

Jordà, Òscar

Working Paper

Inference for impulse responses

Working Paper, No. 07-7

Provided in Cooperation with:

University of California Davis, Department of Economics

Suggested Citation: Jordà, Òscar (2007) : Inference for impulse responses, Working Paper, No. 07-7, University of California, Department of Economics, Davis, CA

This Version is available at:

<https://hdl.handle.net/10419/31350>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Inference for Impulse Responses *

Abstract

Poor identification of individual impulse response coefficients does not necessarily mean that an impulse response is imprecisely estimated. This paper introduces a three-pronged approach on how to communicate uncertainty of impulse response estimates: (1) with Wald tests of joint significance; (2) with conditional t-tests of individual marginal coefficient significance; and (3) with fan charts based on the percentiles of the joint Wald statistics. The paper also shows how to anchor the impulse response analysis with a priori economic restrictions that can be formally tested and used to tighten structural identification. These methods are universal and do not depend on how the impulse responses are estimated. An empirical application illustrates the techniques in practice.

- *Keywords:* impulse response function, local projections, vector autoregressions.
- *JEL Codes:* C32, E47, C53.

Òscar Jordà
Department of Economics
U.C. Davis
One Shields Ave.
Davis, CA 95616
Phone: (530) 752 7021
e-mail: ojorda@ucdavis.edu

*The hospitality of the Federal Reserve Bank of San Francisco during the preparation of this manuscript is gratefully acknowledged. I wish to thank comments by Colin Cameron, Fabio Canova, Michele Cavallo, Jordi Galí, James Hamilton, Massimiliano Marcellino, Paul Ruud, and Aaron Smith, as well as seminar participants at the Bank of Italy, BBVA, Bilkent University, CEMFI, Federal Reserve Banks of Dallas, Philadelphia, and San Francisco, Koç University, Sabancı University, Università Bocconi - IGIER, U.C. Berkeley, and Universitat Pompeu Fabra.

1 Introduction

A symptom of poor identification is when a model's coefficient estimates are highly correlated. However, lack of identification of individual effects does not necessarily imply that the joint effect is imprecisely estimated: while t-statistics may be very “low,” an F-statistic may still be very “high.” In fact, such an observation is the canonical diagnostic of multicollinearity in standard linear regression.

An autocorrelated vector time-series will naturally have impulse response coefficients that are correlated. Thus, it is premature to conclude that impulse responses are imprecisely estimated when traditional two standard-error bands are wide. After all, these bands are the visual representation of the sequence of t-tests associated with each impulse response coefficient. Ideally, one should check the null of joint significance with a formal test to conclude whether or not the impulse response is accurately estimated.

In a perfect world, one would display the 95% confidence ellipsoid derived from the joint distribution of the impulse response function but this is clearly impossible in two-dimensional space. Sims and Zha (1999) are aware of this point and propose a principal component decomposition of the impulse response's covariance matrix as a way of parsing the information contained in the joint distribution. Unfortunately, the associated factors are difficult to interpret: as their paper shows, they can result in confidence bands that intersect, and they can provide varying probability coverage of the impulse responses depending on the application.

A different way of assessing the uncertainty associated with each impulse response coeffi-

cient is to orthogonalize each impulse response’s covariance matrix according to a triangular factorization. Unlike the common short-run identification assumptions used for structural identification of VARs, the passage of time provides a natural and unique way to project each impulse response coefficient on its past. The resulting decomposition translates the Wald statistic of the null of joint significance into the sum of the conditional t-statistics of each coefficient. Such statistics have several advantages: (1) they have a natural statistical interpretation; (2) they can be easily plotted to display appropriate conditional confidence bands; and (3) they can be used to assess the contribution of each individual coefficient toward the joint significance of the impulse response path.

Traditional two standard-error bands have another undesirable feature: they include many paths with very little chance of being observed while excluding paths that fall well inside conventional confidence levels. One way to avoid the complacency of interpreting the confidence band as if “everything inside the band goes” is to plot the contours of the distribution of the Wald statistic of joint significance for different quantiles and hence construct a fan chart bounding the possible impulse response paths associated with each probability level.

The joint distribution can also be used to anchor the impulse response analysis with testable, a priori economic restrictions. For example, the hypothesis that the response of one variable is completely exogenous to shocks in another can be used to refine the identification of the system’s impulse responses. The hypothesis can then be formally tested with a Wald test and, if it is not rejected, it can be used to tighten the uncertainty on the remaining

impulse responses considerably.

It is important to differentiate that the paper is not a discussion about how best to obtain estimates of impulse response coefficients with good small-sample properties. Whatever method of estimation the end user chooses, determines the specifics of how these ideas are implemented in practice, not their substance. For this reason, the paper compiles asymptotic results already available for VARs and complements them with results for local projections (Jordà, 2005) obtained under a variety of assumptions. Deriving asymptotic results for local projections is not only useful in providing analytic formulas for large-sample approximations, it is necessary to formally justify the validity of the bootstrap (see Horowitz, 2001). Additionally and since large-sample approximations can be poor guides in small samples, I will also very briefly discuss resampling methods by which to obtain some of the statistics that I propose although an exhaustive investigation of which alternative works best in practical situations is left for future research.

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 article 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 is based on a sample of quarterly data beginning 1960:I that I extend to 2007:I. Like them, identification 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.

As a proponent of the method of local projections and to complement the Stock and Watson (2001) analysis, I estimate impulse responses by local projections over the 24 periods they investigate. Local-projection impulse responses are similar to those reported in figure 1 of Stock and Watson (2001), albeit slightly more imprecisely estimated. This turns out to be a virtue since the techniques introduced here produce sharp answers (and with considerably simpler formulas) despite this estimation imprecision.

2 Inference for Impulse Responses

Suppose we are investigating the system of impulse responses of a vector times series \mathbf{y}_t of dimension $r \times 1$ over $h = 0, 1, \dots, H$ horizons so that

$$\Phi(0, H) = \begin{bmatrix} \Phi_0 \\ \vdots \\ \Phi_H \end{bmatrix}$$

is an $r(H+1) \times r$ matrix that collects the structural impulse response coefficients for the system. Let $\hat{\phi}_T = \text{vec}(\hat{\Phi}(0, H))$ denote impulse response estimates based on a sample of size T in $r^2(H+1) \times 1$ vector form constructed by stacking the columns of $\hat{\Phi}(0, H)$ with the vec operator. Further, suppose that we can establish the asymptotic result

$$\sqrt{T}(\hat{\phi}_T - \phi_0) \xrightarrow{d} N(0, \Omega_\phi). \quad (1)$$

The availability of such a result is all that is needed for the derivations that I introduce below. Later on, I show that one can arrive at expression (1) with standard VARs or with local projections under very general conditions. In addition to the analytic formulas that I provide, I also discuss throughout resampling methods based on the availability of $\hat{\phi}_T^s$ for $s = 1, 2, \dots, S$ resamples of size T . Naturally, there are several ways of implementing the bootstrap in practice and the reader is referred to Kilian (1998, 1999) and Lütkepohl (2005) for appropriate discussions and references.

Traditional error bands are offered as visual cues about the uncertainty of impulse responses and are commonly constructed from the marginal distribution of each individual impulse response coefficient so that, for a 95% confidence level, they are

$$\hat{\phi}_T \pm 1.96 \times \text{diag} \left(\hat{\Omega}_\phi \right)^{1/2}$$

where $\text{diag} \left(\hat{\Omega}_\phi \right)$ is the $r^2 (H + 1) \times 1$ vector of diagonal elements of $\hat{\Omega}_\phi$.

To understand why reporting impulse response uncertainty with these bands is problematic, we must begin by recognizing that impulse response coefficients are usually highly correlated. High correlation among coefficient estimates is a symptom that individual effects are poorly *identified* but not necessarily that the joint effect is *imprecisely* estimated. The situation is analogous to linear regression with colinear regressors. As an example, figure 1 displays the impulse response and two standard error bands of the response of prices (P) to a shock in the federal funds rate (FF) from the Stock and Watson (2001) VAR. The table underneath the figure reports the correlation matrix among impulse response coefficients.

Since the 95% critical value of statistical significance is 0.15, it is readily apparent that the majority of these correlations are quite high.

Naturally, a formal characterization of the uncertainty about the impulse response function requires its joint distribution yet no single statistic can convey all the features of an impulse response path. What I propose in this paper is a three pronged approach. The first prong is to report the Wald test of joint significance as an obvious summary statistic on estimation precision. Define the selector matrix $S_{ij} = e_j' \otimes (I_{H+1} \otimes e_i)'$ for $i, j = 1, \dots, r$ where e_k is an $r \times 1$ vector of zeros with a 1 in the k^{th} row for $k = i, j$. Consequently, the impulse response of variable i to a shock in variable j is $\hat{\phi}_{ij} = S_{ij} \hat{\phi}_T$ and its associated covariance matrix is $\Omega_{ij} = S_{ij} \Omega_\phi S_{ij}'$ so that the null hypothesis $H_0 : \phi_{ij} = \mathbf{0}$ can be tested with the Wald statistic

$$\widehat{W}_{ij} = \widehat{\phi}_{ij}' \widehat{\Omega}_{ij}^{-1} \widehat{\phi}_{ij} \quad (2)$$

which under expression (1) is such that $\widehat{W}_{ij} \xrightarrow{d} \chi_{H+1}^2$.

The 95% confidence ellipsoid associated with this null appropriately summarizes that which traditional two standard error bands approximately only poorly. As an example, the top panel of figure 2 displays a 95% confidence ellipse for an impulse response with only two coefficients. Traditional two standard error bands result in the box overlaid on the ellipse. Thus, when there is correlation (so that the ellipse is tilted as in figure 2), the area described by the intersection between the ellipse and the box that contains point A, reflects trajectories inside the two standard error bands with less than 5% chance of being observed. Meanwhile,

the area containing point B reflects trajectories excluded by the bands but with better than a 5% chance of being observed. The trajectories implied by points A and B are displayed in the bottom panel of figure 2 (in the more traditional impulse response graph with two standard error bands) to illustrate the point more clearly.

Sims and Zha (1999) recognized these difficulties and propose decomposing Ω_{ij} into its first few principal components. The orthogonality between these factors makes calculation of the factor's 95% confidence bands simple. However, the percentage of the variance explained by these factors will vary from application to application and hence, so will the coverage implied for the original impulse response. Additionally, the bands for some of these factors intersect (as figures 6, 11, and 12 of their paper show). Finally, the principal component decomposition does not resolve satisfactorily how one should think about the uncertainty associated with the impulse response shape as I do below.

Instead, the remaining two prongs of my approach consist, first, in realizing that the original impulse response can be orthogonalized with a Cholesky factorization that has a very natural statistical interpretation: it decomposes the null of joint significance into the sum of squared conditional t-statistics for each individual coefficient in the impulse response. This orthogonalization allows one to isolate those individual coefficients that are statistically significant from those that are not by ordering the coefficients according to the time dimension. The third prong consists in constructing quantile bounds (not bands) of the joint distribution of the impulse response coefficients to generate fan charts that summarize the set of possible impulse response trajectories associated with each probability level. The next

two subsections examine the details of these two prongs.

2.1 Conditional Confidence Bands

Expression (2) states that the Wald statistic of the null hypothesis $H_0 : \phi_{ij} = \mathbf{0}, i, j = 1, \dots, r$ is

$$\widehat{W}_{ij} = \widehat{\phi}_{ij}' \widehat{\Omega}_{ij}^{-1} \widehat{\phi}_{ij} \xrightarrow{d} \chi_{H+1}^2$$

under standard regularity conditions (made more precise below). An orthogonalization of $\widehat{\phi}_{ij}$ alternative to that in Sims and Zha (1999) is to realize that $\widehat{\Omega}_{ij}$ is positive-definite and symmetric and hence admits a Cholesky decomposition

$$\widehat{\Omega}_{ij} = \widehat{A}_{ij} \widehat{D}_{ij} \widehat{A}_{ij}'$$

where \widehat{A}_{ij} is lower triangular with ones in the main diagonal and \widehat{D}_{ij} is a diagonal matrix with entries \widehat{d}_{ij}^h for $h = 0, 1, \dots, H; i, j = 1, \dots, r$. Although, there are as many valid decompositions as there are orderings of the elements of $\widehat{\phi}_{ij}$, time provides a natural and unique sorting mechanism. In particular, notice that $\widehat{\psi}_{ij} \equiv \widehat{A}_{ij}^{-1} \widehat{\phi}_{ij}$ transforms the original impulse response, $\widehat{\phi}_{ij}$, into an orthogonal coordinate system whose covariance matrix, \widehat{D}_{ij} , is diagonal and whose elements have the statistical interpretation

$$\widehat{\psi}_{ij} = E_L \left(\phi_{ij}^h | \widehat{\phi}_{ij}^{h-1}, \dots, \widehat{\phi}_{ij}^0 \right); h = 0, 1, \dots, H,$$

that is, the linear projection of each impulse response coefficient on its past. Thus, \widehat{d}_{ij}^h is the variance of the h^{th} coefficient of the original impulse response, conditional on the historical average path $\widehat{\phi}_{ij}^{h-1}, \dots, \widehat{\phi}_{ij}^0$; $h = 0, 1, \dots, H$.

Another way to see this is to recast the Wald statistic as

$$\begin{aligned}\widehat{W}_{ij} &= \widehat{\phi}_{ij}' \widehat{\Omega}_{ij}^{-1} \widehat{\phi}_{ij} = \widehat{\phi}_{ij}' \left(\widehat{A}_{ij} \widehat{D}_{ij} \widehat{A}_{ij}' \right)^{-1} \widehat{\phi}_{ij} \\ &= \left(\widehat{A}_{ij}^{-1} \widehat{\phi}_{ij} \right)' \widehat{D}_{ij}^{-1} \left(\widehat{A}_{ij}^{-1} \widehat{\phi}_{ij} \right) \\ &= \widehat{\psi}_{ij}' \widehat{D}_{ij}^{-1} \widehat{\psi}_{ij} = \sum_{h=0}^H \left(\frac{\widehat{\psi}_{ij}^h}{\widehat{d}_{ij}^h} \right)^2 = \sum_{h=0}^H t_h^2(i, j),\end{aligned}$$

that is, the null that the impulse response is jointly significant (under Gaussianity) is equivalent to the sum of squares of the t-statistics of the null that each coefficient in the impulse response at time h is significant, given its past. The orthogonality of the $\widehat{\psi}_{ij}$ means that a 95% confidence interval for each individual coefficient can be constructed as $\widehat{\psi}_{ij}^h \pm 1.96 \widehat{d}_{ij}^h$ for $h = 0, 1, \dots, H$ and hence conditional two standard error bands can be displayed along with the original estimated impulse response $\widehat{\phi}_{ij}$, as is done in figure 3.

Figure 3 displays the response of the U.S. unemployment rate (UN) to a shock in inflation (P) in the Stock and Watson (2001) VAR. The figure 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 figure 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 impulse response displayed in figure 3 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 zero for all but six out of the 24 periods displayed). Moreover and as Stock and Watson (2001) themselves do, it is common to report one (rather than two) standard error bands (for a 68% confidence level coverage, which is a somewhat peculiar confidence level choice in statistics).

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) leave little doubt that the response is significant: the impulse response does not wiggle around the zero line like, say, the plot of a white noise process would; instead it is decidedly positive over all but one of the 24 periods displayed. The conditional error bands suggest that, although the response is mostly indistinguishable from zero during the first three to five quarters after impact, it is distinctly positive thereafter and for the duration of the remaining periods.

Instead of the analytic formulas provided, conditional error bands could be constructed with resampling methods. There are many ways to proceed and the reader is referred to Kilian, 1999 for a very informative discussion on the finite-sample properties and advantages of various ways to implement the bootstrap. Here I showcase two natural approaches for the purposes of illustration but not as a comprehensive guide. One option is, given a sample of S bootstrap estimates $\hat{\phi}_T^s$ $s = 1, \dots, S$, to compute (with the analytic formulas provided

below), S estimates of $\widehat{\Omega}_\phi^s$ and hence $\widehat{\Omega}_{ij}^s$. Then conditional error bands can be constructed from the percentiles of the Cholesky decomposition of the S estimates $\widehat{\Omega}_{ij}^s$. Another option is to rely less heavily on asymptotic results and to use the S bootstrap replicates to obtain a bootstrap-sample-based estimate of Ω_{ij} , say $\widetilde{\Omega}_{ij}$, whose Cholesky decomposition can then be used to construct the conditional confidence bands.

It is important to be clear that these conditional error bands (or graphical conditional t-tests) are informative about the individual contribution of each estimated impulse response coefficient to the overall significance of the impulse response path (given by the joint Wald statistic). However, these bands remain silent about the overall uncertainty on the impulse response shape. For that purpose, I discuss response percentile bounds and fan charts in the next subsection.

2.2 Response Percentile Bounds

Let δ_{H+1}^α denote the critical value of a chi-squared distributed random variable with $H + 1$ degrees of freedom at an α -percentile level, that is

$$P(W_{ij} \leq \delta_{H+1}^\alpha) = \alpha \quad (3)$$

Thus, the $\alpha\%$ confidence ellipsoid for $\widehat{\phi}_{ij}$ is given by all the possible vectors ϕ_{ij}^α that satisfy the equation

$$\left(\widehat{\phi}_{ij} - \phi_{ij}^\alpha\right)' \Omega_{ij}^{-1} \left(\widehat{\phi}_{ij} - \phi_{ij}^\alpha\right) = \delta_{H+1}^\alpha \quad (4)$$

When ϕ_{ij} is of dimension 1×1 , condition (4) is simply the square of the t-statistic and the impulse response estimate plus or minus 1.96 times its standard error defines a 95% confidence interval (under Gaussianity). When ϕ_{ij} involves more than one coefficient, then obviously there are infinite combinations of intervals that would satisfy expression (4). In two dimensions, the summary of all these possible combinations is displayed by the top panel of figure 2 as the 95% confidence ellipse. For more than two coefficients, a plot of the 95% confidence ellipsoid is obviously not practical. To solve this problem I consider instead a set of conservative bounds.

The intuition on how to construct these bounds is best understood by returning to figure 2. The slope of the diagonal line that goes through the origin is given by the ratio of the standard errors for each of the two coefficients displayed. The intersection of this diagonal with the α -percentile ellipse (in the figure, the usual 95%) occurs at two points: the point in the northeast (NE) quadrant represents a point in which the original coefficient estimates (the center of the ellipse) are displaced in proportion to their standard errors (as would be the case in a conventional confidence interval for a single coefficient). Similarly, the intersection in the southwest (SW) quadrant reflects a proportional shift but of a negative nature.

The coordinates of these two points are: $\left(\hat{\phi}_1 + \frac{\sqrt{\delta_2^\alpha}}{2}\sigma_{11}; \hat{\phi}_2 + \frac{\sqrt{\delta_2^\alpha}}{2}\sigma_{22}\right)$ for the NE point, and $\left(\hat{\phi}_1 - \frac{\sqrt{\delta_2^\alpha}}{2}\sigma_{11}; \hat{\phi}_2 - \frac{\sqrt{\delta_2^\alpha}}{2}\sigma_{22}\right)$ for the SW point. Since $\delta_2^{0.95} = 5.99$, then $\frac{\sqrt{\delta_2^{0.95}}}{2} = 1.73$, which is slightly smaller than the well worn value of 1.96 we are accustomed to use to construct 95% confidence intervals. The justification for the difference is that the ellipsoid considers the joint occurrence of events whereas the 1.96 value refers to the 95% confidence

level of an individual event.

In general, the simplest way to construct these bounds for any α -percentile level is to exploit the Cholesky decomposition of $\widehat{\Omega}_{ij}$ as follows. Normalize the conditional coordinate system $\widehat{\psi}_{ij}$ defined above to have unit variance as well, that is, $\widetilde{\psi}_{ij} \equiv \widehat{A}_{ij}^{-1} \widehat{D}_{ij}^{-1/2} \widehat{\phi}_{ij}$. The orthogonality and unit variance of this coordinate system simplify the calculation of the point in the α -percentile ellipse corresponding to a displacement of the origin in proportion to the standard deviations of each coefficient, which is readily seen to be

$$\widetilde{\psi}_{ij} \pm \sqrt{\frac{\delta_{H+1}^\alpha}{H+1}} \mathbf{i}_{H+1}$$

where δ_{H+1}^α is defined in (3) and \mathbf{i}_{H+1} is an $(H+1) \times 1$ vector of ones. Therefore, the α -percentile bounds for the original impulse response can be recovered as

$$\widehat{\phi}_{ij} \pm \widehat{A}_{ij} \widehat{D}_{ij}^{1/2} \sqrt{\frac{\delta_{H+1}^\alpha}{H+1}} \mathbf{i}_{H+1} \quad (5)$$

so that for different values of α , one could plot each percentile bound to form a fan chart as is done in figure 4, which plots the same impulse response as figure 3.

If resampling methods are preferred, one way to construct this percentile chart with the bootstrap is as follows. Suppose bootstrap replications $\widehat{\phi}_{ij}^s$, $s = 1, \dots, S$ are available. Then a bootstrap-based sample estimate of the covariance matrix can be readily computed, which I denoted previously as $\widetilde{\Omega}_{ij}$. Next, construct S Wald statistics as in expression (4), that is

$$\widehat{W}_{ij}^s = \left(\widehat{\phi}_{ij} - \widehat{\phi}_{ij}^s \right)' \widetilde{\Omega}_{ij}^{-1} \left(\widehat{\phi}_{ij} - \widehat{\phi}_{ij}^s \right), s = 1, \dots, S$$

which can be used to rank the S bootstrap impulse response paths according to their 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 the \widehat{W}_{ij}^s . If one prefers to rely more heavily on the asymptotic results provided below, then it is straightforward to instead construct S replications of expression (5) and then use the desired percentiles to construct the fan chart.

It is impossible to communicate all the features of a multidimensional object (the 95% ellipsoid) in two-dimensional space. The conditional error bands convey information about the significance of individual coefficients given the historical average path but remain silent about the set of possible shapes the impulse response could have. A fan chart based on α -percentile bounds gives better information on that front. Rather than presenting the 95th-percentile bounds, it is more persuasive to plot the fan chart for several percentiles to avoid the common visual complacency of thinking that any trajectory within the bounds is inside the 95% confidence ellipsoid.

3 Anchoring Impulse Responses with Testable Economic Restrictions

In any econometric model, imposing coefficient restrictions that are “true” often results in more efficient estimates. In addition, coefficient restrictions in a structural impulse response analysis can boost the structural identification of the responses. Economic interpretability is

commonly achieved by imposing restrictions on the contemporaneous conditional correlation matrix of the system's variables. Often, such restrictions “just-identify” the system (such as the common Cholesky decomposition) meaning that enough information is introduced to achieve identification, but not enough to formally test the restrictions imposed with conventional statistics.¹

It seems natural that if economic theory is brought forward to achieve contemporaneous identification that it should also be used to further anchor the impulse response exercise with implied restrictions on the impulse response paths of certain variables. 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.

Accordingly, suppose we want to anchor an impulse response exercise around an assumption on the path of variable k when it responds to a shock in variable l , for any $k, l \in \{1, \dots, r\}$. Call this assumed path $\phi_{k,l}^c$. Then, under assumption (1) and the standard properties of projections in a linear context, it is easy to see that, for any $i, j \in \{1, \dots, r\}$ except the pair $\{k, l\}$,

$$\widehat{\phi}_{ij}|\phi_{kl}^c = \widehat{\phi}_{ij} + S_{ij}\widehat{\Omega}_\phi S'_{kl} \left(S_{kl}\widehat{\Omega}_\phi S'_{kl} \right)^{-1} \left(\phi_{kl}^c - \widehat{\phi}_{kl} \right),$$

where the selector matrices S_{ij} , and S_{kl} have been defined above (i.e., $S_{mn} = e'_n \otimes (I_{H+1} \otimes e_m)'$)

¹ An exception is the work by Granger and Swanson (1997) and Demiralp and Hoover (2003). Both of these papers use graph theory to refine statistical statements about the underlying causal structure of the system.

for e_m a vector of zeros with a one in the m^{th} position and $m = i, k; n = j, l; i, j, k, l \in \{1, \dots, r\}$). Denote the variance of $\widehat{\phi}_{ij}|\phi_{kl}^c$ as $\widehat{\Omega}_{ij|kl}$ then the same projection properties used above suggest that

$$\widehat{\Omega}_{ij|kl} = S_{ij}\widehat{\Omega}_\phi S'_{ij} - \left(S_{ij}\widehat{\Omega}_\phi S'_{kl}\right) \left(S_{kl}\widehat{\Omega}_\phi S'_{kl}\right)^{-1} \left(S_{kl}\widehat{\Omega}_\phi S'_{ij}\right).$$

Notice that the first term of this expression is simply $\widehat{\Omega}_{ij}$ and that the second term is a positive semi-definite matrix so that $tr\left(\widehat{\Omega}_{ij|kl}\right) \leq tr\left(\widehat{\Omega}_{ij}\right)$, that is, the variance of $\widehat{\phi}_{ij}|\phi_{kl}^c$ is smaller than the variance of $\widehat{\phi}_{ij}$. The reason is that the unknown path for ϕ_{kl} is replaced by our assumption ϕ_{kl}^c . It is important to remark that it would have been just as easy to condition the exercise on any linear restrictions for any subset of the entire vector of impulse responses.

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

$$\widehat{W}_{kl}^c = \left(\widehat{\phi}_{kl} - \phi_{kl}^c\right)' \left(S_{kl}\widehat{\Omega}_\phi S'_{kl}\right)^{-1} \left(\widehat{\phi}_{kl} - \phi_{kl}^c\right) \xrightarrow{d} \chi_{H+1}^2 \quad (6)$$

Similarly, notice that in the special case in which $S_{ij}\widehat{\Omega}_\phi S'_{kl} = \mathbf{0}$ then ϕ_{ij} is independent of ϕ_{kl} (under Gaussianity) and any constraint ϕ_{kl}^c imposed on ϕ_{kl} will not affect $\widehat{\phi}_{ij}$, a natural consequence of exogeneity. One way to determine which impulses are most sensitive to assumptions on ϕ_{kl}^c is with the Wald statistic

$$\widehat{W}_{ij}^c = \left(\widehat{\phi}_{ij} | \phi_{kl}^c - \widehat{\phi}_{ij} \right)' \widehat{\Omega}_\phi(i, j | k, l)^{-1} \left(\widehat{\phi}_{ij} | \phi_{kl}^c - \widehat{\phi}_{ij} \right) \xrightarrow{d} \chi_{H+1}^2. \quad (7)$$

finite-sample statistics based on resampling methods can be easily applied to expression (6) and (7).

Another tempting interpretation of the role of ϕ_{kl}^c is that of a counterfactual experiment along the lines of Leeper and Zha's (2003) "modest policy interventions." Thus, \widehat{W}_{ij}^c would provide a formal measure about the "modesty" of the counterfactual ϕ_{kl}^c , which is expressed in terms of the impulse response function rather than in terms of restrictions on some of the coefficients of the VAR representation of the data. However, to justify such an interpretation requires a more careful analysis than is provided here and it is left for future research.

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 5 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. The new response is represented by the line in circles in the bottom panel of figure 5.

The p-value of the Wald test of the plausibility of this restriction (given by the Wald test in expression (6)) is 0.217 so that the restriction is 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 5 along with their respective two conditional standard error bands. It is interesting to see that the conditional response of interest rates is shifted upwards (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 5 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 5 combine a basic Cholesky assumption with a restriction on how prices should respond to interest rates and whose plausibility we formally tested.

4 Asymptotic Distribution of Impulse Responses

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 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 $\hat{\mathbf{b}}_T = \text{vec}\left(\hat{B}_T(0, H)\right)$ first. Then, I derive estimates of P under short-run (Cholesky) and long-run (Blanchard and Quah, 1989; Galí, 1999) recursive identification assumptions. Given $\hat{\mathbf{b}}_T$ and \hat{P}_T , then it is straight-forward to derive the distribution of $\text{vec}(\Phi(0, H)) = \hat{\phi}_T = \left(\hat{P}_T' \otimes I\right) \hat{\mathbf{b}}_T$. 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).

4.1 The Reduced-Form Estimators

Let \mathbf{y}_j for $j = H, \dots, 1, 0, -1, \dots, -K$ be the $(T - K - H) \times r$ matrix of stacked observations for the $1 \times r$ vector \mathbf{y}_{t+j}' . Hence, let $Y_h \equiv (\mathbf{y}_0, \dots, \mathbf{y}_h)$ be the $(T - K - H) \times rh$ matrix of dependent variables for any $h = 0, 1, \dots, H$. Next, define the matrix of regressors $X_k \equiv (\mathbf{y}_{-1}, \dots, \mathbf{y}_{-k})$, which is of dimensions $(T - K - H) \times rk$ for $k = 1, \dots, K$. Let $\mathbf{1}_T$ denote a vector of ones (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 $Z_k \equiv (\mathbf{1}_T, \mathbf{y}_{-2}, \dots, \mathbf{y}_{-k})$ be an $(T - K - H) \times r(k - 1) + 1$ matrix for $k = 2, \dots, K$ with associated projection matrix $M_z =$

$I_{T-K-H} - Z_k (Z_k' Z_k)^{-1} Z_k'$ 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

$$\hat{A}_T \equiv \begin{bmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_K \end{bmatrix} = (Y_1' M_1 X_K) (X_K' M_1 X_K)^{-1} \quad (8)$$

and with the recursions $\hat{B}_h = \sum_{j=1}^h \hat{B}_{h-1} \hat{A}_j$ for $h = 1, 2, \dots, H$ 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 (9) below.

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

$$\hat{B}_T(1, H) = (Y_H' M_z X_1) (X_1' M_z X_1)^{-1} \quad (9)$$

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.

4.2 Assumptions for Covariance-Stationary Processes

A1. Let \mathbf{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. $E(\boldsymbol{\varepsilon}_t) = 0$ and $\boldsymbol{\varepsilon}_t$ are i.i.d.

A3. $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Sigma_{\boldsymbol{\varepsilon}}$ and finite.
 $r \times r$

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

A5. \mathbf{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. $\det \left(\sum_{j=0}^{\infty} B_j z^j \right) \neq 0$ for $|z| \leq 1$; and $\sum_{j=1}^{\infty} j^{1/2} \|B_j\| < \infty$.

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

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

Notice that assumption *A1* includes as a special case finite-dimensional processes since for a finite value K , we can set $A_{K+j} = \mathbf{0}$ 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 \mathbf{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-A6* cover a wide class of models that includes the well-trodden finite dimensional VAR but also include finite-dimensional VARMA models and potentially other covariance-stationary processes.

4.3 Assumptions for Systems with Unit Roots

A1'. Let \mathbf{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 $C_s \equiv -[A_{s+1} + \dots + A_K]$ for $s = 1, 2, \dots, K-1$ and $C_0 = A_1 + \dots + A_K$ 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 $\text{rank}(C_0 - I) = g$, 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 $g = r$ then the system is stationary in the levels and we can revert back to assumptions *A1-A8*.

$A2'$. Same as assumption $A2$.

$A3'$. Same as assumption $A3$.

$A4'$. If $0 < g < r$ then we can rewrite $C_0 - I = \alpha\beta'$ where α and β are $r \times g$. Let α_\perp and β_\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 non-singular. If $g = 0$, then assume that all values z satisfying $|I_r - C_1 z - \dots - C_{K-1} z^{K-1}| = 0$ lie outside the unit circle.

$A5'$ - $A6'$. $\Delta \mathbf{y}_t$ 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 \mathbf{y}_t has a Beveridge-Nelson decomposition given by

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

where $B_j^* = -\sum_{i=j+1}^{\infty} B_i$ for $j = 0, 1, \dots$; \mathbf{y}_0 and $\boldsymbol{\varepsilon}_{-j}$ are initial conditions; and $\sum_{j=0}^{\infty} |B_j^*| < \infty$.

Often times the covariance-stationary assumption in $A1$ is violated in macroeconomic data. When the source of non-stationarity are 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.

4.4 Propositions

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

$$\begin{aligned} \Sigma_v &\equiv \Psi_B (I_{H+1} \otimes \Sigma_\varepsilon) \Psi_B' \\ \Psi_B &\equiv \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & \dots & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & \dots & \mathbf{0}_r \\ \mathbf{0}_r & B_1 & I_r & \dots & \mathbf{0}_r \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0}_r & B_{H-1} & B_{H-2} & \dots & I_r \end{bmatrix} \end{aligned} \quad (10)$$

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:

$$\hat{\Sigma}_\varepsilon = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(T - K - H)} \quad (11)$$

where $\hat{\varepsilon}$ are the least-squares residuals of the VAR(K) estimates in expression (8). The following two propositions provide the asymptotic distribution of the reduced-form impulse response coefficients.

Proposition 1 Let $\widehat{\mathbf{b}}_T^{VAR}$ denote $\text{vec}(\widehat{B}_T(0, H))$ for $\widehat{B}_T(0, H)$ estimated from the VAR(K) estimates in expression (8) and the recursion $\widehat{B}_h = \sum_{j=1}^h \widehat{B}_{h-j} \widehat{A}_j$ for $h = 1, 2, \dots, H$ and $B_0 = I_r$. Then under assumptions A1-A8

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

where Ω_B can be consistently estimated with $\widehat{\Omega}_B = \left(\widehat{\Sigma}_\varepsilon^{-1} \otimes \widehat{\Sigma}_v \right)$ based on expressions (10) and (11).

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 $\widehat{\mathbf{b}}_T^{LP}$ denote $\text{vec}(\widehat{B}_T(0, H))$ for $\widehat{B}_T(0, H)$ estimated by local projections as in expression (9). Then, under assumptions A1-A8

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

where Ω_B can be consistently estimated with $\widehat{\Omega}_B = \left((X_1' M_z X_1)^{-1} \otimes \widehat{\Sigma}_v \right)$ based on expressions (10) and (11).

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 only guaranteed 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 $B_h = \sum_{j=1}^h B_{h-j} A_j$, then consistency of the \widehat{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 only require that the residuals be approximately

uncorrelated and can be specified with more parsimonious lag length choices.² 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 heteroskedascity-robust covariance estimator such as White (1980). The reader is referred to Kuersteiner (2001, 2002) and Gonçalves and Kilian (2006) for related applications.

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

Proposition 3 *Let $\widehat{\mathbf{b}}_T^{VAR}$ denote $\text{vec}(\widehat{B}_T(0, H))$ for $\widehat{B}_T(0, H)$ estimated from the $\text{VAR}(K)$ estimates in expression (8) and the recursion $\widehat{B}_h = \sum_{j=1}^h \widehat{B}_{h-j} \widehat{A}_j$ for $h = 1, 2, \dots, H$ and $B_0 = I_r$. Then under assumptions A1'-A6'*

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

where the i, j^{th} block of Ω_B can be consistently estimated with $\widehat{G}_i \widehat{\Sigma}_\alpha \widehat{G}_j'$ for $i, j = 1, \dots, r$ with $G_i \equiv \frac{\partial \text{vec}(B_i)}{\partial \text{vec}(A)}$ 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 all contained in Hamilton (1994) chapter 18 and Lütkepohl (2005) chapter 7 (more specifically corollary 7.1.1 and proposition 3.6). ■

Proposition 4 *Let $\widehat{\mathbf{b}}_T^{LP}$ denote $\text{vec}(\widehat{B}_T(0, H))$ for $\widehat{B}_T(0, H)$ estimated by local projections as in expression (9). Then, under assumptions A1'-A6'*

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

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

² A good practical way to choose lag lengths in vector autoregressive processes is Hurvich and Tsai's (1989) AIC_c criterion. This criterion is a correction to Akaike's information criterion based on a second order expansion of the Kulback-Leibler information that is tailored to autoregressive processes and is shown to have better small sample properties than either AIC or SIC .

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 either estimate the vector-error correction form or to estimate 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 non-standard (such as would be required to obtain the long-run cumulated response, for example).

4.5 Structural Impulse Responses

The residuals ε_t in assumptions *A1-A1'* 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 non-zero 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.

4.5.1 Short-Run Identification

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

$$\Omega_\phi = \frac{\partial \phi}{\partial b} \Omega_B \frac{\partial \phi}{\partial b'} + \frac{\partial \phi}{\partial vec(P)} \frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)} \Omega_\Sigma \frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)'} \frac{\partial \phi}{\partial vec(P)'} \quad (12)$$

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 $\Phi(0, h) = B(0, h)P$ then it is easy to see that

$$\begin{aligned} \frac{\partial \phi}{\partial b} &= (P' \otimes I_{h+1}) \\ \frac{\partial \phi}{\partial vec(P)} &= (I_r \otimes B(0, h)) \end{aligned}$$

Lütkepohl (2005), chapter 3 provides the additional results

$$\begin{aligned} \frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)} &= L_r' \{L_r(I_{r^2} + K_{rr})(P \otimes I_r) L_r'\}^{-1} \\ \sqrt{T} \left(vech(\widehat{\Sigma}_\varepsilon) - vech(\Sigma_\varepsilon) \right) &\xrightarrow{d} N(0, \Omega_\Sigma) \\ \Omega_\Sigma &= 2D_r^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon) D_r^{+'} \end{aligned} \quad (13)$$

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

$D_r^+ \text{vec}(A) = \text{vech}(A)$. 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,

$$\begin{aligned} \sqrt{T} \left(\hat{\phi}_T - \phi_0 \right) &\xrightarrow{d} N(0, \Omega_\phi) \\ \Omega_\phi &= (P' \otimes I_{h+1}) \Omega_B (P \otimes I_{h+1}) + \\ &2 (I_r \otimes B(0, h)) C D_r^+ (\Sigma_\varepsilon \otimes \Sigma_\varepsilon) D_r^{+'} C' (I_r \otimes B(0, h))' \\ C &= L_r' \{ L_r (I_{r^2} + K_{rr}) (P \otimes I_r) L_r' \}^{-1} \end{aligned}$$

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.

4.5.2 Long-Run Identification

The infinite order process in assumption A1 can be rewritten, without loss of generality, as

$$\mathbf{y}_t = \sum_{j=1}^{\infty} C_j \Delta \mathbf{y}_{t-j} + C_0 \mathbf{y}_{t-1} + \varepsilon_t \quad (14)$$

with $C_i = -\sum_{j=i}^{\infty} A_j$ and $C_0 = \sum_{j=1}^{\infty} A_j$. 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 $P\mathbf{u}_t = \boldsymbol{\varepsilon}_t$, 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

$$\Phi_\infty \Phi'_\infty = \Pi^{-1}PP'\Pi'^{-1} = \Pi^{-1}\Sigma_\varepsilon\Pi'^{-1} = QQ' \quad (15)$$

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 (14). Assuming the system is covariance-stationary, this estimate will be asymptotically normally distributed with covariance matrix, say, Ω_π . If the system contains unit roots, then the estimate will be super-consistent and it will not affect the distribution of the structural impulse responses.

The structural impulse responses can be constructed as

$$\widehat{\Phi}(0, h) = \widehat{B}(0, h)\widehat{\Pi}\widehat{Q} \quad (16)$$

where the asymptotic normality of each element ensures that

$$\sqrt{T} \left(\hat{\phi}_T - \phi_0 \right) \xrightarrow{d} N(0, \Omega_\phi)$$

but where now

$$\begin{aligned} \Omega_\phi = & \frac{\partial \hat{\phi}_T}{\partial \hat{b}_T} \Omega_B \frac{\partial \hat{\phi}_T}{\partial \hat{b}'_T} + \frac{\partial \hat{\phi}_T}{\partial \hat{\pi}_T} \Omega_\pi \frac{\partial \hat{\phi}_T}{\partial \hat{\pi}'_T} + \\ & \frac{\partial \hat{\phi}_T}{\partial \hat{q}_T} \left[\frac{\partial \hat{q}_T}{\partial \hat{\pi}_T} \Omega_\pi \frac{\partial \hat{q}_T}{\partial \hat{\pi}'_T} + \frac{\partial \hat{q}_T}{\partial \text{vech}(\hat{\Sigma}_\varepsilon)} \Omega_\Sigma \frac{\partial \hat{q}_T}{\partial \text{vech}(\hat{\Sigma}_\varepsilon)'} \right] \frac{\partial \hat{\phi}_T}{\partial \hat{q}_T} \end{aligned} \quad (17)$$

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

It is easy to see that the partial derivatives in (17) are:

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

When Π is less than full rank but non-zero, 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 (17) simplifies

considerably to essentially the expression we had for the short-run recursive case, that is

$$\Omega_\phi = \frac{\partial \hat{\phi}_T}{\partial \hat{b}_T} \Omega_B \frac{\partial \hat{\phi}_T}{\partial \hat{b}'_T} + \frac{\partial \hat{\phi}_T}{\partial \hat{q}_T} \left[\frac{\partial \hat{q}_T}{\partial \text{vech}(\hat{\Sigma}_\varepsilon)} \Omega_\Sigma \frac{\partial \hat{q}_T}{\partial \text{vech}(\hat{\Sigma}_\varepsilon)'} \right] \frac{\partial \hat{\phi}_T}{\partial \hat{q}_T} \quad (18)$$

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

5 Conclusion

If impulse response coefficients were independent of each other, we would expect impulse response plots to look rather noisy, much like the plot of the error series from a regression. Seldom is this the case: impulse response paths are rather smooth, a manifestation of the high degree of correlation among the coefficients of the response. High colinearity makes identification of individual effects difficult even when collectively, there may be little ambiguity about the overall effect. Understanding and communicating the sources of uncertainty associated to such objects requires statistics based on their joint distribution.

The major contribution of this paper is to alert the profession of this seemingly self-evident observation and to provide a collection of intuitive statistical tools with which to determine what is learnt from an empirical impulse response exercise. These tools are independent of the method used to estimate the impulse responses in the sense that the formulas rely on the availability of the joint distribution, not on how this distribution is arrived at.

The paper collects a variety of asymptotic results, not out of preference for large-sample approximations necessarily, but rather because these are foundational results that are needed to justify resampling, finite-sample methods in general and because they provide the neces-

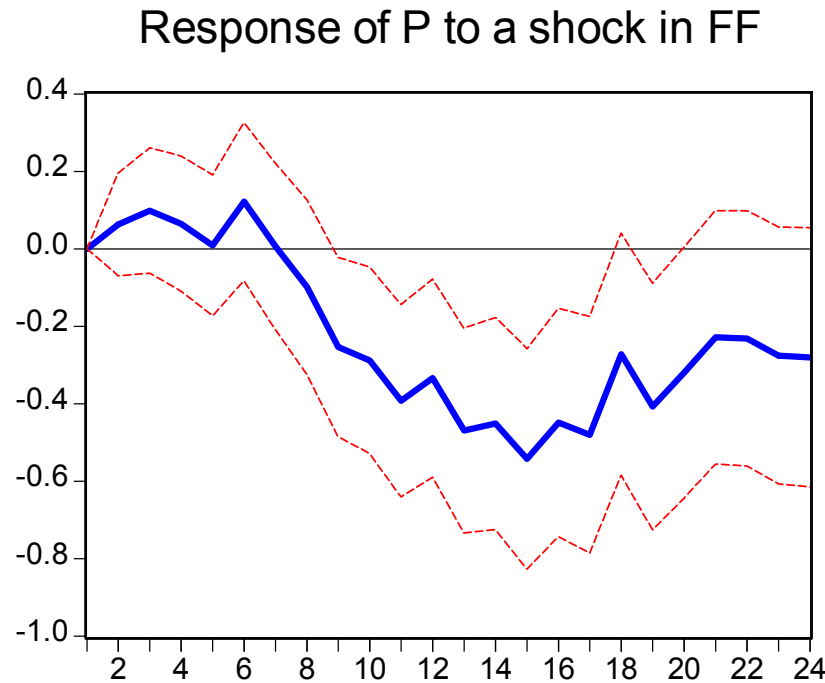
sary formulas to carry out the bootstrap with asymptotic refinements.

References

- Anderson, Theodore W. (1994) **The Statistical Analysis of Time Series Data**. New York, New York; Wiley-Interscience.
- Blanchard, Olivier J. and Danny Quah (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4): 655-673.
- Demiralp, Selva and Kevin D. Hoover (2003) "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics*, 65(supplement):745-767.
- Galí, Jordi (1999) "Technology, Improvement and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89(1):249-271.
- Granger, Clive W. J. and Norman R. Swanson (1997) "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association*, 92:357-367.
- Gonçalves, Silvia and Lutz Kilian (2006) "Asymptotic and Bootstrap Inference for $AR(\infty)$ Processes with Conditional Heteroskedasticity," *Econometric Reviews*, forthcoming.
- Hamilton, James D. (1994) **Time Series Analysis**. Princeton, NJ: Princeton University Press.
- Horowitz, Joel L. (2001) "The Bootstrap" in James J. Heckman and Edward Leamer (eds), **Handbook of Econometrics**, v. 5. Amsterdam: North-Holland, 3159-3228.
- Hurvich, Clifford M. and Chih-Ling Tsai (1989) "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76(2): 297-307.
- Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1): 161-182.
- Jordà, Òscar and Sharon Kozicki (2006) "Projection Minimum Distance: An Estimator for Dynamic Macroeconomic Models," U.C. Davis, working paper 6-23.
- Kilian, Lutz (1998) "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80(2):218-230.
- Kilian, Lutz (1999) "Finite-Sample Properties of Percentile and Percentile- t Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81(4): 652-660.

- Kuersteiner, Guido M. (2001) "Optimal Instrumental Variables Estimation for ARMA Models," *Journal of Econometrics*, 104:359-405.
- Kuersteiner, Guido M. (2002) "Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity," *Econometric Theory*, 18:547-583.
- Leeper, Eric M. and Tao Zha (2003) "Modest Policy Interventions," *Journal of Monetary Economics*, 50(8), 1673-1700.
- Lewis, R. A. and Gregory C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16(33):393-411.
- Lütkepohl, Helmut (2005) **New Introduction to Multiple Time Series**. Berlin, Germany: Springer-Verlag.
- Lütkepohl, Helmut and P. S. Poskitt (1991) "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," *Econometric Theory*, 7:487-496.
- Sims, Christopher A. (1992) "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36(10):975-1000.
- Sims, Christopher A. and Tao Zha (1999) "Error Bands for Impulse Responses," *Econometrica*, 67(5):1113-1155.
- Stock, James H. and Mark W. Watson (2001) "Vector Autoregressions," *Journal of Economic Perspectives*, 15(4):101-115.
- White, Halbert (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4):817-838.

Figure 1 – Correlation Among Impulse Response Coefficients: Response of Inflation to a shock in the Federal Funds Rate

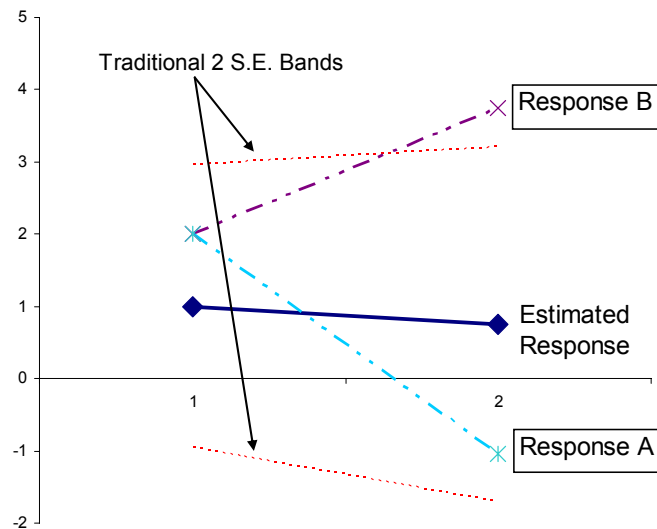
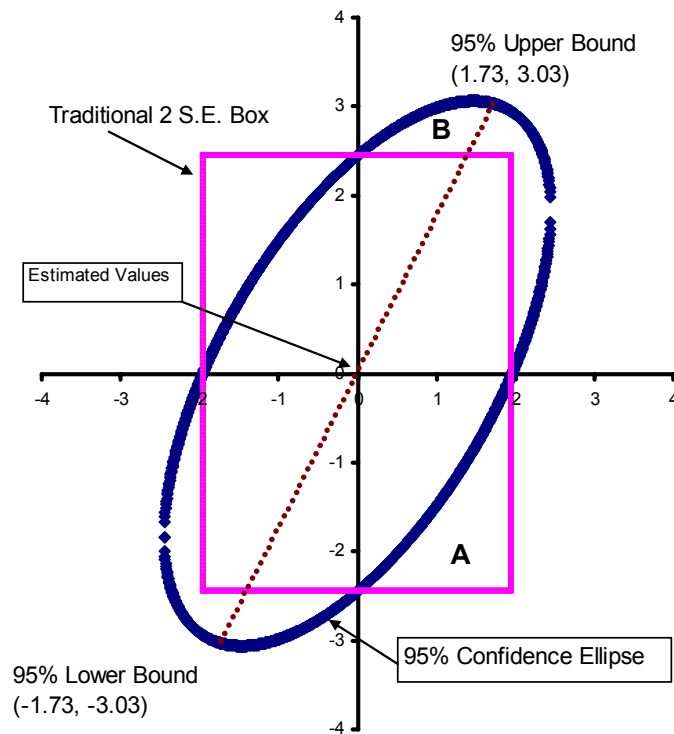


Correlation Matrix: asymptotic critical value of significance is 0.15

	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

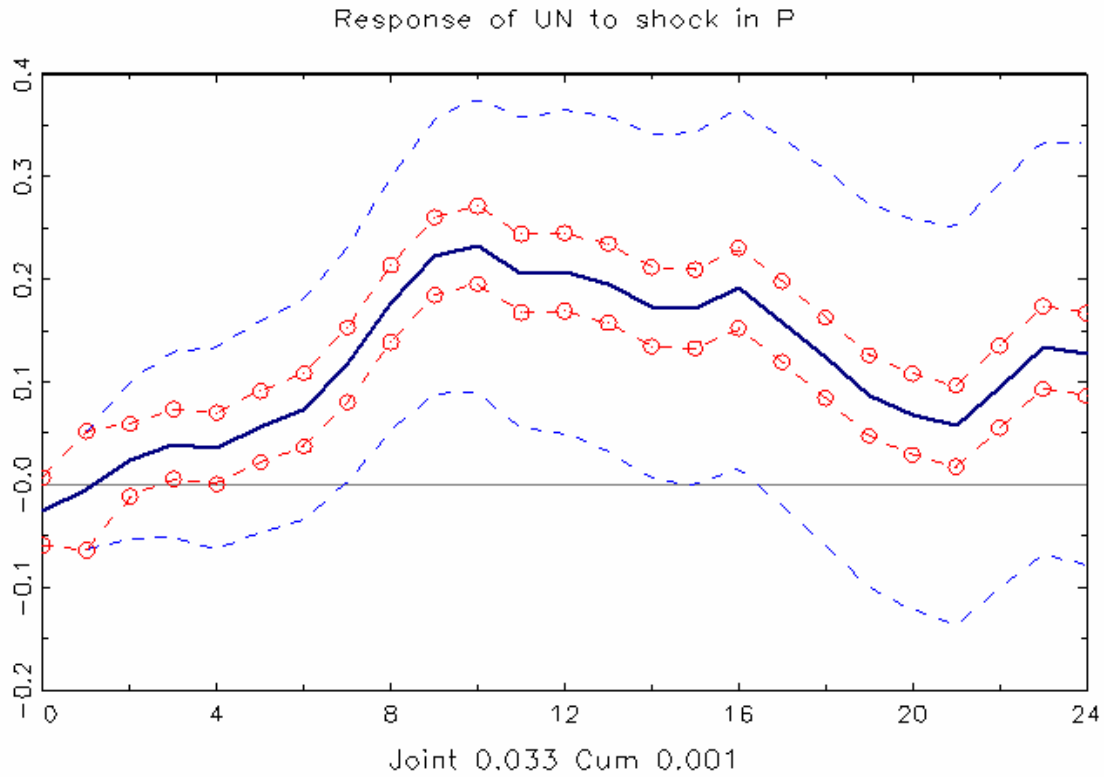
Notes. Top Panel: impulse response calculated by local projections with 6 lags for the Stock and Watson (2001) system. Traditional, marginal two, standard-error bands displayed. Bottom Panel: correlation matrix of impulse response coefficients. Asymptotic critical value of significance is 0.15.

Figure 2 – 95% Confidence Ellipse for Two Impulse Response Coefficients



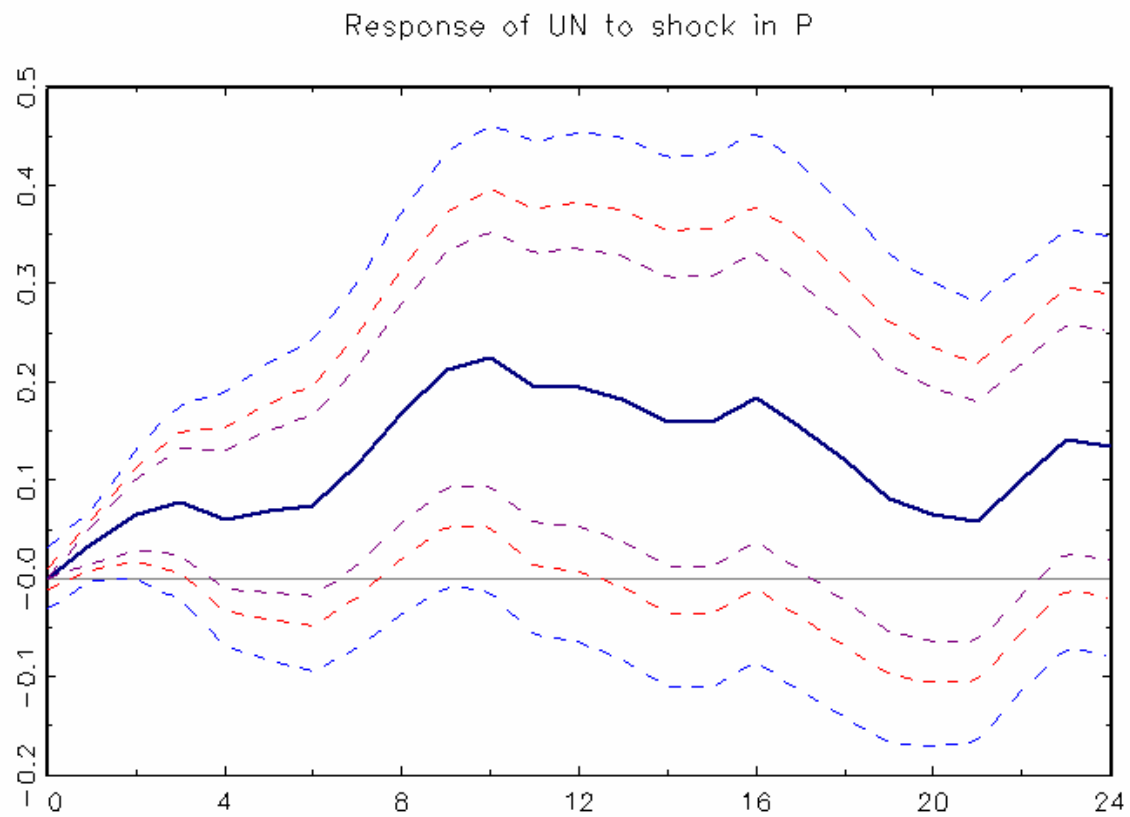
Notes. Top Panel: 95% confidence ellipse for two positively correlated impulse response coefficients. Box represents traditional two standard error bands. Coefficient estimates are centered at their mean values. Bottom Panel: representation of estimated impulse response and paths A and B from the top panel along with two standard error bands.

Figure 3 – 95% Conditional Error Bands for Response of Unemployment to Inflation Shock



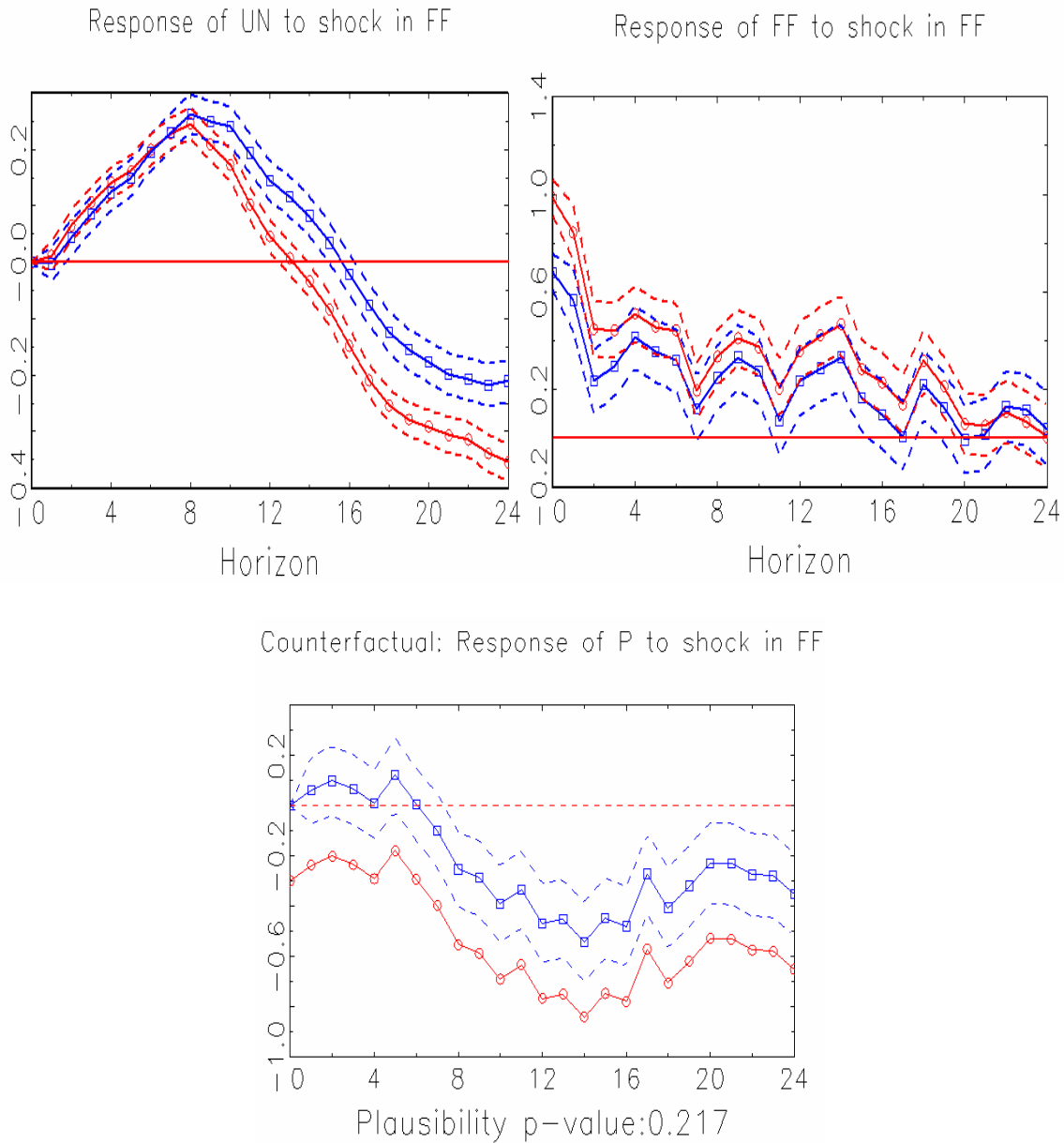
Notes: Dashed lines are traditional two marginal standard error bands. Dashed lines with circles are two conditional standard error bands. “Joint 0.033” refers to the p-value of the null hypothesis that all the response coefficients are jointly zero. “Cum 0.001” is the p-value of the null that the accumulated impulse response after 24 periods is equal to zero. Impulse response calculated by local projections with 6 lags on the Stock and Watson (2001) system.

Figure 4 – Fan Chart of Percentile Bounds for Response of Unemployment to Inflation Shock



Notes: Percentile bounds for 95th, 25th, and 1st percentiles of the Wald test of joint significance. Impulse response calculated by local projections with 6 lags on the Stock and Watson (2001) system.

Figure 5 – Anchoring Experiment: Making the Initial Response of Prices to a Shock in the Federal Funds Rate Negative



Notes: 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 6 lags.