Jordà, Òscar

Working Paper

Joint inference and counterfactual experimentation for impulse response functions by local projections

Working Paper, No. 06-24

Provided in Cooperation with:
University of California Davis, Department of Economics

Suggested Citation: Jordà, Òscar (2007) : Joint inference and counterfactual experimentation for impulse response functions by local projections, Working Paper, No. 06-24, University of California, Department of Economics, Davis, CA

This Version is available at:
https://hdl.handle.net/10419/31365

Joint Inference and Counterfactual experimentation for Impulse Response Functions by Local Projections

Oscar Jorda
University of California, Davis

Oscar Jorda
University of California, Davis

February 15, 2007

Paper # 06-24

This paper provides three measures of the uncertainty associated to an impulse response path: (1) conditional confidence bands which isolate the uncertainty of individual response coefficients given the temporal path experienced up to that point; (2) response percentile bounds} which provide bounds on the universe of permissible paths at a given probability level; and (3) Wald tests of joint significance and joint cumulative significance. These results rely on general assumptions for the joint distribution of the system's impulse responses. Given this distribution, the paper then shows how to construct counterfactual experiments formally; provides a test on the likelihood of observing the counterfactual; and derives the distribution of the system's responses conditional on the counterfactual. The paper then derives the asymptotic joint distribution of structural impulse responses identified by either short- or long-run recursive assumptions and estimated by local projections (Jorda, 2005). An application to a two country system implements all of these new methods.

February 15, 2007

Joint Inference and Counterfactual Experimentation for Impulse Response Functions by Local Projections*

Abstract

This paper provides three measures of the uncertainty associated to an impulse response path: (1) *conditional confidence bands* which isolate the uncertainty of individual response coefficients given the temporal path experienced up to that point; (2) *response percentile bounds* which provide bounds on the universe of permissible paths at a given probability level; and (3) Wald tests of joint significance and joint cumulative significance. These results rely on general assumptions for the joint distribution of the system's impulse responses. Given this distribution, the paper then shows how to construct counterfactual experiments formally; provides a test on the likelihood of observing the counterfactual; and derives the distribution of the system's responses conditional on the counterfactual. The paper then derives the asymptotic joint distribution of structural impulse responses identified by either short- or long-run recursive assumptions and estimated by local projections (Jordà, 2005). An application to a two country system implements all of these new methods.

- *Keywords:* impulse response, local projection, time-profile band, counterfactual simulation.
- *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, James Hamilton, Massimiliano Marcellino, Paul Ruud, and Aaron Smith, as well as seminar participants at BBVA, Università Boconni - IGIER, CEMFI, Bank of Italy, Federal Reserve Bank of San Francisco, and U.C. Berkeley.

1 Introduction

Estimation uncertainty on impulse responses is almost always reported by displaying two standard-error bands based on the marginal distribution of each individual coefficient (a notable exception being Sims and Zha, 1999). One can think of this visual device as the equivalent to graphically displaying the sequence of associated t-statistics. However, the coefficients of the impulse response are usually very highly correlated over time. In a traditional linear regression context, we tend to favor testing the joint significance of highly colinear regressors rather than relying on individual t-tests – individual coefficients are poorly identified and hence estimated imprecisely. Similarly, I will show that impulse response coefficients are often imprecisely estimated but that the impulse response path is not.

A natural consequence of this discussion is the desire to display impulse response uncertainty with the joint distribution rather than with the individual marginals so as to account for possible correlation. Ideally, one would display the 95% confidence, multi-dimensional ellipse associated with all the coefficients of the impulse response: this is clearly impossible in two-dimensional (and even three-dimensional) space.

However, the coefficients of an impulse response have a natural and unique temporal ordering. It turns out that the Cholesky decomposition of the covariance matrix of the impulse path translates the original responses into an orthogonal system of uncorrelated variates. This decomposition has several virtues. Uncertainty on the orthogonal system can be displayed with two standard error bands since, by construction, its joint distribution is the product of its marginals. Moreover, the sum of the sequence of these conditional t-tests

squared is the Wald statistic of the null hypothesis that the coefficients of the response path are jointly zero. Thus, these conditional t-tests not only identify which of the individual coefficients is more or less likely to be zero statistically, they represent the uncertainty associated to each coefficient conditional on the response so far experienced. I will call the graphical display of the sequence of these conditional t-tests *conditional confidence bands.*

In addition to each individual coefficient's conditional uncertainty, we are also interested in an overall measure of uncertainty for the response's path that summarizes all the individual possibilities at each horizon. For a given probability level α, there are obviously infinite possible paths so it seems sensible to focus on those that are most extreme to provide a sense about the boundaries of what we can reasonably expect to observe. For any probability level α, these can be easily constructed with the orthogonalized variates and then translated back to the original coordinate system. I will call these *response percentile bounds.*

Knowledge of the joint distribution allows one to construct joint tests of significance based on the Wald principle. In addition to these, I will discuss how to conveniently formulate tests on the null that the cumulative effect of the response is zero; tests that two responses can be considered equal from a statistical point of view; and tests that the cumulative effect of two responses is equal.

Leeper and Zha (2003) discuss a method of counterfactual simulation based on feeding alternative sequences of errors into an estimated VAR and then construct a statistic that allows one to examine whether the counterfactual is "modest" (that is, unlikely to violate the Lucas critique; Lucas, 1976). Instead, I propose a method of experimentation in which a

counterfactual path replaces the historical path and where the "modesty" of the counterfactual is measured by whether or not the counterfactual path would be rejected by the Wald statistic of the joint hypothesis measuring the distance between the counterfactual and the historical paths. This approach has two obvious virtues. First, it allows for counterfactuals based on simultaneously experimenting with more than one alternative path. Second, the method allows one to calculate the distribution of the system of impulse responses conditional on the counterfactual so that one can formally evaluate the effects of the experiment statistically.

All of these statistics require knowledge of the joint distribution of the impulse responses. This paper derives the asymptotic distribution of structural impulse responses identified either with short-run (e.g. Cholesky) or long-run (e.g. Blanchard and Quah, 1989) recursive identification restrictions and estimated semiparametrically by local projections (Jordà, 2005). Local projections have several advantages over impulse responses derived from vector autoregressions (VARs). Among others, they are more robust to lag length and other forms of misspecification discussed in Jordà (2005). Also, they are a natural building block from which to estimate impulse responses with more flexible, possibly nonlinear models. Consequently, deriving convergence results to a multivariate Gaussian distribution with analytic formulas for the covariance matrix that are closed form and analytically tractable is not only important to construct many statistics of interest discussed below, but also as a starting point for further generalizations of the method. Since these results are not directly available elsewhere, this becomes another contribution of the paper.

The paper proposes a statistical protocol to analyze vector time series with impulse responses and of necessity contains a number of econometric derivations, many of which result from familiar statistical principles. I view this simplicity as an advantage rather than a detraction for the methods presented here – basic least squares results, while seemingly unsophisticated, tend to be very robust. The goal of the paper is not to muddle the presentation with excessive econometric wizardry but rather to provide empirical practitioners with a set of clear, robust and easily implementable tools of analysis. For this reason, the paper contains a detailed empirical application of a two-country basic macroeconomic system involving U.S. and U.K. data illustrating all of these techniques. For clarity, the assumptions required to derive the main results are kept as simple as possible. Where appropriate, I discuss without proof the practical implications of relaxing some of these assumptions.

2 Joint 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 $h = 0, 1, ..., H$ 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. Neither the method of structural identification, nor the method by which these

coefficients are estimated is relevant here. All that is needed now is an available result such that, if $\widehat{\phi}_T = vec\left(\widehat{\Phi}\left(0, H\right)\right)$ (that is, the vectorized estimates of the impulse response coefficients from a sample of size T), then at least asymptotically

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} N\left(0, \Omega_\phi\right). \tag{1}$$

As an example, below I provide such a result for impulse responses estimated by local projections under either short-run or long-run recursive identification assumptions.

Error bands for impulse responses are often used as visual cues about the uncertainty of the possible time profiles that the impulse responses can follow. Traditionally, these bands have been constructed with the standard errors of each individual response coefficient, that is, for 95% confidence bands

$$\widehat{\phi}_T \pm 1.96 \times diag(\widehat{\Omega}_\phi)^{1/2},$$

were $diag\left(\widehat{\Omega}_\phi\right)$ is the $r^2\left(H+1\right) \times 1$ vector that contains the diagonal elements of $\widehat{\Omega}_\phi$.

One interpretation of these bands is as the graphical equivalent to displaying the sequence of individual t-tests associated to each impulse response coefficient. These are, strictly speaking, correct and valid statistics on the unconditional uncertainty of each individual coefficient under general assumptions but they ignore two elementary observations. The first observation is that impulse response coefficients are highly colinear. To illustrate this point, figure 1 displays the response of U.S. unemployment to a shock in the U.S. federal funds rate from the empirical application in section 5. The correlation between the response

coefficient and those for periods 2-12 is displayed in the top right panel of figure 1 and ranges from 0.7 between periods 1 and 2 to 0.2 between periods 1 and 12. The bottom panel of figure 1 displays the entire correlation matrix where many entries have values above 0.5. Even for an impulse response that would traditionally be considered as essentially zero on the basis of the 95% unconditional confidence bands, the impulse response coefficients are very highly correlated indeed.

In the context of traditional linear regression, we favor tests of joint significance of highly colinear regressors rather than relying on individual t-tests – individual coefficients are poorly identified and hence estimated with wide standard errors. Similarly, the uncertainty associated to the impulse response path is summarized by its joint distribution and at a 95% confidence level, it results in a multi-dimensional ellipsoid. Unfortunately, such an ellipsoid cannot be displayed in two- or even three-dimensional space, although clearly, one could construct a Wald statistic for the null hypothesis of joint significance. I will return to the issue of testing joint hypothesis momentarily.

2.1 Conditional Confidence Bands

The second observation is that impulse response coefficients have a natural temporal ordering: the value of the impulse response today determines the possible trajectories of the response in future periods but future periods cannot affect the current path of the response. Therefore, it is natural to translate the original response into a system of orthogonal variates that preserves this temporal ordering while simplifying the task of constructing error bands.

Specifically, suppose we are interested in the response of variable i to a shock in variable j and associated with the coefficients $\widehat{\phi}_{ij}^{0}, \widehat{\phi}_{ij}^{1}, ..., \widehat{\phi}_{ij}^{H}$, which can be collected compactly into the vector $\widehat{\boldsymbol{\phi}}_{ij}$. Let $\Omega_{\phi}(i,j)$ denote the covariance matrix associated to this impulse response which can be constructed by choosing the appropriate rows and columns of the matrix Ω_{ϕ}, as I will show below.

The Wald statistic of the null hypothesis $H_0 : \widehat{\boldsymbol{\phi}}_{ij} = 0$ can be easily constructed as

$$\widehat{W}_{ij} = \left(\widehat{\boldsymbol{\phi}}_{ij} - \mathbf{0}\right)' \widehat{\Omega}_{\phi}(i,j)^{-1} \left(\widehat{\boldsymbol{\phi}}_{ij} - \mathbf{0}\right) \xrightarrow{d} \chi^2_{H+1}$$

Since the covariance matrix $\widehat{\Omega}_{\phi}(i,j)$ is positive-definite and symmetric, it admits a unique Cholesky decomposition such that

$$\widehat{\Omega}_{\phi}(i,j) = \widehat{A}(i,j)\widehat{D}(i,j)\widehat{A}(i,j)'$$

where $A(i,j)$ is lower triangular with ones in the main diagonal and $D(i,j)$ is a diagonal matrix with positive entries that represent the variances of $\widehat{\phi}_{ij}^{h}|\widehat{\phi}_{ij}^{h-1}, ..., \widehat{\phi}_{ij}^{0}$. The way to see this is to realize that the matrix $A(i,j)$ projects each impulse response coefficient $\widehat{\phi}_{ij}^{h}$ onto the response coefficients $\widehat{\phi}_{ij}^{h-1}, ..., \widehat{\phi}_{ij}^{0}$.

Define $\widehat{\psi}_{ij}^{h} \equiv \widehat{\phi}_{ij}^{h}|\widehat{\phi}_{ij}^{h-1}, ..., \widehat{\phi}_{ij}^{0}$, that is, the impulse response of variable i to a shock in j at period h, conditional on the path $\widehat{\phi}_{ij}^{h-1}, ..., \widehat{\phi}_{ij}^{0}$; then $\widehat{\boldsymbol{\psi}}_{ij} = \widehat{A}(i,j)^{-1}\widehat{\boldsymbol{\phi}}_{ij}$ and the covariance matrix of the vector $\widehat{\boldsymbol{\psi}}_{ij}$ is simply $\widehat{D}(i,j)$. Consequently, a 95% confidence region for these conditional impulse responses can be easily calculated as

$$\widehat{\boldsymbol{\psi}}_{ij} \pm 1.96 \times diag\left(\widehat{D}(i,j)\right) \tag{2}$$

and since $\widehat{D}\,(i,j)$ is a diagonal matrix, the region just defined covers approximately the same area as the formal 95% confidence ellipsoid.[1] Further, notice that the Wald statistic of the joint null $H_0 : \widehat{\boldsymbol{\phi}}_{ij} = \mathbf{0}$ can be recast as follows

$$\begin{aligned}
\widehat{W}_{ij} &= \left(\widehat{\boldsymbol{\phi}}_{ij} - \mathbf{0}\right)' \widehat{\Omega}_{\phi}\,(i,j)^{-1} \left(\widehat{\boldsymbol{\phi}}_{ij} - \mathbf{0}\right) \\
&= \widehat{\boldsymbol{\phi}}_{ij}{}'(\widehat{A}(i,j)\widehat{D}(i,j)\widehat{A}(i,j)')^{-1}\widehat{\boldsymbol{\phi}}_{ij} \\
&= \widehat{\boldsymbol{\psi}}_{ij}{}'\widehat{D}(i,j)^{-1}\widehat{\boldsymbol{\psi}}_{ij} = \\
&= \sum_{h=0}^{H} \left(\frac{\widehat{\boldsymbol{\psi}}_{ij}^{h}}{\widehat{d}_{ij}^{h}}\right)^2 = \sum_{h=0}^{H} \left(t_{ij}^{h}\right)^2
\end{aligned} \tag{3}$$

where $\widehat{d}_{ij}^{h}$ is the h^{th} diagonal element of the matrix $\widehat{D}(i,j)$ and t_{ij}^{h} is the t-ratio of the null hypothesis $H_0 : \psi_{ij}^{h} = \phi_{ij}^{h} | \widehat{\phi}_{ij}^{h-1}, ..., \widehat{\phi}_{ij}^{0} = 0$ and $t_{ij}^{h} \xrightarrow{d} N\,(0,1)$. In other words, the Wald statistic of the joint null that the coefficients of the response of i to a shock in j is equivalent to the sum of squared t-tests of the null that the h^{th} conditional impulse response coefficient is zero.

The conditional t-tests and the associated conditional error bands appropriately summarize the uncertainty associated to each impulse response coefficient given its correlation with coefficients at previous horizons. In fact, the conditional standard errors correctly summa-

[1] As I will shortly, formally this area is $\widehat{\psi}_{ij} \pm \sqrt{\frac{\chi^2_{H+1}(0.95)}{H+1}}\mathbf{i}_{H+1}$ where $\chi^2_{H+1}\,(0.95)$ refers to the value of the chi-square with $H+1$ degrees of freedom at a 95% confidence level and $\mathbf{i}_{H+1}$ is an $H+1 \times 1$ vector of ones. Intuitively, we are interested in ensuring that the event consisting on the union of the individual conditional events have probabiliity 0.95, not that each of the constituent events have probability 0.95.

rize the uncertainty of each coefficient conditional on any path followed up to that point, not just the historical average observed. However, because at each horizon there is uncertainty on the coefficient's true value, an overall measure on the path's uncertainty would be desirable. This is what traditional two standard error bands would provide absent any correlation between response coefficients.

The top panel of figure 2 displays the impulse response of U.S. unemployment to a shock in U.S. inflation along with the usual unconditional 95% confidence bands and the conditional 95% confidence bands just discussed. In addition, the bottom of the panel shows the p-value of the joint significance test to be 0.05, whereas the p-value of the joint cumulative test is 0.048. I will discuss these in more detail shortly. Meanwhile, the main message of figure 2 is to show that, while traditional error bands suggest the response borders significance only for a few intermediate horizons, the joint tests clearly suggest the response is significantly different from zero. The conditional error bands support this assessment, suggesting that, while the response starts out insignificantly for the first three periods after impact, it is clearly significant thereafter.

Formally, an overall measure of uncertainty requires all the possible paths the impulse response could follow inside the 95% confidence ellipse (and described by the Wald statistic in expression (3)) be considered. Clearly, this multi-dimensional object cannot be displayed in two or even three dimensions. An alternative approach that I propose is to compute response percentile bounds.

2.2 Response Percentile Bounds

Any vector of values on the conditional α%-confidence boundary can be translated into an actual response α-percentile bound by translating the boundary values in the conditional coordinate system $\widehat{\boldsymbol{\psi}}_{ij}$ into the original system $\widehat{\boldsymbol{\phi}}_{ij}$. The orthogonality of the elements of $\widehat{\boldsymbol{\psi}}_{ij}$ means that its joint distribution is simply the product of its marginals so that the $\pm$ α-percentile conditional bounds are

$$\widehat{\boldsymbol{\psi}}_{ij} \pm \sqrt{\frac{\chi^2_{H+1}(\alpha)}{H+1}}\mathbf{i}_{H+1}$$

where $\chi^2_{H+1}(\alpha)$ is the value of a chi-square with $H+1$ degrees of freedom at an α probability level and $\mathbf{i}_{H+1}$ is an $H+1 \times 1$ vector of ones.

Since $\widehat{\boldsymbol{\phi}}_{ij} = \widehat{A}_{ij}\widehat{\boldsymbol{\psi}}_{ij}$, then the original response α-percentile bounds are

$$\widehat{\boldsymbol{\phi}}_{ij} \pm A_{ij}\sqrt{\frac{\chi^2_{H+1}(\alpha)}{H+1}}\mathbf{i}_{H+1}$$

Thus, for different values of α, one could plot each percentile bound to form a fan chart. Several remarks deserve comment. First, the response percentile bounds contain many paths with less than $1-\alpha$ probability of being observed (hence the name bounds instead of bands). That is because there will be paths inside the bounds that do not conform with the pattern of correlation between impulse response coefficients observed in the data. Second, when the correlation among the elements of $\widehat{\boldsymbol{\phi}}_{ij}$ is zero, then $\widehat{\boldsymbol{\phi}}_{ij} = \widehat{\boldsymbol{\psi}}_{ij}$ and conditional two standard error bands, traditional two standard error bands and response 95^{th}-percentile bounds all coincide. Third, α-percentile bounds are an answer to a natural "worst-case" scenario type

of question but specific economic applications may suggest a different choice (obviously, with a different interpretation). For example, one could imagine a situation where the bound of interest is related to an event in which the conditional response coefficients all sit on the positive α-percentile boundary for the first few periods and then on the negative α-percentile boundary for the remaining periods. This would obviously lead to a different fan chart.

A fan chart based on the response percentile bounds for the impulse response of figure 2 discussed in the previous section, is displayed in the bottom panel of that figure. It is important to note that the outer bounds fan out more widely than even the unconditional 95% confidence bands. This seemingly peculiar result is easily explained: unconditional 95% confidence bands contain many response paths with less than 5% chance of being observed while at the same time excluding many response paths with more than a 5% chance of occurring. This disparity is caused by the correlation between coefficients, which is ignored in the unconditional confidence bands. Therefore, a plot of the fan chart provides a better sense of the set of admissible paths.

2.3 Joint Hypothesis Tests

Conditional confidence bands and response percentile bounds are two alternative methods of assessing an impulse response exercise empirically. Knowledge of the approximate joint distribution of $\widehat{\phi}_T$ given by expression (1) affords a third natural alternative: tests of joint hypotheses. Specific applications will generate specific hypotheses, however, at least four hypotheses are likely to be popular: (1) tests of the null of joint significance; (2) tests of joint cumulative significance; (3) tests of response equality; and (4) tests of cumulative

response equality.

To formulate these tests in practice it is convenient to introduce a selector matrix. Let R_i be an $r \times 1$ column vector of zeroes with a one in the i^{th} entry; similarly, let C_j be an $r \times 1$ column vector of zeroes with a one in the j^{th} entry; then define $S_{ij} = C_j \otimes (I_{H+1} \otimes R_i)$ so that $S_{ij}\widehat{\phi}_T = \boldsymbol{\phi}_{ij}$ and $S_{ij}\widehat{\Omega}_\phi S'_{ij} = \widehat{\Omega}_\phi(i,j)$. Interest is in tests of the generic null

$$H_0 : QS\phi = q,$$

where Q is a matrix of linear combinations, S is a selector matrix to be defined momentarily, and q is a $J \times 1$ vector. The Wald statistic for any null that can be crafted in this format is readily seen to be

$$\widehat{W} = \left(QS\widehat{\phi}_T - q\right)' \left(S\widehat{\Omega}_\phi S'\right)^{-1} \left(QS\widehat{\phi}_T - q\right) \xrightarrow{d} \chi^2_J$$

In particular, for each of the four hypotheses just considered:

1. *Joint Significance Test:* choose $S = S_{ij}$; $Q = I_{H+1}$; $q = \mathbf{0}_{H+1\times 1}$. This test evaluates the null that the impulse response path is jointly indistinguishable from a zero path. The p-values of this test appear at the bottom of both panels of figure 2 and in figures 3 and 4.

2. *Joint Cumulative Significance Test:* choose $S = S_{ij}$; $Q = \mathbf{i}_{H+1}$; $q = 0$. This test evaluates the cumulative impact of the impulse response against a zero null and is reported in figures 2, 3 and 4.

3. *Joint Test of Equality:* choose $S = (S_{ij} \quad S_{lk})'$; $Q = (I_{H+1} \quad -I_{H+1})$; $q = \mathbf{0}_{H+1\times 1}$. This test compares any two impulse responses in the system and assesses whether they are equal. An example is given in section 5.

4. *Joint Test of Cumulative Equality:* choose $S = (S_{ij} \quad S_{lk})'$; $Q = (\mathbf{i}_{H+1} \quad -\mathbf{i}_{H+1})$; $q = 0$. This test compares whether the cumulative impact of any two impulse responses is the same and an example is given in section 5.

3 Counterfactual Experimentation

The Lucas Critique (Lucas, 1976) warns of the dangers of counterfactual experimentation with empirical models. In real economies, rational economic agents immediately adapt to the new environment generated by the counterfactual in ways the empirical models alone cannot anticipate. In essence, the parameters of the empirical model are not constant to the counterfactual – they are functions of deep parameters only a behavioral model can uncover. Hence, predictions based on keeping the parameters constant will be unreliable even though, mechanically speaking, the construction of the counterfactual poses no technical difficulties.

Hoover and Jordà (2001) and Leeper and Zha (2003) examine the empirical aspects of the Lucas Critique in the context of Cochrane's (1998) model. Cochrane (1998) allows for a mixed composition of adaptive and rational behavior to then argue that small deviations from the full rational expectations paradigm can generate mostly adaptive responses and hence approximately correct counterfactuals. Hoover and Jordà (2001) test Cochrane's (1998) theoretical results by using breaks in the policy equation of a VAR and the subsequent

changes in the parametric structure of the data to provide confirmatory evidence in support of Cochrane's (1998) propositions.

Leeper and Zha (2003) instead examine the agents' ability to discern policy interventions from a statistical point of view by arguing that "modest" policy interventions may not result in agents revising their behavior. Their approach consists in selecting a sequence of shocks that can be reasonably expected to be drawn from the distribution of historical fitted disturbances (the basis of their "modesty" test). Given this sequence, they then generate a set of forecasts alternative to those generated without the intervention.

The approach that I introduce is to examine instead response paths that are drawn from the empirical distribution of historical paths. It should be clear that the more uncertain the data, the more difficult it is to ascertain changes in the economy that would cause agents to revise their behavioral rules – in the limit, a parameter with an infinite variance would be essentially unknowable and small variations could hardly be expected to affect optimal economic behavioral rules. Accordingly, the types of experiments I have in mind consist of alternative response paths whose "modesty" can be formally judged with a Wald metric by the distance between the counterfactual and the historical paths in probability units.

Several aspects of this approach make it appealing. First, experimentation is done directly on to the response paths themselves, which are easier to interpret economically. Second, the assumption that $\widehat{\phi}_T$ is normally distributed makes construction of the counterfactual a straightforward application of well known properties of the multivariate normal distribution. In turn, these properties provide the distribution of the responses conditional on

the counterfactual so that formal inference of the type introduced in the previous sections, can be readily applied. Finally, the counterfactual can be constructed to involve alternative paths for more than one response simultaneously since the Wald principle required to assess its validity and the conditional formulas are directly based on $\widehat{\phi}_T$. Thus, for example, one could ask how does inflation respond to different shocks in an economy where GDP and interest rates are simultaneously made less responsive to oil shocks.

3.1 Is the Counterfactual Prudent?

Before I discuss how to conduct the counterfactual, it seems sensible to establish first how best to evaluate its probity. Accordingly, suppose we want to examine how the systems' responses change when we consider a counterfactual path for the response of variable k to a shock in variable l (denoted $\boldsymbol{\phi}_{kl}^c$). I begin by noticing that the probity of this counterfactual, and hence the inherent likelihood that rational agents would revise their behavior, can be assessed with the following Wald statistic

$$\widehat{W}_{kl}^c = \left(\widehat{\boldsymbol{\phi}}_{kl} - \boldsymbol{\phi}_{kl}^c\right)' \widehat{\Omega}_\phi(k,l)^{-1} \left(\widehat{\boldsymbol{\phi}}_{kl} - \boldsymbol{\phi}_{kl}^c\right) \xrightarrow{d} \chi^2_{H+1}$$

where all the elements of the test can be constructed as described in previous sections. One minus the p-value of this test can be thought of as the probability that the counterfactual violates the historical average response path observed in the data. This is a natural metric that is easy to communicate. A p-value of less than 0.05 (or a probability that the counterfactual violates the data greater than 0.95) is not only problematic from the point of view of possibly violating the Lucas Critique, but more generally, it would stretch the lim-

its of the estimated model: the counterfactual would involve extrapolating the model into regions where very little or no data has ever been observed to occur. An example of this test is used in section 5 and displayed in figure 5. In that example, the distance between the counterfactual and the historical paths is 27% in probability units.

3.2 Estimation and Inference of the Responses to the Counterfactual

The fundamentals required to conduct the counterfactual experiment are based on well known results for the multivariate normal distribution. Specifically, we know that if $\mathbf{y}_1$ and $\mathbf{y}_2$ are two random vectors of generic dimensions with joint normal distribution

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \tag{4}$$

then the conditional distribution of $\mathbf{y}_1$ given $\mathbf{y}_2 = \mathbf{y}_2^c$ is

$$\mathbf{y}_1 | \mathbf{y}_2 = \mathbf{y}_2^c \sim N\left(\mu_{1|2}; \Sigma_{11|2}\right)$$

with

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{y}_2^c - \mu_2\right) \tag{5}$$

$$\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Once we have assessed the probity of the counterfactual with the Wald statistic $\widehat{W}_{kl}^c$, suppose we are interested in examining the effect of the counterfactual onto the response of variable

i to a shock in j, where there are no restrictions on the possible values of $i, j, k, l \in \{1, ..., r\}$. Using the selector matrix defined in the previous section, we know that $S_{ij}\widehat{\phi}_T = \widehat{\boldsymbol{\phi}}_{ij}$; $S_{kl}\widehat{\phi}_T = \widehat{\boldsymbol{\phi}}_{kl}$ and hence, the covariance matrix for $\widehat{\boldsymbol{\phi}}_{ij}$ and $\widehat{\boldsymbol{\phi}}_{kl}$ is

$$\begin{pmatrix} S_{ij} \\ S_{kl} \end{pmatrix} \widehat{\Omega}_\phi \begin{pmatrix} S'_{ij} & S'_{kl} \end{pmatrix} = \begin{bmatrix} S_{ij}\widehat{\Omega}_\phi S_{ij'} & S_{ij}\widehat{\Omega}_\phi S'_{kl} \\ S_{kl}\widehat{\Omega}_\phi S'_{ij} & S_{kl}\widehat{\Omega}_\phi S'_{kl} \end{bmatrix}$$
$$= \begin{bmatrix} \widehat{\Omega}_\phi(i,j) & S_{ij}\widehat{\Omega}_\phi S'_{kl} \\ S_{kl}\widehat{\Omega}_\phi S'_{ij} & \widehat{\Omega}_\phi(k,l) \end{bmatrix}$$

with the obvious correspondence to expression (4). Therefore, the conditional path of $\widehat{\boldsymbol{\phi}}_{ij}$ given $\boldsymbol{\phi}_{kl} = \boldsymbol{\phi}^c_{kl}$ is therefore a direct application of expression (5),

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

with conditional variance

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

Several remarks deserve comment. First, under the assumption of normality, $\widehat{\Omega}_\phi(i,j|k,l)$ is all we need to do hypothesis tests on $\widehat{\boldsymbol{\phi}}_{ij}|\boldsymbol{\phi}^c_{kl}$ or any of the derivations described in previous sections. Second, notice that the second term in the expression of the conditional variance is a positive definite matrix so that $tr(\widehat{\Omega}_\phi(i,j|k,l)) \leq tr(\widehat{\Omega}_\phi(i,j))$, that is, the variance conditional on the counterfactual is smaller than the unconditional variance. The reason is that the unknown path of the estimated response $\boldsymbol{\phi}_{kl}$ is being replaced with a fixed value

given by $\boldsymbol{\phi}^c_{kl}$. Third, the counterfactual is not limited to responses originating from a shock in the same variable – the conditioning arguments do not impose any restrictions on the math. It is also straightforward to experiment with counterfactuals involving more than one response at a time by simply extending the selector matrices appropriately. Fourth, when the correlation between the counterfactual estimated response, $\widehat{\boldsymbol{\phi}}_{kl}$, and the response whose conditional distribution we are interested in computing, $\widehat{\boldsymbol{\phi}}_{ij}$, is zero then $\widehat{\boldsymbol{\phi}}_{ij}|\boldsymbol{\phi}^c_{kl} = \widehat{\boldsymbol{\phi}}_{ij}$. In some instances, it may be of interest to test this hypothesis (e.g. is the economy's response to an oil shock different when the central bank does not raise interest rates in response to an oil shock?). Such a hypothesis can be easily tested since under the null $\widehat{\Omega}_\phi(i,j|k,l) = \widehat{\Omega}_\phi(i,j)$ and hence

$$\widehat{W}^0_{ij|kl} = \left(\widehat{\boldsymbol{\phi}}_{ij}|\boldsymbol{\phi}^c_{kl} - \widehat{\boldsymbol{\phi}}_{ij}\right)' \widehat{\Omega}_\phi(i,j)^{-1} \left(\widehat{\boldsymbol{\phi}}_{ij}|\boldsymbol{\phi}^c_{kl} - \widehat{\boldsymbol{\phi}}_{ij}\right) \xrightarrow{d} \chi^2_{H+1}.$$

In section 5, I provide an application of a counterfactual experiment that is displayed in figure 5. Without laboring the details of how the figure is constructed, the panel displays the historical and counterfactual paths along with conditional confidence bands constructed from the historical and counterfactual distributions respectively.

4 The Joint Asymptotic Distribution of Impulse Responses by Local Projections

The previous sections describe several new methods of inference and counterfactual simulation that require knowledge of the joint distribution of the system's vector of impulse responses, $\widehat{\phi}_T$. In this section I provide such a result for structural impulse responses esti-

mated by local projections and whose identification has been achieved by either short- or long-run recursive assumptions. Local projections are a relatively novel method for which these results have not been previously established. Similar results for VARs can probably be collated together from different sources and Lütkepohl (2005) and references therein is probably a good place to start. Hence I will not review these here. Accordingly, consider a covariance-stationary $r \times 1$ vector of time series $\mathbf{y}_t$, whose Wold decomposition is given by

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} B_j \boldsymbol{\varepsilon}_{t-j} \tag{6}$$

where for simplicity no deterministic terms beyond the constant term are considered. Further assume the following.

Assumption 1:

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

(ii) $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \underset{r \times r}{\Sigma_\varepsilon}$

(iii) $\sum_{j=0}^{\infty} \|B_j\| < \infty$ where $\|B_j\|^2 = tr(B_j' B_j)$ and $B_0 = I_r$

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

Then from the Wold decomposition theorem (see e.g. Anderson, 1994) the process in (6) can also be written as:

$$\mathbf{y}_t = \mathbf{m} + \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \tag{7}$$

such that the following holds.

Result 1:

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

(ii) $A(z) = I_r - \sum_{j=1}^{\infty} A_j z^j = B(z)^{-1}$

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

In what follows, I take expression (7) as primitive in describing the class of models whose impulse responses we are interested in characterizing. These assumptions are quite general and include as a special case invertible vector autoregressive moving average models (VARMA) and traditional VARs. The assumption that the $\boldsymbol{\varepsilon}_t$ are $i.i.d.$ allows derivations with traditional central limit theorems and reduces the number of technical conditions needed. Where appropriate, I discuss the consequences of relaxing this assumption.

Given this set-up, one could consider estimating a truncated VAR and then inverting its estimates to obtain the impulse responses. This is the more familiar approach and results on the asymptotic distribution of the marginal distributions of the impulse response coefficients are available in Hamilton (1994) and Lütkepohl (2005), for example. Instead, I estimate impulse responses by local projections (Jordà, 2005) for several reasons. First, I am interested in the joint distribution (rather than the collection of marginals) of the system's impulse responses. Estimates from VARs are nonlinear functions of estimated coefficients that require (not always reliable) delta method approximations and considerably complex algebraic manipulations. The method of local projections is a direct estimate of the impulse response coefficients so that familiar least-squares formulas is all that is needed to compute

the joint asymptotic distribution. Second, Jordà (2005) shows that local projections are more robust than VARs to certain types of misspecification, including lag length. Third, unlike VARs, local projections can be easily generalized to nonlinear models so that deriving asymptotic results is a natural building block for further extensions. Finally, the asymptotic results presented below are not widely available and hence serve to complement the literature.

Jordà's (2005) local projection method is based on the expression that results from simple recursive substitution in the $VAR(\infty)$ representation of expression (7), specifically

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + \boldsymbol{\varepsilon}_{t+h} + B_1 \boldsymbol{\varepsilon}_{t+h-1} + ... + B_{h-1} \boldsymbol{\varepsilon}_{t+1} \tag{8}$$

where:

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

(ii) $A_j^h = B_{h-1} A_j + A_{j+1}^{h-1}$ where $h \geq 1$; $A_{j+1}^0 = 0$; $B_0 = I_r$; and $j \geq 1$.

Now consider truncating the infinite lag expression (8) at lag k

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \tag{9}$$

$$\mathbf{v}_{k,t+h} = \sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j}.$$

Momentarily assume that $A_j = 0$ for $j > k$ so that the model is a VAR(k) and the first term in the expression for $\mathbf{v}_{k,t+h}$ vanishes. Hence, consider estimating the system in expression (9) for $h = 0, ..., H$ and define $\mathbf{y}_j$ for $j = H$, ..., 1, 0, –1, ..., –k as the $(T - k - H) \times r$ matrix of

stacked observations of the $1 \times r$ vector $\mathbf{y}'_{t+j}$. Additionally, define the $(T-k-H) \times r(H+1)$ matrix $Y \equiv (\mathbf{y}_0, ..., \mathbf{y}_H)$; the $(T-k-H) \times r$ matrix $X \equiv \mathbf{y}_0$; the $(T-k-H) \times r(k-1)+1$ matrix $Z \equiv \left(\mathbf{1}_{(T-k-H)\times 1}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k+1}\right)$ and the $(T-k-H) \times (T-k-H)$ matrix $M_z = I_{T-k-H} - Z\left(Z'Z\right)^{-1} Z'$. Notice that the inclusion of $\mathbf{y}_0$ in Y is a simplifying notational trick that has no other effect than to ensure that the first block of coefficients is I_r. This will be convenient when deriving the structural impulse response function in the next section. Using standard properties of least-squares, impulse responses over h horizons can be jointly estimated as

$$\widehat{B_T}(0, H) = \begin{bmatrix} I_r \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_H \end{bmatrix} = \left[Y' M_z X\right] \left[X' M_z X\right]^{-1} \tag{10}$$

and it is straight-forward to see that $\widehat{b}_T = vec(\widehat{B_T}(0, H))$ converges in distribution to

$$\sqrt{T}\left(\widehat{b}_T - b_0\right) \xrightarrow{d} N\left(0, \Omega_B\right) \tag{11}$$

where Ω_B can be estimated with $\widehat{\Omega}_B = \left\{\left[X' M_z X\right]^{-1} \otimes \widehat{\Sigma}_v\right\}$. Although properly speaking the equations associated with $B_0 = I_r$ have zero variance, I find it notationally more compact and mathematically equivalent to calculate the residual covariance matrix $\widehat{\Sigma}_v$ as

$$\widehat{\Sigma}_v = \widehat{\Psi}_B \left(I_{H+1} \otimes \widehat{\Sigma}_\epsilon\right) \widehat{\Psi}'_B,$$

where $\widehat{\Psi}_B$ is

$$\underset{r(H+1)\times r(H+1)}{\widehat{\Psi}_B} = \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & \widehat{B}_1 & I_r & ... & \mathbf{0}_r \\ \vdots & \vdots & \vdots & ... & \vdots \\ \mathbf{0}_r & \widehat{B}_{H-1} & \widehat{B}_{H-2} & ... & I_r \end{bmatrix} \tag{12}$$

with, $\widehat{\Sigma}_\epsilon = \frac{\widehat{\mathbf{v}_1}'\widehat{\mathbf{v}_1}}{T-k-H}$; $\widehat{\mathbf{v}_1} = M_z\mathbf{y}_1 - M_z\mathbf{y}_0\widehat{B}_1$. Therefore, $\widehat{\Omega}_B$ is a simple estimate of the analytic asymptotic covariance matrix of impulse responses across time and across variables.

Several results deserve comment. First, Jordà and Kozicki (2006) show that least-squares estimates of expression (8) produce consistent and asymptotically normal estimates of $A_1^h = B_h$ for $h \geq 1$ when $k \to \infty$ as long as, among other technical conditions made explicit in that paper,

1. *k is chosen as a function of the sample size T such that*

$$\frac{k^3}{T} \to 0; T, k \to \infty$$

2. *k is chosen as a function of T such that*

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

and with the same asymptotic distribution just derived. Second, the assumption that the $\boldsymbol{\varepsilon}_t$ 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 heterostedascity-robust

covariance estimator such as White (1980). The reader is referred to Kuersteiner (2001, 2002) and Gonçalves and Kilian (2006) for related applications.

4.1 The Distribution of Structural Impulse Responses

The Wold decomposition for $\mathbf{y}_t$ in expression (6) does not assume that the residuals $\boldsymbol{\varepsilon}_t$ are orthogonal to each other and therefore $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \Sigma_\varepsilon$ 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 $\boldsymbol{\varepsilon}_t$ given by $P\mathbf{u}_t = \boldsymbol{\varepsilon}_t$, where $E\left(\mathbf{u}_t\mathbf{u}_t'\right) = I_r$ 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, $\Phi_\infty = \sum_0^\infty \Phi_j$.

Consequently, given some estimate $\widehat{P}$ the structural impulse responses Φ_i can be calculated as follows:

$$\widehat{\Phi}\left(0,h\right) = \begin{bmatrix} \widehat{\Phi}_0 \\ \widehat{\Phi}_1 \\ \vdots \\ \widehat{\Phi}_h \end{bmatrix} = \widehat{B}\left(0,h\right)\widehat{P} = \begin{bmatrix} I_r \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_h \end{bmatrix} \widehat{P} = \begin{bmatrix} \widehat{P} \\ \widehat{B}_1\widehat{P} \\ \vdots \\ \widehat{B}_h\widehat{P} \end{bmatrix} \tag{13}$$

Let $\widehat{\phi}_T = vec\left(\widehat{\Phi}(0,h)\right)$, we want to determine the asymptotic covariance matrix Ω_ϕ since it is clear that

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

4.1.1 Short-Run Identification

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

$$\Omega_\phi = \frac{\partial \phi}{\partial b}\Omega_B \frac{\partial \phi}{\partial b'} + \frac{\partial \phi}{\partial vec(P)}\frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)}\Omega_\Sigma \frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)'}\frac{\partial \phi}{\partial vec(P)'} \tag{14}$$

with $\Omega_\Sigma \equiv E\left[vech(\Sigma_\varepsilon)\, vech(\Sigma_\varepsilon)'\right]$ and $E\left[b, vech(\Sigma_\varepsilon)\right] = 0$ since $E\left[X' M_z \mathbf{v}_1/(T-k-h)\right] \xrightarrow{p} 0$. 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

$$\begin{aligned} \frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)} &= L_r' \left\{L_r \left(I_{r^2} + K_{rr}\right)(P \otimes I_r) L_r'\right\}^{-1} \\ & \qquad \sqrt{T}\left(vech\left(\widehat{\Sigma}_\varepsilon\right) - vech(\Sigma_\varepsilon)\right) \xrightarrow{d} N(0, \Omega_\Sigma) \\ \Omega_\Sigma &= 2D_r^+ (\Sigma_\varepsilon \otimes \Sigma_\varepsilon) D_r^{+'} \end{aligned} \tag{15}$$

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,

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

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

4.1.2 Long-Run Identification

If instead structural identification is based on long-run identification assumptions, the infinite order process in expression (7) can be rewritten, without loss of generality, as

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \Psi_j \Delta \mathbf{y}_{t-j} + \Pi \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{16}$$

with $\Psi_i = -\sum_{j=i}^{\infty} A_j$ and $\Pi = \sum_{j=1}^{\infty} A_j$. Under condition (v), then $(I - \Pi)^{-1}$ is the reduced-form, long-run impact matrix. For P the structural rotation matrix such that $P\mathbf{u}_t = \boldsymbol{\varepsilon}_t$, then the structural long-run impact matrix is

$$\Phi_\infty = \left(P^{-1} - P^{-1}\Pi\right)^{-1} = (I - \Pi)^{-1} P.$$

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

$$\Phi_\infty \Phi_\infty' = (I - \Pi)^{-1} PP' (I - \Pi')^{-1} = (I - \Pi)^{-1} \Sigma_\varepsilon (I - \Pi')^{-1} = QQ' \tag{17}$$

and hence $P = (I - \Pi) Q$.

An estimate of Π can be easily obtained by redefining M_z in expression (10) as $\widetilde{Z} \equiv \left(\mathbf{1}_{(T-k)\times 1},\ \Delta\mathbf{y}_{t-1},\ ...,\ \Delta\mathbf{y}_{t-k+1}\right)$ with $\widetilde{M}_z = I_{T-k} - \widetilde{Z}\left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}$ so that

$$\widehat{\Pi} = \left(\mathbf{y}_0' \widetilde{M}_z \mathbf{y}_{-1}\right)\left(\mathbf{y}_{-1}\widetilde{M}_z\mathbf{y}_{-1}\right)^{-1}$$

with

$$\sqrt{T}\left(\widehat{\pi} - \pi_0\right) \xrightarrow{d} (0, \Omega_\pi)$$

where $\pi = vec(\Pi)$ and

$$\Omega_\pi = \left(\mathbf{y}_{-1}\widetilde{M}_z\mathbf{y}_{-1}\right)^{-1} \otimes \Sigma_\varepsilon \tag{18}$$

Σ_ε can then be estimated as $\widehat{\Sigma}_\varepsilon = \frac{\widetilde{\mathbf{v}}_1\widetilde{\mathbf{v}}_1'}{T-k}$; $\widetilde{\mathbf{v}}_1 = \widetilde{M}_z\mathbf{y}_0 - \widetilde{M}_z\mathbf{y}_{-1}\widehat{\Pi}$.

Given the estimates $\widehat{\Pi}, \widehat{\Sigma}_\varepsilon$, then an estimate of Q can be obtained from the Cholesky decomposition described in expression (17) and the structural impulse responses can be constructed as

$$\widehat{\Phi}(0,h) = \widehat{B}(0,h)\left(I - \widehat{\Pi}\right)\widehat{Q} \tag{19}$$

where the asymptotic normality of each element ensures that

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

but where now

$$\begin{aligned}\Omega_\phi &= \frac{\partial\widehat{\phi}_T}{\partial\widehat{b}_T}\Omega_B\frac{\partial\widehat{\phi}_T}{\partial\widehat{b}'_T} + \frac{\partial\widehat{\phi}_T}{\partial\widehat{\pi}_T}\Omega_\pi\frac{\partial\widehat{\phi}_T}{\partial\widehat{\pi}'_T} + \\ &\quad \frac{\partial\widehat{\phi}_T}{\partial\widehat{q}_T}\left[\frac{\partial\widehat{q}_T}{\partial\widehat{\pi}_T}\Omega_\pi\frac{\partial\widehat{q}_T}{\partial\widehat{\pi}'_T} + \frac{\partial\widehat{q}_T}{\partial vech\left(\widehat{\Sigma}_\varepsilon\right)}\Omega_\Sigma\frac{\partial\widehat{q}_T}{\partial vech\left(\widehat{\Sigma}_\varepsilon\right)'}\right]\frac{\partial\widehat{\phi}_T}{\partial\widehat{q}'_T}\end{aligned} \tag{20}$$

with $\widehat{q}_T = vech\left(\widehat{Q}_T\right)$; Ω_Σ is the covariance matrix of $vech\left(\widehat{\Sigma}_\varepsilon\right)$ and we make use of the fact that $\widehat{q}_T$ and $vech\left(\widehat{\Sigma}_\varepsilon\right)$ are uncorrelated since $E\left[\mathbf{y}'_{-1}\widetilde{M}_z\widetilde{\mathbf{v}}_1/(T-k)\right] \xrightarrow{p} 0$.

The appendix explains how the following results are derived:

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

and L_r is the elimination matrix introduced in expression (15); Ω_B is given by expression (11); Ω_Σ is given by expression (15); and Ω_π is given by expression (18).

When condition (v) is violated, it is instructive to rewrite expression (16) as

$$\Delta \mathbf{y}_t = \sum_{j=1}^{\infty} \Psi_j \Delta \mathbf{y}_{t-j} + \Gamma \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\Gamma = -(I - \Pi)$. Violation of condition (v) occurs when $rank(\Gamma) < r$, in which case Γ has a non-standard asymptotic distribution and Γ is superconsistent, i.e., convergence in distribution occurs at rate T instead of the conventional rate $\sqrt{T}$. Such rank conditions can be tested with a Johansen cointegration test (see Hamilton, 1994, chapters 19 and 20). If $rank(\Gamma) = 0$, then the system has exactly r unit roots and clearly the long-run impact matrix is simply I. The superconsistency of Γ simplifies the derivation of (20): since $\widehat{\Gamma}$ (and hence $\widehat{\Pi}$) converges at rate T, then the distribution of $\widehat{\phi}_T$ is dominated by the terms converging at rate $\sqrt{T}$ and hence expression (20) simplifies, considerably, to

$$\Omega_\phi = \frac{\partial \widehat{\phi}_T}{\partial \widehat{b}_T} \Omega_B \frac{\partial \widehat{\phi}_T}{\partial \widehat{b}_T'} + \frac{\partial \widehat{\phi}_T}{\partial \widehat{q}_T} \left[\frac{\partial \widehat{q}_T}{\partial vech\left(\widehat{\Sigma}_\varepsilon\right)} \Omega_\Sigma \frac{\partial \widehat{q}_T}{\partial vech\left(\widehat{\Sigma}_\varepsilon\right)'} \right] \frac{\partial \widehat{\phi}_T}{\partial \widehat{q}_T'} \tag{21}$$

where the formulas for each of the terms in the previous expression are the same as those already derived above. Expression (21) is therefore parallel to expression (14) and serves to highlight that any identification scheme based on constraints that do not depend on parameter estimates (irrespective of whether these are zero coefficient restrictions or some other form of linear restriction) will generate a structural covariance matrix that can be calculated on the basis of expression (14). Even when the restrictions imposed depend on

coefficient estimates (such as long-run identification restrictions but not limited to these), expression (14) is still valid as long as these coefficients are superconsistent and have no effect on the distribution of terms converging at rate $\sqrt{T}$.

5 Policy Trade-offs in the U.S. and in the U.K.

This section examines the economies and policy choices of the U.S. and the U.K. with a global system of seven variables divided into two sub-systems of three variables each (which can be thought of a small scale version of a New-Keynesian model, see Walsh, 2003) and an equation for the bilateral exchange rate between the two countries. There are several reasons why I choose this application. Having the subsystems of two similarly developed economies with comparable economic institutions naturally elicits joint tests based on comparing the responses of both economies to similar stimuli. The relative size of the U.S. and the U.K. economies in turn provide a natural relative ordering of the two sub-systems that I exploit for structural identification (see Keating, 1996, on block-recursive identification assumptions). Within blocks, I follow the standard practice of recursively ordering economic activity, prices and then interest rates (e.g. see Christiano, Eichenbaum and Evans, 1999, and references therein). The system's relative high dimensionality also serves to showcase the robustness of the methods introduced above.

The data includes the unemployment rate (in percent), consumer price inflation (in percent) and the federal funds rate (in percent) for the U.S.; the unemployment rate (in percent), retail price inflation (in percent), and the Bank of England's lending rate (in percent) for

the U.K.; and the U.S. Dollar/British Pound exchange rate (in logs). The data is available monthly, beginning January 1971 and ending December 2005.

This system results in 49 impulse responses that I estimate jointly by local projections with equation (10) over a horizon of twelve periods (one year). The lag length for the projections is selected automatically by AIC_c[2] to be four. Figures 3 and 4 display these 49 impulse response functions. In figure 3, each panel displays the following information: (1) conditional 95% confidence bands; (2) unconditional 95% confidence bands; (3) the p-value of the joint significance test of the responses labeled "Joint" and; (4) the p-value of the cumulative joint significance test labeled "Cum." The narrower bands always correspond to the conditional confidence bands. Figure 4 displays fan charts based on the response percentile bounds instead.

A number of questions of economic interest arise from this system, for example: (1) what is the sensitivity of the policy interest rate to shocks in the unemployment, inflation, and exchange rates; (2) what is the response of unemployment, inflation and exchange rates to monetary shocks; (3) what is the sensitivity of domestic policy rates to shocks in the foreign policy rate; (4) what is the response of exchange rates to shocks in inflation; and (5) what is the response of the unemployment rate when the central bank responds more aggressively to inflation shocks, to list a few.

Answers to question (1) can be used to compare the relative emphasis that each central

[2] AIC_c refers to the correction to AIC introduced in Hurvich and Tsai (1989), which is specifically designed for autoregressive models. There were no significant differences when using SIC or the traditional AIC.

bank places on growth and price stabilization. Question (2) establishes the effectiveness of monetary policy; question (3) determines whether policy changes in one central bank influence policy changes in the other; question (4) speaks loosely about the relative merits of the purchasing power parity condition; and question (5) measures how different is the response of the unemployment rate when the central bank's response to shocks in inflation is more aggressive. Each of these five questions requires different types of inference based on joint hypothesis tests and counterfactual simulations of the type introduced in previous sections.

5.1 Results

The response of the U.S. unemployment rate to a shock in U.S. inflation and displayed in the first row, second column of the panel of impulse response functions in figures 3 and 4, showcases the importance of joint inference over unconditional statements based on individual coefficients. Traditional 95% confidence bands suggest this response is essentially zero except perhaps between four to seven months after impact. Meanwhile, the much narrower conditional 95% confidence bands suggest the response is clearly positive three months after impact and thereafter, a conclusion that is well supported by the p-values of the joint significance and the joint cumulative tests with values 0.05 and 0.048 respectively.

A different illustration is provided by the response of the U.S. Dollar/British Pound exchange rate to a shock in U.S. inflation and displayed in row seven, column two of figures 3 and 4. This is an example of a response where the joint significance test has a p-value of 0.344 but the joint cumulative test has a p-value of 0.055, a result that confirms the

pattern displayed by the conditional bands, which suggest that the U.S. Dollar significantly depreciates between the third and the eleventh/twelfth month after impact.

Returning now to the questions posed in the previous subsection, recall that in question (1) we are trying to assess the relative sensitivity of the central bank to shocks in unemployment, inflation and exchange rates. This information is summarized in columns one, two and seven of figures 3 and 4, row three for the U.S. and row six for the U.K. In both countries, interest rates drop significantly by about 50 basis points in response to a shock in the unemployment rate, with both significance tests and profile bands indicating a statistically significant response (the p-values are essentially zero). However, tests of the joint and cumulative equality of the responses between the two countries are rejected pretty decisively (with p-values of 0.000 and 0.000 respectively). Perhaps this last result is not surprising: the U.S. drops interest rates more quickly and keeps them low for a longer period than the U.K. does. In contrast, both countries do not appear to respond to a shock in inflation: both significance tests have p-values well above the normative 0.05 value and the conditional bands are generally not significant at any horizon. Finally, while joint significant tests do not indicate that interest rates respond to fluctuations in exchange rates, the U.S. displays a significant cumulative effect of interest rates (p-value = 0.015) in response to a depreciation of the U.S. Dollar, a result that is corroborated by the conditional bands.

Question (2) examines the relative effectiveness of monetary policy. The relevant panels are rows one, two and seven of column three for the U.S. and rows four, five and seven in column six for the U.K. The response of unemployment to a shock in interest rates is

very similar in both countries although only the joint significance test for the U.K. has a p-value bordering on significance at 0.075. Unemployment remains essentially flat six to seven months after impact and then steadily climbs (in both countries, the conditional bands suggest this climb is significant). More formally, joint and cumulative equality tests have p-values of 0.52 and 0.83 which confirm both countries respond in a similar way.

Inflation in both countries tends to climb in response to an interest rate shock but only significantly in the U.K. (the joint cumulative test has a p-value of 0.004). The joint equality test cannot reject the null, attaining a p-value of 0.43 although the joint cumulative equality test does, with a p-value of 0.055. Both results seem at odds with what economic theory would predict although this price puzzle has been detected many times before in the U.S. (see e.g. Sims, 1992).

Before discounting these results, it is important to examine the response of the exchange rate. The U.S. Dollar tends to appreciate somewhat throughout the year after impact when interest rates increase (the joint cumulative test has a p-value of 0.090 and the conditional bands border on the zero line). The British pound significantly depreciates on impact but then appreciates for the remainder of the year, although not in a statistically significant way. Hence, although the initial responses of the exchange rate are consistent, the uncovered interest rate parity condition seems to hold only somewhat for the U.K.

In terms of the effect that each central bank has on the other (row three, column six for the U.S. and row six, column three for the U.K.), neither country exhibits a significant response although these tend to remain on the positive side, suggesting that both countries

tend to move interest rates in the same direction.

The final question has to do with the response of the exchange rate to a shock in inflation and is displayed by the panels in row seven, column two for the U.S. and column five for the U.K. The U.S. Dollar tends to depreciate in response to a positive shock in inflation, as purchasing power parity would predict: the joint cumulative test has a p-value of 0.055 and the conditional bands are significant for about ten out of the twelve months displayed. In contrast, the British Pound does not display any appreciable response at any horizon, an observation confirmed by the joint significance and cumulative tests with p-values of 0.78 and 0.64 respectively.

Finally, consider a counterfactual simulation that investigates the effects of a more aggressive response of monetary policy to inflation shocks in the U.S. The panel in the third row, second column of figures 3 and 4 displays the response of the U.S. federal funds rate to a shock in the U.S. inflation rate. This response is not statistically significant: the joint significance and cumulative tests have p-values of 0.950 and 0.546 respectively. Hence, I choose a counterfactual experiment that constrains the response of the fed funds rate to an inflation shock to be that corresponding to the upper, unconditional two standard-error band. Before reporting on the counterfactual itself, notice that the response of the U.S. unemployment rate to a positive inflation shock results in a statistically significant response of unemployment as evinced in row one, column 2 of figure 5 by the conditional confidence bands and joint and cumulative tests, with p-values 0.050 and 0.048 respectively. A test for the probity of this counterfactual has a p-value of 0.73 (that is, one cannot reject the coun-

terfactual is statistically equal to the historical path) or said differently, the counterfactual is 27% away from the historical path in probability units, which is on the conservative side.

The counterfactual experiment itself is displayed in figure 5 and shows that a more aggressive response of the central bank to an inflation shock surprisingly causes the unemployment rate to be lower than it would otherwise be. Figure 5 displays the original and the counterfactual responses along with the appropriate conditional bands under the counterfactual described in section 3. The cumulative increase of the unemployment rate in response to a 0.30% increase in inflation is 0.46% historically but only 0.12% in the counterfactual. The conditional error bands show some overlap early on but suggest that the historical and counterfactual responses are quite different six months after impact and beyond. These results are surprising considering that an increase in the federal funds rate generates higher unemployment according to the panel in row one, column three of figures 3 and 4.

6 Conclusion

If impulse response coefficients were drawn from independent distributions, we would expect their 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 individual estimates of these coefficients imprecise even when collectively, there is little ambiguity about the overall path. 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 easily understandable 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 how this distribution is arrived at.

While I leave to the reader to decide how best to arrive at this point (and I expect others will experiment with the bootstrap, Bayesian MCMC and other simulation techniques), I provide the necessary results for impulse responses estimated by local projections. I do this for several reasons having to do with the approximation properties of local projections over traditional VARs (see Jordá, 2005); because it is considerably simpler to derive the results with this estimator; and because these results have not been collected elsewhere in the literature. Further, insofar as local projections can be easily extended for nonlinearities, it is useful to provide results that can serve as a building block for further research.

Impulse responses are well-defined and rather accurately estimable summary statistics of the data, much like the sample mean or a sample correlation. The tools provided in this paper will look familiar even to readers with only informal knowledge of time series techniques. In my view, this transparency is a major asset in improving the wide applicability and communication of experiments among researchers.

7 Appendix

I provide here the detailed derivations required to derive the covariance of the structural impulse responses derived by imposing long-run identification assumptions. First notice that there are three different ways of expressing the vec version of expression (19), specifically

$$\widehat{\phi}_T = vec\left(\widehat{\Phi}\left(0,h\right)\right) = \begin{cases} \left(\widehat{Q}'\left(I-\widehat{\Pi}\right)\otimes I\right)\widehat{b}_T \\ \left(\widehat{Q}\otimes\widehat{B}\left(0,h\right)\right)vec\left(I-\widehat{\Pi}\right) \\ \left(I\otimes\widehat{B}\left(0,h\right)\left(I-\widehat{\Pi}\right)\right)vec\left(\widehat{Q}\right) \end{cases}$$

from where we obtain $\frac{\partial\widehat{\phi}_T}{\partial\widehat{\pi}_T}$, $\frac{\partial\widehat{\phi}_T}{\partial\widehat{\pi}_T}$, and $\frac{\partial\widehat{\phi}_T}{\partial\widehat{q}_T}$ by realizing that $dvec\left(I-\widehat{\Pi}\right) = -dvec\left(\widehat{\Pi}\right)$ and since Q is lower triangular, $vec\left(Q\right) = L'vech(Q)$.

Next, I derive the expressions for $\frac{\partial\widehat{q}_T}{\partial\widehat{\pi}_T}$ and $\frac{\partial\widehat{q}_T}{\partial vech\left(\widehat{\Sigma}_\varepsilon\right)}$ by first noticing that

$$\left(I-\Pi\right)^{-1}\Sigma_\varepsilon\left(I-\Pi'\right)^{-1} = QQ'$$

so that

$$d\left(I-\Pi\right)^{-1}\Sigma_\varepsilon\left(I-\Pi'\right)^{-1} + \left(I-\Pi\right)^{-1}\Sigma_\varepsilon d\left(I-\Pi'\right)^{-1} + \tag{22}$$

$$\left(I-\Pi\right)^{-1}d\Sigma_\varepsilon\left(I-\Pi'\right)^{-1} = dQQ + QdQ'$$

Begin by setting $d\left(I-\Pi\right)^{-1} = 0$ then

$$\left(I-\Pi\right)^{-1}d\Sigma_\varepsilon\left(I-\Pi'\right)^{-1} = dQQ + QdQ'$$

Taking the vec operator on both sides of this expression

$$\begin{aligned}\left[(I-\Pi)^{-1}\otimes(I-\Pi)^{-1}\right]dvec\left(\Sigma_{\varepsilon}\right) &= (Q\otimes I)\,dvec\left(Q\right)+(I\otimes Q)\,K_{rr}dvec(Q)\\ dvec\left(\Sigma_{\varepsilon}\right) &= \left[(I-\Pi)\otimes(I-\Pi)\right]\left(I_{r^2}+K_{rr}\right)(Q\otimes I)\,dvec\left(Q\right)\end{aligned}$$

using the same rule for the right hand side as in the derivation of the short-run identification case. Finally, using the elimination matrix L_r introduced in expression (15) and noticing that since Q is lower triangular then $L_r'vech\left(Q\right)=vec\left(Q\right)$, we arrive at the desired result

$$\frac{\partial q}{\partial vech\left(\Sigma_{\varepsilon}\right)}=\left\{L\left[(I-\Pi)\otimes(I-\Pi)\right]\left(I_{r^2}+K_{rr}\right)(Q\otimes I)\,L'\right\}^{-1}$$

To derive $\frac{\partial\widehat{q}_T}{\partial\widehat{\pi}_T}$, return to expression (22) and instead set $d\Sigma_{\varepsilon}=0$ so that

$$d\left(I-\Pi\right)^{-1}\Sigma_{\varepsilon}\left(I-\Pi'\right)^{-1}+\left(I-\Pi\right)^{-1}\Sigma_{\varepsilon}d\left(I-\Pi'\right)^{-1}=dQQ+QdQ'$$

Taking the vec operator on both sides of the expression and using similar manipulations as in the previous derivation, it is easy to see that we arrive at

$$\left[\left(I-\Pi\right)^{-1}\Sigma_{\varepsilon}\otimes I\right]dvec\left\{\left(I-\Pi\right)^{-1}\right\}=(Q\otimes I)\,dvec(Q)$$

where the term $(I_{r^2}+K_{rr})$ cancels on both sides of the previous expression. It is straight forward to see then that

$$dvec\left\{\left(I-\Pi\right)^{-1}\right\}=\left[\left(I-\Pi\right)^{-1}\otimes\left(I-\Pi'\right)^{-1}\right]dvec\left(\Pi\right)$$

and since $L_r' vech\,(Q) = vec\,(Q)$, then we arrive at the desired result,

$$\frac{\partial q}{\partial \pi} = \left\{ (Q \otimes I)\, L_r' \right\}^{-1} \left\{ (I - \Pi)^{-1} \Sigma_\varepsilon \otimes I \right\} \left\{ (I - \Pi')^{-1} \otimes (I - \Pi)^{-1} \right\}$$

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.

Christiano, Lawrence J., Martin Eichenbaum and Charles L. Evans (1999) "Monetary Policy Shocks: What Have We Learned and to What End?" in **Handbook of Macroeconomics, v. 1**, John B. Taylor and Michael Woodford (eds.). Amsterdam: North-Holland.

Cochrane, John H. (1998) "What Do the VARs Mean?: Measuring the Output Effects of Monetary Policy," *Journal of Monetary Economics,* 41(2), 277-300.

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

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

Hoover, Kevin D. and Òscar Jordà (2001) "Measuring Systematic Monetary Policy," *Review*, Federal Reserve Bank of st. Louis, 83(4): 113-138.

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.

Keating, John (1996) "Structural Information in Recursive VAR Orderings," *Journal of Economic Dynamics and Control,* 20, 1557-1580.

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.

Lucas, Robert E., Jr. (1976) "Econometric Policy Evaluation: A Critique," *Journal of Monetary Economics*, Supplementary Series 1976, 1(7): 19-46.

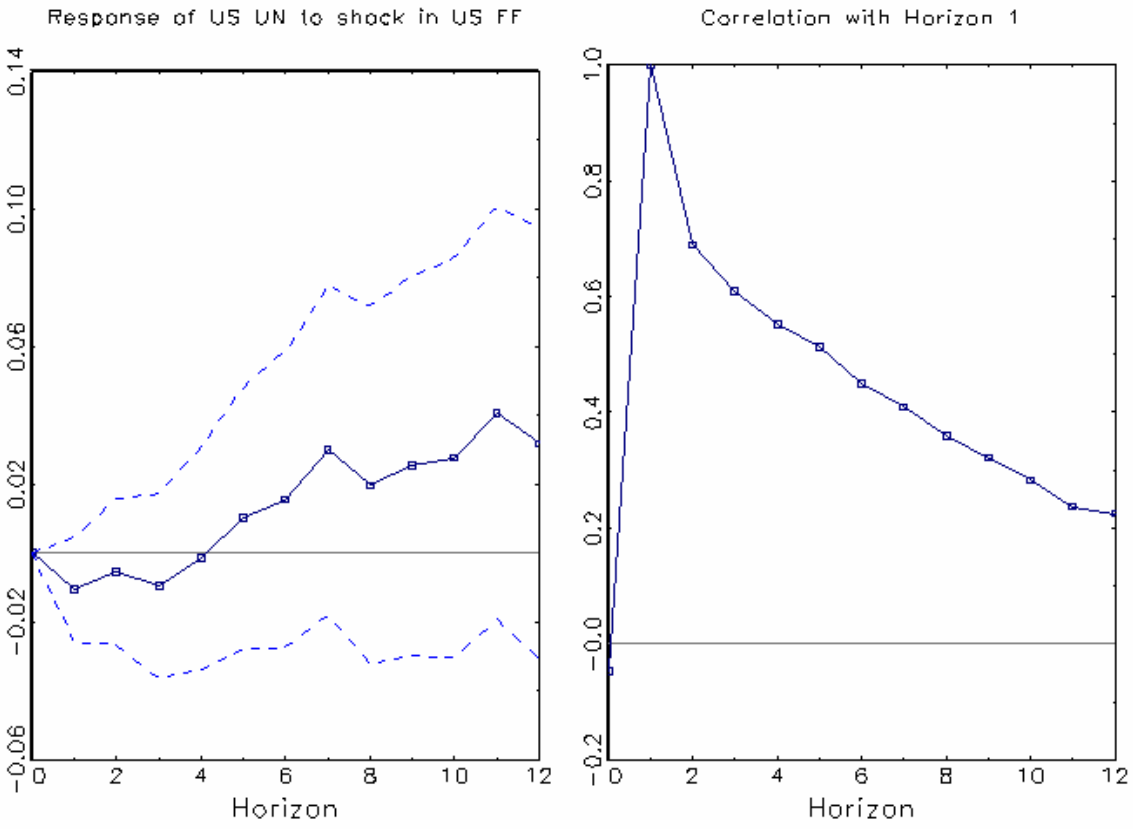
Lütkepohl, Helmut (2005) **New Introduction to Multiple Time Series**. Berlin, Germany: Springer-Verlag.

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.

Walsh, Carl E. (2003) **Monetary Theory and Policy, second edition.** Cambridge, MA: MIT Press.

Figure 1 – Correlation Among Impulse Response Coefficients

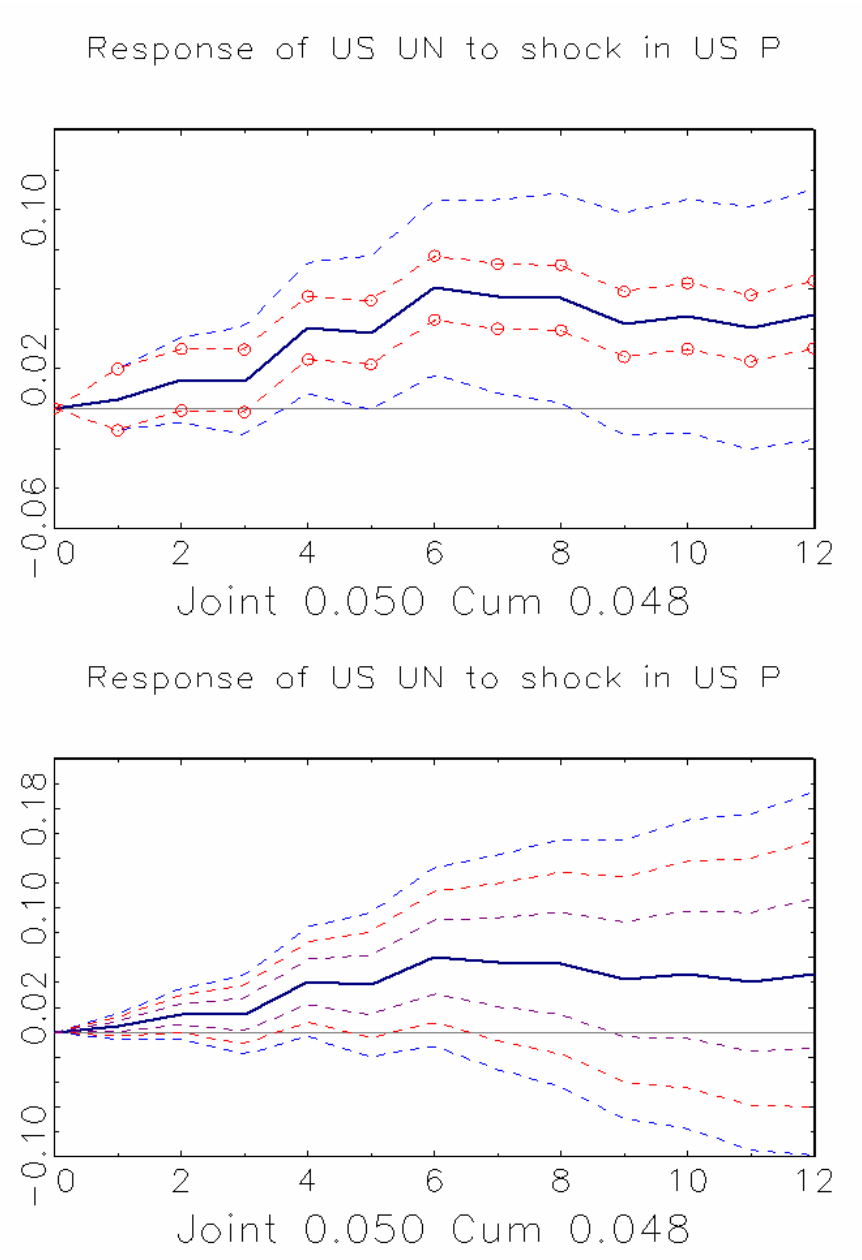


Correlation Matrix of Impulse Response Coefficients

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.7	0.6	0.6	0.5	0.4	0.4	0.4	0.3	0.3	0.2	0.2
2	-0.1	0.7	1.0	0.8	0.7	0.7	0.6	0.6	0.5	0.5	0.4	0.4	0.3
3	0.0	0.6	0.8	1.0	0.9	0.8	0.8	0.7	0.6	0.6	0.5	0.5	0.4
4	0.0	0.6	0.7	0.9	1.0	0.9	0.9	0.8	0.8	0.7	0.6	0.6	0.5
5	0.0	0.5	0.7	0.8	0.9	1.0	0.9	0.9	0.8	0.8	0.7	0.7	0.6
6	0.0	0.4	0.6	0.8	0.9	0.9	1.0	0.9	0.9	0.9	0.8	0.7	0.7
7	0.0	0.4	0.6	0.7	0.8	0.9	0.9	1.0	0.9	0.9	0.9	0.8	0.7
8	0.0	0.4	0.5	0.6	0.8	0.8	0.9	0.9	1.0	1.0	0.9	0.9	0.8
9	0.0	0.3	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	0.9	0.9
10	0.0	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	0.9
11	0.0	0.2	0.4	0.5	0.6	0.7	0.7	0.8	0.9	0.9	1.0	1.0	1.0
12	0.0	0.2	0.3	0.4	0.5	0.6	0.7	0.7	0.8	0.9	0.9	1.0	1.0

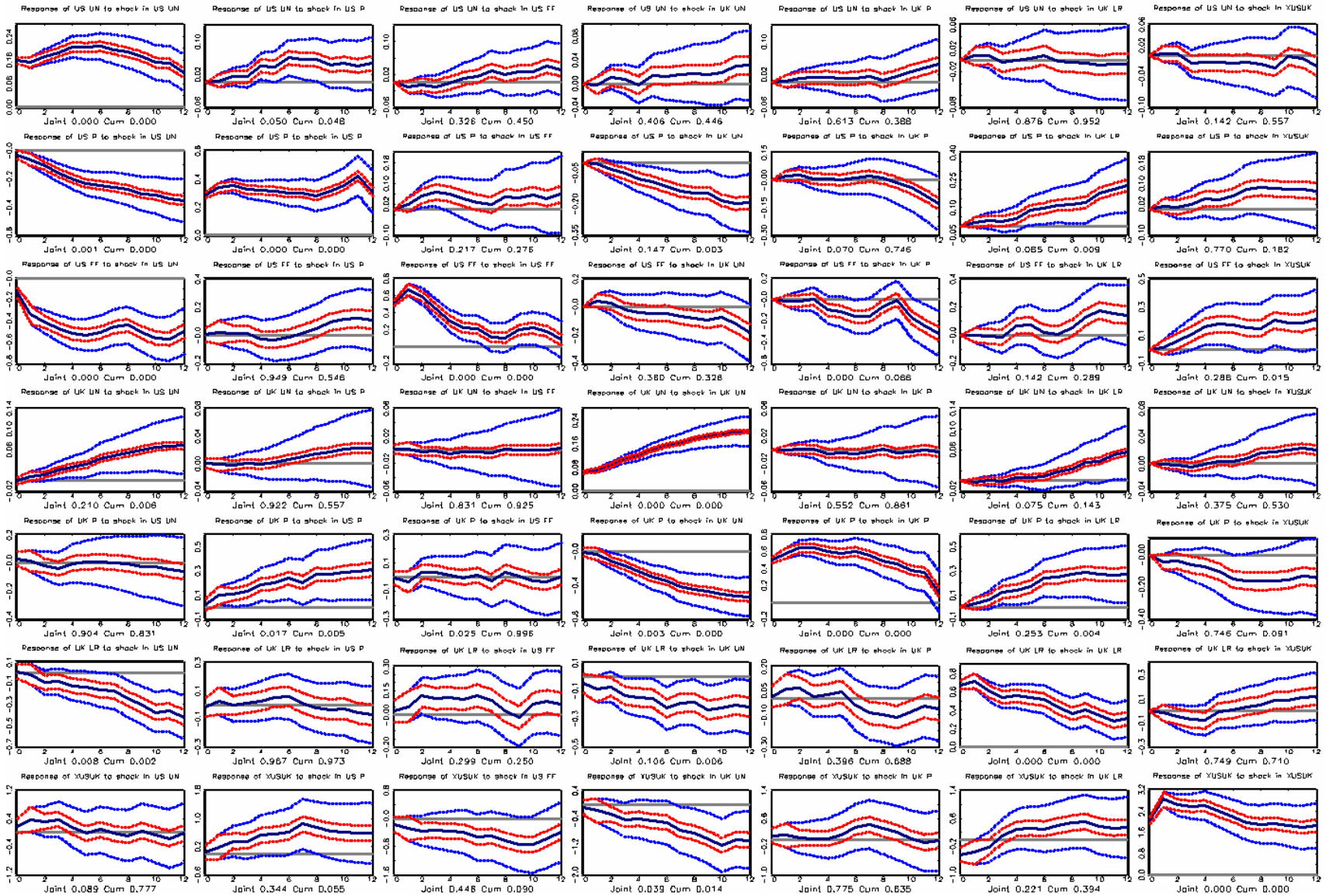
Notes: First panel displays impulse response and traditional two standard-error bands. Correlation matrix displays correlations among all the elements of the impulse response displayed.

Figure 2 – 95% Conditional Confidence Bands versus 95 % Unconditional Confidence Bands and Response Percentile Bounds



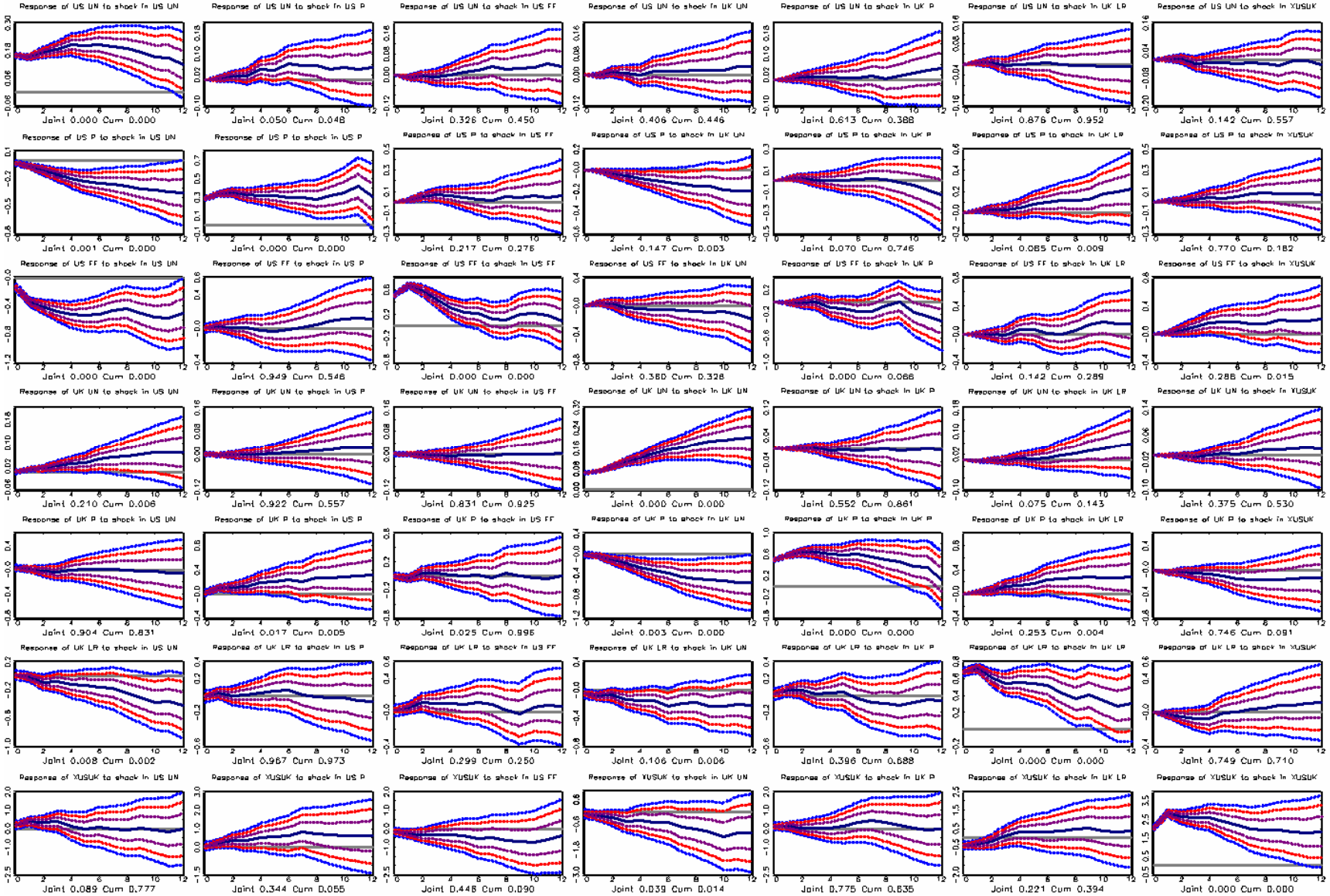
Notes: Top panel: The solid line indicates the estimates impulse response. The narrow bands with circles indicate the 95% conditional confidence bands. The dashed bands indicate the usual 95% unconditional confidence bands. "Joint" refers to the p-value of the null that all the coefficients of the impulse response are zero. "Cum" refers to the p-value of the test that the cumulative effect is jointly zero. Bottom panel: fan chart with the response 95th , 50th , and 1st percentile bounds.

Figure 3 – Impulse Responses and 95% Conditional and Unconditional Confidence Bands



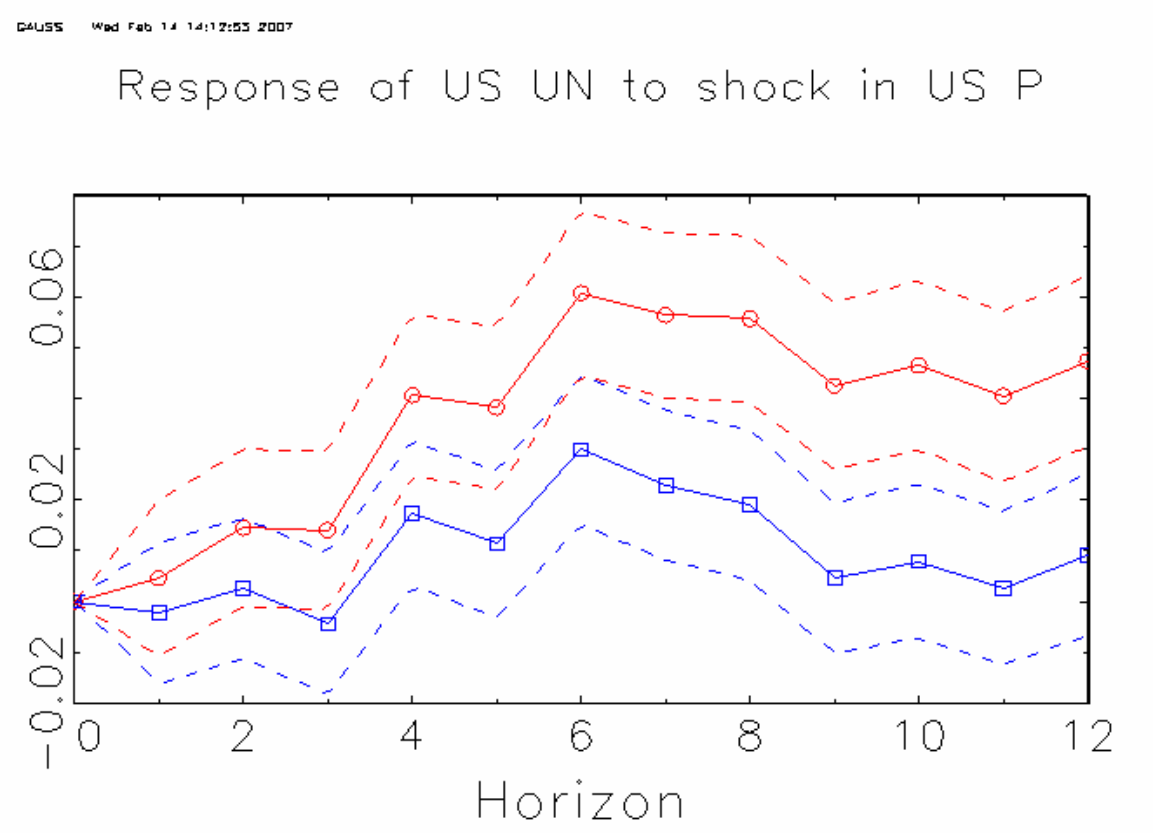
Notes: narrow bands are conditional confidence bands, wide bands are the usual unconditional confidence bands. "Joint" refers to the p-value of the joint significance test. "Cum" refers to the p-value of the joint cumulative effect.

Figure 4 – Impulse Response Fan Charts: Response 95th , 50th , and 1st percentile Bounds



Notes: "Joint" refers to the p-value of the joint significance test. "Cum" refers to the p-value of the joint cumulative effect.

Figure 5 – Counterfactual Experiment: Response of U.S. Unemployment to a U.S. Inflation Shock when the Response of the Federal Funds Rate to an Inflation Shock is made more aggressive



p-value of plausibility test: 0.73

Notes: Solid line with circles is the original impulse response. Solid line with squares is the counterfactual response. The bands around the responses correspond to 95% conditional confidence bands. Note that the counterfactual bands are calculated with the counterfactual covariance matrix, not the original covariance matrix. In probability metric, the counterfactual response of the U.S. federal funds rate to a shock in U.S. inflation is only 27% away (in probability units) from the historical path.