SIMULTANEOUS CONFIDENCE REGIONS FOR IMPULSE RESPONSES

Òscar Jordà*

Abstract—Inference about an impulse response is a multiple testing problem with serially correlated coefficient estimates. This paper provides a method to construct simultaneous confidence regions for impulse responses and conditional bands to examine significance levels of individual impulse response coefficients given propagation trajectories. The paper also shows how to constrain a subset of impulse response paths to anchor structural identification and how to formally test the validity of such identifying constraints. Simulation and empirical evidence illustrate the new techniques. A broad summary of asymptotic analytic formulas is provided to make the methods easy to implement with commonly available statistical software.

I. Introduction

THE plot of an impulse response function is a visual summary of the dynamic propagation mechanisms characterizing a vector time series. Occasionally we are interested in specific, individual coefficients, but generally it is the impulse response shapes that most concern us. Is the response of output to a monetary shock hump shaped, and is this a preeminent feature that a theoretical macroeconomic model should have? Is the response of prices to a monetary shock strictly negative in all periods? Do exchange rates respond to price and interest rate shocks with the celerity that purchasing power parity and uncovered interest rate parity would predict? Therefore, assessing the impulse responses' shapes, the precision with which they are estimated, and their statistical and economic significance figures high in the list of questions we would like to answer.

Inference about an impulse response is a multiple testing problem with coefficient estimates that are serially correlated. Statements about the set of possible impulse response trajectories require construction of simultaneous confidence regions that account for this serial correlation. For approximately multivariate normally distributed random variables, such regions are the multidimensional ellipsoids associated with traditional Wald statistics. Unfortunately, such ellipsoids cannot be easily displayed in a two-dimensional graph.

Lütkepohl's (2005) recommendation to address this problem is to construct Bonferroni bounds, although Lütkepohl himself suggests that these bounds are unnecessarily conservative. In a Bonferroni bound involving several parameters (such as in an impulse response), the bound concentrates the probability mass of the entire parameter vector onto each individual element, one at a time. Thus, these bounds guard against such extreme situations as when the true impulse response may coincide exactly with the estimated response at every point except one. A Bonferroni bound may be the right tool to guard against large but improbable risks, but it is not very useful for thinking about variations in the shape of the impulse response as a whole.

Sims and Zha (1999) instead addressed the problem of serial correlation in the impulse response coefficient estimates (see section 6 of their paper), along with a host of other issues related to the small-sample behavior of typical vector autoregressive (VAR) asymptotic approximations over long horizons and Bayesian solutions that emphasize the likelihood. The first issue is central to this paper, while discussion of the second (which others such as Rossi, 2005, and Pesavento & Rossi, 2006 have attacked with near-to-unity asymptotic theory) is left for future research.

The solution Sims and Zha (1999) proposed to the serial correlation problem consists of a principal component decomposition of the impulse response estimates' covariance matrix so as to parse the information contained in the joint distribution into a few individual factors. Unfortunately these factors are difficult to interpret: as their paper shows, they often result in bands that cross over each other (e.g., see figures 6, 11, and 12 of their paper), and they can provide varying probability coverage depending on each application's factor decomposition.

This paper provides a solution to the two problems discussed in Lütkepohl (2005) and Sims and Zha (1999) based on Scheffé's (1953) S-method of simultaneous inference and on orthogonal linear projections of the impulse responses. The end result is that I introduce two new sets of bands: Scheffé bands designed to represent uncertainty about the shape of the impulse response rather than on individual coefficients and conditional bands designed to examine the individual significance of coefficients in a given trajectory.

These new bands rely on the assumption that the coefficients of the impulse response are approximately multivariate normally distributed. Such a result is a natural consequence of impulse responses estimated with traditional VARs under general conditions (see, e.g., Hamilton, 1994; Lütkepohl, 2005), but also when impulse responses are estimated with local projections (Jordà, 2005) instead. I summarize asymptotic results for both methods that make calculating the new bands simple for practitioners.

These approximate multivariate Gaussian results and the Wald principle are also useful in anchoring structural identification of the system. Just as we use economic theory to restrict the nature of contemporaneous relations, we can

Received for publication February 15, 2007. Revision accepted for publication February 26, 2008.

* Department of Economics, University of California, Davis

The hospitality of the Federal Reserve Bank of San Francisco during the preparation of this manuscript is gratefully acknowledged. Comments by Colin Cameron, Fabio Canova, Michele Cavallo, David Drucker, Paul Gaggl, Jordi Galí, James Hamilton, Massimiliano Marcellino, Paul Ruud, and Aaron Smith, as well as by seminar participants at the Bank of Italy, Bank of Korea, BBVA, Bilkent University, CEMFI, Federal Reserve Banks of Dallas, Philadelphia, and San Francisco, Koç University, Korea University, Sabancı University, Università Bocconi–IGIER, University of California, Berkeley, and Universitat Pompeu Fabra were very helpful. The editor and two anonymous referees were specially patient and helped me improve the final version of the paper considerably.

The Review of Economics and Statistics, August 2009, 91(3): 629–647

restrict some of the impulse response trajectories (such as with nonpositivity restrictions on the response of prices to an interest rate shock). These restrictions are easily imposed and can be formally tested to help refine inference on the dimensions of the system we know less about a priori.

Throughout the presentation, I illustrate the techniques introduced with impulse responses derived from the well-known, three-variable monetary VAR that Stock and Watson (2001) use in their review of vector autoregressions. Their system contains three variables: inflation (measured by the chain-weighted GDP price index), unemployment (measured by the civilian unemployment rate), and the average federal funds rate; and it is based on a sample of quarterly data beginning 1960:I that I extend to 2007:I. As they did, identification here is achieved with the short-run recursive ordering in which the variables are reported. Further details on the specifics of the exercise are available from their paper.

In addition, this paper examines the small sample properties of the asymptotic approximations provided here with Monte Carlo simulations that are built around the Stock and Watson (2001) VAR. Further, I briefly discuss basic methods to implement bootstrap versions of the relevant statistics when the accuracy of large sample approximations is a concern. However, such issues as long-horizon asymptotic distortions in short-order VARs (Sims & Zha, 1999; Rossi, 2005; Pesavento & Rossi, 2006) and distortions of common implementations of the bootstrap (Kilian, 1998a, 1998b, 1999, 2001) require more space than is available in this paper. They are left for further research.

II. Simultaneous Confidence Regions for Impulse Responses

This section asks not how to estimate impulse responses, but rather how to summarize the distributional characteristics of a given set of estimates. For this reason, I will be intentionally vague about the method of estimation, its small sample properties, structural identification assumptions, and other aspects of the impulse response analysis that have consumed rivers of ink in the literature. Instead, the starting point of the discussion is an assumption that, at least approximately, a given set of impulse response estimates is multivariate Gaussian. Then the methods that I propose come directly from general inferential principles for multiple testing problems (see, e.g., Lehmann & Romano, 2005). Asymptotic multivariate Gaussian approximation results for impulse responses estimated with VARs (e.g., Hamilton, 1994; Lütkepohl, 2005) or with local projections (Jordà, 2005), under general assumptions about the data generating process (DGP), and for common methods of short-run and long-run structural identification are provided for completeness.

Suppose the system of impulse responses over $h = 1, \ldots, H$ horizons associated with the $r \times 1$ vector times series $\mathbf{y}_t$ is

$$\Phi(1, H) = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_H \end{bmatrix},$$

so that $\Phi(1, H)$ is an $rH \times r$ matrix and the impulse response of the ith variable to a shock in the jth variable at horizon h corresponds to the (i, j) element of the $r \times r$ matrix Φ_h. In reduced form, $\Phi_0 = I_r$, which is nonstochastic. For structurally identified systems, Φ_0 collects the contemporaneous relations between the elements of $\mathbf{y}_t$ and is orthogonal to Φ_h, $h = 1, \ldots, H$ by construction (as we will see more formally below). For this reason but without loss of generality, I find it more convenient to present the methods with $\Phi(1, H)$ alone and ignore Φ_0 in the presentation.

Now suppose estimates of $\Phi(1, H)$ based on a sample of T observations of $\mathbf{y}_t$ are available, and let $\hat{\boldsymbol{\Phi}}_T = vec(\hat{\boldsymbol{\Phi}}(1, H))$ denote the $r^2H \times 1$ vector resulting from stacking the columns of the matrix of structural impulse response estimates, $\hat{\Phi}(1, H)$. Further, assume that, at least asymptotically, this estimator has the property

$$\sqrt{T}\,(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \xrightarrow{d} N(0, \Omega_\phi). \tag{1}$$

Traditionally, significance of the impulse response estimates is reported by graphically displaying a two-standard-error, marginal rectangular interval around each coefficient estimate of the impulse response considered. Specifically, denote $\phi_h(i, j)$ as the (i, j) element of Φ_h; $\hat{\phi}_h(i, j)$ the associated estimate based on a sample of size T; and $\hat{\sigma}_h(i, j)$ as the estimate of the standard deviation $\hat{\phi}_h(i, j)$, which is the square root of the appropriate diagonal entry of the covariance matrix $\hat{\Omega}_\phi$. Then individual marginal rectangular intervals come from the realization that

$$\Pr\left[\left|\frac{\hat{\phi}_h(i,j)}{\hat{\sigma}_h(i,j)}\right| \leq z_{\alpha/2}\right] = 1 - \alpha, \tag{2}$$

since the associated t-ratio,

$$\hat{t}_h(i, j) = \frac{\hat{\phi}_h(i,j) - \phi_h(i,j)}{\hat{\sigma}_h(i,j)} \rightarrow N(0, 1), \tag{3}$$

is readily seen to be asymptotically standard normal given assumption (1), and where $z_{\alpha/2}$ denotes the appropriate critical value of a standard normal random variable at a $100(1 - \alpha)$% confidence level.

This confidence interval is another way to present the decision problem associated with testing the null hypothesis:

$$H_0 : \phi_h(i, j) = 0 \text{ vs. } H_1 : \phi_h(i,j) \neq 0 \text{ for any } i, j = 1, \ldots, r \text{ and } h = 1, \ldots, H. \tag{4}$$

Insofar as the primitive assumption (1) is correct then, the marginal, individual error bands described by (2) are the correct answer for the test of the individual nulls described by (4).

However, researchers are often interested in the shape of the path described by the impulse response, not the particular value that one of its coefficients may attain. Denote $\boldsymbol{\phi}(i, j)$ as the (i, j) impulse response function over the next 1 to H periods, and let $g(\cdot) : \mathbb{R}^H \rightarrow \mathbb{R}^k$ for $k \leq H$ be a first-order differentiable function with an $H \times k$ invertible Jacobian denoted $G(\cdot)$. The decision problem associated with the shape of the impulse response can be summarized by the null hypothesis $H_0 : g(\boldsymbol{\phi}(i, j)) = g_0$ for any i, $j = 1, \ldots, r$ and where g_0 is a $k \times 1$ vector.

The Gaussian assumption in (1), the Wald principle, and the delta method (or classical minimum distance, see, e.g., Ferguson, 1958) suggest that this null hypothesis can be tested with the usual statistic,

$$\hat{W}(i, j) = (g(\hat{\boldsymbol{\phi}}(i,j)) - g_0)'(\hat{G}'\hat{\boldsymbol{\Omega}}(i,j)\hat{G})^{-1} \times (g(\hat{\boldsymbol{\phi}}(i,j)) - g_0) \xrightarrow{d} \chi^2_k, \tag{5}$$

where $\hat{G}$ denotes the Jacobian evaluated at $\hat{\boldsymbol{\phi}}(i, j)$. I find it convenient to construct these quantities with the help of a selector matrix, S_{ij}. This matrix is particularly convenient for programming and can be easily constructed as $S_{ij} = e_j' \otimes (I_H \otimes e_i)'$ for $i, j = 1, \ldots, r$ where e_m is the mth column of I_r for $m = i, j$. Accordingly, $\hat{\boldsymbol{\phi}}(i, j) = S_{ij}\hat{\boldsymbol{\Phi}}_T$; $\hat{\Omega}(i, j) = S_{ij}\hat{\Omega}_\phi S_{ij}'$, and assume $\hat{\Omega}_\phi \xrightarrow{p} \Omega_\phi$.

For example, an accurate assessment of the statistical significance of the (i, j) impulse response would consist of setting $g(\hat{\boldsymbol{\phi}}(i, j)) = \hat{\boldsymbol{\phi}}(i, j)$, $g_0 = \mathbf{0}_{H\times 1}$, and reporting the p-value of the resulting $\hat{W}(i, j)$ statistic. This p-value gives a precise probabilistic answer to this specific joint null of significance. However, one of the attractions of an impulse response analysis is the ability to graphically convey statistical information about the dynamic behavior of the system under consideration. The natural counterpart to the marginal, rectangular, confidence intervals associated with the individual nulls in (4) is the simultaneous $100(1 - \alpha)\%$ confidence region given by

$$\Pr[\hat{W}(i,j) \leq c_\alpha^2] = 1 - \alpha, \tag{6}$$

where c_α^2 is the critical value of a χ^2_H distributed random variable. Unfortunately, the resulting confidence region is a multidimensional ellipsoid that cannot be easily displayed in two-dimensional space.

A. *Scheffé's S-Method of Simultaneous Inference*

One solution to this problem is to construct a simultaneous rectangular region with Scheffé's (1953) S-method of simultaneous inference. The intuition behind this method is to exploit the Cauchy-Schwarz inequality to transform the Wald statistic in (5) from L_2 metric into L_1 metric. Let me explain this more carefully.

I find it convenient to begin the exposition of this technique with a simpler, but less general, motivating example. Suppose the elements of $\hat{\boldsymbol{\phi}}(i, j)$ were uncorrelated with one another, so that $\Omega(i, j)$ would be a diagonal matrix. The null hypothesis of joint significance $H_0 : \boldsymbol{\phi}(i, j) = \mathbf{0}$ for any i, $j = 1, \ldots, r$ can then be tested with the statistic

$$\hat{W}(i, j) = \hat{\boldsymbol{\phi}}(i,j)'\hat{\Omega}(i,j)^{-1}\hat{\boldsymbol{\phi}}(i,j) \xrightarrow{d} \chi^2_H.$$

Since $\Omega(i, j)$ is diagonal (in this example) with entries $\sigma_h^2(i, j)$ for $h = 1, \ldots, H$ and $i, j = 1, \ldots, r$, it is easy to see that

$$\hat{W}(i, j) = \sum_{h=1}^{H} \hat{t}_h^2(i, j);\ \hat{t}_h(i, j) \xrightarrow{d} N(0, 1),$$

so that the associated simultaneous confidence region becomes

$$\Pr[\hat{W}(i,j) \leq c_\alpha^2] = \Pr\left[\sum_{h=1}^{H} \hat{t}_h^2(i,j) \leq c_\alpha^2\right].$$

Scheffé's (1953) S-method consists in realizing that a direct consequence of Bowden's (1970) lemma is that

$$\max\left\{\frac{\left|\sum_{h=1}^{H} \dfrac{\hat{t}_h(i,j)}{h}\right|}{\sqrt{\sum_{h=1}^{H} \dfrac{1}{h^2}}};\ |h| < \infty\right\} = \sqrt{\sum_{h=1}^{H} \hat{t}_h^2(i,j)},$$

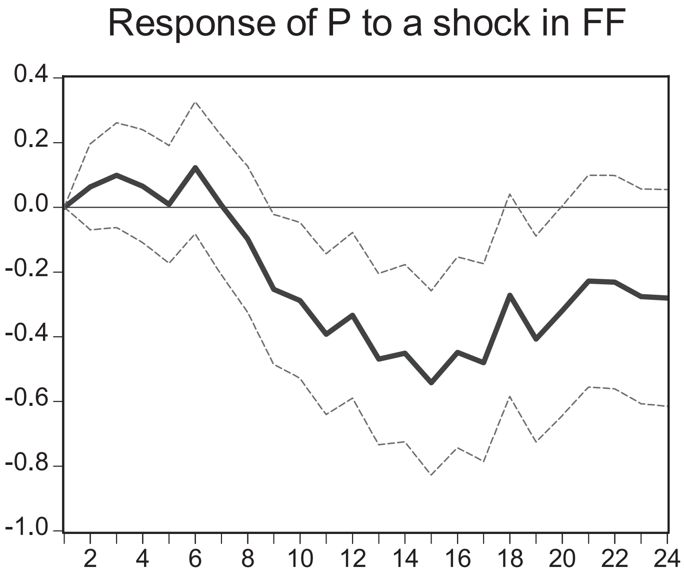
which implies that

$$\Pr\left[\left|\sum_{h=1}^{H} \frac{\hat{t}_h(i,j)}{h}\right| \leq \sqrt{\frac{c_\alpha^2}{H}}\right] \simeq \Pr\left[\sum_{h=1}^{H} \hat{t}_h^2(i,j) \leq c_\alpha^2\right] = 1 - \alpha. \tag{7}$$

I will denote the bands resulting from the first term in expression (7) as Scheffé bands to differentiate them from traditional error bands reported in the VAR literature.

Geometric arguments help clarify the method further. Suppose we consider a standardized impulse response function over two periods so that its mean is centered at zero, its variances are normalized to 1, and its covariances are 0. The plot of an $100(1 - \alpha)\%$ confidence region is clearly seen to be a circle whose radius is c_α and whose center is (0,0). Now consider moving from the center of this circle along the main diagonal by shifting each coefficient by an equal

FIGURE 1.—CORRELATION AMONG IMPULSE RESPONSE COEFFICIENTS: RESPONSE OF INFLATION TO A SHOCK IN THE FEDERAL FUNDS RATE



Correlation Matrix: Asymptotic Critical Value of Significance Is 0.15

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	0.52	0.34	0.27	0.37	0.22	0.21	0.14	0.20	0.10	0.13	0.08	0.09	0.07	0.11	0.13	0.11	0.07	0.08	0.08	0.04	0.04	0.03	0.08
2	0.52	1	0.61	0.43	0.45	0.46	0.33	0.26	0.23	0.22	0.13	0.15	0.10	0.09	0.11	0.14	0.14	0.09	0.08	0.10	0.08	0.05	0.05	0.05
3	0.34	0.61	1	0.64	0.52	0.51	0.50	0.36	0.31	0.24	0.23	0.15	0.15	0.10	0.11	0.13	0.15	0.13	0.10	0.09	0.10	0.09	0.06	0.06
4	0.27	0.43	0.64	1	0.67	0.55	0.54	0.52	0.39	0.32	0.26	0.24	0.16	0.16	0.12	0.13	0.14	0.15	0.14	0.11	0.09	0.10	0.09	0.07
5	0.37	0.45	0.52	0.67	1	0.70	0.59	0.54	0.54	0.38	0.31	0.26	0.23	0.15	0.17	0.14	0.14	0.12	0.15	0.14	0.11	0.10	0.10	0.10
6	0.22	0.46	0.51	0.55	0.70	1	0.73	0.60	0.56	0.53	0.37	0.32	0.26	0.22	0.16	0.18	0.14	0.12	0.12	0.15	0.14	0.12	0.11	0.11
7	0.21	0.33	0.50	0.54	0.59	0.73	1	0.74	0.62	0.56	0.52	0.39	0.33	0.26	0.24	0.18	0.19	0.13	0.13	0.13	0.15	0.15	0.13	0.12
8	0.14	0.26	0.36	0.52	0.54	0.60	0.74	1	0.75	0.64	0.57	0.54	0.41	0.35	0.29	0.27	0.20	0.20	0.15	0.15	0.14	0.17	0.17	0.15
9	0.20	0.23	0.31	0.39	0.54	0.56	0.62	0.75	1	0.76	0.66	0.60	0.56	0.44	0.39	0.33	0.30	0.22	0.23	0.18	0.17	0.17	0.19	0.20
10	0.10	0.22	0.24	0.32	0.38	0.53	0.56	0.64	0.76	1	0.77	0.68	0.62	0.59	0.47	0.42	0.35	0.32	0.25	0.25	0.20	0.19	0.19	0.22
11	0.13	0.13	0.23	0.26	0.31	0.37	0.52	0.57	0.66	0.77	1	0.78	0.69	0.64	0.61	0.50	0.45	0.38	0.35	0.27	0.27	0.22	0.21	0.22
12	0.08	0.15	0.15	0.24	0.26	0.32	0.39	0.54	0.60	0.68	0.78	1	0.79	0.71	0.66	0.63	0.51	0.46	0.40	0.36	0.29	0.29	0.24	0.24
13	0.09	0.10	0.15	0.16	0.23	0.26	0.33	0.41	0.56	0.62	0.69	0.79	1	0.80	0.73	0.68	0.65	0.53	0.48	0.42	0.37	0.31	0.31	0.27
14	0.07	0.09	0.10	0.16	0.15	0.22	0.26	0.35	0.44	0.59	0.64	0.71	0.80	1	0.82	0.75	0.69	0.66	0.55	0.50	0.43	0.39	0.33	0.33
15	0.11	0.11	0.11	0.12	0.17	0.16	0.24	0.29	0.39	0.47	0.61	0.66	0.73	0.82	1	0.83	0.75	0.70	0.67	0.56	0.50	0.44	0.40	0.35
16	0.13	0.14	0.13	0.13	0.14	0.18	0.18	0.27	0.33	0.42	0.50	0.63	0.68	0.75	0.83	1	0.83	0.76	0.71	0.68	0.57	0.51	0.45	0.43
17	0.11	0.14	0.15	0.14	0.14	0.14	0.19	0.20	0.30	0.35	0.45	0.51	0.65	0.69	0.75	0.83	1	0.83	0.77	0.72	0.68	0.57	0.52	0.47
18	0.07	0.09	0.13	0.15	0.12	0.12	0.13	0.20	0.22	0.32	0.38	0.46	0.53	0.66	0.70	0.76	0.83	1	0.84	0.77	0.72	0.68	0.57	0.53
19	0.08	0.08	0.10	0.14	0.15	0.12	0.13	0.15	0.23	0.25	0.35	0.40	0.48	0.55	0.67	0.71	0.77	0.84	1	0.84	0.77	0.72	0.68	0.58
20	0.08	0.10	0.09	0.11	0.14	0.15	0.13	0.15	0.18	0.25	0.27	0.36	0.42	0.50	0.56	0.68	0.72	0.77	0.84	1	0.84	0.77	0.72	0.69
21	0.04	0.08	0.10	0.09	0.11	0.14	0.15	0.14	0.17	0.20	0.27	0.29	0.37	0.43	0.50	0.57	0.68	0.72	0.77	0.84	1	0.84	0.78	0.72
22	0.04	0.05	0.09	0.10	0.10	0.12	0.15	0.17	0.17	0.19	0.22	0.29	0.31	0.39	0.44	0.51	0.57	0.68	0.72	0.77	0.84	1	0.85	0.78
23	0.03	0.05	0.06	0.09	0.10	0.11	0.13	0.17	0.19	0.19	0.21	0.24	0.31	0.33	0.40	0.45	0.52	0.57	0.68	0.72	0.78	0.85	1	0.85
24	0.08	0.05	0.06	0.07	0.10	0.11	0.12	0.15	0.20	0.22	0.22	0.24	0.27	0.33	0.35	0.43	0.47	0.53	0.58	0.69	0.72	0.78	0.85	1

Note: (Top) Impulse response calculated by local projections with six lags for the Stock and Watson (2001) system. Traditional marginal two, standard-error bands displayed. (Bottom) Correlation matrix of impulse response coefficients. Asymptotic critical value of significance is 0.15.

amount $\pm\delta$ along the 45 degree line (this shift is proportional to the variance of each estimate because of the standardization). What would δ need to be in order for the shift to lie on the boundary of the $100(1 - \alpha)\%$ confidence circle where it intersects the 45 degree line? Clearly, δ meets the condition

$$\Pr[\delta^2 + \delta^2 = c_\alpha^2] = 1 - \alpha,$$

from where $\delta = \sqrt{c_\alpha^2/2}$, which is exactly the critical value suggested by Scheffé's method.

B. *Dealing with Impulse Response Serial Correlation*

In general, the elements of $\boldsymbol{\phi}(i, j)$ are serially correlated. Figure 1 provides an example with the response of inflation to a shock in the federal funds rate in the Stock and Watson (2001) VAR displayed in the top panel and its associated correlation matrix displayed in the bottom panel. Since the asymptotic 95% confidence level critical value for these correlations is 0.15, it is clear that most of its entries are significantly different from 0. From a different angle, the correlation between impulse response coefficients is easy to compute analytically for a simple autoregressive model of order 1 (AR(1)) example with autoregressive parameter $|\rho| < 1$. In that case, notice that $\phi_h = \rho^h$ and cov $(\phi_1, \phi_h) = \rho^{(h-1)}(1 - \rho^2)$.

A natural solution to this serial correlation problem is to orthogonalize the impulse response path. As I discussed in section I, Sims and Zha (1999) proposed decomposing $\Omega(i, j)$ into its first few principal components, but this has serious drawbacks. Instead, I propose orthogonalizing the impulse response path by projecting $\hat{\phi}_h(i, j)$ on to $\hat{\phi}_{h-1}(i, j), \ldots, \hat{\phi}_1(i, j)$ for $h = 1, \ldots, H$ so that each impulse response at time h is transformed into the hth impulse response conditional on its past path from 1 to $h - 1$. Mechanically, the method is easier to understand by appealing to the Wald principle and the Cholesky decomposition,

$$\hat{\Omega}(i, j) = \hat{A}_{ij}\hat{D}_{ij}\hat{A}'_{ij};\ i, j = 1, \ldots, r,$$

where A_{ij} is lower triangular with 1s in the main diagonal and D_{ij} is a diagonal matrix. Notice that unlike applications of the Cholesky decomposition for structural identification in VARs, here there is no ambiguity about the ordering: the impulse response traces how a shock is propagated into the future, and I use this propagation over time as the ordering principle to construct the projections. It turns out that this method provides another interesting statistical interpretation that I now discuss.

Recall that the null hypothesis of joint significance H_0 : $\boldsymbol{\phi}(i, j) = \mathbf{0}_{H\times 1}$ for $i, j = 1, \ldots, r$ can be tested with the Wald statistic

$$\hat{W}(i, j) = \hat{\boldsymbol{\phi}}(i, j)'\hat{\Omega}(i, j)^{-1}\hat{\boldsymbol{\phi}}(i, j) \xrightarrow{d} \chi^2_H,$$

with $\hat{\Omega}(i, j) \xrightarrow{p} \Omega(i, j)$. With the help of the Cholesky decomposition, this statistic can be recast as

$$\begin{aligned}\hat{W}(i, j) &= \hat{\boldsymbol{\phi}}(i, j)'(\hat{A}_{ij}\hat{D}_{ij}\hat{A}'_{ij})^{-1}\hat{\boldsymbol{\phi}}(i, j)\\ &= (\hat{A}_{ij}^{-1}\hat{\boldsymbol{\phi}}(i, j))'\hat{D}_{ij}^{-1}(\hat{A}_{ij}^{-1}\hat{\boldsymbol{\phi}}(i, j)).\end{aligned}$$

Defining

$$\hat{\psi}_h(i, j) = E_P[\hat{\phi}_h(i, j)|\hat{\phi}_{h-1}(i, j), \ldots, \hat{\phi}_0(i, j)];$$
$$i, j = 1, \ldots, r;\ h = 1, \ldots, H,$$

where E_P is the linear projection operator and noticing that $\hat{D}_{ij}$ is a diagonal matrix whose entries correspond to the variances of $\hat{\psi}_h(i, j)$, say, $\tilde{\sigma}^2_h(i, j)$, then

$$\hat{W}(i, j) = \sum_{h=1}^{H}\left(\frac{\hat{\psi}_h(i, j)}{\tilde{\sigma}_h(i, j)}\right)^2 = \sum_{h=1}^{H}\hat{t}^2_{h|h-1,\ldots,1}(i, j) \xrightarrow{d} \chi^2_H \quad (8)$$

$$\hat{t}_{h|h-1,\ldots,1}(i, j) \xrightarrow{d} N(0, 1) \quad (9)$$

where $\hat{t}_{h|h-1,\ldots,1}(i, j)$ is the t-ratio of the conditional impulse response coefficient at time h given the impulse response path up to time $h - 1$. It is now apparent why this choice of orthogonalization is convenient: it transforms the decision problem of testing the joint null of significance of correlated impulse response coefficients into the sum of the t-statistics of the individual nulls of significance of the conditional impulse response coefficients.

C. *Error Bands for Impulse Responses*

Expressions (7) and (8) lead us to the construction of two visual devices with which to report the characteristics of the joint distribution of the impulse response path. The first device is a direct application of Scheffé's (1953) method to expression (7) in order to construct Scheffé bands with approximate simultaneous $100(1 - \alpha)$%-confidence coverage. These bands are simply

$$\hat{\boldsymbol{\phi}}(i, j) \pm \hat{A}_{ij}\hat{D}_{ij}^{1/2}\sqrt{\frac{c^2_\alpha}{H}}\,\mathbf{i}_H, \quad (10)$$

where $\mathbf{i}_H$ is an $H \times 1$ vector of ones. The reader whose habit is to estimate impulse responses with a VAR with conventional statistical software will have no difficulty in constructing Scheffé bands with expression (10), since all that is required is the Cholesky decomposition of $\hat{\Omega}(i, j)$. It is important to remark that because of serial correlation, the area inside the Scheffé bands contains impulse response trajectories with less than an α probability of being observed. For this reason, it may be helpful to use a fan chart consisting of plots of the Scheffé bands in expression (10) for different values of α.

Finally, notice that individual conditional (not marginal) rectangular regions can be constructed for each impulse response coefficient by noticing that

$$\Pr[|t_{h|h-1,\ldots,1}(i, j)| \leq z_{\alpha/2}] = 1 - \alpha.$$

The resulting conditional uncertainty for each impulse response coefficient can be summarized by constructing the bands as

$$\hat{\boldsymbol{\phi}}(i, j) \pm z_{\alpha/2}\ diag\ (\hat{D}_{ij}^{1/2}) \quad (11)$$

where $diag\ (\hat{D}_{ij}^{1/2})$ is an $H \times 1$ vector with the diagonal entries of $\hat{D}_{ij}^{1/2}$. Like Scheffé bands, conditional bands can be trivially constructed from available econometric software output and with expression (11). Conditional bands can be useful to pinpoint individual impulse response coefficients. The t-ratios on which they are based represent the individual contribution of each coefficient in the test of joint significance, as can be easily appreciated from expression (8). Further, their variability represents the variability associated with the hth period once variability from the correlation with previous $(h - 1)$ periods has been sterilized.

Let me return to the AR(1) example with autoregressive parameter $|\rho| < 1$ to clarify the relation between marginal, Scheffé, and conditional bands. Specifically, the impulse response over three periods for this AR(1) is

$$\Phi(1, 3) = \begin{bmatrix} \rho \\ \rho^2 \\ \rho^3 \end{bmatrix}; \ \Omega_\phi = (1 - \rho^2)$$
$$\times \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & (1 + \rho^2) & \rho(1 + \rho^2) \\ \rho^2 & \rho(1 + \rho^2) & (1 + \rho^2 + \rho^4) \end{bmatrix},$$

where the variance of the least squares estimate of ρ, say $\hat{\rho}$, is asymptotically $(1 - \rho^2)$. In practice, we would substitute the population moments for sample counterparts, of course. The traditional marginal error bands are

$$\begin{bmatrix} \hat{\rho} \\ \hat{\rho}^2 \\ \hat{\rho}^3 \end{bmatrix} \pm z_{\alpha/2}(1 - \rho^2)^{1/2} \begin{bmatrix} 1 \\ (1 + \rho^2)^{1/2} \\ (1 + \rho^2 + \rho^4)^{1/2} \end{bmatrix}.$$

Instead, the Scheffé bands can be easily obtained by direct application of expression (10) and are readily seen to be

$$\begin{bmatrix} \hat{\rho} \\ \hat{\rho}^2 \\ \hat{\rho}^3 \end{bmatrix} \pm \sqrt{\frac{c_\alpha^2}{3}}(1 - \rho^2)^{1/2} \begin{bmatrix} 1 \\ (1 + \rho) \\ (1 + \rho + \rho^2) \end{bmatrix},$$

where even if $\sqrt{\frac{c_\alpha^2}{3}}$ were equal to $z_{\alpha/2}$, it is clear that in general, $(1 + \rho + \ldots + \rho^h) \neq (1 + \rho^2 + \ldots + \rho^{2h})^{1/2}$ and serves to illustrate the differences in coverage between marginal and Scheffé bands.

The conditional bands that result from applying expression (11) are instead

$$\begin{bmatrix} \hat{\rho} \\ \hat{\rho}^2 \\ \hat{\rho}^3 \end{bmatrix} \pm z_{\alpha/2}(1 - \rho^2)^{1/2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The intuition for the conditional bands comes from realizing that in this model, $\hat{\phi}_h = \hat{\phi}_{h-1}\hat{\rho}$, that is, given $\hat{\phi}_{h-1}$, the value of the impulse response in the next period is determined by the propagation forward of the impulse response through the AR(1) process. This propagation entails a scaling of the impulse response by $\hat{\rho}$, and the uncertainty associated with this estimate is $(1 - \rho^2)^{1/2}$, which is precisely the asymptotic standard deviation from the usual least squares estimate of the AR(1) parameter.

D. *Practitioner's Corner*

Scheffé bands provide a rectangular approximation to the elliptical simultaneous confidence region implied by the Wald principle and hence provide an approximation of the set of possible impulse response trajectories for a given probability level. A fan chart based on Scheffé bands evaluated at different confidence levels may be useful in sorting out what particular shapes the impulse responses can have given their serial correlation: 95% Scheffé bands alone require the reader to be disciplined enough to avoid erroneously assuming that any path between the bands has a 5% or better chance of being the true path.

Conditional error bands distill the variability of each impulse response estimate from the variability caused by the serial correlation with previous impulse response coefficients. Conditional bands therefore give a better sense about the significance of individual impulse response coefficients in a manner consistent with the joint null of significance. As we saw in expression (8), their sum of squares is precisely the usual chi-square test. When the impulse response function is naturally uncorrelated, marginal and conditional bands are equivalent.

Scheffé bands and conditional bands can be constructed with existing statistical software. Most researchers accustomed to estimating impulse responses with VARs need only request the covariance matrix from their software and then apply the appropriate formulas (expressions 10 and 11) to construct the desired bands. This simplicity means that the methods described here can be immediately applied by practitioners. In later sections, asymptotic results with analytic expressions for local projections and for VARs under a variety of scenarios are provided for completeness.

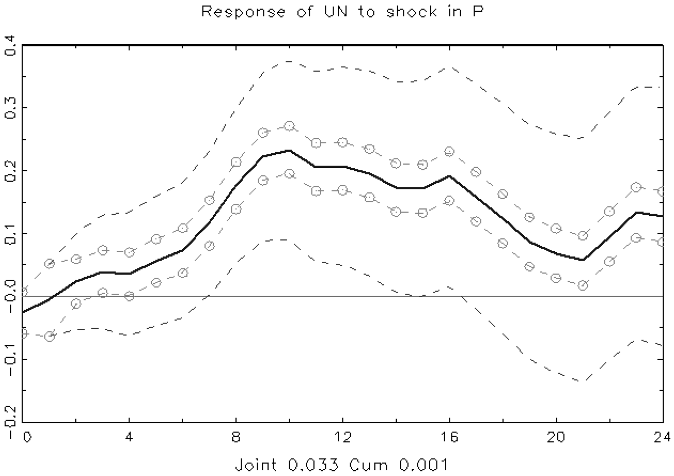
Concern about the validity of the asymptotic approximations in an application with a small sample suggests the use of resampling techniques instead. Kilian (1998a, 1999) and Lütkepohl (2005) discuss several ways of implementing the bootstrap in practice. Here I showcase two natural approaches for the purposes of illustration, but not as a comprehensive guide, since this would require an entirely new paper.

Suppose that a sample of B bootstrap estimates $\hat{\boldsymbol{\Phi}}_T^b$ for $b = 1, \ldots, B$ is available. This can be the result of applying the bootstrap to a VAR or to local projections; the end result is independent of the estimation method. Then there are two choices. One is to rely on the asymptotic formulas provided below to construct b estimates of $\hat{\Omega}_\Phi^b$ and hence $\hat{\Omega}^b(i, j)$. Then the Cholesky decomposition of each $\hat{\Omega}^b(i, j)$ can be used to construct a b resample of the Scheffé bands, the conditional bands, and the fan charts. Under this option, the bootstrap provides any pivotal statistics (such as the conditional bands) an asymptotic refinement. The second option does not rely on asymptotic expressions (and hence does not provide asymptotic refinements but may behave better in small samples) for the covariance matrix and instead uses a bootstrap-sample-based estimate of $\Omega(i, j)$, say $\tilde{\Omega}^B(i, j)$, whose typical entry will be

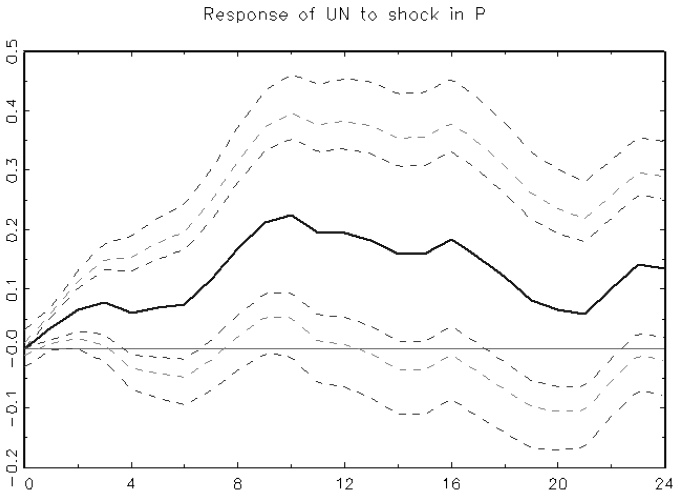
$$cov\,(\hat{\phi}_h^B(i, j); \hat{\phi}_g^B(i, j))$$
$$= \frac{\sum_{b=1}^B \left(\hat{\phi}_h^b(i, j) - \frac{\sum_{b=1}^B \hat{\phi}_h^b(i, j)}{B}\right)\left(\hat{\phi}_g^b(i, j) - \frac{\sum_{b=1}^B \hat{\phi}_g^b(i, j)}{B}\right)}{B - 1}$$

for $h, g = 1, \ldots, H$; $i, j = 1, \ldots, r$. Given $\tilde{\Omega}^B(i, j)$, its Cholesky decomposition can then be used to construct

FIGURE 2.—95% CONDITIONAL ERROR BANDS FOR RESPONSE OF UNEMPLOYMENT TO INFLATION SHOCK



Note: Dashed lines are traditional two marginal standard error bands. Dashed lines with circles are two conditional standard error bands. "Joint 0.033" refers to the p-value of the null hypothesis that all the response coefficients are jointly zero. "Cum 0.001" is the p-value of the null that the accumulated impulse response after 24 periods is equal to 0. Impulse response calculated by local projections with six lags on the Stock and Watson (2001) system.



Note: Percentile bounds for 95th, 25th, and 1st percentiles of the Wald test of joint significance. Impulse response calculated by local projections with six lags on the Stock and Watson (2001) system.

Scheffé and conditional bands with expressions (10) and (11).

A fan chart for different α levels can be constructed by ranking the B resamples of each impulse response function with the auxiliary Wald metric,

$$\tilde{W}^b(i, j) = (\hat{\boldsymbol{\phi}}^b(i,j) - \hat{\boldsymbol{\phi}}(i,j))'\tilde{\Omega}^B(i,j)^{-1}(\hat{\boldsymbol{\phi}}^b(i,j) - \hat{\boldsymbol{\phi}}(i,j));\ b = 1, \ldots, B;\ i, j = 1, \ldots, r.$$

This metric ranks each bootstrap resample according to its Wald distance to the sample estimate. Then a fan chart can be constructed with the bootstrap sample paths corresponding to each desired percentile of the ranking of $\tilde{W}^b(i, j)$.

As an example, figure 2 displays the response of the U.S. unemployment rate (UN) to a shock in inflation (P) in the Stock and Watson (2001) VAR and serves to illustrate what Scheffé and conditional bands look like in practice. The top panel displays the estimated impulse response along with conventional (marginal) two standard error bands (the wider bands in the figure) and the just introduced conditional, two standard error bands (the narrower bands). The bottom values in that panel refer to the p-value of the joint significance test (Joint 0.033) and the p-value of the significance test of the accumulated response after 24 periods (Cum 0.001). The bottom panel displays a Scheffé fan chart of the same impulse response.

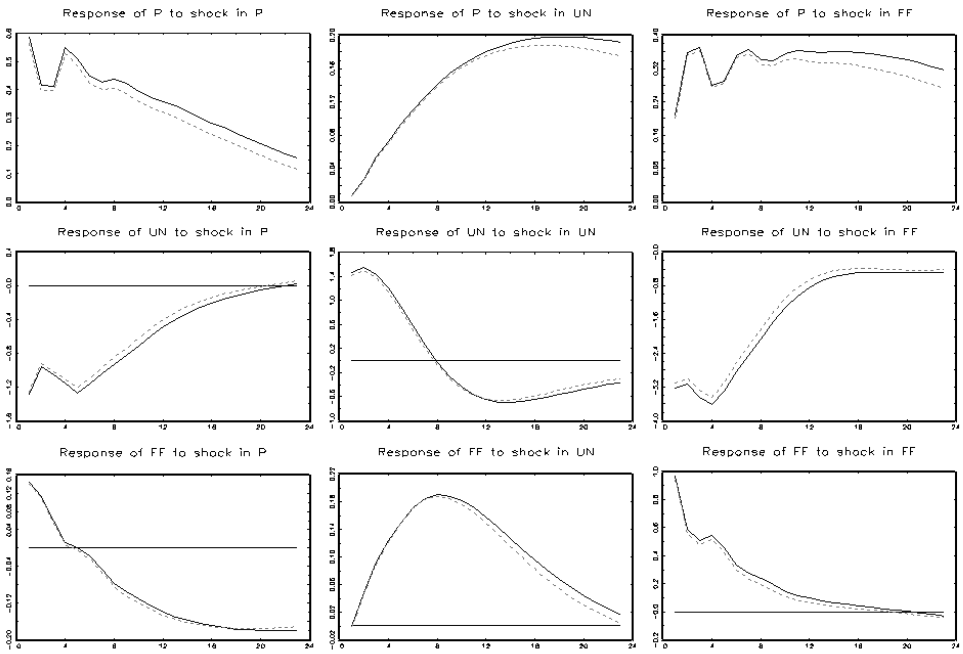
The impulse response displayed in the top panel is emblematic of the VAR literature: the width of the marginal error bands is often taken as evidence that there is little information in the sample about the relationship between unemployment and prices (the marginal error bands include 0 for all but 6 out of the 24 periods displayed). In fact the opposite conclusions are true. The p-value of the joint tests of significance and cumulative significance (0.033 and 0.001, respectively) leaves little doubt that the response is significantly different from 0. The impulse response does not wiggle around the zero line as, say, the plot of a white noise process would; instead, it is decidedly positive over all but 1 of the 24 periods displayed. The conditional error bands suggest that although the response is mostly indistinguishable from 0 during the first three to five quarters after impact, it is distinctly positive thereafter and for the duration of the remaining periods, suggesting that the effect on unemployment is delayed by about four periods. The fan chart in the bottom panel of figure 2 further confirms the high degree of serial correlation in the impulse response estimates and the improbability that the impulse response would fluctuate widely between positive and negative values, which one would have expected by looking at the usual marginal error bands.

E. *Small Sample Monte Carlo Experiments*

A full-scale investigation of the small sample properties of Scheffé and conditional bands deserves far more space than is available here. Instead, by using as background the Stock and Watson (2001) VAR, I hope to remove arbitrariness in the selection of DGP and provide evidence for a situation that is likely to be typical in many practical applications. This section provides Monte Carlo evidence based on simulations generated with the Stock and Watson (2001) VAR, where I begin by estimating the model over the expanded sample (1960:I to 2007:I, 188 observations) using four lags and saving the resulting coefficient estimates to then generate 1,000 samples of observations of varying length (T = 50, 100, 200, and 400) that are always initialized with the first few observations in the original sample.

For each replication, I estimate a VAR(4); compute its impulse responses over 6, 12, and 24 horizons (and thus

FIGURE 3.—TRUE AND MONTE CARLO AVERAGE ESTIMATES OF STOCK AND WATSON'S (2001) BASED IMPULSE RESPONSES



Dotted lines are the true impulse responses generated with the Stock and Watson (2001) DGP. Dashed lines are the Monte Carlo averages over 1,000 impulse response estimates based on VARs estimated on samples of 200 synthetic observations.

generate nine responses for each replication and each horizon choice); and then compute the covariance matrix for each of the nine impulse responses (with the Monte Carlo replications to abstract from distortions generated by computational method). Figure 3 displays the theoretical structural impulse responses over 24 periods and the Monte Carlo average for models estimated with a sample of $T = 200$ observations (approximately the size of the original sample). This figure is directly comparable to figure 1 in Stock and Watson (2001) (except that I do not display the $h = 0$ coefficients) and serves to illustrate both that the impulse responses I generate are virtually identical to those in the Stock and Watson (2001) VAR and that Monte Carlo estimation biases are negligible except perhaps at very long horizons (notice that here, 24 periods are six years of quarterly data).

The first set of experiments is collected in tables 1, 2 and 3 and compares the joint probability coverage of Scheffé and traditional marginal bands for probability levels 50%, 68% (one standard deviation), and 95% (two standard deviations). The comparisons are made along the joint probability coverage implied by the Wald statistic in expression (8). For each impulse response, I calculate this Wald statistic and then report the proportion of impulse responses that fall within the cut-off values implied by the Scheffé and the marginal bands.

Several results deserve comment. First, it is very clear from the tables that the marginal bands provide inadequate probability coverage regardless of the sample size or the length of the impulse response (and it is particularly poor for the 50% and 68% probability levels). The differences are less dramatic for 95% nominal coverage, and in some cases the effective coverage of the marginal bands is close to the nominal level. However, even then, the differences in coverage from one impulse response to the other can vary quite significantly. On the other hand, and even though Scheffé bands are meant to be approximate, the effective coverage is very close to nominal for impulse response horizons 6 and 12, even for relatively small samples. At 24 horizons, there appear to be some distortions in the coverage (85% effective versus 95% nominal) that are perhaps to be expected (see Pesavento & Rossi, 2006; Sims & Zha, 1999).

The second set of experiments is reported in tables 4, 5, and 6. Because of space considerations, I have omitted some of the combinations of sample size and impulse response horizon, which are available from me on request. The experiments report the empirical power and size (by simulating a white noise process) of the individual coefficient significance tests implied by the marginal and he conditional bands. The bottom row of each table reports the power of the joint significance test (which is the same regardless of the type of band and is therefore repeated to make it easier to read) as a guide to the results and for completeness. The reader may also wish to glance back at figure 3 to look at the impulse responses themselves.

The tables clearly indicate that conditional bands have superior power to marginal bands. For example, impulse response (3,1), which corresponds to the response of the federal funds rate to a price shock, would not appear to be significant if one had a sample $T = 50$ and the marginal bands. Such a conclusion would seem to be consistent with the joint chi-square test reported at the bottom of that

TABLE 1.—PROBABILITY COVERAGE COMPARISON: MARGINAL VERSUS SCHEFFÉ BANDS
IMPULSE RESPONSE HORIZONS: 6

$T = 50$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	2%	2%	2%	3%	2%	3%	4%	2%	2%	51%	52%	54%	53%	51%	52%	55%	53%	51%
68%	10%	9%	7%	13%	11%	11%	17%	10%	11%	71%	69%	71%	68%	70%	69%	69%	71%	69%
95%	71%	64%	63%	73%	68%	64%	75%	69%	68%	94%	94%	94%	94%	94%	93%	92%	93%	94%
$T = 100$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	2%	4%	2%	2%	2%	1%	2%	2%	2%	53%	51%	49%	52%	53%	55%	53%	53%	50%
68%	9%	13%	9%	14%	9%	8%	11%	11%	9%	69%	69%	69%	69%	69%	70%	69%	69%	68%
95%	69%	71%	63%	75%	67%	63%	71%	65%	65%	94%	95%	94%	94%	94%	93%	95%	94%	95%
$T = 200$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	3%	2%	1%	2%	3%	1%	3%	1%	1%	51%	51%	49%	49%	52%	52%	51%	51%	52%
68%	11%	9%	7%	9%	11%	8%	13%	6%	7%	68%	69%	68%	70%	68%	69%	69%	69%	70%
95%	71%	69%	60%	72%	70%	63%	73%	61%	66%	94%	95%	95%	94%	94%	95%	94%	95%	94%
$T = 400$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	2%	1%	1%	2%	2%	1%	3%	1%	1%	52%	51%	52%	50%	52%	51%	51%	52%	50%
68%	10%	9%	7%	10%	10%	7%	10%	6%	9%	68%	68%	68%	67%	69%	68%	68%	68%	68%
95%	72%	66%	61%	67%	71%	62%	71%	61%	63%	94%	96%	94%	96%	94%	94%	95%	94%	94%

Notes: DGP is Stock and Watson (2001) three variable, VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the impulse response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. Probability coverage calculated as the percentage of Monte Carlo impulse responses (derived from VAR estimates at each replication) within the Wald statistic implied by the bands considered. 1,000 Monte Carlo replications. Contemporaneous responses are excluded.

TABLE 2.—PROBABILITY COVERAGE COMPARISON: MARGINAL VERSUS SCHEFFÉ BANDS
IMPULSE RESPONSE HORIZONS: 12

$T = 50$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	1%	0%	1%	1%	0%	1%	1%	0%	64%	60%	62%	61%	61%	61%	64%	60%	60%
68%	1%	8%	2%	4%	11%	3%	6%	9%	4%	74%	71%	71%	71%	72%	70%	72%	70%	71%
95%	43%	69%	38%	46%	77%	48%	56%	66%	47%	90%	89%	88%	87%	88%	88%	87%	87%	89%
$T = 100$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	1%	0%	0%	1%	0%	0%	2%	0%	64%	60%	60%	59%	61%	60%	60%	60%	62%
68%	3%	10%	2%	4%	12%	1%	3%	12%	2%	73%	71%	70%	71%	72%	71%	72%	72%	72%
95%	51%	73%	46%	56%	82%	43%	48%	76%	48%	89%	89%	88%	89%	89%	89%	88%	89%	89%
$T = 200$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	1%	0%	0%	1%	0%	0%	3%	0%	61%	59%	58%	59%	59%	57%	62%	59%	60%
68%	5%	12%	1%	3%	13%	1%	4%	22%	2%	71%	70%	70%	73%	70%	71%	72%	70%	70%
95%	63%	83%	49%	59%	84%	44%	56%	88%	48%	90%	89%	90%	90%	89%	90%	90%	88%	89%
$T = 400$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	3%	0%	0%	0%	0%	1%	5%	0%	60%	56%	56%	59%	58%	55%	61%	58%	58%
68%	5%	24%	1%	3%	13%	1%	4%	33%	2%	73%	70%	70%	70%	69%	70%	73%	69%	71%
95%	71%	96%	45%	61%	88%	38%	59%	96%	51%	90%	91%	91%	91%	90%	92%	91%	89%	90%

Notes: DGP is Stock and Watson (2001) three variable, VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the impulse response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. Probability coverage calculated as the percentage of Monte Carlo impulse responses (derived from VAR estimates at each replication) within the Wald statistic implied by the bands considered. 1,000 Monte Carlo replications. Contemporaneous responses are excluded.

TABLE 3.—PROBABILITY COVERAGE COMPARISON: MARGINAL VERSUS SCHEFFÉ BANDS
IMPULSE RESPONSE HORIZONS: 24

$T = 50$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	0%	0%	1%	1%	0%	1%	1%	0%	75%	72%	72%	72%	71%	71%	73%	69%	71%
68%	2%	6%	2%	3%	8%	3%	5%	8%	2%	79%	76%	78%	77%	77%	75%	78%	74%	78%
95%	43%	58%	34%	35%	60%	37%	49%	56%	36%	87%	85%	86%	85%	84%	84%	84%	84%	85%
$T = 100$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	1%	0%	1%	0%	0%	0%	1%	0%	73%	71%	70%	72%	73%	73%	73%	74%	73%
68%	3%	8%	2%	3%	8%	1%	2%	10%	1%	77%	76%	76%	76%	78%	78%	78%	78%	78%
95%	52%	65%	44%	55%	67%	39%	43%	68%	39%	86%	86%	84%	85%	87%	87%	85%	86%	86%
$T = 200$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	8%	0%	0%	37%	1%	0%	8%	0%	72%	70%	70%	71%	70%	70%	73%	72%	72%
68%	8%	38%	6%	3%	74%	12%	6%	40%	2%	77%	75%	76%	76%	75%	75%	77%	77%	77%%
95%	66%	91%	66%	56%	98%	78%	60%	89%	48%	85%	85%	85%	85%	85%	86%	86%	85%	86%
$T = 400$	Marginal									Scheffé								
Nom	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
50%	0%	70%	1%	0%	93%	41%	1%	50%	0%	72%	69%	69%	72%	68%	69%	72%	69%	71%
68%	9%	93%	16%	4%	99%	81%	9%	82%	7%	78%	74%	75%	78%	73%	75%	78%	75%	77%
95%	79%	100%	85%	68%	100%	99%	72%	99%	72%	87%	85%	87%	86%	84%	87%	87%	84%	87%

Notes: DGP is Stock and Watson (2001) three variable, VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the impulse response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. Probability coverage calculated as the percentage of Monte Carlo impulse responses (derived from VAR estimates at each replication) within the Wald statistic implied by the bands considered. 1,000 Monte Carlo replications. Contemporaneous responses are excluded.

impulse response, whose power is 0.17. However, the conditional bands explain both of these results: while there is some ambiguity about the early horizons of the impulse response, these are clearly nonzero over time. The larger sample $T = 200$ provides further confirmation of this suspicion.

Summarizing, the Monte Carlo evidence suggests that the new methods introduced provide adequate guides to questions about the shape of the impulse response and the individual significance of coefficients, with some distortions in the coverage for impulse response horizons that are long relative to the lag structure of the system and the sample size. In any case, the new methods appear to be far more reliable than traditional marginal bands for the specific questions that the new methods were designed to answer.

III. Anchoring Structural Identification with Testable Restrictions

Structural identification in an impulse response exercise is commonly achieved by imposing restrictions on the contemporaneous conditional covariance matrix of the system's variables. Often such restrictions just-identify the system (such as when identification is achieved with a Wald causal ordering and the ubiquitous Cholesky decomposition), meaning that enough zero restrictions are introduced to achieve identification, but not enough to formally test the implied restrictions. Exceptions to this tradition are papers that use overidentifying restrictions or the papers by Granger and Swanson (1997) and Demiralp and Hoover (2003), who develop formal statistical methods based on graph theory.

It seems natural that if economic theory is brought forward to achieve identification through the nature of contemporaneous correlations, it should also be used to anchor the impulse response exercise with implied restrictions on the impulse response paths themselves. Examples of such constraints may include zero impulse response restrictions, restrictions that the impulse response path of a certain variable is strictly positive (or negative), or, in general, any linear constraint on the coefficients of a subset of the system's impulse responses.

The assumption that the vector of impulse responses is at least approximately multivariate Gaussian can be particularly useful here. Suppose economic theory constrains the way variable k responds to a shock to variable l, $k, l \in \{1, \ldots, r\}$. Call the resulting constrained path $\boldsymbol{\phi}^c(k, l)$. Then, assumption (1) and standard properties of linear projections for multivariate Gaussian random variables can be used to show that for any $i, j \in \{1, \ldots, r\}$ except the pair (k, l),

$$\hat{\boldsymbol{\phi}}(i, j)|\boldsymbol{\phi}^c(k, l) = \hat{\boldsymbol{\phi}}(i, j) + S_{ij}\hat{\Omega}_{\phi}S'_{kl}\hat{\Omega}(k, l)^{-1}(\boldsymbol{\phi}^c(k, l) - \hat{\boldsymbol{\phi}}(k, l)),$$

TABLE 4.—POWER AND SIZE COMPARISON: MARGINAL VERSUS CONDITIONAL BANDS
IMPULSE RESPONSE HORIZONS: 6

$T = 50$	Marginal										Conditional									
H	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size
1	0.62	0.05	0.16	0.47	1.00	0.89	0.09	0.05	0.92	0.05	0.62	0.05	0.16	0.47	1.00	0.89	0.09	0.05	0.92	0.05
2	0.26	0.05	0.28	0.22	0.99	0.68	0.05	0.10	0.20	0.06	0.38	0.40	0.57	0.29	1.00	0.88	0.10	0.42	0.44	0.06
3	0.22	0.06	0.22	0.25	0.88	0.71	0.06	0.16	0.10	0.05	0.37	0.52	0.61	0.37	0.99	0.94	0.12	0.60	0.38	0.05
4	0.45	0.08	0.13	0.30	0.61	0.62	0.06	0.20	0.11	0.05	0.69	0.56	0.58	0.60	0.96	0.96	0.13	0.78	0.50	0.05
5	0.32	0.10	0.13	0.32	0.27	0.44	0.06	0.25	0.07	0.05	0.71	0.72	0.73	0.70	0.88	0.91	0.32	0.89	0.58	0.06
χ^2	0.50	0.11	0.16	0.30	1.00	0.74	0.12	0.18	0.94		0.50	0.11	0.16	0.30	1.00	0.74	0.12	0.18	0.94	

$T = 100$	Marginal										Conditional									
H	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size
1	0.99	0.06	0.31	0.80	1.00	1.00	0.16	0.05	1.00	0.04	0.99	0.06	0.31	0.80	1.00	1.00	0.16	0.05	1.00	0.04
2	0.79	0.08	0.58	0.43	1.00	0.99	0.10	0.14	0.72	0.06	0.89	0.43	0.85	0.51	1.00	1.00	0.15	0.44	0.89	0.06
3	0.73	0.13	0.52	0.54	1.00	0.99	0.06	0.24	0.40	0.05	0.86	0.61	0.84	0.66	1.00	1.00	0.14	0.75	0.76	0.05
4	0.96	0.18	0.36	0.61	0.95	0.96	0.06	0.39	0.47	0.05	0.99	0.71	0.77	0.87	1.00	1.00	0.23	0.92	0.91	0.05
5	0.87	0.25	0.35	0.61	0.65	0.88	0.04	0.58	0.31	0.05	0.99	0.89	0.89	0.92	0.99	1.00	0.44	0.98	0.88	0.05
χ^2	0.99	0.25	0.39	0.67	1.00	1.00	0.15	0.53	1.00		0.99	0.25	0.39	0.67	1.00	1.00	0.15	0.53	1.00	

$T = 200$	Marginal										Conditional									
H	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size
1	1.00	0.06	0.58	0.98	1.00	1.00	0.30	0.05	1.00	0.05	1.00	0.06	0.58	0.98	1.00	1.00	0.30	0.05	1.00	0.05
2	0.99	0.11	0.89	0.69	1.00	1.00	0.18	0.25	0.99	0.05	1.00	0.48	0.98	0.76	1.00	1.00	0.24	0.61	1.00	0.06
3	0.99	0.24	0.81	0.83	1.00	1.00	0.07	0.46	0.87	0.05	1.00	0.72	0.97	0.93	1.00	1.00	0.15	0.85	0.98	0.05
4	1.00	0.41	0.64	0.88	1.00	1.00	0.04	0.66	0.90	0.06	1.00	0.89	0.93	0.99	1.00	1.00	0.27	0.97	1.00	0.06
5	1.00	0.59	0.66	0.89	0.96	1.00	0.05	0.86	0.78	0.06	1.00	0.98	0.97	0.99	1.00	1.00	0.51	1.00	0.99	0.06
χ^2	1.00	0.63	0.76	0.93	1.00	1.00	0.38	0.93	1.00		1.00	0.63	0.76	0.93	1.00	1.00	0.38	0.93	1.00	

Notes: DGP is Stock and Watson (2001) three variable VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. The column "Size" controls the size of the test for an impulse response that is theoretically zero at all horizons. Power calculated as 1 minus the proportion of impulse response coefficients inside the corresponding marginal or conditional 95% confidence interval at each horizon. Hence, theoretical size is 5%. 1,000 Monte Carlo replications. Contemporaneous responses are excluded (marginal and conditional bands are the same).

TABLE 5.—POWER AND SIZE COMPARISON: MARGINAL VERSUS CONDITIONAL BANDS
IMPULSE RESPONSE HORIZONS: 24

$T = 50$	Marginal										Conditional									
H	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size
1	0.62	0.05	0.16	0.47	1.00	0.89	0.09	0.05	0.92	0.05	0.62	0.05	0.16	0.47	1.00	0.89	0.09	0.05	0.92	0.05
2	0.26	0.05	0.28	0.22	0.99	0.68	0.05	0.10	0.20	0.06	0.38	0.40	0.57	0.29	1.00	0.88	0.10	0.42	0.44	0.06
3	0.22	0.06	0.22	0.25	0.88	0.71	0.06	0.16	0.10	0.05	0.37	0.52	0.61	0.37	0.99	0.94	0.12	0.60	0.38	0.05
4	0.45	0.08	0.13	0.30	0.61	0.62	0.06	0.20	0.11	0.05	0.69	0.56	0.58	0.60	0.96	0.96	0.13	0.78	0.50	0.05
5	0.32	0.10	0.13	0.32	0.27	0.44	0.06	0.25	0.07	0.05	0.71	0.72	0.73	0.70	0.88	0.91	0.32	0.89	0.58	0.06
6	0.24	0.12	0.14	0.23	0.10	0.29	0.06	0.29	0.04	0.07	0.68	0.79	0.79	0.72	0.78	0.87	0.46	0.94	0.60	0.07
7	0.19	0.16	0.15	0.18	0.05	0.21	0.07	0.28	0.05	0.06	0.68	0.85	0.81	0.73	0.76	0.85	0.56	0.94	0.66	0.06
8	0.19	0.18	0.13	0.15	0.10	0.14	0.07	0.25	0.05	0.07	0.74	0.90	0.83	0.72	0.84	0.83	0.63	0.95	0.73	0.08
9	0.16	0.21	0.11	0.12	0.17	0.11	0.08	0.21	0.05	0.06	0.76	0.92	0.85	0.73	0.91	0.82	0.70	0.94	0.77	0.08
10	0.14	0.23	0.11	0.10	0.23	0.08	0.09	0.17	0.06	0.07	0.75	0.94	0.85	0.76	0.94	0.83	0.73	0.96	0.81	0.09
11	0.11	0.22	0.10	0.09	0.24	0.07	0.08	0.13	0.06	0.06	0.74	0.96	0.89	0.76	0.96	0.82	0.76	0.93	0.82	0.08
12	0.11	0.22	0.09	0.07	0.24	0.06	0.08	0.10	0.06	0.06	0.76	0.96	0.88	0.78	0.96	0.86	0.79	0.93	0.85	0.09
13	0.08	0.19	0.08	0.07	0.21	0.05	0.08	0.08	0.05	0.06	0.77	0.97	0.89	0.80	0.96	0.87	0.83	0.94	0.84	0.08
14	0.08	0.18	0.08	0.06	0.17	0.05	0.07	0.07	0.05	0.06	0.77	0.97	0.90	0.83	0.97	0.89	0.83	0.94	0.86	0.09
15	0.07	0.17	0.07	0.06	0.15	0.05	0.07	0.07	0.05	0.06	0.76	0.98	0.90	0.83	0.96	0.91	0.84	0.95	0.89	0.07
16	0.07	0.16	0.07	0.06	0.12	0.06	0.07	0.06	0.05	0.05	0.81	0.97	0.90	0.84	0.97	0.91	0.86	0.95	0.89	0.09
17	0.06	0.15	0.07	0.06	0.09	0.06	0.06	0.06	0.05	0.06	0.80	0.97	0.89	0.86	0.97	0.92	0.86	0.94	0.90	0.08
18	0.06	0.14	0.06	0.05	0.08	0.05	0.06	0.06	0.06	0.05	0.82	0.97	0.90	0.86	0.97	0.93	0.86	0.95	0.91	0.09
19	0.05	0.12	0.06	0.05	0.07	0.05	0.06	0.06	0.06	0.05	0.85	0.97	0.92	0.87	0.97	0.94	0.88	0.96	0.92	0.07
20	0.05	0.11	0.06	0.05	0.06	0.05	0.06	0.06	0.05	0.05	0.86	0.96	0.91	0.89	0.97	0.95	0.88	0.96	0.92	0.09
21	0.05	0.10	0.06	0.04	0.06	0.05	0.05	0.05	0.05	0.05	0.86	0.98	0.92	0.89	0.98	0.94	0.88	0.96	0.92	0.08
22	0.05	0.10	0.05	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.88	0.98	0.93	0.89	0.98	0.94	0.87	0.97	0.93	0.06
23	0.05	0.08	0.05	0.04	0.06	0.05	0.05	0.05	0.05	0.04	0.89	0.98	0.94	0.91	0.97	0.95	0.89	0.97	0.93	0.08
χ^2	0.21	0.17	0.16	0.20	0.83	0.33	0.17	0.19	0.54		0.21	0.17	0.16	0.20	0.83	0.33	0.17	0.19	0.54	

Notes: DGP is Stock and Watson (2001) three variable VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. The column "Size" controls the size of the test for an impulse response that is theoretically zero at all horizons. Power calculated as 1 minus the proportion of impulse response coefficients inside the corresponding marginal or conditional 95% confidence interval at each horizon. Hence, theoretical size is 5%. Last row (labeled χ^2) is the power of the joint significance test. 1,000 Monte Carlo replications. Contemporaneous responses are excluded (marginal and conditional bands are the same).

TABLE 6.—POWER AND SIZE COMPARISON: MARGINAL VERSUS CONDITIONAL BANDS
IMPULSE RESPONSE HORIZONS: 24

$T = 200$	Marginal										Conditional									
H	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Size
1	1.00	0.06	0.58	0.98	1.00	1.00	0.30	0.05	1.00	0.05	1.00	0.06	0.58	0.98	1.00	1.00	0.30	0.05	1.00	0.05
2	0.99	0.11	0.89	0.69	1.00	1.00	0.18	0.25	0.99	0.05	1.00	0.48	0.98	0.76	1.00	1.00	0.24	0.61	1.00	0.06
3	0.99	0.24	0.81	0.83	1.00	1.00	0.07	0.46	0.87	0.05	1.00	0.72	0.97	0.93	1.00	1.00	0.15	0.85	0.98	0.05
4	1.00	0.41	0.64	0.88	1.00	1.00	0.04	0.66	0.90	0.06	1.00	0.89	0.93	0.99	1.00	1.00	0.27	0.97	1.00	0.06
5	1.00	0.59	0.66	0.89	0.96	1.00	0.05	0.86	0.78	0.06	1.00	0.98	0.97	0.99	1.00	1.00	0.51	1.00	0.99	0.06
6	0.99	0.76	0.78	0.76	0.55	0.97	0.06	0.96	0.43	0.06	1.00	1.00	0.99	0.98	0.98	1.00	0.73	1.00	0.96	0.06
7	0.97	0.88	0.78	0.60	0.11	0.87	0.09	0.98	0.27	0.06	1.00	1.00	0.99	0.99	0.92	1.00	0.83	1.00	0.92	0.06
8	0.97	0.92	0.70	0.46	0.08	0.71	0.14	0.98	0.21	0.08	1.00	1.00	0.99	0.97	0.95	1.00	0.89	1.00	0.92	0.08
9	0.93	0.94	0.70	0.35	0.35	0.48	0.19	0.98	0.14	0.06	1.00	1.00	0.99	0.97	0.99	0.99	0.94	1.00	0.93	0.07
10	0.86	0.96	0.71	0.25	0.59	0.29	0.22	0.96	0.09	0.07	1.00	1.00	0.99	0.97	1.00	0.99	0.96	1.00	0.95	0.10
11	0.81	0.97	0.71	0.18	0.74	0.20	0.26	0.91	0.07	0.07	1.00	1.00	1.00	0.97	1.00	0.98	0.97	1.00	0.97	0.08
12	0.75	0.98	0.69	0.13	0.77	0.15	0.29	0.82	0.06	0.06	1.00	1.00	1.00	0.98	1.00	0.98	0.99	1.00	0.97	0.07
13	0.68	0.98	0.67	0.10	0.78	0.12	0.33	0.69	0.05	0.05	1.00	1.00	1.00	0.98	1.00	0.99	1.00	1.00	0.98	0.08
14	0.60	0.97	0.66	0.07	0.74	0.11	0.36	0.53	0.05	0.06	1.00	1.00	1.00	0.98	1.00	0.98	1.00	1.00	0.99	0.09
15	0.53	0.97	0.63	0.06	0.68	0.11	0.38	0.39	0.05	0.06	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	0.99	0.07
16	0.45	0.96	0.61	0.06	0.61	0.11	0.40	0.26	0.05	0.05	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	0.99	0.08
17	0.38	0.94	0.58	0.06	0.52	0.12	0.40	0.18	0.05	0.05	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	0.99	0.08
18	0.32	0.91	0.56	0.05	0.45	0.13	0.41	0.13	0.05	0.05	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.08
19	0.28	0.89	0.53	0.05	0.38	0.15	0.41	0.10	0.05	0.05	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.07
20	0.23	0.87	0.50	0.05	0.32	0.16	0.42	0.07	0.06	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06
21	0.18	0.84	0.48	0.05	0.27	0.17	0.41	0.06	0.06	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08
22	0.15	0.82	0.44	0.06	0.25	0.18	0.41	0.06	0.06	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08
23	0.12	0.77	0.40	0.06	0.23	0.18	0.40	0.06	0.07	0.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.07
χ^2	1.00	0.62	0.43	0.52	1.00	1.00	0.24	0.54	1.00		1.00	0.62	0.43	0.52	1.00	1.00	0.24	0.54	1.00	

Notes: DGP is Stock and Watson (2001) three variable VAR(4) orthogonalized with the Cholesky decomposition using the same Wald order. (i, j) refers to the response of variable i to a shock in variable j, where prices = 1, unemployment = 2, and federal funds rate = 3. T is the sample size. The column "Size" controls the size of the test for an impulse response that is theoretically 0 at all horizons. Power calculated as 1 minus the proportion of impulse response coefficients inside the corresponding marginal or conditional 95% confidence interval at each horizon. Hence theoretical size is 5%. Last row (labeled χ^2) is the power of the joint significance test. 1,000 Monte Carlo replications. Contemporaneous responses are excluded (marginal and conditional bands are the same).

where S_{ij} and S_{kl} are selector matrices that have been described previously. Furthermore, the covariance matrix for $\hat{\boldsymbol{\phi}}(i, j)|\boldsymbol{\phi}^c(k, l)$ can be obtained from the same projection results as

$$\hat{\Omega}(i, j)|(k, l) = \hat{\Omega}(i, j) - (S_{ij}\hat{\Omega}_{\phi}S'_{kl})\hat{\Omega}(k, l)^{-1}(S_{kl}\hat{\Omega}_{\phi}S'_{ij}).$$

Notice that the second term is a positive semidefinite matrix so that $tr\,(\hat{\Omega}(i, j)|(k, l)) \leq tr\,(\hat{\Omega}(i, j))$, that is, the variance of $\hat{\boldsymbol{\phi}}(i, j)|\boldsymbol{\phi}^c(k, l)$ is smaller than the variance of $\hat{\boldsymbol{\phi}}(i, j)$. The reason is that the unknown path for $\boldsymbol{\phi}(k, l)$ is replaced by our assumption $\boldsymbol{\phi}^c(k, l)$. It is important to remark that it would have been just as easy to condition the exercise on any other linear restrictions and for any subset of the entire vector of impulse responses (from conditioning on the paths of two impulse responses at the same time to conditioning on the first few horizons only, and so on).

An attractive feature of this type of experiment is that its validity can be formally assessed with standard statistics. Specifically, the null hypothesis $H_0 : \boldsymbol{\phi}(k, l) = \boldsymbol{\phi}^c(k, l)$ can be tested with the Wald statistic:

$$\hat{W}^c(k, l) = (\hat{\boldsymbol{\phi}}(k, l) - \boldsymbol{\phi}^c(k, l))'\hat{\Omega}(k, l)^{-1}(\hat{\boldsymbol{\phi}}(k, l) - \boldsymbol{\phi}^c(k, l)) \xrightarrow{d} \chi^2_H. \quad (12)$$

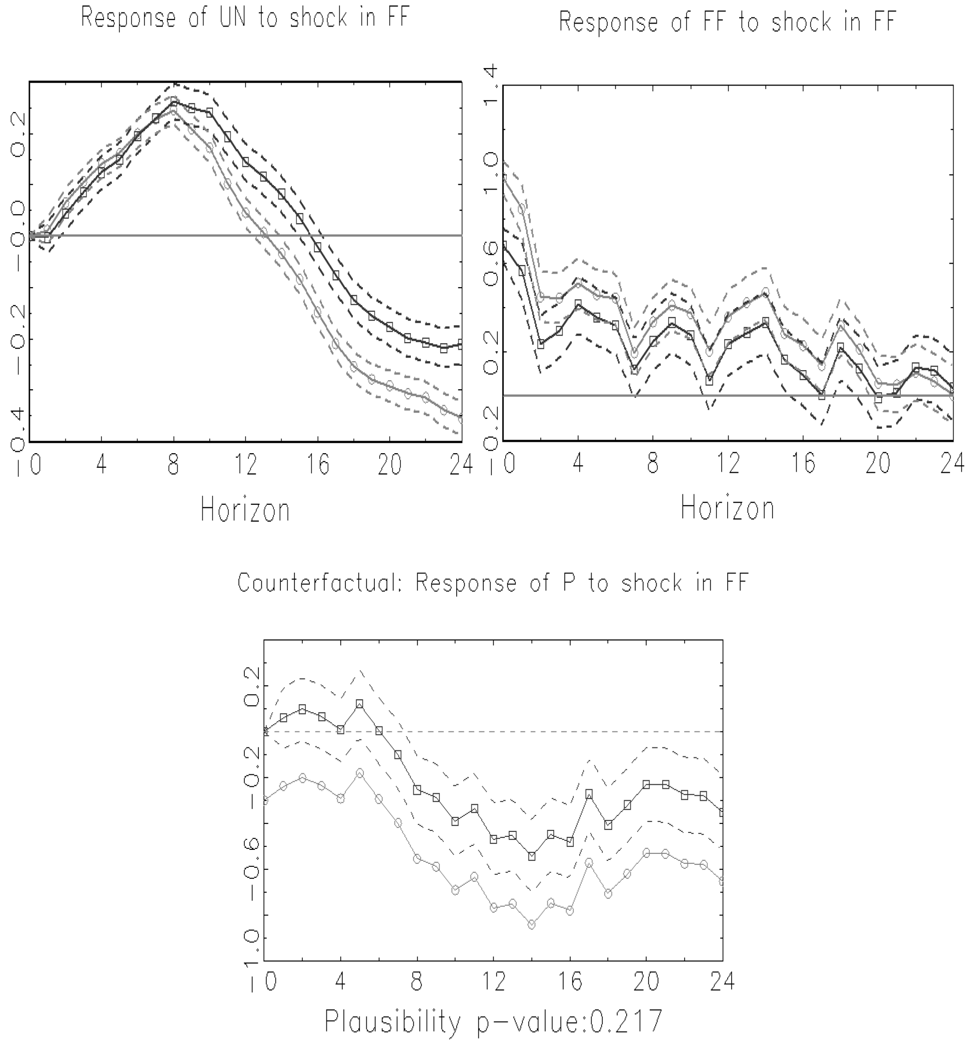
Similarly, notice that in the special case in which $S_{ij}\hat{\Omega}_{\phi}S'_{kl} = \mathbf{0}$, $\boldsymbol{\phi}(i, j)$ is independent of $\boldsymbol{\phi}(k, l)$ (under Gaussianity) and any constraint $\boldsymbol{\phi}^c(k, l)$ imposed on $\boldsymbol{\phi}(k, l)$ will not affect $\hat{\boldsymbol{\phi}}(i, j)$, a natural consequence of exogeneity. This suggests that one way to categorize which impulses are most sensitive to assumptions on $\boldsymbol{\phi}^c(k, l)$ is with a rank order of the Wald metric:

$$\hat{W}^c(i, j) = (\hat{\boldsymbol{\phi}}(i, j)|\boldsymbol{\phi}^c(k, l) - \hat{\boldsymbol{\phi}}(i, j))'\hat{\Omega}_{\phi}(i, j|k, l)^{-1} (\hat{\boldsymbol{\phi}}(i, j)|\boldsymbol{\phi}^c(k, l) - \hat{\boldsymbol{\phi}}(i, j)). \quad (13)$$

The assumption $\boldsymbol{\phi}^c(k, l)$ represents a conjecture about the true value of the unobservable $\boldsymbol{\phi}(k, l)$. The available sample of data provides us with an estimate, $\hat{\boldsymbol{\phi}}(k, l)$, and its sampling distribution. Thus, as long as $\boldsymbol{\phi}^c(k, l)$ is close to $\hat{\boldsymbol{\phi}}(k, l)$ in the statistical sense, conditioning on $\boldsymbol{\phi}^c(k, l)$ is a way of focusing the multivariate distribution of all impulse responses along the direction implied by the conditioning event $\boldsymbol{\phi}^c(k, l)$. This is different from counterfactual experimentation in the sense discussed by Lucas (1976) and, more recently, Leeper and Zha's (2003) "modest policy interventions."

It is tempting to think of $\boldsymbol{\phi}^c(k, l)$ as a counterfactual, which, if chosen to be sufficiently near $\hat{\boldsymbol{\phi}}(k, l)$, would not incur in the Lucas critique by virtue of being a trajectory that would fall within the distribution of the agents' expectations. However, to make such a statement requires careful consideration about the relation between the sampling distribution of the impulse responses and the distribution that reflects the public's and the policymaker's expectations. In

FIGURE 4.—ANCHORING EXPERIMENT: MAKING THE INITIAL RESPONSE OF PRICES TO A SHOCK IN THE FEDERAL FUNDS RATE NEGATIVE



Note: Solid lines with squares and companion dashed lines are the original impulse responses with 95% conditional confidence error bands. Solid line with circles is the counterfactual response in the bottom graph and the conditional responses given this counterfactual for the top panels with associated 95% conditional error bands. All impulse responses come from Stock and Watson's (2001) system estimated by local projections with six lags.

addition, one needs to formally determine the amount of allowable variability in the counterfactual that would not cause changes in the behavior of agents from an optimal control point of view. These are clearly important issues that require research left for another paper.

I now return to the Stock and Watson (2001) VAR to illustrate how this type of restriction can be used in practice. In particular, consider the impulse response of inflation (P) in response to a shock in the federal funds rate (FF) displayed in the bottom panel of figure 4 as the impulse response in squares (along with two standard error conditional bands). As is often the case, prices appear to respond positively to a positive shock in interest rates, at least for the first few quarters, in what is now commonly dubbed as the "price puzzle" in the monetary economics literature (e.g., Sims, 1992). Suppose now, that given our theory on how economies should behave, we impose that this response is negative starting in the first period. As an example, I have subtracted 0.25% to every coefficient in that impulse response to construct the conditioning path. The new response is represented by the line in circles in the bottom panel of figure 4.

The p-value of the Wald test measuring the distance between the conditioning event and the sample estimates (given by the Wald test in expression (12)) is 0.217, so that the restriction is clearly not rejected by the data. The original and the conditional impulse responses of unemployment (UN) and interest rates (FF) in response to an interest rate shock are plotted in the top panels of figure 4, along with their respective two conditional standard error bands. It is interesting to see that the conditional response of interest rates is shifted upward (on impact, interest rates go up by a full percentage point instead of 0.65% and remain approximately 20 basis points higher than the unconditional response throughout), whereas the response of unemployment during the first two years is approximately the same but with a much sharper decline for the conditional response thereafter. Interestingly, the conditional impulse responses that I report in figure 4 correspond rather well to the impulse responses in figure 2 of Stock and Watson (2001). However in that paper, Stock and Watson achieve these results by imposing a version of the Taylor rule on the contemporaneous structure of their VAR. Instead, the results in figure 4 combine a basic Cholesky assumption with a restriction on how prices should respond to interest rates and whose likelihood has been formally tested.

IV. Summary of Asymptotic Results

Large-sample approximations provide analytic expressions of statistics of interest with minimal assumptions and serve to justify the validity of finite-sample calculations with resampling methods (e.g., for the bootstrap, see Horowitz, 2001). This section presents asymptotic results for structural impulse responses under a variety of assumptions (summarized in the appendix) and estimation methods based on least-squares techniques. Although many of the derivations are scattered elsewhere in the literature, it is perhaps useful to catalog the main results together and compile a brief guide for practitioners here.

Recall that the $r(H+1) \times 1$ matrix $\Phi(0, H)$ collects the structural impulse responses of a system $\mathbf{y}_t$ of r variables over $H + 1$ horizons. These are constructed as $\Phi(0, H) = B(0, H)P$ where $B(0, H)$ is an $r(H+1) \times r$ matrix of reduced-form impulse responses and P is the $r \times r$ rotation matrix required for structural identification. Accordingly, I begin by discussing results for $\hat{\mathbf{b}}_T = vec\,(\hat{B}_T(0, H))$ first. Then I derive estimates of P under short-run (Cholesky) and long-run (Blanchard & Quah, 1989; Galí, 1999) recursive identification assumptions. Given $\hat{\mathbf{b}}_T$ and $\hat{P}_T$, it is straightforward to derive the distribution of $vec\,(\Phi(0, H)) = \hat{\boldsymbol{\phi}}_T = (\hat{P}'_T \otimes I)\hat{\mathbf{b}}_T$. Many of the results in this section are derived, with a little bit of work, directly from standard references such as Hamilton (1994) and Lütkepohl (2005).

A. The Reduced-Form Estimators

Let $\mathbf{y}_j$ for $j = H, \ldots, 1, 0 -1, \ldots, -K$ be the $(T - K - H) \times r$ matrix of stacked observations for the $1 \times r$ vector $\mathbf{y}'_{t+j}$. Hence, let $Y_h \equiv (\mathbf{y}_0, \ldots, \mathbf{y}_h)$ be the $(T - K - H) \times r(h + 1)$ matrix of dependent variables for any $h = 0, 1, \ldots, H$. Next, define the matrix of regressors $X_k \equiv (\mathbf{y}_{-1}, \ldots, \mathbf{y}_{-k})$, which is of dimensions $(T - K - H) \times rk$ for $k = 1, \ldots, K$. Let $\mathbf{1}_T$ denote a vector of 1s (meant for the constant term) of dimension $(T - K - H) \times 1$ and the associated projection matrix $M_1 = I_{T-K-H} - \mathbf{1}_T\mathbf{1}'_T$. Let $Z_k \equiv (\mathbf{1}_T, \mathbf{y}_{-2}, \ldots, \mathbf{y}_{-k})$ be an $(T - K - H) \times r(k - 1) + 1$ matrix for $k = 2, \ldots, K$ with associated projection matrix $M_z = I_{T-K-H} - Z_k(Z'_kZ_k)^{-1}Z'_k$ of dimensions $(T - K - H) \times (T - K - H)$.

Estimates of the reduced-form impulse response coefficients based on a VAR(K) can be obtained from the least-squares estimates,

$$\hat{A}_T \equiv \begin{bmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_K \end{bmatrix} = (X'_KM_1X_K)^{-1}(X'_KM_1Y_0)^{-1}, \tag{14}$$

and with the recursions $\hat{B}_h = \sum_{j=1}^{h} \hat{B}_{h-1}\hat{A}_j$ for $h = 1, 2, \ldots, H$ and $B_0 = I_r$ from which it is straightforward to construct $\hat{B}_T(0, H)$. I will denote with $\hat{\varepsilon}$ the $(T - K - H) \times r$ the matrix of residuals from this VAR(K), which coincide exactly with the residuals for the first local projection in expression (15) below.

Instead, impulse response coefficients can be obtained directly by local projections with the least-squares estimates,

$$\hat{B}_T(1, H) = (Y'_HM_zX_1)(X'_1M_zX_1)^{-1}, \tag{15}$$

and by setting $B_0 = I_r$. The large-sample distributions of these two estimators will depend on a set of assumptions for covariance-stationary but possibly infinite-order processes, or a set of assumptions for finite-dimensional but possibly integrated vector processes.

B. Propositions

Before stating the relevant results, it is useful to define the following auxiliary expressions:

$$\Sigma_v \equiv \Psi_B(I_{H+1} \otimes \Sigma_\varepsilon)\Psi'_B$$

$$\underset{r(H+1)\times r(H+1)}{\Psi_B} \equiv \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & \ldots & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & \ldots & \mathbf{0}_r \\ \mathbf{0}_r & B_1 & I_r & \ldots & \mathbf{0}_r \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \mathbf{0}_r & B_{H-1} & B_{H-2} & \ldots & I_r \end{bmatrix} \tag{16}$$

and notice that a consistent and asymptotically normal estimate of Σ_ε (see, e.g., proposition 15.2 in Lütkepohl, 2005) can be obtained as

$$\hat{\Sigma}_\varepsilon = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(T-K-H)}, \tag{17}$$

where $\hat{\varepsilon}$ are the least-squares residuals of the VAR(K) estimates in expression (14). The following two propositions provide the asymptotic distribution of the reduced-form impulse response coefficients.

Proposition 1. *Let $\hat{\mathbf{b}}_T^{VAR}$ denote vec $(\hat{B}_T(0, H))$ for $\hat{B}_T(0, H)$ estimated from the VAR(K) estimates in expression (14) and the recursion $\hat{B}_h = \sum_{j=1}^{h} \hat{B}_{h-j}\hat{A}_j$ for $h = 1, 2, \ldots, H$ and $B_0 = I_r$. Then under assumptions* A1–A8 *in the appendix,*

$$\sqrt{(T-K)}\,(\hat{\mathbf{b}}_T^{VAR} - \mathbf{b}_0) \xrightarrow{d} N(\mathbf{0}, \Omega_B),$$

where Ω_B can be consistently estimated with $\hat{\Omega}_B = (\hat{\Sigma}_\varepsilon^{-1} \otimes \hat{\Sigma}_v)$ based on expressions (16) and (17).

Proof. The proof is a direct result of proposition 15.4 in Lütkepohl (2005) and is mostly based on results by Lewis and Reinsel (1985) and Lütkepohl and Poskitt (1991).

Proposition 2. *Let $\hat{\mathbf{b}}_T^{LP}$ denote vec $(\hat{B}_T(0, H))$ for $\hat{B}_T(0, H)$ estimated by local projections as in expression (15). Then, under assumptions* A1–A8 *in the appendix,*

$$\sqrt{(T-K-H)}\,(\hat{\mathbf{b}}_T^{LP} - \mathbf{b}_0) \xrightarrow{d} N(\mathbf{0}, \Omega_B),$$

where Ω_B can be consistently estimated for $\hat{\Omega}_B = ((X_1'M_zX_1)^{-1} \otimes \hat{\Sigma}_v)$ based on expressions (16) and (17).

Proof. The proof is a direct consequence of theorem 3 in Jordà and Kozicki (2006) and Lewis and Reinsel (1985).

Several results deserve comment. First, the consistency of the VAR(K) coefficients A_j is guaranteed only up to lag K by the conditions that lead to proposition 1 (specifically, assumptions A7–A8). Since impulse responses estimated with a VAR(K) are $B_h = \sum_{j=1}^{h} B_{h-j}A_j$, consistency of the $\hat{B}_h$ requires that the truncation lag K be chosen to be such that $K \geq H$. Thus, while efficiency may suffer in small samples, consistency of the impulse response function suggests a preference for VAR specifications with relatively long lags. In contrast, local projection estimates of B_h require only that the residuals be approximately uncorrelated and can be specified with more parsimonious lag length choices. Second, the assumption that the ε_t are i.i.d. could be replaced by the assumption that they are instead a conditionally heteroskedastic martingale difference sequence of errors. The basic consequence of this alternative assumption would be to replace the estimate of Σ_ε with a heteroskedasticity-robust covariance estimator such as White (1980). The reader is referred to Kuersteiner (2001, 2002) and Gonçalves and Kilian (2007) for related applications.

Propositions 1 and 2 extend to systems with unit roots as follows:

Proposition 3. *Let $\hat{\mathbf{b}}_T^{VAR}$ denote vec $(\hat{B}_T(0, H))$ for $\hat{B}_T(0, H)$ estimated from the VAR(K) estimates in expression (14) and the recursion $\hat{B}_h = \sum_{j=1}^{h} \hat{B}_{h-j}\hat{A}_j$ for $h = 1, 2, \ldots, H$ and $B_0 = I_r$. Then under assumptions* A1′–A6′ *in the appendix,*

$$\sqrt{(T-K)}\,(\hat{\mathbf{b}}_T^{VAR} - \mathbf{b}_0) \xrightarrow{d} N(\mathbf{0}, \Omega_B),$$

where the i, jth block of Ω_B can be consistently estimated with $\hat{G}_i\hat{\Sigma}_\alpha\hat{G}_j'$ for $i, j = 1, \ldots, r$ with $G_i \equiv \dfrac{\partial vec\,(B_i)}{\partial vec\,(A)}$ as in proposition 3.6 in Lütkepohl (2005) and Σ_α as given in corollary 7.1.1 in Lütkepohl (2005).

Proof. The relevant proofs and discussion are contained in Hamilton (1994, chapter 18) and Lütkepohl (2005, chapter 7—more specifically corollary 7.1.1 and proposition 3.6).

Proposition 4. *Let $\hat{\mathbf{b}}_T^{LP}$ denote vec $(\hat{B}_T(0, H))$ for $\hat{B}_T(0, H)$ estimated by local projections as in expression (15). Then, under assumptions* A1′–A6′ *in the appendix,*

$$\sqrt{(T-K-H)}\,(\hat{\mathbf{b}}_T^{LP} - \mathbf{b}_0) \xrightarrow{d} N(\mathbf{0}, \Omega_B).$$

Proof. The proof is based on applying the results in Hamilton (1994, chapter 18) and by realizing that the asymptotic distribution of the impulse response coefficients is dominated by the terms converging at rate $\sqrt{T}$ so that terms converging at rate T do not affect the resulting asymptotic distribution.

Several results deserve comment. For systems with exactly r unit roots, impulse responses based on the system in the differences have the same distribution as that obtained for covariance-stationary processes under assumptions A1–A8. When there is cointegration, the only alternatives are to estimate either the vector error correction form or the system in levels (i.e., without imposing cointegrating restrictions). Propositions 3 and 4 deal with the latter case where the most important caveat is to keep in mind that Ω_B is reduced rank and that although the distribution of the $\hat{A}_j$ is asymptotically normal, the distribution of quantities based on $\sum_{j=1}^{K} \hat{A}_j$ is nonstandard (such as would be required to obtain the long-run cumulated response, for example). Rossi (2005) and Pesavento and Rossi (2006) raise an important caveat to these results that I do not discuss here. The caveat arises when the impulse response horizon is relatively long. In that case, the exponential decay of the components of the impulse response that converge at $\sqrt{T}$ relative to the decay of the components with unit roots or near-to-unit roots (which converge at rate T) will cause distortions in the Gaussian approximation.

C. *Structural Impulse Responses*

The residuals ε_t in assumptions A1–A1′ in the appendix are not assumed to be orthogonal to each other, and therefore $E(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon$ is a symmetric, positive-definite matrix with possibly nonzero entries in the off-diagonal terms. Let the structural residuals $\mathbf{u}_t$ be the rotation of the reduced-form residuals ε_t given by $P\mathbf{u}_t = \varepsilon_t$, where $E(\mathbf{u}_t\mathbf{u}_t') = I_r$, and hence $\Sigma_\varepsilon = PP'$. Notice that the decomposition of Σ_ε is not unique: Σ_ε contains $r(r+1)/2$ distinct terms, but P contains r^2 terms, and therefore $r(r-1)/2$ additional conditions are required to achieve just-identification of the terms in P. Traditional methods of estimating P consist in exogenously imposing $r(r-1)/2$, ad hoc, constraints. Two common approaches are identification via the Cholesky decomposition of Σ_ε (which is equivalent to imposing $r(r-1)/2$ zero restrictions on P) and identification with long-run restrictions that impose $r(r-1)/2$ zero restrictions on the long-run matrix of structural responses.

Short-run identification. When identification is achieved by imposing short-run identification assumptions via the Cholesky decomposition, then

$$\Omega_\phi = \frac{\partial \boldsymbol{\phi}}{\partial b}\,\Omega_B\,\frac{\partial \boldsymbol{\phi}}{\partial b'} + \frac{\partial \boldsymbol{\phi}}{\partial vec\,(P)}\,\frac{\partial vec\,(P)}{\partial vech\,(\Sigma_\varepsilon)}\,\Omega_\Sigma\,\frac{\partial vec\,(P)}{\partial vech\,(\Sigma_\varepsilon)'}\,\frac{\partial \boldsymbol{\phi}}{\partial vec\,(P)'}, \tag{18}$$

with $\Omega_\Sigma \equiv E[vech\,(\Sigma_\varepsilon)\,vech\,(\Sigma_\varepsilon)']$ and $E[b,\,vech\,(\Sigma_\varepsilon)] = 0$ since $E[X_1'M_1\varepsilon/(T-K-H)] \xrightarrow{p} 0$. Since $\Phi(0,h) = B(0,h)P$, it is easy to see that

$$\frac{\partial \boldsymbol{\phi}}{\partial b} = (P' \otimes I_{r(h+1)});\ \frac{\partial \boldsymbol{\phi}}{\partial vec\,(P)} = (I_r \otimes B(0,h)).$$

Lütkepohl (2005, chapter 3) provides the additional results,

$$\begin{aligned}\frac{\partial vec\,(P)}{\partial vech\,(\Sigma_\varepsilon)} &= L_r'\{L_r(I_{r^2}+K_{rr}) \\ &\quad\times (P \otimes I_r)L_r'\}^{-1}\sqrt{T}\,(vech\,(\hat{\Sigma}_\varepsilon) \\ &\quad- vech\,(\Sigma_\varepsilon)) \xrightarrow{d} N(0,\Omega_\Sigma) \\ \Omega_\Sigma &= 2D_r^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)D_r^{+\prime},\end{aligned} \tag{19}$$

where L_r is the elimination matrix such that for any square $r \times r$, matrix A, $vech\,(A) = L_r\,vec\,(A)$, K_{rr} is the commutation matrix such that $vec\,(A') = K_{rr}\,vec\,(A)$, and $D_r^+ = (D_r'D_r)^{-1}D_r$, where D_r is the duplication matrix such that $vec\,(A) = D_r\,vech\,(A)$ and hence $D_r^+\,vec\,(A) = vech\,(A)$. Notice that $D_r^+ = L_r$ only when A is symmetric, but does not hold for the more general case in which A is just a square (but not necessarily symmetric) matrix.

Putting together all of these results, we have,

$$\begin{aligned}&\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \xrightarrow{d} N(0,\Omega_\phi) \\ &\Omega_\phi = (P' \otimes I_{r(h+1)})\Omega_B(P \otimes I_{r(h+1)}) \\ &\quad + 2(I_r \otimes B(0,h)) \\ &CD_r^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)D_r^{+\prime}C'(I_r \otimes B(0,h)') \\ &C = L_r'\{L_r(I_{r^2}+K_{rr})(P \otimes I_r)L_r'\}^{-1},\end{aligned}$$

where in practice, $\hat{\Omega}_\phi$ can be calculated by plugging the sample estimates $\hat{B}(0,h)$, $\hat{\Omega}_B$, $\hat{P}$, and $\hat{\Sigma}_\varepsilon$ into the previous expression.

Long-run identification. The infinite order process in assumption A1 in the appendix can be rewritten, without loss of generality, as

$$\mathbf{y}_t = \sum_{j=1}^{\infty} C_j\Delta\mathbf{y}_{t-j} + C_0\mathbf{y}_{t-1} + \varepsilon_t, \tag{20}$$

with $C_i = -\Sigma_{j=1}^{\infty} A_j$ and $C_0 = \Sigma_{j=1}^{\infty} A_j$. Let $\Pi = (I - C_0)$. Then Π^{-1} is the reduced-form, long-run impact matrix. Notice that if the system has unit roots, Π is not full rank (in the case of cointegration), and it is exactly $\mathbf{0}$ if there are r unit roots in the system. Thus, when unit roots are present, care should be exercised since some recursive long-run restrictions may not actually carry any proper identifying information on the true structure in the system. This is a point often overlooked in the literature. Of course, a similar situation would arise if Σ_ε were a diagonal matrix already, and one were to impose short-run recursive assumptions to achieve identification. For these reasons, I assume that Π is full rank and briefly discuss below what happens if it is reduced rank.

If P is the structural rotation matrix such that $P\mathbf{u}_t = \varepsilon_t$, then the structural long-run impact matrix is

$$\Phi_\infty = \Pi^{-1}P.$$

Lütkepohl (2005) then shows that long-run identification assumptions can be easily imposed by applying the Cholesky decomposition to

$$\Phi_\infty\Phi_\infty' = \Pi^{-1}PP'\Pi'^{-1} = \Pi^{-1}\Sigma_\varepsilon\Pi'^{-1} = QQ', \tag{21}$$

and hence $P = \Pi Q$.

A direct estimate of Π can be easily obtained with the least-squares estimate of a truncated version of the Beveridge-Nelson decomposition of expression (20). Assuming the system is covariance-stationary, this estimate will be asymptotically normally distributed with covariance matrix, say, Ω_π.

The structural impulse responses can be constructed as

$$\hat{\Phi}(0,h) = \hat{B}(0,h)\hat{\Pi}\hat{Q}, \tag{22}$$

where the asymptotic normality of each element ensures that

$$\sqrt{T}\,(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \xrightarrow{d} N(0,\Omega_\phi),$$

but where now

$$\Omega_\phi = \frac{\partial \hat{\Phi}_T}{\partial \hat{b}_T} \Omega_B \frac{\partial \hat{\Phi}_T}{\partial \hat{b}'_T} + \frac{\partial \hat{\Phi}_T}{\partial \hat{\pi}_T} \Omega_\pi \frac{\partial \hat{\Phi}_T}{\partial \hat{\pi}'_T} + \frac{\partial \hat{\Phi}_T}{\partial \hat{q}_T} \left[\frac{\partial \hat{q}_T}{\partial \hat{\pi}_T} \Omega_\pi \frac{\partial \hat{q}_T}{\partial \hat{\pi}'_T} + \frac{\partial \hat{q}_T}{\partial vech\,(\hat{\Sigma}_\varepsilon)} \Omega_\Sigma \frac{\partial \hat{q}_T}{\partial vech\,(\hat{\Sigma}_\varepsilon)'} \right] \frac{\partial \hat{\Phi}_T}{\partial \hat{q}'_T}, \tag{23}$$

with $\hat{q}_T = vech\,(\hat{Q}_T)$. Ω_Σ is the covariance matrix of $vech\,(\hat{\Sigma}_\varepsilon)$, and we make use of the fact that $\hat{q}_T$ and $vech\,(\hat{\Sigma}_\varepsilon)$ are uncorrelated.

It is easy to see that the partial derivatives in (23) are:

- $\frac{\partial \hat{\Phi}_T}{\partial \hat{b}_T} = \hat{Q}'\hat{\Pi}' \otimes I$
- $\frac{\partial \hat{\Phi}_T}{\partial \hat{\pi}_T} = -(\hat{Q} \otimes \hat{B}(0, h))$
- $\frac{\partial \hat{\Phi}_T}{\partial \hat{q}_T} = I \otimes \hat{B}(0, h)\hat{\Pi} L'_r$
- $\frac{\partial \hat{q}_T}{\partial \hat{\pi}_T} = \{(\hat{Q} \otimes I) L'_r\}^{-1} \{\hat{\Pi}^{-1}\hat{\Sigma}_\varepsilon \otimes I\} \{\hat{\Pi}'^{-1} \otimes \hat{\Pi}^{-1}\}$
- $\frac{\partial \hat{q}_T}{\partial vech\,(\hat{\Sigma}_\varepsilon)} = \{L[\hat{\Pi} \otimes \hat{\Pi}](I_{r^2} + K_{rr})(\hat{Q} \otimes I) L'\}^{-1}.$

When Π is less than full rank but nonzero, we are dealing with a cointegrated system. The immediate consequence of this is that Π converges at rate T, and the distribution of $\hat{\Phi}_T$ is then dominated by the terms converging at rate $\sqrt{T}$ so that expression (23) simplifies considerably to essentially the expression we had for the short-run recursive case, that is,

$$\Omega_\phi = \frac{\partial \hat{\Phi}_T}{\partial \hat{b}_T} \Omega_B \frac{\partial \hat{\Phi}_T}{\partial \hat{b}'_T} + \frac{\partial \hat{\Phi}_T}{\partial \hat{q}_T} \left[\frac{\partial \hat{q}_T}{\partial vech\,(\hat{\Sigma}_\varepsilon)} \Omega_\Sigma \frac{\partial \hat{q}_T}{\partial vech\,(\hat{\Sigma}_\varepsilon)'} \right] \frac{\partial \hat{\Phi}_T}{\partial \hat{q}'_T}, \tag{24}$$

where the formulas for each of the terms in the previous expression are the same as those already derived above.

V. Conclusion

This paper introduces simultaneous confidence regions for impulse responses that summarize uncertainty about its shape and uncertainty about the individual significance of its coefficients. These results require solving two problems: construction of appropriate rectangular regions for a multiple testing problem and adjustments to the serial correlation of the response coefficients. The solutions introduced rely on Scheffé's (1953) S-method and on orthogonal linear projections. The resulting regions can be trivially constructed with available statistical software, which should facilitate their diffusion among practitioners.

Underlying these derivations are conditions under which multivariate Gaussian approximations appropriately characterize the joint distribution of the impulse response coefficients. The paper synthesizes a number of asymptotic justifications for impulse responses based on VAR estimates and introduces a number of novel results when the impulse responses are estimated by local projections (Jordà, 2005) instead. An advantage of the multivariate Gaussian approximation is that it allows anchoring of the impulse response exercise by extending assumptions on the nature of contemporaneous correlations to testable assumptions on the impulse response paths themselves.

Numerous important topics have been raised in this paper that have been postponed for future research. Impulse responses are a type of conditional forecast, so it is natural to think of extending the methods introduced here to forecasting. In ongoing work with Massimiliano Marcellino (Jordà & Marcellino, 2007), we investigate evaluation of forecast paths by adjusting for simultaneity and forecast serial correlation and introducing new measures of forecasting accuracy and predictive ability testing.

Several factors can affect the small sample properties of asymptotic approximations, and numerous papers have dealt with such issues in the context of impulse responses. An important source of such distortions is the existence of unit of near-to-unit root behavior, especially for long impulse response horizons (e.g., Rossi, 2005). Similarly, optimal ways to implement the bootstrap for VARs and calculation of impulse response marginal error bands have been extensively discussed (e.g., Kilian, 1999). Thus, it is natural to examine how these issues affect the new confidence regions and results introduced here.

Finally, developing conditions under which counterfactual experimentation with empirical models is valid would be an important contribution. Leeper and Zha (2003) have made some progress in this direction, but as the discussion in this paper makes clear, more work is needed to establish appropriate formal conditions along the lines of the anchoring exercise presented here.

REFERENCES

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

Blanchard, Olivier J., and Danny Quah, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* 79:4 (1989), 655–673.

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

Demiralp, Selva, and Kevin D. Hoover, "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics* 65:suppl. (2003), 745–767.

Ferguson, Thomas S., "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics* 29 (1958), 1046–1062.

Galí, Jordi, "Technology, Improvement and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89:1 (1999), 249–271.

Gonçalves, Silvia, and Lutz Kilian, "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity," *Econometric Reviews* 26:6 (2007), 609–641.

Granger, Clive W. J., and Norman R. Swanson, "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association* 92 (1997), 357–367.
Hamilton, James D., *Time Series Analysis* (Princeton, NJ: Princeton University Press, 1994).
Horowitz, Joel L., "The Bootstrap" (pp. 3159–3228), in James J. Heckman and Edward Leamer (eds.), *The Handbook of Econometrics,* vol. 5 (Amsterdam: North-Holland, 2001).
Jordà, Òscar, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review* 95:1 (2005), 161–182.
Jordà, Òscar, and Sharon Kozicki, "Projection Minimum Distance: An Estimator for Dynamic Macroeconomic Models," University of California, Davis, working paper 6-23 (2006).
Jordà, Òscar, and Massimiliano Marcellino, "Path Forecast Evaluation," University of California, Davis, mimeographed (2007).
Kilian, Lutz, "Small-Sample Confidence Intervals for Impulse Response Functions," this REVIEW 80:2 (1998a), 218–230.
——— "Confidence Intervals for Impulse Responses Under Departures from Normality," *Econometric Reviews* 17:1 (1998b), 1–29.
——— "Finite-Sample Properties of Percentile and Percentile-*t* Bootstrap Confidence Intervals for Impulse Responses," this REVIEW 81:4 (1999), 652–660.
——— "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," *Journal of Forecasting* 20:3 (2001), 161–179.
Kuersteiner, Guido M., "Optimal Instrumental Variables Estimation for ARMA Models," *Journal of Econometrics* 104 (2001), 359–405.
——— "Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity," *Econometric Theory* 18 (2002), 547–583.
Leeper, Eric M., and Tao Zha, "Modest Policy Interventions," *Journal of Monetary Economics* 50:8 (2003), 1673–1700.
Lehmann, E. L., and Joseph P. Romano, *Testing Statistical Hypothesis* (Berlin: Springer-Verlag, 2005).
Lewis, R. A., and Gregory C. Reinsel, "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis* 16:33 (1985), 393–411.
Lucas, Robert E. Jr., "Econometric Policy Evaluation: A Critique," *Journal of Monetary Economics,* Suppl. Ser. 1:2 (1976), 19–46.
Lütkepohl, Helmut, *New Introduction to Multiple Time Series* (Berlin: Springer-Verlag, 2005).
Lütkepohl, Helmut, and P. S. Poskitt, "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," *Econometric Theory* 7 (1991), 487–496.
Pesavento, Elena, and Barbara Rossi, "Small-Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons," *Journal of Applied Econometrics* 21 (2006), 1135–1155.
Rossi, Barbara, "Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle," *International Economic Review* 46:1 (2005), 61–92.
Scheffé, Henry, "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika* 40 (1953), 87–104.
Sims, Christopher A., "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review* 36:10 (1992), 975–1000.
Sims, Christopher A., and Tao Zha, "Error Bands for Impulse Responses," *Econometrica* 67:5 (1999), 1113–1155.
Stock, James H., and Mark W. Watson, "Vector Autoregressions," *Journal of Economic Perspectives* 15:4 (2001), 101–115.
White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48:4 (1980), 817–838.

APPENDIX

Assumptions for covariance-stationary processes

A1. *Let y_t be an $r \times 1$, covariance-stationary vector time series generated by the infinite order process*

$$\mathbf{y}_t = \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t,$$

where the constant term and other deterministic terms have been omitted for simplicity but without loss of generality.

A2. $E(\varepsilon_t) = 0$ *and ε_t are i.i.d.*

A3. $E(\varepsilon_t \varepsilon_t') = \underset{r \times r}{\Sigma_\varepsilon}$ *and finite.*

A4. $\sum_{j=1}^{\infty} \|A_j\| < \infty$ *where* $\|A_j\|^2 = tr\,(A_j' A_j)$

A5. *y_t can also be represented by the infinite moving average process,*

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j \boldsymbol{\varepsilon}_{t-j},$$

with $B_0 = I_r$ and the constant term and other deterministic terms omitted for simplicity.

A6. $\det\,(\sum_{j=0}^{\infty} B_j z^j) \neq 0$ *for* $|z| \leq 1$; *and* $\sum_{j=1}^{\infty} j^{1/2} \|B_j\| < \infty$.

A7. $K^3/T \to \infty$; $K, T \to \infty$. *K is the truncation lag of the VAR(K).*

A8. $K^{1/2} \sum_{j=K+1}^{\infty} \|A_j\| \to 0$ *as* $K, T \to \infty$.

Notice that assumption A1 includes as a special case finite-dimensional processes since for a finite value K, we can set $A_{K+j} = 0$ for $j \geq 1$. Assumption A1 imposes covariance-stationarity but below, I show that in the finite-dimensional case, the asymptotic results carry through as long as the process y_t admits a Beveridge-Nelson decomposition. Assumption A2 is more stringent than is necessary, and it can often be relaxed to accommodate general forms of heteroskedasticity. I will remark on the effect of relaxing the assumption where appropriate. Assumption A5 is really a consequence of assumptions A4 and A6 (see, e.g., Anderson, 1994). Assumptions A1 to A6 cover a wide class of models that includes the well-worn finite dimensional VAR and also finite-dimensional VARMA models and potentially other covariance-stationary processes.

Assumptions for systems with unit roots

A1′. *Let y_t be an $r \times 1$ vector time series generated by the process*

$$\mathbf{y}_t = \sum_{j=1}^{K} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t$$

with K finite, and define $C_s \equiv -[A_{s+1} + \ldots + A_K]$ for $s = 1, 2, \ldots, K-1$ and $C_0 = A_1 + \ldots + A_K$ so that the system can be expressed in the differences as

$$\Delta \mathbf{y}_t = \sum_{j=1}^{K-1} C_j \Delta \mathbf{y}_{t-j} + (C_0 - I)\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Depending on rank $(C_0 - I) = g$, we can have that if $g = 0$, then the system has exactly r unit roots and $C_0 - I = 0$; if $0 < g < r$, then the system is cointegrated; and if $g = r$, the system is stationary in the levels and we can revert back to assumptions A1 *to* A8.

A2′. *Same as assumption* A2.

A3′. *Same as assumption* A3.

A4′. *If $0 < g < r$, we can rewrite $C_0 - I = \alpha\beta'$, where α and β are $r \times g$. Let $\alpha_\perp$ and $\beta_\perp$ denote the space spanned by vectors orthogonal to the space spanned by α and β, respectively. Then assume that*

$$\alpha_\perp' \left(I_r - \sum_{j=1}^{K-1} C_j \right) \beta_\perp$$

is nonsingular. If $g = 0$, assume that all values z satisfying $|I_r - C_1 z - \ldots - C_{K-1} z^{K-1}| = 0$ lie outside the unit circle.

A5′–A6′. *Δy_t is a stationary process with infinite MA representation,*

$$\Delta \mathbf{y}_t = \sum_{j=0}^{\infty} B_j \boldsymbol{\varepsilon}_{t-j},$$

where $\sum_{j=0}^{\infty} j\|B_j\| < \infty$; $B(1) = \sum_{j=0}^{\infty} B_j \neq 0$ and, hence, y_t has a Beveridge-Nelson decomposition given by

$$\mathbf{y}_t = \mathbf{y}_0 + B(1)\sum_{j=1}^{t} \varepsilon_j + \sum_{j=0}^{\infty} B_j^{*}\boldsymbol{\varepsilon}_{t-j} + \sum_{j=0}^{\infty} B_j^{*}\varepsilon_{-j},$$

where $B_j^{*} = -\sum_{i=j+1}^{\infty} B_j$ for $j = 0, 1, \ldots$; y_0 and ε_{-j} are initial conditions; and $\sum_{j=0}^{\infty} |B_j^{*}| < \infty$.

Often the covariance-stationary assumption in A1 is violated in macroeconomic data. When the source of nonstationarity is the presence of unit roots in the system, it will turn out to have little effect on the large-sample results reported below. Impulse response estimates based on a VAR or on local projections are still asymptotically normally distributed, although the covariance matrix now has reduced rank. This has the only consequence of affecting the rates of convergence and asymptotic distribution of certain linear combinations of parameters (such as the long-run accumulated responses) but leaves all other relevant statistics unchanged. However, when the system has unit roots (or near-to-unit roots) and one considers impulse responses at long horizons, then distortions to the Gaussian approximation will result from the exponential rate of decay of the stationary components relative to the persistent components and the differing rates of asymptotic convergence. This is a subtle and important issue that was discussed in Rossi (2005), but here I operate under the assumption that the impulse response horizon is short relative to the rate of decay of the stationary components so that the elements converging at $\sqrt{T}$ will dominate the distribution for every h considered, and hence the results can be expressed in standard form.