



# EUI Working Papers

ECO 2008/34

Path Forecast Evaluation

Oscar Jordà and Massimiliano Marcellino



**EUROPEAN UNIVERSITY INSTITUTE  
DEPARTMENT OF ECONOMICS**

*Path Forecast Evaluation*

**OSCAR JORDÀ**

**and**

**MASSIMILIANO MARCELLINO**

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

The author(s)/editor(s) should inform the Economics Department of the EUI if the paper is to be published elsewhere, and should also assume responsibility for any consequent obligation(s).

ISSN 1725-6704

© 2008 Oscar Jordà and Massimiliano Marcellino

Printed in Italy  
European University Institute  
Badia Fiesolana  
I – 50014 San Domenico di Fiesole (FI)  
Italy

<http://www.eui.eu/>  
<http://cadmus.eui.eu/>

## 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, error bands.

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

Massimiliano Marcellino  
Università Bocconi, IGER and CEPR  
Via Salasco 5  
20136 Milan, Italy  
**e-mail:** massimiliano.marcellino@uni-bocconi.it

---

\*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 forecast is as important as the forecast itself. When predictions are made over several periods, such uncertainty is encapsulated by the joint predictive density of the *path forecast*. There are many questions of interest that can be answered with the marginal distribution of the forecasts at each individual horizon. These are the questions that have received the bulk of attention in the literature and are coded into most commercial econometric packages. For example, mean-squared forecast errors (MSFE) that are reported for each forecast horizon individually; two standard-error band plots that are based on the marginal distribution of each individual forecast error; and fan charts that are constructed from the percentiles of marginal predictive densities.

The basic message of this paper is that many questions of interest require knowledge of the joint predictive density, not the collection of marginal predictive densities alone. The joint distribution and the covariance matrix of the path forecast thus play a prominent role in our discussion. They can be obtained either by simulation methods, see e.g. Garratt, Pesaran and Shin (2003), or analytically for a variety of cases as Section 4 will show.

Information about the range of possible paths the predicted variable may follow is con-

tained in a simultaneous confidence region. Thus, a 95% confidence multi-dimensional ellipse based on the joint distribution of the forecast path is an accurate representation of this uncertainty, but it is impossible to display in two-dimensional space. A first contribution of our paper is to introduce several methods to improve the communication of such joint uncertainty to the end-user based on Scheffé's (1953) S-method of simultaneous inference. In particular, Section 2 shows how to construct simultaneous confidence bands (which we will call Scheffé bands); conditional confidence bands for the uncertainty associated with individual forecast horizons; and fan charts based on the quantiles of the joint predictive density. These results parallel similar developments for impulse response functions in Jordà (2008).

Another commonly used method to evaluate the predictive properties of forecasts in a system of variables is to experiment and report forecasts where, for example, one variable follows an alternative path of interest. For example, monetary authorities often report two-year inflation and GDP forecasts under a variety of assumptions about interest rate paths (see, e.g. the Bank of England's Inflation Reports available from their website). The joint predictive density is the natural vehicle with which to provide formal support to these experiments and Section 3 discusses several simple metrics with which to measure the degree of coherence between the experiments and the historical experience, and the degree of exogeneity of a subspace of the system to these alternative experiments.

The small sample properties of the methods we propose are investigated via Monte Carlo simulations in Section 5. 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 6 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 7, which summarizes the main results of the paper and draws some conclusions.

## 2 Measuring Path Forecast Uncertainty

This section considers the problem of providing a measure of uncertainty around the forecast path of the  $j^{\text{th}}$  variable in the  $k$ -dimensional vector  $\mathbf{y}_t$ . An elementary ingredient of this problem requires the joint density of the system's forecasts 1 to  $H$  periods into the future. For clarity, we present our derivations with an approximate multivariate Gaussian joint distribution and then derive the theoretically optimal simultaneous confidence region from which a rectangular approximation can be obtained with Scheffé's (1953) S-method. The purpose of this rectangular approximation is so that uncertainty for the path forecast can be displayed in two-dimensional space. These approximations can be created for any quantile of the joint distribution to produce fan charts with approximately correct coverage at each probability level.

Section 4 and the appendix contain large sample Gaussian approximation results obtained for rather general data generating processes (DGPs) that could include infinite-dimensional and heterogeneous processes with various mixing and stability conditions. These derivations are provided to assist the reader with some basic results that have simple closed-form analytic expressions. However, we wish to highlight that the procedures we derive from Scheffé's

(1953) S-method apply, largely unchanged, when the covariance matrix of the path forecast is obtained with simulation techniques such as the bootstrap, or as a way to summarize the multivariate posterior density of the path forecast obtained with Bayesian simulation techniques instead. Investigation of the properties of these alternative computational methods is beyond the scope of this paper, however, we trust the reader will be able to adapt our procedures to suit his favorite approach.

## 2.1 Simultaneous Confidence Regions for Path Forecasts

Let  $\hat{\mathbf{y}}_T(h)$  be the forecast for  $\mathbf{y}_{T+h}$ , and let  $\hat{Y}_T(H)$  and  $Y_{T,H}$  be the forecast and actual paths for  $h = 1, \dots, H$ , so that

$$\hat{Y}_T(H) = \begin{bmatrix} \hat{\mathbf{y}}_T(1) \\ \vdots \\ \hat{\mathbf{y}}_T(H) \end{bmatrix}_{kH \times 1}; Y_{T,H} = \begin{bmatrix} \mathbf{y}_{T+1} \\ \vdots \\ \mathbf{y}_{T+H} \end{bmatrix}_{kH \times 1},$$

with, say, large-sample approximate distribution

$$\sqrt{T} \left( \hat{Y}_T(H) - Y_{T,H} \right) \xrightarrow{d} N(\mathbf{0}; \Xi_H). \quad (1)$$

An example of the specific analytic form of  $\Xi_H$  is provided in the section 4 when the DGP is a VAR and for forecasts generated by either the standard iterative method or by direct estimation (e.g. Jordà, 2005; Marcellino, Stock and Watson, 2006). Other relevant references for specific results on  $\Xi_H$  are Clements and Hendry (1993) and Lütkepohl (2005).

Define the selector matrix  $S_j \equiv (I_H \otimes \mathbf{e}_j)$  where  $\mathbf{e}_j$  is a  $1 \times k$  vector of zeros with a 1 in the  $j^{\text{th}}$  column. Then based on (1), the asymptotic distribution for the path forecast of the

$j^{\text{th}}$  variable in  $\mathbf{y}_t$  is readily seen to be

$$\sqrt{T} \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right) \xrightarrow{d} N(\mathbf{0}; \boldsymbol{\Xi}_{j,H}), \quad (2)$$

where  $\widehat{Y}_{j,T}(H) = S_j \widehat{Y}_T(H)$ ;  $Y_{j,T,H} = S_j Y_{T,H}$ ; and  $\boldsymbol{\Xi}_{j,H} = S_j \boldsymbol{\Xi}_H S_j'$ .

The conventional way of reporting forecast uncertainty consists of displaying two standard-error marginal bands constructed from the square roots of the diagonal entries of  $\boldsymbol{\Xi}_{j,H}$ . The confidence region described by these bands is therefore equivalent to testing a joint null hypothesis with the collection of t-statistics associated with the individual elements of the joint null. It is easy to see that such an approach ignores the simultaneous nature of the problem as well as any correlation that may exist among the forecasts across horizons, thus providing incorrect probability coverage.

In general, let  $g(\cdot) : \mathbb{R}^H \rightarrow \mathbb{R}^M$  be a first order differentiable function where  $H \geq M$  and with an  $H \times M$  invertible Jacobian denoted  $G(\cdot)$ . The decision problem associated with this transformation of the forecast path can be summarized by the null hypothesis  $H_0 : E[g(Y_{j,T,H})] = g_0$  for any  $j = 1, \dots, k$ ; sample  $T$ ; and forecast horizon  $H$  and where  $g_0$  is an  $M \times 1$  vector. Well-known principles based on the Gaussian approximation in expression (2), the Wald principle, and the Delta-method (or more generally, classical Minimum Distance, see, e.g. Ferguson, 1958), suggest that tests of this generic joint null hypothesis can be evaluated with the statistic

$$W_H = T \left( g(\widehat{Y}_{j,T}(H)) - g_0 \right)' \left( \widehat{G}'_{j,T,H} \boldsymbol{\Xi}_{j,H} \widehat{G}_{j,T,H} \right)^{-1} \left( g(\widehat{Y}_{j,T}(H)) - g_0 \right) \xrightarrow{d} \chi_H^2 \quad (3)$$

where  $\widehat{G}_{j,T,H}$  denotes the Jacobian evaluated at  $\widehat{Y}_{j,T}(H)$  and as usual,  $\boldsymbol{\Xi}_{j,H}$  can be replaced by its finite-sample estimate. From expression (3), a traditional null of joint significance can

be evaluated by setting  $g(\widehat{Y}_{j,T}(H)) \equiv \widehat{Y}_{j,T}(H)$ ; and  $g_0 \equiv \mathbf{0}_{H \times 1}$  so that a confidence region at an  $\alpha$ -significance level is represented by the values of  $Y_{j,T,H}$  that satisfy

$$\Pr [W_H \leq c_\alpha^2(H)] = 1 - \alpha$$

where  $c_\alpha^2(H)$  is the critical value of a random variable distributed  $\chi_H^2$  at a  $100(1 - \alpha)\%$  confidence level. This confidence region is a multi-dimensional ellipsoid that, in general, cannot be displayed graphically and thus makes communication of forecast uncertainty to the end-user of the forecast difficult. However, for  $H = 2$ , this region can be displayed in two-dimensional space as is done in figure 1.

The top panel of figure 1 displays the 95% confidence region associated with one- and two-period ahead forecasts from an AR(1) model with known autoregressive coefficient  $\rho = 0.75$  and error variance  $\sigma^2 = 1$ . Overlaid on this ellipse is the traditional two standard-error box. The figure makes clear why this box provides inappropriate probability coverage: it contains/excludes forecast paths with less/more than 5% chance of being observed. Further, the top panel of figure 2 illustrates that the correlation across horizons increases with the forecast horizon – the correlation between the two- and three-period ahead forecast errors is larger than that between the one- and two-period forecast errors. The larger the correlation between forecast errors, the larger the size distortion of two-standard-error rectangular intervals. Moreover, adding an MA component with a positive coefficient to the AR(1) model, further distorts the probability coverage, as the bottom panel of figure 2 shows. These two examples are of singular practical relevance since medium-horizon forecasts are of interest for policy making and a positive MA component is statistically significant for several macro-economic time series (see e.g., Marcellino et al., 2006).

## 2.2 Scheffé Confidence Bands for Forecast Paths

In order to reconcile the inherent difficulty of displaying multi-dimensional ellipsoids with the inadequate probability coverage provided by the more easily displayed marginal error bands, we propose constructing simultaneous rectangular regions with Scheffé's (1953) S-method of simultaneous inference (see also Lehmann and Romano, 2005) and use Holm's (1979) step-down procedure to obtain appropriate refinements. Briefly, the S-method exploits the Cauchy-Schwarz inequality to transform the Wald statistic in expression (3) from  $L_2$ -metric into  $L_1$ -metric and thus facilitate construction of a rectangular confidence interval.

We begin by noticing that the covariance matrix of  $\widehat{Y}_{j,T}(H)$  is positive-definite and symmetric and hence admits a Cholesky decomposition  $T^{-1}\Xi_{j,H} = PP'$ , where  $P$  is a lower triangular matrix. The passage of time provides a natural and unique ordering principle so that  $P$  is obtained unambiguously – the result of projecting the  $h^{th}$  forecast on to the path of the previous  $h - 1$  horizons. Notice then that

$$\begin{aligned}
\Pr \left[ T \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right)' \Xi_{j,H} \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right) \leq c_\alpha^2(H) \right] &= 1 - \alpha \\
\Pr \left[ \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right)' (PP')^{-1} \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right) \leq c_\alpha^2(H) \right] &= 1 - \alpha \\
\Pr \left[ \widehat{V}_{j,T}(H)' \widehat{V}_{j,T}(H) \leq c_\alpha^2(H) \right] &= 1 - \alpha \\
\Pr \left[ \sum_{h=1}^H \widehat{v}_{j,T}(h)^2 \leq c_\alpha^2(H) \right] &= 1 - \alpha \quad (4)
\end{aligned}$$

where  $\widehat{V}_{j,T}(H) = P^{-1}\widehat{Y}_{j,T}(H)$  and  $\widehat{v}_{j,T}(h) \stackrel{d}{\rightarrow} N(0, 1)$  are independent across  $h$ , by construction.

Consider now the problem of formulating the rectangular confidence region for the average

path forecast

$$\Pr \left[ \left| \sum_{h=1}^H \frac{\widehat{v}_{j,T}(h)}{h} \right| \leq \delta_\alpha \right] = 1 - \alpha.$$

A direct consequence of Bowden's (1970) lemma is that

$$\max \left\{ \frac{\left| \sum_{h=1}^H \frac{\widehat{v}_{j,T}(h)}{h} \right|}{\sqrt{\sum_{h=1}^H \frac{1}{h^2}}} : |h| < \infty \right\} = \sqrt{\sum_{h=1}^H \widehat{v}_{j,T}(h)^2}$$

which can be applied directly to the bottom line of expression (4) to obtain

$$\Pr \left[ \left| \sum_{h=1}^H \frac{\widehat{v}_{j,T}(h)}{h} \right| \leq \sqrt{\frac{c_\alpha^2(H)}{H}} \right] = 1 - \alpha, \quad (5)$$

which in turn implies that  $\delta_\alpha = \sqrt{\frac{c_\alpha^2(H)}{H}}$ . Expression (5) and  $\widehat{V}_{j,T}(H) = P^{-1}\widehat{Y}_{j,T}(H)$  suggest that a simultaneous confidence region for the path forecast  $\widehat{Y}_{j,T}(H)$  could then be constructed

as

$$\widehat{Y}_{j,T}(H) \pm P \sqrt{\frac{c_\alpha^2(H)}{H}} \mathbf{i}_H \quad (6)$$

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

Notice that the width of the bands constructed according to (6) would depend on the number of horizons the researcher chooses to display. This is a somewhat undesirable feature that we resolve by complementing Scheffé's S-method with a Holm's (1979) type step-down sequential procedure (see, e.g. Lehmann and Romano, 2005) where instead, we recommend that Scheffé bands (as we name our proposed confidence bands) be constructed as

$$\widehat{Y}_{j,T}(H) \pm P \left[ \sqrt{\frac{c_\alpha^2(h)}{h}} \right]_{h=1, \dots, H} \quad (7)$$

where the last term (in brackets) is an  $H \times 1$  vector whose  $h^{th}$  entry is  $\sqrt{\frac{c_\alpha^2(h)}{h}}$ . Hence, for  $H = 1$ , Scheffé and marginal bands are identical since  $\sqrt{\frac{c_\alpha^2(1)}{1}} = z_{\alpha/2}$ . This refinement on

Scheffé’s (1953) S-method provides for a more intuitive construction of confidence bands with better probability coverage rates, as our Monte Carlo experiments in Section 5 will show.

Geometric intuition further clarifies how the method works. In a traditional marginal band, its boundaries represent the largest shift away from the original forecasts such that the resulting region has a pre-specified probability coverage. Thus, the boundary of the marginal band comes from the appropriately variance-scaled critical values of the standard normal density of a region with symmetric  $100(1 - \alpha)\%$  coverage, specifically,  $\hat{y}_{j,T}(h) \pm z_{\alpha/2} \hat{\Sigma}_{j,(h,h)}^{1/2}$ . Instead, consider now a *simultaneous* variance-scaled shift in all the elements of the path forecast: What would the appropriate critical value be? It is easier to answer this question with the orthogonal coordinate system  $\hat{V}_{j,T}(H)$  first, to isolate the answer from the issue of correlation in the forecasts. From expression (4) and denoting this shift  $\delta_\alpha$ , then  $\delta_\alpha$  must meet the condition

$$\Pr [\delta_\alpha^2 + \frac{H}{H} + \delta_\alpha^2 = c_\alpha^2] = 1 - \alpha$$

which implies that  $\delta_\alpha = \sqrt{\frac{c_\alpha^2}{H}}$ . In two dimensions, figure 1 displays the diagonals intersecting the origin of both ellipses, for the original (top panel) and the orthogonalized (bottom panel) path forecasts. The slopes of these diagonals reflect the relative variance of each forecast, thus in the bottom panel the orthogonalization ensures the variances are the same and the diagonal is the 45 degree line representing  $\pm\delta_\alpha$  for all values of  $\alpha$ . The Cholesky factor  $P$  therefore provides the appropriate scaling for  $\delta_\alpha$  since it scales the orthogonal system by the individual variances of its elements and accounts for their correlation.

The literature has previously recognized the problem of simultaneity so one could consider constructing, for example, confidence intervals with Bonferroni’s procedure. This procedure

proposes the construction of a  $(1 - \frac{\alpha}{H})$  confidence interval for each  $y_{j,T}(h)$ ,  $h = 1, \dots, H$  so that the union of these individual confidence intervals generates a region that includes  $Y_{j,T,H}$  with *at least*  $(1 - \alpha)$  probability. Specifically, the Bonferroni confidence region (BCR) is

$$\widehat{Y}_{j,T}(H) \pm z_{\alpha/2H} \times \text{diag}(\Xi_{j,H})^{1/2},$$

where  $z_{\alpha/2H}$  denotes the critical value of a standard normal random variable at an  $\alpha/2H$  significance level and  $\text{diag}(\Xi_{j,H})^{1/2}$  is an  $H \times 1$  vector with the square roots of the diagonal entries of  $\Xi_{j,H}$ . Notice that  $z_{\alpha/2H} \rightarrow \infty$  as  $H \rightarrow \infty$  and therefore, the BCR can be significantly more conservative than our simultaneous confidence region, specially when the correlation between forecasts across horizons is low. The region tends to be overly conservative for low values of  $h$ , and not sufficiently inclusive for long-range forecasts, a feature we demonstrate in our simulation study of Section 5.

The orthogonalization in expression (4) suggests another measure of uncertainty complementary to Scheffé's bands. Notice that  $T^{-1}\Xi_{j,H} = PP' = QDQ'$  where  $Q$  is lower triangular with ones in the main diagonal and  $D$  is a diagonal matrix. Hence, expression (4) can be rewritten as

$$\begin{aligned} W_H &= \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right)' (QDQ')^{-1} \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right) \\ &= \widetilde{V}_{j,T}(H)' D^{-1} \widetilde{V}_{j,T}(H) \\ &= \sum_{h=1}^H \frac{\widetilde{v}_{j,T}(h)^2}{d_{hh}} = \sum_{h=1}^H t_{h|h-1, \dots, 1}^2 \rightarrow \chi_H^2 \end{aligned}$$

where  $\widetilde{V}_{j,T}(H) = Q^{-1} \left( \widehat{Y}_{j,T}(H) - Y_{j,T,H} \right)$  is the unstandardized version of  $\widehat{V}_{j,T}(H)$ ; and  $d_{hh}$  is the  $h^{\text{th}}$  diagonal entry of  $D$ , which is the variance of  $\widetilde{v}_{j,T}(h)$ . In other words, the Wald statistic  $W_H$  of the joint null on  $Y_{j,T,H}$  is equivalent to the sum of the squares of the *conditional*  $t$ -

statistics of the individual nulls of significance of the path forecast. Therefore, a  $100(1 - \alpha)\%$  confidence region for the  $h^{th}$  forecast that sterilizes the uncertainty from the preceding 1 to  $h - 1$  forecasts, can be easily constructed with the bands

$$\widehat{Y}_{j,T}(H) \pm z_{\alpha/2} \times \text{diag}(D)^{1/2}$$

where  $z_{\alpha/2}$  refers to the critical value of a standard normal random variable at an  $\alpha/2$  significance level and  $\text{diag}(D)$  is the  $H \times 1$  vector of diagonal terms of  $D$ .

### 3 Other Methods to Evaluate a Forecasting Exercise

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.

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  $k_1$  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  $\tilde{Y}_T^1(H)$  denote the alternative paths considered for the  $k_1$  variables and let  $\tilde{Y}_T^0(H)$  denote the paths of the  $k_0$  variables given  $\tilde{Y}_T^1(H)$ , that is

$$\tilde{Y}_T^0(H) = S_0 \hat{Y}_T(H) + S_0 \Xi_H S_1' (S_1 \Xi_H S_1')^{-1} \left( \tilde{Y}_T^1(H) - S_1 \hat{Y}_T(H) \right)$$

with covariance matrix

$$\Xi_H^0 = S_0 \Xi_H S_0' - S_0 \Xi_H S_1' (S_1 \Xi_H S_1')^{-1} S_1 \Xi_H S_0$$

In practice, the approximate nature of the predictive density of  $\hat{Y}_T(H)$  indicates that the accuracy of these calculations depends on several factors such as the value of  $H$  relative to

the estimation sample  $T$ , possible nonlinearities in the data, and the distance between  $\tilde{Y}_T^1(H)$  and  $S_1\hat{Y}_T(H)$ , among the more important factors.

The last observation suggests that it is useful to properly evaluate the distance between  $\tilde{Y}_T^1(H)$  and  $S_1\hat{Y}_T(H)$  and this can be easily accomplished with the Wald score

$$W_1 = T(S_1\hat{Y}_T(H) - \tilde{Y}_T^1(H))' (S_1\Xi_H S_1')^{-1} (S_1\hat{Y}_T(H) - \tilde{Y}_T^1(H))$$

This score will have an approximate chi-square distribution with  $k_1H$  degrees of freedom under the same assumptions that would allow one to obtain the approximate predictive density of  $\hat{Y}_T(H)$ . Thus, one minus the  $p$ -value of this score provides an easy to communicate distance metric in probability units between the predicted paths  $S_1\hat{Y}_T(H)$  and the alternative scenarios  $\tilde{Y}_T^1(H)$ . 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 basic linearity assumptions being true.

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 the Wald score

$$W_0 = T \left( S_0\hat{Y}_T(H) - \tilde{Y}_T^0(H) \right)' (S_0\Xi_H S_0')^{-1} \left( S_0\hat{Y}_T(H) - \tilde{Y}_T^0(H) \right)$$

Under the same conditions as before, this Wald 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. Consequently, it seems prudent for any forecasting report to include both  $W_0$  and  $W_1$  scores when experimenting with alternative scenarios.

## 4 Asymptotic Distribution of the Forecast Path

This section characterizes the asymptotic distribution of the path forecast under the assumption that the 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

$$\begin{aligned} \mathbf{y}_t &= \mathbf{m} + \sum_{j=1}^p A_j \mathbf{y}_{t-j} + \mathbf{w}_t \\ \mathbf{w}_t &= \sum_{j=p+1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \end{aligned} \tag{8}$$

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

$$\begin{aligned}\mathbf{y}_{t+h} &= \mathbf{m}_h + \sum_{j=0}^{p-1} A_j^h \mathbf{y}_{t-j} + \mathbf{v}_{t+h} \\ \mathbf{v}_{t+h} &= \sum_{j=p}^{\infty} A_j^h \mathbf{y}_{t-j} + \mathbf{u}_{t+h} + \sum_{j=1}^{h-1} \Phi_j \mathbf{u}_{t+h-j} \quad \text{for } h = 1, \dots, H\end{aligned}\tag{9}$$

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  $\Gamma(j) \equiv E(\mathbf{y}_t \mathbf{y}'_{t+j})$  with  $\Gamma(-j) = \Gamma(j)'$  and define:  $X_{t,p} = (\mathbf{1}, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$ ;  $\widehat{\Gamma}_{1-p,h} = (T-p-h)^{-1} \sum_{t=p}^T X_{t,p} \mathbf{y}'_{t+h}$ ; and  $\widehat{\Gamma}_p = (T-p-h)^{-1} \sum_{t=p}^T X_{t,p} X'_{t,p}$ . Then, the least-squares estimate of the VAR( $p$ ) in expression (8) is given by the formula

$$\widehat{A}(p) = \left( \widehat{\mathbf{m}}, \widehat{A}_1, \dots, \widehat{A}_p \right) = \widehat{\Gamma}'_{1-p,0} \widehat{\Gamma}_p^{-1},\tag{10}$$

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

$$\widehat{A}(p, h) = \left( \widehat{\mathbf{m}}_h, \widehat{A}_1^h, \dots, \widehat{A}_p^h \right) = \widehat{\Gamma}'_{1-p,h} \widehat{\Gamma}_p^{-1}; \quad h = 1, \dots, H.\tag{11}$$

Then, corollary 1 in the appendix shows that the parameter estimates in expressions (10) and (11) are consistent and asymptotically Gaussian.

Next, denote with  $\mathbf{y}_T(h)$  the forecast of the vector  $\mathbf{y}_{T+h}$  assuming the coefficients of the infinite order process (16) were known, that is

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

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

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

where  $\widehat{\mathbf{y}}_T(h-j) = \mathbf{y}_{T+h-j}$  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  $h = 1, \dots, H$  by stacking each of the quantities  $\widehat{\mathbf{y}}_T(h)$ ,  $\mathbf{y}_T(h)$ , and  $\mathbf{y}_{T+h}$  as follows

$$\widehat{\mathbf{Y}}_T(H)_{kH \times 1} = \begin{bmatrix} \widehat{\mathbf{y}}_T(1) \\ \vdots \\ \widehat{\mathbf{y}}_T(H) \end{bmatrix}; Y_T(H)_{kH \times 1} = \begin{bmatrix} \mathbf{y}_T(1) \\ \vdots \\ \mathbf{y}_T(H) \end{bmatrix}; Y_{T,H} = \begin{bmatrix} \mathbf{y}_{T+1} \\ \vdots \\ \mathbf{y}_{T+H} \end{bmatrix}.$$

Our interest is in finding the asymptotic distribution for  $\widehat{\mathbf{Y}}_T(H) - Y_{T,H} = [\widehat{\mathbf{Y}}_T(H) - Y_T(H)] + [Y_T(H) - Y_{T,H}]$ .

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

$$\Omega_H_{kH \times kH} \equiv E[(Y_T(H) - Y_{T,H})(Y_T(H) - Y_{T,H})'] = \Phi(I_H \otimes \Sigma_u) \Phi'. \quad (12)$$

where

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

Furthermore, since the parameter estimates are based on a sample of size  $T$  and hence  $\mathbf{u}_t$  for  $t = p + h, \dots, T$  while the term  $Y_T(H) - Y_{T,H}$  only involves  $\mathbf{u}_t$  for  $T + 1, \dots, T + H$ , then it should be clear that to derive the asymptotic distribution of  $[\widehat{Y}_T(H) - Y_T(H)]$ , 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 (12) but the covariance between these terms will be zero.

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

$$\begin{aligned} & \sqrt{\frac{T-p-H}{p}} \text{vec}(\widehat{Y}_T(H) - Y_T(H)) \xrightarrow{d} N(0, \Psi_H) \quad (13) \\ \Psi_H & \equiv \frac{\partial \text{vec}(\widehat{Y}_T(H))}{\partial \text{vec}(\widehat{\mathbf{A}})} \Sigma_A \frac{\partial \text{vec}(\widehat{Y}_T(H))}{\partial \text{vec}(\widehat{\mathbf{A}})'} \end{aligned}$$

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

$$\widehat{\mathbf{A}} = \begin{bmatrix} \widehat{A}(p, 1) \\ \vdots \\ \widehat{A}(p, H) \end{bmatrix}. \quad (14)$$

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

$$\begin{aligned} \Xi_H &= \sqrt{\frac{T-p-H}{p}} \text{vec} \left( \widehat{Y}_T(H) - Y_{T,H} \right) \xrightarrow{d} N(\mathbf{0}; \Xi_H) \\ \Xi_H &= \left\{ \frac{p}{T-p-H} \Omega_H + \Psi_H \right\} \\ \Omega_H &= \Phi(I_H \otimes \Sigma_u) \Phi' \end{aligned}$$

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

## 5 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 data generating process (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.

Stock and Watson's (2001) well-cited review article on vector autoregressions (VARs) seems like an appropriate choice then. 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 3). 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). The smaller sample of 100 observations is approximately half of the available estimation sample and given the number of parameters to be estimated, a good representation of a relatively small sample with few degrees of freedom. The larger sample of 400 observations is approximately twice the size of the sample available for estimation and hence, considerably closer to the theoretical asymptotic ideal. We constructed 1,000 Monte Carlo replications of each sample size in this fashion.

In order to be as faithful as possible in replicating a typical practical environment, 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).<sup>1</sup> Next, each replication involves estimating a VAR and direct forecasts by least-squares and hence generating appropriate forecast error variances for forecast paths of varying length (specifically for  $H = 1, 4, 8$ , and

---

<sup>1</sup> Hurvich and Tsai (1993) show that  $AIC_c$  has better small sample properties than  $AIC$ ,  $SIC$  and other common information criteria.

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 shows. 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 correspond approximately 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 two metrics. The first metric looks at the proportion of paths that fall completely within the bands. 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.” This type of metric controls the family-wise error rate (FWER) as defined in, for example, Lehmann and Romano (2005).

The second 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 of a 12-period forecast path that had one element outside the bands, such a path would be counted

as “covered” as long as its Wald score was lower than that for the bands. In other words, this metric controls the size of the joint test directly rather than the FWER. Such metric is related to control of the false discovery rate as defined in Benjamini and Hochberg (1995).

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 3 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.

Before commenting on the results in the tables, it is useful to comment on figure 3 first. For one-period ahead forecasts, all three bands attain the same value. However, as the forecast horizon increases, Bonferroni and Scheffé bands fan out wider than marginal bands, the former more conservatively than the latter although after three periods Scheffé bands fan wider than Bonferroni bands.

From Tables 1 and 2, for one-period ahead forecasts (where the three methods coincide), coverage rates 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 simultaneous 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-seventies although *Wald* coverage can sometimes be in the low 20's%). Bonferroni's procedure generates bands that are generally more conservative in terms of *FWER* control 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 simultaneous Wald coverage, with empirical levels around 40% for 68% nominal coverage and  $H = 12$ .

Scheffé bands are designed from a rectangular approximation to the Wald statistic and hence provide the most accurate match between empirical and nominal coverage rates, at all horizons, and at all confidence levels; yet the bands have small distortions in *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 ( $\rho$ ) 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 (which are only approximately correct when this correlation is zero). 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  $\rho$ .

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 in some cases but may provide poor coverage in terms of simultaneous Wald scores. In contrast, Scheffé bands manage to strike a nice balance between *FWER* and simultaneous Wald control and their coverage is relatively robust to all sorts of coverage levels and forecast horizon choices. In addition, they seem specially appropriate if one is interested in constructing fan charts that accurately represent all depicted nominal coverage levels since either marginal or Bonferroni bands can be quite a ways off when different nominal levels are considered.

## 6 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 4 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 but, over the two-year period examined, they tend to be relatively close to the traditional 95% marginal bands (specially for U.S. GDP). 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 the alternative scenario that consists in 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 of 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 in that it is not based on any information coming from the data. 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 it.

The results of this experiment are reported in figure 5. 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 of 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.

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.

## **7 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 increasingly distant horizons – a path forecast – this paper shows that error bands should be derived from the joint predictive density. 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 values a variable can take over time, given a finite sample of data and a statistical model. This solution is 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 by applying a refinement based on Holm's (1973) step-down testing procedure.

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 false discovery rate based on the simultaneous Wald score (the probability that, jointly, the elements of the path are "close" in probability distance units even if one or more elements of the path are not strictly within the bands). 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, sometimes 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 its internal consistency. The approximate joint predictive density can be quite useful in this respect, even when forecasts are produced from a variety of different methods. Thus, 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 model over regions where the sample provides no training, we provide a simple Wald score that measures the probability distance to their conditional mean path. In addition, the Wald score can be used to measure the sensitivity of each variable in the system to the proposed scenarios, thus providing another metric to assess the results of the experiments with alternative scenarios.

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 alternative measures to the commonly used MSFE that integrate the correlation patterns in a path forecast, as well as tests of predictive ability along the lines of Giacomini and White (2006) based on multivariate Wald scores.

## 8 Appendix

We begin by stating our assumptions on the DGP described in section 4 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

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Phi_j' \mathbf{u}_{t-j}, \quad (15)$$

where the moving-average coefficient matrices  $\Phi_j$  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  $\|\Phi_j\|^2 = \text{tr}(\Phi_j' \Phi_j)$  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 (15) can also be written as an infinite VAR process (see, e.g. Anderson, 1994),

$$\mathbf{y}_t = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \quad (16)$$

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 (16). 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. We refer the reader to Gonçalves and Kilian (2007) and references therein for a discussion of these issues. The most significant implication of allowing for these alternative, more flexible assumptions is that a robust covariance estimator along the lines of White (1980) is advised. For now, 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  $T$  such that

$$\frac{p^3}{T} \rightarrow 0 \text{ as } T, p \rightarrow \infty.$$

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

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

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 1** *Given assumptions 1 and 2, the VAR( $p$ ) and  $p^{\text{th}}$  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)  $\sqrt{\frac{T-p-h}{p}} \text{vec} \left( \widehat{A}(p) - A(p) \right) \xrightarrow{d} N(0, \Sigma_a)$  where  $\Sigma_a = \Gamma_p^{-1} \otimes \Sigma_u$

(c)  $\sqrt{\frac{T-p-h}{p}} \text{vec} \left( \widehat{A}(p, h) - A(p, h) \right) \xrightarrow{d} N(0, \Sigma_\alpha)$  where  $\Sigma_\alpha = \Gamma_p^{-1} \otimes \Omega_h$  and  $\Omega_h = \Phi (I_h \otimes \Sigma_u) \Phi'$

where

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

- (d) Let  $\hat{\mathbf{u}}(p)_t \equiv y_t - \hat{\mathbf{m}} - \sum_{j=1}^p \hat{A}_j y_{t-j}$  so that  $\hat{\Sigma}_u(p) = (T-p)^{-1} \sum_{t=1}^T \hat{\mathbf{u}}(p)_t \hat{\mathbf{u}}(p)_t'$  then  $\sqrt{T}(\hat{\Sigma}_u(p) - \Sigma_u) \rightarrow N(0, \Omega_\Sigma)$  where  $\Omega_\Sigma$  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  $p^2/T \rightarrow 0, T, p \rightarrow \infty$  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  $\hat{Y}_T(H) = \hat{\mathbf{A}}X_{T,p}$  and hence

$$\frac{\partial \text{vec}(\hat{Y}_T(H))}{\partial \text{vec}(\hat{\mathbf{A}})} = \frac{\partial \text{vec}(\hat{\mathbf{A}}X_{t,p})}{\partial \text{vec}(\hat{\mathbf{A}})} = \begin{matrix} (X'_{T,p} \otimes I_{kH}), \\ kH \times k^2 H p + kH \end{matrix} \quad (17)$$

which combined with corollary 1(c) results in

$$\sqrt{\frac{T-p-H}{p}} \left( \text{vec} \left( \widehat{\mathbf{A}} - \mathbf{A} \right) \right) \xrightarrow{d} N(0, \Sigma_A) \quad (18)$$

$$\begin{matrix} \Sigma_A \\ k^2 H p + k H \times k^2 H p + k H \end{matrix} = \begin{matrix} \Gamma_p^{-1} \otimes \Omega_H; \\ k H \times k H \end{matrix} \Omega_H; \quad \Omega_H = \mathbf{\Phi} (I_H \otimes \Sigma_u) \mathbf{\Phi}'$$

Putting together expressions (12), (13), (17) and (18), we arrive at the following corollary.

**Corollary 2** *Under assumptions 1 and 2 and expressions (13), (12), (17) and (18), the asymptotic distribution of the forecast path generated with the local projections approach described in assumption 1 is*

$$\sqrt{\frac{T-p-H}{p}} \text{vec} \left( \widehat{Y}_T(H) - Y_{T,H} \right) \xrightarrow{d} N(\mathbf{0}; \Xi_H) \quad (19)$$

$$\Xi_H = \left\{ \frac{p}{T-p-H} \Omega_H + \Psi_H \right\}$$

$$\Omega_H = \mathbf{\Phi} (I_H \otimes \Sigma_u) \mathbf{\Phi}'$$

$$\Psi_H = (X'_{T,p} \otimes I_{kH}) [\Gamma_p^{-1} \otimes \Omega_H] (X_{T,p} \otimes I_{kH})$$

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 (8). For this purpose, we find it easier to work with each element of the vector  $\widehat{Y}_T(H)$  individually, so that we begin by examining the derivation of

$$\sqrt{\frac{T-p-H}{p}} \text{vec} \left( \widehat{\mathbf{y}}_T(h) - \mathbf{y}_T(h) \right) \xrightarrow{d} N(\mathbf{0}; \Psi_{h,h})$$

$$\Psi_{h,h} = \frac{\partial \text{vec} \left( \widehat{\mathbf{y}}_T(h) \right)}{\partial \text{vec} \left( \widehat{A}(p) \right)} \Sigma_a \frac{\partial \text{vec} \left( \widehat{\mathbf{y}}_T(h) \right)}{\partial \text{vec} \left( \widehat{A}(p) \right)}$$

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

$$\Psi_{i,j} = \frac{\partial \text{vec}(\widehat{\mathbf{y}}_T(i))}{\partial \text{vec}(\widehat{A}(p))} \Sigma_a \frac{\partial \text{vec}(\widehat{\mathbf{y}}_T(j))}{\partial \text{vec}(\widehat{A}(p))}$$

which is all we need to construct all the elements in the asymptotic covariance matrix of  $\widehat{Y}_T(H)$ , namely  $\Psi_H$ . An expression for  $\widehat{\mathbf{y}}_T(h)$  generated from the VAR( $p$ ) in expression (8) can be obtained as

$$\widehat{\mathbf{y}}_T(h) = SB^h X_{T,p}$$

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

$$B_{kp+1 \times kp+1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \mathbf{m} & A_1 & A_2 & \dots & A_{p-1} & A_p \\ 0 & I_k & 0 & \dots & 0 & 0 \\ 0 & 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I_k & 0 \end{pmatrix},$$

$$S_{k \times kp+1} = (\mathbf{0}_{k \times 1}, I_k, \mathbf{0}_{k \times k}, \dots, \mathbf{0}_{k \times k}).$$

Therefore, notice that

$$\frac{\partial \text{vec}(\widehat{\mathbf{y}}_T(h))}{\partial \text{vec}(\widehat{A}(p))} = \frac{\partial \text{vec}(SB^h X_{t,p})}{\partial \text{vec}(\widehat{A}(p))} = \sum_{i=0}^{h-1} X'_{T,p}(B')^{h-1-i} \otimes \Pi_i, \quad \Pi_i = SB^i S'.$$

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

**Corollary 3** Under assumptions 1 and 2, the asymptotic distribution of the forecast path  $\widehat{Y}_T(H)$  generated from the VAR( $p$ ) in expression (8) is given by

$$\begin{aligned} & \sqrt{\frac{T-p-H}{p}} \text{vec} \left( \widehat{Y}_T(H) - Y_{T,H} \right) \xrightarrow{d} N(\mathbf{0}; \Xi_H) \\ \Xi_H &= \left\{ \frac{p}{T-p-H} \Omega_H + \Psi_H \right\} \\ \Omega_H &= \Phi(I_H \otimes \Sigma_u) \Phi' \\ \Psi_{i,j} &= \frac{p}{T-p-H} \sum_{k=0}^{i-1} \sum_{s=0}^{j-1} E(X'_{T,p} (B')^{i-1-k} \Gamma_p^{-1} B^{j-1-s} X_{T,p}) \otimes \Pi_k \Sigma_u \Pi'_s \\ &= \frac{p}{T-p-H} \sum_{k=0}^{i-1} \sum_{s=0}^{j-1} \text{tr}((B')^{i-1-k} \Gamma_p^{-1} B^{j-1-s} \Gamma_p) \Pi_k \Sigma_u \Pi'_s \end{aligned} \quad (20)$$

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.
- Bank of England, “Inflation Report,” available quarterly since 1997 at <http://www.bankofengland.co.uk/publications/inflationreport/index.htm>.
- Benjamini, Yoav and Yosef Hochberg (1995) “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society: Series B*, 57(1): 289-300.
- 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.
- Ferguson, Thomas S. (1958) “A Method of Generating Best Asymptotically Normal Estimates with Application to Estimation of Bacterial Densities,” *Annals of Mathematical Statistics*, 29:1046-1062.
- 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.
- Gonçalves, Silvia and Lutz Kilian (2007) “Asymptotic and Bootstrap Inference for  $AR(\infty)$  Processes with Conditional Heteroskedasticity,” *Econometric Reviews*, 26(6): 609-641.
- 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.
- 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.
- 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.
- 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.
- Scheffé, Henry (1953) A Method for Judging All Contrasts in the Analysis of Variance,” *Biometrika*, 40: 87-104.

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

White, Halbert (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4): 817-838.

**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.
<b>T=100</b>	<b>P</b>	67.5	67.5	67.5	67.5	67.5	67.5	93.8	93.8	93.8	93.8	93.8	93.8
	<b>UN</b>	69.5	69.5	69.5	69.5	69.5	69.5	95.8	95.8	95.8	95.8	95.8	95.8
	<b>FF</b>	68.4	68.4	68.4	68.4	68.4	68.4	94.6	94.6	94.6	94.6	94.6	94.6
<b>T=400</b>	<b>P</b>	66.9	66.9	66.9	66.9	66.9	66.9	93.6	93.6	93.6	93.6	93.6	93.6
	<b>UN</b>	69.7	69.7	69.7	69.7	69.7	69.7	96.0	96.0	96.0	96.0	96.0	96.0
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	32.8	78.7	58.4	20.5	70.2	67.1	85.6	95.9	92.8	80.3	95.0	95.4
	<b>UN</b>	43.6	82.4	63.8	15.3	60.8	67.5	88.1	96.5	93.8	72.0	90.8	94.1
	<b>FF</b>	37.0	79.7	61.0	15.8	65.3	68.0	86.4	95.8	93.7	76.1	92.5	94.6
<b>T=400</b>	<b>P</b>	29.8	76.7	56.7	21.5	73.0	67.0	83.9	95.5	92.4	82.9	96.8	96.6
	<b>UN</b>	43.8	83.2	64.2	15.4	62.2	68.6	88.7	97.2	94.2	73.0	91.9	95.1
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	16.7	81.8	56.2	1.5	50.4	63.1	78.7	95.8	91.8	43.9	86.6	93.7
	<b>UN</b>	27.9	84.9	63.2	2.1	58.6	65.7	82.0	96.6	93.7	52.0	91.3	95.6
	<b>FF</b>	24.4	84.0	63.0	1.9	50.8	66.1	80.9	96.3	93.6	44.0	87.4	95.4
<b>T=400</b>	<b>P</b>	13.5	79.9	54.5	1.4	52.5	65.4	77.0	95.7	91.8	45.7	90.6	96.4
	<b>UN</b>	27.6	85.8	63.8	1.9	60.2	67.8	82.8	97.2	94.2	53.0	93.4	96.9
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	12.4	84.2	57.2	0.2	37.3	61.7	74.2	96.2	91.9	21.5	77.5	92.8
	<b>UN</b>	19.1	85.7	62.1	0.4	66.3	66.1	77.0	96.4	92.3	46.6	93.7	96.1
	<b>FF</b>	15.7	85.0	62.4	0.1	39.9	64.9	76.2	96.4	93.1	22.2	81.3	95.4
<b>T=400</b>	<b>P</b>	9.1	83.3	55.5	0.1	37.0	65.8	72.5	96.6	92.2	19.9	81.4	96.0
	<b>UN</b>	18.2	86.4	63.1	0.2	71.2	69.4	77.2	97.0	93.3	49.7	96.8	97.7
	<b>FF</b>	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<sub>c</sub>. 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.
<b>T=100</b>	<b>P</b>	67.5	67.5	67.5	67.5	67.5	67.5	93.8	93.8	93.8	93.8	93.8	93.8
	<b>UN</b>	69.5	69.5	69.5	69.5	69.5	69.5	95.8	95.8	95.8	95.8	95.8	95.8
	<b>FF</b>	68.4	68.4	68.4	68.4	68.4	68.4	94.6	94.6	94.6	94.6	94.6	94.6
<b>T=400</b>	<b>P</b>	66.9	66.9	66.9	66.9	66.9	66.9	93.6	93.6	93.6	93.6	93.6	93.6
	<b>UN</b>	69.8	69.8	69.8	69.8	69.8	69.8	96.0	96.0	96.0	96.0	96.0	96.0
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	30.5	76.4	55.9	24.9	76.3	68.0	83.8	95.0	91.7	85.4	97.1	96.4
	<b>UN</b>	41.7	80.7	63.1	16.7	63.6	68.7	86.9	95.7	93.6	74.5	92.0	94.8
	<b>FF</b>	34.6	77.2	59.4	18.0	69.2	69.0	84.4	94.7	93.0	79.3	94.2	95.5
<b>T=400</b>	<b>P</b>	29.5	76.3	56.2	22.1	74.0	67.1	83.6	95.3	92.2	83.8	97.1	96.7
	<b>UN</b>	43.3	82.8	64.1	15.6	62.7	68.8	88.4	97.0	94.2	73.6	92.1	95.3
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	13.5	78.6	52.9	3.1	65.1	68.2	75.3	94.4	90.0	58.6	93.9	96.0
	<b>UN</b>	25.1	81.2	61.2	3.9	69.7	68.9	78.7	95.2	92.7	63.4	95.1	96.6
	<b>FF</b>	21.1	80.6	60.7	3.3	64.0	70.6	77.3	94.7	92.5	57.4	93.5	96.9
<b>T=400</b>	<b>P</b>	12.8	79.2	53.9	1.6	56.2	66.6	76.3	95.3	91.6	49.2	92.3	96.7
	<b>UN</b>	26.8	84.9	63.5	2.3	62.9	68.5	81.9	96.9	94.1	55.9	94.3	97.2
	<b>FF</b>	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.
<b>T=100</b>	<b>P</b>	9.3	80.4	52.8	0.7	58.6	69.6	69.3	94.8	89.4	39.4	90.7	95.9
	<b>UN</b>	16.6	82.2	55.7	2.3	83.3	71.2	72.9	94.8	87.3	68.7	97.7	97.1
	<b>FF</b>	12.8	81.1	58.7	0.5	65.0	72.8	71.4	94.6	91.3	45.0	93.7	97.5
<b>T=400</b>	<b>P</b>	8.3	82.2	54.7	0.2	42.5	67.7	71.0	96.2	91.9	24.0	85.5	96.6
	<b>UN</b>	17.3	85.3	62.0	0.3	76.2	70.6	75.9	96.6	92.3	56.0	97.7	97.9
	<b>FF</b>	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 AICc. 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.

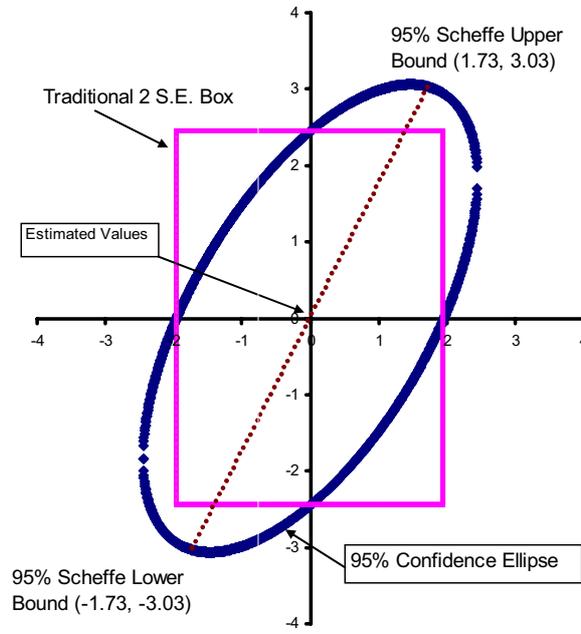
**Table 3. Coverage Rates of Marginal, Bonferroni, and Scheffé Bands in Simple AR(1) Model**

<b>Nominal Coverage Level: 68%</b>										
<b>Horizon = 1</b>						<b>Horizon = 4</b>				
<b>FWER</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	68	67.5	67.8	68.5	68.2	27.3	28.3	28.1	34.2	33.2
<b>Bonf.</b>	68	67.5	67.8	68.5	68.2	73.8	77.3	75	79	77.8
<b>Schef.</b>	68	67.5	67.8	68.5	68.2	53	54.3	55.9	62.3	60.6
<b>WALD</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	68	67.5	67.8	68.5	68.2	26.3	24.7	20.5	20.6	15.5
<b>Bonf.</b>	68	67.5	67.8	68.5	68.2	83	79	72.6	69.1	61.8
<b>Schef.</b>	68	67.5	67.8	68.5	68.2	67.4	69.2	65.7	68.2	66.6
<b>Nominal Coverage Level: 68%</b>										
<b>Horizon = 8</b>						<b>Horizon = 12</b>				
<b>FWER</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	6.8	8.9	9.9	15.4	22.4	1.8	3.6	3.8	6.6	11.6
<b>Bonf.</b>	76.8	76	76.1	79.5	82.8	74.6	75.2	79.4	80.5	85.3
<b>Schef.</b>	42.3	49.6	52.3	59.8	59.3	33.3	42.2	53.4	59.5	59.5
<b>WALD</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	8.2	6.2	3.1	2.2	1	3.3	1.7	0.4	0.3	0.1
<b>Bonf.</b>	93	85.4	71.9	62.7	50.7	98	90.6	78	56.1	37.1
<b>Schef.</b>	69.4	68.4	65.9	68.3	69.2	69.2	67.2	69.6	68.5	69.9
<b>Nominal Coverage Level: 95%</b>										
<b>Horizon = 1</b>						<b>Horizon = 4</b>				
<b>FWER</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	94.7	94.8	95.3	96.2	94.4	82.6	85.9	84.1	86.7	84.2
<b>Bonf.</b>	94.7	94.8	95.3	96.2	94.4	95.2	95.5	95.9	96.6	95.5
<b>Schef.</b>	94.7	94.8	95.3	96.2	94.4	90.4	93.3	93.8	95.2	92.9
<b>WALD</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	94.7	94.8	95.3	96.2	94.4	91	87.7	83.3	80.2	74.9
<b>Bonf.</b>	94.7	94.8	95.3	96.2	94.4	99.3	97.9	96.8	96.5	93.1
<b>Schef.</b>	94.7	94.8	95.3	96.2	94.4	98.1	97.2	97	97.5	95.3
<b>Nominal Coverage Level: 95%</b>										
<b>Horizon = 8</b>						<b>Horizon = 12</b>				
<b>FWER</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	72.6	70.5	71.2	75	79.4	57.8	59.1	64	69.9	75.7
<b>Bonf.</b>	95.7	95.3	96	92.4	97.6	95.4	95.6	95.3	96.8	97.2
<b>Schef.</b>	87.7	91.7	92.2	95.9	95.2	80.2	87.5	92.2	92.3	93.9
<b>WALD</b>	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
<b>Marg.</b>	89.1	80.3	65.4	56.3	43.7	89.1	72.9	55.8	34.5	18.7
<b>Bonf.</b>	99.9	99.6	97.8	92.4	90	100	99.8	98.8	93.2	81
<b>Schef.</b>	98.6	98.4	97.8	95.9	97.2	99.6	98.4	98.3	97.3	97.3

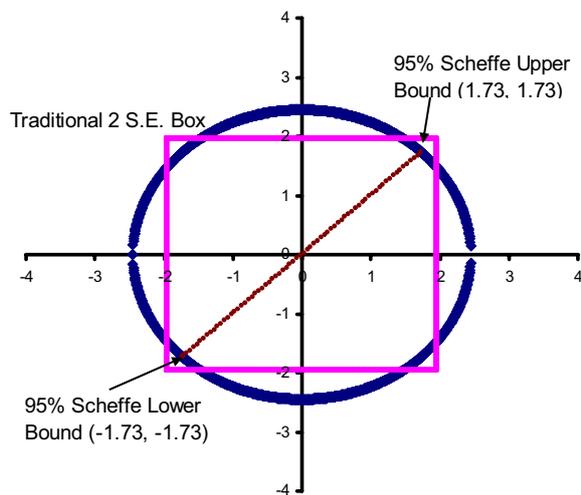
**Notes:** Theoretical values of the forecast error variance (excluding parameter estimation uncertainty) are used to construct three sets of bands (marginal, Bonferroni, and Scheffé). Then 1,000 Monte Carlo replications from the DGP are generated. FWER stands for “family-wise error rate” and computes 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% Scheffe bounds for AR(1) Forecast Path over Two Horizons**

**Panel 1 – Standard confidence bands, confidence ellipse, and Scheffé Bounds**



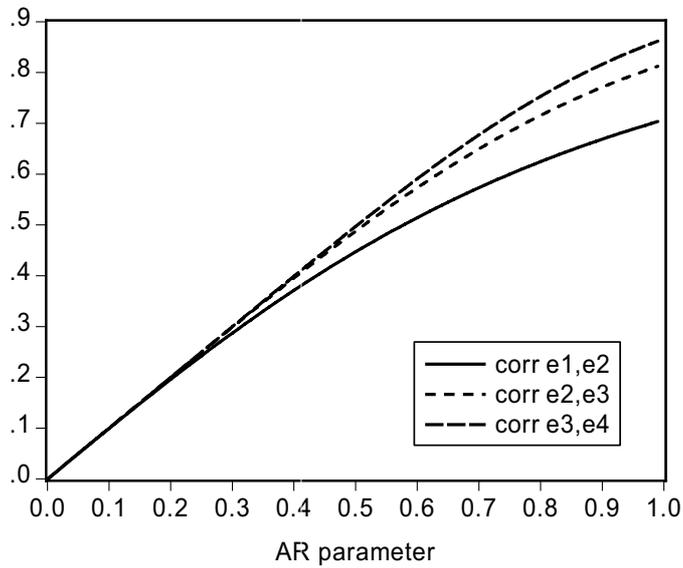
**Panel 2 – 95% Confidence Circle for Orthogonalized Forecast Path**



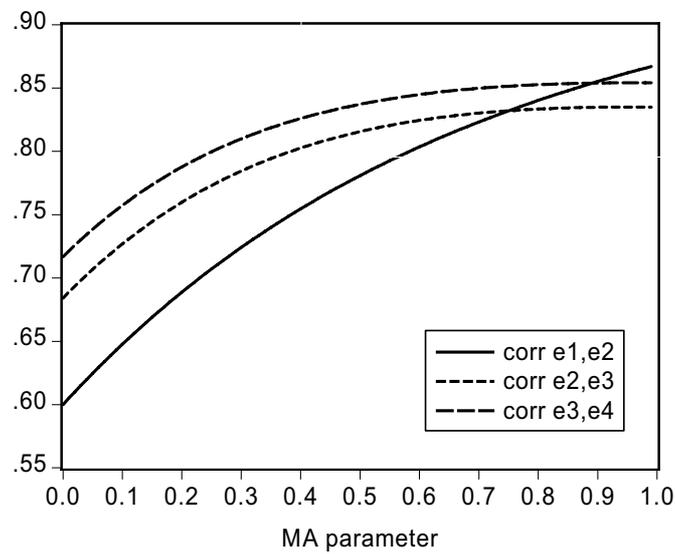
**Notes: AR Coefficient = 0.75, Error Variance = 1**

**Figure 2 – Correlation pairs between 1,2,3,and 4-step ahead forecast errors, AR(1) and ARMA(1,1)**

**Panel 1 – AR model**

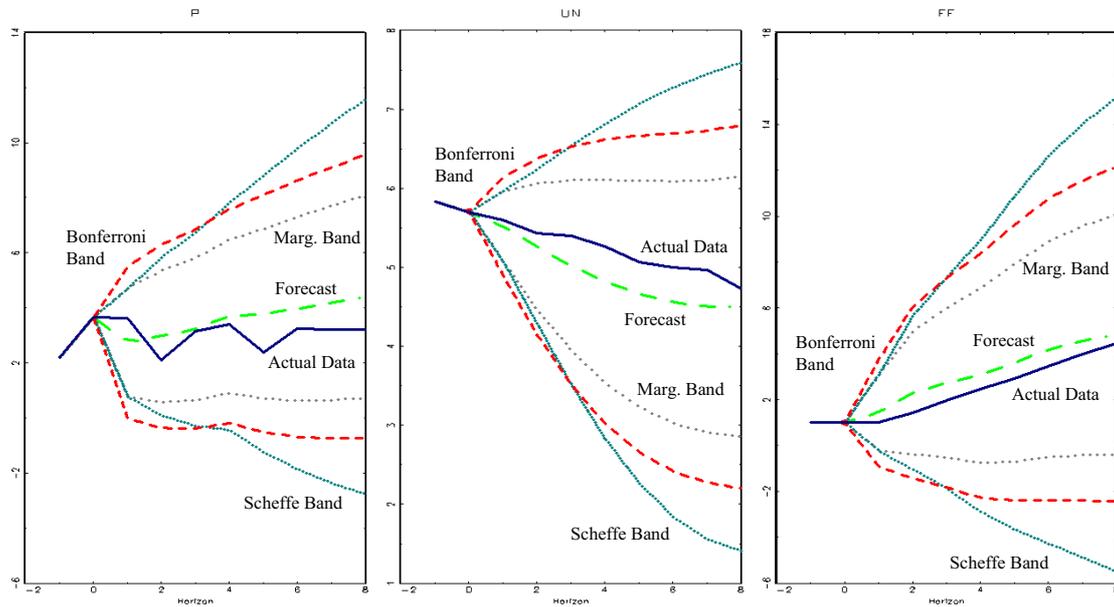


**Panel 2 – ARMA(1,1) model, AR parameter = 0.75**



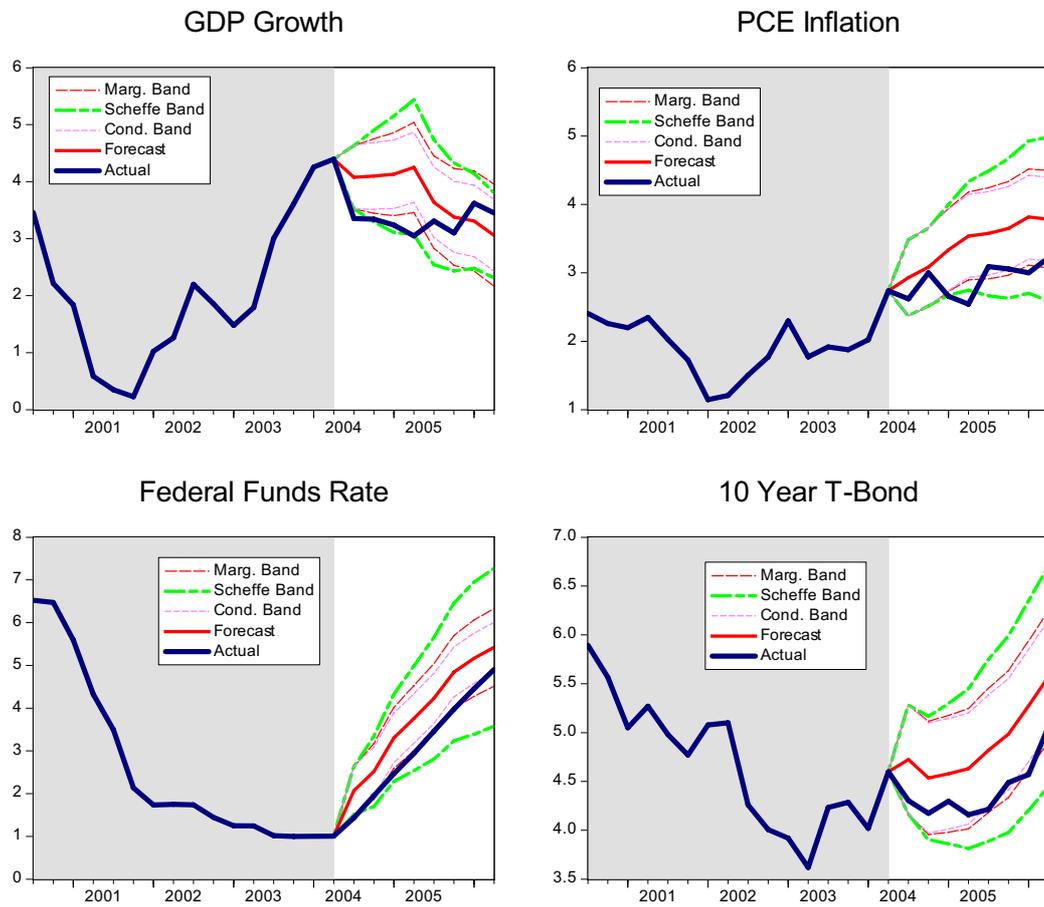
**Notes:** Panel 1 displays the correlation between forecast error pairs in an AR(1) model as a function of the AR parameter. Panel 2 displays the correlation between forecast error pairs of an ARMA(1,1) model as a function of the MA parameter with the AR parameter fixed at 0.75.

**Figure 3. Stock and Watson (2001) Out-of-Sample Forecasts, 8-periods Ahead**



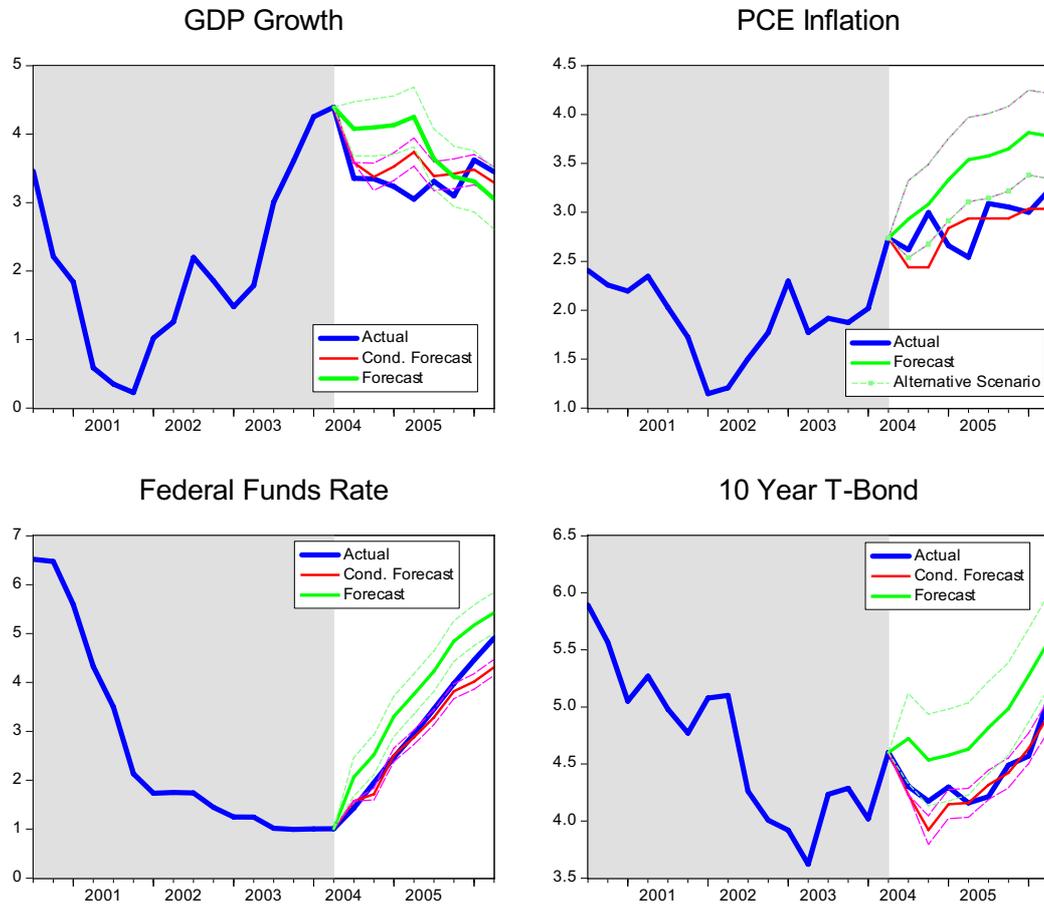
**Notes:** Out-of-sample forecasts for the Stock and Watson (2001) VAR. Estimation sample 1960:I-2004:IV. Prediction sample 2005:I-2007:I. Predictions based on VAR(4). P stands for inflation (measured by the chain-weighted GDP price index), UN stands for unemployment (measured by the civilian unemployment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

**Figure 4. 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 5. 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