Evaluating Conditional Forecasts from Vector Autoregressions *

Todd E. Clark
Federal Reserve Bank of Cleveland

Michael W. McCracken
Federal Reserve Bank of St. Louis

March 2014
(preliminary and incomplete)

Abstract

Many forecasts are conditional in nature. For example, a number of central banks routinely report forecasts conditional on particular paths of policy instruments. Even though conditional forecasting is common, there has been little work on or with methods for evaluating conditional forecasts. This paper provides analytical, Monte Carlo, and empirical evidence on tests of predictive ability for conditional forecasts from estimated models. In the empirical analysis, we consider forecasts of growth, unemployment, and inflation from VAR models, based on conditions on the short-term interest rate. Throughout the analysis, we focus on tests of bias, efficiency, and equal accuracy applied to conditional forecasts from VAR models.

JEL Nos.: C53, C52, C12, C32
Keywords: Prediction, forecasting, out-of-sample

* *Clark*: Economic Research Dept.; Federal Reserve Bank of Cleveland; P.O. Box 6387; Cleveland, OH 44101; todd.clark@clev.frb.org. *McCracken* (corresponding author): Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. The authors gratefully acknowledged helpful comments from .... The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, Federal Reserve Bank of St. Louis, Federal Reserve System, or any of its staff.

1 Introduction

Since the seminal work of Litterman (1986), vector autoregressions (VARs) have been widely known to be useful for out-of-sample forecasting. In many applications, the forecasts of interest are unconditional. Clark and McCracken (2013b) review methods for evaluating such forecasts.

VARs are also commonly used to construct conditional forecasts. In such applications, the models are used to predict variables such as GDP growth and inflation conditional on, e.g., an assumed path of monetary policy variables or an assumed path of oil prices. Examples of VAR forecasts conditional on policy paths include Sims (1982), Doan, Litterman, and Sims (1984), and Meyer and Zaman (2013). Giannone, et al. (2010) use VARs to construct forecasts of inflation conditional on paths for oil and other price indicators. Schorfheide and Song (2013) and Aastveit, et al. (2014) use VARs to produce multi-step forecasts of growth, inflation, and other macroeconomic variables conditional on current-quarter forecasts obtained from other, judgmental sources (the Federal Reserve's Greenbook for the former and Survey of Professional Forecasters for the latter). In all cases, the conditional forecasts of the variables of interest (e.g., growth and inflation) may be very different from (unconditional) forecasts associated with the most likely path of the conditioning variables if the most likely path of the conditioning variables differs from the conditioning path.

Events since the severe recession of 2007-2009 may have increased the practical use of conditional forecasts. For some time, short-term interest rates have been stuck near the zero lower bound. Moreover, some central banks, including the Federal Reserve, have provided forward guidance that indicates policy rates are likely to remain unchanged for a considerable period of time. In these circumstances, a forecaster may find it valuable to condition his or her projection for the macroeconomy on a path of very low, positive interest rates, to both avoid forecasts of negative interest rates over the forecast horizon and reflect the guidance from policymakers.

In light of common need for conditional forecasts, the attention paid to them, and their use in conveying the future stance of monetary policy, one would like to have a feel for their quality. Unfortunately, evaluating the quality of these forecasts is more complex than evaluating unconditional forecasts. To see why, consider a very simple example of a conditional forecast, in which we forecast inflation (y_t) conditioned on a path for the federal funds rate (x_t). Suppose that the data-generating process for inflation and the funds rate is a zero-mean stationary VAR(1) taking the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ v_t \end{pmatrix},$$

with i.i.d. N(0,1) errors with contemporaneous correlation ρ. Following the approach taken in Doan, Litterman, and Sims (1984), conditional on a path for the funds rate of $x_{t+2} = x_{t+1} = x_t$, the minimum mean square error (MSE) one- and two-step ahead forecasts of y_t are as follows:

$$\begin{aligned}
\hat{y}^c_{t+2} &= \hat{y}^u_{t+2} + \rho(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) = \hat{y}^u_{t+2} + (1-c^2)\rho x_t \\
\hat{y}^c_{t+1} &= \hat{y}^u_{t+1} + \rho(a\rho - c)(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) + \rho(\hat{x}^c_{t+1} - \hat{x}^u_{t+1}) \\
&= \hat{y}^u_{t+1} + (\rho(a\rho - c)(1-c^2) + \rho(1-c))x_t,
\end{aligned}$$

where the superscripts c and u denote conditional and unconditional forecasts, respectively. In both cases the conditional forecasts of y are comprised of the standard, unconditional MSE-optimal forecast plus additional terms that capture the hypothetical future "appropriate monetary policy" as captured in this example by the constant future path of the x variable.

With the conditional forecast errors given by $\hat{\varepsilon}^c_{t+2} = y_{t+2} - \hat{y}^c_{t+2}$ and $\hat{\varepsilon}^c_{t+1} = y_{t+1} - \hat{y}^c_{t+1}$, it is immediately clear that some of the most basic measures of the quality of unconditional forecasts fail when applied directly to conditional forecasts. For example, the one-step ahead forecast error is serially correlated of infinite order so long as $c \neq 0$. In addition, both forecast errors are correlated with the forecasts themselves and hence fail the most basic standards of a Mincer-Zarnowitz efficiency regression. Finally, insofar as the conditioning is hypothetical, there is no *a priori* reason to suspect that the conditional forecast is accurate in an MSE sense. As such, much of the literature on tests of predictability, including Diebold and Mariano (1995), West and McCracken (1998), Clark and McCracken (2001, 2005, 2014), and Clark and West (2006, 2007), are not applicable when the forecasts are conditional.

Accordingly, in this paper, we develop and apply methods for the evaluation of conditional forecasts from VARs. More specifically, we provide analytical, Monte Carlo, and empirical evidence on tests of predictive ability for conditional forecasts from estimated models. In the empirical analysis, we consider forecasts of growth, unemployment, and inflation from VAR models, based on conditions on the short-term interest rate. Throughout the analysis, we consider tests of bias and efficiency applied to conditional forecasts from

VAR models. We also consider some new tests of equal accuracy, of conditional versus unconditional forecasts.

While we focus on forecasts from time series models, conditional forecasts are routine in professional forecasting. The forecasts from the Federal Reserve's Open Market Committee that have been published since 1979 are produced conditional on "appropriate monetary policy" as viewed by the respective individual members of the FOMC. In effect, the individual member of the FOMC is asked to produce a point forecast of, say, inflation over the next year given that the federal funds rate takes values over the coming year that are appropriate in the eyes of that individual FOMC member. Similarly, staff forecasts from the Federal Reserve Board are typically conditional on a future path of the funds rate as well as other variables.[1] The Bank of England, Riksbank, Norges Bank, and ECB all produce and release forecasts to the public that are conditional on a hypothetical future stance of monetary policy in one way or another.

While explicitly conditional, the forecasts released by the FOMC and policymakers at other central banks seem to be regularly analyzed by the public and financial markets without taking account of the conditional nature of the forecasts. The same can also be said for academics. For example, Romer and Romer (2008) use standard MSE-based encompassing regressions to infer that Greenbook forecasts have an informational advantage over the corresponding FOMC forecasts. Patton and Timmermann (2012) apply newly developed tests of MSE-based forecast optimality to Greenbook forecasts and find evidence against forecast rationality. Gavin (2003), Gavin and Mandel (2003), Gavin and Pande (2008), Joutz and Stekler (2000), and Reifschneider and Tulip (2007) each evaluate the quality of FOMC forecasts in the context of MSEs.

However, two previous studies have developed methods for the evaluation of some form of conditional forecasts. Motivated by an interest in evaluating the efficiency of Greenbook forecasts, Faust and Wright (2008) develop a regression-based test of predictive ability that accounts for the conditioning of the Greenbook forecasts on a pre-specified path of the federal funds rate over the forecast horizon. For the purpose of evaluating forecasts from DSGE models, Herbst and Schorfheide (2012) develop Bayesian methods for check the accuracy of point and density forecasts. More specifically, Herbst and Schorfheide consider the Bayesian tool of posterior predictive checks and forecasts of each variable conditioned

[1] These forecasts are often referenced as Greenbook forecasts.

on the actual future path of another, selected variable. Our paper differs from these in that we focus on conditional forecasts from VARs and frequentist inference.

The remainder of the paper proceeds as follows. Section 2 describes the two different approaches to conditional forecasting that we consider. Section 3 provides our theory results. Section 4 presents Monte Carlo evidence on the finite-sample accuracy of our proposed methods for evaluating conditional forecasts from VARs. Section presents a practical application, to macroeconomic forecasts conditioned on fixed interest rate paths. Section 6 concludes.

2 Conditional Forecasting Approaches

In generating and evaluating conditional forecasts, we consider two approaches that seem to be common in VAR and DSGE forecasting. The theoretical results on evaluation in the next section apply to both approaches.

The standard in VAR forecasting is based on the textbook problem of conditional projection, as could be handled with a state space formulation of the VAR and the Kalman filter and smoother (see, e.g., Clarida and Coyle (1984) or Giannone, et al (2012)). The conditions on the variables of interest are contained in the measurement vector and equation; the data vector of the VAR is the state vector of the transition equation. The projection problem is one of predicting the state vector given the measurements (conditions). Doan, Litterman, and Sims (1984) developed an alternative approach to solving this formulation of the conditional forecasting problem, which consists of solving a least squares problem to pick the shocks needed to satisfy the conditions. In the context of conditioning on a policy path, this approach to conditional forecasting can be seen as consisting of the following: determining the set of shocks to the VAR that, by a least squares metric, best meet the conditions on the policy rate. In practice, this approach may mean that, in a given episode, an unchanged path of the policy rate could, for example be due to shocks to output. Under this approach, the conditional forecasts are **not** dependent on the identification of structural shocks in the VAR. Accordingly, through the remainder of the paper, we will refer to forecasts obtained under this approach as *conditional-reduced form* forecasts. However, as we note elsewhere, under this approach, the forecast for each period can be affected by the imposition of conditions at all periods. For example, if we impose conditions for two periods, the forecast for period $t+1$ will be affected by the conditions on both period $t+1$

and $t + 2$.

In the interest of simplicity, in our implementation of conditional-reduced form forecasting, we abstract from the enhancements developed in Waggoner and Zha (1999).[2] In our implementation, as is typical, we form the posterior distribution of VAR parameters without taking the forecast conditions into account. Waggoner and Zha develop a Gibbs sampling algorithm that provides the exact finite-sample distribution of the conditional forecasts, by taking the conditions into account when sampling the VAR coefficients. Our reasons for abstracting from their extension are primarily computational. With the size of the large model we use, their Gibbs sampling algorithm would be extremely slow, due to computations of an extremely large VAR coefficient variance-covariance matrix. Moreover, based on a check we performed with our small model, their approach to conditional forecasting doesn't seem to affect the conditional forecasts much.

In DSGE forecasting, the more standard approach for achieving conditions on the policy path rests on feeding in structural shocks to monetary policy needed to hit the policy path (see, e.g., Del Negro and Schorfheide (2013)). We use the following approach to implementing such an approach with our VARs. Under this approach, the scheme for identifying policy shocks does matter for the conditional forecasts. With our focus on conditioning on the policy path, it is the identification of monetary policy that matters. Accordingly, we refer to the forecasts we obtain under this approach as *conditional-policy shock* forecasts. Following common precedent with models such as ours (including, e.g., Bernanke and Blinder (1992), Christiano, Eichenbaum, and Evans (1996), and Giannone, Lenza, and Primiceri (2012)), we rely on identification of monetary policy via a recursive ordering of model innovations. Conditional forecasting under this approach involves the steps listed below. Note that, under this approach, as applied with a VAR, the forecast in period $t + \tau$ is not affected by conditions imposed on future periods $t + \tau + 1$, etc.

1. Using a model estimated with data through period t, at period $t + 1$, draw a set of reduced form shocks to the VAR and form 1-step ahead forecasts.

2. Compute the structural shock needed to make the federal funds rate equal the conditioning rate. Using A = the Choleski factor of the variance-covariance matrix of structural shocks, re-compute the reduced form shocks as $(I + A)$ times the original draw of reduced-form shocks.

[2] Jarocinski (2010) simplifies the formula of Waggoner and Zha (1999) and develops an extension.

3. Re-compute 1-step ahead forecasts.

4. Move forward a step, and repeat steps 1-3 using the preceding period's forecast as data for the preceding period.

5. Continue through period $t + \tau$ to obtain a draw of forecasts from $t + 1$ through $t + \tau$.

In the cases in which we forecast without simulation of the model, the vector of reduced form shocks in step 1 is set to 0, and the single time series of forecast replaces draws of time series of forecasts.

3 Theory for Regression-Based Tests for Bias and Efficiency

In this section we delineate our regression-based approach to inference in the context of the extant literature on regression-based tests of predictive ability — two of which are particularly relevant for this paper. The first is West and McCracken (1998). That paper provides primitive conditions under which the t-statistics associated with coefficients from simple-to-implement OLS regressions, along with standard normal critical values, provide asymptotically valid inference for tests of bias (zero mean prediction error), efficiency, encompassing, and zero first order serial correlation. Each of these regressions take the form

$$\hat{\varepsilon}_{t+\tau} = \hat{z}_t'\gamma + \xi_{t+\tau},$$

where $\hat{y}_{t+\tau}$ denotes a $\tau-$step ahead forecast made at time t, $\hat{\varepsilon}_{t+\tau} = y_{t+\tau} - \hat{y}_{t+\tau}$ denotes the forecast error, and $\hat{z}_t$ denotes an appropriately chosen vector of instruments determined by the null hypothesis of interest. For example, $\hat{z}_t = 1$ for tests of zero bias while $\hat{z}_t = (1, \hat{y}_{t+\tau})'$ or $\hat{z}_t = \hat{y}_{t+\tau}$ for tests of efficiency.

Note that, while we focus on conditional forecasts, in our Monte Carlo and empirical analyses we also include results for unconditional forecasts. In the case of unconditional forecasts from a recursive estimation scheme, the results of West (1996) and West and McCracken (1998) imply that bias and efficiency tests can be computed from standard regression output without a need to correct standard errors for the effects of model estimation error.

As to the conditional forecasts of interest, while the results in West and McCracken (1998) are broad enough to be applied to unconditional forecasts from a wide range of parametric models — including VARs — the assumptions of West and McCracken effectively eliminate application to conditional forecasts. In the context of the tests we consider here, Assumptions (1b) and (1c) of that paper require that the forecast errors are uncorrelated with the forecasts and more generally, anything in the time t information set. As we saw in the simple example from the introduction, even well-constructed conditional forecasts will be correlated with the associated forecast error.

This point has not been completely overlooked in the literature and, in particular, is addressed in a narrower context in Faust and Wright (2008). Faust and Wright are interested in evaluating the efficiency of Greenbook forecasts constructed by the staff at the Federal Reserve Board of Governors. Since these forecasts are constructed conditional on a pre-specified path of the federal funds rate over the forecast horizon, the results in West and McCracken (1998) are not applicable. Faust and Wright therefore develop a new regression-based test of predictive ability that accounts for the conditioning. Let $\hat{y}^c_{t+\tau}$ denote the conditional forecast of interest (for GDP growth, for example), $\hat{\varepsilon}^c_{t+\tau}$ denote the associated forecast error, $\hat{x}^c_{t+\tau-i}$ denote a condition imposed on the interest rate forecast, and $\hat{x}^u_{t+\tau-i}$ denote an unconditional forecast of the interest rate. Their basic regression takes the form

$$\begin{aligned}\hat{\varepsilon}^c_{t+\tau} &= \alpha_0 + \alpha_1 \hat{y}^c_{t+\tau} + \sum_{i=1}^{\tau} \delta_j (\hat{x}^c_{t+\tau-i} - \hat{x}^u_{t+\tau-i}) + \xi_{t+\tau} \\ &= \hat{z}'_t \gamma + \xi_{t+\tau}.\end{aligned}$$

While this regression is similar to that in West and McCracken (1998), it differs via the augmenting variables $\hat{x}^c_{t+\tau-i} - \hat{x}^u_{t+\tau-i}$, $i = 1, ..., \tau$, which represent differences in the unconditional and conditional forecasts of the interest rate. The logic of the Faust-Wright regression is based on a modification of the typical null of forecast efficiency. The question of interest is no longer whether the forecast error $\hat{\varepsilon}^c_{t+\tau}$ is uncorrelated with the forecast $\hat{y}^c_{t+\tau}$ but rather whether the conditional forecast error is uncorrelated with $\hat{y}^c_{t+\tau}$ *given* the augmenting variables $\hat{x}^c_{t+\tau-i} - \hat{x}^u_{t+\tau-i}$ $i = 1, ..., \tau$. As Faust and Wright discuss, this particular test can only be conducted if the unconditional forecast of the interest rate is available. For the case in which the unconditional forecast is not available, they develop a related test regression that uses the actual interest rate in place of the unconditional forecast and instruments for the differential between the actual interest rate and conditional forecast.

The augmented regression approach of Faust and Wright (2008) permits a test of forecast efficiency using conditional forecasts but suffers from a few weaknesses. First, they assume that the conditional forecasts are constructed using a policy shock framework akin to that used in the DSGE literature. As a practical matter this implies that, for example, the time t one-step ahead forecast of y_{t+1} is not influenced by the conditional path of the funds rate at times $t+2$, $t+3$, ..., $t+\tau$. This is in contrast to the reduced form methodology pioneered in Doan, Litterman, and Sims (1984), in which the time t forecasts of y_{t+1}, y_{t+2},...,$y_{t+\tau}$ are constructed jointly accounting for the entire conditional path of the funds rate. Second, the results in Faust and Wright (2008) abstract from the effect that parameter estimation error has on the asymptotic distribution of their test statistic. To be fair, they do so because the model behind the Greenbook forecast errors is unknown and is at least in part influenced by the judgement of the Federal Reserve Board staff. As such there is no way to use the results developed in West (1996) to correct the standard errors for estimation error.

In this paper we extend the results of West and McCracken (1998) and Faust and Wright (2008) to an environment in which a known parametric forecasting model is being used to construct conditional forecasts. We consider a range of evaluation tools, including tests for the bias or efficiency of conditional forecasts and tests for equality of the accuracy of the conditional and unconditional forecasts. In each case we establish that a regression coefficient $\hat{\alpha}_1$ is asymptotically normal with an asymptotic variance that has a component due to estimation error. We then discuss two methods for inference: one in which the standard errors are calculated directly and standard normal critical values are used and a second in which a parametric bootstrap is used to estimate the critical values.

3.1 Tests of bias and efficiency

More specifically, to construct tests of bias and efficiency, respectively, for conditional forecasts, we consider using the regressions

$$\hat{\varepsilon}^c_{t+\tau} = \alpha_1 + \xi_{t+\tau} \tag{1}$$

and

$$\begin{aligned} \hat{\varepsilon}^c_{t+\tau} &= \alpha_0 + \alpha_1 \hat{y}^c_{t+\tau} + \delta(\hat{y}^u_{t+\tau} - \hat{y}^c_{t+\tau}) + \xi_{t+\tau} \\ &= \hat{z}_t'\gamma + \xi_{t+\tau}, \end{aligned} \tag{2}$$

where, in the latter regression, $\hat{y}^u_{t+\tau}$ denotes the unconditional forecast of $y_{t+\tau}$.

Note that our efficiency regression is not so very different from that proposed by Faust and Wright (2008) in situations where $\hat{y}^u_{t+\tau} - \hat{y}^c_{t+\tau}$ is just a linear combination of $\hat{x}^c_{t+\tau-i} - \hat{x}^u_{t+\tau-i}$, $i = 1, ..., \tau$ but is more general. Because we have access to the parametric model being used to construct forecasts we necessarily also have access to the model-implied unconditional forecasts. While perhaps not immediately obvious, this simple distinction allows us to avoid assuming that the conditional forecasts are constructed via the policy shock approach: both the policy shock and the reduced form approach of Doan et al. (1984) can be used. In addition, knowing the model allows us to account for estimation error via the theoretical results in West (1996).

Given the above two regressions, for testing bias and efficiency of conditional forecasts we wish to conduct inference on the OLS estimated regression coefficients $\hat{\alpha}_1$ under the null hypothesis that H_0: $\alpha_1 = 0$. In contrast to the case of unconditional forecasts, interpretation of this null can be a bit confusing when conditional forecasts are being used. To see this, first consider the simple example in the introduction. The one-step ahead conditional forecast error takes the form $\hat{\varepsilon}^c_{t+1} = \hat{\varepsilon}^u_{t+1} - \rho(a\rho - c)(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) - \rho(\hat{x}^c_{t+1} - \hat{x}^u_{t+1})$. Taking expectations and evaluating at the population parameters of the VAR, it is obvious that α_1 is a linear combination of Ee_{t+2} and the expected deviation between the unconditional and conditional forecasts of x. As such, a test of the null of zero bias can be seen as a joint test that the unconditional forecasts is unbiased and that, on average, the conditioning path for x_{t+1} and x_{t+2} is "modest." Here, the term "modest", defined in Doan, Litterman, and Sims (1984) and discussed in detail by Leeper and Zha (2003), is used to mean that on average the conditioning path is not unreasonable (extremely unlikely?) by historical standards, and hence $E(\hat{x}^c_{t+j} - \hat{x}^u_{t+j}) = 0$, $j = 1, 2$.

Now consider the test of efficiency in equation (2). The population value of the coefficient α_1 is the second element of the vector $\gamma = (Ez_t z_t')^{-1}(Ez_t \varepsilon^c_{t+\tau})$. Unfortunately this does not simplify to a clean representation as in the case of zero bias. Nevertheless, in the context of the simple example above we are able to establish that the null holds in two readily interpretable cases. First, if the conditioning events $\hat{x}^c_{t+1}$ and $\hat{x}^c_{t+2}$ are known at time t (e.g., $\hat{x}^c_{t+j} = x_t$, $j = 1, 2$), $\alpha_1 = 0$ so long as the unconditional model is efficient and in particular so long as ε^u_{t+1} is uncorrelated with $\hat{y}^u_{t+1}$ and $\hat{x}^c_{t+j} = x_t$. In this case the test of efficiency of the conditional forecast is just an augmented variant of the standard Mincer-Zarnowitz regression-based test of efficiency. If, however, the conditioning events $\hat{x}^c_{t+1}$ and $\hat{x}^c_{t+2}$ are not known at time t (perhaps arising from a stress testing exercise?), α_1 can still be zero if, in addition to unconditional efficiency, the degree of immodesty is uncorrelated with time t information. Specifically, given an efficient unconditional forecasting model we find that $\alpha_1 = 0$ so long as both $\hat{x}^c_{t+2} - \hat{x}^u_{t+2}$ and $\hat{x}^c_{t+1} - \hat{x}^u_{t+1}$ are uncorrelated with $\hat{y}^u_{t+1}$. In this case the test of efficiency of the conditional forecast becomes a joint test of efficiency of the unconditional forecast and a test of whether time t information can be used to predict the degree of immodesty of the conditioning paths.

3.2 Equal accuracy

We consider three distinct approaches to testing whether the unconditional and conditional forecasts of the variable of interest $y_{t+\tau}$ are equally accurate under quadratic loss. While conditional forecasts are not necessarily intended to be accurate in an MSE-sense, they could be more accurate than unconditional forecasts if the condition reflects a superior judgmental forecast or other information about, for example, the likely path of policy. In today's circumstances, such information could come from verbal forward guidance provided by a central bank. Analytically, to see how such a hypothesis might arise, we again return to the example from the introduction and note that

$$\begin{aligned}\hat{\varepsilon}^c_{t+1} &= \hat{\varepsilon}^u_{t+1} - \rho(a\rho - c)(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) - \rho(\hat{x}^c_{t+1} - \hat{x}^u_{t+1}) \\ &= \hat{\varepsilon}^u_{t+1} - (\hat{\varepsilon}^u_{t+1} - \hat{\varepsilon}^c_{t+1}),\end{aligned}$$

which implies

$$(\hat{\varepsilon}^c_{t+1})^2 - (\hat{\varepsilon}^u_{t+1})^2 = 2\hat{\varepsilon}^u_{t+1}(\hat{\varepsilon}^c_{t+1} - \hat{\varepsilon}^u_{t+1}) + (\hat{\varepsilon}^c_{t+1} - \hat{\varepsilon}^u_{t+1})^2.$$

The null of equal accuracy $E(\varepsilon^c_{t+1})^2 - (\varepsilon^u_{t+1})^2 = 0$ obviously requires $2E\varepsilon^u_{t+1}(\varepsilon^c_{t+1} - \varepsilon^u_{t+1}) + E(\varepsilon^c_{t+1} - \varepsilon^u_{t+1})^2 = 0$. At first blush this might seem an odd hypothesis but it can occur in the two cases discussed in the context of the tests of bias and efficiency. If the conditioning events $\hat{x}^c_{t+1}$ and $\hat{x}^c_{t+2}$ are known at time t (e.g., $\hat{x}^c_{t+j} = x_t$, $j = 1, 2$), the null can hold so long as $E\varepsilon^u_{t+1}(\varepsilon^c_{t+1} - \varepsilon^u_{t+1}) = -0.5E(\varepsilon^c_{t+1} - \varepsilon^u_{t+1})^2$. Since this requires that the unconditional model is inefficient, a test of equal accuracy can be seen as a test of the degree of efficiency of the unconditional model relative to the magnitude of the degree of immodesty measured by $E(\varepsilon^c_{t+1} - \varepsilon^u_{t+1})^2$. In contrast, if the conditioning events are not known at time t, equal accuracy can arise even if the unconditional model is efficient so long as the degree of immodesty $\varepsilon^c_{t+1} - \varepsilon^u_{t+1}$ is negatively correlated with ε^u_{t+1}.

For the test of equal accuracy $E(\varepsilon^c_{t+\tau})^2 - (\varepsilon^u_{t+\tau})^2 = 0$, we first consider a direct, Diebold and Mariano (1995)–West (1996) test of equal MSE, applied to the unconditional and conditional forecast errors, $\hat{\varepsilon}^u_{t+\tau}$ and $\hat{\varepsilon}^c_{t+\tau}$. One problem with this approach is that a rejection tells us nothing about why the tests are equally accurate. A rejection can arise due to: (i) inefficiency in the unconditional model, (ii) very immodest conditioning, or (iii) predictable immodesty.

To get around this issue, we also consider two regression-based tests of equal accuracy that allow us to disentangle the sources of differences in forecast accuracy. The first builds on the observation that the forecasts are equally accurate — that is, $E(\varepsilon_{t+\tau}^c)^2 - E(\varepsilon_{t+\tau}^u)^2 = 0$ — if and only if

$$\frac{E\varepsilon_{t+\tau}^u(\varepsilon_{t+\tau}^c - \varepsilon_{t+\tau}^u)}{E(\varepsilon_{t+\tau}^c - \varepsilon_{t+\tau}^u)^2} = -\frac{1}{2}.$$

To test this restriction, we can use OLS to estimate the regression

$$\hat{\varepsilon}_{t+\tau}^u = \alpha_1(\hat{\varepsilon}_{t+\tau}^c - \hat{\varepsilon}_{t+\tau}^u) + \eta_{t+\tau} \tag{3}$$

and use the t-statistic associated with $\hat{\alpha}_1 - (-1/2)$ to test the equal accuracy null of H_0:$\alpha_1 = -1/2$. But to help with interpretation of the test result, we can also consider a regression augmented with a constant:

$$\hat{\varepsilon}_{t+\tau}^u = \alpha_0 + \alpha_1(\hat{\varepsilon}_{t+\tau}^c - \hat{\varepsilon}_{t+\tau}^u) + \eta_{t+\tau} \tag{4}$$

In this case, we can test the joint null of H_0: $\alpha_0 = 0$, $\alpha_1 = -1/2$. The joint restriction will be satisfied if the unconditional forecast is unbiased, the expected difference in unconditional and conditional forecast errors is 0, and the slope coefficient $\alpha_1 = -1/2$. Accordingly, this second regression can be seen as a joint test of the null that the unconditional model is unbiased, the conditioning is modest in expectation, and the models are equally accurate because model inefficiency (or conditioning on a future event) balances with the expected square of immodesty.

Our second regression-based approach to testing equal accuracy builds on one developed in Granger and Newbold (1977) that takes the form

$$\hat{\varepsilon}_{t+\tau}^c - \hat{\varepsilon}_{t+\tau}^u = \alpha_0 + \alpha_1(\hat{\varepsilon}_{t+\tau}^c + \hat{\varepsilon}_{t+\tau}^u) + \eta_{t+\tau}. \tag{5}$$

As was the case above, the null is tested using the t-statistic associated with $\hat{\alpha}_1$ but with the difference that now H_0: $\alpha_1 = 0$. In recent years this statistic has been used significantly less than that of Diebold and Mariano (1995)–West (1996) for non-nested model comparisons and has been shown by Clark and McCracken (2001) and McCracken (2007) to have considerably less power than other statistics when the models are nested. Nevertheless, this approach is very useful in the context of conditional forecasts because it allows us to parse out why the models are not equally accurate if we happen to reject. In particular the intercept term gives an indication of the presence of immodest conditioning and hence a joint test of modest conditioning and equal accuracy can be conducted based on a Wald-statistic that tests the joint null H_0:$\alpha_0 = \alpha_1 = 0$. If that joint null is rejected then one can look at the individual t-statistics to get an indication of whether the rejection is due to severely immodest conditioning or whether the models are simply not equally accurate despite modest conditioning.

3.3 Analytical results

While not necessary for our results, it is instructive to present the asymptotics in an environment in which OLS-estimated VARs are used to construct τ-step ahead conditional forecasts across the forecast origins $t = R, ..., T - \tau = R + P - \tau$. Specifically, suppose that the model takes the form

$$Y_t = C + A(L)Y_{t-1} + \varepsilon_t,$$

where $Y = (y_1, y_2, ..., y_n)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ and $A(L) = \sum_{j=1}^{m} A_j L^j$ for $n \times 1$ and $n \times n$ parameter matrices C and A_j, $j = 1, ..., m$, respectively. This is equivalent to

$$Y_t = \Lambda x_{t-1} + u_t = (x'_{t-1} \otimes I_n)\beta + \varepsilon_t,$$

if we define $x_t = (1, Y'_t, ..., Y'_{t-m+1})'$, $\beta = vec(\Lambda)$, and $\Lambda = (C, A_1, ..., A_m)$. Since the model is estimated by OLS we obtain

$$\hat{\Lambda}_t = (t^{-1} \sum_{s=1}^{t-1} Y_{s+1} x'_s)(t^{-1} \sum_{s=1}^{t-1} x_s x'_s)^{-1}$$

and hence

$$vec(\hat{\Lambda}_t) = \hat{\beta}_t = \beta + B(t)H(t),$$

where $B(t) = ((t^{-1} \sum_{s=1}^{t-1} x_s x'_s)^{-1} \otimes I_n)$ and $H(t) = t^{-1} \sum_{s=1}^{t-1} vec(\varepsilon_{s+1} x'_s) = t^{-1} \sum_{s=1}^{t-1} h_{s+1}$.

These parameter estimates are used to construct both unconditional and conditional forecasts of $Y_{t+\tau}$, denoted $\hat{Y}^u_{t+\tau}$ and $\hat{Y}^c_{t+\tau}$, which in turn imply forecast errors $\hat{\varepsilon}^u_{t+\tau}$ and $\hat{\varepsilon}^c_{t+\tau}$. If we are specifically interested in the jth element of $Y_{t+\tau}$, define $\hat{y}_{j,t+\tau} = \iota'_j \hat{Y}_{t+\tau}$ for the vector ι_j with zeros everywhere but with a 1 in the jth position. Define $\hat{\varepsilon}^c_{j,t+\tau} = \iota'_j \hat{\varepsilon}^c_{t+\tau} = y_{j,t+\tau} - \hat{y}^c_{j,t+\tau}$ and $\hat{\varepsilon}^u_{j,t+\tau} = \iota'_j \hat{\varepsilon}^u_{t+\tau} = y_{j,t+\tau} - \hat{y}^u_{j,t+\tau}$, accordingly.

Our asymptotic results follow from an application of West's (1996) Theorem 4.1. In the context of the test for zero bias this is straightforward since $\hat{\alpha}_1$ is just a sample average of forecast errors and hence $\hat{\alpha}_1 = P^{-1}\sum_{t=R}^{T-\tau}\hat{\varepsilon}^c_{j,t+\tau}$. As such, if we define the function $f_{t+\tau}(\hat{\beta}_t) = \hat{\varepsilon}^c_{j,t+\tau}$, and maintain assumptions A1 - A5 of West (1996) we find that under the null $H_0 : E\varepsilon^c_{j,t+\tau} = 0$ for all t,

$$P^{1/2}\hat{\alpha}_1 \to^d N(0, \Omega),$$

where Ω captures the estimation error associated with all the elements of $\hat{\varepsilon}^c_{j,t+\tau}$ and hence takes the form

$$\Omega = S_{ff} + 2(1 - \pi^{-1}\ln(1 + \pi))(FBS_{fh} + FBS_{hh}B'F'),$$

where

$$\begin{aligned}
B &= ((Ex_s x_s')^{-1} \otimes I_n),\ S_{ff} = \lim_{P,R\to\infty} Var(P^{-1/2}\sum_{t=R}^{T-\tau}\varepsilon^c_{j,t+\tau}), \\
S_{fh} &= \lim_{P,R\to\infty} Cov(P^{-1/2}\sum_{t=R}^{T-\tau}\varepsilon^c_{j,t+\tau}, P^{-1/2}\sum_{t=R}^{T-\tau}\varepsilon_{t+1}x_t'), \\
F &= E\partial f_{t+\tau}(\beta)/\partial\beta, \text{ and } \lim_{P,R\to\infty} P/R = \pi.
\end{aligned}$$

Delineating the asymptotic distribution for $\hat{\alpha}_1$ for the efficiency regression in equation (2) is significantly more complicated but can still be mapped into the framework of West (1996) with a bit of thought. First note that $\hat{\alpha}_1 = \iota_2'\hat{\gamma} = \iota_2'(P^{-1}\sum_{t=R}^{T-\tau}\hat{z}_t\hat{z}_t')^{-1}(P^{-1}\sum_{t=R}^{T-\tau}\hat{z}_t\hat{\varepsilon}^c_{j,t+\tau})$. Hence there exists a continuously differentiable function $g(\cdot)$ such that

$$\hat{\alpha}_1 = g(P^{-1}\sum_{t=R}^{T-\tau} vech_{-1}(\hat{z}_t\hat{z}_t')', P^{-1}\sum_{t=R}^{T-\tau}\hat{z}_t'\hat{\varepsilon}_{j,t+\tau}),$$

where the notation $vech_{-1}$ denotes the $vech$ operator but omits the first element — that associated with the intercept. As such, if we define the function $f_{t+\tau}(\hat{\beta}_t) = (vech_{-1}(\hat{z}_t\hat{z}_t')', \hat{z}_t'\hat{\varepsilon}_{j,t+\tau})'$, and maintain assumptions A1-A5 of West (1996), the Delta method along with Theorem 4.1 of West (1996) implies that

$$P^{1/2}\hat{\alpha}_1 \to^d N(0, \Omega)$$

under the null $H_0 : \alpha_1 = g(Ef_{t+\tau}) = 0$ for all t. The asymptotic variance Ω is a bit more complicated due to the use of the Delta method and takes the form $\Omega = \nabla g'(Ef_{t+\tau})V\nabla g(Ef_{t+\tau})$, where

$$V = S_{ff} + 2(1 - \pi^{-1}\ln(1 + \pi))(FBS_{fh} + FBS_{hh}B'F')$$

and

$$
\begin{aligned}
B &= ((Ex_s x_s')^{-1} \otimes I_n),\ S_{ff} = \lim_{P,R\to\infty} Var(P^{-1/2}\sum_{t=R}^{T-\tau}(vech_{-1}(z_t z_t' - Ez_t z_t')', z_t'\varepsilon_{j,t+\tau} - Ez_t'\varepsilon_{j,t+\tau})),\\
S_{fh} &= \lim_{P,R\to\infty} Cov(P^{-1/2}\sum_{t=R}^{T-\tau}(vech(z_t z_t' - Ez_t z_t')', z_t'\varepsilon_{j,t+\tau} - Ez_t'\varepsilon_{j,t+\tau}), P^{-1/2}\sum_{t=R}^{T-\tau}\varepsilon_{s+1}x_s'),\\
F &= E\partial f_{t+\tau}(\beta)/\partial\beta,\ \lim_{P,R\to\infty} P/R = \pi,\ \nabla g'(Ef_{t+\tau}) = \partial g(Ef_{t+\tau})/\partial Ef_{t+\tau}.
\end{aligned}
$$

While we don't spell out the details in the interest of brevity, we obtain similar results for our proposed tests of equal MSE.[3]

Broadly, in these results, parameter estimation error (PEE) has to be accounted for in the asymptotic variances of the statistics of interest. However, the algebra of checking the special conditions that could make PEE asymptotically irrelevant is difficult in general VAR settings. Instead, to provide a sense of when PEE will be asymptotically relevant, we have analyzed the simple case of 1-step ahead conditional forecasts from a VAR(1). Consider a test for no bias in the conditional forecast of y_{t+1}, in which case the condition is that x_{t+1} remain unchanged from period t, such that $x^c_{t+1} = x_t$. In this example, PEE does not matter under the recursive estimation scheme because it just so happens that $-FBS_{fh} = FBS_{hh}B'F'$ even though $F \neq 0$. But PEE does matter under the rolling or fixed window estimation schemes, and since $S_{ff} = -FBS_{fh} = FBS_{hh}B'F'$, the degree of estimation error-driven size distortion increases monotonically with P/R if PEE is ignored in the calculation of standard errors. In the case of the rolling scheme, tests for bias will become increasing undersized as P/R rises. Now consider the case of tests of bias in conditional forecasts obtained using true future information about x_{t+1}. Specifically, suppose that the condition is $x^c_{t+1} = x_{t+1} + n_{t+1}$, such that the conditioning value is the true value for next period plus a standard normally distributed noise term. In this case, the matrices F, B, and S_{hh} are the same as in the backward-looking conditioning case. But S_{ff} and S_{fh} change, such that $-FBS_{fh} \neq FBS_{hh}B'F'$. As a result, under forward-looking conditioning, PEE matters under all estimation schemes.

3.4 Inference with standard normal critical values

For both tests of bias and efficiency we are able to establish that $P^{1/2}\hat{\alpha}_1$ is asymptotically normal. We are also able to establish asymptotic normality of the test of equal MSE (or

[3]However, the tests of equal MSE require an assumption to rule out error correlations of zero that would make the unconditional and conditional forecasts exactly the same.

chi-square if a joint test of modesty and equal MSE). As such, one approach to inference is to construct a consistent estimate of the asymptotic variance and then compare the standardized statistic $P^{1/2}\hat{\alpha}_1/\hat{\Omega}^{1/2}$ to standard normal critical values. As discussed in West (1996) as well as West and McCracken (1998), many of the components of the asymptotic variance are easily estimated, while other elements can be more complicated. Among the easiest components are $\hat{\pi} = P/R$ and $\hat{B} = ((T^{-1}\sum_{s=1}^{T-1} x_s x_s')^{-1} \otimes I_n)$. In addition, standard HAC estimators can be used to construct estimates of S_{ff} and S_{fh} given $f_{t+\tau}(\hat{\beta}_t)$ and $\hat{h}_{t+1} = vec(\hat{\varepsilon}_{t+1}x_t')$, where $\hat{\varepsilon}_{t+1}$ denotes the residuals from the OLS-estimated VAR.

As a practical matter, it is significantly harder to estimate both $\nabla g(Ef_{t+\tau})$ and F. This is particularly true for F, which, especially at longer horizons, can be a very complicated function of the $n + mn^2$-dimensioned vector β. Rather than pursue a closed form solution for F, an alternative approach is to use numerical approximations as discussed in Newey and McFadden (1994) but adapted to an out-of-sample environment. Specifically, for each element of the function $f_{t+\tau}(\hat{\beta}_t)$ $i = 1, ..., l = \dim(f_{t+\tau}(\beta))$[4], an estimate of F_i can be built by constructing $k = \dim(\beta)$ individual elements of the vector F_i for each $j = 1, ..., k$ according to

$$\hat{F}_{i,j} = P^{-1}\sum_{t=R}^{T-\tau} \frac{[\hat{f}_{i,t+\tau}(\hat{\beta}_t + \iota_j r(P)) - \hat{f}_{i,t+\tau}(\hat{\beta}_t - \iota_j r(P))]}{2r(P)},$$

where ι_j is zero everywhere but the jth position and is 1 in that spot. As shown in McCracken (2001) this will provide a consistent estimate of F for some scalar non-stochastic function $r(P)$ satisfying $\lim_{P\to\infty} r(P) = 0$ and $\lim_{P\to\infty} P^{1/2}r(P) = \infty$.

Estimating $\nabla g(Ef_{t+\tau})$ is less complex than F but still requires a significant amount of algebra to construct directly using the estimator $\nabla g(P^{-1}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t))$. Alternatively the numerical approximation approach could be implemented for each of the $i = 1, ..., l$ elements of ∇g

$$\nabla \hat{g}_l = \frac{g(P^{-1}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t) + \iota_l r(P)) - g(P^{-1}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t) - \iota_l r(P))}{2r(P)},$$

where ι_l is zero everywhere but the lth position and is 1 in that spot.

3.5 A bootstrap approach to inference

The discussion in the previous section provides a means of assessing the statistical significance of the coefficient $\hat{\alpha}_1$ of the bias and efficiency regressions and the coefficients of the

[4]For the test of zero bias there is only one element of $f_{t+\tau}(\hat{\beta}_t)$. For the efficiency test there are eight. Hence $F = F_1$ for the bias test while F is a matrix of concatenated vectors F_1 through F_8.

equal MSE tests. Specifically, if $\hat{\Omega}$ is a consistent estimate of Ω, it immediately follows that $P^{1/2}\hat{\alpha}_1/\hat{\Omega}^{1/2} \rightarrow^d N(0,1)$, and therefore one can use standard normal critical values to test the null that $\alpha_1 = 0$ against an alternative in which $|\alpha_1| > 0$. In the test of zero bias this approach to inference might not be too difficult to implement, especially for shorter forecast horizons. But for the test of efficiency, constructing the standard errors quickly becomes impractical. The same applies to the tests of equal MSE. Moreover, as noted in the example of the introduction, conditional forecast errors can contain high order serial correlation, which will necessitate autocorrelation-consistent variances that may be difficult to estimate with precision in finite samples.

As a result, we develop and suggest a parametric bootstrap approach to inference. The bootstrap is patterned on the approach first used in Kilian (1999) and applied in the univariate forecast evaluation of such studies as Clark and McCracken (2001, 2005). The bootstrap DGP is an estimate of a VAR fit to the full sample of data on the variables of interest, from which we obtain a vector of residuals $\hat{\varepsilon}_t$. We then draw a time series of i.i.d. N(0,1) innovations, denoted $n_t,\ t = 1, \ldots, T$. We obtain a bootstrapped draw of the time series of the vector of residuals as $n_t\hat{\varepsilon}_t,\ t = 1, \ldots, T$. We use these residuals and the autoregressive structure of the VAR to obtain an artificial time series y_t^*.[5] In each bootstrap replication, the bootstrapped data are used to (recursively) estimate the VAR forecasting model and generate artificial, out-of-sample forecasts. These forecasts and associated forecast errors are used to run bias, efficiency, and equal MSE regressions and compute the associated test statistics. Critical values are simply computed as percentiles of the bootstrapped test statistics.

3.6 Tests applied to conditional forecasts under Giacomini-White asymptotics

The theoretical results in West (1996) and West and McCracken (1998) focus on testing a null hypothesis of the form $H_0 : Ef(X_{t+\tau}, \beta^*) = 0$. In words, this hypothesis is designed to evaluate the predictive ability of a model if the true parameter values β of the model are known. But one might prefer to evaluate the predictive content of the model accounting for the fact that the model parameters must be estimated and will never be known. Doing so changes the null hypothesis to something akin to H_0: $Ef(X_{t+\tau}, \hat{\beta}_t) = 0$, where the estimation error associated with the parameter estimates $\hat{\beta}_t$ is now captured under the null

[5] The initial observations are selected by sampling from the actual data as in Stine (1987).

hypothesis.

Even though the null hypothesis H_0: $Ef(X_{t+\tau}, \hat{\beta}_t) = 0$ is reasonably intuitive, the results in West (1996) and West and McCracken (1998) do not apply under this null and hence their theory cannot be used to infer the asymptotic distribution of $(P - \tau + 1)^{-1/2} \sum_{t=R}^{T-\tau} f(X_{t+\tau}, \hat{\beta}_t)$ under this null. Instead, theoretical results in Giacomini and White (2006) can be used to conduct inference. Giacomini and White show that so long as all parameter estimates are estimated using a small rolling window of observations (small in the sense that R is finite and P diverges to infinity), $(P - \tau + 1)^{-1/2} \sum_{t=R}^{T-\tau} f(X_{t+\tau}, \hat{\beta}_t) \rightarrow^d N(0, S_{\hat{f}\hat{f}})$ under the null hypothesis

$$\lim_{P\to\infty} (P - \tau + 1)^{-1/2} \sum_{t=R}^{T-\tau} Ef(X_{t+\tau}, \hat{\beta}_t) = 0,$$

where $S_{\hat{f}\hat{f}} = \lim_{P\to\infty} Var(P^{-1/2} \sum_{t=R}^{T-\tau} f(X_{t+\tau}, \hat{\beta}_t))$.

The theoretical results in Giacomini and White (2006) can be used for a wide range of applications, including zero bias, forecast efficiency, and equal accuracy. So long as the statistic takes the form $(P - \tau + 1)^{-1/2} \sum_{t=R}^{T-\tau} f(X_{t+\tau}, \hat{\beta}_t)$, and all parameters are estimated using a small rolling window of observations, normal critical values can be used to conduct inference on the null hypothesis $\lim_{P\to\infty} (P - \tau + 1)^{-1/2} \sum_{t=R}^{T-\tau} Ef(X_{t+\tau}, \hat{\beta}_t) = 0$, without any additional correction of standard errors for the effects of estimation of the forecasting model's parameters. However, in light of the simple example of our introduction, it remains the case that conditional forecast errors will have high-order serial correlation. As noted above, this may create problems in accurate HAC estimation in finite samples. In practice, as much as it may be asymptotically valid (under a rolling estimation scheme for the forecasting model) to compare bias, efficiency, and equal accuracy tests against normal distributions without correction for parameter estimation error, HAC estimation problems may still make a bootstrap like the one we described above preferable to standard normal-based inference.

4 Monte Carlo Analysis

This section presents a Monte Carlo analysis of the finite sample properties of tests for bias, efficiency, and equal MSE applied to unconditional and conditional forecasts from VAR models. In most cases, the conditioning we consider is similar to the monetary policy-motivated conditioning we described above and will use below in our application to macroeconomic forecasting. In particular, we produce and evaluate forecasts conditioned on a path for a pseudo-policy variable that holds the value of the variable at its most recent (relative to the forecast origin) value. However, we also provide some results for conditional forecasts obtained by conditioning on a path for a pseudo-policy variable that is the true future value plus measurement error. Using the simulated data, we consider the properties of bias, efficiency, and equal MSE tests, based on both normal (or chi square) distributions with standard errors that abstract from PEE and bootstrapped test distributions.

In these experiments, we use bivariate and trivariate VARs as data-generating processes (DGPs). To form forecasts, we consider both OLS estimation and Bayesian estimation, under a Normal-inverted Wishart prior and posterior. In the OLS case, point forecasts are obtained by using OLS estimates of coefficients and simple iteration of forecasts based on the OLS coefficients. In the Bayesian estimation case, we also obtain forecasts by iteration, using just the posterior mean coefficient estimates obtained with the analytical solution for the Normal-Wishart posterior. At horizons greater than 1 period, proper Bayesian methods would require simulation of the posterior distribution, but Carriero, Clark, and Marcellino (2012) find that multi-step point forecasts based on just the posterior mean coefficients are essentially the same as properly simulated point forecasts. While we focus on forecasts computed under a recursive (expanding window) estimation scheme, we also provide some results for a rolling window estimation scheme.

In all simulations, based on 2000 Monte Carlo draws, we report the percentage of Monte Carlo trials in which the null of no bias, efficiency, or equal MSE is rejected — the percentage of trials in which the sample test statistics fall outside (two-sided) critical values. In the reported results, the tests are compared against 10% critical values. Using 5% critical values yields similar findings.

We proceed by detailing first the data-generating processes and other aspects of experiment design and then our implementation of Bayesian methods. We then present the results.

4.1 Monte Carlo design

For each DGP, we generate data using independent draws of innovations from the normal distribution and the autoregressive structure of the DGP. The initial observations necessitated by the lag structure of each DGP are generated with draws from the unconditional normal distribution implied by the DGP. With quarterly data in mind, for some DGPs we report results for forecast horizons of just 1 period, while for others we report results for forecast horizons of 1, 2, and 4 periods. We consider sample sizes of R, P = 50,100; 50,150; 100,50; 100,100; and 100,150. We use DGPs 1-3 to evaluate size properties and DGP 4 to evaluate power.

DGP 1 is a bivariate VAR(1), with regression coefficients given in the first panel of Table 1 and an error variance-covariance matrix of:

$$\text{var}(\varepsilon_t) = \begin{pmatrix} 1.0 & \\ -0.5 & 1.0 \end{pmatrix}.$$

DGP 2 takes the same form, with regression coefficients given in the second panel of Table 1 and an error variance-covariance matrix of:

$$\text{var}(\varepsilon_t) = \begin{pmatrix} 1.0 & \\ -0.3 & 0.2 \end{pmatrix}.$$

In experiments with DGPs 1 and 2, to keep the conditioning simple and avoid entangling the effects of conditioning with possible effects of additional serial correlation in multi-step forecast errors, we only report results for a 1-step ahead forecast horizon. In the main results for the conditional case, we forecast $y_{1,t+1}$, the first variable in period $t+1$, conditional on a value for $y_{2,t+1}$, under two approaches spelled out below. In light of this conditioning specification, we only report test results for forecasts of the first variable, $y_{1,t+1}$. For these DGPs, we only report results for conditional-reduced form forecasts. If, for identification, we use a recursive ordering with $y_{1,t+1}$ before $y_{2,t+1}$, at a 1-period ahead forecast horizon, under the policy shock approach to conditional forecasting, the condition on the second variable would have no implications for the forecast of the first. We should also note that DGPs 1 and 2 have been deliberately parameterized so that the 1-step ahead conditional forecast error is more serially correlated with DGP 2 than DGP 1.

As to the conditioning approaches used with DGPs 1 and 2, we consider two different ones. First, we provide results for forecasts conditioned on $y_{2,t+1}$ not changing in period $t+1$ — that is, conditional on $\hat{y}^c_{2,t+1} = y_{2,t}$. Second, we provide results for forecasts conditioned on the actual future value of $y_{2,t+1}$ plus measurement error. We include and parameterize the measurement error to make the unconditional and conditional forecasts equally accurate in population. Specifically, in this case, we produce forecasts conditioned on $\hat{y}^c_{2,t+1} = y_{2,t+1} + n_{t+1}$, where n_{t+1} is normal with mean 0 and variance equal to the variance of innovations to $y_{2,t}$ in the DGP.

DGP 3 is a trivariate VAR(2), with regression coefficients given in the third panel of

Table 1 and an error variance-covariance matrix of:

$$\text{var}(\varepsilon_t) = \begin{pmatrix} 9.265 & & \\ 0.296 & 1.746 & \\ 0.553 & 0.184 & 0.752 \end{pmatrix}. \tag{6}$$

We set the parameters of DGP 3 to equal OLS estimates of a VAR in GDP growth, inflation less the survey-based trend used in our application, and the federal funds rate less the survey-based trend for inflation, over a sample of 1961-2007.

In experiments with DGP 3, we extend the conditioning horizon, and we report results for horizons of 1, 2, and 4 quarters ahead. In this case, we forecast variables $y_{1,t+h}$ and $y_{2,t+h}$ for $h = 1, \ldots, 4$ conditional on on $y_{3,t+h}$ not changing in periods $t+1$ through $t+4$ — that is, conditional on $\hat{y}^c_{3,t+h} = y_{3,t}, h = 1, \ldots, 4$. In some unreported experiments, we also conducted DGP 3 experiments with conditioning horizons of just 1 and 2 periods and obtained results for those horizons similar to those we report for the conditioning horizon of 4 periods.

To evaluate power properties, DGP 4 imposes some coefficient breaks on the specification of the bivariate DGP 1. The breaks consist of one-time shifts in the intercept of the $y_{1,t}$ equation and the slope and intercept coefficients of the $y_{2,t}$ equation. The pre- and post-break coefficients are given in the last panel of Table 1; the error variance-covariance matrix is kept constant, at the setting used with DGP 1 (see above). The break is imposed to occur at period $R+1$, the date of the first out-of-sample forecast.

4.2 Implementation of Bayesian estimation methods

In the experiments in which the forecasting model is estimated by Bayesian methods, the BVAR(m) takes the form detailed in section 5:

$$\begin{aligned} Y_t &= C + A(L)Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma) \\ &= \Lambda x_{t-1} + \varepsilon_t. \end{aligned}$$

For this model, we use the Normal-Wishart prior and posterior detailed in such studies as Kadiyala and Karlsson (1997) and Banbura, Giannone, and Reichlin (2010).

The prior takes a conventional Minnesota form, without cross-variable shrinkage, in which the prior for the VAR coefficients is normally distributed:

$$\text{vec}(\Lambda) \sim N(\text{vec}(\underline{\mu}_\Lambda), \underline{\Omega}_\Lambda).$$

Letting $\Lambda_l^{(ij)}$ denote the coefficient on lag l of variable j in equation i and $\Lambda_l^{(ij)}$ with l=0 denote the intercept of equation i, we define the prior mean and variance as follows:

$$\begin{aligned} \underline{\mu}_\Lambda \text{ such that } E[\Lambda_l^{(ij)}] &= 0 \; \forall \; i, j, l \qquad (7) \\ \underline{\Omega}_\Lambda \text{ such that } V[\Lambda_l^{(ij)}] &= \begin{cases} \frac{\theta^2}{l^2} \frac{\sigma_i^2}{\sigma_j^2} & \text{for } l > 0 \\ \varepsilon^2 \sigma_i^2 & \text{for } l = 0 \end{cases} \qquad (8) \end{aligned}$$

In our BVAR implementation, we set the tightness hyperparameter θ at a standard value of 0.2 and the intercept hyperparameter ε to 1000, and we set the scale parameters σ_i^2 at estimates of residual variances from AR(m) models estimated with the data sample available at each forecast origin. For the evaluation of point forecasts from the BVAR, we obtain forecasts using the posterior mean value of the coefficients and the iterative approach to forecasting at horizons greater than 1 period.

4.3 Test implementation

To test bias, we regress forecast errors (either unconditional or conditional) on a constant and form the t-statistic for the null of a coefficient of zero. The associated regressions are given in section 3's equations (1) and (2). With unconditional forecasts, the forecast error for horizon τ should follow an MA($\tau - 1$) process. At the 1-step horizon, we form the test statistic using the simple OLS estimate of the variance. At longer horizons, we use the HAC estimator of Newey and West (1987), with $2(\tau - 1)$ lags, following studies such as Kim and Nelson (1993). But with conditional forecasts, as we noted with the introduction's example, even 1-step ahead forecast errors will be serially correlated, potentially at long lags, depending on the features of the DGP. For the tests applied to all conditional forecasts, we use the HAC estimator of Newey and West (1987), with $6 + 2(\tau - 1)$ lags.[6]

In testing efficiency, in the unconditional case we regress the forecast error on a constant and the unconditional forecast and form the t-statistic for the null of a coefficient of zero on the unconditional forecast. In the conditional case, we run the regression of equation (2) and test $\alpha_1 = 0$.

Note: Results for tests of equal MSE are still in process.

[6]In some preliminary experiments, in computing all test statistics, we used the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992) instead of the Newey and West (1987) estimator. However, this conceptually preferable estimator didn't yield materially better results, and the additional calculations (to pre-whiten, compute the optimal bandwidth, etc.) significantly increased the CPU time requirements of each experiment.

4.4 Results

Tables 2 through 5 provide size results from Monte Carlo experiments with DGPs 1, 2, and 3. Under OLS estimation, our experiments yield the following results. In these results, the conditional forecasts are produced under a condition that the last variable of the VAR is equal to its last observed value as of the forecast origin. Results for forecasts conditioned on actual future values of the last variable of the VAR are still in process.

- With unconditional forecasts, tests of bias based on standard normal critical values range from being about correctly sized (DGPs 1 and 2) to oversized (DGP 3). Oversizing is more likely at longer forecast horizons, likely because of difficulties with the finite sample precision of HAC variance estimates.

- With unconditional forecasts, tests of efficiency based on standard normal critical values are roughly correctly sized at the 1-step horizon with larger sample sizes and DGPs 1 and 2. However, at smaller samples, 1-step efficiency tests can be modestly oversized. At longer forecast horizons, the efficiency test can be severely oversized, again likely because of HAC estimation challenges. For example, for forecasts of $y_{2,t}$ in DGP 3, $\tau = 4$, and $R, P = 100{,}100$, the rejection rate for the normal-based efficiency test is 44.0% (see Table 4).

- With unconditional forecasts, basing inference on bootstrap critical values largely eliminates any size distortions in bias and efficiency tests. In fact, tests of efficiency based on bootstrap critical values can be a little undersized. Continuing with the same example, the bootstrap-based rejection rate for the efficiency test is 8.3% (Table 4).

- With standard normal critical values, rejection rates for tests of efficiency are higher for conditional forecasts obtained under the conditional-reduced form approach than for unconditional forecasts, especially with DGP 2, which creates relatively high correlation in conditional forecast errors.

- However, comparing the conditional efficiency test against bootstrap critical values greatly improves size, and yields tests that are either correctly sized or a little undersized. For example, forecasts of $y_{2,t}$ in DGP 3, $\tau = 4$, and $R, P = 100{,}100$, the rejection rate for the normal-based efficiency test of conditional-reduced form forecast is 56.4%, while the bootstrap based rejection rate is 8.6% (see Table 4).

- Results for conditional-policy shock forecasts (just for DGP 3; and, by construction, the policy shock conditioning only affects forecasts at horizons of 2 and 4 periods) are qualitatively very similar to those for conditional-reduced forecasts.

Under Bayesian estimation of the forecasting models, the results are qualitatively very similar to those based on OLS estimation. Most notably, it continues to be the case that a bootstrap approach to inference seems to be reliable in the sense of yielding rejection rates reasonably close to nominal size. Broadly, the main differences under Bayesian estimation are that the rejection rates of bias tests are a little higher for Bayesian-estimated models than OLS-estimated models, while efficiency test rejection rates are a little lower for Bayesian-estimated models than OLS-estimated models. These patterns likely reflect the anticipated effects of shrinkage that pushes all parameters to zero,increasing the chances of small bias in forecasts but reducing serial correlation in conditional forecasts by reducing slope coefficient estimates. In any case, using bootstrap critical values yields tests that are about correctly sized.

The results for DGP 4 in Table 6 show that the bias and efficiency tests have power against their associated null hypotheses when the DGP suffers a coefficient break. As might be expected, given that the mean of the predictand shifts when forecasting starts, power is greater for the test of bias than the test of efficiency. For example, with $R, P = 100, 100$ and standard normal critical values, the rejection rate is 77.9% for a test for bias in the unconditional forecast and 43.0% for a test of efficiency in the same forecast. As might be expected, given the evidence of some size distortions in the tests based on normal critical values (see discussion above), power is somewhat lower with bootstrap critical values than standard normal critical values. In the same example, the corresponding bootstrap-based rejection rates are 74.1% (bias) and 27.7% (efficiency). For conditional forecasts, the rejection rate is somewhat higher (compared to unconditional forecasts) for the test of bias (95.3% with the bootstrap) and about the same for the test of efficiency (28.6% with the bootstrap).

5 Empirical Application

In our empirical application, we examine forecasts of U.S. GDP growth, the unemployment rate, and PCE inflation obtained with models of 5 and 22 variables. The smaller model is similar to one of those in Clark (2011), using variables transformed for stationarity and a prior on steady states (and constant volatility of innovation variances). As detailed below, the smaller model is also specified to use information on trend inflation, measured with long-run survey forecasts of inflation, and the steady-state rate of unemployment, measured with the Congressional Budget Office's estimate of the potential unemployment rate. The larger model is taken from Giannone, Lenza, and Primiceri (2012), using variables in levels or log levels. We chose the smaller model with a steady state prior on the basis of evidence that such models often reduce bias in inflation forecasts (e.g., Clark 2011) and the larger model on the basis of evidence that larger models often forecast more accurately than smaller models (e.g., Banbura, Giannone, and Reichlin 2010).

This section proceeds by detailing the models and our implementation (some additional details are included in the appendix), data and forecasting details, and results.

Note: The application results are incomplete in some important results. First, we do not yet have complete results for the small model. Second, we need to add results of tests of equal MSE.

5.1 Models and implementation

5.1.1 Small model

The small model (BVAR5-SSP) takes a form similar to Clark's (2011) specification of a VAR with some time-varying trends and a steady-state prior. The table below lists the variables included and their transformations. Let y_t denote the $n \times 1$ vector of model variables (detailed below) and d_t denote a $q \times 1$ vector of deterministic variables. In this paper, the only variable in d_t is a constant. Let $\Pi(L) = I_n - \Pi_1 L - \Pi_2 L^2 - \cdots - \Pi_m L^m$ and $\Psi =$ a $n \times q$ matrix of coefficients on the deterministic variables. The BVAR5-SSP model takes the form

$$\Pi(L)(y_t - \Psi d_t) = v_t,\ v_t \sim N(0, \Sigma). \tag{9}$$

In our implementation, we depart from the specification of Clark (2011) in a few respects. To help with the identification of monetary policy, we include in the model (adding to Clark's specification) commodity price inflation. To simplify forecasting, we make some

Variables in small BVAR

variable	transformation	prior mean on AR(1) coefficient
real GDP	$\Delta \ln$	0
unemployment rate (less CBO potential)	level	0.7
spot commodity price index	$\Delta \ln$	0
PCE price index (less long-run inflation expectation)	$\Delta \ln$	0.7
federal funds rate (less long-run inflation expectation)	level	0.7

adjustments to the treatment of trends. First, we drop the survey-based long-run inflation expectation from the model's set of endogenous variables. In the research with the original model, making the expectation endogenous had little effect on the point forecasts or forecast distributions. Second, to capture the trend in unemployment, we use the Congressional Budget Office's (CBO) estimate of the potential rate of unemployment instead of an exponentially smoothed trend. Finally, for measuring the long-run inflation expectation, we use the series included in the Federal Reserve Board's FRB/US model instead of a series based largely on the Blue Chip consensus. The FRB/US series, denoted PTR, splices econometric estimates of inflation expectations from Kozicki and Tinsley (2001a) early in the sample to 5- to 10-year-ahead survey measures compiled by Richard Hoey and, later in the sample, to 10-year-ahead values from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters. With these choices, the variable vector y_t of the small VAR includes GDP growth, the unemployment rate less the natural rate (CBO potential estimate), inflation less the long-run inflation expectation, the rate of growth in commodity prices, and the federal funds rate less the long-run inflation expectation. The variables are listed in the table above. One final implementation detail to note is that we set the lag length of the VAR to 4.

The prior for the BVAR5-SSP model takes the Normal-diffuse form used in Villani (2009), which yields conditional posterior distributions that are multivariate normal for Π and Ψ and inverted Wishart for Σ. The appendix details the estimation steps.

The prior for the VAR slope coefficients $\Pi(L)$ is based on a Minnesota-style specification. The prior means suppose each variable follows an AR(1) process, with coefficients of 0 for GDP growth and commodity prices (less persistent variables) and 0.7 for the other variables (more persistent variables). Prior standard deviations are controlled by the usual hyperparameters, with overall tightness of 0.2, cross–equation tightness of 1.0, and linear decay in the lags. The standard errors used in setting the prior are estimates from univariate

AR(4) models fit over the estimation sample, in keeping with common practice for Minnesota priors.

Priors are imposed on the deterministic coefficients Ψ to push the steady-states toward certain values, of: (1) GDP growth, 3.0 percent; (2) unemployment less the natural rate, 0.0; (3) inflation less the long-run inflation expectation, 0.0; (4) commodity price inflation, 3.0 percent; and (5) federal funds rate less the long-run inflation expectation, 2.0. Accordingly, in the prior for the elements of Ψ, all means are zero, except as follows: GDP growth, intercept coefficient of 3.0; commodity price inflation, intercept coefficient of 3.0; and fed funds rate, intercept coefficient of 2.0. We set the following standard deviations on each element of Ψ: GDP growth, 0.5; unemployment less natural rate, 0.15; inflation less long-run expectation, 0.15; commodity price inflation, 0.5; and fed funds rate less long-run inflation expectation, 0.5. We set lower prior standard deviations for a few variables for which there is more confidence in steady state and higher standard deviations for which the steady state value is a priori more uncertain.

For each post-burn draw in the MCMC chain, we draw forecasts from the posterior distribution as follows: for each period of the forecast horizon, we sample shocks to the VAR with a variance equal to the associated draw of Σ and compute the forecast draw of Y_{t+h} from the VAR structure (using the associated draws of Π and Ψ) and drawn shocks. For simplicity, in forming the forecasts, we abstract from uncertainty surrounding trend inflation and the natural rate. Over each forecast interval, we hold trend inflation and unemployment constant at the last observed values (the values at the date of the forecast origin). For each post-burn draw of parameters, the model is used to forecast GDP growth, unemployment less trend, inflation less the long-run inflation expectation, commodity price inflation, and the funds rate less the long-run inflation expectation. The random-walk forecasts of the level of the inflation expectation are then added to the forecasts of inflation less the expectation and the funds rate less the expectation to obtain forecasts of the levels of inflation and the funds rate. The same approach is used for the unemployment rate. In unreported results of Clark (2011), this simple approach yielded results very similar to those obtained for the paper by taking account of uncertainty about future values of trends. Finally, we report posterior estimates based on 2000 draws (retained after discarding a burn sample of 500 draws). Point forecasts are constructed as posterior means of the MCMC distributions.

Variables in large BVAR

variable	transformation	prior mean on AR(1) coefficient
real GDP	ln	1
real personal consumption expenditures (PCE)	ln	1
real business fixed investment (BFI)	ln	1
real residential investment (RESINV)	ln	1
industrial production (IP)	ln	1
capacity utilization in manufacturing (CU)	ln	1
total hours worked in the nonfarm business sector (HOURS)	ln	1
payroll employment (PAYROLLS)	ln	1
unemployment rate	level	1
consumer sentiment	level	1
spot commodity price index	ln	1
PCE price index	ln	1
price index for gross private domestic investment	ln	1
GDP price index	ln	1
real hourly compensation in the nonfarm business sector	ln	1
federal funds rate	level	1
M2	ln	1
total reserves	ln	1
S&P 500 index of stock prices	ln	1
1-year (constant maturity) Treasury bond yield	level	1
5-year (constant maturity) Treasury bond yield	level	1
real effective exchange rate for major currencies	ln	1

5.1.2 Large model

The large model takes a form similar to the large model specification of Giannone, Lenza, and Primiceri (2012). This specification is a VAR in 22 variables, in levels with Minnesota-style unit root priors, sums of coefficient priors, and initial observations priors, patterned after Sims and Zha (1998), for example. The table below lists the variables included and their transformations.

More specifically, the large BVAR takes the form:

$$\begin{aligned} Y_t &= \Pi_c + \Pi(L)Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,\Sigma) \\ &= \Lambda x_{t-1} + \varepsilon_t. \end{aligned} \tag{10}$$

The variables included in the model, listed in the table above, are very similar to those of Giannone, Lenza, and Primiceri (2012). The lag length is set to 5.

The prior for the BVAR22-NW model takes the conjugate Normal-inverted Wishart used in studies such as Banbura, Giannone, and Reichlin (2010), which yields a Normal-inverted

Wishart posterior. The appendix details the estimation steps.

The prior for the VAR slope coefficients $\Pi(L)$ is based on a Minnesota specification, supplemented with sums of coefficients and dummy initial observations priors. More specifically, in the Minnesota-style component of the prior, we impose the prior expectation and standard deviation of the coefficient matrices to be:

$$E[\Pi_k^{(ij)}] = \begin{cases} \Pi^* & \text{if } i = j,\ k = 1 \\ 0 & \text{otherwise} \end{cases}, \text{ st. dev.}[\Pi_k^{(ij)}] = \begin{cases} \frac{\lambda_1}{k}\frac{\sigma_i}{\sigma_j}, & k = 1, ..., m, \\ \lambda_0 \sigma_i, & k = 0 \end{cases} \tag{11}$$

where $\Pi_k^{(ij)}$denotes the element in position (i, j) in the matrix Π_k. The prior mean Π^* is set to 1 for all variables. For the intercept we assume an informative prior with mean 0 and standard deviation $\lambda_0\sigma_i$. The shrinkage parameter λ_1 measures the overall tightness of the prior: when $\lambda_1 \to 0$ the prior is imposed exactly and the data do not influence the estimates, while as $\lambda_1 \to \infty$ the prior becomes loose and the prior information does not influence the estimates, which will approach the standard OLS estimates. To set each scale parameter σ_i we follow common practice (see e.g. Litterman, 1986; Sims and Zha, 1998) and set it equal to the standard deviation of the residuals from a univariate autoregressive model.

Doan et al. (1984) and Sims (1993) have proposed complementing standard Minnesota priors with additional priors which favor unit roots and cointegration, and introduce correlations in prior beliefs about the coefficients in a given equation. Accordingly, in our benchmark specification, we also include the "sum of coefficients" and "dummy initial observation" priors proposed in Doan et al. (1984) and Sims (1993), respectively. Both these priors can be implemented by augmenting the system with dummy observations. The details of our implementation of this priors (including the definition of hyperparameters λ_3 and λ_4) are contained in the appendix.

Reflecting typical settings in the literature, we set the hyperparameters that govern the tightness of the prior as follows:

$$\lambda_0 = 1; \lambda_1 = 0.2;\ \lambda_3 = 1;\ \lambda_4 = 1.$$

The prior specification is completed by choosing v_0 and S_0 so that the prior expectation of Σ is equal to a fixed diagonal residual variance $E[\Sigma] = diag(\sigma_1^2, ..., \sigma_n^2)$. In particular, following Kadiyala and Karlsson (1997), we set the diagonal elements of S_0 to $s_{0ii} = (v_0 - n - 1)\sigma_i^2$ and $v_0 = n + 2$.

To streamline computations, particularly with a bootstrap approach to inference, we compute forecasts from the BVAR22-NW without simulation. We use the posterior mean of coefficients and the error covariance matrix (Σ, which we use in forming conditional forecasts) to compute point forecasts based on just the posterior mean coefficients, iterating forward from 1-step ahead forecasts. Carriero, Clark, and Marcellino (2012) found point forecasts obtained with this approach to be essentially the same as point forecasts obtained from Monte Carlo simulation.

5.2 Data and sample

In generating and evaluating forecasts, we use current vintage data taken from the FAME database of the Federal Reserve Board.[7] The quarterly data on industrial production, capacity utilization, payroll employment, the unemployment rate, consumer sentiment, commodity prices, interest rates, M2, reserves, stock prices, and the exchange rate are constructed as simple within-quarter averages of the source monthly data (in keeping with the practice of, e.g., Blue Chip and the Federal Reserve). All growth and inflation rates are measured as annualized log changes (from $t-1$ to t).

In keeping with common practice in recent VAR forecasting analyses (e.g., Giannone, Lenza, and Primiceri 2012, Koop 2013) and with the FOMC's forecast reporting practices, we provide results for a subset of the variables included in our models: GDP growth, the unemployment rate, and PCE inflation. Unlike these studies, we do not report forecasts for the federal funds rate because we are interested in forecasts of the other variables conditional on particular paths of the federal funds rate.

Motivated in part by the Federal Reserve's current forward guidance that links the setting of the federal funds rate to forecasts of inflation one to two years ahead, we provide results for forecast horizons of four and eight quarters, as well as one and two quarters ahead. In keeping with common central bank practice, for growth and inflation rates at the four and eight quarter horizons, we average quarterly growth and inflation forecasts over four quarter intervals to obtain forecasts of four quarter growth and inflation rates, four and eight quarters ahead. For the unemployment rate at these longer horizons, we use forecasts of the quarterly level.

We generate and evaluate forecasts for a sample of 1991 through 2007. We start fore-

[7]In light of the complexities of inference for data subject to revision, we leave the analysis of real time data to future research. However, qualitatively, we obtained forecast results in real time data similar to those we report for current data.

casting in 1991 and not sooner because, through the 1980s, inflation was still trending down, presumably due to a deliberate disinflation effort by the Federal Reserve. We stop our forecast evaluation in 2007:Q4 because of the zero lower bound constraints that became relevant in subsequent years. In the period since 2007, unconditional forecasts of the federal funds rate from typical VAR models have sometimes fallen well below zero. This problem can be circumvented by conditioning on a non-zero, positive path of the funds rate. However, as we detail below, testing the efficiency of the conditional forecasts of growth, unemployment, and inflation requires unconditional forecasts of the funds rate. To reliably extend our analysis beyond 2007, it would be necessary to modify the models to incorporate some allowance for the zero lower bound. While there is some research with structural term structure models that uses a shadow rate concept to do so (e.g., Krippner 2012, Kim and Singleton 2012 Christensen and Rudebusch 2013), we're not aware of any work that effectively deals with the zero lower bound in (unconditional) VAR forecasting.

In all cases, we estimate models with a sample that starts in 1961:Q1. We begin by estimating models with data from 1961:Q1 through 1990:Q4 and forming forecasts for 1991:Q1 through 1992:Q4. We then proceed by moving to a forecast origin of 1991:Q2, estimating models with data from 1961:Q1 through 1991:Q1 and forming forecasts for 1991:Q2 through 1993:Q1. We proceed similarly through time, up through 2007:Q3, to obtain a sample of forecasts from 1991 through 2007.

5.3 Results: basic properties of forecasts

Before formally evaluating the forecasts, it is useful to consider how conditioning on the path of the federal funds rate affects the forecasts of growth, unemployment, and inflation. Figures 1 and 2 provide the time paths of the 8-quarter ahead forecasts and actual values. The forecasts are from the BVAR22-NW model; our smaller model yields qualitatively similar results, in terms of the effects of conditioning. As noted above, at the 8-quarter horizon, growth and inflation are computed as 4-quarter averages, so that the forecasts of growth and inflation shown for period t refer to growth or inflation over the four quarters from $t-4$ to t, made using data through period $t-8$.

The results in Figure 1 indicate that conditioning sometimes has small effects on the forecasts of growth and unemployment and other times has large effects. The size of the effects is of course in part a function of the divergence between the conditioning path of the federal funds rate and the unconditional path. The closer the conditional path of the interest rate is to the unconditional forecast path, the smaller the effects on the conditional forecasts of growth or unemployment relative to the unconditional forecasts are likely to be. For example, in the late 1990s and early 2000s, the unconditional and conditional forecasts were relatively close together. But from about 2004 through 2007, the differences between unconditional and conditional forecasts of growth and unemployment often differed substantially. One other feature evident in Figure 1 is that different approaches to conditioning (reduced form vs. policy shocks) can, at times (not always), have different effects on the forecast. For instance, in 2006, the conditional-reduced form approach to conditioning produced a forecast below the unconditional projection of unemployment, while the policy shock approach produced a forecast above the unconditional projection of unemployment. Finally, Figure 2's results for inflation show that, under the policy shock approach, conditioning on the path of the federal funds rate does not often have much effect on the forecast (relative to the unconditional projection). This likely reflects the weak and sometimes problematic link between monetary policy and inflation in conventional VARs, manifest in the prize puzzle of impulse response analysis.

5.4 Results: bias and efficiency

Tables 7 and 8 provide bias and efficiency test results for growth, unemployment, and inflation forecasts from the BVAR5-SSP model. Tables 9 and 10 provide the same for the BVAR22-NW model. For each variable and horizon, each table provides the estimate of the coefficient of interest (intercept in the bias regression and slope coefficient on the forecast in the efficiency regression), its t-statistic, and bootstrap critical values for the 10% significance level.

In the case of the BVAR22-NW model, while normal critical values would indicate the inflation forecasts — both unconditional and conditional — to be biased, the bootstrap critical values that our Monte Carlo analysis show to be more reliable indicate the null of no bias should not be rejected. Consider, for example, 8-quarter ahead forecasts of inflation. On average, the unconditional and conditional forecasts exceed actual inflation by about 0.60 percentage point (on an annualized basis). The associated t-statistics are -2.794, -2.106, and -2.459 for, respectively, the unconditional, conditional-reduced form, and conditional-policy shock forecasts. These would all imply rejection of the null of no bias with normal critical values. However, the associated left tail bootstrap critical values (associated with a 2-sided significance level of 10%), are -3.467, -3.684, and -3.553, respectively. Based on these bootstrap critical values, the null cannot be rejected. Therefore, over the 1991-2007 period, conditioning the inflation forecasts on a fixed path of the federal funds rate does not introduce any bias into the forecasts of inflation or the forecasts of GDP growth and unemployment.

As to efficiency tests, it continues to be the case that using bootstrap critical values implies considerably less evidence of inefficiency than the (inaccurate) normal critical values would imply. However, even with bootstrap-based inference, there is some evidence of inefficiency in forecasts of inflation, in both the unconditional and conditional forecasts. More specifically, in the case of GDP growth forecasts, using normal critical values would yield rejection of the null of efficiency at almost all horizons, for all three types of forecasts. But based on bootstrap critical values, the null generally cannot be rejected for unconditional or conditional forecasts of GDP growth. There is, however, some evidence of efficiency in both unconditional and conditional forecasts of inflation. In the case of unconditional forecasts of inflation, the null of efficiency is rejected at horizons of 1, 2, and 4 quarters ahead. The same is true for conditional-reduced form forecasts. The same is also true for conditional-policy shock forecasts. Overall, based on the more reliable bootstrap, conditioning on an interest rate path does not seem to effect the efficiency of forecasts of growth and unemployment. With efficiency of inflation forecasts generally rejected for all three types of forecasts, it may also be said that conditioning doesn't affect the efficiency of the inflation forecasts, either. However, there seems to be some mis-specification of the model with respect to inflation.

5.5 Results: equal MSE

As a further prelude to formal analysis of forecast bias and efficiency, it is also useful to compare the broad accuracy of the forecasts, as measured by RMSEs, provided in Table 11. Somewhat surprisingly, conditioning the forecasts of GDP growth, unemployment, and inflation on an unchanged path of the federal funds rate over a two year period from the forecast origin does not consistently reduce forecast accuracy (subject to the caveat that the differences in accuracy are often small). For some variables at some horizons, the conditional forecasts have higher RMSEs than the unconditional forecasts, while for other variable-horizon combinations, the reverse is true. Consider, for example, 8-quarter ahead forecasts of the unemployment rate. In the BVAR5-SSP results, the RMSE of the conditional-reduced form forecast is a little lower (at 0.793) than the RMSE of the unconditional forecast (which is 0.720), while the RMSE of the conditional-policy shock forecast is, as expected, higher (at 0.935) than the RMSE of the unconditional forecast. In the BVAR22-NW specification, it is again the case that the conditional-reduced form forecast (RMSE of 0.801) is a little more accurate than the unconditional forecast (RMSE of 0.836), but the conditional-policy shock forecast is yet more accurate (RMSE of 0.687)

6 Conclusions

Motivated in part by the potential value of conditional forecasting in today's environment in which some central banks have indicated policy rates are likely to remain very low fairly far into the future and the lack of a developed theory for evaluating conditional forecasts, this paper provides analytical, Monte Carlo, and empirical evidence on tests of predictive ability for conditional forecasts from estimated models. While our theory extends to other tests, we focus on two basic, commonly-used tests of forecast model specification: tests of bias (zero mean prediction error) and efficiency (no correlation between the prediction error and prediction).

For these tests, we establish asymptotic normality in the context of VAR-based conditional forecasts. Our results follow from an application of West (1996) and as such establish the role of estimation error on the asymptotic distribution. As a practical matter, the standard errors can be quite complex and as such we consider a bootstrap approach to inference that is valid when even estimation error contributes to the asymptotic variance of the test statistic. Monte Carlo evidence suggests that the tests can be reasonably well sized in samples of the size often seen in macroeconomic applications.

Building on these results, we evaluate unconditional and conditional forecasts from common macroeconomic VARs. We produce conditional forecasts of GDP growth, unemployment, and inflation on an assumption that the federal funds rate is unchanged over an eight quarter forecast horizon. While simply comparing bias and efficiency tests to standard normal critical values would yield a wide range of rejections of efficiency, comparing the tests against bootstrap critical reduces the evidence of misspecification problems. However, even with bootstrap critical values, we find evidence of misspecification in unconditional and conditional forecasts of inflation, represented by rejections of forecast efficiency.

7 Appendix

7.1 Estimation procedure for BVAR5-SSP

The model is estimated with a 3-step Gibbs sampler, as in Villani (2009).

Step 1: Draw the slope coefficients Π conditional on Ψ and Σ.

For this step, the VAR is recast in demeaned form, using $Y_t = y_t - \Psi d_t$:

$$Y_t \quad = \quad (I_p \otimes X_t') \cdot \text{vec}(\Pi) + v_t, , \tag{12}$$

where X_t contains the appropriate lags of Y_t and vec(Π) contains the VAR slope coefficients.

The vector of coefficients is sampled from a normal posterior distribution with mean $\bar{\mu}_\Pi$ and variance $\bar{\Omega}_\Pi$, based on prior mean μ_Π and Ω_Π, where:

$$\bar{\Omega}_\Pi^{-1} \quad = \quad \Omega_\Pi^{-1} + \Sigma^{-1} \otimes X'X \tag{13}$$

$$\bar{\mu}_\Pi \quad = \quad \bar{\Omega}_\Pi \left\{ \text{vec}(X'^{-1}) + \Omega_\Pi^{-1}\mu_\Pi \right\}. \tag{14}$$

Step 2: Draw the steady state coefficients Ψ conditional on Π and Σ.

For this step, the VAR is rewritten as

$$q_t = \Pi(L)\Psi d_t + v_t, \text{ where } q_t \equiv \Pi(L)y_t. \tag{15}$$

The dependent variable q_t is obtained by applying to the vector y_t the lag polynomial estimated with the preceding draw of the Π coefficients. The right-hand side term $\Pi(L)\Psi d_t$ simplifies to $\Theta\bar{d}_t$, where, as in Villani (2009) with some modifications, $\bar{d}_t$ contains current and lagged values of the elements of d_t, and Θ is defined such that vec(Θ) $= U$vec(Ψ):

$$\bar{d}_t = (d_t', -d_{t-1}', -d_{t-2}', -d_{t-3}', \ldots, -d_{t-k}')' \tag{16}$$

$$U = \begin{pmatrix} I_{pq\times pq} \\ I_q \otimes \Pi_1 \\ I_q \otimes \Pi_2 \\ I_q \otimes \Pi_3 \\ \vdots \\ I_q \otimes \Pi_k \end{pmatrix}. \tag{17}$$

The vector of coefficients Ψ is sampled from a normal posterior distribution with mean $\bar{\mu}_\Psi$ and variance $\bar{\Omega}_\Psi$, based on prior mean μ_Ψ and Ω_Ψ, where:

$$\bar{\Omega}_\Psi^{-1} \quad = \quad \Omega_\Psi^{-1} + U' \left(D'^{-1}\right) U \tag{18}$$

$$\bar{\mu}_\Psi \quad = \quad \bar{\Omega}_\Psi \left\{ U'\text{vec}(\Sigma^{-1}Y'D) + \Omega_\Psi^{-1}\mu_\Psi \right\}. \tag{19}$$

Step 3: Draw Σ conditional on Π and Ψ.

The sampling of Σ, the variance-covariance matrix of innovations to the VAR, is based on an inverted Wishart posterior. The posterior scaling matrix is the sum of squared residuals, a combination of the prior and the sample moment $\sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$, and the degrees of freedom is the size of the estimation sample, T.

7.2 Estimation procedure for BVAR22-NW

By grouping the coefficient matrices in the $N \times M$ matrix $\Pi' = [\Pi_c \ \Pi_1 \ ... \ \Pi_p]$ and defining $x_t = (1 \ y_{t-1}' \ ... \ y_{t-p}')'$ as a vector containing an intercept and p lags of y_t, the VAR can be written as:

$$y_t = \Pi' x_t + \varepsilon_t. \tag{20}$$

An even more compact notation is:

$$Y = X\Pi + E, \tag{21}$$

where $Y = [y_1, .., y_T]'$, $X = [x_1, .., x_T]'$, and $E = [\varepsilon_1, .., \varepsilon_T]'$ are, respectively, $T \times N$, $T \times M$ and $T \times N$ matrices.

With the large model, we use the conjugate Normal-inverted Wishart prior:

$$\Pi|\Sigma \sim N(\Pi_0, \Sigma \otimes \Omega_0), \ \Sigma \sim IW(S_0, v_0). \tag{22}$$

As the N-IW prior is conjugate, the conditional posterior distribution of this model is also N-IW (Zellner 1971):

$$\Pi|\Sigma, Y \sim N(\bar{\Pi}, \Sigma \otimes \bar{\Omega}), \ \Sigma|Y \sim IW(\bar{S}, \bar{v}). \tag{23}$$

Defining $\hat{\Pi}$ and $\hat{E}$ as the OLS estimates, we have that $\bar{\Pi} = (\Omega_0^{-1} + X'X)^{-1}(\Omega_0^{-1}\Pi_0 + X'Y)$, $\bar{\Omega} = (\Omega_0^{-1} + X'X)^{-1}$, $\bar{v} = v_0 + T$, and $\bar{S} = \Pi_0 + \hat{E}'\hat{E} + \hat{\Pi}'X'X\hat{\Pi} + \Pi_0'\Omega_0^{-1}\Pi_0 - \bar{\Pi}'\bar{\Omega}^{-1}\bar{\Pi}$. Sources such as Kadiyala and Karlsson (1997) and Banbura, Giannone, and Reichlin (2010) provide additional detail on the N-IW prior and posterior.

7.3 Additional details on the prior for BVAR22-NW

The "sum of coefficients" prior expresses a belief that when the average of lagged values of a variable is at some level $\bar{y}_{0i}$, that same value $\bar{y}_{0i}$ is likely to be a good forecast of future observations, and is implemented by augmenting the system in (21) with the dummy observations Y_{d_1} and X_{d_1} with generic elements:

$$y_d(i,j) = \begin{cases} \bar{y}_{0i}/\lambda_3 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \; ; \quad x_d(i,s) = \begin{cases} \bar{y}_{0i}/\lambda_3 & \text{if } i = j,\ s < M \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

where i and j go from 1 to N while s goes from 1 to M. When $\lambda_3 \to 0$ the model tends to a form that can be expressed entirely in terms of differenced data, there are as many unit roots as variables and there is no cointegration.

The "dummy initial observation" prior introduces a single dummy observation such that all values of all variables are set equal to the corresponding averages of initial conditions up to a scaling factor $(1/\lambda_4)$. It is implemented by adding to the system in (21) the dummy variables Y_{d_2} and X_{d_2} with generic elements:

$$y_d(j) = \bar{y}_{0j}/\lambda_4; \quad x_d(s) = \begin{cases} \bar{y}_{0j}/\lambda_4 & \text{for } s < M \\ 1/\lambda_4 & \text{for } \ s = M, \end{cases} \tag{25}$$

where j goes from 1 to N while s goes from 1 to M. As $\lambda_4 \to 0$ the model tends to a form in which either all variables are stationary with means equal to the sample averages of the initial conditions, or there are unit root components without drift terms, which is consistent with cointegration.

References

Aastveit, Knut Are, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino (2014), "Have Standard VARs Remained Stable Since the Crisis?" manuscript.

Amisano, Gianni, and Rafaella Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics* 25, 177-190.

Andrews, Donald W.K., and J. Christopher Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60, 953-966.

Banbura, Michele, Domenico Giannone, and Lucrezia Reichlin (2010), "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics* 25, 71-92.

Bernanke, Ben S., and Alan S. Blinder (1992), "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review* 82, 901-921.

Carriero Andrea, Todd E. Clark, and Massimiliano Marcellino (2012), "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, forthcoming.

Christensen, Jens H.E., and Glenn D. Rudebusch (2013), "Estimating Shadow-Rate Term Structure Models with Near-Zero Yields," *Federal Reserve Bank of San Francisco Working Paper 2013-07.*

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (1996), "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics* 78, 16-34.

Clarida, Richard H., and Coyle, Diane (1984), "Conditional Projection by Means of Kalman Filtering." Mimeo, National Bureau of Economic Research1984.

Clark, Todd E. (2011), "Real-Time Density Forecasts from BVARs with Stochastic Volatility," *Journal of Business and Economic Statistics* 29, 327-341.

Clark, Todd E., and Michael W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.

Clark, Todd E., and Michael W. McCracken (2005), "Evaluating Direct Multistep Forecasts," *Econometric Reviews* 24, 369-404.

Clark, Todd E. and Michael W. McCracken (2008), "Forecasting with Small Macroeconomic VARs in the Presence of Instabilities," in Forecasting in the Presence of Structural Breaks and Model Uncertainty, ed. D.E. Rapach and M.E. Wohar, (Amsterdam: Elsevier).

Clark, Todd E., and Michael W. McCracken (2009), "Tests of Equal Predictive Ability with Real-Time Data," *Journal of Business and Economic Statistics* 27, 441-454.

Clark, Todd E., and Michael W. McCracken (2010), ""Averaging Forecasts from VARs with Uncertain Instabilities," *Journal of Applied Econometrics* 25, 5-29.

Clark, Todd E., and Michael W. McCracken (2011a), "Nested Forecast Model Comparisons:

A New Approach to Testing Equal Accuracy," manuscript, Federal Reserve Bank of St. Louis.

Clark, Todd E., and Michael W. McCracken (2011b), "Testing for Unconditional Predictive Ability," in *Oxford Handbook of Economic Forecasting*, Michael P. Clements and David F. Hendry, eds., Oxford: Oxford University Press.

Clark, Todd E., and Michael W. McCracken (2012a), "Reality Checks and Comparisons of Nested Predictive Models," *Journal of Business and Economic Statistics* 30, 53-66.

Clark, Todd E., and Michael W. McCracken (2013a), "Advances in Forecast Evaluation," in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, Amsterdam: North-Holland.

Clark, Todd E., and Michael W. McCracken (2013b), "Evaluating the Accuracy of Forecasts from Vector Autoregressions," in *Advances in Econometrics: VAR models in Macroeconomics — New Developments and Applications*, Thomas Fomby, Lutz Kilian, and Anthony Murphy (eds.), Emerald, 117-168.

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

Clark, Todd E., and Kenneth D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* 138, 291-311.

Del Negro, Marco, and Frank Schorfheide (2013), "DSGE Model-Based Forecasting," in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, Amsterdam: North-Holland.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Doan, T., Litterman, R., and Sims, C., 1984. Forecasting and Conditional Projection Using Realistic Prior Distributions, *Econometric Reviews* 3, 1-100.

Faust, Jon, and Jonathan H. Wright (2008), "Efficient Forecast Tests for Conditional Policy Forecasts," *Journal of Econometrics* 146, 293-303.

Gavin, William T. (2003), "FOMC Forecast: Is All the Information in the Central Tendency?" *Federal Reserve Bank of St. Louis Review*, May/June, 85, 27-48.

Gavin, William T. and Mandal, Rachel J. (2003), "Evaluating FOMC Forecasts," *International Journal of Forecasting*, 19, 655-67.

Gavin, William T. and Geetanjali Pande (2008), "FOMC Consensus Forecasts," *Federal Reserve Bank of St. Louis Review*, May/June, 90 (Part 1), 146-163.

Giacomini, Rafaella, and Halbert White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74, 1545-1578.

Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2012), "Prior Selection for Vector Autoregressions," mimeo, Free University of Brussels.

Giannone, D., Lenza, M., Momferatou, D., and Onorante L., 2010. Short-Term Inflation

Projections: A Bayesian Vector Autoregressive Approach, manuscript, European Central Bank.

Goncalves, Silvia, and Lutz Kilian (2004), “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form,” *Journal of Econometrics*, 123, 89-120.

Granger, Clive W.J., and Paul Newbold (1977), Forecasting Economic Time Series, Academic Press Inc., London.

Herbst, Edward, and Frank Schorfheide (2012), “Evaluating DSGE Model Forecasts of Comovements,’ ’*Journal of Econometrics*, 171, 152-166.

Jarocinski, Marek (2010), “Conditional Forecasts and Uncertainty about Forecast Revisions in Vector Autoregressions,” *Economics Letters* 108, 257-259.

Joutz, Fred and Herman Stekler (2000), “An Evaluation of the Predictions of the Federal Reserve,” *International Journal of Forecasting* 16, 17-38.

Kadiyala, K. Rao, and Sune Karlsson (1997), ”Numerical Methods for Estimation and Inference in Bayesian VAR-Models,” *Journal of Applied Econometrics* 12, 99-132.

Karlsson, Sune (2012), “Forecasting with Bayesian VAR models,” forthcoming in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, Amsterdam: North-Holland.

Kilian, Lutz (1999), “Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?” *Journal of Applied Econometrics* 14, 491-510.

Kim, Don H., and Kenneth J. SIngleton (2012), “Term Structure Models and the Zero Bound: An Empirical Investigation of Japanese Yields,” *Journal of Econometrics* 170, 32-49.

Kim, Myung J., and Charles R. Nelson (1993), “Predictable Stock Returns: The Role of Small Sample Bias,” *Journal of Finance* 48, 641-661.

Koop, Gary (2012), “Forecasting with Medium and Large Bayesian VARs,” *Journal of Applied Econometrics*, forthcoming.

Krippner, Leo (2013), “Measuring the Stance of Monetary Policy in Zero Lower Bound Environments,” *Economics Letters* 118, 135-138.

Leeper, Eric M., and Tao Zha (2003), “Modest Policy Interventions,” *Journal of Monetary Economics* 50, 1673-1700.

Leeper, E.M., Sims, C., Zha, T., 1996. What Does Monetary Policy Do? *Brookings Papers on Economic Activity* 27(2),1-78.

Litterman, Robert B. (1986), “Forecasting with Bayesian Vector Autoregressions-Five Years of Experience,” *Journal of Business and Economic Statistics* 4, 25-38.

McCracken, Michael W. (2000), “Robust Out-of-Sample Inference,” *Journal of Econometrics* 99, 195-223.

McCracken, Michael W. (2004), “Parameter Estimation Error and Tests of Equal Forecast Accuracy Between Non-nested Models,” *International Journal of Forecasting* 20, 503-

514.

McCracken, Michael W. (2007), "Asymptotics for Out-of-Sample Tests of Granger Causality," *Journal of Econometrics* 140, 719-752.

Meyer, Brent, and Saeed Zaman (2013), "It's Not Just for Inflation: The Usefulness of the Median CPI in BVAR Forecasting," Federal Reserve Bank of Cleveland Working Paper Working Paper 13-03.

Newey, Whitney K., and Daniel L. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in R.F. Engle and D.L. Newey (eds.) *Handbook of Econometrics*, 4, 2113-2247, Elsevier Science..

Newey, Whitney K., and Kenneth D. West (1987), "A simple, positive semi–definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Patton, Andrew J., and Allan Timmermann (2012), "Forecast Rationality Tests Based on Multi-Horizon Bounds" *Journal of Business and Economic Statistics* 30, 1-17.

Reifschneider, David, and Peter Tulip (2007), "Gauging the Uncertainty of the Economic Outlook from Historical Forecasting Errors" *FEDS WP* 2007-60.

Romer, Christina D., and David H. Romer (2008), "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?" *American Economic Review* 98, 230-235.

Romer, Christina D., and David H. Romer (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review* 90, 429-457.

Schorfheide, Frank, and Dongho Song (2013), "Real Time Forecasting with a Mixed-Frequency VAR," NBER Working Paper 19712.

Sims, Christopher A. (1982), "Policy Analysis with Econometric Models," *Brookings Papers on Economic Activity*, No. 1, 107-152.

Sims, Christopher A. (1980), "Macroeconomics and Reality," *Econometrica* 48, 1-48.

Sims, C., and Zha, T., 1998. Bayesian Methods for Dynamic Multivariate Models, *International Economic Review* 39(4), 949–68.

Stine, Robert A. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association* 82, 1072-1078.

Waggoner D.F., and Zha T. 1999. Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics* 81(4), 639-651

West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084.

West, Kenneth D. (2006), "Forecast Evaluation," in *Handbook of Economic Forecasting*, Elliott G., Granger C.W.J., Timmermann, A. (eds), North Holland.

West, Kenneth D., and Michael W. McCracken (1998), "Regression-based Tests of Predictive Ability," *International Economic Review* 39, 817-840.

Zellner, Arnold (1971), An Introduction to Bayesian Inference in Econometrics, J. Wiley and Sons, Inc., New York.

White, Halbert (2000), "A Reality Check For Data Snooping," *Econometrica* 68, 1097-1127.

Table 1. Monte Carlo DGP coefficients

explanatory variable	$y_{1,t}$ equation	$y_{2,t}$ equation	$y_{3,t}$ equation
DGP 1 (size)			
$y_{1,t-1}$	0.50	0.00	
$y_{2,t-1}$	0.10	0.80	
intercept	0.00	0.00	
DGP 2 (size)			
$y_{1,t-1}$	0.25	0.00	
$y_{2,t-1}$	-0.50	0.80	
intercept	0.00	0.00	
DGP 3 (size)			
$y_{1,t-1}$	0.234	0.029	0.059
$y_{1,t-2}$	0.164	-0.039	0.031
$y_{2,t-1}$	-0.134	0.575	0.038
$y_{2,t-2}$	-0.150	0.138	0.019
$y_{3,t-1}$	-0.057	0.200	1.006
$y_{3,t-2}$	-0.165	-0.184	-0.087
intercept	2.425	0.054	-0.110
DGP 4 (power)			
$y_{1,t-1}$, pre-break	0.50	0.00	
$y_{1,t-1}$, post-break	0.50	0.25	
$y_{2,t-1}$, pre-break	0.10	0.80	
$y_{2,t-1}$, post-break	0.10	0.40	
intercept, pre-break	0.00	0.00	
intercept, post-break	0.50	0.40	

Notes:
1. The table provides the coefficients of Monte Carlo DGPs 1-4.
2. The variance-covariance matrix of innovations and other aspects of the Monte Carlo design are described in section 4.1.

Table 2: Monte Carlo Results on Size, DGP 1

(nominal size = 10%, forecast horizon = 1 period)

test	*source of critical values*	R=50 P=100	R=50 P=150	R=100 P=50	R=100 P=100	R=100 P=150
OLS estimation						
bias, uncond.	Normal	0.128	0.105	0.130	0.110	0.111
bias, condit.-RF	Normal	0.100	0.087	0.123	0.096	0.078
efficiency, uncond.	Normal	0.133	0.131	0.107	0.114	0.119
efficiency, condit.-RF	Normal	0.188	0.168	0.245	0.168	0.152
bias, uncond.	bootstrap	0.109	0.090	0.107	0.091	0.097
bias, condit.-RF	bootstrap	0.108	0.100	0.101	0.098	0.087
efficiency, uncond.	bootstrap	0.097	0.100	0.095	0.090	0.114
efficiency, condit.-RF	bootstrap	0.066	0.080	0.077	0.072	0.089
BVAR estimation						
bias, uncond.	Normal	0.154	0.129	0.152	0.127	0.124
bias, condit.-RF	Normal	0.118	0.106	0.144	0.107	0.089
efficiency, uncond.	Normal	0.081	0.080	0.091	0.089	0.103
efficiency, condit.-RF	Normal	0.130	0.118	0.200	0.143	0.142
bias, uncond.	bootstrap	0.112	0.098	0.108	0.092	0.095
bias, condit.-RF	bootstrap	0.107	0.101	0.104	0.104	0.088
efficiency, uncond.	bootstrap	0.109	0.105	0.108	0.097	0.115
efficiency, condit.-RF	bootstrap	0.072	0.082	0.079	0.076	0.086

Notes:
1. The data generating process is a bivariate VAR(1), with coefficients given in Table 1 and error variance matrix given in section 4.1.
2. For each artificial data set, forecasts of $y_{1,t+1}$ are formed recursively using estimates of a bivariate VAR(1). The top panel of the table provides results for models estimated by OLS; the lower panel provides results for models estimated with Bayesian methods, under a Normal-inverted Wishart prior detailed in section 4.2. We consider both unconditional forecasts and conditional forecasts, obtained under the DLS approach. The conditional forecasts of $y_{1,t+1}$ are based on a condition of $\hat{y}^c_{2,t+1} = y_{2,t}$.
3. These forecasts are then used to form bias and efficient tests, detailed in sections 3 and 4.3. For unconditional forecasts, the test statistics use OLS-estimated variances. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of 6.
4. R and $\tilde{P}$ refer to the number of in–sample observations and forecasts, respectively.
5. In each Monte Carlo replication, the simulated test statistics are compared against standard normal critical values and critical values obtained with a wild bootstrap of the VAR, as described in section 3.3.
6. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 3: Monte Carlo Results on Size, DGP 2

(nominal size = 10%, forecast horizon = 1 period)

OLS estimation						
	source of	R=50	R=50	R=100	R=100	R=100
test	*critical values*	P=100	P=150	P=50	P=100	P=150
bias, uncond.	Normal	0.128	0.109	0.132	0.107	0.109
bias, condit.-RF	Normal	0.066	0.053	0.090	0.066	0.051
efficiency, uncond.	Normal	0.153	0.140	0.158	0.112	0.108
efficiency, condit.-RF	Normal	0.617	0.626	0.466	0.474	0.486
bias, uncond.	bootstrap	0.106	0.097	0.104	0.089	0.089
bias, condit.-RF	bootstrap	0.101	0.094	0.103	0.092	0.088
efficiency, uncond.	bootstrap	0.100	0.100	0.105	0.096	0.093
efficiency, condit.-RF	bootstrap	0.062	0.053	0.068	0.059	0.050
BVAR estimation						
bias, uncond.	Normal	0.156	0.145	0.150	0.132	0.129
bias, condit.-RF	Normal	0.079	0.067	0.104	0.071	0.060
efficiency, uncond.	Normal	0.081	0.070	0.121	0.081	0.082
efficiency, condit.-RF	Normal	0.320	0.304	0.315	0.263	0.238
bias, uncond.	bootstrap	0.107	0.096	0.105	0.095	0.093
bias, condit.-RF	bootstrap	0.109	0.102	0.108	0.098	0.090
efficiency, uncond.	bootstrap	0.111	0.104	0.116	0.102	0.097
efficiency, condit.-RF	bootstrap	0.062	0.050	0.062	0.051	0.058

Notes:
1. The data generating process is a bivariate VAR(1), with coefficients given in Table 1 and error variance matrix given in section 4.1.
2. For each artificial data set, forecasts of $y_{1,t+1}$ are formed recursively using estimates of a bivariate VAR(1). The top panel of the table provides results for models estimated by OLS; the lower panel provides results for models estimated with Bayesian methods, under a Normal-inverted Wishart prior detailed in section 4.2. We consider both unconditional forecasts and conditional forecasts, obtained under the DLS approach. The conditional forecasts of $y_{1,t+1}$ are based on a condition of $\hat{y}^c_{2,t+1} = y_{2,t}$.
3. These forecasts are then used to form bias and efficient tests, detailed in sections 3 and 4.3. For unconditional forecasts, the test statistics use OLS-estimated variances. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of 6.
4. R and $\tilde{P}$ refer to the number of in–sample observations and forecasts, respectively.
5. In each Monte Carlo replication, the simulated test statistics are compared against standard normal critical values and critical values obtained with a wild bootstrap of the VAR, as described in section 3.3.
6. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 4: Monte Carlo Results on Size, DGP 3, OLS Estimation
(nominal size = 10%)

test	*source of critical values*	forecasts of $y_{1,t}$ $\tau = 1$	$\tau = 2$	$\tau = 4$	forecasts of $y_{2,t}$ $\tau = 1$	$\tau = 2$	$\tau = 4$
R,P = 50,100							
bias, uncond.	Normal	0.130	0.149	0.152	0.151	0.180	0.178
bias, condit.-RF	Normal	0.081	0.092	0.151	0.121	0.130	0.155
bias, condit.-pol. shock	Normal	0.149	0.158	0.186	0.147	0.153	0.173
efficiency, uncond.	Normal	0.409	0.327	0.333	0.259	0.392	0.610
efficiency, condit.-RF	Normal	0.518	0.420	0.485	0.307	0.472	0.706
effic., condit.-pol. shock	Normal	0.433	0.393	0.391	0.296	0.413	0.695
bias, uncond.	bootstrap	0.101	0.093	0.100	0.103	0.105	0.098
bias, condit.-RF	bootstrap	0.097	0.095	0.095	0.105	0.101	0.100
bias, condit.-pol. shock	bootstrap	0.103	0.093	0.104	0.104	0.102	0.095
efficiency, uncond.	bootstrap	0.071	0.072	0.074	0.089	0.088	0.073
efficiency, condit.-RF	bootstrap	0.050	0.067	0.088	0.083	0.087	0.087
effic., condit.-pol. shock	bootstrap	0.062	0.070	0.082	0.083	0.090	0.089
R,P = 50,150							
bias, uncond.	Normal	0.123	0.143	0.138	0.157	0.178	0.174
bias, condit.-RF	Normal	0.073	0.079	0.127	0.120	0.123	0.142
bias, condit.-pol. shock	Normal	0.118	0.136	0.155	0.142	0.141	0.160
efficiency, uncond.	Normal	0.433	0.339	0.300	0.252	0.360	0.615
efficiency, condit.-RF	Normal	0.476	0.360	0.438	0.274	0.432	0.670
effic., condit.-pol. shock	Normal	0.442	0.362	0.317	0.275	0.366	0.639
bias, uncond.	bootstrap	0.097	0.098	0.092	0.115	0.110	0.107
bias, condit.-RF	bootstrap	0.094	0.097	0.098	0.118	0.114	0.106
bias, condit.-pol. shock	bootstrap	0.093	0.093	0.096	0.108	0.107	0.106
efficiency, uncond.	bootstrap	0.068	0.077	0.070	0.099	0.082	0.074
efficiency, condit.-RF	bootstrap	0.059	0.060	0.072	0.088	0.092	0.085
effic., condit.-pol. shock	bootstrap	0.065	0.068	0.071	0.089	0.096	0.079
R,P = 100,50							
bias, uncond.	Normal	0.129	0.160	0.193	0.139	0.194	0.218
bias, condit.-RF	Normal	0.116	0.144	0.212	0.151	0.172	0.217
bias, condit.-pol. shock	Normal	0.160	0.192	0.246	0.172	0.188	0.238
efficiency, uncond.	Normal	0.201	0.234	0.291	0.207	0.335	0.520
efficiency, condit.-RF	Normal	0.375	0.368	0.456	0.342	0.474	0.623
effic., condit.-pol. shock	Normal	0.281	0.297	0.372	0.312	0.419	0.608
bias, uncond.	bootstrap	0.094	0.091	0.097	0.097	0.102	0.107
bias, condit.-RF	bootstrap	0.099	0.104	0.100	0.101	0.099	0.099
bias, condit.-pol. shock	bootstrap	0.089	0.092	0.093	0.099	0.102	0.103
efficiency, uncond.	bootstrap	0.078	0.081	0.089	0.090	0.088	0.085
efficiency, condit.-RF	bootstrap	0.076	0.070	0.091	0.081	0.092	0.112
effic., condit.-pol. shock	bootstrap	0.076	0.070	0.090	0.081	0.084	0.105

Table 4, continued: Monte Carlo Results on Size, DGP 3, OLS Estimation
(nominal size = 10%)

		forecasts of $y_{1,t}$			forecasts of $y_{2,t}$		
test	*source of critical values*	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 4$
R,P = 100,100							
bias, uncond.	Normal	0.112	0.136	0.157	0.141	0.172	0.184
bias, condit.-RF	Normal	0.080	0.097	0.153	0.126	0.123	0.157
bias, condit.-pol. shock	Normal	0.126	0.146	0.184	0.140	0.146	0.169
efficiency, uncond.	Normal	0.241	0.220	0.234	0.197	0.274	0.440
efficiency, condit.-RF	Normal	0.338	0.288	0.376	0.258	0.352	0.564
effic., condit.-pol. shock	Normal	0.280	0.247	0.287	0.243	0.311	0.523
bias, uncond.	bootstrap	0.095	0.092	0.101	0.103	0.096	0.092
bias, condit.-RF	bootstrap	0.105	0.107	0.108	0.116	0.101	0.101
bias, condit.-pol. shock	bootstrap	0.100	0.098	0.102	0.104	0.095	0.096
efficiency, uncond.	bootstrap	0.096	0.097	0.092	0.099	0.088	0.083
efficiency, condit.-RF	bootstrap	0.079	0.069	0.085	0.088	0.101	0.086
effic., condit.-pol. shock	bootstrap	0.077	0.082	0.086	0.090	0.095	0.074
R,P = 100,150							
bias, uncond.	Normal	0.124	0.141	0.141	0.136	0.168	0.172
bias, condit.-RF	Normal	0.073	0.080	0.135	0.110	0.115	0.132
bias, condit.-pol. shock	Normal	0.127	0.146	0.172	0.136	0.141	0.157
efficiency, uncond.	Normal	0.251	0.218	0.217	0.183	0.265	0.436
efficiency, condit.-RF	Normal	0.327	0.262	0.336	0.226	0.312	0.546
effic., condit.-pol. shock	Normal	0.272	0.261	0.246	0.224	0.275	0.514
bias, uncond.	bootstrap	0.100	0.097	0.092	0.107	0.106	0.106
bias, condit.-RF	bootstrap	0.102	0.097	0.100	0.101	0.104	0.103
bias, condit.-pol. shock	bootstrap	0.095	0.099	0.100	0.106	0.103	0.106
efficiency, uncond.	bootstrap	0.099	0.101	0.094	0.098	0.104	0.100
efficiency, condit.-RF	bootstrap	0.081	0.079	0.098	0.097	0.099	0.090
effic., condit.-pol. shock	bootstrap	0.090	0.095	0.095	0.099	0.105	0.099

Notes:
1. The data generating process is a trivariate VAR(2), with coefficients given in Table 1 and error variance matrix given in section 4.1.
2. For each artificial data set, forecasts of $y_{1,t+\tau}$ and $y_{2,t+\tau}$ are formed recursively using OLS estimates of a trivariate VAR(2) and an iterative approach to computing multi-step forecasts. We consider both unconditional forecasts and conditional forecasts, obtained under both the DLS approach and the policy shock approach. The conditional forecasts of $y_{1,t+\tau}$ and $y_{2,t+\tau}$ are based on a condition of $\hat{y}^c_{3,t+h} = y_{3,t}$, $h = 1, \ldots, 4$.
3. These forecasts are then used to form bias and efficient tests, detailed in sections 3 and 4.3. For unconditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $2(\tau - 1)$. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $6 + 2(\tau - 1)$.
4. R and $\tilde{P}$ refer to the number of in–sample observations and forecasts, respectively.
5. In each Monte Carlo replication, the simulated test statistics are compared against standard normal critical values and critical values obtained with a wild bootstrap of the VAR, as described in section 4.3.
6. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 5: Monte Carlo Results on Size, DGP 3, BVAR Estimation

(nominal size = 10%)

		forecasts of $y_{1,t}$			forecasts of $y_{2,t}$		
test	*source of critical values*	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 4$
R,P = 50,100							
bias, uncond.	Normal	0.155	0.155	0.142	0.246	0.229	0.199
bias, condit.-RF	Normal	0.088	0.101	0.147	0.166	0.164	0.178
bias, condit.-pol. shock	Normal	0.140	0.149	0.178	0.180	0.181	0.186
efficiency, uncond.	Normal	0.127	0.149	0.209	0.102	0.154	0.434
efficiency, condit.-RF	Normal	0.186	0.259	0.382	0.143	0.251	0.589
effic., condit.-pol. shock	Normal	0.140	0.208	0.308	0.139	0.187	0.524
bias, uncond.	bootstrap	0.102	0.106	0.102	0.124	0.120	0.109
bias, condit.-RF	bootstrap	0.094	0.098	0.097	0.119	0.115	0.108
bias, condit.-pol. shock	bootstrap	0.101	0.100	0.100	0.113	0.110	0.104
efficiency, uncond.	bootstrap	0.088	0.081	0.075	0.107	0.096	0.083
efficiency, condit.-RF	bootstrap	0.062	0.071	0.110	0.086	0.088	0.097
effic., condit.-pol. shock	bootstrap	0.073	0.071	0.087	0.084	0.088	0.096
R,P = 50,150							
bias, uncond.	Normal	0.148	0.155	0.136	0.233	0.227	0.195
bias, condit.-RF	Normal	0.078	0.087	0.128	0.163	0.151	0.160
bias, condit.-pol. shock	Normal	0.118	0.130	0.157	0.175	0.170	0.175
efficiency, uncond.	Normal	0.111	0.120	0.152	0.109	0.129	0.403
efficiency, condit.-RF	Normal	0.135	0.201	0.329	0.130	0.171	0.482
effic., condit.-pol. shock	Normal	0.116	0.167	0.222	0.127	0.145	0.407
bias, uncond.	bootstrap	0.105	0.107	0.100	0.130	0.122	0.112
bias, condit.-RF	bootstrap	0.098	0.097	0.104	0.127	0.126	0.116
bias, condit.-pol. shock	bootstrap	0.098	0.099	0.099	0.118	0.118	0.114
efficiency, uncond.	bootstrap	0.088	0.080	0.071	0.117	0.084	0.065
efficiency, condit.-RF	bootstrap	0.067	0.070	0.092	0.093	0.081	0.081
effic., condit.-pol. shock	bootstrap	0.077	0.074	0.094	0.092	0.087	0.088
R,P = 100,50							
bias, uncond.	Normal	0.146	0.169	0.199	0.216	0.243	0.245
bias, condit.-RF	Normal	0.126	0.153	0.225	0.196	0.212	0.243
bias, condit.-pol. shock	Normal	0.162	0.193	0.245	0.209	0.231	0.263
efficiency, uncond.	Normal	0.119	0.185	0.267	0.117	0.211	0.421
efficiency, condit.-RF	Normal	0.249	0.293	0.414	0.217	0.348	0.566
effic., condit.-pol. shock	Normal	0.203	0.243	0.330	0.210	0.301	0.531
bias, uncond.	bootstrap	0.102	0.099	0.113	0.117	0.118	0.117
bias, condit.-RF	bootstrap	0.102	0.101	0.105	0.110	0.112	0.112
bias, condit.-pol. shock	bootstrap	0.093	0.092	0.102	0.110	0.107	0.110
efficiency, uncond.	bootstrap	0.093	0.090	0.090	0.104	0.092	0.087
efficiency, condit.-RF	bootstrap	0.081	0.076	0.094	0.082	0.098	0.106
effic., condit.-pol. shock	bootstrap	0.081	0.077	0.102	0.088	0.092	0.108

Table 5, continued: Monte Carlo Results on Size, DGP 3, BVAR Estimation

(nominal size = 10%)

		forecasts of $y_{1,t}$			forecasts of $y_{2,t}$		
test	*source of critical values*	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 4$
R,P = 100,100							
bias, uncond.	Normal	0.134	0.153	0.159	0.212	0.223	0.213
bias, condit.-RF	Normal	0.091	0.105	0.158	0.161	0.167	0.181
bias, condit.-pol. shock	Normal	0.131	0.147	0.186	0.176	0.182	0.198
efficiency, uncond.	Normal	0.105	0.142	0.182	0.112	0.147	0.310
efficiency, condit.-RF	Normal	0.165	0.194	0.317	0.160	0.208	0.446
effic., condit.-pol. shock	Normal	0.140	0.182	0.246	0.160	0.187	0.377
bias, uncond.	bootstrap	0.101	0.107	0.110	0.118	0.111	0.108
bias, condit.-RF	bootstrap	0.102	0.104	0.110	0.125	0.116	0.108
bias, condit.-pol. shock	bootstrap	0.104	0.103	0.111	0.110	0.105	0.104
efficiency, uncond.	bootstrap	0.098	0.095	0.089	0.107	0.095	0.080
efficiency, condit.-RF	bootstrap	0.077	0.074	0.101	0.090	0.102	0.088
effic., condit.-pol. shock	bootstrap	0.080	0.080	0.088	0.092	0.090	0.088
R,P = 100,150							
bias, uncond.	Normal	0.139	0.149	0.141	0.194	0.198	0.181
bias, condit.-RF	Normal	0.078	0.084	0.135	0.137	0.139	0.152
bias, condit.-pol. shock	Normal	0.124	0.143	0.166	0.157	0.157	0.164
efficiency, uncond.	Normal	0.104	0.131	0.162	0.107	0.143	0.304
efficiency, condit.-RF	Normal	0.146	0.161	0.281	0.153	0.168	0.404
effic., condit.-pol. shock	Normal	0.123	0.157	0.193	0.144	0.164	0.344
bias, uncond.	bootstrap	0.108	0.105	0.100	0.115	0.113	0.112
bias, condit.-RF	bootstrap	0.104	0.096	0.103	0.108	0.112	0.111
bias, condit.-pol. shock	bootstrap	0.100	0.104	0.103	0.108	0.104	0.107
efficiency, uncond.	bootstrap	0.101	0.100	0.100	0.109	0.105	0.101
efficiency, condit.-RF	bootstrap	0.087	0.075	0.104	0.100	0.092	0.093
effic., condit.-pol. shock	bootstrap	0.093	0.075	0.086	0.095	0.105	0.085

Notes:
1. The data generating process is a trivariate VAR(2), with coefficients given in Table 1 and error variance matrix given in section 4.1.
2. For each artificial data set, forecasts of $y_{1,t+\tau}$ and $y_{2,t+\tau}$ are formed recursively using Bayesian estimates (under a Normal-inverted Wishart prior detailed in section 4.2) of a trivariate VAR(2) and an iterative approach to computing multi-step forecasts. We consider both unconditional forecasts and conditional forecasts, obtained under both the DLS approach and the policy shock approach. The conditional forecasts of $y_{1,t+\tau}$ and $y_{2,t+\tau}$ are based on a condition of $\hat{y}^c_{3,t+h} = y_{3,t}$, $h = 1, \ldots, 4$.
3. These forecasts are then used to form bias and efficient tests, detailed in sections 3 and 4.3. For unconditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $2(\tau - 1)$. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $6 + 2(\tau - 1)$.
4. R and $\tilde{P}$ refer to the number of in–sample observations and forecasts, respectively.
5. In each Monte Carlo replication, the simulated test statistics are compared against standard normal critical values and critical values obtained with a wild bootstrap of the VAR, as described in section 3.3.
6. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 6: Monte Carlo Results on Power, DGP 4

(nominal size = 10%, forecast horizon = 1 period)

OLS estimation						
test	*source of critical values*	R=50 P=100	R=50 P=150	R=100 P=50	R=100 P=100	R=100 P=150
bias, uncond.	Normal	0.532	0.551	0.682	0.779	0.806
bias, condit.-RF	Normal	0.811	0.844	0.929	0.964	0.978
efficiency, uncond.	Normal	0.497	0.604	0.270	0.430	0.498
efficiency, condit.-RF	Normal	0.649	0.732	0.450	0.546	0.622
bias, uncond.	bootstrap	0.476	0.507	0.645	0.741	0.784
bias, condit.-RF	bootstrap	0.748	0.788	0.876	0.953	0.969
efficiency, uncond.	bootstrap	0.279	0.399	0.156	0.277	0.344
efficiency, condit.-RF	bootstrap	0.329	0.444	0.153	0.286	0.389

Notes:
1. The data generating process is a bivariate VAR(1), with coefficient breaks, with coefficients given in Table 1 and error variance matrix given in section 4.1. In each experiment, the coefficient break occurs in period R+1.
2. For each artificial data set, forecasts of $y_{1,t+1}$ are formed recursively using OLS estimates of a bivariate VAR(1). We consider both unconditional forecasts and conditional forecasts, obtained under the DLS approach. The conditional forecasts of $y_{1,t+1}$ are based on a condition of $\hat{y}^c_{2,t+1} = y_{2,t}$.
3. These forecasts are then used to form bias and efficient tests, detailed in sections 3 and 4.3. For unconditional forecasts, the test statistics use OLS-estimated variances. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of 6.
4. R and $\tilde{P}$ refer to the number of in–sample observations and forecasts, respectively.
5. In each Monte Carlo replication, the simulated test statistics are compared against standard normal critical values and critical values obtained with a wild bootstrap of the VAR, as described in section 3.3.
6. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 9. Tests of bias in point forecasts from 22-variable BVAR-NW

	Unconditional forecasts			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.238	-0.281	-0.329	-0.268
t-statistic	-1.056	-1.141	-1.320	-0.535
bootstrap CVs	(-2.400, 1.303)	(-2.568, 1.359)	(-2.692, 1.520)	(-2.402, 1.861)
Unemployment rate				
coefficient estimate	0.008	0.026	0.083	0.165
t-statistic	0.495	0.730	0.843	0.637
bootstrap CVs	(-1.851, 2.418)	(-1.911, 2.534)	(-2.033, 2.888)	(-2.568, 2.745)
PCE inflation				
coefficient estimate	-0.239	-0.321	-0.347	-0.618
t-statistic	-1.818	-1.860	-2.392	-2.794
bootstrap CVs	(-2.271, 1.529)	(-2.510, 1.581)	(-2.715, 1.675)	(-3.467, 1.917)
	Conditional forecasts, reduced form approach			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.232	-0.223	-0.236	-0.189
t-statistic	-0.978	-0.695	-0.787	-0.503
bootstrap CVs	(-2.212, 1.170)	(-2.271, 1.093)	(-2.578, 1.376)	(-2.653, 2.067)
Unemployment rate				
coefficient estimate	0.012	0.026	0.056	0.067
t-statistic	0.458	0.445	0.401	0.301
bootstrap CVs	(-1.480, 2.137)	(-1.494, 2.243)	(-1.555, 2.384)	(-2.208, 2.784)
PCE inflation				
coefficient estimate	-0.235	-0.310	-0.324	-0.502
t-statistic	-1.942	-2.423	-3.032	-2.106
bootstrap CVs	(-2.329, 1.463)	(-2.498, 1.457)	(-2.770, 1.691)	(-3.684, 1.854)
	Conditional forecasts, policy shock approach			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.238	-0.267	-0.300	-0.327
t-statistic	-1.317	-1.211	-1.350	-0.806
bootstrap CVs	(-2.438, 1.334)	(-2.536, 1.404)	(-2.972, 1.694)	(-2.636, 2.352)
Unemployment rate				
coefficient estimate	0.008	0.024	0.073	0.161
t-statistic	0.508	0.700	0.879	0.958
bootstrap CVs	(-1.879, 2.373)	(-2.000, 2.687)	(-2.283, 3.069)	(-3.285, 3.233)
PCE inflation				
coefficient estimate	-0.239	-0.321	-0.343	-0.597
t-statistic	-2.058	-2.483	-2.869	-2.459
bootstrap CVs	(-2.263, 1.469)	(-2.492, 1.510)	(-2.897, 1.730)	(-3.553, 1.947)

Notes:

1. As described in section 5.2, point forecasts of real GDP growth, the unemployment rate, and PCE inflation (all defined at annualized rates) are obtained from recursive estimates of the BVAR22-NW and BVAR5-SSP model specifications. The forecasts include unconditional, conditional obtained under the DLS approach, and conditional obtained under the policy shock approach. At each forecast horizon t, the conditions imposed are that, over an eight quarter forecast horizon from $t+1$ through $t+8$, the federal funds rate remain at its value in period t. The forecasts of growth and inflation at forecast horizons of four and eight quarters refer to four-quarter averages of growth and inflation (obtained by averaging forecasts for each quarter).
2. The table reports the results of bias tests over the 1991-2007 sample, obtained by running a regression of each forecast error on a constant, as detailed in section 3. Each panel provides the estimate of the constant (the bias), the associated t-statistic, and 10% bootstrap critical values. For unconditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $2(\tau-1)$. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $6+2(\tau-1)$.

Table 10. Tests of efficiency of point forecasts from 22-variable BVAR-NW

	Unconditional forecasts			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.415	-0.699	-0.599	-1.342
t-statistic	-2.534	-3.719	-3.299	-5.647
bootstrap CVs	(-3.606, 0.445)	(-3.984, 0.585)	(-5.033, 0.644)	(-6.226, 0.195)
Unemployment rate				
coefficient estimate	0.015	0.022	0.017	-0.471
t-statistic	0.792	0.556	0.180	-1.724
bootstrap CVs	(-3.287, 0.899)	(-3.981, 0.738)	(-5.667, 0.395)	(-8.598, -0.028)
PCE inflation				
coefficient estimate	-0.706	-0.807	-0.782	-0.949
t-statistic	-6.076	-6.634	-7.220	-7.673
bootstrap CVs	(-2.754, 1.180)	(-3.298, 0.999)	(-4.299, 1.118)	(-8.267, 0.604)
	Conditional forecasts, reduced form approach			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.477	-0.693	-0.547	-1.543
t-statistic	-3.411	-3.453	-2.882	-4.710
bootstrap CVs	(-4.218, 0.393)	(-4.375, 0.596)	(-5.971, 0.505)	(-7.127, 0.926)
Unemployment rate				
coefficient estimate	0.017	0.019	-0.019	-0.649
t-statistic	0.902	0.533	-0.207	-4.170
bootstrap CVs	(-3.502, 1.067)	(-4.632, 0.675)	(-6.431, 0.366)	(-10.409, 0.633)
PCE inflation				
coefficient estimate	-0.758	-0.882	-0.870	-0.976
t-statistic	-4.730	-4.275	-6.641	-7.932
bootstrap CVs	(-3.024, 1.325)	(-3.850, 1.079)	(-5.432, 1.276)	(-10.266, 0.664)
	Conditional forecasts, policy shock approach			
	h=1	h=2	h=4	h=8
GDP growth				
coefficient estimate	-0.415	-0.699	-0.548	-1.520
t-statistic	-2.714	-3.210	-3.032	-3.991
bootstrap CVs	(-3.947, 0.489)	(-4.465, 0.801)	(-5.718, 0.868)	(-7.040, 0.834)
Unemployment rate				
coefficient estimate	0.015	0.023	-0.068	-0.681
t-statistic	0.798	0.627	-1.012	-5.719
bootstrap CVs	(-3.602, 0.982)	(-4.562, 0.919)	(-6.722, 0.438)	(-10.902, 0.204)
PCE inflation				
coefficient estimate	-0.706	-0.863	-0.819	-1.021
t-statistic	-5.891	-4.148	-5.368	-8.035
bootstrap CVs	(-2.934, 1.269)	(-3.558, 1.119)	(-5.045, 1.172)	(-10.071, 0.878)

Notes:
1. As described in section 5.2, point forecasts of real GDP growth, the unemployment rate, and PCE inflation (all defined at annualized rates) are obtained from recursive estimates of the BVAR22-NW and BVAR5-SSP model specifications. The forecasts include unconditional, conditional obtained under the DLS approach, and conditional obtained under the policy shock approach. At each forecast horizon t, the conditions imposed are that, over an eight quarter forecast horizon from $t+1$ through $t+8$, the federal funds rate remain at its value in period t. The forecasts of growth and inflation at forecast horizons of four and eight quarters refer to four-quarter averages of growth and inflation (obtained by averaging forecasts for each quarter).
2. The table reports the results of efficiency tests over the 1991-2007 sample, obtained by running a regression of each forecast error on the forecast and other terms as needed, as detailed in section 3. Each panel provides the estimate of the coefficient on the forecast, the associated t-statistic, and 10% bootstrap critical values. For unconditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $2(\tau-1)$. For conditional forecasts, the test statistics use Newey-West variances, computed with a lag length of $6+2(\tau-1)$.

Table 11A. RMSEs of forecasts from 22-variable BVAR-NW

	h=1	h=2	h=4	h=8
GDP growth				
unconditional	1.871	2.019	1.282	2.337
conditional, condit.-RF	1.922	2.075	1.286	2.171
conditional, policy shock	1.871	2.008	1.224	2.215
Unemployment rate				
unconditional	0.140	0.225	0.442	0.836
conditional, condit.-RF	0.154	0.262	0.518	0.801
conditional, policy shock	0.140	0.225	0.418	0.687
PCE inflation				
unconditional	1.110	1.230	0.975	1.425
conditional, condit.-RF	1.106	1.197	0.941	1.329
conditional, policy shock	1.110	1.226	0.976	1.482

Notes:

Table 11B. RMSEs of forecasts from 5-variable BVAR-SSP

	h=1	h=2	h=4	h=8
GDP growth				
unconditional	2.371	2.373	1.793	2.173
conditional, condit.-RF	2.251	2.187	1.617	2.331
conditional, policy shock	2.375	2.344	1.860	2.483
Unemployment rate				
unconditional	0.174	0.299	0.559	0.793
conditional, condit.-RF	0.168	0.280	0.501	0.720
conditional, policy shock	0.174	0.301	0.568	0.935
PCE inflation				
unconditional	0.965	1.007	0.745	1.146
conditional, condit.-RF	0.966	1.014	0.762	1.179
conditional, policy shock	0.965	0.999	0.741	1.157

Notes:
1. As described in section 5.2, point forecasts of real GDP growth, the unemployment rate, and PCE inflation (all defined at annualized rates) are obtained from recursive estimates of the BVAR22-NW and BVAR5-SSP model specifications. The forecasts include unconditional, conditional obtained under the DLS approach, and conditional obtained under the policy shock approach. At each forecast horizon t, the conditions imposed are that, over an eight quarter forecast horizon from $t+1$ through $t+8$, the federal funds rate remain at its value in period t. The forecasts of growth and inflation at forecast horizons of four and eight quarters refer to four-quarter averages of growth and inflation (obtained by averaging forecasts for each quarter).
2. The table reports the RMSEs of the point forecasts over the 1991-2007 sample.