UNIFORM INFERENCE IN SVARS IDENTIFIED WITH EXTERNAL INSTRUMENTS

By José L. Montiel-Olea, James H. Stock, Mark W. Watson[1]

This paper studies Structural Vector Autoregressions identified using *external instruments*. These external instruments are taken to be correlated with the target shock (e.g., the oil instrument is *relevant*) and to be uncorrelated with other macroeconomic shocks of the model (e.g., the oil instrument is *exogenous*).

The correlation between the external instrument and the target structural shock affects the validity of standard inference in large samples. With this observation in mind, we propose a new confidence set for the coefficients of the Structural Impulse-Response function and we show that its asymptotic confidence level is not affected by nuisance parameters. The implementation of our confidence set requires no more work than solving a single-variable quadratic equation.

In an empirical application studying the dynamic effects of a structural oil shock using U.S. monthly data, we compare standard confidence sets with our inference procedure. As the theory suggests, we find substantial differences in cases where the external instrument is weakly correlated with the reduced-form error in the oil price equation.

Keywords: Structural Vector Autoregressions, Uniform Inference, Instrumental Variables.

1. INTRODUCTION

An increasingly important line of research in Structural Vector Autoregressions (SVARs) uses information in variables not included in the system to identify structural dynamic causal effects, which in VAR terminology are *structural impulse-response functions*. The work of Romer and Romer (1989) is the seminal reference in this literature. Their reading of the minutes of the Federal Reserve Board allowed them to pinpoint some moments at which monetary policy decisions were arguably exogenous; i.e., independent of other economic shocks at the time. Their work produced a time series of binary indicators of monetary policy decisions. A large number of subsequent papers have adopted Romer and Romer (1989)'s *narrative approach* to construct time series that are treated as exogenous shocks; see Stock and Watson (2012) for a discussion of a number of papers.

[1]This version: December 28, 2015. First draft: February 2014

Most of the papers in this literature have treated the new external time series as the structural shocks that enter the SVAR model. In our view, these external series are not, strictly speaking, the shocks of interest. Rather, these are random variables presumably correlated with the actual shocks of the model (i.e., the external time series is *relevant* for the target shock) and uncorrelated with other shocks in the economy (i.e., the external time series is *exogenous*). It seems natural, thus, to treat the constructed time series as an *external instrument*: the macroeconometric counterpart of microeconometric instrumental variables constructed using quasi-experiments.

After normalizing the scale of the structural impulse-response function (IRF), it is strightforward to show that the covariance between the external instruments and the SVAR reduced-form innovations can be used to identify the IRF coefficients (see Stock and Watson (2012) p. 107). The idea of explicitly using a constructed series as an instrumental variable to estimate a structural IRF dates at least to Hamilton (2003). The identification framework used in this paper was first suggested in Stock and Watson (2008) and was developed independently by Mertens and Ravn (2013).

The identification result in Stock and Watson (2012) yields a natural plug-in estimator for the IRF coefficients. Unfortunately, this plug-in estimator faces similar limitations as the Instrumental Variables (IV) estimator in a just-identified linear model. In particular, when the correlation between the external instrument and the target structural shock is *small*—a possibility that practitioners cannot rule out a priori—the standard confidence sets based on the plug-in estimator need not have correct coverage even in large samples.[1]

To address this problem, this paper proposes a confidence set for the structural IRF coefficients whose asymptotic confidence level is not affected by nuisance parameters. To formalize this claim, we establish the *uniform consistency in level* of our confidence set over a large class of data generating processes (DGPs). In terms of implementation, we show that reporting our confidence set requires no more work than solving a single variable quadratic equation. The coefficients of such equation depend on the reduced-form VAR parameters and the covariance between reduced-form residuals and external instruments.

[1]Our results show that even if the correlation between the external instrument and the target shock is not small, the asymptotic coverage of the standard confidence set based on the plug-in estimator is differs from the nominal confidence level whenever the vector of contemporaneous responses to a target shock is small.

By reporting our confidence set, practitioners do not need to worry about situtaions in which the correlation between the external instrument and each of the estimated reduced-form residuals is close to zero: there is no need of formal or informal pre-tests for *weak instruments*. For a recent proposal of a pre-test for weak instruments in SVARs identified with external instruments see Lunsford (2015).

The rest of this paper is organized as follows. Section 2 presents the model and the main identification result. Section 3 describes our confidence set and establishes its *uniform consistency in level* over a large class of DGPs. Section 4 discusses the implementation of our confidence set in terms of the roots of a single-variable quadratic equation. Section 5 presents illustrative examples of our approach. All the proofs are collected in the Appendix.

Generic Notation: If A is a matrix, A_{ij} denotes the ij-th element of A, $\text{vec}(A)$ denotes the vectorization of A, and $A \otimes B$ denotes the Kronecker product between matrices A and B. The vector $e_i \in \mathbb{R}^n$ denotes the i-th column of the identity matrix of dimension n. If B is a matrix of dimension $n \times n$, $B_i \equiv Be_i$ denotes its i-th column.

2. BASIC MODEL AND EXTERNAL INSTRUMENTS

2.1. *The Model*

The $n \times 1$ times series Y_t is assumed to follow a reduced-form stationary vector autoregression (VAR) with p lags:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + \eta_t, \tag{2.1}$$

where η_t denotes the *reduced-form* VAR innovations. There is a random vector of *structural innovations*, ε_t, such that:

$$\eta_t = B\varepsilon_t, \tag{2.2}$$

where B is an unknown, invertible matrix of dimension $n \times n$. The structural innovations are assumed to be serially uncorrelated distributed according to some probability measure that satisfies:

$$\mathbb{E}[\varepsilon_t] = \mathbf{0}_{n\times 1}, \quad \mathbb{E}[\varepsilon_t \varepsilon_t'] = D \equiv \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2).$$

In addition, the stochastic process in equation (2.1) is assumed to admit the *structural moving average* representation:

$$Y_t = \sum_{k=0}^{\infty} C_k(A) H \varepsilon_{t-k}, \text{ where } A = (A_1, A_2, \ldots, A_p), \tag{2.3}$$

and

$$C_k(A) \equiv \sum_{m=1}^{k} C_{k-m}(A) A_m, \quad k \in \mathbb{N},\ C_0(A) \equiv \mathbb{I}_n, \tag{2.4}$$

with $A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116.

The object of interest in this paper is the response of variable i—at horizon k—to a shock of 's' standard deviations in $\varepsilon_{j,t}$. Such parameter is given by:

$$\text{IRF}_{k,i,j}(A, B, D, s) \equiv e_i' C_k(A) B e_j (s\sigma_j), \tag{2.5}$$

where Be_j is the j-th column of the matrix B and σ_j is the j-th diagonal element of the matrix D. We refer to the parameter $\text{IRF}_{k,i,j}(A, B, D, s)$ as the (k, i, j) coefficient of the structural impulse-response function with scale 's'.

Target Shock: Our paper focuses on identifying the responses to a single structural shock (e.g., and oil shock). We refer to the shock of interest as the *target shock* and—without loss of generality—we label this shock as $\varepsilon_{1,t}$. Note that the coefficients of the impulse-response function corresponding to the target shock are determined by B_1 (the first column of B).

Normalization: We set the scale of the impulse-response function in terms of the contemporaneous response of some pre-specified variable i^* (e.g., the price of oil). This means that the policy question we have in mind is of the form: *"what is the response of variable i to a target structural shock that increases variable i^* in x percent"*. In this case, the scale for the impulse-response function is given by:

$$s = \frac{x\%}{e_{i^*}' B_1 \sigma_1},$$

where we have assumed implicitly that $e_{i^*}' B_1 \neq 0$. Without loss of generality, we let the pre-specified variable i^* is the first component of Y_t. Therefore, the parameter of interest can be written as:

$$\lambda_{k,i}(A, B_1) = x(e_i' C_k(A) B_1)/(e_1' B_1), \tag{2.6}$$

which no longer depends on D.

2.2. *Identification using an external instrument*

The main assumption of this paper is that the econometrician observes a real-valued random variable z_t that can be used as a 'proxy' for the target shock $\varepsilon_{1,t}$. Examples of proxy variables in the literature include the measures of monetary shocks in Romer and Romer (2004); the measures of oil shocks in Hamilton (2003) and Kilian (2008); and the measure of government spending shocks in Ramey (2011).

Let $\{(Y_t, z_t)\}_{t=1}^{T}$ denote the data observed by the econometrician. Note that the data generating process—which we will denote by P—is controlled by the structural parameters (A, B) and a joint distribution F for the process $\{(\epsilon_t, z_t)\}_{t=1}^{\infty}$, which we assume to be stationary.

Instead of specifying a concrete time series model for $\{(\epsilon_t, z_t)\}_{t=1}^{\infty}$, we assume that F satisfies the following assumptions, which will allow us to treat the proxy variable z_t as an *external instrument.*

Assumption 1 (External Instrument) The distribution F of the process $\{(\epsilon_t, z_t)\}_{t=1}^{\infty}$ satisfies the following properties:

1.1. $\mathbb{E}_F[z_t \epsilon_{1,t}] = \alpha \quad \forall \quad t, \, \alpha \neq 0$

1.2. $\mathbb{E}_F[z_t \epsilon_{j,t}] = 0 \quad \forall \quad t, j \neq 1$

Our main assumption can be thought of as a time series analog of the definition of an instrumental variable in the microeconometrics literature. Assumption 1.1 is a *relevance* requirement for the proxy variable at hand: z_t has to be correlated with the structural shock of interest. Assumption 1.2 is an *exogeneity* assumption for z_t: the proxy variable is correlated with the structural shock $\varepsilon_{1,t}$, but uncorrelated with any other structural shock in the model.

A simple example of a statistical model for $\{(\epsilon_t, z_t)\}_{t=1}^{\infty}$ that gives rise to an external instrument is the following *measurement error* model:

$$z_t = f(\varepsilon_{1,t}) + v_t,$$

where the measurement error $\{v_t\}_{t=1}^{\infty}$ is an stationary process independent of $\{\varepsilon_t\}_{t=1}^{\infty}$.[2]

[2]Another possibility is the model:

$$z_t = f(\varepsilon_{1,t}) + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots a_p Y_{t-p} + v_t.$$

Indentification of the IRF coefficients: Under Assumption 1, the structural parameter B_1 is identified, up to scale, by the covariance between the external instrument z_t and the vector of VAR reduced-form errors:

$$\Gamma \equiv \mathbb{E}_F[z_t \eta_t] = B\mathbb{E}_F[z_t \varepsilon_t] = B(\alpha, 0, 0, \ldots 0)' = \alpha B_1.$$

This implies that the structural parameter of interest $\lambda_{i,k}(A, B_1)$ can be identified from (A, Γ) as:

$$\lambda_{k,i}(A, \Gamma) = x e_i' C_k(A) \mathbb{E}_F[z_t \eta_t] / e_1' \mathbb{E}_F[z_t \eta_t], \tag{2.7}$$

as long as $e_1'\mathbb{E}_F[z_t\eta_t] \neq 0$. This condition implies that our identification strategy relies on two assumptions: $\alpha \neq 0$ and $e_1' B_1 \neq 0$. These conditions hold if and only if $e_1'\mathbb{E}_F[z_t\eta_t] \neq 0$.

Identification of the shock of interest: In SVAR applications it is customary to use the identification results for the structural parameter B to recover the structural shocks of interest. Let $\Sigma \equiv \mathbb{E}_P[\eta_t \eta_t']$. Note that if $\alpha \neq 0$:

$$\begin{aligned}
\Gamma'\Sigma^{-1}\eta_t / e_1'\Gamma &= (\alpha B e_1)'(BDB')^{-1}\eta_t / \alpha e_1' B_1, \\
&\quad (\text{since } \mathbb{E}_P[\eta_t \eta_t'] = B\mathbb{E}_P[\varepsilon_t \varepsilon_t'] B') \\
&= e_1' B' (B')^{-1} D^{-1} B^{-1} B \varepsilon_t / e_1' B_1, \\
&\quad (\text{as } \eta_t \equiv B\varepsilon_t) \\
&= e_1' D^{-1} \varepsilon_t / e_1' B_1, \\
&= \varepsilon_{1,t} / e_1' B_1 \sigma_1^2.
\end{aligned}$$

Consequently,

$$x\Gamma'\Sigma^{-1}\eta_t / e_1'\Gamma = s(\varepsilon_{1,t}/\sigma_1), \tag{2.8}$$

where s is the scale for normalization defined in the previous section.

In both cases, the proxy variable is allowed to be serially correlated.

3. INFERENCE

3.1. *Overview*

We are interested in constructing a confidence set for the parameter $\lambda_{k,i}$ in equation (2.6). In addition to the assumption that z_t is an external instrument, we will impose restrictions on the class of data generating processes $\mathcal{P}$ for $\{Y_t, z_t\}_{t=1}^T$.

Our inference approach will be based on the statistics:

$$\widehat{A}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^T Y_t X_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^T X_t X_t'\Big)^{-1}, \quad X_t \equiv (Y_{t-1}, Y_{t-2}', \dots Y_{t-p}')' \tag{3.1}$$

and

$$\widehat{\Gamma}_T \equiv \frac{1}{T}\sum_{t=1}^T z_t \widehat{\eta}_t, \quad \widehat{\eta}_T = Y_t - \widehat{A}_T X_t. \tag{3.2}$$

The statistic in (3.1) is the Ordinary Least-Squares (OLS) estimator of the VAR autoregressive coefficients. The statistic in (3.2) is the sample covariance between the external instrument and the OLS estimator of the VAR reduced-form residuals.

An intuitive plug-in estimator for the parameter of interest is given by:

$$\widehat{\lambda}_{k.i} = x e_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T / e_1'\widehat{\Gamma}_T. \tag{3.3}$$

It is not difficult to show (and we do it in Appendix B) that if we allow for data generating processes (DGPs) in which either α or $e_1' B_1$ are arbitrarily close to zero, the plug-in estimator in (3.3) is not *uniformly consistent* and the associated Wald statistic is not *regular*—in the sense of Van der Vaart (2000), p. 115. Since such DGPs cannot be excluded a priori, it is necessary to consider alternative procedures to conduct inference for $\lambda_{k,i}$.[3]

Our suggested inference approach builds on a simple observation. Assume, for the sake of exposition, that the parameter A is known, note that:

$$\sqrt{T}(xe_i'C_k(A) - \lambda_{k,i}e_1')\widehat{\Gamma}_T = (xe_i'C_k(A) - \lambda_{k,i}e_1')\sqrt{T}(\widehat{\Gamma}_T - \Gamma) + \sqrt{T}(xe_i'C_k(A) - \lambda_{k,i}e_1')\Gamma.$$

[3]The statistical issues with the plug-in estimator in (3.3) are the same that arise in the estimation of the ratio of two normal means when the denominator is close to zero [Fieller (1954)] or the estimation of a just-identified linear Instrumental Variables regression with a weak instrument [Staiger and Stock (1997)].

Equation (2.7) implies that $\lambda_{k,i} e_1' \Gamma = x e_i' C_k(A) \Gamma$; consequently,

$$\sqrt{T}(xe_i'C_k(A) - \lambda_{k,i} e_1')\widehat{\Gamma}_T = (xe_i'C_k(A) - \lambda_{k,i} e_1')\sqrt{T}(\widehat{\Gamma}_T - \Gamma). \tag{3.4}$$

Thus, if $\sqrt{T}(\widehat{\Gamma}_T - \Gamma)$ can be approximated by a multivariate normal distribution then we can build a confidence set for $\lambda_{k,i}$ by collecting all the values of λ for which (3.4) is small. This idea is analogous to the principle behind the Anderson and Rubin (1949) statistic in a linear Instrumental Variables model.

Confidence Set: The confidence set for $\lambda_{k,i}$ proposed in this paper has the form:

$$\mathrm{CS}_T(1-\alpha;\lambda_{k,i}) \equiv \Big\{ \lambda \in \mathbb{R} \ : \ [\sqrt{T}(xe_i'C_k(\widehat{A}_T) - \lambda e_1')\widehat{\Gamma}_T]^2 / \widehat{\sigma}_T^2(\lambda) \leq z_{1-\alpha/2}^2 \Big\} \tag{3.5}$$

where $z_{1-\alpha/2}$ denotes the upper $1-\alpha$ quantile of a standard normal random variable, and $\widehat{\sigma}_T^2(\lambda)$ is given by the formula:

$$\widehat{\sigma}_T^2(\lambda) \equiv \widehat{d}_T(\lambda)' \widehat{S}_T' \widehat{\Omega}_T \widehat{S}_T \widehat{d}_T(\lambda) \tag{3.6}$$

with

$$\widehat{d}_T(\lambda) \equiv [(\widehat{\Gamma}_T' \otimes e_i') x \partial \mathrm{vec}(C_k(\widehat{A}_T))/\partial \mathrm{vec}(A)', (xe_i'C_k(\widehat{A}_T) - \lambda e_1')]', \tag{3.7}$$

and

$$\widehat{S}_T \equiv \begin{bmatrix} \Big(\Big(\frac{1}{T} \sum_{t=1}^{T} X_t X_t' \Big)^{-1} \otimes \mathbb{I}_n \Big) & \mathbf{0}_{n^2 p \times n} \\ -\Big(\frac{1}{T} \sum_{t=1}^{T} X_t' z_t \otimes \mathbb{I}_n \Big) & \mathbb{I}_n \end{bmatrix}'. \tag{3.8}$$

The matrix $\widehat{\Omega}_T$ is an estimator for the long-run variance of the process:

$$m_t(P) \equiv \begin{pmatrix} X_t \\ z_t \end{pmatrix} \otimes \eta_t - \mathbb{E}_P\Big[\begin{pmatrix} X_t \\ z_t \end{pmatrix} \otimes \eta_t \Big], \quad P \in \mathcal{P},$$

In the next subsection we provide conditions on $\mathcal{P}$ to guarantee that our confidence set is uniformly consistent in level; that is:

$$\liminf_{T\to\infty} \inf_{P \in \mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha;\lambda_{k,i}) \Big) = 1-\alpha.$$

3.2. *High-level Assumptions*

In this section, we present the two-main high-level assumptions for the class $\mathcal{P}$. These assumptions are sufficient to establish the uniform consistency in level of our confidence set. The class $\mathcal{P}$ under consideration allows for i) arbitrary small values of the correlation between the external instrument and the target structural shock; and ii) arbitrary small values of the contemporaneous response of $Y_{1,t}$ to an impulse in $\varepsilon_{1,t}$.

Consider a kernel estimator for the long-run variance of m_t given by:

$$\widehat{\Omega}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\widehat{m}_t\widehat{m}_t' + \sum_{j=1}^{B_T} k(j,B_T)\Big(\frac{1}{T}\sum_{t=1}^{T-j}\widehat{m}_t\widehat{m}_{t-j}' + \frac{1}{T}\sum_{t=1}^{T-j}\widehat{m}_{t-j}\widehat{m}_t'\Big). \tag{3.9}$$

where

$$\widehat{m}_t \equiv \begin{pmatrix} X_t \\ z_t \end{pmatrix} \otimes \widehat{\eta}_t - \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} X_t \\ z_t \end{pmatrix} \otimes \widehat{\eta}_t.$$

We say that a sequence of random vectors $c_t(P)$ indexed by $P \in \mathcal{P}$ is $o_{\mathcal{P}}(1)$ if

$$\lim_{T\to\infty} \sup_{P\in\mathcal{P}} P(||c_t(P)|| > \epsilon) \to 0, \quad \forall\, \epsilon > 0.$$

We say that a sequence of random vectors $c_t(P)$ is bounded from below in $\mathcal{P}$ if there exists $\delta > 0$ such that:

$$\lim_{T\to\infty} \inf_{P\in\mathcal{P}} P(||c_t(P)|| > \delta) \to 1.$$

Our first high-level assumption to conduct inference on $\lambda_{k,i}$ is as follows:

Assumption 2 (Asymptotic Behavior of $m_t(P)$ over the class $\mathcal{P}$) The class $\mathcal{P}$, the kernel $k(j, B_T)$, and the bandwidth B_T are such that for any $q \in \mathbb{R}$ and any random vector $c_T(P)$ bounded from below in $\mathcal{P}$:

$$\limsup_{T\to\infty} \sup_{P\in\mathcal{P}} P\Big[\Big(c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)\Big)^2 / c_T(P)'\widehat{\Omega}_T c_T(P) > q^2\Big]$$

equals

$$\mathbb{P}(N(0,1)^2 > q^2).$$

Comment 1: Assumption 2 can be established by imposing the following restrictions: i) a Central Limit Theorem for $m_t(P)$ uniformly over the class $\mathcal{P}$; ii) the long-run variance of $m_t(P)$ can be consistently estimated uniformly over $\mathcal{P}$ with a kernel estimator, and iii) the smallest eigenvalue of the long-run variance matrix (which depends on P) is bounded away from zero uniformly over $\mathcal{P}$. Our choice to state a high-level assumptions instead of primitive conditions aims to simplify the description of the theory behind our inference procedure.

The next high-level assumption will allows us to approximate the asymptotic distribution of $C_k(\widehat{A}_T)$ (properly centered and scaled), in terms of the asymptotic distribution of $\sqrt{T}(\widehat{A}_T - A(P))$.

Assumption 3 (Delta method for $C_k(\widehat{A}_T)$ uniformly over $\mathcal{P}$) The class $\mathcal{P}$ is such that:

$$\sqrt{T}\text{vec}(C_k(\widehat{A}_T) - C_k(A(P))) = G_k(A(P))\sqrt{T}\text{vec}(\widehat{A}_T - A(P)) + o_{\mathcal{P}}(1),$$

where $G_k(A) \equiv \partial\text{vec}(C_k(A))/\partial\text{vec(A)'}$ is the $n^2 \times n^2 p$ matrix defined in Lütkepohl (1990), p. 118 as:

$$G_k(A) = \sum_{m=0}^{k-1} J(\mathcal{A}')^{k-1-m} \otimes C_m(A),$$

with $J \equiv [\mathbb{I}_n, \mathbf{0}_{n\times(n-1)p}]$ and

$$\mathcal{A} \equiv \begin{pmatrix} A_1 & \dots & A_{p-1} & A_p \\ & \mathbb{I}_{n(p-1)} & & \mathbf{0}_{n(p-1)\times n} \end{pmatrix}.$$

Comment 2: Assumption 3 requires a uniform delta-method in probability to hold for $C_k(\widehat{A}_T)$. It is well known that the Gaussian approximation for the non-linear transformation $C_k(\widehat{A}_T)$ can be problematic in certain parts of the parameter space [Benkwitz, Neumann, and Lütekpohl (2000)]. In a simple VAR(1), this can happen whenever the autoregressive matrix A_1 is close to zero. Assumption 3 assumes these problems away.[4]

Main Result: Let

[4]An alternative approach would replace the estimator $C_k(\widehat{A}_T)$ for the reduced-form MA coefficients by the a uniformly normal estimator for $C_k(A)$. One possibility could be the local projection estimators proposed by Jordà (2005).

$$\widehat{d}_T(\lambda) \equiv [(\widehat{\Gamma}'_T \otimes e'_i)x\partial\text{vec}(C_k(\widehat{A}_T))/\partial\text{vec}(A)', (xe'_iC_k(\widehat{A}_T) - \lambda e'_1)]',$$

and

$$d_T(\lambda) \equiv [(\widehat{\Gamma}'_T \otimes e'_i)x\partial\text{vec}(C_k(A_T))/\partial\text{vec}(A)', (xe'_iC_k(A_T) - \lambda e'_1)]'.$$

Define $\sigma^2_T(\lambda) \equiv d_T(\lambda)'\widehat{S}'_T\widehat{\Omega}_T\widehat{S}_Td_T(\lambda)$ and let $\widehat{\sigma}^2_T(\lambda) \equiv \widehat{d}_T(\lambda)'\widehat{S}'_T\widehat{\Omega}_T\widehat{S}_T\widehat{d}_T(\lambda)$.

Proposition 1 (Uniform Consistency in Level) *Suppose that*

$$\lim_{T\to\infty} \sup_{P\in\mathcal{P}} \sup_{\lambda\in\mathbb{R}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} - 1\Big| > \epsilon\Big) \to 0$$

and suppose there is $\delta > 0$ such that

$$\lim_{T\to\infty} \inf_{P\in\mathcal{P}} \inf_{\lambda\in\mathbb{R}} P(||\widehat{S}_Td_T(\lambda)|| > \delta) = 1.$$

for any $\epsilon > 0$. If Assumptions 1 to 3 hold then:

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in CS_T(1-\alpha; \lambda_{k,i})\Big) = 1 - \alpha.$$

Proof: See Appendix A.1.

4. IMPLEMENTATION

Our confidence set, with confidence level $1-\alpha$, is given by the set of all $\lambda \in \mathbb{R}$ such that:

$$\frac{\left[\sqrt{T} x e_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T - \sqrt{T}\lambda e_1'\widehat{\Gamma}_T\right]^2}{\widehat{d}_T(\lambda)'\widehat{W}_T\widehat{d}_T(\lambda)} \leq z^2_{1-\alpha/2},$$

where $\widehat{W}_T \equiv \widehat{S}_T'\widehat{\Omega}_T\widehat{S}_T$. In this section, we show that our confidence set can be expressed in terms of the roots of the quadratic equation:

$$\lambda^2 \widehat{a}_{1-\alpha} + \lambda \widehat{b}_{1-\alpha} + \widehat{c}_{1-\alpha} = 0. \tag{4.1}$$

Additional Notation: Partition the matrix $\widehat{W}_T$ as

$$\widehat{W}_T = \begin{pmatrix} \widehat{W}_{1,1} & \widehat{W}_{1,2} \\ \widehat{W}_{1,2}' & \widehat{W}_{2,2} \end{pmatrix},$$

where $\widehat{W}_{1,1}$ is of dimension $n^2p \times n^2p$. Note that the quadratic form $\widehat{d}_T(\lambda)'\widehat{W}_T\widehat{d}_T(\lambda)$ can be written as the sum of the following three terms:

$$x^2(\widehat{\Gamma}_T' \otimes e_i')G_k(\widehat{A}_T)\widehat{W}_{1,1}G_k(\widehat{A}_T)'(\widehat{\Gamma}_T \otimes e_i), \quad \text{(does not depend on } \lambda) \tag{4.2}$$

$$2x(\widehat{\Gamma}_T' \otimes e_i')G_k(\widehat{A}_T)\widehat{W}_{1,2}[C_k(\widehat{A}_T)'e_i x - \lambda e_1], \quad \text{(linear in } \lambda) \tag{4.3}$$

$$[C_k(\widehat{A}_T)'e_i x - \lambda e_1]'\widehat{W}_{2,2}[C_k(\widehat{A}_T)e_i x - \lambda e_1], \quad \text{(quadratic in } \lambda). \tag{4.4}$$

Some simple algebra shows that $\lambda \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})$ if and only if:

$$\lambda^2 \widehat{a}_{1-\alpha} + \lambda \widehat{b}_{1-\alpha} + \widehat{c}_{1-\alpha} \leq 0, \tag{4.5}$$

where:

$$\begin{aligned} \widehat{a}_{1-\alpha} &\equiv T(e_1'\widehat{\Gamma}_T)^2 - z^2_{1-\alpha/2} e_1'\widehat{W}_{2,2}e_1, \\ \widehat{b}_{1-\alpha} &\equiv -2Tx(e_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T)(e_1'\widehat{\Gamma}_T) + 2z^2_{1-\alpha/2}x(\widehat{\Gamma}_T' \otimes e_i')G_k(\widehat{A}_T)\widehat{W}_{1,2}e_1 \end{aligned}$$

$$\begin{aligned}
\widehat{c}_{1-\alpha} &\equiv (\sqrt{T} x e_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T)^2 \\
&- z^2_{1-\alpha/2} x^2 (\widehat{\Gamma}_T' \otimes e_i') G_k(\widehat{A}_T) \widehat{W}_{1,1} G_k(\widehat{A}_T)' (\widehat{\Gamma}_T \otimes e_i), \\
&- 2 z^2_{1-\alpha/2} x^2 (\widehat{\Gamma}_T' \otimes e_i') G_k(\widehat{A}_T) \widehat{W}_{1,2} (C_k(\widehat{A}_T)' e_i) \\
&- 2 z^2_{1-\alpha/2} x^2 e_i' C_k(\widehat{A}_T) \widehat{W}_{2,2} C_k(\widehat{A}_T) e_i.
\end{aligned}$$

Implementation: Our confidence set can be written as the following quadratic inequality:

$$\text{CS}_T(1-\alpha;\lambda_{k,i}) = \Big\{ \lambda \in \mathbb{R} \, : \, \lambda^2 \widehat{a}_{1-\alpha} + \lambda \widehat{b}_{1-\alpha} + \widehat{c}_{1-\alpha} \leq 0 \Big\}.$$

An obvious observation is that whenever $\widehat{a}_{1-\alpha} \neq 0$, we can represent the quadratic inequality above as:

$$\Big\{ \lambda \in \mathbb{R} \, : \, \widehat{a}_{1-\alpha} \Big(\lambda + \frac{1}{2} \frac{\widehat{b}_{1-\alpha}}{\widehat{a}_{1-\alpha}} \Big) - \Big(\frac{\Delta_{1-\alpha}}{4 \widehat{a}_{1-\alpha}} \Big) \leq 0 \Big\}, \qquad \Delta_{1-\alpha} \equiv \widehat{b}^2_{1-\alpha} - 4 \widehat{a}_{1-\alpha} \widehat{c}_{1-\alpha},$$

which can be thought of as the set of values in which a parabola takes negative values. Therefore, whenever $\widehat{a}_{1-\alpha} \neq 0$, our confidence set can take one of the following forms:

1. If $\widehat{a}_{1-\alpha} > 0, \Delta_{1-\alpha} \geq 0$:

$$\text{CS}_T(1-\alpha, \lambda_{k,i}) = \Big[\frac{-\widehat{b}_{1-\alpha} - \sqrt{\Delta_{1-\alpha}}}{2 \widehat{a}_{1-\alpha}} \, , \, \frac{-\widehat{b}_{1-\alpha} + \sqrt{\Delta_{1-\alpha}}}{2 \widehat{a}_{1-\alpha}} \Big].$$

2. If $\widehat{a}_{1-\alpha} < 0, \Delta_{1-\alpha} \geq 0$:

$$\text{CS}_T(1-\alpha, \lambda_{k,i}) = \Big(-\infty, \frac{-\widehat{b}_{1-\alpha} + \sqrt{\Delta_{1-\alpha}}}{2 \widehat{a}_{1-\alpha}} \Big] \cup \Big[\frac{-\widehat{b}_{1-\alpha} - \sqrt{\Delta_{1-\alpha}}}{2 \widehat{a}_{1-\alpha}}, \infty \Big).$$

3. If $\widehat{a}_{1-\alpha} > 0, \Delta_{1-\alpha} \leq 0$:[5]

$$\text{CS}_T(1-\alpha, \lambda_{k,i}) = \phi,$$

[5] This case, however, cannot happen as the plug-in estimator:

$$\widehat{\lambda}_{k.i} = x e_i' C_k(\widehat{A}_T) \widehat{\Gamma}_T / e_1' \widehat{\Gamma}_T$$

is always an element of our confidence set.

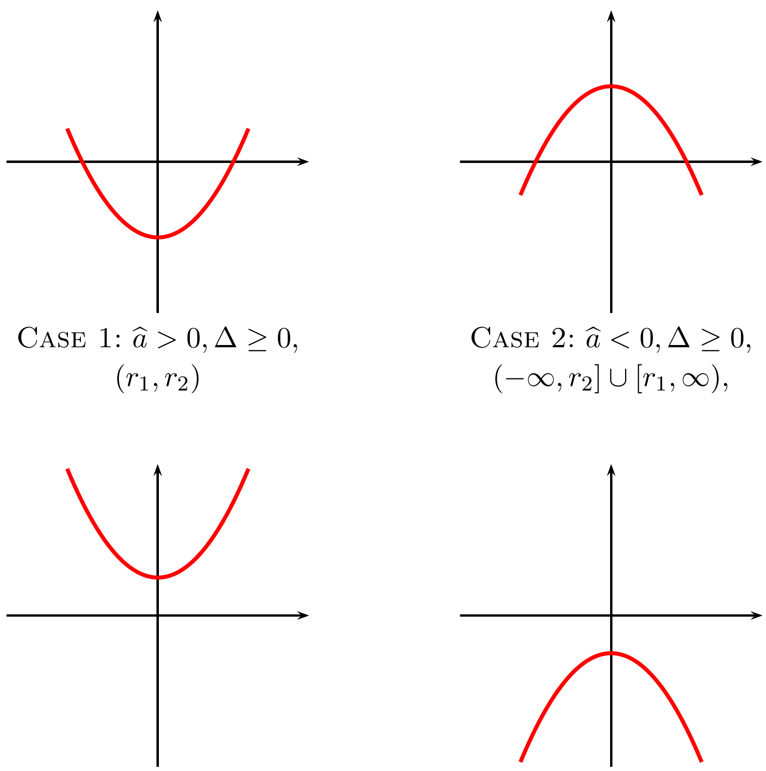
4. If $\widehat{a}_{1-\alpha} < 0, \Delta_{1-\alpha} \leq 0$:

$$\mathrm{CS}_T(1-\alpha, \lambda_{k,i}) = (-\infty, \infty).$$

Graphically, these four cases can be represented as follows:

Figure 1: $\mathrm{CS}_T(1-\alpha, \lambda) = \{\lambda \in \mathbb{R} \,:\, \lambda^2 \widehat{a} + \lambda \widehat{b} + \widehat{c} < 0\}$

$(r_1 \equiv (-\widehat{b} - \sqrt{\Delta})/2\widehat{a},\; r_2 \equiv (-\widehat{b} + \sqrt{\Delta})/2\widehat{a})$



Case 1: $\widehat{a} > 0, \Delta \geq 0$, (r_1, r_2)

Case 2: $\widehat{a} < 0, \Delta \geq 0$, $(-\infty, r_2] \cup [r_1, \infty)$,

Case 3: $\widehat{a} > 0, \Delta \leq 0$, $\emptyset$ (Cannot happen)

Case 4: $\widehat{a} < 0, \Delta \leq 0$, $(-\infty, \infty)$

Description: Figure 1 presents a graphical representation of the set of $\lambda \in \mathbb{R}$ for which the quadratic equation $\lambda^2 \widehat{a} + \lambda \widehat{b} + \widehat{c}$ can take negative values.

Finally, if $\widehat{a}_{1-\alpha} = 0$ (which happens with probability zero) the confidence set is of the form:

$$\mathrm{CS}_T(1-\alpha, \lambda_{k,i}) \equiv (-\infty, \widehat{c}_{1-\alpha}/\widehat{b}_{1-\alpha}],$$

if $\widehat{b}_{1-\alpha} > 0$ or

$$\mathrm{CS}_T(1-\alpha, \lambda_{k,i}) \equiv [\widehat{c}_{1-\alpha}/\widehat{b}_{1-\alpha}, +\infty),$$

if $\widehat{b}_{1-\alpha} < 0$.

5. ILLUSTRATIVE EXAMPLES

In this section we conduct inference about the coefficients of the structural IRF in two popular SVAR models. First we study the dynamic effects of a structural oil shock using a 4-variable SVAR estimated with U.S. monthly data (January 1973 to September 2004). We consider two different external instruments for this model: Hamilton (2003)'s maximum deviation series and Kilian (2008)'s shortfall in OPEC's oil production.

Second, we study the dynamic effects of a structural monetary shock using a 4-variable Monetary SVAR also estimated with U.S. monthly data (January 1970 to December 1996). We consider two different external instruments: Romer and Romer (2004)'s measure of monetary shocks and also the time series of montetary shocks estimated by Sims and Zha (2006).

5.1. *Oil SVAR*

The Oil SVAR includes the following four variables:

1. First-Difference of the natural logarithm of the Crude Petroleum item of the Producer Price Index (PPI). This time series is obtained directly from the Bureau of Labor Statistics (BLS, series WPU0561).

2. First Difference of the natural logarithm of the Consumer Price Index (CPI). The CPI is obtained from the Federal Reserve Economic Data (FRED) website of the Federal Reserve Bank of St. Louis.

3. First Difference of the natural logarithm of the Gross Domestic Product (interpolated monthly GDP series).

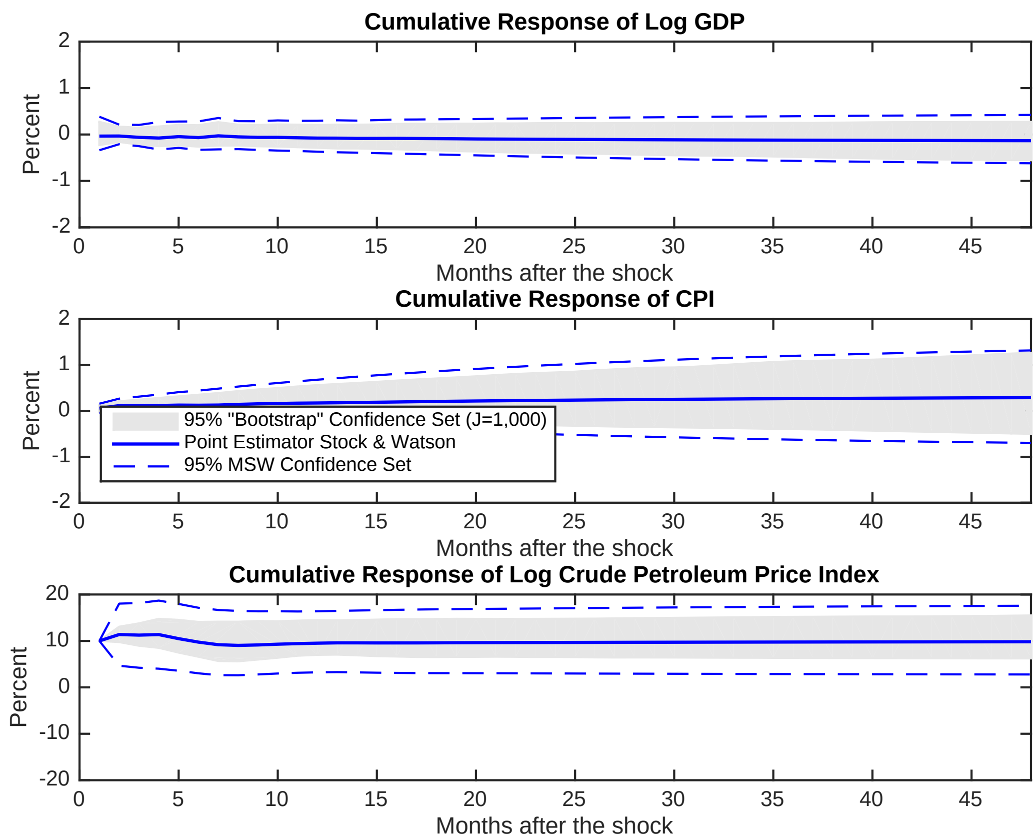
4. First Difference of the Federal Funds Rate. This time series is obtained from the FRED website of the Federal Reserve Bank of St. Louis.

Our benchmark specification uses 6 lags. The Bayes Information Criterion selects 2 lags, the Hannan-Quinn Information Criterion selects 3 lags, and the Akaike Information Criterion selects 9 lags.

The target shock is an oil shock. The variable used for normalization is the price of oil. The target shock is assumed to increase the first difference of the natural logarithm of the price of oil in 10 percent.

Figure 2 reports the cumulative effects of GDP, CPI and the price of oil to a structural oil shock identified using Hamilton (2003)'s measure of oil shocks. This external instrument is constructed by taking the maximum deviation of the crude petroleum PPI relative to its previous twelve-month maximum. If the difference is negative, the instrument takes the value of zero.

Figure 2: Dynamic effects of a Structural Oil Shock (Hamilton's Instrument)



The solid blue line in the pictures above is the plug-in estimator $\widehat{\lambda}_{k,i}$ in equation (3.3). As we mentioned before, the plug-in estimator depends on the data only through $\widehat{A}_T$ and $\widehat{\Gamma}_T$.

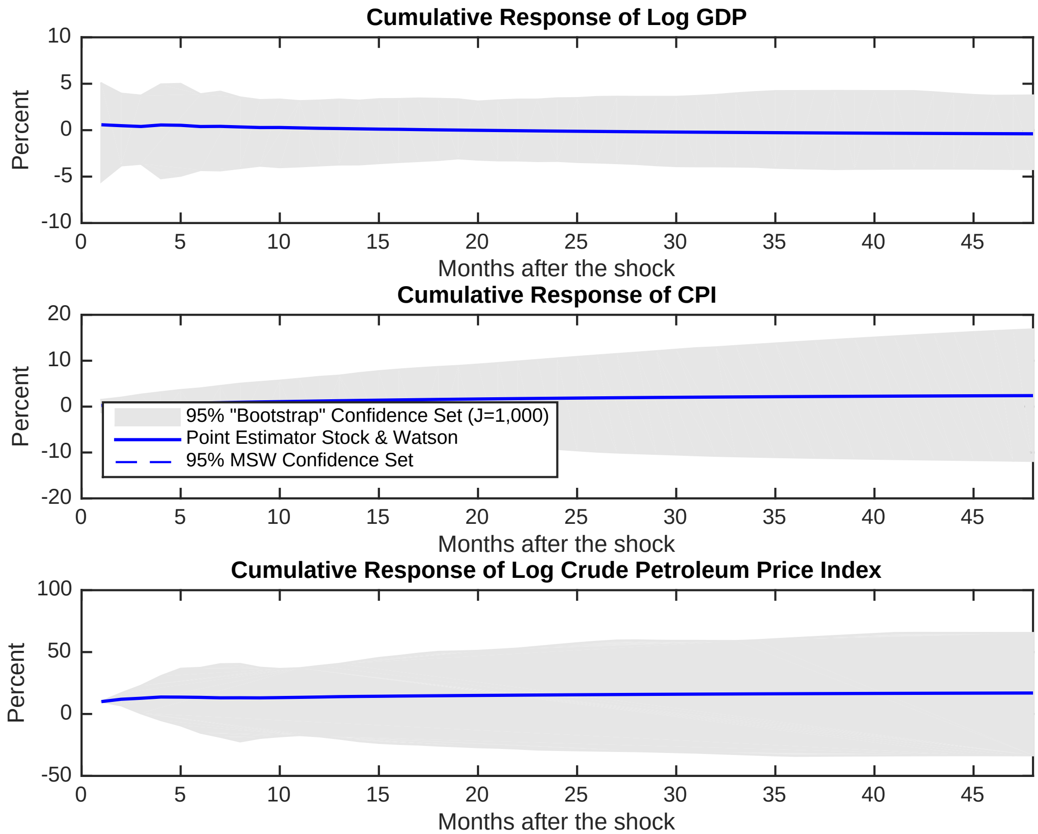
The gray area is a "bootstrap"-type 95% confidence set for $\widehat{\lambda}_{k,i}$ based on samples of the asymptotic distribution of $\widehat{A}_T$ and $\widehat{\Gamma}_T$ (assumed Gaussian with the covariance matrix given by $\widehat{W}_T$ in Section 4 computed using a Newey-West kernel with a bandwidth of 0). We generate $J = 1,000$ samples from this distribution and compute the plug-in estimator for each of the samples. The gray area represents the region between the .025 and .975 quantiles.

The blue dashed lines correspond to the 95% confidence region proposed in Section 3 of this paper. To compute the end points of our confidence set, we solve for the roots of the single-variable quadratic equation with the coefficients defined in Section 4.

In this application, the "bootstrap"-type 95% confidence set and our confidence set are remarkably similar. In the appendix we show that this feature follows from the fact that the covariance between the external instrument and the price of oil—which is the variable defining the scale of the IRFs—is far away from zero (the Wald statistic for the covariance between the external instrument and the price of oil is 15.9061).

Figure 3 reports the cumulative effects of an oil shock over GDP, CPI, and the price of oil using Kilian (2008)'s measure of oil shocks. These shocks are estimated as the expected shortfall in OPEC's production. For every variable and every horizon, our confidence set is $(-\infty, \infty)$, which is significantly different to the 95% "bootstrap"-type confidence set. We note that Wald Statistic for the covariance between the external instrument and the price of oil is for this model is 0.1881.

Figure 3: Dynamic effects of a Structural Oil Shock (Kilian's Instrument)



5.2. *Monetary SVAR*

The Monetary SVAR includes the following four variables:

1. First-Difference of the natural logarithm of the Industrial Production Index (IP). This time series is obtained from the Board of Governors website (series IP.B50001.SIP.B50001.S).

2. First Difference of the natural logarithm of the Producer Price Index (PPI). The PPI is obtained from the Bureau of Labor Statistics (BLS, series WPUS-SOP3000).

3. First Difference of the Federal Funds Rate. This time series is obtained from

the FRED website of the Federal Reserve Bank of St. Louis.

4. First Difference of the natural logarithm of the commodity price index used in the monetary SVAR in Stock and Watson (2012).

Our benchmark specification uses 6 lags. The Bayes Information Criterion selects 1 lags, the Hannan-Quinn Information Criterion selects 2 lags, and the Akaike Information Criterion selects 3 lags.

The target shock is a monetary shock. The variable used for normalization is the federal funds rate. The target shock is assumed to increase the federal funds rate in 1 percent.

Figure 4: Dynamic effects of a Structural Monetary Shock (Romers' Instrument)

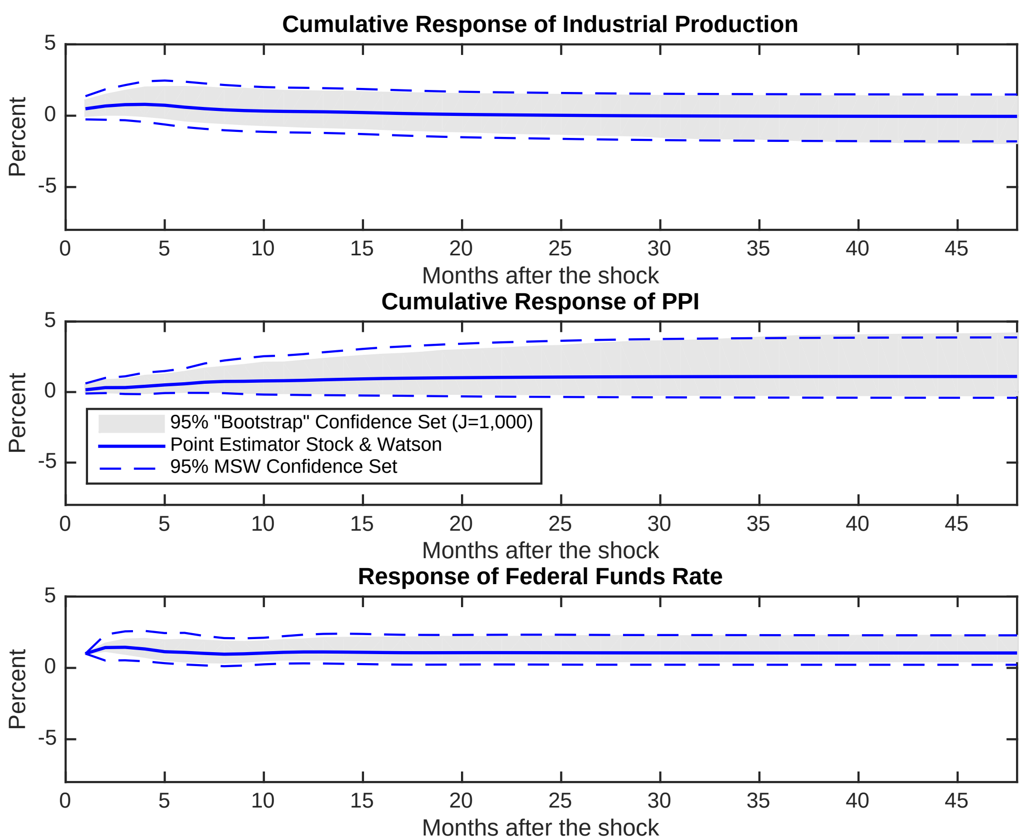


Figure 4 reports the cumulative response of IP, PPI and the federal funds rate to a structural monetary shock identified using Romer and Romer (2004)'s instrument.

Their instrument is constructed in two steps. First, they construct an 'intended federal funds rate' based on a careful reading of the Record of Policy Actions of the Open Market Committee (FOMC). Second, they use the FOMC's *greenbook* forecasts to 'clean the intended funds' rate from anticipatory movements. Broadly speaking, the instrument is given by the residual of a regression of the intended federal funds rate on the greenback forecasts of output and inflation.

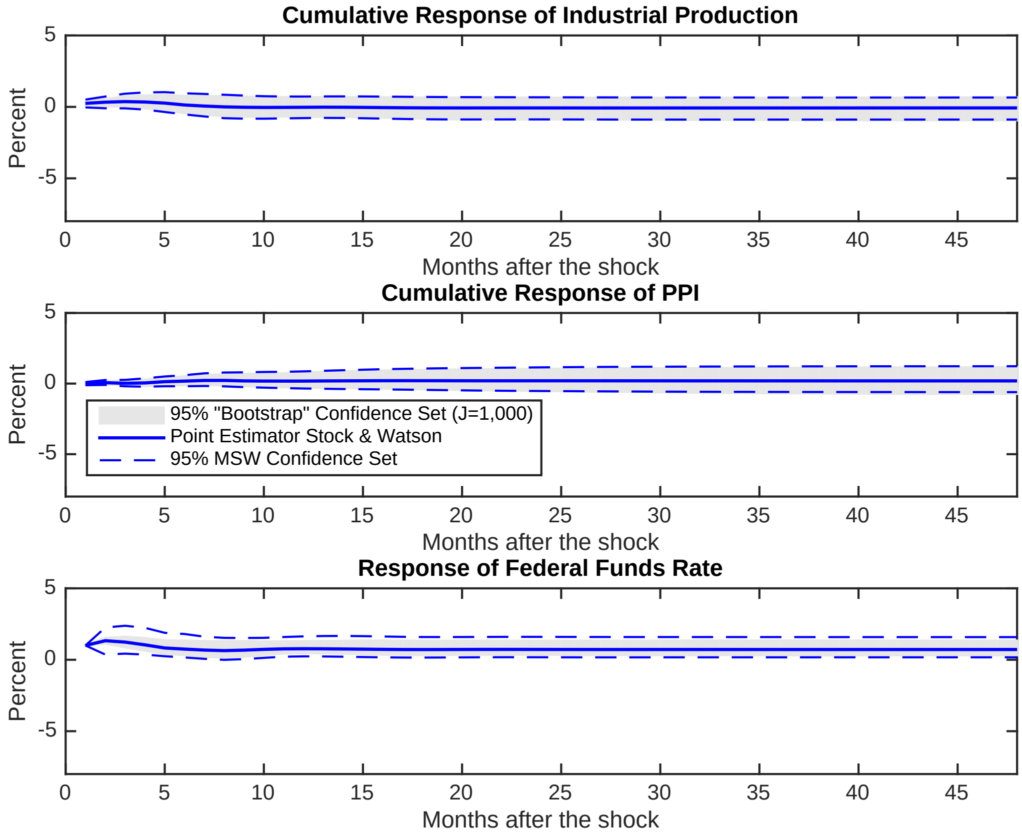
Once again, we note that the "bootstrap"-type 95% confidence set and our confidence set lie close to each other. The Wald statistic for the covariance between the external instrument and the price of oil is 12.8721). We note that despite the inclusion of a commodity price index to account for the so-called price-puzzle (Bernanke and Blinder (1992), Christiano, Eichenbaum, and Evans (1996)), the confidence sets reported in Figure 4 do not rule out the possibility that prices react positively to a contractionary monetary shock (at any horizon). This feature is robust to several specifications (different lags and also different measures of economic activity/prices).

Finally, we report the cumulative response of IP, PPI and the federal funds rate to a structural monetary shock identified using Sims and Zha (2006)'s measure of monetary shocks.

We note that although there some quantitive differences, the IRFs estimated using the Romers' instrument and Sims and Zha's monetary shock are qualitatively similar. We note that both approaches lead to estimated IRFs that exhibit the price puzzle. Finally, the Wald statistic for the covariance between the external instrument and the price of oil is 11.8103.

Figure in Appendix C reports the estimated monetary shocks using each of the external instruments.

Figure 5: Dynamic effects of a Structural Monetary Shock (Sims and Zha's Instrument)



6. CONCLUSION

This paper studied SVARs identified using external instruments. These external instruments were taken to be correlated with the target shock (e.g., the oil instrument is *relevant*) and to be uncorrelated with other macroeconomic shocks of the model (e.g., the oil instrument is *exogenous*).

We proposed a confidence set for the coefficients of the Structural Impulse-Response function and we established its *uniform consistency in level* over a large class of data generating processes. The implementation of our confidence set requires no more work than solving a single-variable quadratic equation. The coefficients of such equation depend on the reduced-form VAR parameters and the covariance between reduced-form residuals and external instruments. Our paper provides formulas for these coefficients.

By reporting our confidence set, practitioners do not need to worry about situations in which the correlation between the external instrument and each of the estimated reduced-form residuals is close to zero: there is no need of formal or informal pre-tests for *weak instruments*. Thus, we recommend practitioners to report the plug-in estimator in Stock and Watson (2012) along with our confidence bands.

In an empirical application studying the dynamic effects of a structural oil shock using U.S. monthly data, we compared standard confidence sets with our inference procedure. As the theory suggests, we found substantial differences in cases where the external instrument is weakly correlated with the reduced-form error in the oil price equation.

We also analyzed a monetary SVAR identified with two popular measures of monetary shocks: Romer and Romer (2004) and Sims and Zha (2006). The estimated IRFs (and the estimated structural monetary shock) exhibit similar features regardless of the instrument used to identify the model. Interestingly, none of the external instruments rule out the price puzzle.

Our paper focused on the case in which the dynamic responses to *one* structural shock are identified with only *one* external instrument. We leave the 'over-identified' case (more than one instrument for one shock) for future research.

REFERENCES

ANDERSON, T. AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63.

BENKWITZ, A., M. H. NEUMANN, AND H. LÜTEKPOHL (2000): "Problems related to confidence intervals for impulse responses of autoregressive processes," *Econometric Reviews*, 19, 69–103.

BERNANKE, B. S. AND A. S. BLINDER (1992): "The federal funds rate and the channels of monetary transmission," *The American Economic Review*, 901–921.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. EVANS (1996): "The effects of monetary policy shocks: some evidence from the flow of funds," *The Review of Economics and Statistics*, 78, 16–34.

FIELLER, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185.

HAMILTON, J. D. (2003): "What is an oil shock?" *Journal of econometrics*, 113, 363–398.

JORDÀ, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American economic review*, 161–182.

KILIAN, L. (2008): "Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?" *The Review of Economics and Statistics*, 90, 216–240.

LUNSFORD, K. G. (2015): "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy," *Working paper, Federal Reserve Bank of Cleveland.*

LÜTKEPOHL, H. (1990): "Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models," *The Review of Economics and Statistics*, 116–125.

MERTENS, K. AND M. O. RAVN (2013): "The dynamic effects of personal and corporate income tax changes in the United States," *American Economic Review*, Forthcoming.

RAMEY, V. (2011): "Identifying Government Spending Shocks: Its All in the Timing," *Quarterly Journal of Economics.*

ROMER, C. AND D. ROMER (1989): "Does monetary policy matter? A new test in the spirit of Friedman and Schwartz," in *NBER Macroeconomics Annual 1989, Volume 4*, MIT Press, 121–184.

ROMER, C. D. AND D. H. ROMER (2004): "A new measure of monetary shocks: Derivation and implications," *American Economic Review*, 94.

SIMS, C. A. AND T. ZHA (2006): "Were there regime switches in US monetary policy?" *The Amer-*

ican Economic Review, 54–81.

STAIGER, D. AND J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. AND M. W. WATSON (2008): "NBER Summer Institute Minicourse 2008: Whats New in Econometrics–Time Series, Lecture 7: Structural VARs," *Cambridge, Mass.: National Institute for Economic Research. www. nber. org/minicourse_2008. html*.

——— (2012): "Disentangling the Channels of the 2007-2009 Recession," *Brookings Papers on Economic Activity*.

VAN DER VAART, A. (2000): *Asymptotic statistics*, 3, Cambridge Univ Pr.

Appendix

APPENDIX A: PROOF OF MAIN RESULTS

A.1. *Proof of Proposition 1*

In order to prove Proposition 1 we establish to simple lemmas.

Lemma 1 (Representation of $\widehat{A}_T$ and Γ_T in terms of $m_t(P)$) *If Y_t admits the structural MA representation in (2.3) and the structural innovations ε_t are serially uncorrelated, then:*

$$\begin{pmatrix} \sqrt{T}vec(\widehat{A}_T - A(P)) \\ \\ \sqrt{T}\Big(\widehat{\Gamma}_T - \mathbb{E}_P[z_t\eta_t]\Big) \end{pmatrix} = \widehat{S}'_T \frac{1}{\sqrt{T}} \sum_{t=1}^T m_t(P), \quad \widehat{S}'_T \equiv \begin{bmatrix} \Big(\Big(\frac{1}{T}\sum_{t=1}^T X_t X_t'\Big)^{-1} \otimes \mathbb{I}_n\Big) & \mathbf{0}_{n^2p\times n} \\ \Big(\frac{1}{T}\sum_{t=1}^T X_t' z_t \otimes \mathbb{I}_n\Big) & \mathbb{I}_n \end{bmatrix}.$$

Proof: Note first that

$$\begin{aligned} \sqrt{T}\text{vec}(\widehat{A}_T - A(P)) &= \Big[\Big(\frac{1}{T}\sum_{t=1}^T X_t X_t'\Big)^{-1} \otimes \mathbb{I}_n\Big] \frac{1}{\sqrt{T}} \sum_{t=1}^T X_t \otimes \eta_t \\ &= \Big[\Big(\frac{1}{T}\sum_{t=1}^T X_t X_t'\Big)^{-1} \otimes \mathbb{I}_n\Big] \frac{1}{\sqrt{T}} \sum_{t=1}^T X_t \otimes \eta_t - \mathbb{E}_P[X_t \otimes \eta_t], \end{aligned}$$

where the last line follows from the fact that Y_t admits the structural MA representation in (2.3) and the structural innovations ε_t are serially uncorrelated.

Note now that:

$$\begin{aligned} \sqrt{T}\Big(\widehat{\Gamma}_T - \mathbb{E}_P[z_t\eta_t]\Big) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t\widehat{\eta}_t - \mathbb{E}_P[z_t\eta_t], \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t(Y_t - \widehat{A}_T X_t) - \mathbb{E}_P[z_t\eta_t], \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t(\eta_t - (\widehat{A}_T - A(P))X_t) - \mathbb{E}_P[z_t\eta_t]. \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t\eta_t - \mathbb{E}_P[z_t\eta_t] - \sqrt{T}(\widehat{A}_T - A(P))\frac{1}{T}\sum_{t=1}^T z_t X_t, \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t\eta_t - \mathbb{E}_P[z_t\eta_t] - \Big(\frac{1}{T}\sum_{t=1}^T X_t' z_t \otimes \mathbb{I}_n\Big)\sqrt{T}\text{vec}(\widehat{A}_T - A(P)). \end{aligned}$$

Consequently,

$$\begin{pmatrix} \sqrt{T}\text{vec}(\widehat{A}_T - A(P)) \\ \\ \sqrt{T}\Big(\widehat{\Gamma}_T - \mathbb{E}_P[z_t\eta_t]\Big) \end{pmatrix} = \widehat{S}'_T \frac{1}{\sqrt{T}} \sum_{t=1}^T m_t(P), \quad \widehat{S}'_T \equiv \begin{bmatrix} \Big(\Big(\frac{1}{T}\sum_{t=1}^T X_t X_t'\Big)^{-1} \otimes \mathbb{I}_n\Big) & \mathbf{0}_{n^2p\times n} \\ -\Big(\frac{1}{T}\sum_{t=1}^T X_t' z_t \otimes \mathbb{I}_n\Big) & \mathbb{I}_n \end{bmatrix}.$$

Q.E.D.

Lemma 2 (Asymptotic Representation for $\sqrt{T}(xe_i'C_k(\widehat{A}_T)-\lambda_{k,i}(P)e_1')\widehat{\Gamma}_T$) *Suppose that the Assumptions of Lemma 1 hold. If Assumption 3 also holds and z_t is an external instrument then:*

$$\sqrt{T}(xe_i'C_k(\widehat{A}_T)-\lambda_{k,i}(P)e_1')\widehat{\Gamma}_T = d_T(\lambda_{k,i}(P))'\widehat{S}_T'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}m_t(P)+o_{\mathcal{P}}(1),$$

where

$$d_T \equiv [x(\widehat{\Gamma}_T'\otimes e_i')G_k(A(P)),(e_i'C_k(A(P))-\lambda_{k,i}(P)e_1')]'.$$

Proof: Since z_t is an external instrument and $e_1'B(P)\neq 0$, then

$$\lambda_{k,i}(P)=xe_i'C_k(A(P))\mathbb{E}_P[z_t\eta_t]/e_1'\mathbb{E}_P[z_t\eta_t].$$

This implies that:

$$\begin{aligned}
\sqrt{T}(xe_i'C_k(\widehat{A}_T)-\lambda_{k,i}(P)e_1')\widehat{\Gamma}_T &= \sqrt{T}(xe_i'[C_k(\widehat{A}_T)-C_k(A(P))])\widehat{\Gamma}_T\\
&+ \sqrt{T}(xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1')(\widehat{\Gamma}_T-\mathbb{E}_P[z_t\eta_t]),\\
&\quad \text{(where the last line follow from the fact that}\\
&\quad \sqrt{T}(xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1')\mathbb{E}_P[z_t\eta_t]=0)\\
&= x(\widehat{\Gamma}_T'\otimes e_i')\sqrt{T}\text{vec}(C_k(\widehat{A}_T)-C_k(A(P)))\\
&+ \sqrt{T}(xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1')(\widehat{\Gamma}_T-\mathbb{E}_P[z_t\eta_t]),\\
&= x(\widehat{\Gamma}_T'\otimes e_i')G_k(A(P))\sqrt{T}(\text{vec}(\widehat{A}_T)-\text{vec}(A(P)))+o_{\mathcal{P}}(1),\\
&\quad \text{(by Assumption 3)}\\
&+ (xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1')\sqrt{T}(\widehat{\Gamma}_T-\mathbb{E}_P[z_t\eta_t]).
\end{aligned}$$

Therefore $\sqrt{T}(xe_i'C_k(\widehat{A}_T)-\lambda_{k,i}(P)e_1')\widehat{\Gamma}_T$ equals:

$$(x(\widehat{\Gamma}_T'\otimes e_i')G_k(A(P)),(xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1'))\begin{pmatrix}\sqrt{T}\text{vec}(\widehat{A}_T-A(P))\\ \\ \sqrt{T}\Big(\widehat{\Gamma}_T-\mathbb{E}_P[z_t\eta_t]\Big)\end{pmatrix}+o_{\mathcal{P}}(1), \tag{A.1}$$

and, by Lemma 1, it follows that (A.1) can be written as:

$$d_T(\lambda_{k,i}(P))'\widehat{S}_T'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}m_t(P)+o_{\mathcal{P}}(1),$$

where

$$d_T(\lambda_{k,i}(P))\equiv[x(\widehat{\Gamma}_T'\otimes e_i')G_k(A(P)),(xe_i'C_k(A(P))-\lambda_{k,i}(P)e_1')]'\in\mathbb{R}^{n^2p+n}.$$

Q.E.D.

Proof of Proposition 1: We break the proof into two steps. We will use Lemma 1 and 2.

Step 1: We show first that

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) \geq 1-\alpha.$$

Let $c_T(P) \equiv d_T(\lambda_{k,i}(P))\widehat{S}_T$, where $d_T(\lambda_{k,i}(P))$ is defined as in Lemma 1.

$$\begin{aligned} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) &= 1 - P(\lambda_{k,i}(P) \notin \mathrm{CS}_T(1-\alpha, \lambda_{k,i})), \\ &= 1 - P\Big([\sqrt{T}(xe_i'C_k(\widehat{A}_T) - \lambda_{k,i}(P)e_1')\widehat{\Gamma}_T]^2/\widehat{\sigma}^2_T(\lambda_{k,i}(P)) > z^2_{1-\alpha/2}\Big), \\ &= 1 - P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\widehat{\sigma}^2_T(\lambda_{k,i}(P)) > z^2_{1-\alpha/2}\Big), \\ &\quad \text{(by Lemma 2, which requires Assumption 3)} \\ &= 1 - P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > \frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} z^2_{1-\alpha/2}\Big). \end{aligned}$$

Note that, for any $\epsilon > 0$, the following probability:

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > \frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} z^2_{1-\alpha/2}\Big)$$

equals the sum of

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > \frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} z^2_{1-\alpha/2} \text{ and } \Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| > \epsilon\Big)$$

and

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > \frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} z^2_{1-\alpha/2} \text{ and } \Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| \leq \epsilon\Big)$$

The monotonocity of the probability measure implies that the first term is smaller than or equal to

$$P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| > \epsilon\Big),$$

and the second term is smaller than or equal to

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > (1-\epsilon) z^2_{1-\alpha/2}\Big).$$

Therefore:

$$\begin{aligned} \inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) &\geq 1 - \sup_{P\in\mathcal{P}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| > \epsilon\Big) \\ + \quad \sup_{P\in\mathcal{P}} P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) &+ o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > (1-\epsilon) z^2_{1-\alpha/2}\Big). \end{aligned}$$

Note that

$$\inf_{P\in\mathcal{P}} P(||c_T(P)|| > \delta) = \inf_{P\in\mathcal{P}} P(||d_T(\lambda)\widehat{S}_T|| > \delta) \geq \inf_{P\in\mathcal{P}} \inf_{\lambda\in\mathbb{R}} P(||d_T(\lambda)\widehat{S}_T|| > \delta).$$

Since, by assumption, the last term converges to 1 then the sequence $c_T(P)$ is bounded from below in $\mathcal{P}$.

Assumption 2 and the fact that $\sup_{P\in\mathcal{P}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| > \epsilon\Big) \leq \sup_{P\in\mathcal{P}}\sup_{\lambda\in\mathbb{R}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda}{\sigma^2_T(\lambda)} - 1\Big| > \epsilon\Big) \to 0$ imply that

$$\liminf_{T\to\infty}\inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) \geq 1-\alpha.$$

This concludes the proof of Step 1.

Step 2: We now show that

$$\liminf_{T\to\infty}\inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) \leq 1-\alpha.$$

To see this, note that for any $\epsilon > 0$:

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda) > \frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} z^2_{1-\alpha/2}\Big)$$

is larger than or equal to

$$P\Big([d_T'\widehat{S}_T'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda) > \frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} z^2_{1-\alpha/2} \text{ and } \Big|\frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} - 1\Big| \leq \epsilon\Big).$$

Once again, the monotonicity of the probability measure implies that the expression above is larger than or equal than:

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda) > (1-\varepsilon) z^2_{1-\alpha/2} \text{ and } \Big|\frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} - 1\Big| \leq \epsilon\Big).$$

since $P(A\cap B) = P(A) + P(B) - P(A\cup B) \geq P(A) + P(B) - 1$, the expression above is larger than or equal to:

$$P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda) > (1-\varepsilon) z^2_{1-\alpha/2}\Big) + P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} - 1\Big| > \epsilon\Big).$$

Therefore:

$$\inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) \leq 1 - \sup_{P\in\mathcal{P}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda_{k,i}(P))}{\sigma^2_T(\lambda_{k,i}(P))} - 1\Big| > \epsilon\Big)$$
$$+ \sup_{P\in\mathcal{P}} P\Big([c_T(P)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(P) + o_{\mathcal{P}}(1)]^2/\sigma^2_T(\lambda_{k,i}(P)) > (1+\epsilon) z^2_{1-\alpha/2}\Big).$$

We have shown that $c_{T(P)}$ is bounded from below in $\mathcal{P}$. Assumption 2 and the fact that $\sup_{P\in\mathcal{P}}\sup_{\lambda\in\mathbb{R}} P\Big(\Big|\frac{\widehat{\sigma}^2_T(\lambda)}{\sigma^2_T(\lambda)} - 1\Big| > \epsilon\Big) \to 0$ imply that

$$\liminf_{T\to\infty}\inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \mathrm{CS}_T(1-\alpha, \lambda_{k,i})\Big) \leq 1-\alpha.$$

This concludes the proof of Step 2.

CONCLUSION: Step 1 and Step 2 implies that under Assumption 1 and 2

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} P\Big(\lambda_{k,i}(P) \in \text{CS}_T(1-\alpha, \lambda_{k,i})\Big) = 1-\alpha.$$

Q.E.D.

APPENDIX B: ASYMPTOTIC DISTRIBUTION OF THE PLUG-IN ESTIMATOR

In this section, we derive the asymptotic distribution of the plug-in estimator:

$$\lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T) \equiv x e_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T / e_1'\widehat{\Gamma}_T,$$

under different *drifting* sequences of parameter values. The objective of this section is twofold.

On the one hand, we show that the plug-in estimator above is not *uniformly consistent* (when evaluated over a reasonable class of data generating processes). The result suggests that applied researchers ought to be cautious with the interpretation of point-estimators for IRFs derived with an external instrument.

On the other hand, we show that the Wald statistic based on the plug-in estimator is not *regular* (at different points in the parameter space). In particular, one problematic area—which researchers cannot rule out a priori—concerns the presence of a weak external instrument. Other problematic points, however, arise when B_1 is local-to-zero.

Set-up: Let P denote the distribution of $\{Y_t, z_t\}_{t=1}^{\infty}$. As we have mentioned before, the distribution P is indexed by the parameters (A, B, F), where F is the joint distribution of $\{z_t, \varepsilon_t\}_{t=1}^{\infty}$. Let $\{(A_T, B_T, F_T)\}_{T=1}^{\infty} \rightarrow (A, B, F)$ be a converging sequence of parameter values. Such sequence induces the sequence of DGPs: $\{P_T\}_{T=1}^{\infty}$. We note that:

$$\mathbb{E}_{P_T}[z_t \eta_T] = B_{1,T}\alpha_T,$$

where $\mathbb{E}_{F_T}[z_t \varepsilon_{1,t}] \equiv \alpha_T$. Throughout this section we work with the following high-level assumption.

Assumption HL: Let $\mathcal{P}$ be a class of DGPs such that for every converging sequence:

$$\{(A_T, B_T, F_T)\}_{T=1}^{\infty}$$

such that $\mathcal{P}_T \in \mathcal{P}$, there is a matrix $W(P)$ (where P is the measure induced by the limiting parameter values) such that:

$$W^{-1/2}(P)\begin{pmatrix} \text{vec}(\widehat{A}_T) - \text{vec}(A_T) \\ \widehat{\Gamma}_T - \alpha_T B_{1,T} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathcal{N}_{n^2p+n}\Big(\mathbf{0}, \mathcal{I}_{n^2p+n}\Big).$$

B.1. *The plug-in estimator is not uniformly consistent*

We use the high-level assumption above to show that the plug-in estimator is not uniformly consistent.

Definition (Uniform Consistency) The plug-in estimator $\lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T)$ is *uniformly consistent* over a class $\mathcal{P}$ if for every sequence of parameter values $\{(A_T, B_T, F_T)\}_{T=1}^{\infty}$ s.t. $P_T \in \mathcal{P}$ for all T:

$$\lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T) - \lambda_{k,i}(A_T, \Gamma_T) \xrightarrow{p} 0$$

Result 1 *(The plug-in estimator proposed by Stock and Watson (2012) is not uniformly consistent) Let $\mathcal{P}$ be a class of DGPs such that Assumption HL holds. If there is a sequence (A, B, F_T)*

inside this class such that

$$\alpha_T = a/\sqrt{T};$$

that is, assume that the covariance between the external instrument and the target shock is local-to-zero. Then:

$$\lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T) - \lambda_{k,i}(A, \Gamma_T) \xrightarrow{d} \Delta^* \equiv \frac{\sigma_1}{\sigma_2} \frac{\mathcal{Z}_1}{\mathcal{Z}_2 + \mu},$$

where

$$\begin{pmatrix} \mathcal{Z}_1 \\ \mathcal{Z}_2 \end{pmatrix} \sim \mathcal{N}_2\Big(\boldsymbol{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big),$$

$$\sigma_1 \equiv \Big((xe_i' C_k(A) - \lambda e_1') W_{2,2}(P) (xe_i' C_k(A) - \lambda e_1')' \Big)^{1/2}, \quad \sigma_2 \equiv \Big(e_1' W_{2,2}(P) e_1 \Big)^{1/2},$$

$$\rho \equiv (xe_i' C_k(A) - \lambda e_1') W_{2,2} e_1,$$

and

$$\mu = e_1' B_1 a / \sigma_2.$$

Comment 1: Result 1 implies that if the class of DGPs under consideration, $\mathcal{P}$, is such that: i) a uniform CLT holds for $\widehat{A}_T$, $\widehat{\Gamma}_T$ and ii) we allow for arbitrary small covariances between the external instrument and the target shock—a feature that cannot be ruled out a priori—; then the plug-in estimator in Stock and Watson (2012) will not be uniformly consistent.[6]

Proof: Define

$$\widehat{\Delta}_T \equiv \lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T) - \lambda_{k,i}(A, \Gamma_T).$$

Note that

$$\begin{aligned}
\widehat{\Delta}_T &= (xe_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T - \lambda_{k,i}(A, \Gamma_T) e_1' \widehat{\Gamma}_T) / e_1' \widehat{\Gamma}_T, \\
&= \Big[xe_i' [C_k(\widehat{A}_T) - C_k(A)] \widehat{\Gamma}_T + (xe_i' C_k(A) - \lambda_{k,i}(A, \Gamma_T) e_1') \widehat{\Gamma}_T \Big] / e_1' \widehat{\Gamma}_T, \\
&= \Big[xe_i' [C_k(\widehat{A}_T) - C_k(A)] \sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) + (xe_i' C_k(A) - \lambda_{k,i}(A, \Gamma_T) e_1') \sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \Big] / \sqrt{T} e_1' \widehat{\Gamma}_T \\
&+ \sqrt{T} \Big[xe_i' [C_k(\widehat{A}_T) - C_k(A)] \Gamma_T + xe_i' C_k(A) \Gamma_T - \lambda_{k,i}(A, \Gamma_T) e_1' \Gamma_T \Big] / \sqrt{T} e_1' \widehat{\Gamma}_T, \\
&\quad \text{(where } \Gamma_T \equiv \alpha_T B_1\text{)}, \\
&= (xe_i' C_k(A) - (e_i' C_k(A) B_1 / e_1' B_1) e_1') \sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) / \Big[e_1' \sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) + e_1' \sqrt{T} \widehat{\Gamma}_T \Big] \\
&\quad \text{(since } xe_i' C_k(A) \Gamma_T - \lambda_{k,i}(A, \Gamma_T) e_1' \Gamma_T = 0 \text{ and } C_k(\widehat{A}_T) \xrightarrow{p} C_k(A)\text{)} \\
&\xrightarrow{d} (xe_i' C_k(A) - \lambda e_1') W_{2,2}(P)^{1/2} \xi_2 / \Big[e_1' W_{2,2}(P)^{1/2} \xi_2 + e_1' B_1 a \Big], \\
&\quad \text{(by Assumption HL and defining } \lambda \equiv e_i' C_k(A) B_1 / e_1' B_1\text{)}
\end{aligned}$$

where $W_{2,2}(P)$ denotes the diagonal $n^2 \times n^2$ block of $W(P)$. Define:

$$\mathcal{Z}_1 \equiv (xe_i' C_k(A) - \lambda e_1') W_{2,2}(P)^{1/2} \xi_2 / \sigma_1, \quad \mathcal{Z}_2 \equiv e_1' W_{2,2}(P)^{1/2} \xi_2 / \sigma_2,$$

[6] We also note that that lack of uniform consistency need not arise only because the external instrument is local-to-zero. Another problematic sequence is $\alpha_T = \alpha$ and $B_1 = b/\sqrt{T}$

with

$$\sigma_1 \equiv \Big((xe_i'C_k(A) - \lambda e_1')W_{2,2}(P)(xe_i'C_k(A) - \lambda e_1')'\Big)^{1/2}, \quad \sigma_2 \equiv \Big(e_1'W_{2,2}(P)e_1\Big)^{1/2}.$$

Note that

$$\begin{pmatrix} \mathcal{Z}_1 \\ \mathcal{Z}_2 \end{pmatrix} \sim \mathcal{N}_2\Big(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big),$$

with

$$p \equiv (xe_i'C_k(A) - \lambda e_1')C_k(A)W_{2,2}e_1.$$

Thus,

$$\widehat{\Delta}_T \xrightarrow{d} \frac{\sigma_1}{\sigma_2}\frac{\mathcal{Z}_1}{\mathcal{Z}_2 + \mu}, \quad \mu = e_1'B_1 a/\sigma_2$$

Q.E.D.

B.2. *The Wald statistic based on the plug-in estimator is not regular*

Let $\widehat{W}_T$ be an estimator for the matrix W in Assumption HL. The Wald statistic for the testing problem:

$$\mathbf{H}_0 : \lambda_{k,i} = \lambda_0 \quad \text{v.s.} \quad \mathbf{H}_1 : \lambda_{k,i} \neq \lambda_0$$

is given by:

$$\text{(B.1)} \qquad \mathcal{W}_T(\lambda_0) = \Big[\sqrt{T}(xe_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T - \lambda_0 e_1'\widehat{\Gamma}_T)\Big]^2 / \widehat{d}_T(\widehat{\lambda}_{k,i})'\widehat{W}_T\widehat{d}_T(\widehat{\lambda}_{k,i})$$

where

$$\widehat{d}_T(\widehat{\lambda}_{k,i}) \equiv [(\widehat{\Gamma}_T' \otimes e_i')x\partial\text{vec}(C_k(\widehat{A}_T))/\partial\text{vec}(A)', (xe_i'C_k(\widehat{A}_T) - \widehat{\lambda}_{k,i}e_1')]'.$$

Definition (Regularity) We say that the Wald statistic in B.1 is regular over a class $\mathcal{P}$ if for any sequence (A_T, B_T, F_T):

$$\mathcal{W}_T(\lambda_{k,i}(A_T, B_T)) \xrightarrow{d} \chi^2_1$$

Result 2 (The Wald statistic based on the plug-in estimator in Stock and Watson (2012) is not regular) *Let $\mathcal{P}$ be a class of DGPs such that Assumption HL holds. If there is a sequence (A, B, F_T) inside this class such that*

$$\alpha_T = a/\sqrt{T},$$

then:

$$\mathcal{W}_T(\lambda_{k,i}(A_T, \Gamma_T)) \xrightarrow{d} \Big[\sigma_1\mathcal{Z}_1\Big]^2 / (xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')W_{2,2}(P)(xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')',$$

where Δ^* is distributed as in Result 1.

PROOF: From the proof of Result 1, it follows that the numerator of the Wald statistic satisfies:

$$\left[\sqrt{T}(xe_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T - \lambda_0 e_1'\widehat{\Gamma}_T)\right]^2 \xrightarrow{d} \left[\sigma_1 \mathcal{Z}_1\right]^2.$$

Using Result 1 and the continuous mapping theorem, the denominator converges to:

$$[\mathbf{0}_{1\times n^2 p}, (xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')]W(P)[\mathbf{0}_{1\times n^2 p}, (xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')]',$$

which can be written as:

$$(xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')W_{2,2}(P)(xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')'.$$

The Continuous Mapping Theorem thus implies that:

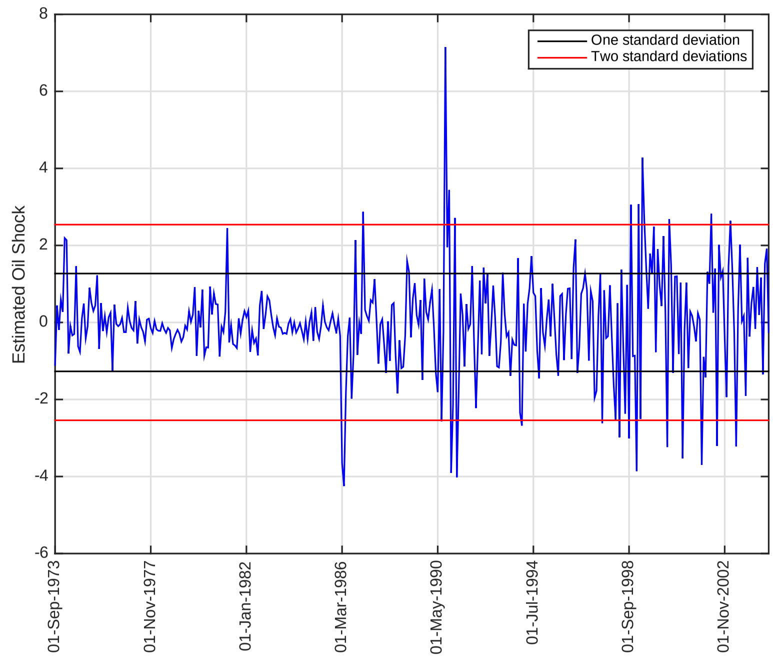
$$\mathcal{W}_T(\lambda_{k,i}(A_T, \Gamma_T)) \xrightarrow{d} \left[\sigma_1 \mathcal{Z}_1\right]^2 / (xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')W_{2,2}(P)(xe_i'C_k(A) - (\lambda_0 + \Delta^*)e_1')',$$

where Δ^* is distributed as in Result 1. *Q.E.D.*

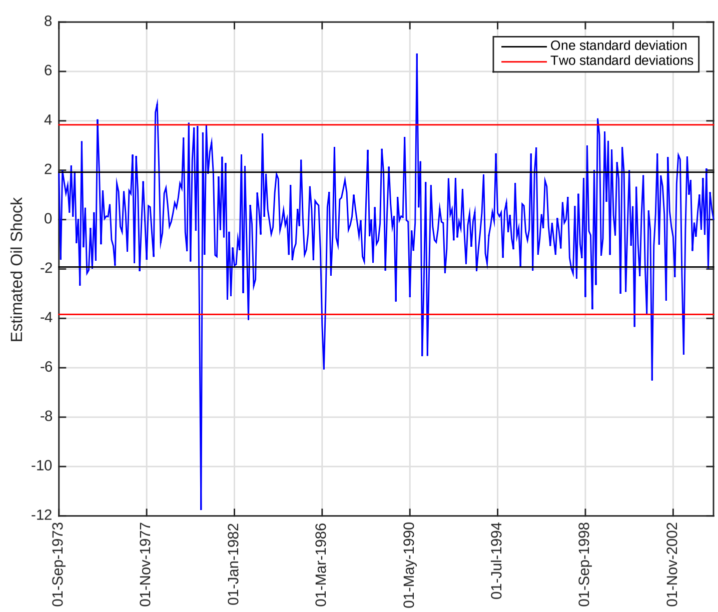
APPENDIX C: EXTRA FIGURES

The last two figures report the estimated structural shocks using the plug-in version of equation (2.8).

Figure 6: Estimated Structural Oil Shock

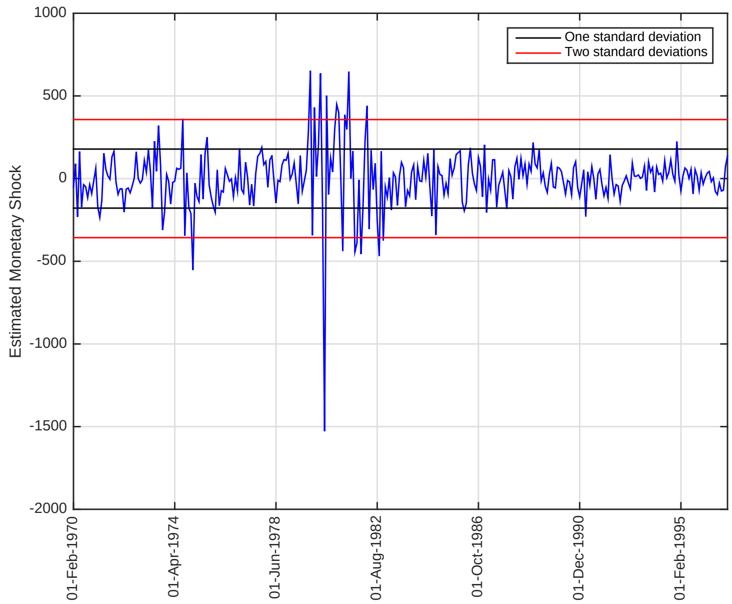


(a) Hamilton

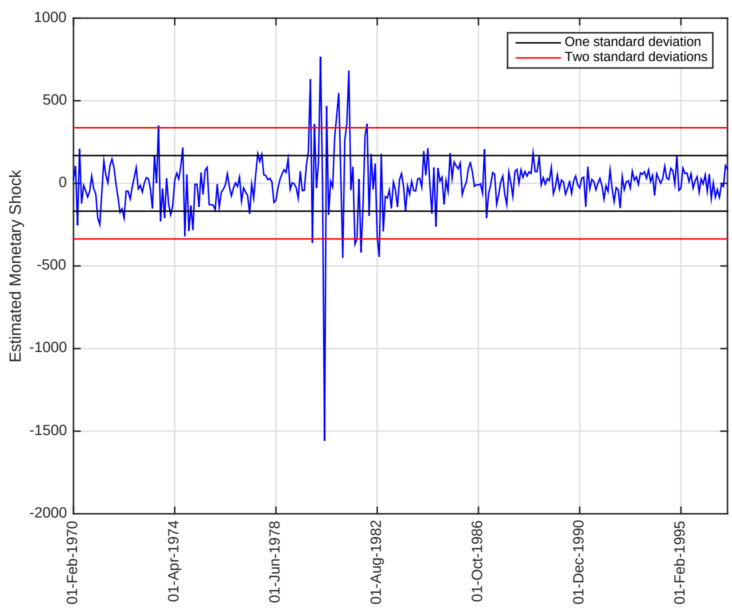


(b) Kilian

Figure 7: Estimated Structural Monetary Shock



(a) Romer and Romer



(b) Sims and Zha