November 2008

Path Forecast Evaluation*

Abstract

A path forecast refers to the sequence of forecasts 1 to H periods into the future. A summary of the range of possible paths the predicted variable may follow for a given confidence level requires construction of simultaneous confidence regions that adjust for any covariance between the elements of the path forecast. This paper shows how to construct such regions with the joint predictive density and Scheffé's (1953) S-method. In addition, the joint predictive density can be used to construct simple statistics to evaluate the local internal consistency of a forecasting exercise of a system of variables. Monte Carlo simulations demonstrate that these simultaneous confidence regions provide approximately correct coverage in situations where traditional error bands, based on the collection of marginal predictive densities for each horizon, are vastly off mark. The paper showcases these methods with an application to the most recent monetary episode of interest rate hikes in the U.S. macroeconomy.

JEL Classification Codes: C32, C52, C53

Keywords: path forecast, simultaneous confidence region, Scheffé S-method, Mahalanobis distance.

Òscar Jordà
Department of Economics
University of California, Davis
One Shields Ave.
Davis, CA 95616-8578
e-mail: ojorda@ucdavis.edu

Massimiliano Marcellino
Department of Economics
European University Institute
Via della Piazzuola 43
50133 Firenze, Italy
e-mail: massimiliano.marcellino@eui.eu

*We are grateful to two referees, Filippo Altissimo, Frank Diebold, Peter Hansen, Hashem Pesaran, Shinichi Sakata, Mark Watson, Jonathan Wright and seminar participants in the 2007 Oxford Workshop on Forecasting, the 2007 Econometrics Workshop at the Federal Reserve Bank of St. Louis, the 2007 ECB Conference on Forecasting, the Bank of Korea, the Federal Reserve Bank of San Francisco, and the University of California, Davis, for useful comments and suggestions. Jordà acknowledges the hospitality of the Federal Reserve Bank of San Francisco during the preparation of this manuscript.

> [...] a central bank seeking to maximize its probability of achieving its goals is driven, I believe, to a risk-management approach to policy. By this I mean that policymakers need to consider not only the most likely future path for the economy but also the distribution of possible outcomes about that path.
>
> *Alan Greenspan, 2003.*

1 Introduction

Understanding the uncertainty associated with a prediction is as important as the prediction itself. When a prediction is made about a collection of future events – which we denominate a *path forecast* – its uncertainty is encapsulated by the path forecast's joint predictive density. Therefore, information about the range of possible paths the predicted variable may follow (for a given probability level) is contained in a simultaneous confidence region. Although such a confidence region represents a correct probability statement about the set of possible paths, it is a multi-dimensional ellipsoid. This makes communication of the path forecast's uncertainty difficult since, for example, it cannot be graphically presented in two-dimensional space (except, of course, when the path is at most two periods in length).

Traditionally, the uncertainty associated with a forecasting exercise has been examined with the marginal predictive density of the forecasts at each individual horizon. This is the approach that has received the bulk of attention in the literature and can be found pre-coded into most commercial econometric packages. The basic message of this paper is that there are many questions of interest that require knowledge of the joint predictive density and appropriate statistics for which the collection of marginal predictive densities and traditional

statistics are inadequate.

The first contribution of our paper is to propose new methods to construct an approximate two-dimensional simultaneous confidence region that summarizes the uncertainty about the variability of the path forecast that is most interesting to end-users of the forecasting exercise. In Section 2 we show that this region focuses the space of alternative hypotheses onto a particular subspace of linear combinations of forecasts that best represents paths of interest rather than providing a probability bound (as the well known Bonferroni procedure does). This region is constructed with Scheffé's (1953, 1959) S-method of simultaneous inference. As a result, our simultaneous confidence bands (which we denominate Scheffé bands) provide approximate control of the family-wise error rate (or $FWER$, which loosely speaking is the probability that one or more of the hypotheses is rejected when the joint null is true[1]) while at the same time increasing the power of the joint null hypotheses implied by the Scheffé bands. This approach is similar to that proposed by Jordà (2008) for impulse response functions.

Another way to evaluate a system's path forecasts is to examine the local consistency of the predictive exercise and measure how the paths predicted for some variables vary in response to alternative assumptions on how the paths for the other variables evolve. For example, the Bank of England's inflation report (available from their website) reports two-year inflation and gross domestic product (GDP) forecasts conditional on a variety of assumptions about the path of interest rates. A second contribution of our paper is to provide a simple protocol with which to evaluate the coherence of the experiment with respect to the

[1] See, e.g., Hochberg and Tahmane, 1987; or Lehmann and Romano, 2005 for a formal definition.

historical experience, as well as the relative exogeneity of the path forecasts to the selected experiments. Coherence and exogeneity can be measured from the path forecasts' predictive density with the Mahalanobis (1936) distance.[2]

Section 3 of the paper derives the asymptotic distribution of the path forecasts for a class of models frequently used in practice, whereas the small sample properties of the methods we propose are investigated via Monte Carlo simulations in Section 4. Specifically, we simulate data from the VAR process discussed in Stock and Watson's (2001) review article and show that using different estimation methods, different forecasting horizons, and different metrics of performance, traditional marginal bands provide very poor and unreliable coverage – a problem that is successfully addressed with the methods that we introduce. Section 5 displays our methods in action with a forecasting exercise of the most recent monetary episode of interest rate hikes experienced in the U.S., beginning June, 2003. Finally, directions for further research are outlined in Section 6, which summarizes the main results of the paper and draws some conclusions.

2 Measuring Path Forecast Uncertainty

This section considers how to measure the uncertainty surrounding a collection of forecasts 1 to H periods into the future (an H-dimensional path forecast) for a variable in a K-dimensional system $\{\mathbf{y}_t\}_{t=1}^{T}$.[3] The simultaneous confidence regions that we construct for this purpose are based on the assumption that the path forecasts are distributed multivari-

[2] The Mahanalobis (1936) distance between to n-dimensional vectors, say u and v, with common $n \times n$ covariance matrix Ω is given by $(u-v)'\Omega^{-1}(u-v)$.

[3] Although the results in this section could have been discussed in the context of univariate path forecasts, we present them for systems instead to incorporate into the forecast information from exogenous variables in a more natural way. In addition, systems provide a natural sageway for the discussion in Section 2.3.

ate Gaussian (at least in large samples); Scheffé's (1953, 1959) S-method of simultaneous inference; and on a step-down testing recursion. In practice, although the multivariate distribution of the data is unknown, for H not too large relative to the estimation sample size R (a mnemonic for "regression," where $R < T$) and conditionally on the information on which the forecasts are produced, the assumption of Gaussianity can be justified under rather general conditions. As an example, in Section 3 we derive asymptotic Gaussian approximations for homogeneous, covariance-stationary processes whose forecasts are generated with either vector autoregressions (VARs) or direct forecasts (see, e.g. Marcellino, Stock and Watson, 2006).

Although the derivations in Section 3 cover many practical situations in applied work, there are many others where one cannot justify common large sample approximations. A simple example of this failure is readily seen for forecasts generated with fixed-size rolling windows (rather than windows whose size is allowed to grow with the sample), or when predicting autoregressive-moving average (ARMA) models with residuals that exhibit generalized autoregressive conditional heteroskedasticity (GARCH). In the latter case, Baillie and Bollerslev (1992) show that the predictive density has to be approximated with, for example, a Cornish-Fisher expansion.

These examples would seem to limit the scope of our results somewhat. However, Ringland (1983) shows that relative to Bonferroni and to modulus multiple inference methods (see, e.g., Hochberg and Tahmane, 1987), the Scheffé (1953, 1959) S-method that forms the basis of our procedures, is robust and stable under general forms of non-normality and for a variety of M-estimators.

2.1 Null Hypotheses and Confidence Regions

Let $\mathbf{y}_\tau$ be a K-dimensional random vector and denote $\widehat{\mathbf{y}}_\tau(h) = \widehat{E}(\mathbf{y}_{\tau+h}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...)$ where the sample size used to estimate the parameters required to generate $\widehat{\mathbf{y}}_\tau(h)$ is R and $\{\tau : R \leq \tau \leq T - H\}$. We do not make primitive assumptions on the stochastic process $\{\mathbf{y}_t\}$ because the focus of this section is not about how to derive Gaussian large sample approximations but rather how to derive appropriate methods of simultaneous inference. Instead, denote $\widehat{Y}_\tau(H)$ and $Y_{\tau,H}$ the predicted and observed paths

$$\widehat{Y}_\tau(H) = \begin{bmatrix} \widehat{\mathbf{y}}_\tau(1) \\ \vdots \\ \widehat{\mathbf{y}}_\tau(H) \end{bmatrix}; \underset{KH\times 1}{Y_{\tau,H}} = \begin{bmatrix} \mathbf{y}_{\tau+1} \\ \vdots \\ \mathbf{y}_{\tau+H} \end{bmatrix}.$$

Then, the derivations that follow are based on the following convenient assumption.

Condition 1

$$\sqrt{R}\left(\widehat{Y}_\tau(H) - Y_{\tau,H}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Xi}_H\right) \text{ as } R \to \infty$$

and an estimate $\widehat{\Xi}_H$ *is available such that (with a slight abuse of notation):* $\widehat{\Xi}_H \xrightarrow{p} \Xi_H$

Specific analytic expressions for Ξ_H are obtained trivially when one assumes directly that $\{\mathbf{y}_t\}$ is Gaussian. In Section 3, we relax this assumption for forecasts generated from a VAR or with direct forecasts (e.g. Jordà, 2005; Marcellino, Stock and Watson, 2006) and provide appropriate analytic expressions. Other relevant references for specific results on Ξ_H include, among others, Clements and Hendry (1993) and Lütkepohl (2005). The assumption of normality is convenient but not crucial so that distributions belonging to the exponential family, or elliptic distributions (see Mitchell and Krzanowski, 1985) would be easily accommodated in our discussion.

In what follows, we focus on the path forecasts for each of the variables in $\mathbf{y}_t$, one at a time. Define the selector matrix $S_k \equiv (I_H \otimes e_k)$ where e_k is the k^{th} row of I_K. Then from condition 1, the distribution of the path forecast for the k^{th} variable in $\mathbf{y}_t$ is

$$\sqrt{R}\left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Xi}_{k,H}\right)$$

where $\widehat{Y}_{k,\tau}(H) = S_k\widehat{Y}_\tau(H); Y_{k,\tau,H} = S_k Y_{\tau,H}$; and $\mathbf{\Xi}_{k,H} = S_k\mathbf{\Xi}_H S_k'$.

Next, we are interested in constructing a confidence region based on inverting the statistic for the null hypothesis

$$\begin{aligned} H_0 &: E\left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) = 0 \text{ versus} \\ H_1 &: E\left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \neq 0 \text{ for } k = 1, ..., K. \end{aligned} \tag{1}$$

This null hypothesis can be expressed as the intersection of the family of hypotheses

$$\begin{aligned} H_{0h} &: E\left(\widehat{y}_{k,\tau}(h) - y_{k,\tau+h}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) = 0 \text{ versus} \\ H_{1h} &: E\left(\widehat{y}_{k,\tau}(h) - y_{k,\tau+h}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \neq 0 \text{ for } k = 1, ..., K \text{ and } h = 1, ..., H. \end{aligned} \tag{2}$$

Clearly, it is the case that $H_0 = \cap_{h=1}^H H_{0h}$ and $H_1 = \cup_{h=1}^H H_{1h}$. The rejection regions of the tests for the individual hypotheses are given by the well-known studentized ratios

$$|T_h| = \frac{|\widehat{y}_{k,\tau}(h) - y_{k,\tau+h}|}{\sqrt{\frac{\Xi_k(h,h)}{R}}} > \xi_h; h = 1, ..., H; k = 1, ..., K$$

where $\Xi_k(h,h)$ refers to the (h,h) diagonal entry of the covariance matrix $\Xi_{k,H}$ and ξ_h can be determined for a given size α as

$$\Pr\left(|T_h| > \xi_h\right) = \alpha.$$

Customarily in forecasting applications, confidence regions for forecasts at different horizons have been constructed by setting $\xi_h = \xi \ \forall h$ and then selecting ξ as the critical value from

a standard normal random variable (or t-distribution when small sample results are invoked instead) with probability level $\alpha/2$, that is $\xi = z_{\alpha/2}$. This tradition, of course, ignores the simultaneous nature of the problem in expressions (1)-(2) as well as any correlation among the T_h. As a consequence, using this procedure, the probability that one or more of the null hypotheses (2) will be rejected increases toward one with H.

If instead one is interested in controlling the probability of making any error given a family of inferences (i.e., control of the family-wise error rate, $FWER$), then the union-intersection method proposed by Roy (1953) can be used to ensure $FWER \leq \alpha$ by choosing a test of the null in (1) that rejects if

$$\max_{1 \leq h \leq H} |T_h| > \xi^*. \tag{3}$$

Roy and Bose (1953) show that

$$\max_{1 \leq h \leq H} T_h^2 = R \left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H} \right)' \Xi_{k,H}^{-1} \left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H} \right) \tag{4}$$

which demonstrates the correspondence between testing the null (1) and the intersection of the family of nulls in (2). Scheffé (1953, 1959) generalizes this result and shows that any test of non-null linear combinations of $\widehat{Y}_{k,\tau}(H)$ has the same distribution – in other words, when the joint null (1) is rejected, it gives a basis for data-snooping which of the elementary hypotheses most likely generated the rejection without requiring further adjustments to the distribution of these statistics. Savin (1980, 1984) provides formal derivations of (4) both in small samples (under the common t-distribution/F-distribution results) and in large samples (with the normal/chi-square approximations) in econometrics.

Without loss of generality and in order to construct confidence bands for the path forecasts $\widehat{Y}_{k,\tau}(H)$, we find it convenient to proceed with the orthogonalized studentized ratios

$$T_h^* = \frac{\widehat{z}_{k,\tau}(h)}{\sqrt{R}} \xrightarrow{d} N(0,1)$$

where $\widehat{z}_{k,\tau}(h)$ is the h^{th} entry of the vector $\widehat{Z}_{k,\tau}(H) = P^{-1}(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H})$ and $\Xi_{k,H} = PP'$ with P the lower triangular Cholesky decomposition of $\Xi_{k,H}$. In practical terms, this is simply a projection of the forecast at time $\tau + h$ onto the vector of forecasts from $\tau + 1$ to $\tau + h - 1$. Under condition 1 and result (4), the null hypothesis (1) can be tested with the familiar Wald statistic

$$W_H = R\left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H}\right)' \Xi_{k,H}^{-1} \left(\widehat{Y}_{k,\tau}(H) - Y_{k,\tau,H}\right) \xrightarrow{d} \chi_H^2. \tag{5}$$

The proposed orthogonalization transforms this Wald statistic into

$$W_H = \sum_{h=1}^{H} T_h^{*^2} \xrightarrow{d} \chi_H^2 \tag{6}$$

where now the $T_h^{*^2}$ have the interpretation of being uncorrelated (and standarized) t-statistics of the null that the average deviation of $\widehat{y}_{k,\tau}(h)$ from $y_{k,\tau+h}$ is zero, conditional on the path forecast over periods 1 to $(h-1)$, i.e., $\widehat{Y}_{k,\tau}(h-1)$.

A simultaneous confidence interval for $\widehat{Z}_{k,\tau}(H)$ (and hence $\widehat{Y}_{k,\tau}(H)$) can therefore be easily constructed by inverting the Wald statistic (6) so that

$$\Pr\left[\sum_{h-1}^{H} T_h^{*^2} \leq c_\alpha^2(H)\right] = 1 - \alpha \tag{7}$$

for any probability level α of interest and with $c_\alpha^2(H)$ the critical value of a random variable distributed as χ_H^2 (or, if invoking finite sample results, by choosing $Hc_\alpha^*(H,R)$ where $c_\alpha^*(H,R)$ is the critical value of a random variable distributed $F_{H,R}$). This confidence region is readily

seen to be an H-dimensional sphere, which for $H = 2$ results in a circumference that is plotted in Figure 1 for $\alpha = 0.05$.

The simultaneous confidence region (7) displayed in Figure 1 makes clear that the induced test for the hypotheses (2) is given by expressions (3) and (4) and specifically for the orthogonalized studentized ratios

$$\max_{1 \leq h \leq H} |T_h^*| > \xi^*$$

with

$$\Pr(\max_{1 \leq h \leq H} |T_h^*| > \xi^*) = \alpha \tag{8}$$

so that $\xi^* = c_\alpha^2(H)$ (see Savin, 1984). Figure 1 clarifies this result: the most T_1^* can stray away from the origin to the 95% confidence circumference is when $T_2^* = 0$, and vice versa.

2.2 Family-Wise versus Simultaneous Error Rate Control

Choosing $\xi^* = c_\alpha^2(H)$ ensures that $FWER \leq \alpha$ but this is not a very useful metric for paths. The reason is that control of $FWER$ guards against such extremes as when all forecasts in a path exactly match the future realizations of the predicted variable except for one specific horizon, onto which the entire probability mass of the forecast path is concentrated. For some financial applications where one is interested in insuring against extreme risks, control of $FWER$ may make sense. However, in most other applications of path forecasting (specially in macroeconomics), power can be improved by appropriately restricting the range of possible alternatives and for this reason, we introduce two modifications to the previous procedures.

The first modification consists in limiting the space of alternative hypotheses to those that spread the uncertainty in the orthogonalized path forecast evenly over all horizons. We

achieve this by maximizing the joint probability surface of the orthogonalized system implied by the T_h^* for $h = 1, ..., H$ (rather than maximizing the Tschebyshev distance as is done in expression (8)) and hence by directly controlling the Mahalanobis (1936) distance (or *MD*) between $\widehat{Y}_{k,\tau}(H)$ and the range of possible paths. *MD* corresponds to the distance implied by the Wald statistic[4] in expression (5). The second modification comes from the realization that the bounds implied in (8) depend on the choice of H: for example, the size of the one-period ahead error bands will vary depending on whether the path forecast is two, three, four, ..., H periods in length. This is somewhat unappealing but can be easily corrected as we will show. Incorporating these two modifications inevitably deteriorates control of $FWER$ but as the Monte Carlo experiments of Section 4 show, the trade-off turns out to be very modest and the new procedures control the simultaneous Mahalanobis (or *Wald*) distance very well.

We begin by discussing the first of our modifications. The desire to spread uncertainty evenly across the path can be recast by modifying expression (8) where interest is now in finding δ^* such that

$$\Pr\left(\max\left|\sum_{h=1}^{H}\frac{1}{H}T_h^*\right| > \delta^*\right) = \alpha. \tag{9}$$

The objective of maximizing the average allocation of the path's uncertainty across periods has the effect of maximizing the probability region covered by the bands in terms of *MD*. The example in Figure 1 can be useful in clarifying this concept. *MD* compares the square distances between two vectors by weighting the vectors with the inverse of their covariance

[4] Although Prasanta Mahalanobis published his paper in (1936) it was Abraham Wald's (1943) paper that provided the statistical foundations of what today is known as the Wald statistic.

matrix, the same way that the Wald statistic would for a typical joint hypothesis test as is done in expression (5). In Figure 1, notice that the T_h^* $h = 1, 2$ have been orthogonalized so that the weighting matrix is the identity and hence, the objective is to find T_1^* and T_2^* such that

$$\Pr\left(\max\left|\frac{1}{2}T_1^* + \frac{1}{2}T_2^*\right| \leq \delta^*\right) = 1 - \alpha.$$

This probability region is maximized when $T_1^* = T_2^*$ so that the bands for each test can be easily calculated since, for example from (5)

$$\Pr\left(2T_1^{*^2} \leq c_\alpha^2(2)\right) = \Pr\left(|T_1^*| \leq \sqrt{\frac{c_\alpha^2(2)}{2}}\right) = 1 - \alpha$$

and similarly for T_2^* from where it is clear that $\delta^* = \sqrt{c_\alpha^2(2)/2}$.

Scheffé's (1953, 1959) S-method gives simultaneous confidence regions for any non-null linear combination $\sum_{h=1}^{H} u_h T_h^*$ where $\mathbf{u} = (u_1, ..., u_h)$, is such that $\mathbf{u} \in U$, the set of non-null linear combinations of $(T_1^*, ..., T_h^*)$. In order to calculate δ^* in expression (9), we use Bowden's (1970) lemma, which states that

$$\max\left\{\frac{\left|\sum_{h=1}^{H} u_h T_h^*\right|}{\sqrt{\sum_{h=1}^{H} u_h^2}} : |u_h| < \infty\right\} = \sqrt{\sum_{h=1}^{H} T_h^{*^2}}.$$

Hence, from the confidence region derived in (7) and out of the Wald statistic in (5), recall that we have that

$$\Pr\left(\sum_{h=1}^{H} T_h^{*^2} \leq c_\alpha^2(H)\right) = 1 - \alpha$$

and by Bowden's (1970) lemma we can rewrite this expression as

$$\Pr\left(\max\left\{\frac{\left|\sum_{h=1}^{H} u_h T_h^*\right|}{\sqrt{\sum_{h=1}^{H} u_h^2}}\right\} \leq \sqrt{c_\alpha^2(H)}\right) = 1 - \alpha$$

which for $u_h = 1/H$ becomes

$$\Pr\left(\max\left\{\left|\sum_{h=1}^{H}\frac{1}{H}T_h^*\right|\right\} \leq \sqrt{c_\alpha^2(H)}\sqrt{\sum_{h=1}^{H}\frac{1}{H^2}}\right) = 1-\alpha$$

so that clearly

$$\delta^* = \sqrt{\frac{c_\alpha^2(H)}{H}}$$

and therefore a simultaneous (1-α)% confidence region for the original path forecast would be

$$\widehat{Y}_{k,\tau}(H) \pm P\left(\sqrt{\frac{c_\alpha^2(H)}{H}}\mathbf{i}_H\right) \tag{10}$$

where $\mathbf{i}_H$ is an $H \times 1$ vector of ones.

Error bands derived from expression (10) are simple to construct but, as discussed earlier, have the somewhat undesirable feature that they depend on H. For example, the width of the one-period ahead error bands declines as a function of H because as we increase H, we are considering the joint variation of a path with more and more elements and this limits how much any one of these individual elements can fluctuate. Therefore, the second modification that we propose is to refine our simultaneous confidence region with a step-down recursive procedure (for examples of step-down procedures see Lehmann and Romano, 2005. Perhaps the best known of such procedures is Holm, 1979).

Notice that for any H, $\sqrt{c_\alpha^2(h)/h} \geq \sqrt{c_\alpha^2(H)/H}$ for $h \leq H$ and hence

$$\left\{Y_{k,\tau,H} : Y_{k,\tau,H} \in \widehat{Y}_{k,\tau}(H) \pm P\sqrt{\frac{c_\alpha^2(H)}{H}}\mathbf{i}_H\right\} \subset \left\{Y_{k,\tau,H} : Y_{k,\tau,H} \in \widehat{Y}_{k,\tau}(H) \pm P\left[\sqrt{\frac{c_\alpha^2(h)}{h}}\right]_{h=1}^{H}\right\}$$

where $\left[\sqrt{\frac{c_\alpha^2(h)}{h}}\right]_{h=1}^{H}$ refers to the $H \times 1$ vector whose h^{th} entry is $\sqrt{c_\alpha^2(h)/h}$ and hence the region defined by

$$\widehat{Y}_{k,\tau}(H) \pm P\left[\sqrt{\frac{c_\alpha^2(h)}{h}}\right]_{h=1}^{H} \tag{11}$$

includes all paths in (10). However in (11) the bands for forecasts at horizon h are no longer functions of the overall length of the forecast vector the researcher chooses to plot thereafter.

We end this section with two final notes. First, notice that by projecting each forecast on the antecedent forecast path (as is done to orthogonalize the T_h into T_h^* in expression (6)), allows one to consider the forecast uncertainty associated with each individual horizon irrespective of the path followed to reach that particular period. Hence, it is easy to construct the region that corresponds to the individual tests of the conditional nulls

$$\begin{aligned} H_{0h}^c &: E\left(\widehat{y}_{k,\tau}(h) - y_{k,\tau+h} | \widehat{Y}_{k,\tau}(h-1), \mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) = 0 \text{ versus} \\ H_{1h}^c &: E\left(\widehat{y}_{k,\tau}(h) - y_{k,\tau+h} | \widehat{Y}_{k,\tau}(h-1), \mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \neq 0 \text{ for } k = 1, ..., K \text{ and } h = 1, ..., H. \end{aligned} \tag{12}$$

as

$$\widehat{Y}_{k,\tau}(H) \pm z_{\alpha/2}\text{diag}(P)$$

where the operator diag(P) takes the diagonal elements of the matrix P and stacks them into a $H \times 1$ vector. These conditional error bands summarize the uncertainty associated to each point forecast independently of the uncertainty associated with previous forecasts in the path. Second, in order to evaluate the statistical properties of our testing procedures, Section 4 below provides Monte Carlo experiments where coverage is calculated with respect to control of $FWER$ as is convention, but where control of the error rate of the joint test implied by the Mahalanobis (1936) distance in (9) is clearly more appropriate for paths. For this reason, our Monte Carlo experiments report both metrics.

2.3 Conditional Path Forecast Evaluation with the Mahalanobis Distance

Scheffé confidence bands, whether reported for a given $100(1-\alpha)\%$ confidence level or reported in the form of a fan chart for a collection of different confidence levels, are a natural way for the professional forecaster to communicate the accuracy of the forecasting exercise. However, when the exercise involves more than one predicted variable, it is often of interest for the end-user to have a means to evaluate the local internal consistency of forecasts across variables. For example, the Bank of England's quarterly Inflation Report (available from their web-site) provides GDP and inflation, two-year ahead projections based on "market interest rate expectations" and projections based on "constant nominal interest rate" paths. Alternatively, it is not difficult to envision a policy maker's interest in examining inflation forecasts based on an array of different assumptions on the future path of crude oil prices, for example. Obviously such checks are not meant to uncover the nature of structural relations between variables, nor provide guidance about the effects of specific policy interventions, both of which, from a statistical point of view, fall into the broad theme of the treatment evaluation literature (see, e.g. Cameron and Trivedi, 2005 for numerous references) and are not discussed here.[5]

Rather, the objective is to investigate the properties of the forecast exercise in a local neighborhood. Accordingly, for a given K-dimensional vector of path forecasts, it will be of interest: (1) to derive how forecasts for a k_0-dimensional subset of variables vary if the path forecasts of the remaining k_1 variables in the system (i.e. $K = k_0 + k_1; 1 \leq k_1 < K$) are set to follow paths different from those originally predicted; (2) to evaluate whether the k_1

[5] An example of papers with this bend are Leeper and Zha (2003) and Waggoner and Zha (1999), which examine counterfactual experimentation in the context of forecasting models using Bayesian techniques.

alternative paths considered deviate substantially from the observed historical record; and (3) to examine how sensitive the k_0 variables are to variations in these alternative scenarios.

Mechanically speaking, an approximate answer to question (1) can be easily derived from the multivariate Gaussian large-sample approximation to the joint predictive density and the linear projection properties of the multivariate normal distribution. Specifically, define the selector matrices $S_0 = I_H \otimes E_0$; and $S_1 = I_H \otimes E_1$ where E_0 and E_1 are $k_0 \times K$ and $k_1 \times K$ matrices formed from the rows of I_K corresponding to the indices in k_0 and k_1 respectively. Let $\widetilde{Y}_\tau^1(H)$ denote the alternative paths considered for the k_1 variables and let $\widetilde{Y}_\tau^0(H)$ denote the paths of the k_0 variables given $\widetilde{Y}_\tau^1(H)$, that is

$$\widetilde{Y}_\tau^0(H) = S_0\widehat{Y}_\tau(H) + S_0\Xi_H S_1' \left(S_1\Xi_H S_1'\right)^{-1} \left(\widetilde{Y}_\tau^1(H) - S_1\widehat{Y}_\tau(H)\right)$$

with covariance matrix $\Xi_H^0 = S_0\Xi_H S_0' - S_0\Xi_H S_1' \left(S_1\Xi_H S_1'\right)^{-1} S_1\Xi_H S_0$. In practice, the approximate nature of the predictive density of $\widehat{Y}_\tau(H)$ indicates that the accuracy of these calculations depends on several factors such as the value of H relative to the estimation sample R, possible nonlinearities in the data, and the distance between $\widetilde{Y}_\tau^1(H)$ and $S_1\widehat{Y}_\tau(H)$, among the more important factors.

The last observation suggests that it is useful to properly evaluate the Mahalanobis distance between $\widetilde{Y}_\tau^1(H)$ and $S_1\widehat{Y}_\tau(H)$ and this can be easily accomplished with

$$W_1 = R(S_1\widehat{Y}_\tau(H) - \widetilde{Y}_\tau^1(H))' \left(S_1\Xi_H S_1'\right)^{-1} (S_1\widehat{Y}_\tau(H) - \widetilde{Y}_\tau^1(H))$$

This score will have an approximate chi-square distribution with $k_1 H$ degrees of freedom when condition 1 can be invoked. Thus, one minus the p-value of this score provides and easy to communicate distance metric in probability units between the predicted paths $S_1\widehat{Y}_\tau(H)$

and the alternative scenarios $\widetilde{Y}_\tau^1(H)$. The bigger this probability distance, the more the alternative scenarios strain the forecasting exercise toward regions in which the model has received little to no training by sample and the more one has to rely on the estimated econometric model being the true unknown DGP.

Similarly, it is of interest to evaluate which path forecasts from the k_0 variables are most sensitive to the alternative scenarios of the k_1 variables. This sensitivity can be evaluated with

$$W_0 = R\left(S_0\widehat{Y}_\tau(H) - \widetilde{Y}_\tau^0(H)\right)'\left(S_0\Xi_H S_0'\right)^{-1}\left(S_0\widehat{Y}_\tau(H) - \widetilde{Y}_\tau^0(H)\right)$$

When condition 1 can be invoked, this score will have an approximate chi-square distribution with k_0H degrees of freedom. Thus, p-values of this score below conventional significance values (say 0.05 for 95% confidence levels) indicate that the k_0 forecast paths are not exogenous to variations in the forecast paths of the k_1 variables and hence care should be taken that the W_1 score is kept sufficiently low. Thus, it seems wise that any forecasting report should routinely include both the W_0 and W_1 scores since they can be easily calculated in practice.

3 Asymptotic Distribution of the Forecast Path

This section characterizes the asymptotic distribution of the path forecast under the assumption that the data generating process (DGP) is possibly of infinite order while the forecasts are generated by finite-order VARs or finite-order direct forecasts. This DGP is sufficiently general to represent a large class of problems of practical interest, and VARs and direct forecasts are the two most commonly used forecasting strategies. Formal presentation of assumptions, corollaries and proofs are reserved for the appendix. Here we sketch the main

ideas.

Suppose the K-dimensional vector of weakly stationary variables $\{\mathbf{y}_t\}$ has a possibly infinite VAR representation given by

$$\mathbf{y}_t = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t$$

whose statistical properties are collected in assumptions 1 and 2 in the appendix. Given this DGP, one can either estimate a VAR(p), such as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{m} + \sum_{j=1}^{p} A_j \mathbf{y}_{t-j} + \mathbf{w}_t \\ \mathbf{w}_t &= \sum_{j=p+1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \end{aligned} \tag{13}$$

from which forecasts can be constructed with standard available formulas (see, e.g. Hamilton, 1994). Alternatively, forecasts could be constructed with a sequence of direct forecasts given by

$$\begin{aligned} \mathbf{y}_{t+h} &= \mathbf{m}_h + \sum_{j=0}^{p-1} A_j^h \mathbf{y}_{t-j} + \mathbf{v}_{t+h} \\ \mathbf{v}_{t+h} &= \sum_{j=p}^{\infty} A_j^h \mathbf{y}_{t-j} + \mathbf{u}_{t+h} + \sum_{j=1}^{h-1} \Phi_j \mathbf{u}_{t+h-j} \quad \text{for } h = 1, ..., H \end{aligned} \tag{14}$$

where $A_1^h = \Phi_h$ for $h \geq 1$; $A_j^h = \Phi_{h-1} A_j + A_{j+1}^{h-1}$ for $h \geq 1$; $A_{j+1}^0 = 0; \Phi_0 = I_K$; and $j \geq 1$. Let $\Gamma(j) \equiv E\left(\mathbf{y}_t \mathbf{y}_{t+j}'\right)$ with $\Gamma(-j) = \Gamma(j)'$ and define: $X_{t,p} = \left(\mathbf{1}, \mathbf{y}_{t-1}', ..., \mathbf{y}_{t-p}'\right)'$; $\underset{Kp+1 \times K}{\widehat{\Gamma}_{1-p,h}} = (R-p-h)^{-1} \sum_{t=p}^{R} X_{t,p} \mathbf{y}_{t+h}'$; and $\underset{K(p+1) \times K(p+1)}{\widehat{\Gamma}_p} = (R-p-h)^{-1} \sum_{t=p}^{R} X_{t,p} X_{t,p}'$. Then, the least-squares estimate of the VAR(p) in expression (13) is given by the formula

$$\underset{K \times Kp+1}{\widehat{A}(p)} = \left(\widehat{\mathbf{m}}, \widehat{A}_1, ..., \widehat{A}_p\right) = \widehat{\Gamma}_{1-p,0}' \widehat{\Gamma}_p^{-1}, \tag{15}$$

whereas the coefficients of the mean-squared error linear predictor of $\mathbf{y}_{t+h}$ based on $\mathbf{y}_t, ..., \mathbf{y}_{t-p+1}$ is given by the least-squares formula

$$\underset{K\times Kp+1}{\widehat{A}(p,h)} = \left(\widehat{\mathbf{m}}_h, \widehat{A}_1^h, ..., \widehat{A}_p^h\right) = \widehat{\Gamma}'_{1-p,h}\widehat{\Gamma}_p^{-1}; \quad h = 1, ..., H. \tag{16}$$

Then, corollary 1 in the appendix shows that the parameter estimates in expressions (15) and (16) are consistent and asymptotically Gaussian.

Next, denote with $\mathbf{y}_\tau(h)$ the forecast of the vector $\mathbf{y}_{\tau+h}$ assuming the coefficients of the infinite order process (21) were known. Notice that we have now changed the subscript from t to τ to reflect forecasts obtained in the interval $R \leq \tau \leq T - H$, that is

$$\mathbf{y}_\tau(h) = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_\tau(h-j)$$

where $\mathbf{y}_\tau(h-j) = \mathbf{y}_{\tau+h-j}$ for $h - j \leq 0$. Denote $\widehat{\mathbf{y}}_\tau(h)$ the forecast that relies on coefficients estimated from a sample of size R and based on a finite order VAR or direct forecasts, respectively

$$\begin{aligned}
\widehat{\mathbf{y}}_\tau(h) &= \widehat{\mathbf{m}} + \sum_{j=1}^{p} \widehat{A}_j \widehat{\mathbf{y}}_\tau(h-j) \\
\widehat{\mathbf{y}}_\tau(h) &= \widehat{\mathbf{m}}_h + \sum_{j=0}^{p-1} \widehat{A}_j^h \mathbf{y}_{\tau-j}
\end{aligned}$$

where $\widehat{\mathbf{y}}_\tau(h-j) = \mathbf{y}_{\tau+h-j}$ for $h - j \leq 0$. To economize in notation, we do not introduce a subscript that identifies how the forecast path was constructed as it should be obvious in the context of the derivations we provide. Then, define the forecast path for $h = 1, ..., H$ by stacking each of the quantities $\widehat{\mathbf{y}}_\tau(h)$, $\mathbf{y}_\tau(h)$, and $\mathbf{y}_{\tau+h}$ as follows

$$\underset{KH\times 1}{\widehat{Y}_{\tau}(H)} = \begin{bmatrix} \widehat{\mathbf{y}}_{\tau}(1) \\ \vdots \\ \widehat{\mathbf{y}}_{\tau}(H) \end{bmatrix}; \underset{KH\times 1}{Y_{\tau}(H)} = \begin{bmatrix} \mathbf{y}_{\tau}(1) \\ \vdots \\ \mathbf{y}_{\tau}(H) \end{bmatrix}; \underset{KH\times 1}{Y_{\tau,H}} = \begin{bmatrix} \mathbf{y}_{\tau+1} \\ \vdots \\ \mathbf{y}_{\tau+H} \end{bmatrix}.$$

Our interest is in finding the asymptotic distribution for $\widehat{Y}_{\tau}(H) - Y_{\tau,H} = \left[\widehat{Y}_{\tau}(H) - Y_{\tau}(H)\right] + \left[Y_{\tau}(H) - Y_{\tau,H}\right]$, conditional on information up to time τ, $R \leq \tau \leq T - H$.

It should be clear that $\left[Y_{\tau}(H) - Y_{\tau,H}\right]$ does not depend on the estimation method and hence its mean-squared error can be easily verified to be

$$\underset{KH\times KH}{\Omega_H} \equiv E\left[\left(Y_{\tau}(H) - Y_{\tau,H}\right)\left(Y_{\tau}(H) - Y_{\tau,H}\right)'\right] = \mathbf{\Phi}\left(I_H \otimes \Sigma_u\right)\mathbf{\Phi}'. \tag{17}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} I_K & \mathbf{0} & ... & \mathbf{0} \\ \Phi_1 & I_K & ... & \mathbf{0} \\ \vdots & \vdots & ... & \vdots \\ \Phi_{h-1} & \Phi_{h-2} & ... & I_K \end{bmatrix}.$$

Furthermore, since the parameter estimates are based on a sample of size R and hence $\mathbf{u}_t$ for $t \in \{p+h, ..., R\}$ while the term $Y_{\tau}(H) - Y_{\tau,H}$ only involves $\mathbf{u}_{\tau}$ for $\tau \in \{R+1, ..., T\}$, then it should be clear that to derive the asymptotic distribution of $\left[\widehat{Y}_{\tau}(H) - Y_{\tau}(H)\right]$, the asymptotic covariance of the forecast path will simply be the sum of the asymptotic covariance for this term and the mean-squared error in expression (17) but the covariance between these terms will be zero.

Corollary 1(a) and 1(b) in the appendix and the observation that $\widehat{Y}_{\tau}(H)$ is simply a function of estimated parameters and predetermined variables is all we need to conclude

that

$$\sqrt{\frac{R-p-H}{p}}vec\left(\widehat{Y}_{\tau}\left(H\right)-Y_{\tau}\left(H\right)|\mathbf{y}_{\tau},\mathbf{y}_{\tau-1},...\right)\xrightarrow{d}N\left(0,\Psi_{H}\right) \tag{18}$$

$$\Psi_{H} \equiv \frac{\partial vec\left(\widehat{Y}_{\tau}\left(H\right)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)}\Sigma_{A}\frac{\partial vec\left(\widehat{Y}_{\tau}\left(H\right)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)'}$$

where Σ_A is the covariance matrix for $vec\left(\widehat{\mathbf{A}}\right)$; with $\widehat{\mathbf{A}}=\widehat{A}\left(p\right)$ for estimates from a VAR(p); and for estimates from local projections

$$\widehat{\mathbf{A}}=\begin{bmatrix}\widehat{A}\left(p,1\right)\\ \vdots \\ \widehat{A}\left(p,H\right)\end{bmatrix}. \tag{19}$$

Therefore, corollaries 2 and 3 in the appendix contain the analytic formulas that show that

$$\sqrt{\frac{R-p-H}{p}}vec\left(\widehat{Y}_{\tau}\left(H\right)-Y_{\tau,H}\right)\xrightarrow{d}N\left(\mathbf{0};\Xi_{H}\right)$$

$$\Xi_{H} = \left\{\frac{p}{R-p-H}\Omega_{H}+\Psi_{H}\right\}$$

$$\Omega_{H} = \mathbf{\Phi}(I_{H}\otimes\Sigma_{u})\mathbf{\Phi}'$$

were the specific analytic expression of Ψ_H depends on whether a VAR(p) or direct forecasts are used. The appendix contains the specific formulae in each case.

4 Small Sample Monte Carlo Experiments

This section compares the probability coverage of traditional marginal error bands, bands constructed with the Bonferroni procedure, and Scheffé bands with a small-scale simulation study. In setting up the DGP for the simulations our objective was to choose a forecasting exercise that would be representative of situations researchers will likely encounter in practice.

In addition and to avoid the arbitrary nature of parameter choices and model specifications common to Monte Carlo experiments, we borrowed a well-known empirical specification directly from the literature, specifically Stock and Watson's (2001) well-cited review article on vector autoregressions (VARs).

The specification discussed therein examines a three-variable system (inflation, measured by the chain-weighted GDP price index; unemployment, measured by the civilian unemployment rate; and the average federal funds rate) that is observed quarterly over a sample beginning the first quarter of 1960 and that we extend to the first quarter of 2007 (188 observations). Their VAR is estimated with four lags.

The DGP for our experiments is therefore constructed from this VAR specification as follows. First, we estimate a VAR(4) on the sample of data just described except for the last 12 observations (3 years worth), which we save to do some out-of-sample exercises later on (reported in Figure 2). We collect the least-squares parameter estimates of the conditional means and the residual covariance matrix to generate the simulated samples of data of size $T = 100$, 400 (these are always initialized using the first four observations from the data for consistency across runs). We constructed 1,000 Monte Carlo replications of each sample size in this fashion.

At each replication the VAR's lag length is determined empirically (rather than chosen to be its true value of four) with the information criterion AIC_C – a correction to the traditional AIC, specially designed for VARs by Hurvich and Tsai (1993).[6] Next, each replication involves estimating a VAR and direct forecasts by least-squares and hence gen-

[6] Hurvich and Tsai (1993) show that AIC_c has better small sample properties than AIC, SIC and other common information criteria.

erating appropriate forecast error variances for forecast paths of varying length (specifically for $H = 1, 4, 8$, and 12 or one quarter and one, two and three years ahead) that include forecast error uncertainty as well as estimation error uncertainty as the previous section showed. Thus, each replication produces two sets of estimates (VAR and direct forecasts) with which we construct traditional marginal bands, Bonferroni bands and Scheffé bands; one and two standard deviations in width (the traditional choices in the literature), which roughly correspond with 68% and 95% probability coverage, respectively. These bands and forecasts are computed for each of the three variables (inflation, the unemployment rate and the federal funds rate) in the system and they are reported separately.

In order to assess the empirical coverage of these three sets of bands, we then generated 1,000 draws from the known model and multivariate distribution of the residuals in the DGP and hence constructed 1,000 paths conditional on the last four observations in the data (since the DGP is a VAR(4)). These conditioning observations are used to homogenize the analysis in all the Monte Carlo runs and thus facilitate comparability.

The empirical coverage of each set of bands is then evaluated with the two metrics discussed in Section 2. The *FWER* metric looks at the proportion of paths that fall completely within the bands, without regard of whether the joint variation of the elements respects the *Wald* metric or not. For example, a 12-period ahead forecast path in which, say, only one forecast out of the 12 fell outside the bands, would be considered "not covered."

The Mahalanobis/Wald (*Wald*) metric constructs the value of the Wald statistic associated with the bands and with each of the 1,000 predicted paths. Hence we compute the proportion of predicted paths with Wald scores lower than those for the bands. Using the

previous example, a 12-period forecast path that had one element outside the bands would be counted as "covered" as long as its Wald score was lower than that for the bands. Conversely, a path could have no individual element violate the Bonferroni bound per se, yet taking all elements together as a path, they could violate the Wald score.

The results of these experiments are reported in Tables 1 (for VAR-based forecasts) and 2 (for direct forecasts) for forecast horizons $H = 1, 4, 8$, and 12; for each of the three variables in the VAR (with mnemonics P for inflation, UN for the unemployment rate; and FF for the federal funds rate). In addition, Figure 2 displays what the three types of bands (marginal, Bonferroni and Scheffé) look like for an out-of-sample, two-year ahead path forecast from the VAR estimated with the actual data.

From Tables 1 and 2, for one-period ahead forecasts (where the three methods coincide), coverage rates under either metric are very close to nominal values even in small samples. However, as the forecasting horizon increases, several important results emerge. The most evident is the severely distorted coverage provided by marginal bands. In terms of $FWER$ metric, the empirical coverage is in the neighborhood of 15% for nominal coverage 68%. These distortions are even more dramatic in terms of the *Wald* metric, with empirical coverage below 1% for $H = 12$ and nominal coverage 68%. At higher coverage levels (95%) the distortions are less dramatic although still considerable (for $H = 12$, the $FWER$ empirical coverage is around the mid-70's% although $Wald$ coverage can sometimes be in the low 20's%). Bonferroni's procedure generates bands that correctly control $FWER$ across all forecast horizons and nominal coverage levels and with empirical coverage close to 95% confidence levels even with $H = 12$. However, there are considerable distortions in terms of

Wald coverage, with empirical levels around 40% for 68% nominal coverage and $H = 12$.

Scheffé bands provide the most accurate match between empirical and nominal *Wald* coverage rates, at all horizons, and at all confidence levels; yet these bands have small distortions in terms of $FWER$ metric, usually within 10% of the corresponding nominal values, thus providing the best overall balance between these two metrics and empirical coverage of all three methods (marginal, Bonferroni and Scheffé). Finally, we did not observe significant differences in performance between forecasts generated from VARs or from direct forecasts.

As a complement to these results, we experimented with a simple AR(1) model whose autoregressive coefficient (ρ) was allowed to vary between 0.5 and 0.9. We did not consider smaller values because at longer horizons the forecasts quickly revert to their unconditional mean. For example, if $\rho = 0.5$ notice that $\rho^{12} = 0.000244$. Further, we isolated the effects of parameter uncertainty, model misspecification, and other sources of model uncertainty to focus exclusively on forecasting uncertainty generated from the arrival of shocks. Insofar as the leading root of higher order processes often provides a good summary of its dynamic properties, we felt that this small-scale set of experiments elucidates for practitioners variations in band coverage as a function of the persistence of the process considered. These results are reported in Table 3 and use 1,000 Monte Carlo replications.

The simulations generally replicate the findings of the VAR examples considered above. As one would expect, the more persistence, the more correlation among the elements of the forecast path and the worse the coverage of the marginal bands. The same is true for Bonferroni bands although the distortions are less severe (and at 95% confidence levels, often

behave quite reasonably). Predictably, the same situations that make marginal bands fail (high correlation among elements of the forecast path), are the situations were correcting for this correlation pays-off. Hence Scheffé bands tend to do considerably better the higher the value of ρ.

No Monte Carlo exercise is ever exhaustive of all the situations practitioners may encounter in practice. However, the results from our simulations clearly indicate that traditional marginal bands provide particularly poor coverage, the worse the more persistence in the data. If interest is in controlling the $FWER$, Bonferroni bands work relatively well but may provide poor *Wald* coverage. In contrast, Scheffé bands manage to strike a convenient balance between $FWER$ and *Wald* control and their coverage is relatively robust to all sorts of coverage levels and forecast horizon choices.

5 A Macroeconomic Forecasting Exercise

On June 30, 2004, the Federal Open Market Committee (FOMC) raised the federal funds rate (the U.S. key monetary policy rate) from 1% to 1.25% – a level it had not reached since interest rates were last changed from 1.5% to 1.25% on November 6, 2002. For more than a year before the June 30, 2004 change, the Federal Reserve had kept the federal funds rate fixed at 1%. This section examines forecasts of the U.S. economy on the eve of the first in a series of interest rate increases that would culminate two years later, on June 29, 2006, with the federal funds rate at 5.25%.

Our out-of-sample forecast exercise examines U.S. real GDP growth (in yearly percentage terms, and seasonally adjusted); inflation (measured by the personal consumption expenditures deflator, in yearly percentage terms, and seasonally adjusted); the federal funds rate;

and the 10 year Treasury Bond rate. All data are measured quarterly (with the federal funds rate and the 10 year T-Bond rate averaged over the quarter) from 1953:II to 2006:II and were the last two years are reserved for evaluation purposes only. With these data, we then construct two-year (eight-quarters) ahead forecasts by direct forecasts. The lag length of the projections was automatically selected to be six by AIC_C.

Figure 3 displays these forecasts along with the actual realizations of these economic variables, conditional and marginal 95% confidence bands, and 95% Scheffé bands. Several results deserve comment. First, the 95% Scheffé bands are more conservative and tend to fan out as the forecast horizon increases. Second, the 95% conditional bands are considerably narrower in all cases but they are meant to capture the uncertainty generated by that period's shock, not the overall uncertainty of the path. Third, our simple exercise results in projections for output and inflation that are more optimistic than the actual data later displayed. As a consequence, our forecast for the federal funds rate is more aggressive (after two years we would have predicted the rate to be at 5.5% instead of 5.25%) although the general pattern of interest rate increases is very similar. Not surprisingly, the 10 year T-Bond rate is also predicted to be higher than it actually was although consistent with a higher inflation premium.

At this point, a forecast report may include other experiments that allow the reader to assess the internal coherence of the exercise. As an illustration, we experimented with choosing a more benign inflation path (perhaps because the end of major military operations in Iraq portended more stability in oil markets would be forthcoming or other factors that may be difficult to quantify within the model). Along these lines, we experimented with a

path for inflation that tracks the lower 95% conditional confidence band so that inflation is predicted to be at 3.4% (rather than at 3.8%) after two years. Of course, this is a completely arbitrary choice – it is not based on any information coming from the data – but this is precisely the objective: to stress the forecasting exercise locally along a direction that differs from that originally predicted but that does not stray too far from the historical experience.

The results of this experiment are reported in Figure 4. We remark that this alternative path is very conservative: the *Wald* distance between the alternative and the original inflation forecast path is 29% in probability units, suggesting that such an experiment is well within the experience observed in the historical sample. In all cases, the exogeneity metric indicates that the paths for output, the federal funds rate and the 10-year T-Bond rate are not exogenous to variations in the path of inflation, as might have been expected a priori on economic grounds.

Interestingly, the forecasts obtained by conditioning on this alternative path for inflation are remarkably close to the actual data later observed. In particular, the path of predicted increases in the federal funds rate is virtually identical to the actual path observed, whereas the path of the 10 year T-Bond rate is mostly within the 95% conditional bands. The most significant difference was a slight drop in output after one year to a 3% growth rate that in the conditional exercise was predicted to be closer to 3.5%, but otherwise both paths seem to reconnect at the end of the two year predictive horizon. Obviously, we are not speculating that this alternative scenario reflected the Federal Reserve's view on inflation at the time – ours is not a statement about actual behavior. Rather, it serves to illustrate how staff forecasters could have formally presented small-scale alternative assumptions about the

paths of some of the variables in the forecasting exercise and their effect on the predictions made about the paths of other variables in the system.

6 Conclusions

Error bands around forecasts summarize the uncertainty the professional forecaster has about his predictions and are an elementary tool of communication. When forecasts are generated over a sequence of future periods – a path forecast – this paper shows that error bands should be derived as a simultaneous confidence region. The common practice of building error bands from the marginal distribution of each point forecast does not provide appropriate probability coverage; is a misleading representation of the set of possible paths the predicted variable may take; and should therefore be abandoned.

This paper provides a satisfactory solution to the problem of graphically summarizing the range of possible paths, based on an application of Scheffé's (1953) S-method of simultaneous inference; the realization that the Cholesky decomposition orthogonalizes the forecast path's covariance matrix by projecting each forecast on to its immediate past; and designing a testing procedure that controls the overall variation in the path with the Mahalanobis (1936) distance to ensure each individual point forecast is internally consisten with the path as a whole.

The result is a set of bands (that we call Scheffé bands) which balance the family-wise error rate (the probability that one or more elements of the path will lie outside the bands) with a measure of the maximal joint variation based on the simultaneous *Wald* score (the probability that, jointly, the elements of the path are "close" in probability distance units). Monte Carlo experiments demonstrate that Scheffé bands provide approximately correct

probability coverage under either of these measures whereas marginal bands or bands based on Bonferroni's procedure fail in one or both metrics, often quite substantially.

When path forecasts are reported for more than one variable, another way to evaluate the properties of the forecasting exercise is to examine that the path forecasts across different variables are internally consistent. The coherence of the forecasting exercise can be analyzed by examining alternative scenarios – a common feature in many forecast reports. To ensure that the alternative scenarios do not stress the predictive model over regions where the sample provides no training, we provide a simple *Wald* score that measures the probability distance to the conditional mean path. In addition, the *Wald* score can be used to measure the sensitivity of each variable in the system to the set of alternative paths.

The basic statistical principles discussed in this paper suggest a number of intriguing research directions. In a sequel to this paper, we investigate ways in which predictive ability measures and statistics can be extended to path forecasts. It is well known that, relative to simple specifications, more elaborate models tend to predict well in the short-run and poorly in the long-run. Instead, we are interested in assessing a model's performance with respect to its ability to predict general dynamic patterns even at the cost of imprecision in specific point forecasts. Hence, we are developing measures based on the Mahalanobis (1936) distance alternative to the commonly used MSFE, which are based on the Euclidean distance instead. The new measures integrate the correlation patterns in a path forecast, and serve as the basis to construct tests of predictive ability along the lines of Giacomini and White (2006) based on multivariate *Wald* scores.

7 Appendix

We begin by stating our assumptions on the DGP described in Section 3 to which the reader is referred for any doubts about the notation.

Assumption 1: Suppose the K-dimensional vector of weakly stationary variables, $\mathbf{y}_t$ has a Wold representation given by

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Phi_j' \mathbf{u}_{t-j}, \tag{20}$$

where the moving-average coefficient matrices Φ_j are of dimension $k \times k$, and we assume that:

(i) $E(\mathbf{u}_t) = 0$; and u_t are *i.i.d.* and Gaussian

(ii) $E(\mathbf{u}_t \mathbf{u}_t') = \Sigma_u < \infty$.

(iii) $\sum_{j=0}^{\infty} ||\Phi_j|| < \infty$ where $||\Phi_j||^2 = tr\left(\Phi_j' \Phi_j\right)$ is the equivalent of the Euclidean L_2 norm for matrices and $\Phi_0 = I_k$.

(iv) $\det\{\Phi(z)\} \neq 0$ for $|z| \leq 1$ where $\Phi(z) = \sum_{j=0}^{\infty} \Phi_j z^j$.

Then the process in (20) can also be written as an infinite VAR process (see, e.g. Anderson, 1994),

$$\mathbf{y}_t = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \tag{21}$$

such that,

(v) $\sum_{j=1}^{\infty} ||A_j|| < \infty$.

(vi) $A(z) = I_k - \sum_{j=1}^{\infty} A_j z^j = \Phi(z)^{-1}$.

(vii) $\det\{A(z)\} \neq 0$ for $|z| \leq 1$.

Assumption 1 includes the class of stationary vector autoregressive moving average, VARMA(p,q) processes as a special case. Lewis and Reinsel (1985) derive conditions under which a finite order VAR will provide consistent and asymptotically normal estimates of the p original autoregressive coefficient matrices A_j in expression (21). We will use this result momentarily and extend it for local projections when deriving the asymptotic distribution of the forecast path. The *i.i.d.* assumption could be relaxed to allow for heteroskedasticity so that the consistency and asymptotic normality results in Lewis and Reinsel (1985) are derived with appropriate laws of large numbers and central limit theorems for martingale difference sequences (*m.d.s.*) under more general mixing conditions, but these are not explored here. We prefer to trade-off some sophistication for clarity to illustrate the more important points we discuss below. Similarly, the assumption of Gaussian errors could be relaxed, but then the distribution of the forecast errors would no longer be Normal and should be obtained by means of simulation methods, see e.g. Garratt et al. (2003).

Assumption 2: If $\{\mathbf{y}_t\}$ satisfies conditions (i)-(vii) in assumption 1 and:

(i) $E|u_{it}u_{jt}u_{rt}u_{lt}| < \infty$ for $1 \leq i,j,r,l \leq k$.

(ii) p is chosen as a function of R such that

$$\frac{p^3}{R} \to 0 \text{ as } R,p \to \infty.$$

(iii) p is chosen as a function of R such that

$$p^{1/2} \sum_{j=p+1}^{\infty} ||A_j|| \to 0 \text{ as } R,p \to \infty.$$

Then, a summary of results shown by Lewis and Reinsel (1985), Lütkepohl and Poskitt (1991) and Jordà and Kozicki (2007) are contained in the following corollary.

Corollary 2 *Given assumptions 1 and 2, the VAR(p) and p^{th} order local projections are consistent and asymptotically normal, specifically:*

(a) $\widehat{A}_j \xrightarrow{p} A_j$; $\widehat{A}_j^h \xrightarrow{p} A_j^h$ *and* $\widehat{A}_1^h \xrightarrow{p} \Phi_h$.

(b) $\sqrt{\frac{R-p-h}{p}} vec\left(\widehat{A}(p) - A(p)\right) \xrightarrow{d} N(0, \Sigma_a)$ *where* $\Sigma_a = \Gamma_p^{-1} \otimes \Sigma_u$

(c) $\sqrt{\frac{R-p-h}{p}} vec\left(\widehat{A}(p,h) - A(p,h)\right) \xrightarrow{d} N(0, \Sigma_\alpha)$ *where* $\Sigma_\alpha = \Gamma_p^{-1} \otimes \Omega_h$ *and* $\Omega_h = \Phi(I_h \otimes \Sigma_u)\Phi'$ *where*

$$\mathbf{\Phi} = \begin{bmatrix} I_K & \mathbf{0} & ... & \mathbf{0} \\ \Phi_1 & I_K & ... & \mathbf{0} \\ \vdots & \vdots & ... & \vdots \\ \Phi_{h-1} & \Phi_{h-2} & ... & I_K \end{bmatrix}$$

(d) *Let* $\widehat{\mathbf{u}}(p)_t \equiv y_t - \widehat{\mathbf{m}} - \sum_{j=1}^{p} \widehat{A}_j y_{t-j}$ *so that* $\widehat{\Sigma}_u(p) = (R-p)^{-1} \sum_{t=1}^{R} \widehat{\mathbf{u}}(p)_t \widehat{\mathbf{u}}(p)_t'$ *then* $\sqrt{R}\left(\widehat{\Sigma}_u(p) - \Sigma_u\right) \rightarrow N(0, \Omega_\Sigma)$ *where* Ω_Σ *is the covariance matrix of the residual covariance matrix.*

Several results deserve comment. Technically speaking, condition (ii) in assumption 2 is required for asymptotic normality but not for consistency, where the weaker condition $p^2/R \rightarrow 0$, $R, p \rightarrow \infty$ is sufficient. Results (a)-(c) show that estimators of truncated models are consistent and asymptotically normal. Result (d) is useful if one prefers to rotate the vector of endogenous variables $\mathbf{y}_t$ when providing structural interpretations for the forecast exercise. Here though, we abstain of such interpretation and provide the result only for completeness.

We find it convenient to momentarily alter the order of our derivations and begin by examining forecasts from direct forecasts first, since these are linear functions of parameter estimates and hence can be obtained in a straightforward manner.

First notice that $\widehat{Y}_\tau(H) = \widehat{\mathbf{A}} X_{\tau,p}$ and hence

$$\frac{\partial vec\left(\widehat{Y}_\tau(H)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)} = \frac{\partial vec\left(\widehat{\mathbf{A}} X_{\tau,p}\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)} = \underset{KH \times K^2Hp+KH}{\left(X'_{\tau,p} \otimes I_{KH}\right)}, \tag{22}$$

which combined with corollary 1(c) results in

$$\sqrt{\frac{R-p-H}{p}}\left(vec\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right) \xrightarrow{d} N(0, \Sigma_A) \tag{23}$$

$$\underset{K^2Hp+KH \times K^2Hp+KH}{\Sigma_A} = \Gamma_p^{-1} \otimes \Omega_H; \quad \underset{KH \times KH}{\Omega_H} = \mathbf{\Phi}\left(I_H \otimes \Sigma_u\right)\mathbf{\Phi}'$$

Putting together expressions (17), (18), (22) and (23), we arrive at the following corollary.

Corollary 3 *Under assumptions 1 and 2 and expressions (18), (17), (22) and (23), the asymptotic distribution of the forecast path generated with the local projections approach described in assumption 1 is*

$$\sqrt{\frac{R-p-H}{p}} vec\left(\widehat{Y}_\tau(H) - Y_{\tau,H} | \mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \xrightarrow{d} N(\mathbf{0}; \Xi_H) \tag{24}$$

$$\begin{aligned}
\Xi_H &= \left\{\frac{p}{R-p-H}\Omega_H + \Psi_H\right\} \\
\Omega_H &= \mathbf{\Phi}(I_H \otimes \Sigma_u)\mathbf{\Phi}' \\
\Psi_H &= \left(X'_{\tau,p} \otimes I_{KH}\right)\left[\Gamma_p^{-1} \otimes \Omega_H\right]\left(X_{\tau,p} \otimes I_{KH}\right)
\end{aligned}$$

In practice, all population moments can be substituted by their conventional sample counterparts.

We now return to the more involved derivation of the asymptotic distribution of the forecast path when the forecasts are generated by the VAR(p) in expression (13). For this purpose, we find it easier to work with each element of the vector $\widehat{Y}_\tau(H)$ individually, so

that we begin by examining the derivation of

$$\sqrt{\frac{R-p-H}{p}}vec\left(\widehat{\mathbf{y}}_{\tau}\left(h\right)-\mathbf{y}_{\tau}\left(h\right)|\mathbf{y}_{\tau},\mathbf{y}_{\tau-1},...\right)\xrightarrow{d}N\left(\mathbf{0};\Psi_{h,h}\right)$$

$$\Psi_{h,h} = \frac{\partial vec\left(\widehat{\mathbf{y}}_{\tau}\left(h\right)\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)}\Sigma_{a}\frac{\partial vec\left(\widehat{\mathbf{y}}_{\tau}\left(h\right)\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)}$$

where we remind the reader that from corollary 1(b), $\Sigma_a = \Gamma_p^{-1} \otimes \Sigma_u$. In general, notice that

$$\Psi_{i,j} = \frac{\partial vec\left(\widehat{\mathbf{y}}_{\tau}\left(i\right)\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)}\Sigma_{a}\frac{\partial vec\left(\widehat{\mathbf{y}}_{\tau}\left(j\right)\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)}$$

which is all we need to construct all the elements in the asymptotic covariance matrix of $\widehat{Y}_{\tau}\left(H\right)$, namely Ψ_H. An expression for $\widehat{\mathbf{y}}_{\tau}\left(h\right)$ generated from the VAR(p) in expression (13) can be obtained as

$$\widehat{\mathbf{y}}_{\tau}\left(h\right) = SB^{h}X_{\tau,p}$$

where B simply stacks the VAR(p) coefficients in companion form and S is a selector matrix, both of which are

$$\underset{Kp+1\times Kp+1}{B} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ \mathbf{m} & A_1 & A_2 & ... & A_{p-1} & A_p \\ 0 & I_K & 0 & ... & 0 & 0 \\ 0 & 0 & I_K & ... & 0 & 0 \\ \vdots & \vdots & \vdots & ... & \vdots & \vdots \\ 0 & 0 & 0 & ... & I_K & 0 \end{pmatrix},$$

$$\underset{K\times Kp+1}{S} = (\underset{K\times 1}{\mathbf{0}}, \underset{K\times K}{I_K}, \underset{K\times K}{\mathbf{0}_K}, ..., \underset{K\times K}{\mathbf{0}_K}).$$

Therefore, notice that

$$\frac{\partial vec\left(\widehat{\mathbf{y}}_{\tau}\left(h\right)\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)} = \frac{\partial vec\left(SB^{h}X_{\tau,p}\right)}{\partial vec\left(\widehat{A}\left(p\right)\right)} = \sum_{i=0}^{h-1} X'_{\tau,p}(B')^{h-1-i}\otimes\Pi_i, \quad \Pi_i = SB^iS'.$$

The following corollary characterizes the asymptotic distribution of VAR(p) generated forecasts paths.

Corollary 4 *Under assumptions 1 and 2, the asymptotic distribution of the forecast path* $\widehat{Y}_\tau(H)$ *generated from the VAR(p) in expression (13) is given by*

$$\sqrt{\frac{R-p-H}{p}}vec\left(\widehat{Y}_\tau(H) - Y_{\tau,H}|\mathbf{y}_\tau, \mathbf{y}_{\tau-1}, ...\right) \xrightarrow{d} N(\mathbf{0};\Xi_H) \tag{25}$$

$$\begin{aligned}
\Xi_H &= \left\{\frac{p}{R-p-H}\Omega_H + \Psi_H\right\} \\
\Omega_H &= \mathbf{\Phi}(I_H \otimes \Sigma_u)\mathbf{\Phi}' \\
\Psi_{i,j} &= \frac{p}{R-p-H}\sum_{k=0}^{i-1}\sum_{s=0}^{j-1} E(X'_{\tau,p}(B')^{i-1-k}\Gamma_p^{-1}B^{j-1-s}X_{\tau,p}) \otimes \Pi_k\Sigma_u\Pi'_s \\
&= \frac{p}{R-p-H}\sum_{k=0}^{i-1}\sum_{s=0}^{j-1} tr((B')^{i-1-k}\Gamma_p^{-1}B^{j-1-s}\Gamma_p)\Pi_k\Sigma_u\Pi'_s
\end{aligned}$$

In practice all moment matrices can be substituted by their sample counterparts as usual.

References

Anderson, Theodore W. (1994) **The Statistical Analysis of Time Series Data.** New York, NY: Wiley Interscience.

Baillie, Richard T. and Tim Bollerslev (1992) "Prediction in Dynamic Models with Time-Dependent Covariances," *Journal of Econometrics*, 52: 91-113.

Bank of England, "Inflation Report," available quarterly since 1997 at http://www.bankofengland.co.uk/publications/inflationreport/index.htm.

Bowden, David C. (1970) "Simultaneous Confidence Bands for Linear Regression Models," *Journal of the American Statistical Association*, 65(329): 413-421.

Cameron, A. Colin and Pravin Trivedi (2005) **Microeconometrics: Methods and Applications.** Cambridge, U.K.: Cambridge University Press.

Clements, Michael P. and David F. Hendry (1993) "On the Limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting,* 12: 617-676.

Garratt, A., Lee, K., Pesaran, H. M., and Shin, Y. (2003) "Forecast Uncertainties in Macroeconometric Modelling: An Application to the UK Economy", *Journal of the American Statistical Association*, 98: 829-838.

Giacomini, Raffaella and Halbert White (2006) "Tests of Conditional Predictive Ability," *Econometrica,* 74(6): 1545-1578.

Greenspan, Alan (2003) Remarks at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming on August 29, 2003. Available at: http://www.federalreserve.gov/boarddocs/speeches/2003/20030829/.

Hamilton, James D. (1994). **Time Series Analysis.** Princeton, NJ: Princeton University Press.

Hochberg, Yosef and Ajit C. Tahmane (1987) **Multiple Comparison Procedures.** New York, NY: John Wiley and Sons, Inc.

Holm, S. (1979) "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics,* 6:65-70.

Hurvich, Clifford M. and Chih-Ling Tsai (1993) "A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection," *Journal of Time Series Analysis*, 14: 271-279.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1): 161-182.

Jordà, Òscar (2008) "Simultaneous Confidence Regions for Impulse Responses," *Review of Economics and Statistics,* forthcoming.

Jordà, Òscar and Sharon Kozicki (2007) "Estimation and Inference by the Method of Projection Minimum Distance," U.C. Davis working paper 07-8.

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

Lehmann, E. L. and Joseph P. Romano (2005) **Testing Statistical Hypothesis.** Berlin, Germany: Springer-Verlag.

Lewis, R. A. and Gregory C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis,* 16(33): 393-411.

Lütkepohl, Helmut (2005) **New Introduction to Multiple Time Series**. Berlin, Germany: Springer-Verlag.

Lütkepohl, Helmut and P. S. Poskitt (1991) "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," *Econometric Theory,* 7:487-496.

Mahalanobis, Prasanta C. (1936) "On the Generalized Distance in Statistics," *Proceedings of the National Institute of Science of India,* 12: 49-55.

Marcellino, Massimiliano, James H. Stock and Mark W. Watson (2006) "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics,* 127(1-2): 499-526.

Mitchell, Ann F. S. and Wojtek J. Krzanowski (1985) "The Mahalanobis Distance and Elliptic Distributions," *Biometrika*, 72: 464-467.

Ringland, J. T. (1983) "Robust Multiple Comparisons," *Journal of the American Statistical Association,* 78: 145-151.

Roy, Sumarendra N. (1953) "On a Heuristic Method of Test Construction and its use in Multivariate Analysis," *Annuals of Mathematical Statistics,* 24: 220-238.

Roy, Sumarendra N. and Raj C. Bose (1953) "Simultaneous Confidence Interval Estimation," *Annals of Mathematical Statistics,* 24: 513-536.

Savin, N. Eugene (1980) "The Bonferroni and Scheffé Multiple Comparison Procedures," *Review of Economic Studies,* 47: 225-273.

Savin, N. Eugene (1984) "Multiple Hypothesis Testing," in **Handbook of Econometrics, v.II,** Zvi Griliches and Michael D. Intrilligator (eds.). Amsterdam: Elsevier.

Scheffé, Henry (1953) "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika,* 40: 87-104.

Scheffé, Henry (1959) **Analysis of Variance.** New York, NY: John Wiley and Sons Inc.

Stock, James H. and Mark W. Watson (2001) "Vector Autoregressions," *Journal of Economic Perspectives,* 15(4): 101-115.

Waggoner, Daniel F. and Tao Zha (1999) "Conditional Forecasts in Dynamic Multivariate Models," *Review of Economics and Statistics,* 81(4): 639-651.

Wald, Abraham (1943) "Tests of Statistical Hypothesis Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society,* 54(3): 426-482.

Table 1. Coverage Rates of Marginal, Bonferroni, and Scheffé Bands in Stock and Watson's (2001) VAR(4). Forecasts Obtained with VARs

Forecast Horizon: 1

		Nominal Coverage: 68%						Nominal Coverage: 95%					
			FWER			WALD			FWER			WALD	
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	67.5	67.5	67.5	67.5	67.5	67.5	93.8	93.8	93.8	93.8	93.8	93.8
	UN	69.5	69.5	69.5	69.5	69.5	69.5	95.8	95.8	95.8	95.8	95.8	95.8
	FF	68.4	68.4	68.4	68.4	68.4	68.4	94.6	94.6	94.6	94.6	94.6	94.6
T=400	**P**	66.9	66.9	66.9	66.9	66.9	66.9	93.6	93.6	93.6	93.6	93.6	93.6
	UN	69.7	69.7	69.7	69.7	69.7	69.7	96.0	96.0	96.0	96.0	96.0	96.0
	FF	67.8	67.8	67.8	67.8	67.8	67.8	94.2	94.2	94.2	94.2	94.2	94.2

Forecast Horizon: 4

		Nominal Coverage: 68%						Nominal Coverage: 95%					
			FWER			WALD			FWER			WALD	
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	32.8	78.7	58.4	20.5	70.2	67.1	85.6	95.9	92.8	80.3	95.0	95.4
	UN	43.6	82.4	63.8	15.3	60.8	67.5	88.1	96.5	93.8	72.0	90.8	94.1
	FF	37.0	79.7	61.0	15.8	65.3	68.0	86.4	95.8	93.7	76.1	92.5	94.6
T=400	**P**	29.8	76.7	56.7	21.5	73.0	67.0	83.9	95.5	92.4	82.9	96.8	96.6
	UN	43.8	83.2	64.2	15.4	62.2	68.6	88.7	97.2	94.2	73.0	91.9	95.1
	FF	36.3	79.5	60.8	15.3	65.8	68.3	86.4	96.1	93.4	76.4	92.9	94.9

Forecast Horizon: 8

		Nominal Coverage: 68%						Nominal Coverage: 95%					
			FWER			WALD			FWER			WALD	
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	16.7	81.8	56.2	1.5	50.4	63.1	78.7	95.8	91.8	43.9	86.6	93.7
	UN	27.9	84.9	63.2	2.1	58.6	65.7	82.0	96.6	93.7	52.0	91.3	95.6
	FF	24.4	84.0	63.0	1.9	50.8	66.1	80.9	96.3	93.6	44.0	87.4	95.4
T=400	**P**	13.5	79.9	54.5	1.4	52.5	65.4	77.0	95.7	91.8	45.7	90.6	96.4
	UN	27.6	85.8	63.8	1.9	60.2	67.8	82.8	97.2	94.2	53.0	93.4	96.9
	FF	23.5	84.7	62.9	1.8	50.3	68.1	81.4	96.7	93.8	43.1	89.2	96.9

Forecast Horizon: 12

		Nominal Coverage: 68%						Nominal Coverage: 95%					
			FWER			WALD			FWER			WALD	
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	12.4	84.2	57.2	0.2	37.3	61.7	74.2	96.2	91.9	21.5	77.5	92.8
	UN	19.1	85.7	62.1	0.4	66.3	66.1	77.0	96.4	92.3	46.6	93.7	96.1
	FF	15.7	85.0	62.4	0.1	39.9	64.9	76.2	96.4	93.1	22.2	81.3	95.4
T=400	**P**	9.1	83.3	55.5	0.1	37.0	65.8	72.5	96.6	92.2	19.9	81.4	96.0
	UN	18.2	86.4	63.1	0.2	71.2	69.4	77.2	97.0	93.3	49.7	96.8	97.7
	FF	14.8	85.9	62.5	0.1	40.0	68.4	77.2	97.5	93.6	21.3	84.7	97.7

Notes: 1,000 samples generated on which a VAR is fitted and whose order is selected automatically by AIC_C. Each estimated VAR on these 1,000 samples generates a forecast error variance (which includes estimation uncertainty) for the forecast path and hence the sets of bands (marginal, Bonferroni, and Scheffé) used in the analysis. Hence 1,000 forecast paths from the true DGP are generated and then compared with each set of 1,000 bands to determine the appropriate coverage rates. FWER stands for "family-wise error rate" and simply computes the proportion of paths strictly inside the bands. WALD instead is the proportion of forecast paths whose joint Wald statistic relative to the forecast, attains a value that is lower than that implied by the Wald statistic for the bands.

Table 2. Coverage Rates of Marginal, Bonferroni, and Scheffé Bands in Stock and Watson's (2001) VAR(4). Forecasts Obtained by Direct Forecasts

Forecast Horizon: 1

		Nominal Coverage: 68%						Nominal Coverage: 95%					
		FWER			WALD			FWER			WALD		
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	67.5	67.5	67.5	67.5	67.5	67.5	93.8	93.8	93.8	93.8	93.8	93.8
	UN	69.5	69.5	69.5	69.5	69.5	69.5	95.8	95.8	95.8	95.8	95.8	95.8
	FF	68.4	68.4	68.4	68.4	68.4	68.4	94.6	94.6	94.6	94.6	94.6	94.6
T=400	**P**	66.9	66.9	66.9	66.9	66.9	66.9	93.6	93.6	93.6	93.6	93.6	93.6
	UN	69.8	69.8	69.8	69.8	69.8	69.8	96.0	96.0	96.0	96.0	96.0	96.0
	FF	67.9	67.9	67.9	67.9	67.9	67.9	94.2	94.2	94.2	94.2	94.2	94.2

Forecast Horizon: 4

		Nominal Coverage: 68%						Nominal Coverage: 95%					
		FWER			WALD			FWER			WALD		
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	30.5	76.4	55.9	24.9	76.3	68.0	83.8	95.0	91.7	85.4	97.1	96.4
	UN	41.7	80.7	63.1	16.7	63.6	68.7	86.9	95.7	93.6	74.5	92.0	94.8
	FF	34.6	77.2	59.4	18.0	69.2	69.0	84.4	94.7	93.0	79.3	94.2	95.5
T=400	**P**	29.5	76.3	56.2	22.1	74.0	67.1	83.6	95.3	92.2	83.8	97.1	96.7
	UN	43.3	82.8	64.1	15.6	62.7	68.8	88.4	97.0	94.2	73.6	92.1	95.3
	FF	35.9	79.0	60.5	15.6	66.5	68.4	86.1	95.9	93.7	76.9	93.2	95.1

Forecast Horizon: 8

		Nominal Coverage: 68%						Nominal Coverage: 95%					
		FWER			WALD			FWER			WALD		
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	13.5	78.6	52.9	3.1	65.1	68.2	75.3	94.4	90.0	58.6	93.9	96.0
	UN	25.1	81.2	61.2	3.9	69.7	68.9	78.7	95.2	92.7	63.4	95.1	96.6
	FF	21.1	80.6	60.7	3.3	64.0	70.6	77.3	94.7	92.5	57.4	93.5	96.9
T=400	**P**	12.8	79.2	53.9	1.6	56.2	66.6	76.3	95.3	91.6	49.2	92.3	96.7
	UN	26.8	84.9	63.5	2.3	62.9	68.5	81.9	96.9	94.1	55.9	94.3	97.2
	FF	22.7	83.8	62.6	2.0	53.2	68.8	80.5	96.4	93.6	45.8	90.6	97.2

Forecast Horizon: 12

		Nominal Coverage: 68%						Nominal Coverage: 95%					
		FWER			WALD			FWER			WALD		
		Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.	Marg.	Bonf.	Schef.
T=100	**P**	9.3	80.4	52.8	0.7	58.6	69.6	69.3	94.8	89.4	39.4	90.7	95.9
	UN	16.6	82.2	55.7	2.3	83.3	71.2	72.9	94.8	87.3	68.7	97.7	97.1
	FF	12.8	81.1	58.7	0.5	65.0	72.8	71.4	94.6	91.3	45.0	93.7	97.5
T=400	**P**	8.3	82.2	54.7	0.2	42.5	67.7	71.0	96.2	91.9	24.0	85.5	96.6
	UN	17.3	85.3	62.0	0.3	76.2	70.6	75.9	96.6	92.3	56.0	97.7	97.9
	FF	14.0	85.0	62.0	0.1	45.1	70.2	76.0	97.0	93.4	24.9	87.8	98.1

Notes: 1,000 samples generated on which local projections are fitted and whose order is selected automatically by AIC_C. From each of these 1,000 samples, one obtains the forecast error variance (which includes estimation uncertainty) for the forecast path and hence the sets of bands (marginal, Bonferroni, and Scheffé) used in the analysis. Hence 1,000 forecast paths from the true DGP are generated and then compared with each set of 1,000 bands to determine the appropriate coverage rates. FWER stands for "family-wise error rate" and simply computes the proportion of paths strictly inside the bands. WALD instead is the proportion of forecast paths whose joint Wald statistic relative to the forecast, attains a value that is lower than that implied by the Wald statistic for the bands.

Table 3. Coverage Rates of Marginal, Bonferroni, and Scheffé Bands in Simple AR(1) Model

Nominal Coverage Level: 68%										
	Horizon = 1					**Horizon = 4**				
FWER	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	68	67.5	67.8	68.5	68.2	27.3	28.3	28.1	34.2	33.2
Bonf.	68	67.5	67.8	68.5	68.2	73.8	77.3	75	79	77.8
Schef.	68	67.5	67.8	68.5	68.2	53	54.3	55.9	62.3	60.6
WALD	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	68	67.5	67.8	68.5	68.2	26.3	24.7	20.5	20.6	15.5
Bonf.	68	67.5	67.8	68.5	68.2	83	79	72.6	69.1	61.8
Schef.	68	67.5	67.8	68.5	68.2	67.4	69.2	65.7	68.2	66.6
Nominal Coverage Level: 68%										
	Horizon = 8					**Horizon = 12**				
FWER	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	6.8	8.9	9.9	15.4	22.4	1.8	3.6	3.8	6.6	11.6
Bonf.	76.8	76	76.1	79.5	82.8	74.6	75.2	79.4	80.5	85.3
Schef.	42.3	49.6	52.3	59.8	59.3	33.3	42.2	53.4	59.5	59.5
WALD	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	8.2	6.2	3.1	2.2	1	3.3	1.7	0.4	0.3	0.1
Bonf.	93	85.4	71.9	62.7	50.7	98	90.6	78	56.1	37.1
Schef.	69.4	68.4	65.9	68.3	69.2	69.2	67.2	69.6	68.5	69.9

Nominal Coverage Level: 95%										
	Horizon = 1					**Horizon = 4**				
FWER	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	94.7	94.8	95.3	96.2	94.4	82.6	85.9	84.1	86.7	84.2
Bonf.	94.7	94.8	95.3	96.2	94.4	95.2	95.5	95.9	96.6	95.5
Schef.	94.7	94.8	95.3	96.2	94.4	90.4	93.3	93.8	95.2	92.9
WALD	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	94.7	94.8	95.3	96.2	94.4	91	87.7	83.3	80.2	74.9
Bonf.	94.7	94.8	95.3	96.2	94.4	99.3	97.9	96.8	96.5	93.1
Schef.	94.7	94.8	95.3	96.2	94.4	98.1	97.2	97	97.5	95.3
Nominal Coverage Level: 95%										
	Horizon = 8					**Horizon = 12**				
FWER	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	72.6	70.5	71.2	75	79.4	57.8	59.1	64	69.9	75.7
Bonf.	95.7	95.3	96	92.4	97.6	95.4	95.6	95.3	96.8	97.2
Schef.	87.7	91.7	92.2	95.9	95.2	80.2	87.5	92.2	92.3	93.9
WALD	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Marg.	89.1	80.3	65.4	56.3	43.7	89.1	72.9	55.8	34.5	18.7
Bonf.	99.9	99.6	97.8	92.4	90	100	99.8	98.8	93.2	81
Schef.	98.6	98.4	97.8	95.9	97.2	99.6	98.4	98.3	97.3	97.3

Notes: Theoretical values of the forecast error variance (excluding parameter estimation uncertainty) are used to construct three sets of bands (marginal, Bonferroni, and Scheffé). Then 1,000 Monte Carlo replications from the DGP are generated. FWER stands for "family-wise error rate" and computes the proportion of paths inside the bands. WALD computes the Wald statistic for each path relative to its forecast and computes the proportion whose value is lower than the Wald statistic implied by the bands.

Figure 1 – 95% Confidence Circumference for Orthogonalized Forecast Path

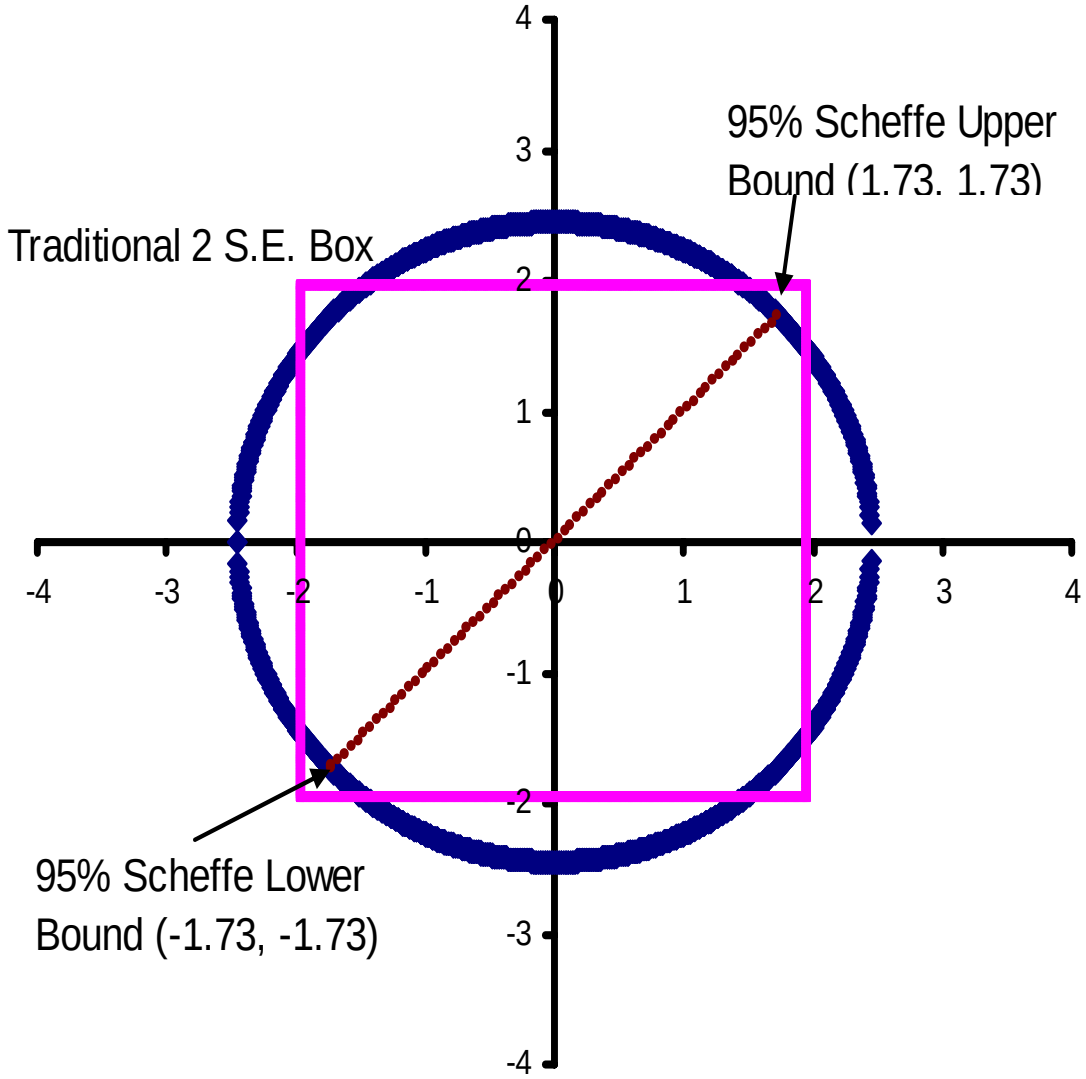
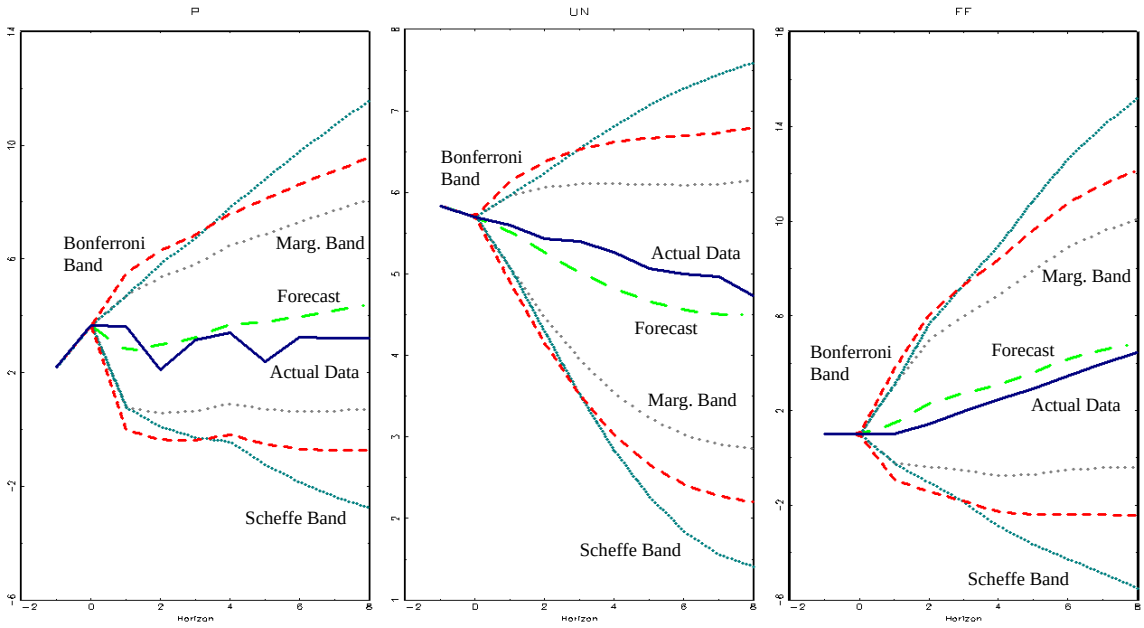
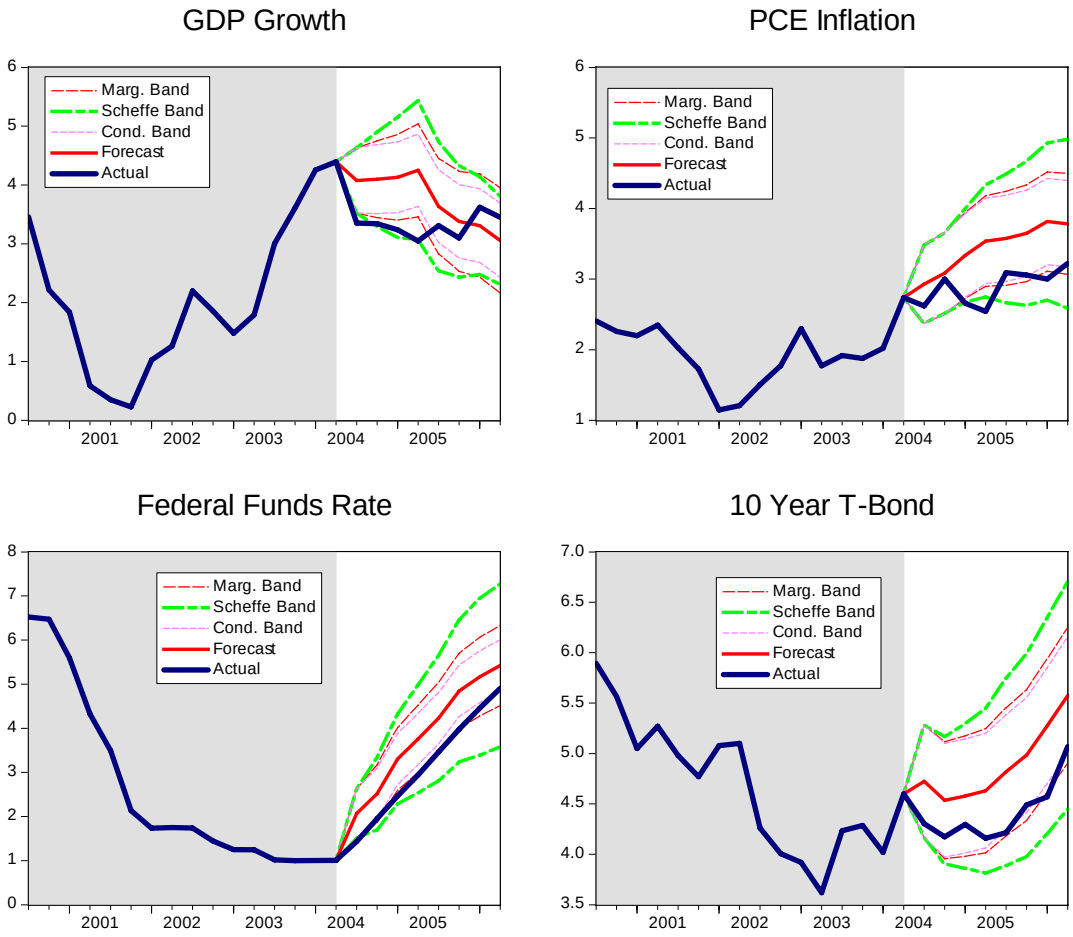


Figure 2. Stock and Watson (2001) Out-of-Sample Forecasts, 8-periods Ahead



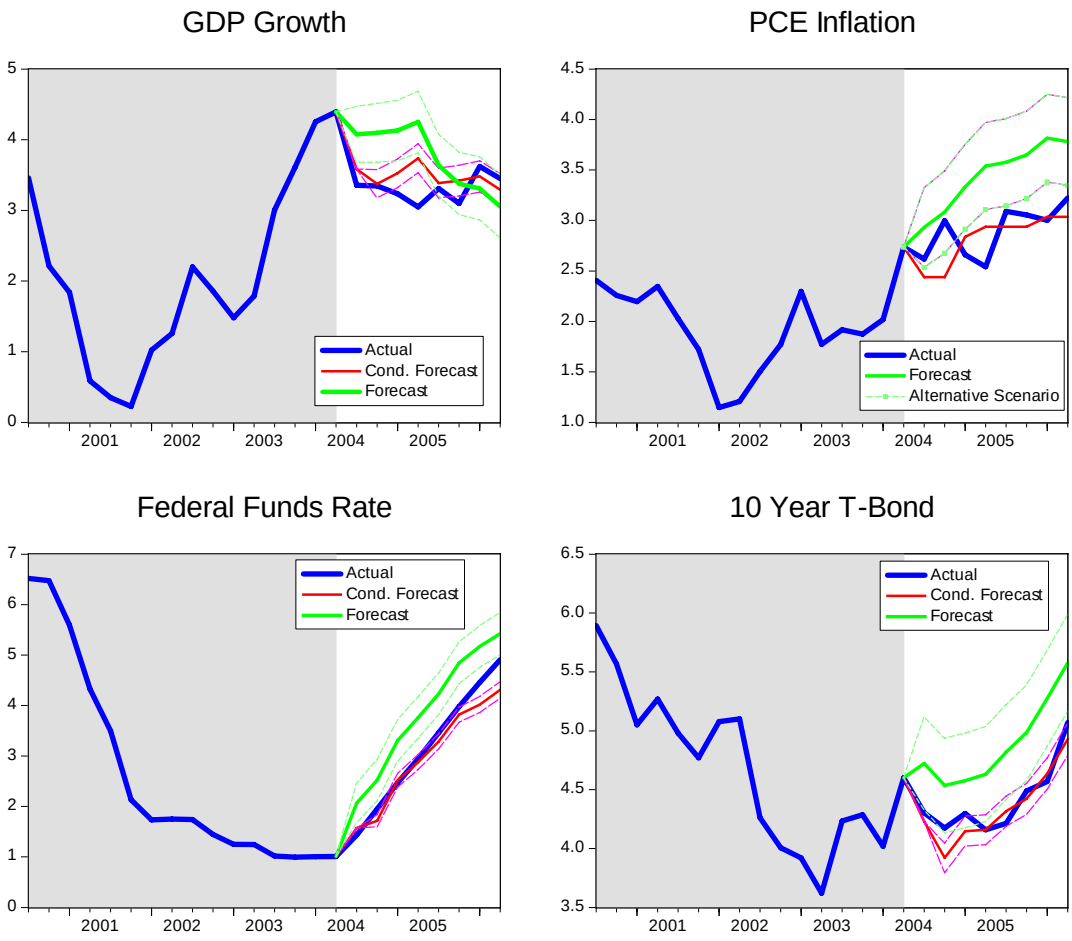
Notes: Out-of-sample forecasts for the Stock and Watson (2001) VAR. Estimation sample 1960:I-2004:IV. Prediction sample 2005:I-2007:I. Predictions based on VAR(4). P stands for inflation (measured by the chain-weighted GDP price index), UN stands for unemployment (measured by the civilian unemployment rate), and FF stands for federal funds rate (average over the quarter).

Figure 3. 95% Marginal, Scheffé and Conditional Error Bands and Forecast



Notes: Estimation sample: 1953:II – 2004:II; out-of-sample forecast period: 2004:II – 2006:II

Figure 4. Forecasts Conditional on Alternative Inflation Path



Notes: Estimation sample: 1953:II – 2004:II; out-of-sample forecast period: 2004:II – 2006:II Conditional bands shown for original forecast and for forecasts conditional on alternative inflation path