time t will be a valid instrument for the $i^c_{t,t+h-j} - i_{t,t+h-j}$ variables in (13). Thus, we can estimate the coefficients of (13) consistently by instrumenting for the $i^c_{t,t+h-j} - i_{t,t+h-j}$, and any valid test of $\tilde{H}_0$: $\tilde{\gamma}_h = 0$ will be a test of the joint validity of H_y and H_i.

There are two differences between the IV-based test and the OLS-based test under transparency. We have replaced $i^u_{t,t+h}$ with the *ex post* realization

[4] While $\tilde{H}_0$ is necessary for H_y and H_i, it is not sufficient. The expression for $\tilde{\gamma}_h$ can be zero even if some of the γs are nonzero.

$i_{t,t+h}$. This forces us to use IV to properly *soak up* the effects of conditioning. Second, the test is now a test of the joint efficiency of the y and i forecasts. The OLS-based test was directed at the y forecast alone.

The principle requirement for valid x_t variables and for valid instruments is the same: both must come from the CB's information set at time t so that under the null they are orthogonal to $\tilde{\varepsilon}_{t,t+h}$. One could partition such variables into x_ts and instruments, but there seems to be no basis upon which to make such a distinction. A natural way to proceed is to lump all the potential xs and instruments together, use them all as instruments, and then estimate equation (13) with no γ terms by IV. One can test forecast efficiency by testing any over-identifying restrictions in this system.

More concretely, we estimate,

$$y^c_{t,t+h} - y_{t+h} = \sum_{j=0}^{h} \beta_{h,j}(i^c_{t,t+h-j} - i_{t,t+h-j}) + u_{t,t+h} \tag{14}$$

by instrumental variables using any variables in the CB's information set at time t as instruments. We then test overidentifying restrictions using the Sargan (1958) test or its generalization as Hansen's J test in the GMM context. Intuitively, this tests whether the instruments explain the residuals from (14); that is, it tests whether the instruments collectively explain the forecast error $y^c_{t,t+h} - y_{t+h}$, above and beyond their ability to explain $i^c_{t,t+h-j} - i_{t,t+h-j}$. Inference can be sharpened if we impose assumption A2 as well, because that restricts the parameter space to the region in which $\Sigma^r_{j=0}\beta_{h,j} \leq 0$.

3.2 Valid Instruments and Identification

We need to instrument for variables of the form $i^c_{t,t+h-j} - i_{t,t+h-j}$, using variables in the central bank's information set at time t. Since the conditioning path for interest rates is deliberately counterfactual and since $i^c_{t,t+h-j}$ is in the information set, one natural instrument is $i^c_{t,t+h-j} - i^f_{t,t+h-j}$ where $i^f_{t,t+h-j}$ is any interest rate forecast in the CB's information set and for which we have data. We will use rates on money market futures as these forecasts. More generally, any variable that helps predict interest rates may be correlated with $i^c_{t,t+h-j} - i_{t,t+h-j}$ and so may be a good instrument.

In the single equation test, we estimate (14) by IV and have $h+1$ variables to instrument for. Identification requires that the projection matrix for projecting the $h+1$ $(i^c_{t,t+h-j} - i_{t,t+h-j})$ variables on the instruments be of rank $h+1$. For h larger than 0 or 1, the structure of the problem suggests that this identification condition may fail or nearly fail.

The $h+1$ endogenous variables are interest rate differentials, $i^c_{t,t+h-j} - i_{t,t+h-j}$, at successive horizons. Each can be written as $i^c_{t,t+h-j} - i^u_{t,t+h-j} + \varepsilon^i_{t,t+h}$ and our instruments will be correlated only with $i^c_{t,t+h-j} - i^u_{t,t+h-j}$. Both i^c and i^u tend to have very smooth paths; thus, $i^c - i^u$ may itself have a variance-covariance matrix that is nearly singular, and our projection matrix may be nearly singular as well.

The rank condition for identification may not be such a problem if we treat the equations as a system and impose the cross-equation restrictions on the βs implied by A1(ii). Consider estimating the system for horizons 0 through H by GMM. Let $z_{t,h}$ be a vector of k_h instruments to be used in forming the moment conditions for equation h. This gives $K = \Sigma^H_{h=0} k_h$ moment conditions of the form

$$E(\tilde{\varepsilon}_{t,t+h} z_{t,h}) = 0$$

The sufficient condition for identification of the $H+1$ β^*s requires that the gradient of the moment conditions with respect to the parameter values be of full column rank. Given the restrictions on the β^*s in A1(ii), this requires that the Kx$(H+1)$ matrix

$$\Pi = \begin{pmatrix} E((i^c_t - i_t)z_{t,0}) & 0 & \dots & 0 \\ E((i^c_{t+1} - i_{t+1})z_{t,1}) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ E((i^c_{t+H} - i_{t+h})z_{t,H}) & \dots & E((i^c_{t+1} - i_{t+1})z_{t,H}) & E((i^c_t - i_t)z_{t,H}) \end{pmatrix}$$

must have rank $H+1$.

This identifying restriction is far easier to meet than the one in the single equation case: it can, for example, be satisfied using the same single instrument for each equation. Specifically, if there is a variable, z_t, for which

$E((i^c_{t,t} - i_t)z_t) \neq 0$ and we use this as an instrument in each equation, then we will satisfy the rank condition. To see this, note that in this case the main diagonal of Π will be a constant and nonzero, which is sufficient for full rank of a lower triangular matrix. This result follows under A1(ii) from the lower triangular nature of B in (6). Even though the system is likely to be formally identified, even for large H, it is still possible, of course, that identification will be weak.

Either in single-equation estimation, or in the system, identification may be weak, meaning that the matrix Π may be rank deficient, or nearly rank deficient. This, in turn, would imply that our tests of overidentifying restrictions would have size distortions. Fortunately, there are approaches to inference that are robust to weak identification and even a complete lack of identification. These weak instrument issues can make inference about the βs quite complicated. In the case at hand, however, we are not much interested in the βs: they are essentially nuisance parameters. We are mainly interested in testing $\gamma = 0$ (or, equivalently, in testing overidentifying restrictions), and robust inference about γ is a bit simpler.

To understand the robust test, it is useful briefly to review inference regarding a parameter vector θ, that could be either the βs in the single equation framework or the β^*s in the system. Suppose that θ_0 is a hypothesized parameter vector in (14) and $S(\theta)$ denotes the continuous updating GMM objective function based on K moment conditions ($S(\theta)$ reduces to LIML in the single equation case, and FIML in the system case). The distribution of $S(\theta_0)$ is asymptotically $\chi^2_{(K)}$ under the null hypothesis of forecast efficiency, regardless of the rank of Π (Anderson and Rubin, 1949; Staiger and Stock, 1997; Stock and Wright, 2000; Stock, Wright and Yogo, 2002). If $c(\alpha, K)$ is the upper α percentile of a $\chi^2_{(K)}$ distribution, then the set

$$\mathcal{S} = \{\theta \in \mathbf{\Theta} : S(\beta) \leq c(\alpha, K)\}$$

where $\mathbf{\Theta}$ is the parameter space, is a confidence set for θ with asymptotic coverage of $100 - \alpha$ percent. Under weak identification, this confidence set has infinite expected volume (Dufour, 1997), and hence is uninformative. If the moment conditions are false, in the sense that no parameter value satisfies them, then $\mathcal{S}$ is empty with probability one asymptotically.

This suggests a valid test of overidentifying restrictions that is robust to weak instruments. We reject if

$$\inf_{\theta \in \Theta} S(\theta) > c(\alpha, K)$$

This test will be asymptotically conservative in that it will reject the null with probability less than or equal to α asymptotically.

Note that making the usual assumption of identification, and using the LIML or FIML estimator, the Sargan test (or Hansen's J-test) of overidentifying restrictions rejects if

$$\inf_{\theta \in \Theta} S(\theta) > c(\alpha, K - p),$$

where p is the dimension of β. Thus, the robust test merely raises the critical value by taking the value from the $\chi^2_{(K)}$ distribution instead of the $\chi^2_{(K-p)}$. Clearly, if the rank condition for identification is satisfied, using the robust test wastes power. In the tables below we report both standard and robust p-values.

Tests are available that may provide some indication as to which p-values may be more appropriate. For example, the rank test of Cragg and Donald (1993) is a generalization of the simple first-stage F statistic to the case of multiple regressors. This test can be used to test the null hypothesis that the matrix Π is rank deficient, and, thus, provides a test of the null of weak identification. It is well known that even if this test rejects a lack of identification at conventional significance levels, weak instrument problems in the second stage may still be severe in small samples (Hall, Rudebusch and Wilcox, 1996; Staiger and Stock, 1997; and Stock and Yogo, 2001). Accordingly, our inclination is to emphasize the robust p-values, except when the rejection from the rank test is particularly strong.

3.3 The tests with possible violations of the assumption A1

While the basic impulse-response structure implied by assumption A1 underlying these tests is probably appropriate, the assumption that these impulse responses are time invariant may not be. In any judgmental system, time invariance might be hard to impose (or to check). Tetlow and Ironside (2006) find evidence of a changing impulse responses in the Fed's largest explicit model of the economy—not to be confused with the implicit model generating the forecast. Faust and Leeper (2005) report results consistent with changing βs in the BOE forecast.

These findings warrant two additional comments. First, we can include an error in (3), which is now written as an identity. We will not detail all the cases about allowable forms for the error, but to get the idea of what is possible, consider the single-equation test under transparency. To retain proper size of the test, there can be an error in (3); it can be correlated with $i^c - i^u$, but not with any x variables.

Second, strictly speaking, we only need A1 to hold under the null hypothesis. Thus, so long as the deviations from A1 are reflective of inefficiency in the forecasting system, rejections stemming from deviations from A1 are appropriate.

4 Empirical application

In this section we apply the tests described above to the Greenbook forecast of the Federal Reserve and to the inflation report forecast of the Bank of England. As noted in the introduction, the properties of these forecasts have been studied before both by the CBs and by others.

Romer and Romer (2000) found that the Greenbook forecast is superior to private sector alternatives; Sims (2002) reaches a similar overall conclusion, while being more critical about a number of particulars. Atkeson and Ohanian (2001) and Faust, et al. (2004) are less supportive of the Greenbook forecast. Work by the Bank of England (2004) shows that their inflation forecast is pretty good by conventional standards, but the GDP forecast is not as good. Pagan (2003) reviews the BOE's forecasting framework and forecast and finds generally favorable results.

Perhaps more relevant to our distinction between the conditional and unconditional forecast, several authors, (e.g., Faust and Henderson, 2004; Goodhart, 2005) have argued that the BOE's forecast has some odd properties when viewed as a conditional forecast.[5] In particular, while the forecast

[5] Other inflation targeting banks show this same phenomenon in their forecasts (Leeper, 2003).

is conditioned on a constant path for policy, forecast inflation tends to return to target over a horizon of two years. This would seem to indicate that the policymakers never expect to need to change policy in order to hit the inflation objective. Policy rate changes were, however, frequent and highly serially correlated over this period (as with most CBs in most periods). Tests explicitly taking account of the conditional nature of the forecast could potentially shed some light on such issues.

4.1 The Data

We use the Greenbook forecast of output and inflation measured as GDP and the GDP deflator for Greenbook forecasts from Oct. 1988 through the last Greenbook in 1999. This gives a total of 85 observations given the FOMC schedule and a few missing Greenbook forecasts. We use the GDP and RPIX inflation forecast data published by the BOE in quarterly inflation reports from 1998:1 through 2003:2, giving 22 observations.[6] See the Appendix for more details on the raw data and data construction.

The Greenbook sample ends about 5 years ago due to confidentiality requirements of the Fed. The BOE sample ends about 2 years ago to allow us to have reasonably mature "final" data for creating forecast errors for GDP. Because we have almost 4 times as many observations for the Greenbook forecast relative to the BOE forecast, we expect to have considerably greater power to reject the efficiency hypothesis if it does not hold.

The inflation and output data are stated as approximate growth rates measured as 400 times the quarterly log change in the variable. While the RPIX inflation rate is not revised, the other three variables are perpetually revised, giving rise to an issue regarding what to treat as the *final* data in computing forecast errors. From a forecasting standpoint, a key issue is whether the revisions are forecastable, and in the U.K. but not the U.S. the revisions are significantly forecastable.[7] For the "final" data, we use the data as they stood about 2 years after the time of the forecast.

Our data also include the interest rate path, i^c, upon which the forecasts are conditioned, obtained from the same sources. These are stated as

[6] Our sample does not span the change in price indices by the BOE.
[7] For a discussion of this issue, see Faust, et al. (2005a).

quarterly average interest rates in annualized percent.

We also require a measure of private sector expectations of the policy interest rate. For the U.S., we use a measure created from Federal Funds rate and eurodollar futures. For the U.K. we use the financial market-based measure of expectations published by the BOE in the inflation report. For more details, see the Appendix.

Figure 1 plots the assumed interest rate paths from selected Greenbooks (the last Greenbook in our sample in each year since 1989). The figure also shows the corresponding market forecasts for policy, constructed from interest rate futures quotes. The Greenbook interest rate assumption is often very close to being an unchanged path for policy, which can deviate substantially from the market forecast. Under the transparency assumption, the market forecast is the unconditional expectation of interest rates, implying that the conditioning assumption is often highly counterfactual. Without the transparency assumption, we cannot tell from these graphs how counterfactual the conditioning assumption was from the CBs perspective, but it still seems reasonable to suppose that it may lead the conditional expectations for inflation or output growth to be quite different from the corresponding unconditional expectations. Similarly, Figure 2 shows the interest rate assumptions used in selected BOE inflation reports (the report in the second quarter of each year). Again, the concurrent market forecasts are also included. The conditioning assumption is by construction always a flat line at the current policy rate beyond the current quarter and can be quite far from the market forecast.

4.2 Concrete varieties of the tests

There is a very large number of tests we could report based on different permutations of testing assumptions, choices of instruments and augmenting forecast variables. To keep this somewhat manageable, we have made the following choices.

First consider the OLS-based tests. Here we take our financial-market-based measure of interest rate expectations, $i^f_{t,t+h}$, and form regressors $\delta(c,f)_{t,k} = i^c_{t,t+k} - i^f_{t,t+k}$. Then, setting $i^u_{t,t+h-j} = i^f_{t,t+h-j}$ in equation (12), we use a conventional (robust) Wald test to test $\gamma_{y,h} = 0$ in the regression,[8]

$$y^c_{t,t+h} - y_{t+h} = \gamma'_{y,h} x_t + \Sigma^h_{j=0} \beta_{h,j} \delta(c,f)_{t,h-j} + \varepsilon^y_{t,t+h}$$

For comparison purposes, we run what we call a *naive* version of the test, which omits the summation term. This is a natural test of efficiency under the assumption that the fact of conditioning can be neglected. The *transparency* version of the test includes the summation terms. We consider two sets of xs, both of which include a constant. First, we include only $y^c_{t,t+h}$. This is the natural generalization of the familiar Mincer-Zarnowitz test to the present case; we call this the R1 set. The second set is called the R2 and uses instruments I2, defined and motivated below, as the augmenting regressors.

Next, turning to the IV single-equation tests without transparency, the estimating equation, from (14), is

$$y^c_{t,t+h} - y_{t+h} = \sum^h_{j=0} \beta_{h,j} \delta(c,a)_{t,h-j} + u_{t,t+h}$$

where $\delta(c,a)_{t,k} = i^c_{t,t+k} - i_{t,t+k}$. We do not need to impose assumption A2, but if we choose to do so then this just means restricting the parameter space to the region where $\Sigma^r_{j=0} \beta_{h,j} \leq 0$ for some value of r. We must instrument for the $\delta(c,a)$s. We always include a constant in the instruments. As argued above, $\delta(c,f)_{t,h-j}$ should be a reasonably efficient predictor of $\delta(c,a)_{t,h-j}$ and is in the CB information set at t. Thus, we always include $\{\delta(c,f)_{t,h-j}\}^h_{j=0}$. We supplement these instruments in 2 different ways. Instrument set I1 supplements the δ variables with $y^c_{t,t+h}$. Instrument set I2 includes $y^c_{t,t+h}$ and also uses the Mallows (1973) criterion in the first-stage regression to choose further instruments from among all combinations of $\delta(c,f)_{t,h-j},\ j = 1, ...H$ and $y^c_{t,t+j},\ j = 0, ...H,\ j \neq h$. The selection is made separately for each equation. [9] We use this method to choose a

[8] We use the Newey-West heteroskedasticity and autocorrelation consistent estimate of the variance-covariance matrix with lag length equal to h.

[9] This procedure is loosely motivated by Donald and Newey (2001), who show that if the available instruments can be ordered from most to least relevant, then choosing the number of instruments from this ordered list to minimize the Mallows criterion has certain asymptotic optimality properties. As we do not posit an ordered list, the result does not strictly apply.

parsimonious set of instruments because, while we are concerned about the possibility of weak identification, we also have a large number of potentially valid instruments that are likely to be highly correlated. This includes any variable in the information set at time t that may predict interest rates.

[The instrument selection problem is somewhat different for the systems estimates. As noted above, no matter how large is our chosen maximum horizon H, one instrument may be sufficient for identification. We report systems of equations for $H = 1, \ldots, 4$. We again consider two choices of instrument sets. With both instrument sets, for each equation, h, in the system, we include a constant and $\delta(c, f)_{t,t+h}$ as instruments. For instrument set I3, we also include $y^c_{t,t+h}$ among the instruments for equation h. For instrument set I4, we instead include $y^c_{t,t}$ as the additional instrument in each equation.[10]—drop this section if not reporting system results]

4.3 Testing Transparency

For Assumption A3, it is necessary that the public's expectations can be measured from financial market proxies. Of course, financial market proxies may diverge from true expectations by a potentially time-varying risk premium. It is well known that this is a very serious issue when using interest rates of over a year (e.g, Chernenko, et al. 2004). But we can, of course, test whether our proxy data for rate expectations constitute efficient forecasts of subsequent interest rates. We report one such test in Table 1. Consistent with other work in this area we find that our financial-market-based proxies for interest rate expectations seem to be efficient predictors out at most one or two quarters. Further ahead, the assumption that the central bank and the public share the same policy expectations seems very strong.

Against this backdrop, we put forward the transparency assumption mainly as an important benchmark. Even though central banks have made great strides in transparency, it is surely too much to assume that CB and private sector expectations of the policy rate entirely coincide. We believe that the results are still of interest, however, if we view the test as a test of the joint hypothesis of transparency and efficiency.

[10] Both the single equation and the system estimates are based on the continuous updating GMM objective function and the Newey-West weight matrix with lag length h in the single equation case, and the largest lag h in the system, for the system case.

Suppose, as we find in some cases below, that the test under transparency rejects CB forecast efficiency, but without the transparency assumption we cannot reject efficiency. Setting aside type I errors, there are two natural explanations. First, the CB is transparent and efficiency fails. The instrument-based test simply has power that is too low to detect the efficiency failure. Second, transparency is incomplete. If the lack of transparency is sufficient to account for the results, it must be that the central bank and public disagree on the likely future path of policy.[11] This disagreement must be sufficiently large to drive significant differences in the forecast paths for output growth or inflation. Each of these possibilities—lack of efficiency and lack of transparency—may be troubling from a public policy perspective.[12]

4.4 The Naive and Transparency Tests

The OLS-based tests under the naive and transparency assumptions for the Greenbook illustrate several important points (Table 2a–b). First, there is a fairly general rejection of forecast efficiency under these assumptions for the Greenbook: throughout the tables, many of the $p-$values are quite small. Second, in many instances, the $p-$values for rejecting forecast efficiency are smaller in the transparency case than in the naive version.[13] One might incorrectly have supposed that, since naively treating the conditional forecast as unconditional can lead to incorrect rejections of forecast efficiency, taking account of conditionality must necessarily lead to fewer rejections. As the Tables illustrate, this need not necessarily be the case. Taking account of the role of the conditioning error may increase the precision of the regression and allow one to more clearly measure the role of the additional xs in the regression.

Third, the Tables also show the ΔR^2 statistics, which are the changes in the uncentered R^2 due to including the x variables, and which form a crude estimate of how much better the forecast might have been had the x variables been used better. In the naive form of the test, these numbers

[11] Unless the financial markets give exceptionally poor measures of private sector rate expectations.

[12] Of course, statistically significant differences in forecasts do not necessarily imply economically meaningful differences.

[13] The relevant comparison is across p-values on any row.

are often quite substantial, in the 0.15 to 0.30 range. These values are much smaller in the transparency form of the test, suggesting that, despite the sharper rejection, the economic significance of the inefficiency may be rather modest.

The U.K. results are broadly similar. Given that we have about one-quarter as many observations in this case, the strength of the rejections might seem surprising. As noted above, the GDP forecast, in particular, has been found to be problematic in the past. In the case of U.K. GDP growth, the ΔR^2 statistics suggest that inefficiency might be nontrivial in economic terms.

In the end, however, we prefer to think of these results mainly as illustrative benchmarks. One must remember that the transparency results are premised on the assumption that public and central bank rate expectations coincide and that our proxy taken from financial markets is an accurate measure of those expectations. Both assumptions are surely too strong and the Table 1 results suggest that rate expectations measured in financial markets are associated with systematic prediction errors.[14]

4.5 The GMM, IV-based tests

The IV-based tests do away with the transparency assumption, at the expense of requiring instrumental variables estimation. The results for single-equation estimation are shown in Table 3. Results are shown for different values of the horizon h, and both imposing Assumption A2 for various values of r, and without imposing this assumption. Recall that r measures the number of periods for which the effect of policy shocks is required to be nonpositive. Where assumption A2 is not imposed, the value of r in the second column is left blank. The rank test reported in the final two columns very generally indicates that weak instrument issues may be serious. The primary exception to this conclusion is for short horizons in the case of the U.S..

At the zero- or one-quarter horizons, the single-equation test of the

[14] Of course, central banks seem to rely on these financial market measures to infer rate expectations. We do not require that the proxy be an efficient forecast, only that it be an accurate reflection of CB rate expectations. See, for example, the discussion of market-rate-based conditional expectations in Faust and Leeper (2005).

Greenbook GDP growth forecast (Table 3a) generally rejects efficiency. At longer horizons, the robust p-values are most appropriate, and provide little if any evidence against the efficiency hypothesis.

For the Greenbook inflation forecast, the story is different. There is not solid evidence against efficiency at any horizon. However, at the horizon of 2 to 3 quarters, some of the conventional $p-$values suggest rejecting efficiency, especially when assumption A2 is imposed for larger values of r. Taking these conventional $p-$values seriously for a moment, it is worth stressing the interpretation. There is relatively strong evidence against efficiency of the inflation forecast, so long as we assume that artificially high interest rates in the conditioning path have nonpositive effects on inflation in the conditional forecast (relative to the unconditional).

As for the U.K. forecasts, there is no evidence against forecast efficiency in the IV-based estimates. This might seem surprising given the OLS-based results, where rejections were plentiful. In part, the explanation comes from the fact that we have very few observations for the U.K. (only 22).

To conserve space, results of the system GMM tests are not shown, but are broadly similar, with some evidence against forecast inefficiency for the Greenbook GDP growth forecast, but no solid evidence against conditional efficiency of any of the other forecasts.

We do not mean to suggest, however, that in light of the OLS estimates, failure to reject efficiency in the IV estimates is simply an issue of lack of power. It is perfectly possible that neither result is in error. As noted above, one way to reconcile the two results is with the conclusion that central bank and private sector expectations for interest rates are very different, and this difference in opinions over rates is of sufficiently great economic significance as to drive the difference in results.

5. Conclusion

Central banks commonly make forecasts conditional on counterfactual paths for the policy interest rate. Evaluating the quality of such forecasts is complicated. Lacking a better method, up until now, these forecasts have generally been evaluated as if the conditional nature could appropriately be neglected.

As our results illustrate, ignoring the conditional nature can lead to either type I or type II errors in tests of forecast efficiency.

We show how tests for efficiency may be constructed under a plausible assumption about the relation between the conditional and underlying unconditional forecasts. We apply these new tests to the Fed's Greenbook forecast and the Bank of England's inflation report forecast. Under a very strong transparency assumption, we find some evidence against efficiency of both the Greenbook and BOE forecast. Under weaker assumptions, we still find some evidence against efficiency of the Greenbook forecast.

These results raise some deeper issues for those seeking to monitor and/or improve forecasting operations. It takes relatively strong assumptions to coherently analyze these conditional forecasts. While our assumptions are plausible, the appropriateness of each is open to question. Even if the assumptions are correct, the nature of the problem is likely to lead to weak instrument issues, requiring a robust approach to inference. Thus, even substantial inefficiency in the forecasting framework of the central bank may be difficult to detect based on conditional forecasts.

In short, the conditional nature of these forecasts presents a significant roadblock to the analysis of their quality. We believe this presents an important challenge to quality control in forecasting operations that are based on conditional forecasts. This issue may be particularly important in the case of the BOE and other inflation targeting central banks, which put the forecasts forward as an important element facilitating monitoring of the central bank. It will be very difficult for the public to gain any clear understanding of the merits of the forecast.

Data Appendix

Our Greenbook data come from internal historical archives of the Greenbook forecast. A version these data (excluding the interest rate assumptions) is available on the Philadelphia Fed web site. The BOE forecast data are available from the BOE website.

For the U.S., our "final" vintage data come from the historical Greenbook archives. The final data are the data as they stood at the time of the first Greenbook at least two years after the from which the forecast is taken.

For the U.K., no vintage data are needed for the RPIX. For GDP growth, we use the vintage of data 2-years after the inflation report in question. Most of these data are posted on the BOE web site. For certain vintages, we augment the BOE data with data typed in from original sources.

The private sector expectations data for the U.S. are constructed from federal funds rate futures and eurodollar futures data in a standard manner described in Faust, et al. (2005b). The BOE expectations data are reported in the relevant inflation reports.

The inflation and GDP data are converted from the basis reported to approximate annualized percent changes computed as $400 \log(x_t/x_{t-1})$.

We start our forecasts at horizon zero. That is, the forecast in quarter t of quarter t data. Both the interest rate conditioning assumption and the public expectations are stated on a quarterly-average-of-daily-data basis. Since these data are available in real time, the forecast for the current quarter made, say, 23 days into the current 91-day quarter, must be constructed as an average of the 23 days of actual data and the forecast (conditional or otherwise) going forward. We construct the data in the appropriate way based on a choice of when the forecast is made in the quarter. For the Greenbook forecast, we know the day that the forecast was completed. In the case of the inflation reports, this is not so clear. We use a date about a week before publication. The actual dates chosen are available from the authors.

References

Anderson, Theodore W. and Herman Rubin (1949): Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *Annals of Mathematical Statistics*, 20, pp.46–63.

Andersson, Michael, Jesper Linde, Mattias Villani and Anders Vredin (2005): Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks, manuscript, *Riksbank*.

Atkeson, Andrew, and Lee E. Ohanian (2001): Are Phillips Curves Useful for Forecasting Inflation? Federal Reserve Bank of Minneapolis Quarterly Review, vol. 25, no. 1, winter, 2–11.

Bank of England (various), *Inflation Report.*

Bank of England (2004), *Inflation Report,* August.

Canzoneri, M.B. (1985): Monetary Policy Games and the Role of Private Information, *American Economic Review*, 75, 1056–70.

Chernenko, Sergey V., Krista B. Schwarz, Jonathan H. Wright (2004): The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk, *International Finance Discussion Paper*, 2004-808.

Cragg, John G. and Stephen G. Donald (1993): Testing Identifiability and Specification in Instrumental Variable Models, *Econometric Theory*, 9, pp.222–240.

Donald, Stephen G. and Whitney K. Newey (2001): Choosing the Number of Instruments, *Econometric*a, 69, pp.1161–1191.

Dufour, Jean-Marie (1997): Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, *Econometrica*, 65, pp.1365–1387.

Faust, Jon (1998): The Robustness of Identified VAR Conclusions About Money, *Carnegie-Rochester Conference Series on Public Policy*, 49, 207-244.

Faust, Jon and Dale Henderson (2004): Is inflation targeting best practice monetary policy?, Federal Reserve Bank of St. Louis *Review* , v.86, n.4, pp. 117–144.

Faust, Jon and Eric Leeper (2005): Forecasts and Inflation Reports: An Evaluation. Manuscript.

Faust, Jon, Rogers, John H., Wright, Jonathan H. (2005a): News and Noise in G-7 GDP Announcements, *Journal of Money, Credit, and Banking*, v.37, n.3, June., 403–419.

Faust, Jon, Rogers, John H., Wang, Shing-Yi, Wright, Jonathan H. (2005b): The High Frequency Response of Exchange Rates and Interest Rates to Macroeconomic Announcements, *Journal of Monetary Economics*, forthcoming.

Faust, Jon, Eric T. Swanson, and Jonathan H. Wright (2004): Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?,*BEPress Contributions to Macroeconomics*, Vol. 4: No. 1, Article 10.

Goodhart, Charles (2005): The Monetary Policy Committee's Reaction Function: An Exercise in Estimation, BEPress Topics in Macroeconomics, 5(1), no. 18..

Hall, Alastair R., Glenn D. Rudebusch and David W. Wilcox (1996): Judging Instrument Relevance in Instrumental Variables Estimation, *International Economic Review*, 37, pp.283–289.

Leeper, Eric M. (2003): An 'Inflation Reports' Report, Sveriges Riksbank *Economic Review* , n.3., pp. 94–118.

Leeper, Eric M. and Zha, Tao (2003): Modest Policy Interventions, *Journal of Monetary Economics,* vol. 50, iss. 8, pp.1673–1700.

Mallows, Colin L. (1983): Some Comments on C_p, *Technometrics*, 15, pp.661–675.

Pagan, Adrian (2003): *Report on modeling and forecasting at the Bank of England* , London: Bank of England.

Romer, Christina and David Romer (2000): Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review* 90, 429–57.

Sargan, John D. (1958): The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica*, 26, pp.393–415.

Sims, Christopher A. (2002): The Role of Models and Probabilities in the Monetary Policy Process, *Brookings Papers on Economic Activity,* iss. 2, pp.1–40.

Staiger, Douglas and James H. Stock (1997): Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, pp.557–586.

Stock, James H. and Jonathan H. Wright (2000): GMM with Weak Identification, *Econometrica*, 68, pp.1055–1096.

Stock, James H., Jonathan H. Wright and Motohiro Yogo (2002): A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20, pp.518–529.

Stock, James H. and Motohiro Yogo (2005): Testing for Weak Instruments in Linear IV Regression, in D.W.K. Andrews and J.H. Stock, eds., "Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg," Cambridge University Press, Cambridge.

Tetlow, Robert J., and Brian Ironside, (2006): Real-time Model Uncertainty in the United States: The Fed from 1996-2003, FEDS working paper 2006-8 (March).

Table 1. Test of forecast efficiency of interest rate expectations data

h:	0	1	2	3	4
			U.S.		
W	0.68	9.85	8.87	13.15	15.32
p-val	0.71	0.01	0.01	0.00	0.00
			U.K.		
W	1.78	0.79	3.66	17.63	38.77
p-val	0.41	0.67	0.16	0.00	0.00

Note: W is the Wald statistic for testing $a = b = 0$ in

$$i^f_{t,t+h} - i_{t+h} = a + b i^f_{t,t+h} + v_t$$

and p-val is the associated p-value from the asymptotic χ^2 distribution of the statistic under the null.

Table 2a. Two OLS-based forecast efficiency tests: U.S. output growth

h	Naive W	Naive p-val.	Naive ΔR^2	Transp. W	Transp. p-val.	Transp. ΔR^2
			R1			
0	14.39	0.00	0.15	14.70	0.00	0.00
1	3.01	0.22	0.06	4.33	0.11	0.01
2	2.94	0.23	0.08	6.50	0.04	0.06
3	4.42	0.11	0.12	9.98	0.01	0.08
4	2.08	0.35	0.09	6.21	0.04	0.07
			R2			
0	14.50	0.00	0.15	5.13	0.16	0.03
1	13.31	0.00	0.16	21.53	0.00	0.13
2	44.85	0.00	0.33	58.30	0.00	0.15
3	10.37	0.01	0.15	25.34	0.00	0.05
4	1.34	0.25	0.05	4.66	0.03	0.00

Note: See notes at end of Table 2d.

Table 2b. Two OLS-based forecast efficiency tests: U.S. inflation

h	Naive W	Naive p-val.	Naive ΔR^2	Transp. W	Transp. p-val.	Transp. ΔR^2
			R1			
0	15.01	0.00	0.15	16.81	0.00	0.00
1	13.49	0.00	0.19	19.30	0.00	0.02
2	13.41	0.00	0.23	17.18	0.00	0.01
3	12.62	0.00	0.26	17.45	0.00	0.03
4	7.58	0.02	0.21	5.63	0.06	0.03
			R2			
0	20.30	0.00	0.16	2.63	0.62	0.01
1	13.58	0.00	0.21	18.27	0.00	0.07
2	12.48	0.00	0.23	0.10	0.75	0.00
3	11.11	0.00	0.25	0.14	0.71	0.00
4	8.69	0.01	0.23	1.95	0.38	0.02

Note: See notes at end of Table 2d.

Table 2c. Two OLS-based forecast efficiency tests: U.K. output growth

h	Naive W	Naive p-val.	Naive ΔR^2	Transp. W	Transp. p-val.	Transp. ΔR^2
			R1			
0	15.01	0.00	0.31	28.38	0.00	0.35
1	23.32	0.00	0.41	7.99	0.02	0.20
2	3.63	0.16	0.21	4.78	0.09	0.13
3	20.10	0.00	0.40	59.37	0.00	0.20
4	18.42	0.00	0.39	45.66	0.00	0.39
			R2			
0	1.83	0.61	0.08	30.84	0.00	0.07
1	16.14	0.00	0.45	10.43	0.01	0.04
2	0.08	0.78	0.01	3.98	0.05	0.00
3	2.64	0.27	0.09	18.62	0.00	0.02
4	0.04	0.85	0.00	22.37	0.00	0.00

Note: See notes at end of Table 2d.

Table 2d. Two OLS-based forecast efficiency tests: U.K. inflation

h	Naive W	Naive p-val.	Naive ΔR^2	Transp. W	Transp. p-val.	Transp. ΔR^2
			R1			
0	0.11	0.94	0.01	0.12	0.94	0.00
1	2.77	0.25	0.06	1.40	0.50	0.04
2	2.05	0.36	0.11	1.08	0.58	0.04
3	2.44	0.30	0.12	5.02	0.08	0.10
4	4.73	0.09	0.08	1.65	0.44	0.02
			R2			
0	1.59	0.81	0.04	1.30	0.86	0.04
1	0.85	0.84	0.03	1.75	0.63	0.01
2	0.04	0.84	0.00	0.38	0.54	0.00
3	0.09	0.76	0.00	0.24	0.62	0.00
4	0.28	0.60	0.01	0.01	0.94	0.00

Note: The naive and transparency tests are OLS-based Wald tests that $\gamma = 0$ in two versions of $y^c_{t,t+h} - y_{t+h} = \gamma'_{y,h} x_t + \Sigma^h_{j=0} \beta_{h,j} \delta(c,f)_{t,h-j} + \varepsilon^y_{t,t+h}$ where the y variable is either output growth or inflation for the U.S. or U.K. as noted in the title of the table. The naive case omits the terms in the summation from the regression. The W columns give the value of the Wald statistic; the p-val. column gives the nominal p-value from the asymptotic χ^2 distribution of the statistic, and the ΔR^2 column gives the change in the uncentered R^2 from including the x terms. The R1 and R2 sets of supplementary variables, x, are described in the text.

Table 3a. Single-equation IV forecast efficiency tests: U.S. output growth

h	r	S	p-val.	robust	rank	p-val.
I1 instruments						
0	-	0.60	0.44	0.90	22.30	0.00
0	0	7.06	0.01	0.07		
1	-	1.53	0.22	0.82	7.92	0.02
1	0	3.04	0.08	0.55		
1	1	3.04	0.08	0.55		
2	-	0.67	0.41	0.98	0.28	0.87
2	0	3.79	0.05	0.58		
2	1	5.02	0.03	0.41		
2	2	6.02	0.01	0.30		
3	-	2.79	0.10	0.84	0.56	0.76
3	0	2.26	0.13	0.89		
3	1	4.86	0.03	0.56		
3	2	8.18	0.00	0.23		
4	-	0.46	0.50	0.99	0.11	0.94
4	0	0.77	0.38	0.99		
4	1	0.96	0.33	0.99		
4	2	1.68	0.19	0.98		
I2 instruments						
0	-	1.91	0.59	0.86	15.71	0.00
0	0	7.30	0.06	0.20		
1	-	3.82	0.28	0.70	7.67	0.10
1	0	6.96	0.07	0.32		
1	1	6.96	0.07	0.32		
2	-	0.68	0.71	0.99	0.28	0.96
2	0	7.35	0.03	0.29		
2	1	8.50	0.01	0.20		
2	2	8.51	0.01	0.20		
3	-	2.77	0.25	0.91	0.47	0.93
3	0	2.97	0.23	0.89		
3	1	3.67	0.16	0.82		
3	2	9.32	0.01	0.23		
4	-	0.46	0.50	0.99	0.11	0.94
4	0	0.77	0.38	0.99		
4	1	0.96	0.33	0.99		
4	2	1.68	0.19	0.98		

Note: See notes at end of Table 3d.

Table 3b. Single-equation IV forecast efficiency tests: U.S. inflation

h	r	S	p-val.	robust	rank	p-val.
		I1 instruments				
0	-	0.35	0.56	0.95	19.98	0.00
0	0	0.35	0.56	0.95		
1	-	0.96	0.33	0.92	9.04	0.01
1	0	0.96	0.33	0.92		
1	1	0.96	0.33	0.92		
2	-	2.30	0.13	0.81	0.98	0.61
2	0	3.05	0.08	0.69		
2	1	3.40	0.07	0.64		
2	2	3.86	0.05	0.57		
3	-	0.50	0.48	0.99	0.64	0.73
3	0	0.50	0.48	0.99		
3	1	0.50	0.48	0.99		
3	2	1.69	0.19	0.95		
4	-	0.73	0.39	0.99	0.17	0.92
4	0	0.73	0.39	0.99		
4	1	1.02	0.31	0.99		
4	2	1.23	0.27	0.99		
		I2 instruments				
0	-	0.94	0.92	0.99	10.81	0.06
0	0	0.94	0.92	0.99		
1	-	3.06	0.38	0.80	6.97	0.14
1	0	8.21	0.04	0.22		
1	1	3.06	0.38	0.80		
2	-	2.30	0.13	0.81	0.98	0.61
2	0	3.05	0.08	0.69		
2	1	3.40	0.07	0.64		
2	2	3.86	0.05	0.57		
3	-	0.50	0.48	0.99	0.64	0.73
3	0	0.50	0.48	0.99		
3	1	0.50	0.48	0.99		
3	2	1.69	0.19	0.95		
4	-	0.32	0.85	0.99	0.16	0.98
4	0	1.08	0.58	0.99		
4	1	1.22	0.54	0.99		
4	2	1.26	0.53	0.99		

Note: See notes at end of Table 3d.

Table 3c. Single-equation IV forecast efficiency tests: U.K. output growth

h	r	S	p-val.	robust	rank	p-val.
			I1 instruments			
0	-	1.82	0.18	0.61	3.61	0.16
0	0	2.81	0.09	0.42		
1	-	1.97	0.16	0.74	1.33	0.51
1	0	2.11	0.15	0.71		
1	1	2.24	0.13	0.69		
2	-	1.53	0.22	0.91	0.31	0.86
2	0	1.75	0.19	0.88		
2	1	1.75	0.19	0.88		
2	2	2.23	0.14	0.82		
3	-	1.02	0.31	0.98	0.00	0.99
3	0	1.02	0.31	0.98		
3	1	0.06	0.81	0.99		
3	2	3.46	0.06	0.75		
4	-	0.10	0.75	0.99	0.03	0.99
4	0	0.28	0.60	0.99		
4	1	1.83	0.18	0.97		
4	2	3.22	0.07	0.86		
			I2 instruments			
0	-	1.99	0.57	0.85	6.08	0.19
0	0	3.84	0.28	0.57		
1	-	2.06	0.36	0.84	1.29	0.73
1	0	2.16	0.34	0.83		
1	1	2.31	0.31	0.80		
2	-	1.53	0.22	0.91	0.31	0.86
2	0	1.75	0.19	0.88		
2	1	1.75	0.19	0.88		
2	2	2.23	0.14	0.82		
3	-	0.35	0.84	0.99	0.04	0.99
3	0	0.35	0.84	0.99		
3	1	0.35	0.84	0.99		
3	2	3.12	0.21	0.87		
4	-	0.10	0.75	0.99	0.03	0.99
4	0	0.28	0.60	0.99		
4	1	1.83	0.18	0.97		
4	2	3.22	0.07	0.86		

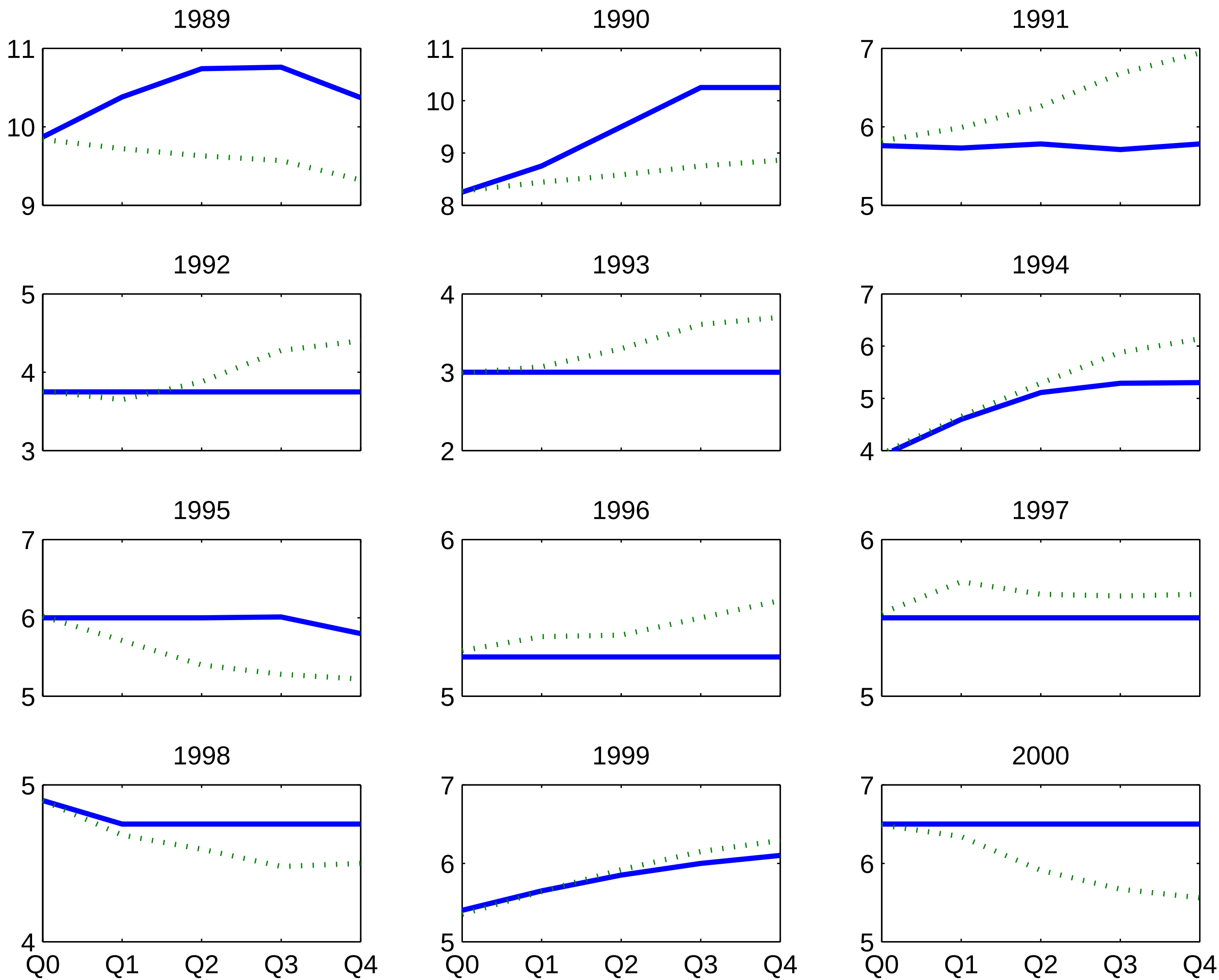
Note: See notes at end of Table 3d.

Table 3d. Single-equation IV forecast efficiency tests: U.K. inflation

h	r	S	p-val.	robust	rank	p-val.
			I1 instruments			
0	-	0.02	0.90	0.99	0.15	0.93
0	0	0.02	0.90	0.99		
1	-	0.02	0.89	0.99	0.10	0.95
1	0	0.02	0.89	0.99		
1	1	0.13	0.72	0.99		
2	-	0.15	0.70	0.99	0.11	0.95
2	0	2.42	0.12	0.79		
2	1	2.29	0.13	0.81		
2	2	2.59	0.11	0.76		
3	-	0.03	0.87	0.99	0.00	0.99
3	0	0.03	0.87	0.99		
3	1	0.05	0.83	0.99		
3	2	2.93	0.09	0.82		
4	-	0.02	0.88	0.99	0.01	0.99
4	0	0.08	0.78	0.99		
4	1	0.12	0.73	0.99		
4	2	0.14	0.71	0.99		
			I2 instruments			
0	-	2.73	0.60	0.84	3.01	0.70
0	0	2.76	0.60	0.84		
1	-	1.70	0.64	0.95	2.31	0.68
1	0	3.02	0.39	0.81		
1	1	3.02	0.39	0.81		
2	-	0.15	0.70	0.99	0.11	0.95
2	0	2.42	0.12	0.79		
2	1	2.29	0.13	0.81		
2	2	2.59	0.11	0.76		
3	-	0.03	0.87	0.99	0.00	0.99
3	0	0.03	0.87	0.99		
3	1	0.05	0.83	0.99		
3	2	2.93	0.09	0.82		
4	-	0.02	0.88	0.99	0.01	0.99
4	0	0.08	0.78	0.99		
4	1	0.12	0.73	0.99		
4	2	0.14	0.71	0.99		

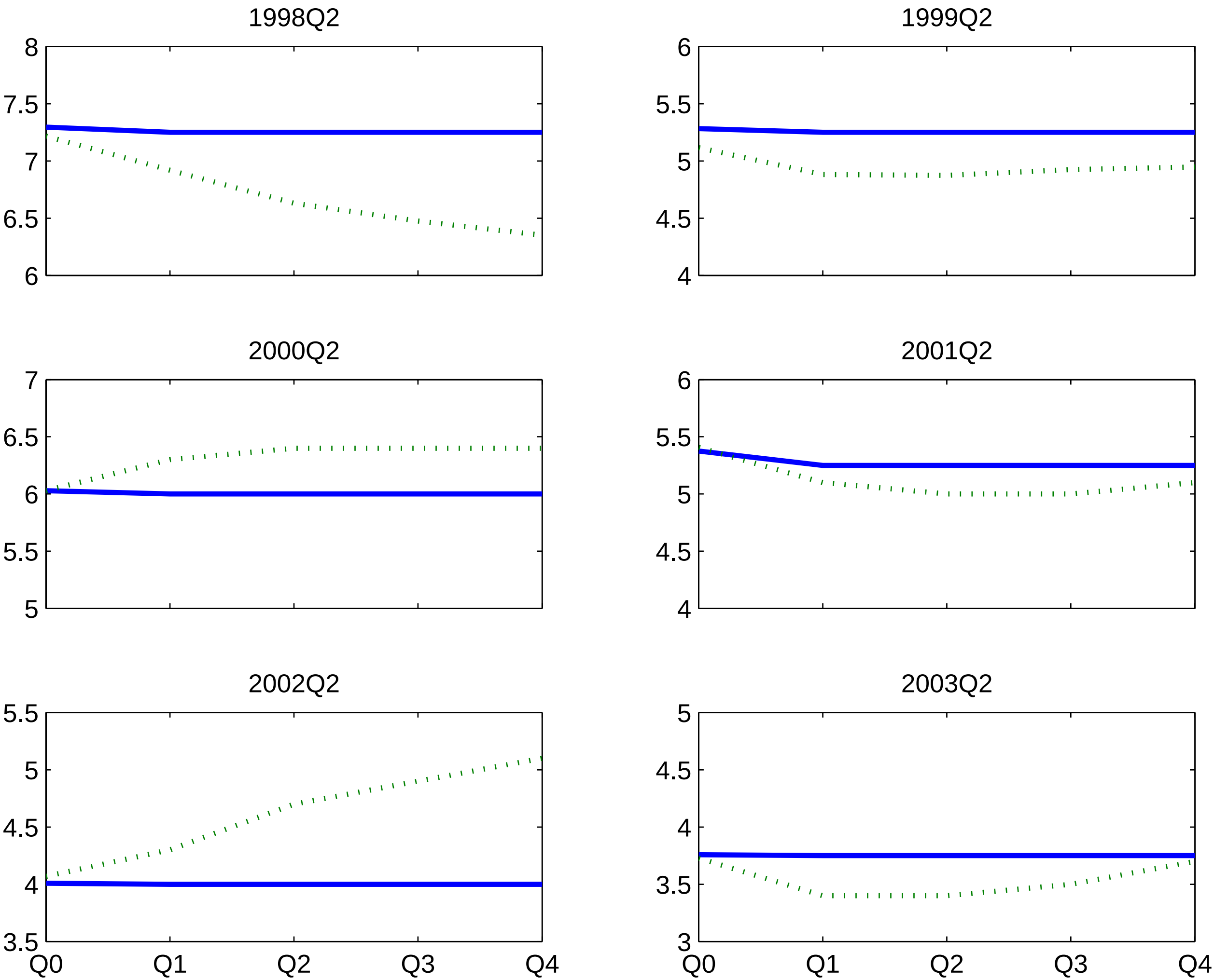
Note: Each row reports results of a continuous-updating GMM estimate of a single equation of the form, $y^c_{t,t+h} - y_{t+h} = \sum_{j=0}^{h} \beta_{h,j} \delta(c,a)_{t,h-j} + u_{t,t+h}$ The βs are constrained as in A2 through horizon r. The columns S, p-val., and *robust* report the Sargan test of overidentifying restrictions and standard and robust p values, respectively. The column labelled "rank" and the following p-val. column give the Cragg and Donald test statistic and p-value for the test of the null hypothesis that the rank condition for identification fails. If the rank condition fails to hold, the robust p-values of the Sargan test will be more reliable. The I1 and I2 instrument sets are described in the text.

Fig. 1: Conditional interest rate path in the final Greenbook each year and concurrent futures path



The Greenbook conditioning assumption is given by the solid line, the futures path is given by the dashed line.

Fig. 2: Conditional interest rate path in selected BoE inflation reports and concurrent futures path



The BoE conditioning assumption is given by the solid line, the futures path is given by the dashed line.