October 2007

Path Forecast Evaluation*

Abstract

A forecast path refers to the vector of forecasts over the next 1 to h periods into the future. These forecasts are correlated across horizons so that to properly understand the uncertainty associated with the forecast path, one requires the joint predictive density of the path rather than the collection of marginal predictive densities for each horizon. This paper derives the joint predictive density for forecasts generated with VARs or with direct forecast methods from possibly infinite order data generating processes. Given this density, we use Scheffé's S-method to construct simultaneous confidence regions for the forecast path and show how to construct path forecasts conditional on assumed paths for a subset of the system's variables, along with their conditional predictive density and a test on the assumed path's likelihood. We then introduce the mean square forecast path metric to compare the predictive ability between competing models, and appropriately modify Diebold-Mariano-West and Giacomini-White predictive ability tests. Finally, as empirical illustrations of the theoretical concepts, we evaluate path forecasts from a system of U.S. macroeconomic variables, and examine the role of monetary aggregates in forecasting inflation.

JEL Classification Codes: C32, C52, C53

Keywords: path forecast, local projections, vector autoregression, simultaneous confidence region.

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, IGIER and CEPR
Via Salasco 5
20136 Milan, Italy
e-mail: massimiliano.marcellino@uni-bocconi.it

---

*We are grateful to 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 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.

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 density of the *forecast path.* There are many questions of interest that can be answered based on 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) 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 density, not the collection of marginal densities alone. The joint distribution and the covariance matrix for the forecast path thus play a prominent role in our discussion, and we begin by deriving appropriate asymptotic results for data generating processes (DGP) of infinite order in Section 2. Vector autoregressions are a natural multivariate method of producing forecasts and we will provide results that complement those available in, e.g., Lütkepohl (2005). Alternatively, direct forecast methods (see, e.g., Bhansali, 2002 and references therein, and more recently Marcellino, Stock and Watson, 2003, 2006) are a natural choice when there is hesitation about the true model characterizing the DGP or when non-linearities make multiple-step ahead forecasts cumbersome to obtain. We derive asymptotic results for direct forecasts based on linear vector autoregressions for infinite order DGPs. We will call forecasts based on this method *local projections,* following the nomenclature in

Jordà (2005). Some of the derivations that we provide will look familiar to those readers acquainted with Lewis and Reinsel (1985), and Kuersteiner (2001, 2002), to cite a few.

A 95% confidence, multi-dimensional ellipse based on the joint distribution of the forecast path is an accurate representation of its uncertainty, but it is impossible to display in two-dimensional space. Another contribution of the paper is to introduce several methods to present such joint uncertainty in a useful manner to the end-user of the forecasting exercise based on Scheffé's (1953) method of simultaneous inference. In particular, in Section 3, we show how to construct simultaneous confidence bands, conditional confidence bands for the uncertainty associated to individual forecast horizons, and fan charts based on the quantiles of the joint predictive density.

The availability of the joint predictive density allows us to construct the distribution of forecasts conditional on future values of one or more of the endogenous variables in the system under consideration. In Section 4 we show that the Wald metric provides a natural statistic to evaluate the likelihood of observing the conditioning paths. These large-sample results are related to Waggoner and Zha's (1999) Bayesian derivations and provide asymptotic justification for bootstrap-based, finite-sample inference (Horowitz, 2001).

A natural consequence of the correlation across forecast horizons is the desire to cast forecasting performance comparisons in terms of forecast paths. Therefore, in Section 5, we introduce the mean squared forecast path (MSFP) as the natural extension of the MSFE by using the Wald metric and the forecast path's correlation matrix. Furthermore, we show how to appropriately use this Wald metric to extend formal testing of predictive ability along the lines of the Diebold-Mariano-West (Diebold and Mariano, 1995; West, 1996) test or the more recent approach in Giacomini and White (2006). These extensions also complement deterministic measures such as the determinant of the covariance matrix of the vector of forecast errors at different horizons proposed by Clements and Hendry (1993).

In Section 6, we provide two empirical examples to illustrate all of the results introduced in the paper. In particular, the first application examines path forecasts for a system of U.S. macroeconomic variables, and provides an example of counterfactual simulation. The second application serves to illustrate the newly introduced measures of forecast path comparison, and examines the role of monetary aggregates in forecasting US and euro area inflation, a topic of much recent debate in central bank circles.

Several final comments are worth making. We hope our paper will pave the way for many natural extensions, the most immediate of which are to: (1) generalize our basic assumptions on the DGP as well as the mixing and heteroskedasticity assumptions of the error process; and (2) to extend our large-sample results to finite-sample inference based on bootstrap or subsampling refinements. We felt space considerations and transparency had to rein in our ambitions. Finally, ours is not a criticism of the status-quo, rather we view our contribution as an addition to existing methods: different hypotheses require different statistics and appropriately tailoring statistics results in more precise answers.

2 Asymptotic Distribution of the Forecast Path

This section characterizes the asymptotic distribution of the forecast path under the assumption that the data generating process (DGP) is of infinite order while the forecasts are generated by finite-order VARs or finite-order local projections. We feel the DGP is sufficiently general to represent a large class of problems of interest and that VARs and local projections are the two most commonly used modeling strategies. We begin by stating our assumptions on the DGP.

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

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Phi_j' \mathbf{u}_{t-j}, \tag{1}$$

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 $i.i.d.$

(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 (1) can also be written as an infinite VAR process (see, e.g. Anderson, 1994),

$$\mathbf{y}_t = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \tag{2}$$

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 (2). 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 (2006) 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.

Given the DGP in expression (2) suppose we estimate a VAR(p) instead. This VAR(p) will be of the form

$$\begin{aligned} \mathbf{y}_t &= \mathbf{m} + \sum_{j=1}^{p} A_j \mathbf{y}_{t-j} + \mathbf{w}_t \\ \mathbf{w}_t &= \sum_{j=p+1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t \end{aligned} \tag{3}$$

Given estimates from this VAR(p) then one could construct forecasts with standard available formulas (see, e.g. Hamilton, 1994). Alternatively, forecasts could be constructed with a sequence of local projections given by

$$\begin{aligned}\mathbf{y}_{t+h} &= \mathbf{m}_h + \sum_{j=0}^{p-1} A_j^h \mathbf{y}_{t-j} + \mathbf{v}_{t+h} \\ \mathbf{v}_{t+h} &= \sum_{j=p}^{\infty} A_j^h \mathbf{y}_{t-j} + \mathbf{u}_{t+h} + \sum_{j=1}^{h-1} \Phi_j \mathbf{u}_{t+h-j} \quad \text{for } h = 1, ..., H\end{aligned} \tag{4}$$

where:

(i) $A_1^h = \Phi_h$ for $h \geq 1$

(ii) $A_j^h = \Phi_{h-1} A_j + A_{j+1}^{h-1}$ 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:

(iii) $X_{t,p} = \left(\mathbf{1}, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-p}\right)'$.

(iv) $\underset{kp+1\times k}{\widehat{\Gamma}_{1-p,h}} = (T-p-h)^{-1} \sum_{t=p}^{T} X_{t,p} \mathbf{y}'_{t+h}$.

(v) $\underset{k(p+1)\times k(p+1)}{\widehat{\Gamma}_p} = (T-p-h)^{-1} \sum_{t=p}^{T} X_{t,p} X'_{t,p}$.

Then, the least-squares estimate of the VAR(p) in expression (3) is given by the formula

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

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

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

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

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

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

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

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

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

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

(c) $\sqrt{\frac{T-p-h}{p}} vec\left(\widehat{A}(p,h) - A(p,h)\right) \xrightarrow{d} N(0, \Sigma_\alpha)$ *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* $\widehat{\Sigma}_u(p) = (T-p)^{-1} \sum_{t=1}^{T} \widehat{\mathbf{u}}(p)_t \widehat{\mathbf{u}}(p)_t'$ *then* $\sqrt{T}\left(\widehat{\Sigma}_u(p) - \Sigma_u\right) \rightarrow N(0, \Omega_\Sigma)$ *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 $p^2/T \rightarrow 0$, $T, p \rightarrow \infty$ 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.

Next, denote with $\mathbf{y}_T(h)$ the forecast of the vector $\mathbf{y}_{T+h}$ assuming the coefficients of the infinite order process (2) 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 local projections, respectively

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

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

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

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

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

$$\underset{kH\times kH}{\Omega_H} \equiv E\left[\left(Y_T(H) - Y_{T,H}\right)\left(Y_T(H) - Y_{T,H}\right)'\right] = \mathbf{\Phi}\left(I_H \otimes \Sigma_u\right)\mathbf{\Phi}'. \tag{5}$$

Furthermore, since parameter estimates are based on a sample of size T and hence $\mathbf{u}_t$ for $t = p+h, ..., T$ while the term $Y_T(H) - Y_{T,H}$ only involves $\mathbf{u}_t$ for $T+1, ..., T+H$, then it should be clear that to derive the asymptotic distribution of $\left[\widehat{Y}_T(H) - Y_T(H)\right]$, 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 (5) but the covariance between these terms will be zero.

Corollary 1(a) and 1(b) 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

$$\sqrt{\frac{T-p-H}{p}}vec\left(\widehat{Y}_T(H)-Y_T(H)\right)\xrightarrow{d}N(0,\Psi_H) \tag{6}$$

$$\Psi_H \equiv \frac{\partial vec\left(\widehat{Y}_T(H)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)}\Sigma_A\frac{\partial vec\left(\widehat{Y}_T(H)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)'}$$

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

$$\widehat{\mathbf{A}}=\begin{bmatrix}\widehat{A}(p,1)\\ \vdots\\ \widehat{A}(p,H)\end{bmatrix}. \tag{7}$$

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

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

$$\frac{\partial vec\left(\widehat{Y}_T(H)\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)}=\frac{\partial vec\left(\widehat{\mathbf{A}}X_{t,p}\right)}{\partial vec\left(\widehat{\mathbf{A}}\right)}=\underset{kH\times k^2Hp+kH}{\left(X'_{T,p}\otimes I_{kH}\right)}, \tag{8}$$

which combined with corollary 1(c) results in

$$\sqrt{\frac{T-p-H}{p}}\left(vec\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right)\xrightarrow{d}N(0,\Sigma_A) \tag{9}$$

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

Putting together expressions (6), (5), (8) and (9), we arrive at the following corollary.

Corollary 2 *Under assumptions 1 and 2 and expressions (6), (5), (8) and (9), the asymptotic distribution of the forecast path generated with the local projections approach described in assumption 1 is*

$$\sqrt{\frac{T-p-H}{p}} vec\left(\widehat{Y}_T(H) - Y_{T,H}\right) \overset{d}{\to} N\left(\mathbf{0}; \Xi_H\right) \qquad (10)$$

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

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

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

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 (3). 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

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

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

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

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

which is all we need to construct all the elements in the asymptotic covariance matrix of $\widehat{Y}_T(H)$, namely Ψ_H. An expression for $\widehat{\mathbf{y}}_T(h)$ generated from the VAR(p) in expression (3) can be obtained as

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

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

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

$$\underset{k\times kp+1}{J} = (\underset{k\times 1}{\mathbf{0}}, \underset{k\times k}{I_k}, \underset{k\times k}{\mathbf{0}_k}, ..., \underset{k\times k}{\mathbf{0}_k}).$$

Therefore, notice that

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

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*

$\widehat{Y}_T\left(H\right)$ *generated from the VAR(p) in expression (3) is given by*

$$\sqrt{\frac{T-p-H}{p}}vec\left(\widehat{Y}_T\left(H\right)-Y_{T,H}\right)\overset{d}{\to}N\left(\mathbf{0};\Xi_H\right) \tag{11}$$

$$\begin{aligned}\Xi_H &= \left\{\frac{p}{T-p-H}\Omega_H+\Psi_H\right\} \\ \Omega_H &= \mathbf{\Phi}(I_H\otimes\Sigma_u)\mathbf{\Phi}' \\ \Psi_{i,j} &= \frac{p}{T-p-H}\sum_{k=0}^{i-1}\sum_{s=0}^{j-1}E(X'_{T,p}(B')^{i-1-k}\Gamma_p^{-1}B^{j-1-s}X_{T,p})\otimes\Pi_k\Sigma_u\Pi'_s \\ &= \frac{p}{T-p-H}\sum_{k=0}^{i-1}\sum_{s=0}^{j-1}tr((B')^{i-1-k}\Gamma_p^{-1}B^{j-1-s}\Gamma_p)\Pi_k\Sigma_u\Pi'_s\end{aligned}$$

In practice all moment matrices can be substituted by their sample counterparts as usual. Notice also that corollaries 2 and 3 can be combined with the delta method to provide the joint predictive density of linear or nonlinear functions of the path forecasts.

In summary, corollaries 2 and 3 provide the necessary results on the joint predictive density of the path forecasts. The next sections exploit these results to compute simultaneous confidence regions for path forecasts, to study the properties of conditional path forecasts, and to propose new methods of path forecast comparison and predictive ability testing.

3 Simultaneous Confidence Regions for Forecast Paths

This section considers the problem of constructing a simultaneous confidence region for the forecast path of the j^{th} variable in the k-dimensional system we have so far examined. Under assumptions 1 and 2, corollaries 2 and 3 show that the asymptotic distribution of $\widehat{Y}_T(H)$ is (with the obvious simplifications relative to (10) or (11)):

$$\sqrt{T}\left(\widehat{Y}_T(H)-Y_{T,H}\right)\overset{d}{\to}N\left(\mathbf{0};\mathbf{\Xi}_H\right). \tag{12}$$

Let $S_j \equiv (I_H \otimes \mathbf{e}_j)$ where $\mathbf{e}_j$ is a $1 \times k$ vector of zeros with a 1 in the j^{th} column. Then the asymptotic distribution for the forecast path of the j^{th} variable in (12) is readily seen to be

$$\sqrt{T}\left(\widehat{Y}_{j,T}(H) - Y_{j,T,H}\right) \xrightarrow{d} N\left(\mathbf{0}; \mathbf{\Xi}_{j,H}\right), \tag{13}$$

where $\widehat{Y}_{j,T}(H) = S_j \widehat{Y}_T(H)$; $Y_{j,T,H} = S_j Y_{T,H}$; and $\mathbf{\Xi}_{j,H} = S_j \mathbf{\Xi}_H S_j'$. The derivations we are about to present do not depend on how one arrives at expression (13). Therefore, to make the results more general, we take expression (13) as our primitive assumption so as to accommodate forecasting environments other than those implied by assumptions 1 and 2.

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

Consider now the first order differentiable function $g(.) : \mathbb{R}^H \to \mathbb{R}^s$ for $H \geq s$ with an $H \times s$ invertible Jacobian denoted $G(.)$. The decision problem associated with this transformation of the forecast path can be summarized by the null hypothesis $H_0 : g(Y_{j,T,H}) = g_0$ for any $j = 1, ..., k$; sample T; and forecast horizon H and where g_0 is an $s \times 1$ vector. The Gaussian expression (13), 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 tested with the statistic

$$W_H = T\left(g(\widehat{Y}_{j,T}(H)) - g_0\right)' \left(\widehat{G}'_{j,T,H} \mathbf{\Xi}_{j,H} \widehat{G}_{j,T,H}\right)^{-1} \left(g(\widehat{Y}_{j,T}(H)) - g_0\right) \xrightarrow{d} \chi^2_H \qquad (14)$$

where $\widehat{G}_{j,T,H}$ denotes the Jacobian evaluated at $\widehat{Y}_{j,T}(H)$. For example, a simple null of joint significance means that $g(\widehat{Y}_{j,T}(H)) = \widehat{Y}_{j,T}(H)$; and $g_0 = \mathbf{0}_{H\times 1}$ so that a confidence region at an α significance level is represented by the values of $Y_{j,T,H}$ that satisfy

$$\Pr\left[W_H \leq c^2_\alpha\right] = 1 - \alpha$$

where c^2_α is the critical value of a random variable distributed χ^2_H at a $1-\alpha$ confidence level. This confidence region is a multi-dimensional ellipsoid that in general, is too complicated to display graphically and makes communication of forecast uncertainty difficult. However, for $H = 2$, this region can be displayed in two-dimensional space as is done in Figure 1.

Figure 1 displays the 95% confidence region associated to a one- and two-period ahead forecasts from an AR(1) model with known autoregressive coefficient $\rho = 0.75$ and $\sigma = 1$. 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. Meanwhile, the top panel of Figure 2 illustrates that the problem worsens with the forecast horizon, in the sense that the correlation between the two- and three-period ahead forecast errors is larger than that between the one- and two-period errors. The larger the correlation in forecast errors, the larger the size distortion of the two-standard-error-box. Moreover, from the bottom panel of Figure 2, adding an MA component with a positive coefficient to the AR(1) model further worsens the situation. These two cases are quite relevant empirically, since medium-horizon forecasts are of interest for policy making and a positive MA component is statistically significant for several macroeconomic time series, e.g., for US inflation.

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 construct a simultaneous rectangular region with Scheffé's (1953) S-method of simultaneous inference (see also Lehmann and Romano, 2005). Briefly, the intuition of the method is to exploit the Cauchy-Schwarz inequality to transform the Wald statistic from L_2-metric into L_1-metric to facilitate construction of a rectangular confidence interval.

Notice 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. Notice then that

$$
\begin{aligned}
\Pr\left[T\left(\widehat{Y}_{j,T}(H)-Y_{j,T,H}\right)'\Xi_{j,H}\left(\widehat{Y}_{j,T}(H)-Y_{j,T,H}\right)\leq c_\alpha^2\right] &= 1-\alpha \\
\Pr\left[\left(\widehat{Y}_{j,T}(H)-Y_{j,T,H}\right)'(PP')^{-1}\left(\widehat{Y}_{j,T}(H)-Y_{j,T,H}\right)\leq c_\alpha^2\right] &= 1-\alpha \\
\Pr\left[\widehat{V}_{j,T}(H)'\widehat{V}_{j,T}(H)\leq c_\alpha^2\right] &= 1-\alpha \\
\Pr\left[\sum_{h=1}^{H}\widehat{v}_{j,T}(h)^2\leq c_\alpha^2\right] &= 1-\alpha \qquad (15)
\end{aligned}
$$

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

Consider now the problem of constructing the rectangular confidence region for the average path-average 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 (15) to obtain

$$\Pr\left[\left|\sum_{h=1}^{H}\frac{\widehat{v}_{j,T}(h)}{h}\right| \leq \sqrt{\frac{c_\alpha^2}{H}}\right] = 1-\alpha, \tag{16}$$

which in turn implies that $\delta_\alpha = \sqrt{\frac{c_\alpha^2}{H}}$. Expression (16) and $\widehat{V}_{j,T}(H) = P^{-1}\widehat{Y}_{j,T}(H)$ imply that a simultaneous confidence region for the forecast path $\widehat{Y}_{j,T}(H)$ can then be constructed as

$$\widehat{Y}_{j,T}(H) \pm \delta_\alpha P\mathbf{i}_H \tag{17}$$

where $\mathbf{i}_H$ is an $H \times 1$ vector of ones. We call these bands Scheffé confidence bands to distinguish then from the usual two (marginal) standard error bands commonly reported.

One way to gain intuition about the Scheffé bands is to establish their relation to traditional marginal error bands. In a traditional error band, the boundaries of the band represent a shift from the mean of the distribution (the parameter estimate) in proportion to its variance. Thus, the boundary is the appropriately scaled critical values of the standard normal density of a region with symmetric $100(1-\alpha)\%$ coverage, that is for example, $\widehat{y}_{j,T}(h) \pm z_{\alpha/2}\widehat{\Xi}^{1/2}_{j,(h,h)}$. Similarly, consider now a *simultaneous* shift in all the elements of the forecast path in proportion to their variances. What would the appropriate critical value be? It is easier to answer this question with the orthogonal coordinate system $\widehat{V}_{j,T}(H)$. From expression (15) and denoting this critical value δ_α, then δ_α must meet the condition

$$\Pr\left[\delta_\alpha^2 + \overset{H}{\dots} + \delta_\alpha^2 = c_\alpha^2\right] = 1 - \alpha$$

which implies that $\delta_\alpha = \sqrt{\frac{c_\alpha^2}{H}}$. In two dimensions, both panels of Figure 3 display diagonals intersecting the origin of both ellipses. The slopes of these diagonals reflect the relative variance of each forecast (in the bottom panel the normalization ensures the variances are the same so the diagonal is the 45 degree line) and represent the $\pm\delta_\alpha$ for all values of α. The Cholesky factor P therefore provides the appropriate scaling for δ_α since it not only scales the orthogonal system by the individual variances of its elements but also accounts for their correlation.

Several results deserve comment. First, when $H = 1$ so that we are considering a one-period ahead forecast, then $c_{0.05}^2 = 3.84$ for a χ_1^2 random variable and hence $\delta_\alpha = \sqrt{3.84} = 1.96$. In this case, $P = \sigma_1$ so that the rectangular confidence interval obtained by Scheffé's S-method corresponds to the traditional two standard-error band. However, when $H = 2$, then $\delta_\alpha = 1.73$, not the usual 1.96. Second, because the relation between the L_2-norm implied by the Wald statistic and the rectangular region implied by the associated L_1-norm holds by Hölder's inequality (rather than with equality), the probability coverage is more conservative. Third, an alternative approach is to construct confidence intervals with Bonferroni's inequality. This inequality suggests to construct a $\left(1 - \frac{\alpha}{H}\right)$ confidence interval for $y_{j,T}(h)$, $h = 1, ..., H$; then, the union of these confidence intervals generates a region that includes $Y_{j,T,H}$ with *at least* $(1 - \alpha)$ probability. Specifically, the Bonferroni confidence region (BCR) is

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

where $z_{\alpha/2H}$ denotes the critical value of a standard normal random variable at a $\alpha/2H$ significance level. Thus, the BCR can be significantly more conservative than our simultaneous confidence region, specially when the correlation between forecasts across horizons is low.

In addition, we offer two complementary ways to report uncertainty about the forecast path. The first is to notice that it is easy to construct a fan chart with the quantiles of the joint predictive density simply by calculating the simultaneous rectangular regions associated with the values that c_α^2 and δ_α take for different values of α. An example of such a chart is provided below in the empirical section.

The second measure is based on the following observation. 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, the Wald statistic in expression (14) can be rewritten as

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

where $\widetilde{V}_{j,T}(H) = Q^{-1}\left(\widehat{Y}_{j,T}(H) - Y_{j,T,H}\right)$ is the unstandarized version of $\widehat{V}_{j,T}(H)$; and d_{hh} is the h^{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 forecast path at time h given the path from 1 to $h-1$ for $h = 1, ..., H$. An implication of this result is that an $100(1-\alpha)\%$ confidence region that sterilizes the uncertainty about the forecast path up to time $h-1$ and summarizes the uncertainty about the h horizon forecast alone, can be easily constructed with the bands

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

where $z_{\alpha/2}$ refers to the critical value of a standard normal random variable at an $\alpha/2$ significance level.

A simple example provides further intuition about the relation between the different confidence regions discussed in this section. Suppose the data were generated by the simple AR(1) model

$$y_t = \rho y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \sigma^2\right)$$

and for simplicity assume that ρ is known rather than estimated. Then, the 95% confidence ellipse results from the associated Wald statistic in expression (14), that is

$$W_2 = \frac{1}{\sigma^2}\left(\widehat{Y}_T(2) - Y_{T,2}\right)' \begin{bmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{bmatrix}^{-1} \left(\widehat{Y}_T(2) - Y_{T,2}\right) \leq 6,$$

since $c^2_{0.05} \simeq 6$ for a χ^2_2 random variable. This is the ellipse displayed in the top panel of Figure 3 for $\sigma = 1$ and $\rho = 0.75$. For this example, the traditional two standard-error box is given by $[-1.96, 1.96]$ and $[-2.45, 2.45]$.

The Cholesky decomposition of this forecast path's covariance matrix is

$$\sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & 1 \end{bmatrix} \tag{18}$$

and it is clear that the orthogonal path's covariance matrix is

$$\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

that is, this covariance simply summarizes the fact that the only source of forecast uncertainty in this simple example comes from the *i.i.d.* shock, ε_t (to see it more clearly suppose that $\rho = 0$). The 95% confidence circle associated to the orthogonalized forecast path is displayed in the bottom panel of Figure 3. The associated rectangular region can be easily constructed by noticing that $\delta_\alpha = \sqrt{c_\alpha^2/2} = 1.73$ and hence the box is given by $[-1.73, 1.73]$ and $[-1.73, 1.73]$. Parenthetically, these also correspond to the conditional two standard error bands.

In order to obtain the simultaneous, rectangular 95% confidence region we need to translate the rectangular box in the orthogonal coordinate system back to the original coordinate system of the forecast path. The Cholesky factor when $\rho = 0.75$ can be easily obtained from expression (18), which multiplied by the orthogonal rectangular values

$$\pm \begin{bmatrix} 1 & 0 \\ 0.75 & 1 \end{bmatrix} \begin{bmatrix} 1.73 \\ 1.73 \end{bmatrix} = \pm \begin{bmatrix} 1.73 \\ 3.03 \end{bmatrix}$$

delivers the simultaneous 95% rectangular region $[-1.73, 1.73]$ and $[-3.03, 3.03]$, which is reported in the top panel of Figure 3. Notice that compared to the traditional two standard-error box $[-1.96, 1.96]$ and $[-2.45, 2.45]$, Scheffé's (1953) method produces a confidence region in which the first period's forecast uncertainty is narrower but the second period's is wider. Figure 4 translates all of the bands discussed so far into a more traditional format to make the interpretation more transparent, assuming for simplicity that $\widehat{y}_{T+1} = 1$.

4 Conditional Path Forecasts

This section discusses a method of path forecast evaluation that exploits the asymptotic Gaussian approximation of the joint predictive density in corollaries 2 and 3; and properties of the multivariate normal distribution and linear projections. The problem that we have in mind is that of constructing forecast paths conditional on alternative hypothetical paths for a subset of variables in the k-dimensional system being considered. An example perhaps provides better intuition about what we mean.

Suppose a policy maker is confronted with a set of path forecasts $\widehat{Y}_T(H)$ about the future behavior of output growth, inflation, oil prices, exchange rates, interest rates, and so on. Given these forecasts, suppose the policy maker wants to stress the model and examine how would these macroeconomic forecasts vary if, for example, a path of oil prices different than that predicted were to take place. It turns out that when the joint predictive density is Gaussian, not only is it simple to obtain what the conditional forecast paths would be, it is straightforward to calculate the associated conditional predictive density. Waggoner and Zha (1999) develop Bayesian methods to compute this distribution in finite samples for VARs whereas Leeper and Zha (2003) further investigate projections based on hypothetical paths of monetary policy.

It is important to remark that this section does not specifically address parameter stability in the face of variability in the hypothetical paths in the sense discussed by Lucas (1976). Such questions are reserved for future research. Rather, we examine hypothetical paths drawn from the joint predictive density and, therefore, a natural starting point is to examine the likelihood of the hypothetical paths proposed. For this reason, we introduce some additional notation.

Specifically, let $k = k_0 + k_1$ where k_0 is the dimension of the subvector of variables whose conditional path forecasts we want to calculate, and k_1 is the subvector of variables whose hypothetical paths are the conditioning set. Define the selector matrices $S_0 = (I_H \otimes E_0)$ and $S_1 = (I_H \otimes E_1)$ where E_0 is a $k_0 \times k$ matrix formed with the k_0 rows of I_k associated with the variables whose conditional paths we wish to calculate, and E_1 is a $k_1 \times k$ matrix whose rows are the k_1 rows of I_k associated with the variables whose hypothetical paths provide the conditioning set. Therefore, if $\widehat{Y}_T^c(H)$ denotes the $kH \times 1$ vector of conditional forecasts, $\widehat{Y}_T^0(H) = S_0\widehat{Y}_T^c(H)$ is the $k_0H \times 1$ vector of forecasts conditional on the hypothetical paths $Y_T^1(H) = S_1\widehat{Y}_T^c(H)$, which are of dimension $k_1H \times 1$ and where the "hat" is omitted because the hypothetical paths are not estimated but rather are assumed.

Recall that under assumptions 1 and 2, corollaries 2 and 3 suggest that the asymptotic predictive density of $\widehat{Y}_T(H)$ is

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

although we remark that different assumptions and forecast environments could produce a similar asymptotic result. The likelihood of the hypothetical paths $Y_T^1(H)$ can be evaluated by testing the null hypothesis $H_0 : S_1 Y_{T,H} = Y_T^1(H)$ with the Wald statistic

$$W_1^c = T\left(S_1\widehat{Y}_T(H) - Y_T^1(H)\right)'\left(S_1\Xi_H S_1'\right)^{-1}\left(S_1\widehat{Y}_T(H) - Y_T^1(H)\right) \xrightarrow{d} \chi^2_{k_1 H}.$$

Hence, one minus the p-value associated with W_1^c is a measure of the distance, in probability units, between the predicted paths and the hypothetical paths of the k_1 variables. Low p-values (i.e. toward the direction of rejecting the null) stress the conditional forecasting exercise toward paths for the k_1 variables that have been rarely observed in the historical sample. In such situations, the resulting conditional forecasts are likely to be more problematic because we are asking about regions of the event space where the model has not been trained by the sample.

Once the likelihood of the hypothetical values has been assessed, we want to calculate the conditional forecast paths $\widehat{Y}_T^0(H)$ and their predictive distribution. Standard properties of linear projections and the multivariate Gaussian distribution are all is needed to conclude that

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

with a Gaussian predictive density whose covariance matrix is

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

It is worth remarking that the Gaussian approximation and Ξ_H^0 is all that is needed to construct simultaneous rectangular confidence regions, conditional error bands and Wald tests of joint hypotheses with the techniques described in previous sections. Section 6 illustrates all of these techniques with an empirical illustration.

5 Model Comparison

5.1 Mean Squared Forecast Path Error

The most commonly used metric of forecast model comparison is the mean squared forecast error (*MSFE*). Assuming for the sake of simplicity that y is univariate, and given an estimation sample of $1, ..., T$ observations and a forecast evaluation sample of $T+1, ..., T+N$ observations, this metric is constructed as

$$MSFE_h = \left(\frac{1}{N} \sum_{i=1}^{N} (\widehat{\mathbf{y}}_{T+i}(h) - \mathbf{y}_{T+i+h})^2 \right).$$

Therefore, an absolute comparison of the overall predictive merits between two competing models for a particular forecast horizon h can be directly obtained by comparing their respective $MSFE_h$.

However, often times a model that predicts well at short horizons will predict badly a longer horizons (and vice versa), thus making an assessment of overall predictive performance difficult. A natural way to overcome this difficulty is to construct a metric that evaluates entire forecast paths jointly. Clements and Hendry (1993) suggest to base comparison on the determinant of the forecast error second moment matrix pooled across horizons of interest, which they call *generalized forecast error second moment* (*GFESM*). Compared to the standard *MSFE*, the *GFESM* has the advantage of being invariant to non-singular, scale-preserving linear transformations. However, *GFESM* does not provide a natural basis on which to build tests of relative predictive ability.

Instead, suppose that the $1, ..., H$ forecast path $\widehat{Y}_T(H)$ has an asymptotic distribution given by

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

Examples of such a result are corollaries 2 and 3 in Section 2. We know then that the associated Wald statistic

$$W_H = T\left(\widehat{Y}_T(H) - Y_{T,H}\right)' \Xi_H^{-1}\left(\widehat{Y}_T(H) - Y_{T,H}\right) \xrightarrow{d} \chi^2_{kH} \tag{19}$$

provides a natural metric of distance between $\widehat{Y}_T(H)$ and $Y_{T,H}$. This metric operates at two levels: (1) the relative efficiency with which each forecast is generated; and (2) the degree of correlation between forecasts. Based on this distance metric, the measure in equivalent units to the $MSFE_h$ is

$$MSFP_{1,H} = \left(\frac{1}{HN}\sum_{i=1}^{N}\left(\widehat{Y}_{T+i}(H) - Y_{T+i,H}\right)' \widehat{\Lambda}_H^{-1}\left(\widehat{Y}_{T+i}(H) - Y_{T+i,H}\right)\right)$$

where Λ_H is the correlation matrix associated to Ξ_H and $\widehat{\Lambda}_H \xrightarrow{p} \Lambda_H$ if $\widehat{\Xi}_H \xrightarrow{p} \Xi_H$ and Ξ_H is invertible, such as in corollaries 2 and 3. The term $MSFP_{1,H}$ stands for the mean squared forecast path error over horizons $1, ..., H$.

For $H = 1$, then $MSFP_{1,1} = MSFE_1$. Similarly, if forecasts at horizons $1, ..., H$ are uncorrelated (Ξ_H is diagonal) then $MSFP_{1,H} = \frac{1}{H}\sum_{h=1}^{H} MSFE_h$, that is, an average of the MSFE over $1, ..., H$.

As an analytical illustration, suppose the data were generated by the simple AR(1) model with known parameters previously discussed in Section 3, so that the forecast path for $h = 1, 2$ is

$$\widehat{Y}_T(2) = \begin{bmatrix} \rho y_T \\ \rho^2 y_T \end{bmatrix}; Y_{T,2} = \begin{bmatrix} \rho y_T + \varepsilon_{T+1} \\ \rho^2 y_T + \varepsilon_{T+2} + \rho\varepsilon_{T+1} \end{bmatrix}.$$

with

$$\widehat{Y}_T(2) - Y_{T,2} = \begin{bmatrix} \varepsilon_{T+1} \\ \varepsilon_{T+2} + \rho\varepsilon_{T+1} \end{bmatrix} \sim N\left(0; \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{bmatrix}\right)$$

Then, the $MSFP_{1,H}$ for a predictive sample of N observations is

$$\begin{aligned} MSFP_{1,2}^{AR} &= \frac{1}{2N}\sum_{i=1}^{N}\left(\widehat{Y}_{T+i}(2) - Y_{T+i,2}\right)' \begin{bmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{bmatrix}^{-1} \left(\widehat{Y}_{T+i}(2) - Y_{T+i,2}\right) \\ &= \frac{1}{2N}\sum_{i=1}^{N}\begin{pmatrix} \varepsilon_{T+i+1} & \varepsilon_{T+i+2} + \rho\varepsilon_{T+i+1} \end{pmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{bmatrix}^{-1} \begin{pmatrix} \varepsilon_{T+i+1} \\ \varepsilon_{T+i+2} + \rho\varepsilon_{T+i+1} \end{pmatrix} \\ &= \frac{1}{2N}\sum_{i=1}^{N}\left(\varepsilon_{T+i+1}^2 + \varepsilon_{T+i+2}^2\right) \xrightarrow{p} \sigma^2 \text{ as } N \to \infty \end{aligned}$$

That is, absent parameter estimation uncertainty and given that ε_t are *i.i.d.* normal, the $MSFP_{1,2}^{AR}$ is simply the predictive sample variance of the shocks hitting the model at each forecast horizon. In other words, this is a pure measure of how well the model fits at each horizon, *distilled from the correlation between how the forecasts are constructed over time.*[1] The next subsection exploits the Wald metric on which $MSFP$ is based to extend available tests of relative predictive ability between models.

[1] Parenthetically, notice that when one omits parameter estimation uncertainty, there is no difference between forecasts made by iterating the AR(1) specification or by local projections so that this result is not method-dependent.

5.2 Tests of Path Predictive Ability

The $MSFP_{1,H}$ metric allows one to easily determine which of two competing forecasting models performs best, *in the available sample of data*. If we want to determine whether differences in forecasting performance are statistically significant, we need to recast common tests of predictive ability in terms of paths. The literature on comparing the predictive ability of competing forecasts given general loss functions was initiated by Diebold and Mariano (1995) and further formalized in West (1996); McCracken (2000); Clark and McCracken (2001); Corradi, Swanson and Olivetti (2001); and Chao, Corradi and Swanson (2001), among others. Giacomini and White (2006) extend this literature even further by considering conditional predictive ability tests based on examining null hypotheses that are expressed in terms of conditional expected forecast loss functions rather than unconditionally, as had previously been done. The literature is obviously very extensive so we cannot presume to explore the details of every possible contingency when testing path predictive ability. However, we think the underlying principle can be succinctly presented for the most common testing scenario, and we leave for further research appropriate generalizations.

Giacomini and White (2006) provide convenient and general conditions for multistep unconditional predictive ability testing à là Diebold-Mariano-West (DMW) that are summarized in theorem 4 of their paper. The forecasting environment is characterized by non-vanishing estimation uncertainty – the sample size remains fixed either because forecasts are obtained from a fixed initial sample, or because a bounded rolling window scheme is used instead. Meanwhile, the evaluation sample is allowed to grow to infinity to provide appropriate asymptotic results. Among other advantages, this setup allows one to consider comparison of nested models, a feature that is not available when using set-ups based on West (1996) assumptions.

Suppose the estimation sample is of size T (although Giacomini and White (2006) allow for bounded but varying-size, rolling samples as in Pesaran and Timmermann, 2006). Thus, estimation with a fixed initial sample is based on observations $t = 1, ..., T$ whereas estimation with rolling windows is based on observations $t = \tau(T) - T + 1, ..., \tau(T)$ for $\tau(T) = T, ..., N - H$ where N is the forecast evaluation sample size. Omitting the index T in $\tau(T)$ and using the same notation as before, then $\widehat{Y}_{\tau}^{j}(H)$ denotes the forecast path corresponding to observations $Y_{\tau,H} = (\mathbf{y}_{\tau+1}, ..., \mathbf{y}_{\tau+H})'$ obtained by method $j = 1, 2$ with an estimation sample of size T (either by a fixed initial sample or with rolling windows).

Let $L_{\tau,H}^{j}$ denote a generic loss function $L_{\tau,H}^{j}\left(Y_{\tau,H} - \widehat{Y}_{\tau}^{j}(H)\right)$, where for reasons that will become clear momentarily, we will focus on a mean squared forecast path metric. A test of unconditional equal path predictive ability consists in a test of the null

$$H_0 : E\left[L_{\tau,H}^{1} - L_{\tau,H}^{2}\right] = 0; \tau = T, ..., N - H$$

against the alternative

$$H_A : \left|E\left[\Delta\overline{L}_{\tau,H}\right]\right| \geq \delta > 0$$

for all N sufficiently large and where

$$\Delta\overline{L}_{\tau,H} = \frac{1}{N}\sum_{\tau=T}^{N-H} L_{\tau,H}^{1} - L_{\tau,H}^{2}.$$

Let $\widehat{\sigma}_N^2$ denote a suitable heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance $\sigma_N^2 = var\left[\sqrt{N}\Delta\overline{L}_{\tau,H}\right]$, then Giacomini and White (2006) show that the statistic

$$t_{N,H} = \frac{\Delta\overline{L}_{\tau,H}}{\widehat{\sigma}_N/\sqrt{N}} \xrightarrow{d} N(0,1)$$

under the null and, for any constant $c \in \mathbb{R}$, and $\Pr[|t_{N,H}| > c] \to 1$ as $N \to \infty$ under the alternative.

These results deserve several comments. First, an essential element of the proof in Giacomini and White (2006) is to provide enough primitive conditions on $\{\mathbf{y}_t\}$ so that $\Delta L_{\tau,H}$ is a martingale difference sequence for which a suitable central limit theorem can be applied. These primitive conditions allow for substantial heterogeneity and persistence, more in fact, than we have allowed in Section 2. Second, a natural choice of loss function given the discussion in the previous subsection is $L_{\tau,H} = MSFP_{1,H}$. Third, typically, in a multistep testing situation, the moving-average structure of the forecast errors justifies the need for HAC covariance estimators. However, forecast comparisons based on paths are simultaneous comparisons of forecasts at all horizons within the path. The Wald metric given in expression (19) represents a natural method of simultaneous comparison that has the virtue of orthogonalizing the sequence of forecast errors at all horizons – hence the convergence to a χ^2 distribution. Therefore, a predictive ability test with loss functions based on $MSFP_{1,H}$ suggests that, in most practical situations, heterogeneity rather than autocorrelation will be the primary concern in ensuring consistency of the asymptotic covariance matrix. Consequently, a low truncation lag for the HAC estimator will often suffice, thus providing additional power to the test, as Giacomini and White (2006) suggest. Finally, note that $MSFP_{1,H}$ requires an estimate of the correlation matrix Λ_H. Given forecasts based on our VAR or local projection set-up, one could estimate this matrix from the estimation sample with the given formulas. However, since the asymptotics are based on a fixed estimation sample, the natural choice is to obtain Λ_H from a sample estimator based on the evaluation sample to ensure consistency. We feel confident the reader will immediately see how the previous discussion can be adapted to Giacomini and White's (2006) multistep conditional predictive ability tests described in theorem 3 of their paper, and for this reason we do not further discuss the details of this or other procedures described therein.

6 Empirical Illustration

This section contains two applications of the techniques presented above. The first application examines path forecasts for a system of U.S. macroeconomic variables and provides an example of a conditional forecast simulation. The second application serves to illustrate the newly introduced measures of forecast path comparison and tests of predictive ability; and examines the role of monetary aggregates in forecasting inflation, a topic of much recent debate in central bank circles.

6.1 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 forecast exercise examines U.S. real GDP growth (on a yearly basis, in percentage terms, and seasonally adjusted); inflation (measured by the personal consumption expenditures deflator on a yearly basis, in 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 2004:II. With these data, we then construct two-year (eight-quarters) ahead forecasts for this system of variables by local projections. The lag length of the projections was automatically selected to be six by AIC_C – a correction to AIC designed for autoregressions and with better properties in small samples than either AIC or SIC (see Hurvich and Tsai, 1989).

Figure 5 displays these forecasts along with the actual realizations of these economic variables, conditional and marginal 95% 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 experiment 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. For completeness, the same forecasts are displayed in Figure 6 with 95^{th}, 50^{th} and 5^{th} simultaneous percentiles to form appropriate fan charts.

In order to make sense of where the differences between our forecasts and the historical record may come from, we experimented with the following hypothetical. Suppose that the Federal Reserve at the time believed that inflation would not run as high as predicted (perhaps because of the end of major military operations in Iraq suggested 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.

The results of this conditional experiment are reported in Figure 7. We begin by remarking that this hypothetical path is very conservative: the Wald test of the null that the hypothetical is statistically equivalent to the unconditional forecast has a p-value of 0.71, that is, the distance between the hypothetical and the forecasted path is only 29% in probability units. Therefore, we feel reasonably certain that such an experiment is still well approximated by our model.

Interestingly, the forecasts obtained by conditioning on this hypothetical path for inflation are remarkably close to the historical record. In particular, the path of increases in the federal funds rate is virtually identical to the actual path 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 was predicted to be closer to 3.5%, but otherwise both paths seem to reconnect at the end of the two year predictive horizon. Whereas we cannot be certain that this hypothetical reflected the Federal Reserve view's on inflation at the time, it serves to illustrate that formal statistical experimentation with alternative scenarios can be easily provided to policy makers with the techniques presented here.

6.2 Do Monetary Aggregates Help Forecast Inflation?

In order to achieve monetary policy goals, central banks constantly monitor risks to price stability. Since its inception, the European Central Bank (ECB) has used a so-called "two pillar" approach that gives a specific role to monetary aggregates over other economic indicators when forecasting inflation. Consequently, the question we ask in this section is whether monetary aggregates help forecast inflation in the U.S. and in the euro area, specially since the U.S. Federal Reserve (Fed) does not reserve a special role for monetary aggregates over other predictors.

Our objective is to evaluate whether there are differences in the ability to predict inflation (measured by the consumer price index minus food and energy) when money growth measures (M2, M3, and in the euro area, M3C growth, a corrected version of M3) are included alongside measures of economic activity (industrial production index growth in the U.S., real GDP growth in the euro area); and interest rates (the federal funds rate in the U.S., and the 4-month Euribor in the euro area). All variables are available monthly except for euro area real GDP, which is available quarterly. The U.S. sample begins in January 1985 and runs until January 2007 (February 2006 when M3 is used since it was discontinued at this date). We have chosen a later starting date to avoid possible structural breaks in the behavior of inflation following the turbulent 1970s. The euro area sample begins January 1997 and runs until September 2006. We note that M3C is only available beginning January 1999. Four years worth of data are reserved for out-of-sample comparisons in the U.S., and two in the euro area due to the shorter available sample. The size of the estimation sample is kept constant by obtaining estimates through a series of rolling windows as prescribed by Giacomini and White (2006).

Figure 8 plots for the U.S., the *MSFE* and the *MSFP* in terms of the percentage forecast improvement of the model that includes money aggregates (M2 in the top panels, M3 in the bottom panels) relative to the model that excludes them. Alongside is a plot of the value of the corresponding Giacomini and White (2006) statistic for loss functions based on the differences in *MSFE* and *MSFP* respectively resulting from including money aggregates as predictors. The forecasting model is based on local projections whose lag length is determined automatically by information criteria (Hurvich and Tsai's (1989) AIC corrected) so that the with and without money aggregates models do not necessarily always have the same number of lags. Similar results are reported for the euro area in Figure 8.

Figures 8 and 9 suggest that both in the U.S. and in the euro area, forecasting performance deteriorates when M2 is included but performance improves when M3 is included instead (notice that in the euro area this is true for M3C but less clear for M3). However, while these differences are not statistically significant in the U.S. (Giacomini and White (2006) statistics are well below critical values irrespective on whether the loss function is based on differences in *MSFE* or *MSFP*), the differences are statistically significant in the euro area, perhaps explaining the motivation behind the ECB's "two-pillar" policy.

More generally, the two metrics of forecasting performance, *MSFE* and *MSFP* can behave quite differently. Typically, *MSFP* is smoother and less volatile than *MSFE* and its behavior is more consistent with the Giacomini and White (2006) tests, even when the difference in loss functions is measured in terms of *MSFE.* Although both metrics tend to coincide in the direction in which the forecasts improve, we noticed that in the case of the euro area with M3, the metrics provided different suggestions with the *MSFE* preferring the model that included M3 and *MSFP* marginally preferring the model without M3.

7 Conclusions

Suppose you are comparing two defective computer monitors. One where the color of each pixel is randomly chosen from a relatively tight distribution centered at the true color for each pixel; the other shifts the color of each pixel by a random shock of higher variance than in the first monitor, but the shock is the same for all pixels (i.e. the shocks across pixels are perfectly correlated). Although the individual pixel, color-error-rate in the first monitor is smaller than in the second, images in the first will be quite blurry while images in the second will be perfectly crisp – albeit with the wrong tint.

By the same token, we believe it is often more sensible to compare the patterns implied by the forecast paths of competing models jointly when assessing predictive ability. To that end, our paper provides a long list of results. We begin by deriving the asymptotic distribution of forecast paths generated by finite order VARs or local projections from potentially infinite order DGPs. Hence our results cover a wide class of situations practitioners are likely to encounter in practice, leaving for further research elaborate extensions from these foundational results.

Summarizing the wealth of information contained in the joint distribution of the forecast path presents a host of new difficulties. We provide several new graphical solutions to this problem based on the observation that the Cholesky decomposition of the covariance matrix of the forecast path orthogonalizes the path into the constituent shocks hitting the forecasting model at each horizon. Hence we introduce simultaneous confidence bands based on Scheffé's (1953) method, conditional bands, and fan charts based on the quantiles of the joint predictive density. In the end, the underlying intuition of our derivations is the same as the intuition in classical linear regression with correlated regressors: while individual coefficients may be imprecisely estimated (low t-statistics), the joint effect could still be quite precisely estimated (high F-statistics).

Knowledge of the joint distribution is also advantageous for counterfactual simulation. However, experimentation with alternative scenarios is complicated in most economic applications since the Lucas Critique warns of the possibility that the forecasting model may be parametrically unstable with respect to the hypothetical path. Furthermore, insofar as the hypothetical paths are far away from the history observed, we are asking an approximate model to make predictions in regions were the model has no training from the sample. To get a grip on these issues, we provide formal statistics on the distance of the hypothetical from the average historical distribution of the paths based on the Wald principle. Once the validity of the hypothetical is formally assessed, we provide simple formulas to derive the paths and their distribution, conditional on the hypothetical.

Returning to the intuition of our computer monitor example, we ask what is the best way to compare the predictive ability of models. We accomplish this in two ways: by extending the traditional mean squared forecast error measures to paths (and hence we create the mean squared forecast path) and by extending Diebold-Mariano-West statistics of equal predictive ability in terms on the joint null over the path rather than on its constituent elements.

References

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

Bhansali, R. J. (2002) “Multi-Step Forecasting,” in **A Companion to Economic Forecasting**, Michael P. Clements and David F. Hendry (eds.). Oxford, U.K.: Blackwell Publishers.

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

Chao, J.C., Valentina Corradi and Norman R. Swanson (2001) “An Out-of-Sample Test for Granger-Causality,” *Macroeconomic Dynamics*, 5: 598-620.

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

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

Corradi, Valentina, Norman R. Swanson and Claudio Olivetti (2001) “Predictive Ability with Cointegrated Variables,” *Journal of Econometrics*, 164: 315-358.

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

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

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.

Gonçalves, Sílvia and Lutz Kilian (2006) “Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity,” *Econometric Reviews,* forthcoming.

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

Horowitz, Joel L. (2001) "The Bootstrap and Hypothesis Tests in Econometrics," *Journal of Econometrics*, 100(1): 37-40.

Hurvich, Clifford M. and Chih-Ling Tsai (1989) "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76(2): 297-307.

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

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

Kuersteiner, Guido (2001) "Optimal Instrumental Variables Estimation for ARMA Models," *Journal of Econometrics,* 104(2): 359-405.

Kuersteiner, Guido (2002) "Efficient Instrumental Variables Estimation for Autoregressive Models with Conditional Heteroskedasticity," *Econometric Theory,* 18: 547-583.

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

Lucas Jr., Robert E. (1976) "Econometric Policy Evaluation: A Critique," in Brunner, Karl and A. H. Meltzer (eds.). *Carnegie-Rochester Conference Series on Public Policy*, 1: 104-130.

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 (2003) "Macroeconomic Forecasting in the Euro Area: Country-Specific versus Area-Wide Information," *European Economic Review,* 47(1): 1-18.

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.

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

Pesaran, M. Hashem and Allan Timmermann (2006) “Selection of Estimation Window in the Presence of Breaks,” *Journal of Econometrics,* 137(1): 134-161.

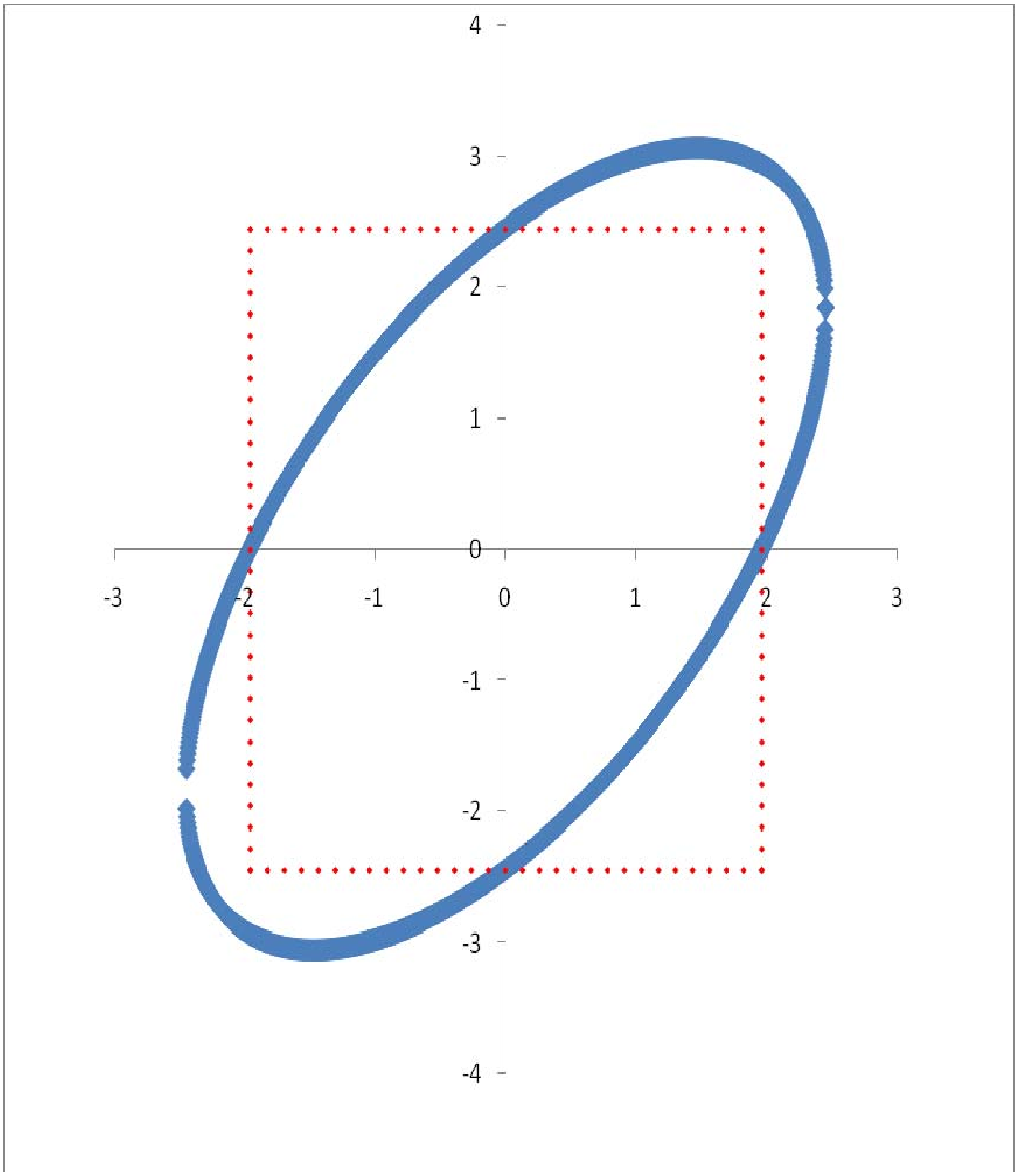
Scheffé, Henry (1953) A Method for Judging All Contrasts in the Analysis of Variance,” *Biometrika,* 40: 87-104.

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

West, Kenneth D. (1996) “Asymptotic Inference about Predictive Ability,” *Econometrica,* 64: 1067-1084.

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

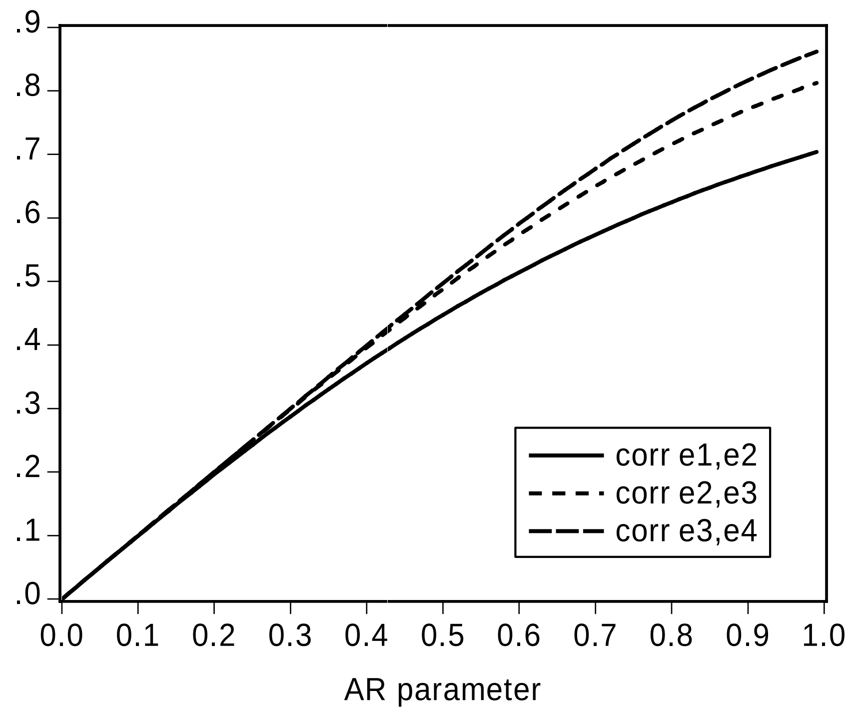
Figure 1 – 95% Confidence Ellipse for AR(1) Forecast Path over Two Horizons



Notes: Standard confidence bands and confidence ellipse (AR Coefficient = 0.75, Error Variance = 1)

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

Panel 1 – AR model



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

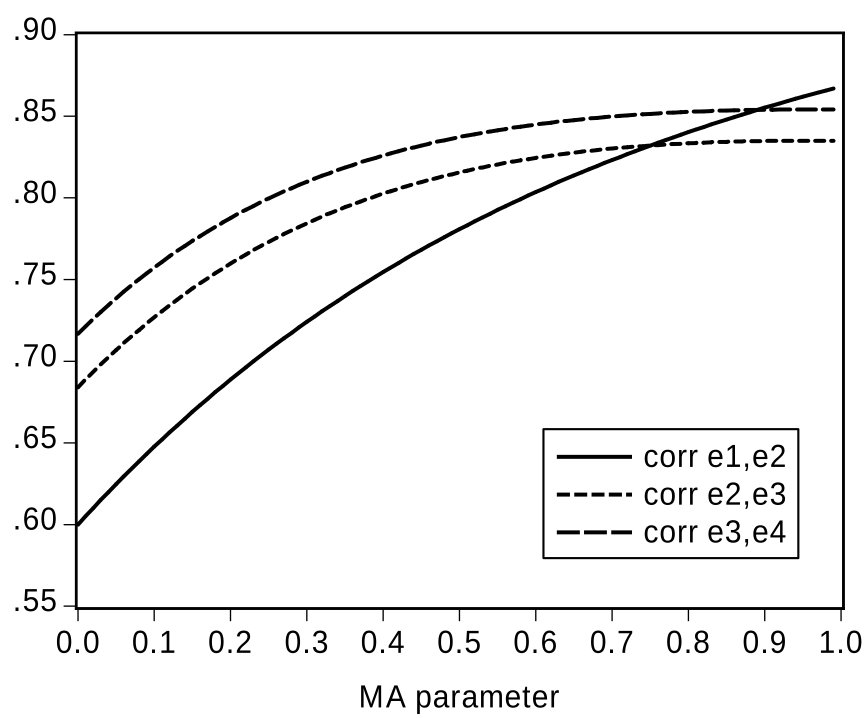
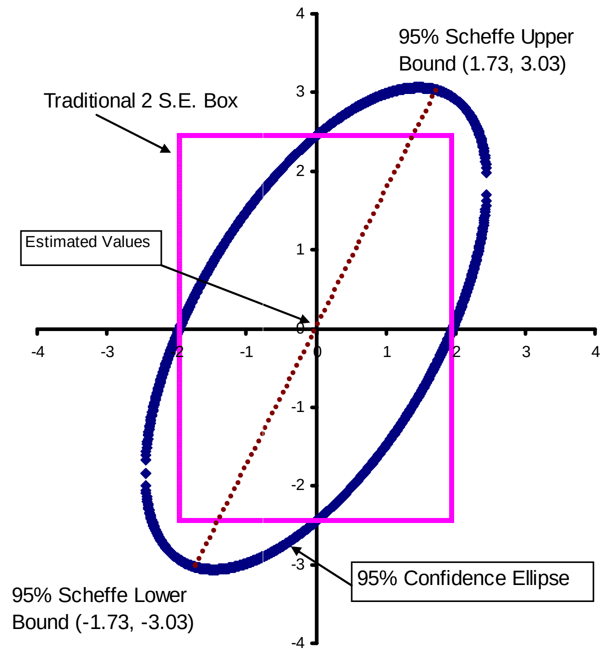
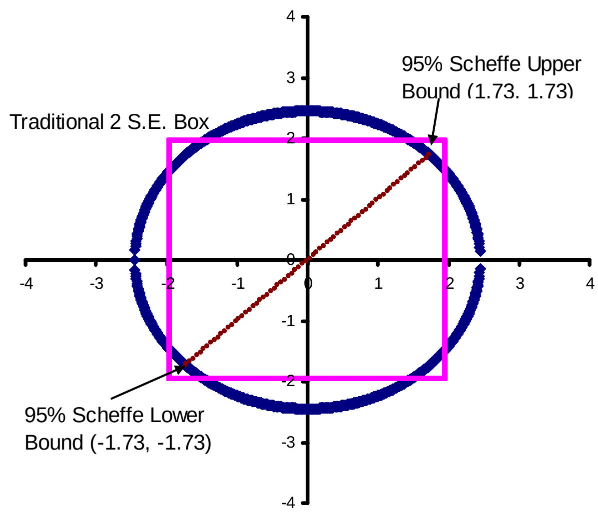


Figure 3 – 95% Scheffe bounds for AR(1) Forecast Path over Two Horizons

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

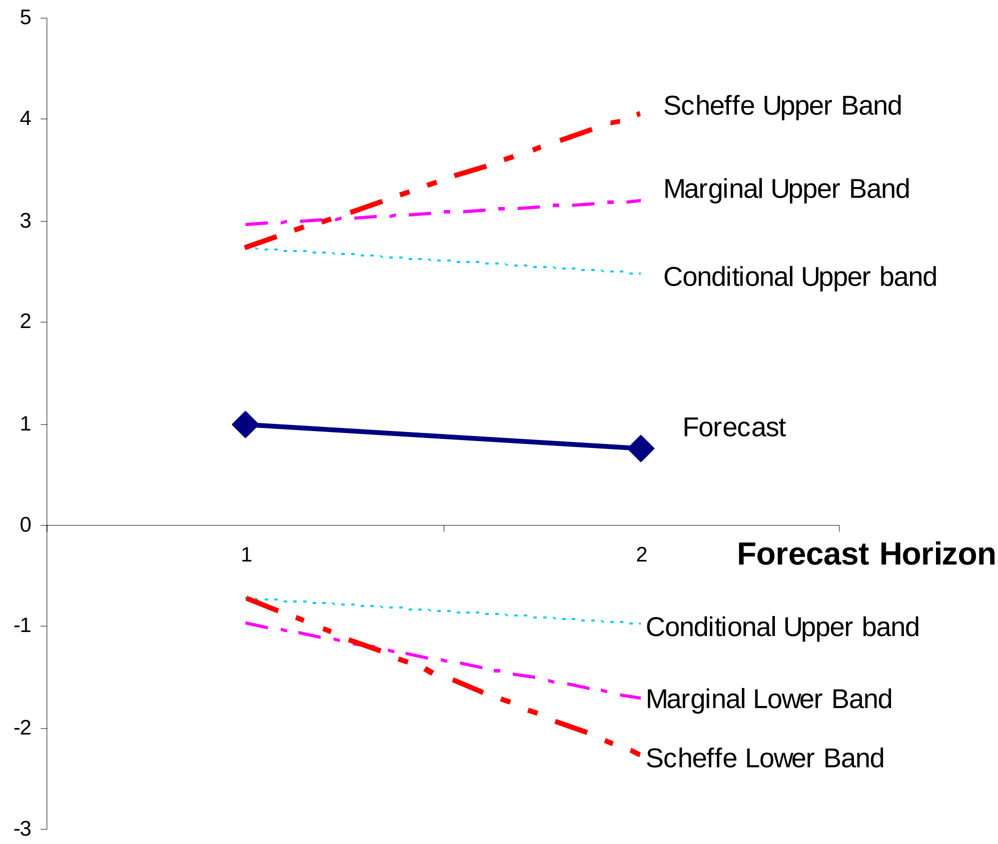


Panel 2 – 95% Confidence Circle for Orthogonalized Forecast Path



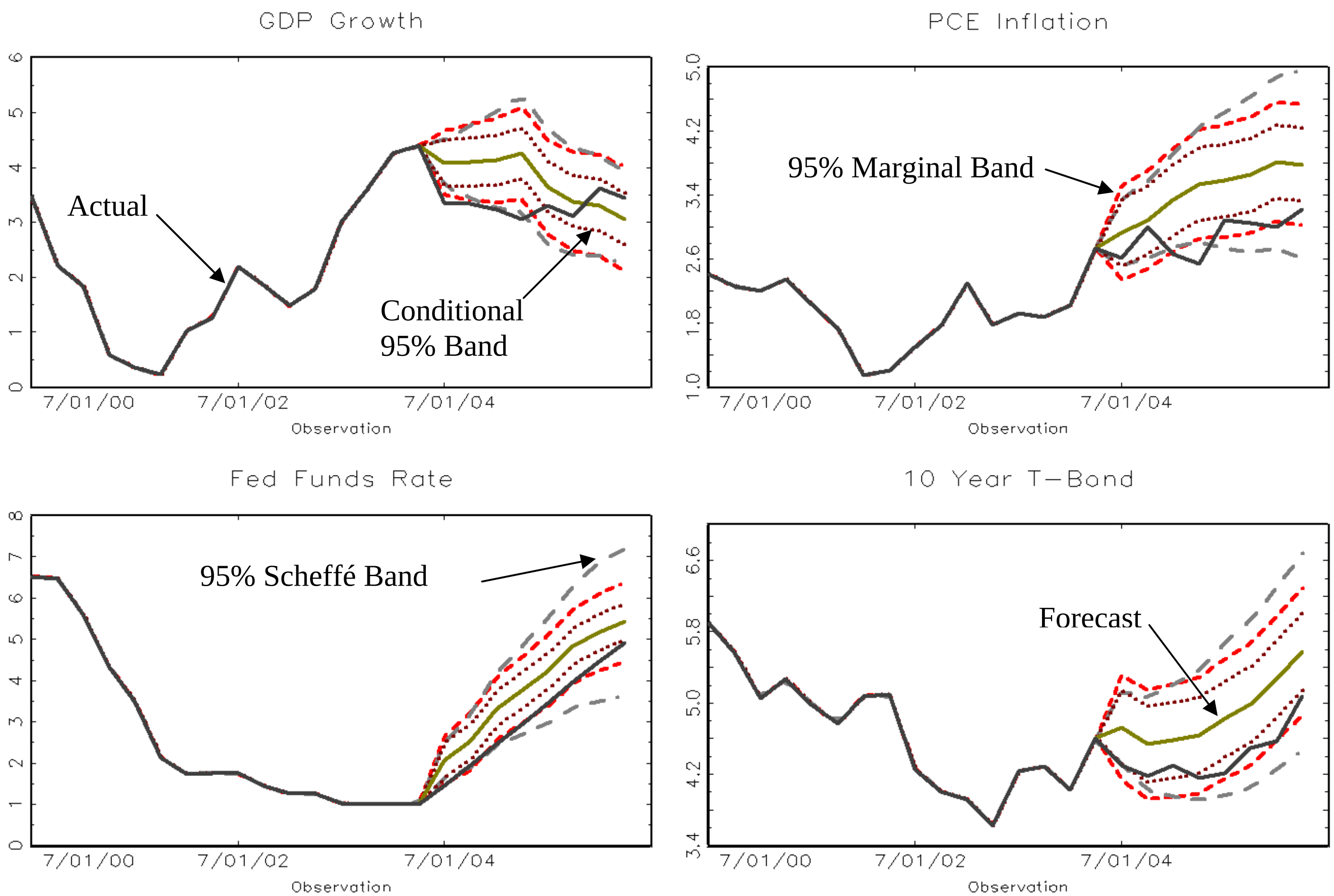
Notes: AR Coefficient = 0.75, Error Variance = 1

Figure 4 – 95% Confidence Scheffé, Marginal and Conditional Bands



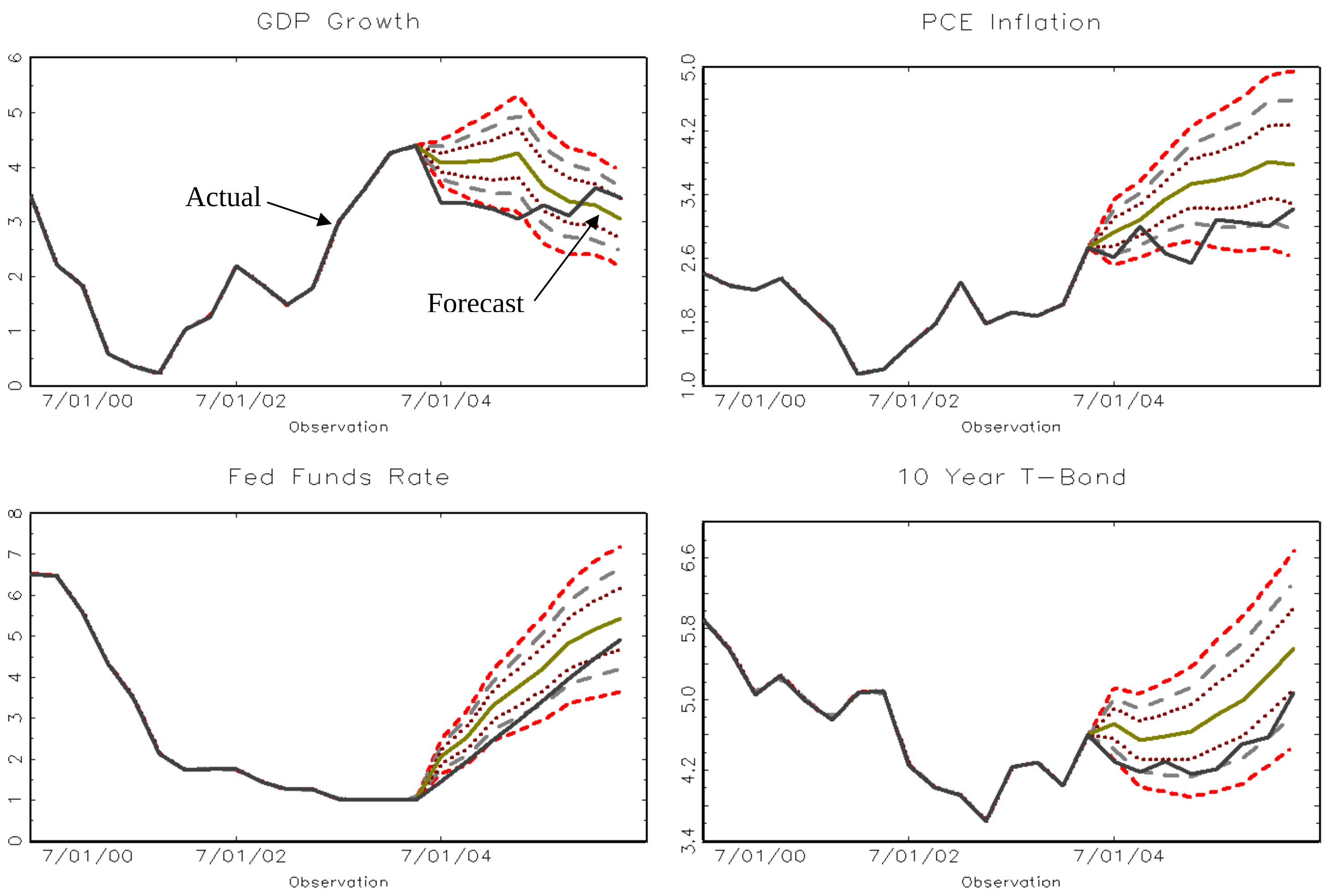
Notes: AR(1) two period ahead forecasts with ρ = 0.75 and σ = 1. This representation corresponds to the two dimensional representation in Figure 3.

Figure 5 – 95% Scheffé, Conditional and Marginal Error Bands for Macroeconomic Forecasts



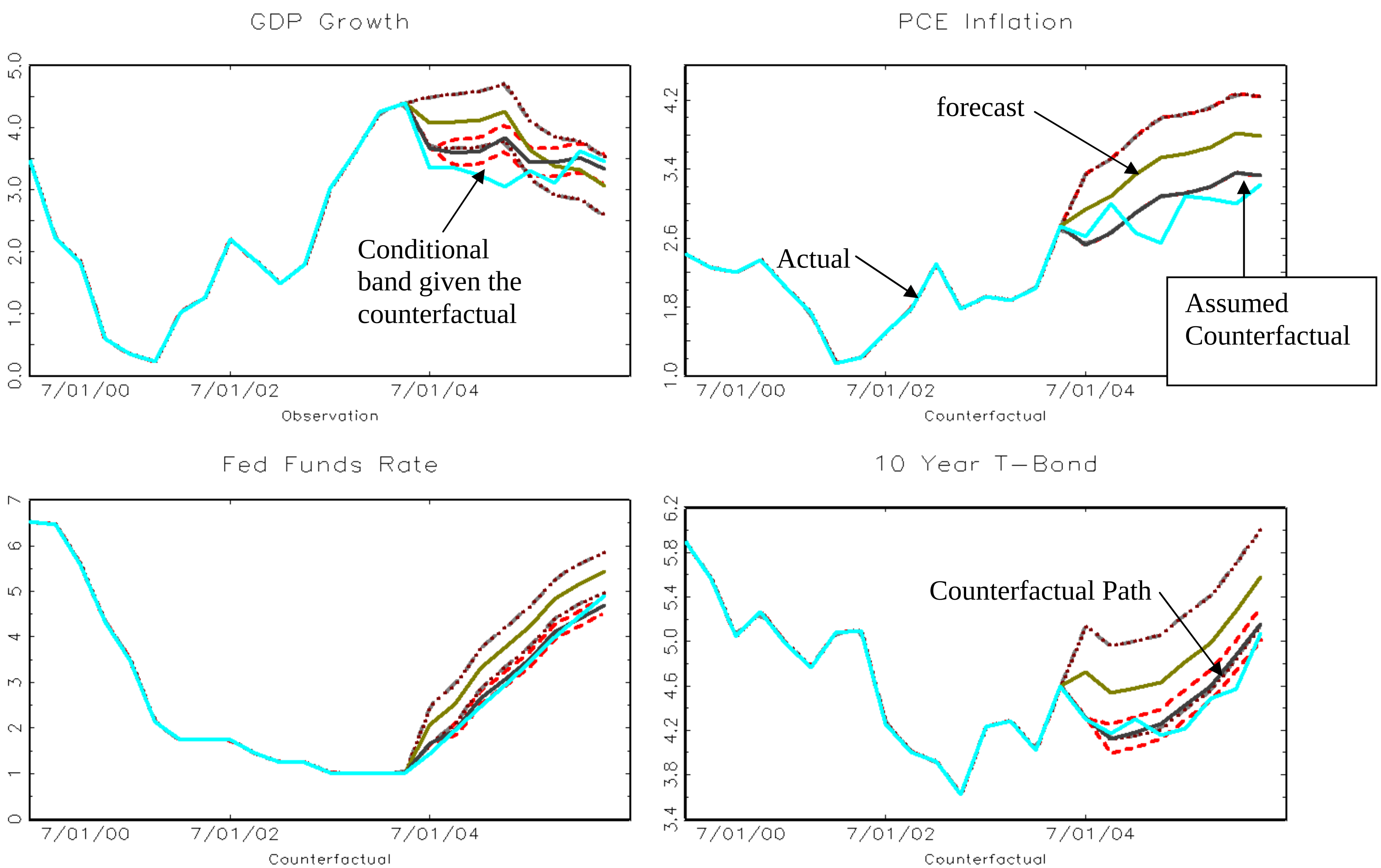
Notes: Sample runs from 1953:II to 2004:II. The forecast horizon runs from 2004:III to 2006:III.

Figure 6 – Fan Charts: 95th, 50th, and 5th percentiles



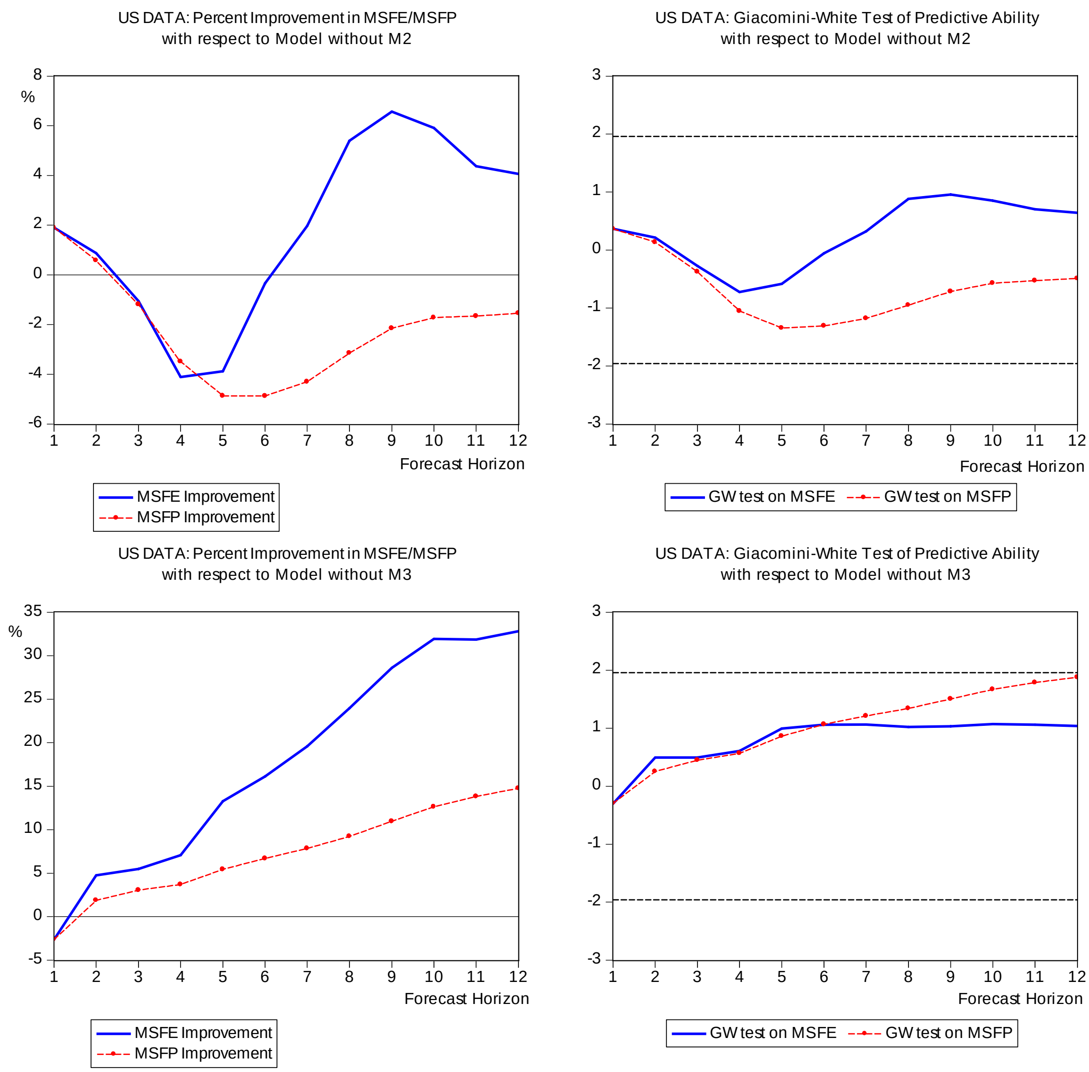
Notes: These are the same forecasts as in figure 4 but with Scheffé percentiles displayed.

Figure 7 – Counterfactual: Inflation set at the lower 95% conditional band value. Sample: 1953:II – 2004:II



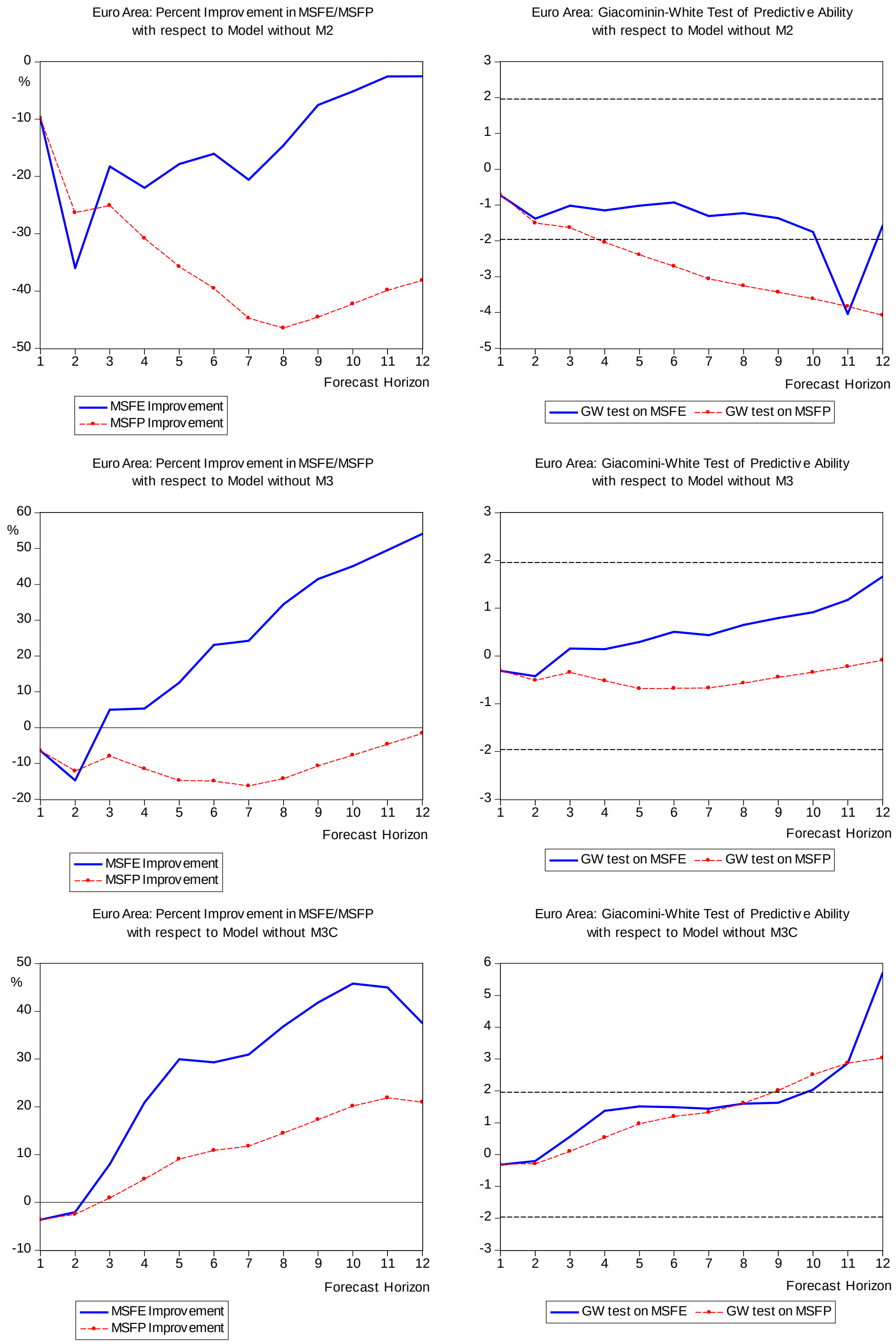
Notes: Distance of the counterfactual from the forecast in probability units is 0.29 (or p-value of the joint test of equality is 0.71).

Figure 8 – Predicting U.S. Inflation with and without Monetary Aggregates
Sample: January 1985 – January 2007



Notes: Error bands in Giacomini-White graphs indicate the critical values of the test.

Figure 9 – Predicting Euro Area Inflation with and without Monetary Aggregates
Sample: January 1997 – September 2006



Notes: Error bands in Giacomini-White graphs indicate the critical values of the test.