Inference in Structural Vector Autoregressions Identified With an External Instrument

First Draft: June 2012
This Draft: November 2018

José L. Montiel Olea
Department of Economics, Columbia University

James H. Stock
Department of Economics, Harvard University
and the National Bureau of Economic Research

and

Mark W. Watson*
Department of Economics and the Woodrow Wilson School, Princeton University
and the National Bureau of Economic Research

*We have benefited from comments from many seminar participants including those at UC Santa Barbara, UC Berkeley, NESG-2012, NBER-NSF Time Series Meetings-2012, Harvard/MIT, Cowles Foundation, LACEA-LAMES, UCL, Erasmus, ASSA-2014, Princeton, UPF, and Columbia. We would like to thank Luigi Caloi and Hamza Husain for excellent research assistance. All errors are our own. A suite of MATLAB programs for carrying out the calculations described in the paper is available at https://github.com/jm4474/SVARIV.

Abstract

This paper studies Structural Vector Autoregressions in which a structural shock of interest (e.g., an oil supply shock) is identified using an external instrument. The external instrument is taken to be correlated with the target shock (the instrument is relevant) and to be uncorrelated with other shocks of the model (the instrument is exogenous). The potential weak correlation between the external instrument and the target structural shock compromises the large-sample validity of standard inference. We suggest a confidence set for impulse response coefficients that is not affected by the instrument strength (i.e., is weak-instrument robust) and asymptotically coincides with the standard confidence set when the instrument is strong.

Keywords: Narrative approach, instrumental variables, weak identification, impulse response functions

1. Introduction

An increasingly important line of research in Structural Vector Autoregressions (SVARs) uses information in variables not included in the system to identify dynamic causal effects, which in VAR terminology are “structural impulse response functions”. The work of Romer and Romer (1989) is a key precursor to this literature. Their reading of the minutes of the Federal Reserve Board allowed them to pinpoint dates 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’s “narrative approach” to construct time series that capture exogenous changes affecting the macroeconomy.[1]

Most of the papers in this literature have treated these exogenous variables as a time series of structural shocks, and estimated their dynamic effects using distributed lag regressions. But these external series are not, strictly speaking, the shocks of interest. Rather, they are variables plausibly correlated with a particular structural shock, and uncorrelated with others. It seems natural, therefore, to treat these exogenous variables as “external instruments”: the macroeconometric counterpart of microeconometric instrumental variables constructed using quasi-experiments. Stock (2008) makes this point and shows how these external instruments can be used to identify structural shocks in SVARs and their impulse response functions.[2] Recent applications of the external-instrument approach to SVAR identification and estimation include Stock and Watson (2012), Mertens and Ravn (2013, 2014), Gertler and Karadi (2015), and Mertens and Montiel Olea (2018).[3]

[1] Notable examples include unanticipated defense spending shocks (Ramey and Shapiro (1998)), monetary policy shocks (Romer and Romer (2004)), oil market shocks (Hamilton (2003), Killian (2008)), tax shocks (Romer and Romer (2010)), and government spending shocks (Ramey (2011)). In a similar vein, asset price changes measured using high frequency data from financial markets have been used to measure exogenous changes attributed to monetary policy; important early examples include Rudebusch (1998), Kuttner (2001). See Ramey (2016) for additional references and discussion.

[2] Stock (2008) refers to this as the “natural experiment approach” to SVAR identification, but it has subsequently become known as the “external instrument approach.” The idea that these exogenous variables can serve as instruments goes back at least as far as Romer and Romer (1989) (see the comments by Blanchard and Sims in the published discussion) and has been used in distributed lag regressions (e.g., Hamilton (2003)). Stock (2008) is the first reference that we are aware of that explicitly incorporates external instruments in SVAR analysis, and that framework has been adopted in the subsequent SVAR literature.

[3] Much recent empirical work has used local-projection methods (Jordà (2005)) in place of SVARs to estimate dynamic causal effects, increasingly using external instruments. This paper focuses on SVARs with external

External instruments impose second moment restrictions that identify SVAR shocks and associated impulse response coefficients, variance decompositions, and other objects of interest in SVAR analysis. Standard inference about these objects can be carried using linear and nonlinear GMM methods; see Mertens and Ravn (2013). However, an important lesson from the use of Instrumental Variables (IV) regression in microeconometrics is that standard methods are unreliable when instruments are only weakly correlated with the variable of interest. A large weak-instrument IV regression literature has developed both diagnostics for weak instruments and weak-instrument robust inference procedures. See Stock, Wright, and Yogo (2002) and I. Andrews, Stock, and Sun (2018) for surveys.

External instruments in macroeconometrics can also be weak, and in this paper we discuss how this potential weakness compromises the validity of standard inference in SVARs. Building on methods that have been successfully used in IV regression, we propose weak-instrument robust inference methods for impulse response coefficients. The primary focus of the paper is on estimating the dynamic effects of single structural shock identified by a single external instrument. We discuss extensions to the overidentified case briefly in the text and in more detail in Appendix A.3.2.

The paper is organized as follows. Section 2 lays out the SVAR and shows how an external instrument can be used to identify the structural shock of interest, its impulse response coefficients, historical effect on the variables in the VAR, and contribution to forecast error variances. Section 3 focuses on inference for impulse response coefficients, studying first the strong-instrument properties of standard estimators, then the distortions caused by a weak instrument. When the instrument is weak, the estimator of the impulse response function is biased towards the Cholesky decomposition impulse response function, with the shock of interest ordered first. Section 4 then presents a confidence set (based on the classical Fieller (1944) and Anderson-Rubin (1949) methods) that retains its validity when the external instrument is weak and coincides with the standard confidence interval when the instrument is strong. This section briefly discusses several other issues, including diagnostic tests for weak instruments, and the extension of the inference methods to allow for multiple instruments. Section 5 includes a brief empirical illustration that focuses on the effect of an oil-supply shock on oil prices using Killian's

---

instruments. Stock and Watson (2018) surveys the recent Local-Projection (LP-IV) contributions and compares SVAR and LP methods.

(2009) 3-variable SVAR. Section 6 presents Monte Carlo evidence illustrating the problems of conducting standard inference in the presence of a weak instrument and the benefits of our proposed method. Section 7 offers a summary and conclusions.

Generic Notation: If A is a matrix, A_{ij} denotes its ij'th element, A_i denotes its i'th column, vec(A) denotes the vectorization of A, and vech(A) vectorizes the lower triangular portion of the symmetric matrix A. The vector e_i denotes the i'th column of I_n, the $n\times n$ identity matrix.

2. Model and Identification

2.1 The Model

The model is the standard stationary finite-order structural vector autoregression. We use the following notation:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + \eta_t\ , \tag{1.1}$$

where Y_t is $n\times 1$, and η_t is a vector of reduced-form VAR innovations. The reduced form innovations are related to a vector of structural shocks, ε_t, via

$$\eta_t = \Theta_0 \varepsilon_t, \tag{1.2}$$

where Θ_0 is a non-singular $n\times n$ matrix; thus, we assume that the structural model is invertible in the sense that the VAR forecast errors at date t are a nonsingular transformation of the structural errors at date t. The structural shocks are assumed to be serially and mutually uncorrelated, with

$$\mathbf{E}(\varepsilon_t) = 0 \ \text{ and } \mathbf{E}(\varepsilon_t\varepsilon_t) = D = \text{diag}(\sigma_1^2,\ \ldots,\ \sigma_n^2).$$

The implied value of the covariance matrix for the reduced form innovations is

$$\mathbf{E}(\eta_t\eta_t') = \Sigma = \Theta_0 D \Theta_0'. \tag{1.3}$$

Y_t has a structural moving average representation given by

$$Y_t = \sum_{k=0}^{\infty} C_k(A)\Theta_0 \varepsilon_{t-k}, \tag{1.4}$$

where the notation $C_k(A)$ emphasizes the dependence of the MA coefficients on the AR coefficients in $A = (A_1, A_2, \ldots, A_p)$. Specifically:

$$C_k(A) = \sum_{m=1}^{k} C_{k-m}(A) A_m, \; k = 1, 2, \ldots \tag{1.5}$$

with $C_0(A) = I_n$ and $A_m = 0$ for $m > p$; see Lütkepohl (1990, 2007).

The structural "impulse response" coefficient is the response of $Y_{i,t+k}$ to a one-unit change in $\varepsilon_{j,t}$, which from (1.4) is

$$\partial Y_{i,t+k}/\partial \varepsilon_{j,t} = e_i' C_k(A)\Theta_0 e_j, \tag{1.6}$$

where e_j denotes the j'th column of the identity matrix I_n.

Target Shock. We focus on identifying the impulse responses to a single structural shock (e.g., an oil supply shock in the empirical illustration in Section 5), and without loss of generality this shock is ordered first, so the shock of interest is $\varepsilon_{1,t}$. The impulse responses with respect to this target shock are determined by $\Theta_0 e_1 = \Theta_{0,1}$, the first column of Θ_0.

Scale Normalization. Because $\eta_t = \Theta_0 \varepsilon_t$, the scales of $\varepsilon_{1,t}$ and $\Theta_{0,1}$ are not separately identified. We normalize the scale of the target shock $\varepsilon_{1,t}$ so that it is interpretable in terms of the observed data Y_t. Specifically, we normalize the size of target shock to have a 1 unit-contemporaneous effect on a pre-specified variable Y_{i^*}, that is $\partial Y_{i^*,t}/\partial \varepsilon_{1,t} = 1$. In the empirical illustration, ε_1 is an oil-supply shock and Y_{i^*} is the percent change in global crude oil production, so we consider an oil supply shock that leads to a 1 percent increase in oil production. Without loss of generality, we order the data so that $i^* = 1$ and because $\partial Y_{1,t}/\partial \varepsilon_{1,t} = \Theta_{0,11}$, the scale normalization sets $\Theta_{0,11} = 1$. This is the "unit effect" normalization discussed in detail in Stock and Watson (2016).

2.2 Using an external instrument to identify impulse responses and other structural parameters

External Instrument. Let z_t denote a scalar random variable that can serve as an instrument (or "proxy") for the target shock. The stochastic process for $\left\{(\varepsilon_t, z_t)\right\}_{t=1}^{\infty}$ is assumed to satisfy

Assumption 1 (External Instrument)

(A1.1) $\mathbf{E}[z_t\, \varepsilon_{1,t}] = \alpha \neq 0$.

(A1.2) $\mathbf{E}[z_t\, \varepsilon_{j,t}] = 0$ for $j \neq 1$.

This assumption is the SVAR analogue of the familiar definition of an instrumental variable: (A1.1) says z_t is correlated with the target shock (the instrument is *relevant*), and (A1.2) says that z_t is uncorrelated with the other shocks (the instrument is *exogenous*).

Identification of the impulse response coefficients. Let $\lambda_{k,i} = \partial Y_{i,t+k}/\partial \varepsilon_{1,t}$ denote an impulse response coefficient of interest. From (1.6), $\lambda_{k,i}$ depends on the VAR coefficients A and the first column of Θ_0, that is $\Theta_{0,1}$. From Assumption 1, $\Theta_{0,1}$ is identified up to scale by the covariance between z_t and the reduced form innovations η_t:

$$\boldsymbol{\Gamma} = \mathbf{E}(z_t \eta_t) = \mathbf{E}(z_t\, \Theta_0 \varepsilon_t) = \alpha\, \Theta_{0,1}\ . \tag{1.7}$$

Using the scale normalization $\Theta_{0,11} = 1$, $\boldsymbol{\Gamma}_{11} = \mathbf{E}(z_t \eta_{1,t}) = \alpha$, so that

$$\Theta_{0,1} = \boldsymbol{\Gamma}/\boldsymbol{\Gamma}_{11} = \boldsymbol{\Gamma}/e_1'\boldsymbol{\Gamma}. \tag{1.8}$$

Thus, the structural impulse response with respect to $\varepsilon_{1,t}$ follows directly from (1.6):

$$\lambda_{k,i} = e_i' C_k(A) \boldsymbol{\Gamma}/e_1'\boldsymbol{\Gamma}. \tag{1.9}$$

***Identification of* $\{\varepsilon_{1,t}\}$.** The instrument can be used to recover the structural shock $\varepsilon_{1,t}$ from the reduced-form innovations η_t. To see how, use $\mathbf{E}(z_t\eta_t) = \Gamma = \alpha\,\Theta_0 e_1$ and $\Sigma = \mathbf{E}(\eta_t\eta_t') = \Theta_0 D\Theta_0'$ to write the projection of z_t onto η_t as

$$\begin{aligned}\text{Proj}\left(z_t \middle| \eta_t\right) &= \Gamma'\Sigma^{-1}\eta_t = \left(\alpha\Theta_0 e_1\right)'\left(\Theta_0 D\Theta_0'\right)^{-1}\eta_t = \left(\alpha\Theta_0 e_1\right)'\left(\Theta_0 D\Theta_0'\right)^{-1}\Theta_0\varepsilon_t \\ &= \alpha e_1 D^{-1}\varepsilon_t = \left(\alpha/\sigma_1^2\right)\varepsilon_{1,t}.\end{aligned} \tag{1.10}$$

This projection determines $\varepsilon_{1,t}$ up to the scale factor (α/σ_1^2); dividing by $\left(\Gamma'\Sigma^{-1}\Gamma\right)^{1/2}$ yields $\varepsilon_{1,t}/\sigma_1$ up to sign.

***Identification of the historical decomposition of* $\{Y_t\}$.** Another object of interest in SVAR analysis is a decomposition of the historical values of Y_t into a component associated with current and lagged values of $\varepsilon_{1,t}$, say $Y_t(\varepsilon_1)$, and a residual component associated with the other structural innovations. The structural moving average (1.4) yields:

$$Y_t(\varepsilon_1) = \sum_{k=0}^{\infty} C_k(A)\Theta_{0,1}\varepsilon_{1,t-k} = \sum_{k=0}^{\infty} C_k(A)\left(\Gamma'\Sigma^{-1}\Gamma\right)^{-1}\Gamma\Gamma'\Sigma^{-1}\eta_{t-k} \tag{1.11}$$

where the second equality follows from $\left(\Gamma'\Sigma^{-1}\Gamma\right)^{-1}\Gamma\Gamma'\Sigma^{-1}\eta_{t-k} = \Theta_{0,1}\varepsilon_{1,t}$.[4]

Identification of the variance decomposition. The variance decomposition measures the fraction of the k-step ahead forecast error variance for $Y_{i,t+k}$ associated with $\varepsilon_{1,t+h}$ for $h = 1, \ldots, k$. Denoting this by $FEVD_{k,i}$, a direct calculation using (1.5) and (1.11) yields:

$$FEVD_{k,i} = \frac{\Gamma'\left(\sum_{s=0}^{k} C_s(A)'e_i e_i'C_s(A)\right)\Gamma}{(\Gamma'\Sigma^{-1}\Gamma)e_i'\left(\sum_{s=0}^{k} C_s(A)\Sigma C_s(A)\right)e_i}. \tag{1.12}$$

[4] Which in turn follows from $\Gamma=\alpha\Theta_{0,1}$ (from (1.8)), $\Gamma'\Sigma^{-1}\eta_t = (\alpha/\sigma_1^2)\varepsilon_{1,t}$ (from (1.10)), and $\Gamma'\Sigma^{-1}\Gamma = \alpha^2/\sigma_1^2$.

3. Inference about impulse response coefficients

3.1 Plug-in estimators and δ-method confidence sets

The plug-in estimator for $\lambda_{k,i}$ replaces A and Γ in (1.9) with the corresponding estimators:

$$\hat{\lambda}_{k,i}(\hat{A}_T, \hat{\Gamma}_T) = e_i' C_k(\hat{A}_T)\hat{\Gamma}_T / e_1'\hat{\Gamma}_T, \tag{2.1}$$

where $\hat{A}_T$ is the least squares estimator of the VAR coefficients and $\hat{\Gamma}_T$ is the sample covariance between z_t and the VAR residuals.[5]

When z_t is a strong instrument, confidence sets for impulse responses can be formed in the usual way. Under standard assumptions $[vec(\hat{A}_T - A), (\hat{\Gamma}_T - \Gamma)]$ has a limiting normal distribution. A δ-method calculation implies that $\sqrt{T}\ [\hat{\lambda}_{k,i}(\hat{A}_T, \hat{\Gamma}_T) - \lambda_{k,i}(A,\Gamma)]$ is approximately distributed $N(0, \sigma^2_{k,i})$ in large samples, where $\sigma^2_{k,i}$ depends on the limiting variance for the estimators $(\hat{A}_T, \hat{\Gamma}_T)$ and the gradient of $\lambda_{k,i}(A,\Gamma)$ with respect to (A,Γ). This leads to the usual $100\times(1\text{-}a)\%$ large sample confidence set for $\lambda_{k,i}$:

$$CS^{Plug-in} = \left\{ \lambda_{k,i} \left| \frac{T\left(\lambda_{k,i}(\hat{A}_T, \hat{\Gamma}_T) - \lambda_{k,i}\right)^2}{\hat{\sigma}^2_{T,k,i}} \le \chi^2_{1,1-a} \right. \right\}, \tag{2.2}$$

---

[5] Letting $S_{ab} = T^{-1}\sum_{t=1}^{T} a_t b_t'$ for matrices a_t and b_t, $\hat{A}_T = S_{YX}S_{XX}^{-1}$ with $X_t = (1, Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-p})'$, $\hat{\Gamma}_T = S_{z\hat{\eta}}$ where $\hat{\eta}_t = Y_t - \hat{A}_T X_t$, and $\hat{\Sigma}_T = S_{\hat{\eta}\hat{\eta}}$.

where $\hat{\sigma}^2_{T,k,i}$ is a consistent estimator for $\sigma^2_{k,i}$ and $\chi^2_{1,1-a}$ is the 1-a percentile of the χ^2_1 distribution.

However, the presence of $e_1'\hat{\Gamma}_T$ in the denominator of (2.1) suggests that the large-sample normal approximation of the distribution of the plug-in estimator may be poor when $e_1'\Gamma$ is small, leading to poor coverage of the resulting δ-method confidence set. We outline the familiar reasoning in the following subsection.

3.2 Weak-instrument asymptotic distributions of plug-in estimators of impulse response coefficients

The vector Γ is proportional to the covariance between the target structural shock, ε_{1t}, and the instrument, z_t, that is $\Gamma = \alpha\, \Theta_{0,1}$. To allow for models in which α can be arbitrarily close to zero, while recognizing that sampling variability depends on the sample size T, consider a sequence of models in which $\mathbf{E}(z_t\varepsilon_{1,t}) = \alpha_T$, where $\alpha_T \rightarrow \alpha$, and $\alpha = 0$ is allowed.[6] This framework allows, for example, strong instruments (with $\alpha_T = \alpha \neq 0$), but also weak instruments as in Staiger and Stock (1997) (with $\alpha_T = a/\sqrt{T}$). Let $\Gamma_T = \alpha_T\Theta_{0,1}$. Under a variety of primitive assumptions, the estimators ($\hat{A}_T, \hat{\Gamma}_T, \hat{\Sigma}_T$) will be asymptotically normally distributed after centering them at the true values (A, Γ_T, Σ) and scaling by $\sqrt{T}$. This is summarized in Assumption 2:

<u>Assumption 2: (Asymptotic normality of reduced-form statistics)</u>

$$\sqrt{T}\begin{pmatrix} vec(\hat{A}_T - A) \\ (\hat{\Gamma}_T - \Gamma_T) \\ vech(\hat{\Sigma}_T - \Sigma) \end{pmatrix} \Rightarrow \begin{pmatrix} \varsigma \\ \xi \\ \varphi \end{pmatrix} \sim N(0, W) \tag{2.3}$$

[6] Formally, this means considering a sequence of stochastic processes, say P_T, for $\{z_t, \varepsilon_t\}_{t=1}^T$, where the expectation is taken with respect this process, so that $\mathbf{E}(z_t\varepsilon_{1,t}) = \alpha_T$ denotes $\mathbf{E}_{PT}(z_t\varepsilon_{1,t}) = \alpha_T = \alpha_T$ and so forth.

In a strong instrument setting, $\Gamma_T = \Gamma \neq 0$, and Assumption 2, along with the δ-method implies that the plug-in estimator is asymptotically normally distributed and this serves as the basis for the plug-in confidence set (2.2).

But suppose instead that the instrument is weak in the sense that $\alpha_T = a/\sqrt{T}$ where a is held fixed as $T \to \infty$. A straightforward calculation (see Appendix A.1.1 for details) then shows that the plug-in estimator has the weak-instrument asymptotic representation,

$$\hat{\lambda}_{k,i}(\hat{A}_T, \hat{\Gamma}_T) \Rightarrow \lambda_{k,i} + \frac{\delta_{k,i}'\xi}{e_1'\xi + a\Theta_{0,11}}\ , \tag{2.4}$$

where $\delta_{k,i} = (e_i' C_k(A) - \lambda_{k,i} e_1')'$ and ξ is defined in (2.3). Thus, the plug-in estimator is equal to the true value of the impulse response plus a ratio of correlated normal random variables. This is the SVAR analogue of Staiger and Stock (1997)'s asymptotic representation of the IV estimator in a just-identified linear model with a single right-hand side endogenous regressor and a single weak instrument.[7] The parameter $(a\Theta_{0,11})^2/Var(e_1'\xi)$ is analogous to the so-called concentration parameter in IV regression.

Just as in the IV model the plug-in estimator (2.1) is not consistent, the usual Wald test for testing the null hypothesis $\lambda_{k,i} = \lambda_0$ does not have the correct size, and the plug-in confidence sets (2.2) (which are based on inverting the Wald test) will not have the proper coverage probability.[8]

When instruments are weak, the plug-in estimator (2.1) is biased toward the probability limit of the estimator of the impulse response coefficient estimated by ordering $Y_{1,t}$ first in a Cholesky decomposition of the innovation variance matrix, that is, when the shock of interest is identified by placing it first in a Wold causal ordering. This result obtains by noting that, under the unit effect normalization, the IV estimator of $\Theta_{0,1}$ is obtained as the IV estimator in the regressions,

[7] The results in Staiger and Stock (1997) imply that whenever the first-stage coefficient of a linear IV model is local-to-zero the IV estimator, denoted β_{IV}, converges in distribution to $\beta + z_1/(z_2 + c)$, where $(z_1\ z_2)$ are bivariate normal, β is the true parameter, and c is the scalar localization parameter.

[8] In Section 6 we provide Monte Carlo evidence, based on a plausible empirical design, showing that the distortions associated to a weak external instrument are not negligible. In our simulations, the estimated coverage of a nominal 95% confidence interval can be as low as 85%.

$$\hat{\eta}_{jt} = \Theta_{0,1j}\hat{\eta}_{1,t} + u_t,\ j = 2,\ldots, n \tag{2.5}$$

using the instrument z_t (or its innovation), where $\hat{\eta}_t$ is the vector of innovations and u_t is a generic error term (see for example Stock and Watson (2018), equation (21)). This formulation of the SVAR-IV estimator of $\Theta_{0,1}$ makes it possible to apply standard results about the bias of the distribution of the IV estimator under weak instruments (c.f., Nelson and Startz (1990), Staiger and Stock (1997). In particular, if z_t is weak, the IV estimator will be biased towards the probability limit of the OLS estimator of (2.5). The OLS estimator of $\Theta_{0,1}$ in (2.5) is the first column of the Cholesky decomposition with the shock of interest ordered first. This result suggests caution in interpreting the near-coincidence of Cholesky and external instrument estimates of impulse responses as evidence in favor of the Cholesky ordering assumption without evidence on instrument strength.

3.3 Weak-instrument asymptotic distributions of plug-in estimators of other objects of interest in SVAR analysis

As shown in equations (1.10), (1.11), and (1.12), the time series of the target shock, its contribution to y_t, and the forecast error decomposition can be written as functions of the reduced-form VAR parameters, (A,Σ), the covariance of the instrument and the reduced form errors, Γ, and (for the target shock and historical decompositions) the data. Inference about the true values of these objects – their values associated with the true value of (A,Σ,Γ) – is standard and straightforward when z_t is a strong instrument. Examination of (1.10), (1.11), and (1.12) shows that, with Γ bounded away from zero, each of these objects is a well-behaved smooth function of (A,Σ,Γ). Assumption 2 and the δ-method then imply that the corresponding plug-in estimators are normally distributed in large samples with a covariance matrix that is readily computed from the large-sample covariance matrix of $(\hat{A}_T, \hat{\Gamma}_T, \hat{\Sigma}_T)$ and the relevant gradient vector.

Equations (1.11) and (1.12) show that the historical and variance decompositions are ratios of quadratic functions of Γ, so (generically) the resulting gradients either converge to zero or diverge as $\Gamma \to 0$. Thus, δ-method inference based on plug-in estimators for the historical and variance decompositions is not robust to weak instruments.

The weak-instrument representation of the estimate of the shock, for use in historical decomposition, and of the FEVD are, respectively,

$$\hat{\varepsilon}_{1,t} \Rightarrow \left(\xi + a\Theta_0\right)' \Sigma^{-1}\eta_t \Big/ \left[\left(\xi + a\Theta_0\right)' \Sigma^{-1}\left(\xi + a\Theta_0\right)\right]^{1/2} \tag{2.6}$$

$$\widehat{FEVD}_{k,t} \Rightarrow \frac{\Gamma^*\left(\sum_{s=0}^{k} C_s(A)e_i e_i' C_s(A)\right)\Gamma^*}{\left(\Gamma^*\Sigma\Gamma^*\right)\sum_{s=0}^{k} e_i' C_s(A)\Sigma C_s(A)' e_i}, \tag{2.7}$$

where $\Gamma^* = \xi + a\Theta_{0,1}$. These expressions are derived in Appendix A.1.2.

4. Weak-instrument robust confidence sets

The analogy between inference in the linear IV model and SVAR impulse responses carries over to the construction of weak-instrument robust confidence sets using analogues of Fieller-method confidence sets for the ratio of two normal means (Fieller (1944)) and the Anderson-Rubin (1949) confidence sets for coefficients in the linear IV model. To see how, it is useful to briefly review Fieller's problem and the Anderson-Rubin confidence set.

Fieller's problem and Anderson-Rubin confidence set. Suppose (X, Y) are bivariate normally distributed with mean ($\beta\pi$, π) and covariance matrix Σ. Fieller's problem is to construct a confidence interval for the ratio of the two means, β. The null hypothesis $\beta = \beta_0$ implies that $X - \beta_0 Y \sim N(0, \sigma(\beta_0)^2)$, where $\sigma(\beta_0)^2 = \sigma_X^2 - 2\beta_0\sigma_{XY} + \beta_0^2\sigma_Y^2$, and $q(\beta_0) = (X - \beta_0 Y)^2/\sigma(\beta_0)^2 \sim \chi_1^2$. With Σ known, the 100%(1-a) Anderson-Rubin (AR) confidence set for β can then be constructed as $CS^{AR} = \{\beta \mid q(\beta) \leq \chi^2_{1,1-a}\}$.[9,10] An important property of the AR confidence set is

[9] In Fieller's (1944) formulation, X and Y correspond to sample means from an i.i.d. normal sample, Σ is unknown and inference is based on the squared Student-t distribution instead of the χ_1^2 distribution. Anderson and Rubin (1949) showed how to extend Fieller's construction to IV regression (a nontrivial extension at the time). In the AR

that it is valid for any value π , including values arbitrarily close to zero. When $\pi = 0$, β is not identified, and (as discussed in footnote 12) the confidence set will have infinite length with probability $1-a$.

4.1 Inference for impulse response coefficients (single structural shock identified by a single external instrument)

To understand how the AR method can be used to form weak-instrument robust confidence sets for the coefficients of impulse response function, suppose the instrument is valid (so that $\alpha_T \neq 0$), but potentially weak ($\alpha_T \to \alpha$, where $\alpha = 0$ is allowed). Let H_T denote the 2×1 vector composed of the numerator and denominator of the expression defining the impulse response coefficient in (1.9):

$$H_T = \begin{bmatrix} e_i' C_k(A)\Gamma_T \\ e_1'\Gamma_T \end{bmatrix}, \tag{2.8}$$

so that $\lambda_{k,i} = H_{T,1}/H_{T,2}$, and let $\hat{H}_T$ denote the plug-in estimator of H_T constructing by replacing (A,Γ_T) with $(\hat{A}_T, \hat{\Gamma}_T)$. Note that H_T is a differentiable function of A and a linear function of Γ_T, so that (from Assumption 2 and the δ-method) $\sqrt{T}(\hat{H}_T - H_T) \Rightarrow \eta \sim$ N(0,Ω), where Ω depends on W and the gradient of $\lim_{T\to\infty} H_T$ with respect to (A,Γ). Importantly, this result follows regardless of the strength of the instrument ($\Gamma_T = \alpha_T \Theta_{0,1} \to \alpha\, \Theta_{0,1} = \Gamma$, with $\Gamma = 0$ allowed).

Large sample theory thus yields the approximation $\hat{H}_T \overset{a}{\sim} N(H_T, T^{-1}\Omega)$, where the parameter of interest is the ratio of the means $H_{T,1}/H_{T,2}$. This is Fieller's problem. The null

formulation, X is the OLS estimator of the regression coefficient of the outcome variable on the instrument and Y is the OLS estimator of the first-stage coefficient.

[10] The inequality $q(\beta) \leq \chi^2_{1,1-a}$ defining the Anderson-Rubin confidence set is quadratic in β, which in standard form can be written as $a\beta^2 + b\beta + c \leq 0$, where (a,b,c) are functions of (X,Y,Σ). The structure of the problem (c.f., Fieller (1944) and Kendall and Stuart (1979 section 20.35)) yields the following features of the confidence set: (1) $\hat{\beta} \in CS^{AR}$; (2) if $a > 0$, the confidence set is the interval $(-b \pm (b^2\text{-}4ac)^{1/2})/2a$; (3) if $a < 0$, the confidence interval includes either the entire real line or the union of the two sets $(-\infty, -[b + (b^2\text{-}4ac)^{1/2}]/2a)$ and $(-[b - (b^2\text{-}4ac)^{1/2}]/2a, \infty)$; (4) when $Y^2/\sigma_Y^2 \leq \chi^2_{1,1-a}$ (so the hypothesis $\mu_Y = 0$ is not rejected), the confidence set for β is the entire real line.

hypothesis $\lambda_{k,i} = \lambda_0$ imposes a linear restriction on the means: $H_{T,1} - \lambda_0 H_{T,2} = 0$, which can be tested using the Wald statistic

$$q_T(\lambda_0) = \frac{T\left(\hat{H}_{T,1} - \lambda_0 \hat{H}_{T,2}\right)^2}{\hat{\omega}_{T,11} - 2\lambda_0\hat{\omega}_{T,12} + \lambda_0^2\hat{\omega}_{T,22}} \quad , \tag{2.9}$$

where $\hat{\omega}_{T,ij}$ are consistent estimators of the elements of the covariance matrix Ω. Inverting this test yields the Anderson-Rubin confidence set

$$CS^{AR} = \{\lambda_{k,i} \mid q_T(\lambda_{k,i}) \leq \chi^2_{1,1-a} \}. \tag{2.10}$$

The weak and strong-instrument validity of the CS^{AR} is summarized in the following:

Proposition 1 (Asymptotic validity of CS^{AR})

Let CS^{AR} (1-a) denote the AR confidence set (2.10) with nominal coverage $1-a$, and let P_T denote the probability distribution for $\{Y_t, z_t\}_{t=1}^T$ under the stochastic process corresponding to α_T. Suppose

(i) Assumptions 1 and 2 are satisfied,

(ii) $\alpha_T \to \alpha$ (which may be 0)

(iii) $\hat{\Omega}_T \xrightarrow{p} \Omega \neq 0$

Then: $\lim_{T\to\infty} P_T(\lambda_{k,i} \in CS^{AR}(1\text{-}a)) = 1\text{-}a$.

Proof: See Appendix A.2.

The covariance matrix in the asymptotic distribution of $\hat{H}_T$ is $\Omega = G(A,\Gamma)WG(A,\Gamma)'$, where G denotes the limit of the gradient of H_T in (2.8) with respect to (A,Γ) and W is asymptotic variance of the estimators from Assumption 2.[11] This suggests the estimator $\hat{\Omega}_T = G(\hat{A}_T, \hat{\Gamma}_T)\hat{W}_T G(\hat{A}_T, \hat{\Gamma}_T)'$, so that (iii) is satisfied if $G(A,\Gamma) \neq 0$ and $\hat{W}_T$ is consistent for W.

A natural question to ask is whether the weak-instrument robustness of the AR confidence set comes at the cost of reduced accuracy (or increased expected length) when the instrument is strong. The next proposition shows that that the "distance" between the Anderson-Rubin confidence set and the δ-method confidence interval converges to zero when the instrument is strong. In this sense, there is no cost from using the robust confidence set.

Let $d_H(A,B)$ denote the Hausdorff distance between two subsets A and B of the real line:

$$d_H(A,B) = \max\left\{ \sup_{x\in A}\inf_{y\in B} d(x,y), \sup_{y\in B}\inf_{x\in A} d(x,y) \right\}.$$

Proposition 2 (Strong-instrument asymptotic equivalence of $CS^{Plug\text{-}in}$ and CS^{AR})

Let $CS^{Plug\text{-}in}$ (1-a) and CS^{AR} (1-a) denote the confidence sets given in (2.2) and (2.10) with nominal coverage $1-a$. Suppose

(i) Assumptions 1 and 2 are satisfied,

(ii) $\alpha_T \to \alpha \neq 0$,

(iii.a) $\hat{\Omega}_T \xrightarrow{p} \Omega \neq 0$, and

(iii.b) $\hat{\sigma}^2_{T,k,i} \overset{p}{\to} \sigma^2_{k,i}$.

Then: $\sqrt{T} d_H\left(CS_T^{AR}(1-a), CS_T^{Plug-in}(1-a) \right) \overset{p}{\to} 0.$

Proof: See Appendix A.2.2

Proposition 2 applies to the just-identified case. Inference for the overidentified case is discussed below.

[11] We derive analytical expressions for $G(A,\Gamma)$ and include them in the MATLAB suite that implements the Anderson-Rubin confidence set.

4.2 Diagnostic for weak instruments

The instrument is weak if $\mathbf{E}(z_t\varepsilon_{1t}) = \alpha$ is small relative to the sampling error in $\hat{\alpha}_T$. The expression for the estimator of $\Theta_{0,1}$ as the IV estimator in (2.5) shows that the heteroskedasticity-robust first-stage F statistic provides a measure of the strength of the instrument in this setting too, where the first-stage regression is of $Y_{1,t}$ against z_t (including VAR lags of Y_t as exogenous controls).[12] The heteroskedasticity-robust first-stage F can be compared to the Stock-Yogo (2005) critical values or to some rule of thumb, such as F>10. When there are multiple instruments and heteroskedasticity is a concern, the Montiel Olea-Pflueger (2013) effective first-stage F is recommended, for the reasons discussed in I. Andrews, Stock, and Sun (2018).

An alternative diagnostic arises from noting that, with $\Theta_{0,11}$ normalized to equal 1, α equals $\Gamma_{1,1}$. Because $\sqrt{T}\left(\hat{\Gamma}_T - \Gamma_T\right) \xrightarrow{d} N(0, W_\Gamma)$, the Wald statistic $\xi_1 = T\hat{\Gamma}_{T,1}^2 / \hat{W}_{\Gamma,11}$ also is a measure of instrument strength. Under weak instrument asymptotics, ξ_1 has the same noncentrality parameter as the heteroskedasticity-robust first-stage F, although algebraic manipulations and numerical simulations suggest ξ_1 will tend to be smaller in finite samples than the first-stage F. The statistic ξ_1 has the feature that the 100%(1-a) Anderson-Rubin (AR) confidence set is a bounded interval if and only if $\xi_1 > \chi^2_{1,1-a}$ (see footnote 10).

4.3 Extensions

Overidentification. If there are $m > 1$ instruments for the target structural shock, it is conceptually straightforward to extend the Anderson-Rubin confidence set (see Appendix A.3.2). In the over-identified case, the Anderson-Rubin confidence set is known to be valid for both weak- and strong-instruments, but inefficient relative to standard confidence sets when the instruments are strong. Appendix A.3.2 also discusses how weak-instrument robust methods developed for over-identified IV regression, such as the Lagrange Multiplier and the Quasi-Conditional Likelihood Ratio test, can be applied for inference about impulse response coefficients in the SVAR model.

Inference about FEVDs and historical decompositions. For inference about impulse responses, the lack of robustness of plug-in δ-methods can be solved using the Anderson-Rubin

[12] Under some circumstances it might be desirable to also add lags of z_t; see Stock and Watson (2018).

method. Broadly speaking, this is possible because Γ enters "linearly" in the numerator and denominator of (1.9). Such a simplification is not possible for historical and variance decompositions because Γ enters the numerator and denominator of (1.11) and (1.12) as quadratic functions. Weak-instrument robust inference for these objects is not addressed in this paper and remains an area of on-going research.[13]

[13] One way to construct a conservative weak-instrument confidence set for the forecast error decomposition is to note (from (1.12)) that $FEVD_{k,i} = \omega' Q_{k,i}(A,\Sigma)\omega$, where $\omega = \Sigma^{1/2}\Gamma/(\Gamma'\Sigma\Gamma)^{1/2}$ and $Q_{k,i}(A,\Sigma)$ is a matrix that depends on the reduced-form parameters only through (A,Σ). Because $\omega'\omega = 1$, $\mathrm{mineig}(Q_{k,i}(A,\Sigma)) \leq FEVD_{k,i} \leq \mathrm{maxeig}(Q_{k,i}(A,\Sigma))$, and a confidence set can be constructed for this interval. However, because $Q_{k,i}(A,\Sigma)$ does not depend on any identifying information in Γ, this is a confidence set for the variance decomposition associated with *any* possible structural shock, and is therefore likely to be extremely conservative.

5. An illustrative example

Killian (2009) used a 3-variable SVAR to investigate the effect of oil-supply and oil-demand shocks on oil production and oil prices. In this section we use Killian's model and data to illustrate the external-instrument methods discussed above.

The three variables in Killian's (2009) SVAR are the percent change in global crude oil production (*prod*), real oil prices (*rpo*), and a global real activity index of dry goods shipments (*rea*). Killian uses these variables to identify three structural shocks – oil supply (ε^{Supply}), aggregate demand ($\varepsilon^{Ag.Demand}$), and oil-specific demand ($\varepsilon^{Oil\text{-}Spec.Demand}$) – using the Wold causal ordering (ε^{Supply}, $\varepsilon^{Ag.Demand}$, $\varepsilon^{Oil\text{-}Spec.Demand}$) in the VAR with variables ordered as (*prod, rea, rpo*). We focus on the oil supply shock identified using the same reduced-form VAR as Killian (2009), but with an external instrument.

We use Killian's (2008) measure of "exogenous oil supply shocks" as the external instrument. The instrument measures shortfalls in OPEC oil production associated with wars and civil disruptions. Because this variable measures shortfalls in production, it is plausibly correlated with the structural oil supply shock ε^{Supply}, and because it measures shortfalls associated with political events such as wars in the Middle East, it is plausibly uncorrelated with the two oil demand shocks. Thus, Killian's (2008) measure plausibly satisfies the conditions for an external instrument given in Assumption 1.

Of course, while Assumption 1 implies that the external instrument is valid, the internal validity of the SVAR depends on additional assumptions, notably (1.1) and (1.2). From (1.1), the VAR coefficients are assumed to be time-invariant, and from (1.2), the structural shocks are contemporaneous linear functions of the VAR reduced-form forecast errors: $\varepsilon_t = \Theta_0^{-1}\eta_t$. The recent empirical literature using SVARs to model the oil market has questioned both of these assumptions (see Stock and Watson (2016) for discussion). We are sympathetic to these concerns and to the post-Killian (2009) literature that expands the variables in the VAR (e.g., Aastveit (2014)), and uses sign restrictions to help identify the dynamic effects of oil supply shocks in both frequentist (e.g, Killian and Murphy (2012)) and Bayes (e.g., Baumeister and Killian (2015)) settings. That said, the simplicity of Killian's (2009) 3-variable time-invariant VAR makes it an ideal framework for illustrating the use of external instruments.

Killian's (2009) analysis used monthly data from 1973:M1-2007:M12. The instrument, Killian's (2008) exogenous oil supply shock series, is available from 1973:M1-2004:M9, and we use the common sample period (1973:M1-2004:M9) for the analysis.[14] Following Killian (2009), the VAR is estimated using $p = 24$ lags and a constant term. The covariance matrix W is estimated using a standard Eicker-White robust estimator (equivalently, a Newey-West HAC estimator with 0 lags). The confidence sets presented in Section 3 were based on δ-method approximations that relied on gradients of particular functions with respect to A and Γ. We have created a Matlab suite to implement our confidence set using analytical formulae for these gradients. We also suggest a simple bootstrap-like method that involves sampling $(\text{vec}(\hat{A}_T), \hat{\Gamma}_T)$ from an estimated normal distribution consistent with Assumption 2. Details are provided in Appendix A.4.[15]

Weak-instrument diagnostics. The statistic $\xi_1 = T\hat{\Gamma}^2_{T,1} / \hat{W}_{\Gamma,11} = 4.4$ and the robust first-stage statistic is 9.4. Both statistics are below the Staiger-Stock value of 10, suggesting that the instrument is weak. However, because $\xi_1 > 3.84$ (the 95% χ^2_1 critical value), the 95% Anderson-Rubin weak-instrument confidence sets for the impulse response coefficients are bounded intervals (see footnote 12).

Impulse response coefficients. Figure 1 shows the estimated impulse response coefficients and corresponding $CS^{Plug\text{-}in}$ and CS^{AR} confidence sets.[16] The 68% weak-instrument robust CS^{AR} confidence sets essentially coincide with the strong-instrument $CS^{Plug\text{-}in}$ intervals, but the 95% CS^{AR} confidence sets suggest considerably more uncertainty than their strong-instrument counterparts. An important finding in Killian (2009), was that Cholesky-identified oil supply shocks had small effects on oil prices (implying highly elastic oil demand). This is evident in panel A, which plots (in red) the estimated impulse response coefficients for the Cholesky-identified shock. The point estimates imply that a Cholesky-identified oil supply shock that

---

[14] We use the common sample period for (y_t, z_t) for convenience. In principle, the entire sample period can be used to estimate the VAR parameters, and a shorter sample period used to estimate Γ. This entails only a modification in estimator used for the covariance matrix W in assumption 2.

[15] The bootstrap method is more computationally intensive than the δ-method (because it requires re-sampling from the reduced-form parameters and constructing quantiles of a test statistic over a grid of possible values for the impulse response coefficients), but does not require analytical computation of the gradient of the expression in equation (2.5).

[16] In appendix A.4 we also compare the CS^{AR} reported in Figure 1 with its bootstrap version.

increases oil supply by 1% on impact, leads to a fall in prices of 0.03% on impact and has a maximum price effect of -0.07% after four months. In contrast, the corresponding supply shock identified using the external instrument leads to fall in prices of 0.14% on impact and maximum price effect of -0.22% after four months. But, while the external-instrument identified price effects are larger than the Cholesky-identified effect, both are small in an absolute sense, and Killian's overall conclusion of small price effects is consistent with the external-instrument estimates and associated weak-instrument robust confidence sets.

6. Monte Carlo Evidence

We conduct a simple Monte Carlo exercise to analyze the coverage of the $CS^{Plug\text{-}in}$ and CS^{AR} confidence sets. The data generating process for the Monte Carlo exercise is parameterized by the matrix of autoregressive coefficients, the matrix of contemporaneous impulse response coefficients, the variance of the structural innovations, and the joint distribution of the external instrument and target shock. We explain our choice of these parameters below.

We consider $T = 356$ observations from a 3-dimensional vector Y_t generated by a reduced-form VAR model with reduced-form parameters (A,Σ) equal to those estimated from Killian's (2008) data. The sample size matches the number of observations in Kilian's application.

For the matrix of contemporaneous impulse response coefficients, Θ_0, we make the first column equal to $e/\sqrt{e'\Sigma^{-1}e}$ where $e = (1\ 1\ \text{-}1)'$. The signs of this vector are in line with the typical interpretation of an expansionary supply shock. The remaining columns of Θ_0 are chosen to satisfy the equation $\Theta_0\Theta_0' = \Sigma$.

We use a linear measurement error model for the external instrument:

$$z_t = \mu_Z + \alpha\varepsilon_{1,t} + \sigma_Z\nu_t$$

The structural shocks $\varepsilon_t = (\varepsilon_{1,t}\ \varepsilon_{2,t}\ \varepsilon_{3,t})$ and ν_t are independent standard normal random variables. The parameters μ_Z and σ_Z are chosen to match the first and second moment of Kilian's external instrument. We vary the parameter α to obtain two different values of the concentration parameter $(T\alpha)^2/Var(z_t\,\eta_{1t})$: 3.7 and 10.09. Our simulations, reported in Figure 2, show that the coverage of the nominal 95% δ-method confidence interval ($CS^{Plug\text{-}in}$) can be as low as 85% for some horizons when the concentration parameter is small. The CS^{AR} confidence exhibits some distortion (presumably because the critical values are based on large sample approximations), but it is never below 90%. As expected, the coverage of $CS^{Plug\text{-}in}$ improves as the concentration parameter increases.

In Appendix A.5 we also report the coverage of the bootstrap version of the CS^{AR}. There is a slight improvement in the coverage of CS^{AR} confidence set, but the difference does not seem substantial. This suggests that although there can be some gain in using critical values that are not computed explicitly using large sample formulae, improved coverage comes from choosing a weak-instrument robust procedure. Finally, we also report simulations for a sample size of T=1500. We use this to show that in a sufficiently large sample the Monte Carlo coverage of CS^{AR} essentially coincides with the nominal level.

7. Conclusions

This paper studied SVARs identified using an external instrument. The external instrument was taken to be correlated with the target shock (e.g., the short-fall of OPEC oil production is correlated with the aggregate oil supply shock) and to be uncorrelated with other shocks in the model. Standard estimators for the model's reduced-form parameters (including the covariance of the instrument and the reduced-form errors) are normally distributed in large samples. We provide formulae for SVAR parameters like impulse response coefficients or variance decompositions as a function of these reduced-form parameters. The analysis shows that the large-sample distribution of such SVAR parameter estimators depends on the strength of the instrument. When the instrument is highly correlated with the target structural shock (so that the instrument is strong), standard δ-method arguments imply that SVAR parameter estimators are approximately normally distributed and the usual Wald tests and associated confidence sets have the correct size and coverage probability. However, when the external instrument is weak, the distribution of SVAR parameter estimators is not well approximated by the Normal distribution, so the usual Wald tests and confidence sets are invalid.

This paper shows that confidence sets for impulse response coefficients constructed using Fieller (1944) and Anderson and Rubin (1949) methods are valid when external instruments are weak and asymptotically coincide with the usual confidence sets when instruments are strong and the model is just identified. Thus, these weak-instrument robust confidence sets should routinely be used for impulse response coefficients identified with an external instrument. Along with our weak-instrument robust confidence sets, we suggest that practitioners report either the Wald statistic for the null hypothesis that the external instrument is irrelevant, or the heteroskedasticity-robust first-stage *F* statistic as described in Section 4.2. Large values of these statistics (e.g., above 10) suggest approximately valid coverage of standard 95% confidence intervals.

References

Aastveit, K.A. (2014). "Oil Price Shocks in a Data-Rich Environment." *Energy Economics* 45, 268-279.

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.

Andrews, I., J.H. Stock, and L. Sun (2018). "Weak Instruments in IV Regression: Theory and Practice," manuscript.

Baumeister, C. and J.D. Hamilton (2018). "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," manuscript.

Fieller, E.C. (1944). "A Fundamental Formula in the Statistics of Biological Assay, and Some Applications," *Quarterly Journal of Pharmacy and Pharmacology,* Vol. 17, 117-123.

Gertler, M. and P. Karadi (2015). "Monetary Policy Surprises, Credit Costs and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

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*, vol. 95(1), 161-182.

Kendall, M. and A. Stuart (1979), *The Advanced Theory of Statistics, Vol. 2: Inference and Relationship*, London: Griffin.

Killian, L. (2008). "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics,* 90 (2), 216-240.

Killian, L. (2009). "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 99 (3), 1053-1069.

Kilian, L. and D.P. Murphy (2012). "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association* 10, 1166-1188.

Kuttner, K.N. (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market." *Journal of Monetary Economics* 47, 523-544.

Mertens, K. and Montiel Olea, J.L. (2018). "Marginal Tax Rates and Income: New Time Series Evidence," *Quarterly Journal of Economics,* Volume 133 (4), 1803-1884.

Montiel Olea, J. and C. Pflueger (2013). "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31: 358-369.

Nelson, C. and R. Startz (1990). "Some further results on the exact small sample properties of the instrumental variable estimator," *Econometrica*, 58, 967–976.

Ramey, V. (2011). "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 126, 1-50.

Ramey, V. A. (2016). "Macroeconomic shocks and their propagation," in J. B. Taylor and H. Uhlig (eds) *Handbook of Macroeconomics* Vol. 2A, Amsterdam: Elsevier.

Ramey, V. A. and M. D. Shapiro (1998). "Costly capital reallocation and the effects of government spending," in *Carnegie-Rochester Conference Series on Public Policy*, Elsevier, vol. 48, 145–194.

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.

Romer, C. D. and D. H. Romer (2010). "The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks," *American Economic Review*, 100, 763–801.

Rudebusch, G.D. (1998). "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review,* 39, 907-931.

Staiger, D. and J. Stock (1997). "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H. (2008). *What is New in Econometrics: Time Series, Lecture 7*, Short course lectures, NBER Summer Institute, at http://www.nber.org/minicourse_2008. html.

Stock, J.H. and M.W. Watson (2012). "Disentangling the Channels of the 2007-2009 Recession," *Brookings Papers on Economic Activity*, No. 1, 81-135.

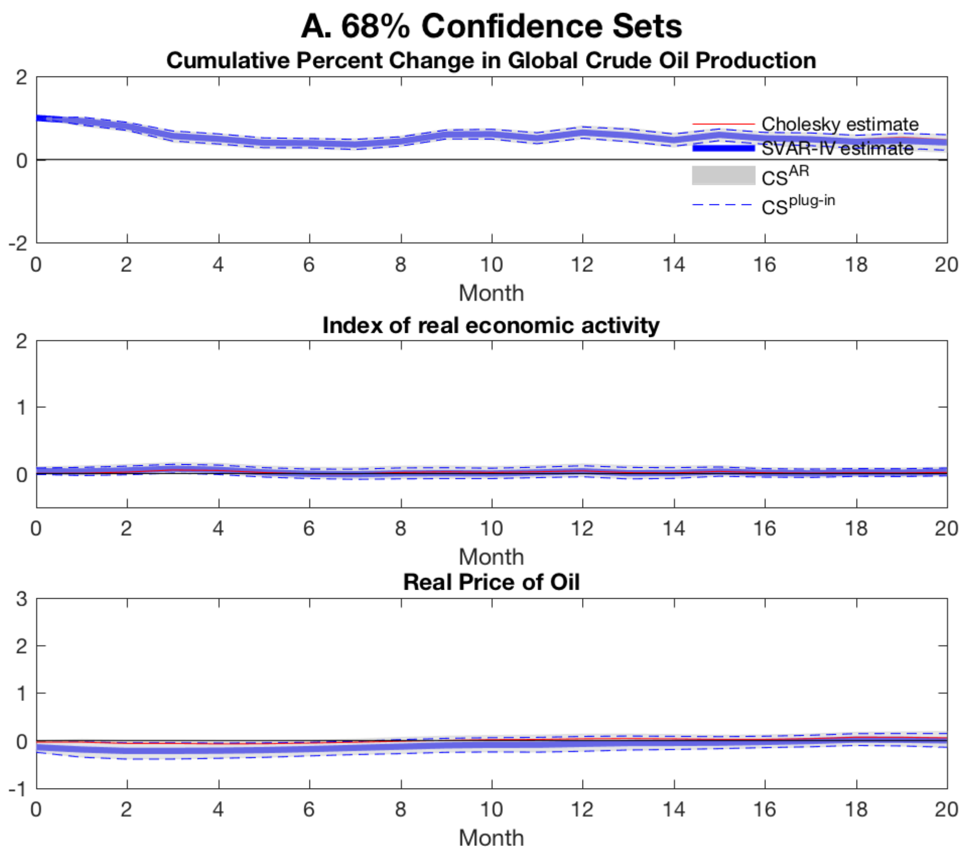
Stock, J.H. and M.W. Watson (2016): "Factor Models and Structural Vector Autoregressions in Macroeconomics," in J. B. Taylor and H. Uhlig (eds) *Handbook of Macroeconomics* Vol. 2A, Amsterdam: Elsevier.

Stock, J.H. and M.W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, 128 (May), 917-948.

Stock, J.H., J.H. Wright, amd M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics,*20(4),518-529.

Stock, J.H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression,"), Ch. 5 in J.H. Stock and D.W.K. Andrews (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge University Press, 80-108.

Figure 1: Impulse response coefficients for an oil-supply shock



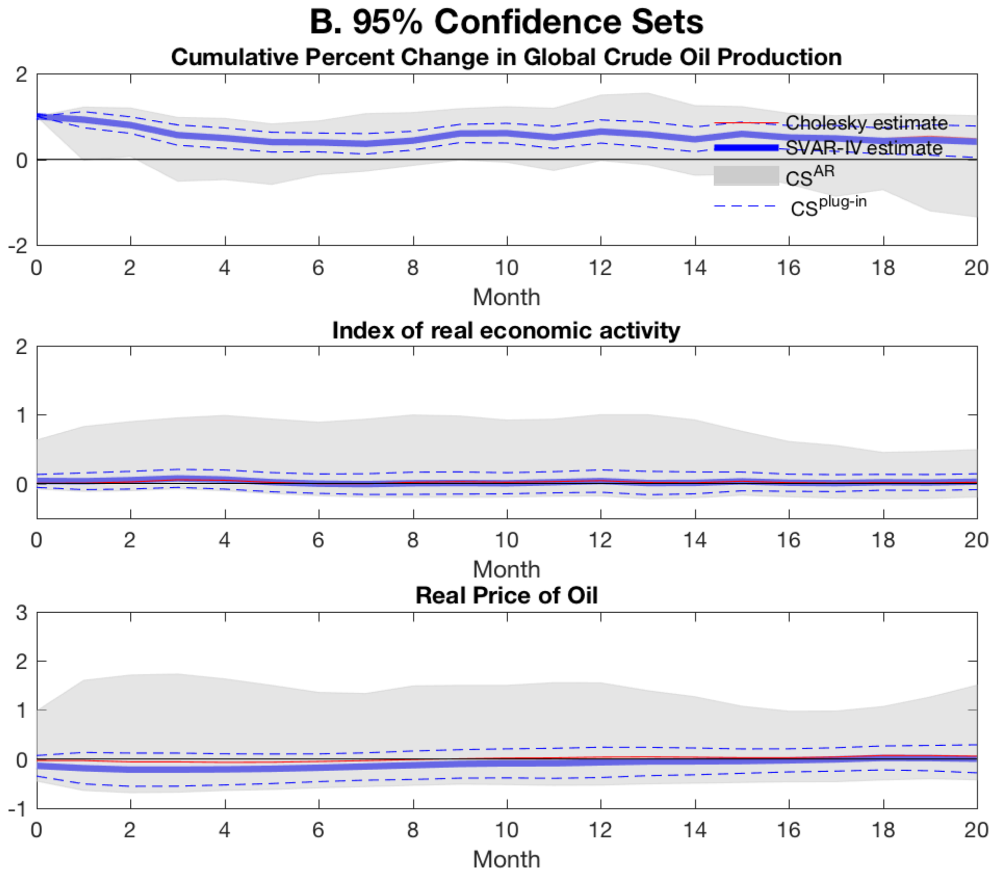
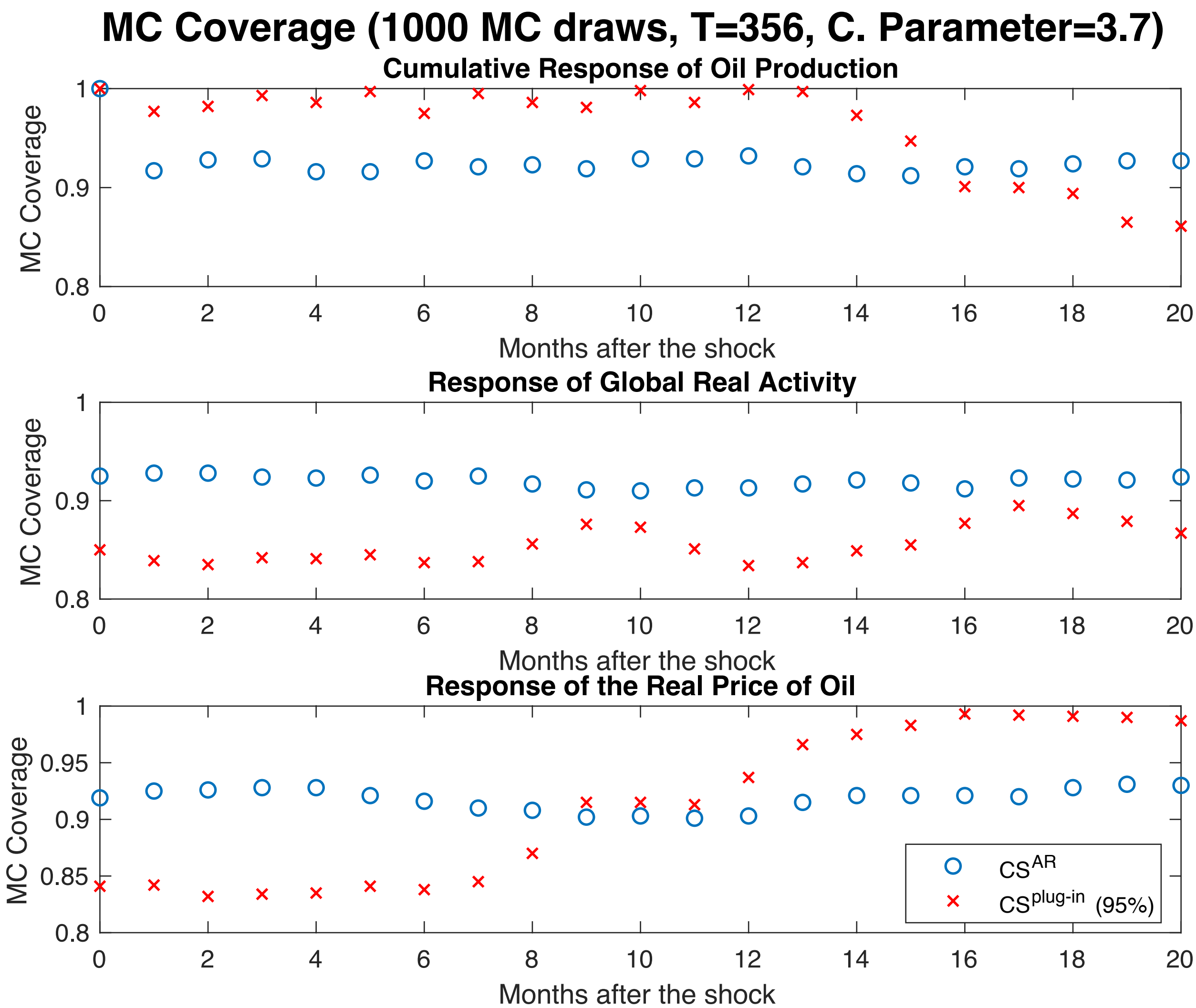
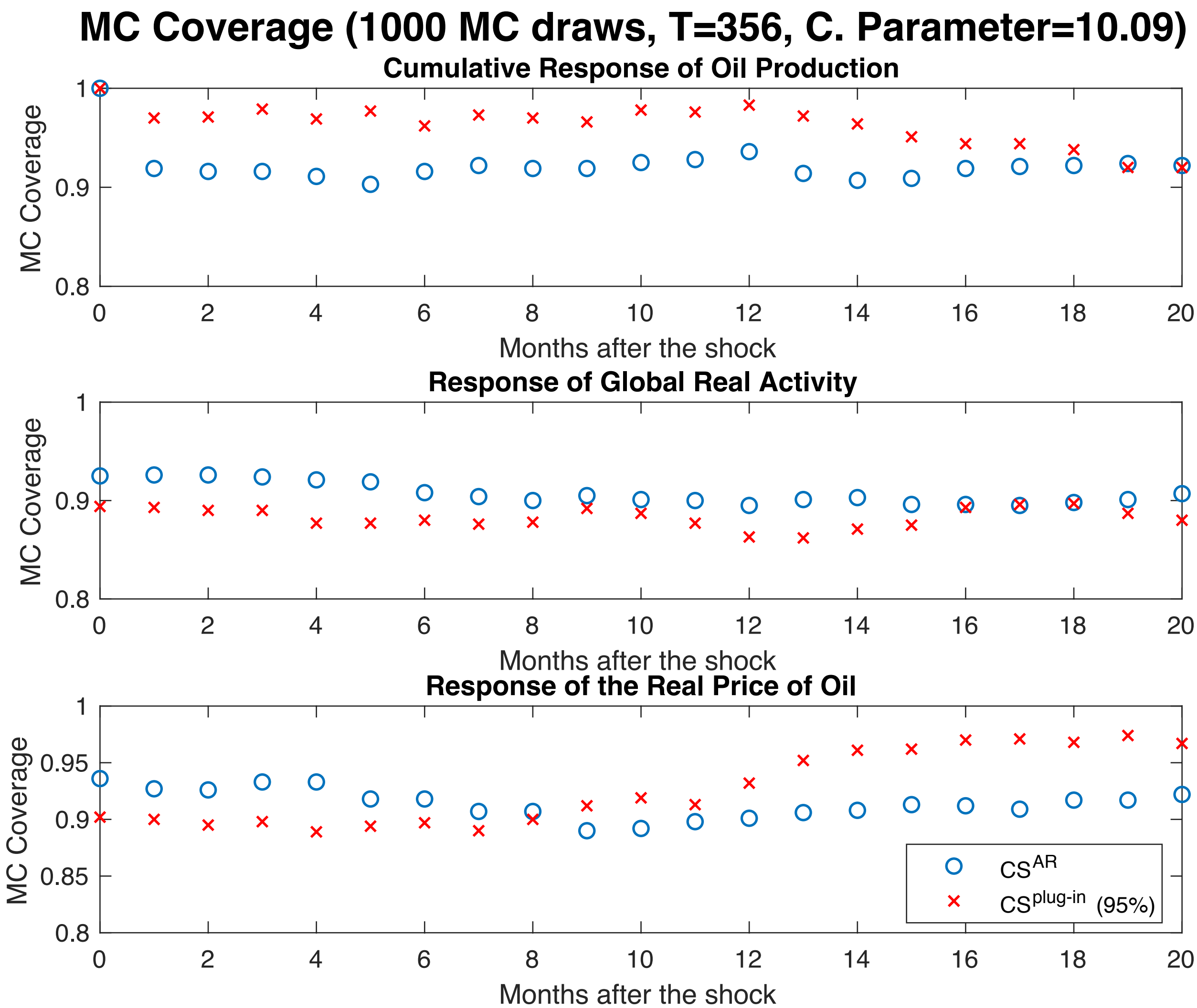

B. 95% Confidence Sets
Cumulative Percent Change in Global Crude Oil Production
Cholesky estimate
SVAR-IV estimate
CS^AR
CS^plug-in
Month
Index of real economic activity
Month
Real Price of Oil
Month

Figure 2: Coverage rates for nominal 95% confidence intervals

A. Concentration Parameter: 3.7



B. Concentration Parameter: 10.09



Notes: These figures show coverage rates for nominal 95% $CS^{Plug\text{-}in}$ and CS^{AR} confidence sets for impulse responses at horizons 0-20 periods (labeled "months" in the figures). The SVAR design is discussed in the text. The experiments use $T = 356$ and 1000 Monte Carlo simulations.