

Inference for VARs Identified with Sign Restrictions

Hyungsik Roger Moon

University of Southern California

Frank Schorfheide*

University of Pennsylvania, CEPR, NBER

Eleonora Granziera

University of Southern California

Mihye Lee

University of Southern California

November 3, 2009

Abstract

There is a growing literature that partially identifies structural vector autoregressions (SVARs) by imposing sign restrictions to the responses of a subset of the endogenous variables to a particular structural shock (sign-restricted SVARs). To date, the methods that have been used are only justified from a Bayesian perspective. This paper develops methods of constructing error bands for impulse response functions of sign-restricted SVARs that are valid from a frequentist perspective. We also provide a comparison of frequentist and Bayesian error bands. (JEL: C1, C32)

KEY WORDS: Bayesian Inference, Frequentist Inference, Partially Identified Models, Sign Restrictions, Structural VARs.

*Correspondence: H.R. Moon, E. Granziera, M. Lee: Department of Economics, University of Southern California, KAP 330C, Los Angeles, CA 90089. Email: moonr@usc.edu, granzier@usc.edu, mi-hyelee@usc.edu. F. Schorfheide, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. Email: schorf@ssc.upenn.edu. We thank Eric Renault as well as seminar participants at the 2009 Canadian Econometric Study Group Meeting and the University of Pennsylvania for helpful comments. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 0617803.

1 Introduction

During the three decades following Sims (1980)'s "Macroeconomics and Reality" structural vector autoregressions (SVARs) have become an important tool in empirical macroeconomics. They have been used for macroeconomic forecasting and policy analysis, to investigate the sources of business cycle fluctuations, and to provide a benchmark against which modern dynamic macroeconomic theories can be evaluated. The most controversial step in the specification of a structural VAR is the mapping between reduced form one-step-ahead forecast errors and orthogonalized, interpretable, structural innovations. Most SVARs in the literature have been constructed by imposing sufficiently many restrictions such that the relationship between structural innovations and forecast errors is one-to-one. However, in the past decade, starting with Faust (1998), Canova and Nicolo (2002), and Uhlig (2005), empirical researchers have used more agnostic approaches that generate bounds on structural impulse response functions by restricting the sign of certain responses. We will refer to this class of models as sign-restricted SVARs. They have been employed, for instance, to measure the effects of monetary policy shocks (Faust, 1998; Canova and Nicolo, 2002; Uhlig, 2005), technology shocks (Dedola and Neri, 2007; Peersman and Straub, 2009), government spending shocks (Mountford and Uhlig, 2008; Pappa, 2009), and oil price shocks (Baumeister and Peersman, 2008; Kilian and Murphy, 2009).

Empirical findings about the dynamic effects of structural economic shocks are typically reported in terms of (point) estimates of impulse response functions, surrounded by error bands. If the autoregressive system is stationary and the SVAR is sufficiently restricted such that the impulse response functions are identifiable, then the reported error bands are typically interpretable from both a frequentist as well as a Bayesian perspective. In large samples they delimit approximately valid frequentist confidence intervals and Bayesian credible sets. Since impulse responses in sign-restricted SVARs can only be bounded, they belong to the class of partially-identified econometric models, using the terminology of Manski (2003). As shown in detail in Moon and Schorfheide (2009), the large-sample numerical equivalence of frequentist confidence sets and Bayesian credible sets breaks down in partially identified models. The error bands for sign-restricted SVARs that have been reported literature thus far, are only meaningful from a Bayesian perspective and cannot be interpreted as frequentist confidence intervals. The contribution of this paper is to provide methods of constructing error bands that delimit valid frequentist confidence intervals.

We construct confidence sets for impulse responses through a point-wise testing procedure. This method dates back to work by Anderson and Rubin (1949) and is widely employed to implement identification-robust inference. It has been used in the weak-instrument literature, e.g. Dufour (1997) and Staiger and Stock (1997), and starting with Chernozhukov, Hong, and Tamer (2007, henceforth CHT) also in the literature on partially identified econometric models. Frequentist inference in partially identified models is a rapidly growing field and the methods are predominantly tailored toward applications in microeconometrics. Rather than providing a comprehensive survey of the literature, we will subsequently place our contribution in the context of three related papers: CHT, Rosen (2008), and Andrews and Guggenberger (2009).

Our approach can be described as follows. We utilize consistently estimable reduced-form parameters that can be converted into a vector ϕ of orthogonalized yet non-structural impulse responses. The object of interest to the econometrician is a vector θ of impulse responses to a structural shock, e.g. a monetary policy shock. Since the sign-restricted SVAR is only partially identified, each ϕ is associated with a set of structural responses $\Theta(\phi)$, which is called the identified set. Starting point for our inference is an estimator $\hat{\phi}$ of the reduced form parameter, which is assumed to have a Gaussian limit distribution. We then choose an objective function of the form

$$Q_T(\theta) = \operatorname{argmin}_{\mu \geq 0, \|q\|=1} \left(m_1(\hat{\phi}, q) - m_2(\theta, \mu) \right)' W_T \left(m_1(\hat{\phi}, q) - m_2(\theta, \mu) \right)$$

which is constructed such that $Q_T(\theta) = 0$ if $\theta \in \Theta(\hat{\phi})$ and $Q_T(\theta) > 0$ otherwise. Here μ corresponds to the slackness in the inequalities defined by the sign restrictions and q is a nuisance parameter on a unit hypersphere that transforms orthogonalized responses into structural responses. Formal definitions of the functions $m_1(\cdot)$ and $m_2(\cdot)$ and the weight matrix W_T are not essential right now and will be provided in Section 4. We show how to construct confidence sets from the contours of $Q_T(\theta)$.

The construction of the confidence set resembles CHT's criterion function approach. However, the data enter our objective function only through the reduced form parameter estimator $\hat{\phi}$, which simplifies the analysis of the statistical properties of $Q_T(\theta)$ considerably. Rosen (2008) and Andrews and Guggenberger (2009) study inference in moment inequality models defined by a condition of the form $\mathbb{E}[m(y_t, \theta)] \geq 0$ and (among other approaches)

construct confidence intervals from the contours of the objective function

$$\tilde{Q}_T(\theta) = \operatorname{argmin}_{\mu \geq 0} \left(\frac{1}{T} \sum_{i=1}^T (m(y_t, \theta) - \mu) \right)' W_T \left(\frac{1}{T} \sum_{i=1}^T (m(y_t, \theta) - \mu) \right).$$

An important difference between our $Q_T(\theta)$ and the objective function $\tilde{Q}_T(\theta)$ studied in the moment-inequality condition literature is the presence of the additional nuisance parameter q , which needs to be concentrated out. As in CHT and Andrews and Guggenberger (2009), we show that our confidence sets are asymptotically valid in a uniform sense.

Frequentist inference in partially-identified models is non-standard because the sampling distribution of criterion functions such as $Q_T(\theta)$ or $\tilde{Q}_T(\theta)$, using our notation, tends to depend on both θ and ϕ through the distance of θ from the boundary of the identified set $\Theta(\phi)$. The construction of asymptotically valid critical values often requires computer-intensive bootstrap or sub-sampling techniques that are impractical in high-dimensional models. We circumvent this problem by bounding $Q_T(\theta)$ with a stochastic function whose quantiles are nuisance parameter free. From a practitioner's perspective an important advantage of our approach is that it is straightforward to obtain critical values that can be used to construct the contour confidence set. These critical values for the bounding function are identical to the ones derived by Rosen (2008) for the objective function $\tilde{Q}_T(\theta)$. A potential disadvantage of our method is that the use of a bound function can make the confidence interval conservative.

The remainder of the paper is organized as follows. Section 2 introduces our setup and notation. Section 3 briefly reviews Bayesian inference for a sign-restricted VAR as used in the above-referenced empirical literature. In Section 4 we develop frequentist inference procedures for impulse responses. To illustrate our methods, we conduct a small Monte-Carlo study and construct confidence intervals for the VAR considered in Uhlig (2005) in Section 6. Section 7 concludes and detailed technical proofs can be found in the Appendix.

We use the following notation throughout the remainder of the paper: “ \xrightarrow{p} ” and “ \implies ” denote convergence in probability and distribution, respectively. “ \equiv ” signifies distributional equivalence. $\mathcal{I}\{x \geq a\}$ is the indicator function that is one if $x \geq a$ and zero otherwise. We use $\operatorname{sgn}(x)$ to denote the sign of x and \propto to indicate proportionality. $0_{n \times m}$ is a $n \times m$ matrix of zeros and I_n is the $n \times n$ identity matrix. \otimes is the Kronecker product, $\operatorname{vec}(\cdot)$ stacks the columns of a matrix, and $\operatorname{tr}[\cdot]$ is the trace operator. If A is a $n \times m$ vector,

then $\|A\|_W = \sqrt{\text{tr}[WA'A]}$. In the special case of a vector, our definition implies that $\|A\|_W = \sqrt{a'Wa}$. If the weight matrix is the identity matrix we omit the subscript. A p -variate normal distribution is denoted by $N_p(\mu, \Sigma)$. A $p \times q$ matrix X is matrix-variate normal $MN_{p \times q}(M, Q \otimes P)$ if $\text{vec}(X) \sim N_{pq}(\text{vec}(M), Q \otimes P)$. A $q \times q$ matrix Σ has the Inverted Wishart $IW_q(S, \nu)$ distribution if $p(\Sigma|S, \nu) \propto |\Sigma|^{-(\nu+q+1)/2} \exp\{-\frac{1}{2}\text{tr}[\Sigma^{-1}S]\}$. If $X|\Sigma \sim MN_{p \times q}(M, \Sigma \otimes P)$ and $\Sigma \sim IW_q(S, \nu)$, we say that $(X, \Sigma) \sim MNIW(M, P, S, \nu)$. If there is no ambiguity about the dimension of the random vectors and matrices we drop the subscripts that signify dimensions. We use χ_m^2 to denote a χ^2 distribution with m degrees of freedom and $c_\tau(\chi_m^2)$ to denote its 100 τ percent critical value.

2 Setup and Notation

The evolution of the $n \times 1$ vector y_t is described by a p 'th order difference equation of the form

$$y_t = \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + u_t. \quad (1)$$

We omitted any deterministic trend terms because they are irrelevant for the subsequent discussion. We refer to (1) as reduced form representation of the VAR because the u_t 's are simply one-step-ahead forecast errors and do not have a specific economic interpretation. In order to characterize the conditional distribution of y_t given its history, one has to make a distributional assumption for u_t . We shall proceed under the assumption that $u_t \sim iidN(0, \Sigma)$. The normality is not essential for the frequentist analysis, but simplifies the derivation of small-sample posterior distributions in a Bayesian framework.

Dynamic macroeconomic theory suggests that the one-step-ahead forecast errors are functions of some fundamental innovations, for instance, to aggregate technology, preferences, or monetary policy. A structural VAR is a model in which the forecast errors are explicitly linked so such fundamental innovations. Let ϵ_t be a vector of orthogonal structural shocks with unit variances and

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t. \quad (2)$$

Here, Σ_{tr} is defined to be the lower triangular Cholesky factor of Σ and Ω is an arbitrary orthogonal matrix. It is straightforward to verify that Ω is not identifiable from the first and second moments of u_t , because by definition $\Omega\Omega' = I_n$.

Assuming that the lag polynomial associated with the VAR in (1) is invertible, we can express y_t as the following infinite-order vector moving average (VMA) process:

$$y_t = \sum_{h=0}^{\infty} C_h u_{t-h} = \sum_{j=0}^{\infty} C_h \Sigma_{tr} v_{t-h}, \quad (3)$$

where v_t is a $n \times 1$ vector of standard normal variates. The matrices of the moving average representation can be interpreted as impulse responses

$$R_h^v = \mathbb{E}[y_{t+h}|v_t = I, \mathcal{F}_t] - \mathbb{E}[y_{t+h}|\mathcal{F}_t] = \frac{\partial y_{t+h}}{\partial \epsilon_t'} = C_h \Sigma_{tr}, \quad h = 0, \dots, H. \quad (4)$$

Here the information set \mathcal{F}_t is comprised of y_t, y_{t-1}, \dots . We define the $n(H+1) \times n$ matrix $R_{0,H}^v$ that stacks the matrices R_h^v , $h = 0, \dots, H$. While responses to the v_t shocks are often not what a researcher is interested in, they can be easily converted into response to the structural shocks ϵ_t :

$$R_{0,H}^e = R_{0,H}^v \Omega. \quad (5)$$

Rather than examining responses to the full vector of structural shocks, ϵ_t , we will focus on responses to one particular shock. Without loss of generality, we assume that the shock of interest is $\epsilon_{1,t}$ and will denote the first column of Ω by the $n \times 1$ vector q . We let θ be a $k \times 1$ vector of responses of selected elements of y_t at certain horizons:

$$\theta = S_\theta R_{0,H}^v q. \quad (6)$$

The vector θ is the object of our inference and S_θ is a $k \times n(H+1)$ selection matrix. We also define the vector $\phi = \text{vec}((R_{0,H}^v)')$. Notice that ϕ can be consistently estimated, as it only depends on the parameters of the reduced form VAR in (1).

At this point the vector q only has to have unit length, because it is the first column of the orthonormal matrix Ω . Restrictions on the sign of the responses of certain elements of y_t at particular horizons can be used to impose further restrictions on q . These sign restrictions either affect responses that are contained in θ , or they affect responses not contained in θ . Hence we express the sign restrictions as

$$M_\theta \theta \geq 0, \quad S_R R_{0,H}^v q \geq 0. \quad (7)$$

Here M_θ is a $r_1 \times k$ matrix that selects (and potentially pre-multiplies them by -1) the elements of θ for which we constrain the sign. The matrix S_R is of dimension $r_2 \times n(H+1)$

and selects the remaining responses (and potentially pre-multiplies them by -1) that are constrained in their sign. Thus, we are imposing a total of $r = r_1 + r_2$ sign restrictions.

Example: Suppose $n = 2$ and $y_t = \Sigma_{tr} v_t$. We restrict our attention to the response at horizon $h = 0$. Let $H = 0$, then $R_0^v = R_{0,0}^v = \Sigma_{tr}$. Suppose that y_t is composed of inflation and aggregate output growth. Moreover, $\epsilon_{1,t}$ is a demand shock that moves output and prices in the same direction. We also impose the normalization that the demand shock is *positive*. The object of interest is the inflation response to such a demand shock. Thus, $k = 1$, $S_\theta = [1, 0]$, $M_\theta = 1$, $S_R = [0, 1]$, and $r_1 = r_2 = 1$. Moreover, we obtain

$$\theta = \Sigma_{11}^{tr} q_1, \quad \theta \geq 0, \quad \Sigma_{21}^{tr} q_1 + \Sigma_{22}^{tr} q_2 \geq 0.$$

It is straightforward to verify that the inequalities constrain the impulse response θ to lie in the set

$$\Theta(\Sigma_{tr}) = \left[0, \Sigma_{11}^{tr} \sqrt{\frac{(\Sigma_{22}^{tr})^2}{(\Sigma_{21}^{tr})^2 + (\Sigma_{22}^{tr})^2}} \right]. \quad \square$$

We refer to the set of impulse responses that are consistent with a particular reduced form parameter vector ϕ as θ as the identified set $\Theta(\phi)$. A general characterization of the identified set can be obtained as follows. We vectorize Equations (6) and (7) and parameterize the slackness in the inequality relationships (7) using the vector μ . Then

$$\begin{aligned} 0 &= (S_\theta \otimes q')\phi - \theta, & 0 &= (S_R \otimes q')\phi - \mu \\ 0 &\leq \mu, & 0 &\leq M_\theta \theta, \quad \|q\| = 1. \end{aligned} \quad (8)$$

Thus, given ϕ , the vector θ is a valid impulse response if there exist vectors q and μ such that the conditions (8) are satisfied. Define the function $G(\theta, q, \mu; \phi)$ as

$$G(\theta, q, \mu; \phi) = \left\| \begin{bmatrix} (S_\theta \otimes q')\phi - \theta \\ (S_R \otimes q')\phi - \mu \end{bmatrix} \right\|_W^2, \quad (9)$$

where W is a positive definite matrix. The function $G(\theta, q, \mu; \phi)$ is constructed such that for $M_\theta \theta \geq 0$

$$\min_{\|q\|=1, \mu \geq 0} G(\theta, q, \mu; \phi) = 0, \quad (10)$$

if and only if $\theta \in \Theta(\phi)$, that is, there exists a μ and q such that the conditions (8) are satisfied. We will use a sample analogue of $G(\theta, q, \mu; \phi)$ to construct frequentist confidence intervals for θ in Section 4.

3 Bayesian Inference

As mentioned in the introduction, empirical researchers who have estimated sign-restricted SVARs have used inference procedures that are only justifiable from a Bayesian perspective. In order to be able to compare the commonly used Bayesian error bands to the frequentist error bands developed in this paper, we briefly review the Bayesian inference in a sign-restricted VAR. Write (1) as a Gaussian linear regression model:

$$y'_t = x'_t \Phi + u'_t, \quad u_t \sim \mathcal{N}(0, \Sigma). \quad (11)$$

Here $x'_t = [y'_{t-1}, \dots, y'_{t-p}]$ and $\Phi = [\Phi_1, \dots, \Phi_p]'$. The matrices Φ and Σ collect the reduced-form parameters of the VAR. In addition, let q be a unit length vector. The joint distribution of data and parameters can be factorized as follows:

$$p(Y, \Phi, \Sigma, q) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma)p(q|\Phi, \Sigma). \quad (12)$$

Here $p(Y|\Phi, \Sigma)$ is the likelihood function associated with (11), which does not depend on q . The absence of q from the likelihood function is a manifestation of the identification problem. We expressed the joint prior density of the triplet (Φ, Σ, q) as the product of a conditional density for q , $p(q|\Phi, \Sigma)$, and a marginal density for the reduced form parameters, $\bar{p}(\Phi, \Sigma)$.

The vector q defines a one-dimensional subspace of \mathbb{R}^n . Sets of lower dimensional subspaces of \mathbb{R}^n are called Grassmann manifolds. Specifying a prior distribution for q can be viewed as specifying a probability distribution on a Grassmann manifold of dimension $(n, 1)$. The uniform distribution on a Grassmann manifold is defined as the unique invariant distribution under the transformations of the manifold induced by orthonormal transformations of \mathbb{R}^n (James, 1954). The prior density $p(q|\Phi, \Sigma)$ is often chosen to imply that the conditional distribution of q is uniform. This uniform distribution is then truncated to ensure that the inequalities (7) are satisfied.

If \tilde{q} has density $p(\tilde{q}) \propto \exp[-(1/2)\tilde{q}'\tilde{q}]$, then the space spanned by \tilde{q} (or equivalently $q = \tilde{q}/\|q\|$) is uniformly distributed on the Grassmann manifold of dimension $(n, 1)$. Thus, \tilde{q} is simply an n -dimensional vector of standard normal random variables. Uhlig (2005) proposed the following joint prior distribution for the VAR parameters, here expressed in terms of \tilde{q} :

$$p(\Phi, \Sigma, \tilde{q}) \propto p_{MNIV}(\Phi, \Sigma) \exp\{-\tilde{q}'\tilde{q}/2\} \mathcal{I}\{(\Phi, \Sigma, \tilde{q}) \in \mathcal{S}\}. \quad (13)$$

$p_{MNIW}(\Phi, \Sigma)$ denoted the density of a MNIW random variable and \mathcal{S} the set of triplets $(\Phi, \Sigma, \tilde{q})$ such that the impulse responses of the corresponding structural VAR satisfy the sign restrictions. Conditional on the reduced form parameters, the indicator function truncates the uniform distribution of q . The marginal prior for the reduced form parameters is given by a “re-weighted” MNIW density:

$$p(\Phi, \Sigma) \propto p_{MNIW}(\Phi, \Sigma) \int \exp\{-\tilde{q}'\tilde{q}/2\} \mathcal{I}\{(\Phi, \Sigma, \tilde{q}) \in \mathcal{S}\} d\tilde{q}. \quad (14)$$

Since q does not enter the likelihood function, we deduce from integrating (12) with respect to q that

$$p(Y, \Phi, \Sigma) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma). \quad (15)$$

Thus, once the marginal prior for Φ and Σ has been obtained according to (14), the calculation of the posterior distribution of the reduced form parameters is not affected by the presence of the non-identifiable matrix q . Moreover, the conditional posterior density of q can be calculated as follows:

$$p(q|Y, \Phi, \Sigma) = \frac{p(Y|\Phi, \Sigma)\bar{p}(\Phi, \Sigma)p(q|\Phi, \Sigma)}{\int p(Y|\Phi, \Sigma)\bar{p}(\Phi, \Sigma)p(q|\Phi, \Sigma)dq} = p(q|\Phi, \Sigma), \quad (16)$$

which implies that the conditional distribution of the non-identifiable parameter q does not get updated in view of the data. This is a well-known property of Bayesian inference in partially identified models, see for instance Kadane (1974), Poirier (1998), and Moon and Schorfheide (2009).¹ Draws from the posterior distribution of the VAR parameters can be obtained with the Acceptance Sampler described in Uhlig (2005).

Example (Cont'd.): Suppose we start from the improper prior $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$. The posterior distribution of Σ is $IW(S, T)$, where T is the sample size and $S = \sum_{t=1}^T y_t y_t'$. It is apparent from the calculations in Section 2 that the inequality constraints (7) do not lead to a truncation of the domain of Σ . For $n = 2$ the vector q can be expressed as $q = [\cos \varphi, \sin \varphi]'$, where $\varphi \in [-\pi/2, \pi/2]$. It can be shown (James, 1954) that q is uniformly distributed on the Grassmann manifold if $\varphi \sim \mathcal{U}[-\pi/2, \pi/2]$. Conditional on Σ

¹Some authors prefer to use a prior of the form $p(\Phi, \Sigma, q) \propto p(\Phi, \Sigma)p(q)\mathcal{I}\{(\Phi, \Sigma, q) \in \mathcal{R}\}$, where $p(\Phi, \Sigma)$ corresponds to the MNIW distribution, $p(q)$ to the distribution that implies that the space spanned by q is uniformly distributed on the Grassmann manifold $(n, 1)$, and \mathcal{R} is now the set of triplets (Φ, Σ, q) for which the sign restrictions are satisfied.

the appropriately truncated prior (and posterior) distribution of φ is given by

$$\varphi|\Sigma \sim \mathcal{U} \left[-\text{sgn}(\Sigma_{21}^{tr}) \arccos \left(\sqrt{\frac{(\Sigma_{22}^{tr})^2}{(\Sigma_{21}^{tr})^2 + (\Sigma_{22}^{tr})^2}} \right), \frac{\pi}{2} \right].$$

If $\Sigma_{21}^{tr} \leq 0$, then the change of variables $\theta = \Sigma_{11}^{tr} \cos \varphi$ implies that

$$p(\theta|\Sigma) = p(\theta|Y, \Sigma) \propto \frac{\mathcal{I} \left\{ 0 \leq \theta \leq \Sigma_{11}^{tr} \sqrt{(\Sigma_{22}^{tr})^2 / ((\Sigma_{21}^{tr})^2 + (\Sigma_{22}^{tr})^2)} \right\}}{\sqrt{1 - (\theta/\Sigma_{11}^{tr})^2}}.$$

While q is uniformly distributed on the Grassman manifold, the object of interest θ is not uniformly distributed over the identified set $\Theta(\Sigma_{tr})$. The density function is monotonically increasing in θ and peaks at the upper end of the identified set. \square

4 Frequentist Inference

We now develop methods for frequentist inference about θ . We proceed in two steps. First, we construct an estimator $\hat{\phi}$ of a vector of reduced form parameters. Second, we conduct inference on θ conditional on the estimator $\hat{\phi}$. Since empirical researchers typically depict error bands for impulse response functions that delimit point-wise credible or confidence intervals, in most applications θ is a scalar, representing the response of variable $i \in \{1, \dots, n\}$ to a shock $j \in \{1, \dots, n\}$ at horizon h . Rather than proceeding with the definition $\phi = \text{vec}((R_{0,H}^v)')$ given in Section 2, we re-define ϕ as the minimal set of elements of $R_{0,H}^v$ that are necessary to construct θ and the responses that are sign-restricted in (7). Moreover, we re-place the expressions $(S_\theta \otimes q')$ and $(S_R \otimes q')$ by functions $S_\theta(q)$ and $S_R(q)$ to preserve the relationships

$$S_\theta(q)\phi = \theta \quad \text{and} \quad S_R(q)\phi \geq 0$$

under the new definition of ϕ . Notice that now the dimension m of the vector ϕ is at most $(k + r_2)n$.

Example (Cont'd.): Under the definition of ϕ given in Section 2 ϕ is a 4×1 vector with elements $\phi_1 = \Sigma_{11}^{tr}$, $\phi_2 = 0$, $\phi_3 = \Sigma_{21}^{tr}$, $\phi_4 = \Sigma_{22}^{tr}$. However, ϕ_2 is by construction zero and not needed to recover $\theta = \phi q_1$ or the sign-restricted response $\phi_3 q_1 + \phi_4 q_2$. Thus, we eliminate the redundant element and re-define $\phi = [\Sigma_{11}^{tr}, \Sigma_{21}^{tr}, \Sigma_{22}^{tr}]'$. \square

Rather than placing low-level restrictions on the VAR coefficient matrices Φ and Σ , as well as the distribution of the reduced-form innovations u_t , and deriving the distribution of $\hat{\phi}$, we directly assume that $\hat{\phi}$ has a Gaussian limit distribution. It is noteworthy that this assumption requires that all roots of the characteristic polynomial associated with the difference equation (1) lie outside of the unit circle. Hence, we are ruling out the presence of unit roots and are implicitly assuming that y_t is trend stationary.

Assumption 1 *There exists an estimator $\hat{\phi}$ of the $m \times 1$ vector ϕ such that $\sqrt{T}(\hat{\phi} - \phi) \rightarrow \mathcal{N}(0, \Lambda)$ uniformly for $\phi \in \mathcal{P}$. The matrix Λ is positive definite. Moreover, there exists an estimator $\hat{\Lambda} \xrightarrow{P} \Lambda$.*

To simplify the subsequent exposition we assume that the covariance matrix Λ is of full rank. This assumption requires that we eliminate the $n(n-1)/2$ zero elements of the lower triangular matrix R_0^v by appropriately defining ϕ . Further rank reductions can occur if the number of elements of $R_{0,H}^v$ necessary to construct θ and the sign-restricted responses exceeds the number of non-redundant coefficients in the matrices Φ and Σ . In this case the subsequent analysis has to be modified by replacing Λ^{-1} with a generalized inverse. Subsequently, we will discuss two methods of constructing confidence sets for θ . We start by taking unions of identified sets in Section 4.1 and then consider a minimum-distance approach in Section 4.2 that delivers a sharper confidence set than the first method.

4.1 Union of Identified Sets

A conceptually straightforward approach of constructing a valid confidence set for a partially identified parameters is to take the union of identified sets $\Theta(\phi)$ over all values of ϕ in a $1 - \tau$ confidence set CS_τ^ϕ for the reduced form parameter:

$$CS_{U,\tau}^\theta = \bigcup_{\phi \in CS_\tau^\phi} \Theta(\phi). \quad (17)$$

This confidence set is valid, because

$$\inf_{\phi \in \mathcal{P}, \theta \in \Theta(\phi)} P_\phi\{\theta \in CS_{U,\tau}^\theta\} \geq \inf_{\phi \in \mathcal{P}} P_\phi\{\phi \in CS_\tau^\phi\} \geq 1 - \tau.$$

The frequentist literature, e.g. Imbens and Manski (2004) and CHT, distinguishes between confidence sets for the parameter θ and confidence sets for the identified set $\Theta(\phi)$. It is straightforward to verify that the set in (17) is also a valid confidence set for $\Theta(\phi)$.

Assumption 1 implies that we can obtain an asymptotically valid confidence set for the reduced form parameter as follows

$$CS_\tau^\phi = \left\{ \phi \in \mathcal{P} \mid T \|\hat{\phi} - \phi\|_{\hat{\Lambda}^{-1}} \leq c_\tau(\chi_m^2) \right\}, \quad (18)$$

where $c_\tau(\chi_m^2)$ is the 100τ percent critical value associated with a χ_m^2 distribution. In practice it is numerically awkward to construct the union of identified sets directly. However, the following “guess-and-verify” procedure will allow a fairly straightforward implementation, which resembles a point-wise testing procedure. Define the function

$$\begin{aligned} G_T^{(1)}(\theta, q, \mu; \hat{\phi}) &= \min_{\phi \in \mathcal{P}} T \|\hat{\phi} - \phi\|_{\hat{\Lambda}^{-1}}^2 \\ \text{s.t.} \quad & 0 = S_\theta(q)\phi - \theta, \quad 0 = S_R(q)\phi - \mu \end{aligned} \quad (19)$$

and let

$$Q_T^{(1)}(\theta) = \min_{\|q\|=1, \mu \geq 0} G_T^{(1)}(\theta, q, \mu; \hat{\phi}). \quad (20)$$

The confidence set (17) for θ can then be constructed as

$$CS_{(1),\tau}^\theta = \left\{ \theta \mid M_\theta \theta \geq 0 \text{ and } Q_T^{(1)}(\theta) \leq c_\tau(\chi_m^2) \right\}. \quad (21)$$

The following lemma states that the confidence set $CS_{(1),\tau}^\theta$ is indeed identical to the union of identified sets $CS_{U,\tau}^\theta$. A formal proof is provided in the Appendix.

Lemma 1 $CS_{(1),\tau}^\theta = CS_{U,\tau}^\theta$.

It is instructive to further simplify the function $G_T^{(1)}(\theta, q, \mu; \hat{\phi})$. Define the $(k+r_2) \times m$ matrix $S(q) = [S'_\theta(q), S'_R(q)]'$. We will illustrate below that the matrix $S(q)$ has a reduced row rank for some values of q . To account for the rank reduction, it is convenient to work with its singular value decomposition:

$$S(q) = V(q)D(q)U'(q) = V_1(q)D_{11}(q)U'_1(q), \quad (22)$$

where $V = [V_1, V_2]$ and $U = [U_1, U_2]$ are orthonormal matrices and D_{11} is a diagonal $l \times l$ matrix with the non-zero singular values of $S(q)$. The dimensions of V_1 and U'_1 are $(k+r_2) \times l$ and $l \times m$, respectively.

Example (Cont'd.): Recall that we defined $\phi = [\Sigma_{11}^{tr}, \Sigma_{21}^{tr}, \Sigma_{22}^{tr}]'$. Then,

$$S(q) = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_1 & q_2 \end{bmatrix}.$$

The row-rank of $S(q)$ reduces from 2 to 1 if $q = [0, 1]'$. The singular value decompositions of $S(q)$ for $q \neq [0, 1]'$ and $q = [0, 1]'$ are given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} q_1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & q_1 & q_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 0 & q_1 & q_2 \end{bmatrix},$$

respectively. \square

The Lagrangian associated with the constrained minimization in (19) that defines $T^{-1}G_T^{(1)}(\theta, q, \mu; \hat{\phi})$ can be written as

$$\mathcal{L} = \|\hat{\phi} - \phi\|_{\hat{\Lambda}^{-1}} - \lambda'(V_1 D_{11} U_1' \phi - \psi),$$

where $\psi = [\theta', \mu']'$. The first-order conditions are

$$\begin{aligned} 0 &= \Lambda^{-1}(\phi - \hat{\phi}) - U_1 D_{11} V_1' \lambda \\ 0 &= V_1 D_{11} U_1' \phi - \psi. \end{aligned}$$

Solving for $V_1' \lambda$ and $\phi - \hat{\phi}$ yields

$$\begin{aligned} V_1' \lambda &= D_{11}^{-1} (U_1' \Lambda U_1)^{-1} D_{11}^{-1} V_1' (\psi - V_1 D_{11} U_1' \hat{\phi}) \\ \phi - \hat{\phi} &= \Lambda U_1 (U_1' \Lambda U_1)^{-1} D_{11}^{-1} V_1' (\psi - V_1 D_{11} U_1' \hat{\phi}), \end{aligned}$$

which implies that we can express $G_T^{(1)}(\theta, q, \mu; \hat{\phi})$ as

$$G_T^{(1)}(\theta, q, \mu; \hat{\phi}) = T \left\| \begin{bmatrix} S_\theta(q) \hat{\phi} - \theta \\ S_R(q) \hat{\phi} - \mu \end{bmatrix} \right\|_{V_1 D_{11}^{-1} (U_1' \Lambda U_1)^{-1} D_{11}^{-1} V_1'}^2, \quad (23)$$

where V_1 , D_{11} , and U_1 are functions of q , defined in (22). The computations to obtain the confidence set $CS_{(1),\tau}^\theta$ can be executed as follows.

Algorithm: Computation of $CS_{(1),\tau}^\theta$.

1. Choose an estimator $\hat{\phi}$ that satisfies Assumption 1 and construct a consistent estimator $\hat{\Lambda}$ of its covariance matrix.
2. Approximate the domain of $\tilde{\theta}$ by a grid $\bar{\Theta}$.
3. For each $\tilde{\theta} \in \bar{\Theta}$ such that $M_\theta \tilde{\theta} \geq 0$ compute $Q_T^{(1)}(\tilde{\theta})$ in (20) and include $\tilde{\theta}$ in $CS_{(1),\tau}^\theta$ if $Q_T^{(1)}(\tilde{\theta}) \leq c_\tau(\chi_m^2)$.

4.2 A Minimum-Distance Approach

An alternative confidence set for θ can be obtained as follows. We replace the function $G(\theta, q, \mu; \phi)$, defined in (9) to characterize the identified set, by a sample analogue, using $\hat{\phi}$ to substitute ϕ and replacing W by a weight matrix that is allowed to depend on q . We then concentrate out q and μ and construct a bound for the concentrated objective function. This bound is chosen to obtain nuisance parameter free critical values that can be used to construct confidence sets. Let

$$G_T^{(2)}(\theta, q, \mu; \hat{\phi}) = T \left\| \begin{bmatrix} S_\theta(q)\hat{\phi} - \theta \\ S_R(q)\hat{\phi} - \mu \end{bmatrix} \right\|_{W_T(q)}^2, \quad (24)$$

where $\{W_T(q)\}$ is a sequence of weight matrices to be defined below. Moreover, define

$$Q_T^{(2)}(\theta) = \min_{\|q\|=1, \mu \geq 0} G_T^{(2)}(\theta, q, \mu; \hat{\phi}). \quad (25)$$

Now consider a particular $\tilde{\theta} \in \Theta(\phi)$ and let \tilde{q} and $\tilde{\mu}$ be a particular solution of (10) given $\tilde{\theta}$ and ϕ , which implies $\tilde{\theta} = S_\theta(\tilde{q})\phi$ and $\tilde{\mu} = S_R(\tilde{q})\phi$. Then

$$G_T^{(2)}(\tilde{\theta}, q, \mu; \hat{\phi}) = T \left\| \begin{bmatrix} S_\theta(q)\hat{\phi} - S_\theta(\tilde{q})\phi \\ S_R(q)\hat{\phi} - S_R(\tilde{q})\phi + S_R(\tilde{q})\phi - \mu \end{bmatrix} \right\|_{W_T(q)}^2. \quad (26)$$

The objective function $Q_T^{(2)}(\theta)$ can be bounded by replacing q in $G_T^{(2)}(\theta, q, \mu; \hat{\phi})$ with the particular \tilde{q} instead of minimizing over the unit hypersphere:

$$Q_T^{(2)}(\theta) \leq \min_{\mu \geq 0} \left\| \begin{bmatrix} S_\theta(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) \\ S_R(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) - \sqrt{T}(\mu - \tilde{\mu}) \end{bmatrix} \right\|_{W_T(\tilde{q})}^2. \quad (27)$$

Define $S(q)$ by stacking $S_\theta(q)$ and $S_R(q)$ and apply a singular value decomposition as in (22), let $\nu = \sqrt{T}(\mu - \tilde{\mu})$, and $M_\nu = [0_{k \times r_2}, I_{r_2}]'$. Then we can write the bound as

$$Q_T^{(2)}(\theta) \leq \min_{\nu \geq -\sqrt{T}\tilde{\mu}} \left\| S(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) - M_\nu \nu \right\|_{W_T(\tilde{q})}^2. \quad (28)$$

The right-hand-side of (28) is essentially isomorphic to the large sample approximation of the objective function $\tilde{Q}_T(\theta)$ for moment inequality problems discussed in Section 1. Using Rosen's (2008) insight that the least favorable case is $\tilde{\mu} = 0$, that is, all inequalities are binding, we obtain

$$Q_T^{(2)}(\theta) \leq \min_{\nu \geq 0} \left\| S(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) - M_\nu \nu \right\|_{W_T(\tilde{q})}^2 = \bar{Q}_T^{(2)}(\tilde{q}). \quad (29)$$

Notice that the bound $\bar{Q}_T^{(2)}(\tilde{q})$ depends on θ only indirectly through \tilde{q} . We now select a sequence of weight matrices that allows us to obtain a nuisance parameter free bound. In particular, we choose the weight matrices such that $W_T(q)^{1/2}S(q)\sqrt{T}(\hat{\phi} - \phi)$ converges in distribution to a vector of standard normal random variables as the sample size increases. Using the singular value decomposition in (22), we let

$$W_T(q) = V_1 D_{11}^{-1} (U_1' \Lambda U_1)^{-1} D_{11}^{-1} V_1'. \quad (30)$$

Under this particular weight matrix the objective function $G_T^{(2)}(\theta, q, \mu; \hat{\phi})$ is identical to $G_T^{(1)}(\theta, q, \mu; \hat{\phi})$ in (23). But unlike in Section 4.1, we will now study its distribution directly to obtain a sharper cut-off level for a confidence set.

While formal analysis of $\bar{Q}_T^{(2)}(q)$ can be found in the Appendix, we subsequently provide a heuristic argument that abstracts from two complications: the potential rank reduction of $S(q)$ and establishing that our approximation is uniformly valid. Let $\Omega(q) = S(q)\hat{\Lambda}S(q)'$ and $W_T(q) = \Omega^{-1}(q)$. Since we abstract from uniformity issues, we drop the (q) -arguments from S and Ω . Partition $S' = [S_1', S_2']$, where S_2 is $1 \times m$. Denote the conforming partitions of Ω by $\Omega_{ij} = S_i \hat{\Lambda} S_j'$. Moreover, factorize $\hat{\Lambda} = LL'$ and let $\nu' = [\nu_1', \nu_2]$, where ν_2 is scalar. Then,

$$\begin{aligned} \bar{Q}_T^{(2)}(q) &= \min_{\nu \geq 0} \|S\sqrt{T}(\hat{\phi} - \phi) - M\nu\|_{\Omega^{-1}}^2 \\ &\leq \|S_1\sqrt{T}(\hat{\phi} - \phi)\|_{\Omega_{11}}^2 \\ &\quad + \min_{\nu_2 \geq 0} \|(S_2 - \Omega_{21}\Omega_{11}^{-1}S_1)\sqrt{T}(\hat{\phi} - \phi) - \nu_2\|_{(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}}^2 \\ &= \hat{\zeta}' P_{A_1} \hat{\zeta} + \min_{\nu_2 \geq 0} \|(A_2'(I - P_{A_1})\hat{\zeta} - \nu_2)\|_{(A_2'(I - P_{A_1})A_2)^{-1}}^2 \\ &\implies \chi_{k+r_2-1}^2 + \mathcal{I}\{Z \geq 0\}Z^2, \quad Z \sim N(0, 1). \end{aligned} \quad (31)$$

The inequality is obtained by setting $\nu_1 = 0$. The third expression on the right-hand side is obtained by defining $A_1 = L'S_1'$, $P_{A_1} = A_1(A_1'A_1)^{-1}A_1'$, and $\hat{\zeta} = L^{-1}\sqrt{T}(\hat{\phi} - \phi)$. Since $\hat{\zeta}$ converges in distribution to a $(k+r_2) \times 1$ vector of standard normals and P_{A_1} is a projection onto a $k+r_2-1$ -dimensional subspace, we obtain the convergence of $\hat{\zeta}'P_{A_1}\hat{\zeta}$ to a $\chi_{k+r_2-1}^2$. The solution to the minimization problem is

$$\hat{\nu}_2 = \mathcal{I}\{A_2'(I - P_{A_1})\hat{\zeta} \geq 0\} \frac{A_2'(I - P_{A_1})\hat{\zeta}}{\sqrt{A_2'(I - P_{A_1})A_2}},$$

which generates the term $\mathcal{I}\{Z \geq 0\}Z^2$ in the limit distribution. Since $P_{A_1}\hat{\zeta}$ and $A_2'(I - P_{A_1})\hat{\zeta}$

are asymptotically uncorrelated, Z is independent of the χ^2 term in the characterization of the asymptotic distribution.

The limit distribution in (31) arises commonly in multivariate generalizations of one-sided hypothesis problems, e.g. Perlman (1969), and its quantiles are used by Rosen (2008) to construct contour confidence sets for moment inequality models. It can be verified that

$$P\left\{\chi_{k+r_2-1}^2 + \mathcal{I}\{Z \geq 0\}Z^2 \leq c\right\} = \frac{1}{2}P\left\{\chi_{k+r_2}^2 \leq c\right\} + \frac{1}{2}P\left\{\chi_{k+r_2-1}^2 \leq c\right\},$$

which makes it straightforward to compute critical values for the construction of confidence intervals. In particular, if c_τ satisfies

$$\frac{1}{2}P\left\{\chi_{k+r_2}^2 \leq c_\tau\right\} + \frac{1}{2}P\left\{\chi_{k+r_2-1}^2 \leq c_\tau\right\} = 1 - \tau, \quad (32)$$

then the set

$$CS_{(2),\tau}^\theta = \left\{\theta \mid M_\theta \theta \geq 0 \text{ and } Q_T^{(2)}(\theta) \leq c_\tau\right\} \quad (33)$$

is an asymptotically valid confidence set. A formal statement is provided in the following theorem, proved in the Appendix.

Theorem 1 *Suppose that Assumption 1 is satisfied and $W_T(\tilde{q})$ is chosen according to (30). Then the confidence interval $CS_{(2),\tau}^\theta$ defined in (33) is an asymptotically valid confidence interval for θ :*

$$\liminf_T \inf_{\phi \in \mathcal{P}, \theta \in \Theta(\phi)} P_\phi\{\theta \in CS_{(2),\tau}^\theta\} \geq 1 - \tau.$$

As mentioned previously, the functions $G_T^{(1)}(\theta, q, \mu; \hat{\phi})$ in (23) and $G_T^{(2)}(\theta, q, \mu; \hat{\phi})$ in (24) are identical if $W_T(q)$ is chosen according to (30). The key difference between the corresponding confidence sets $CS_{(1),\tau}^\theta$ and $CS_{(2),\tau}^\theta$ is the critical value that is used to construct the contour sets. The union-of-identified-sets requires a critical value from a χ^2 distribution with m degrees of freedom, where m could be as large as $n(k + r_2)$. The critical value used for $CS_{(2),\tau}^\theta$ is slightly smaller than a critical value from a χ^2 distribution with $k + r_2$ degrees of freedom. Thus, our minimum-distance approach may lead to a reduction of the degrees of freedom by a factor of n , where n is the number of variables in the SVAR. Thus we obtain the following corollary.

Corollary 1 *Suppose Assumption 1 is satisfied. Then the union-of-identified-sets confidence set $CS_{(1),\tau}^\theta \subset CS_{(2),\tau}^\theta$.*

The computations to obtain the confidence set $CS_{(2),\tau}^\theta$ can be implemented using the algorithm described in Section 4.1 by replacing $Q_T^{(1)}(\theta)$ with $Q_T^{(2)}(\theta)$ in (25) using the critical value defined in (32).

4.3 Discussion

We will subsequently provide a brief discussion of extensions and limitations of the results obtained in Sections 4.1 and 4.2.

Levels versus Growth Rates. Recall the example of an inflation and output growth VAR discussed previously. We assumed that the $n = 2$ dimensional vector y_t is composed of log differences of aggregate prices and output. Suppose that the sign restrictions are specified as follows: in response to a positive demand shock the log level of prices and output will be non-negative in periods 0, 1, and 2. We can easily handle this case by defining $R_h^v = \sum_{\tau=0}^h C_\tau \Sigma_{tr}$, use $R_{0,H}^v$ to stack R_h^v , $h = 0, \dots, H$, and let $\phi = \text{vec}((R_{0,H}^v)')$. Moreover, suppose $H = 2$ and θ is defined as the inflation rate between periods 1 and 2. The inflation response can be obtained as the difference of the level response by defining

$$\theta = (S_\theta \otimes q')\phi \quad \text{with} \quad S_\theta = \begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 0 \end{bmatrix}.$$

Sign-Restrictions Combined with Zero Restrictions. In our framework it is straightforward to sharpen the identified set by combining sign-restrictions with more traditional exclusion restrictions. Suppose that $n = 4$ and y_t is composed of output, inflation, interest rates, and real money balances. A common exclusion restriction, e.g. Boivin and Giannoni (2006), is to assume that private sector variables do not respond to monetary policy shocks contemporaneously. In our example, we could assume that output and inflation do not respond to a monetary policy shock. This assumption generates two linear restrictions for the unit vector q . Since $n = 4$, it leaves q underdetermined. We can then add further restrictions by assuming that interest rates rise, and prices and money balances fall in response to a contractionary policy shock. As before, the inequality restrictions can be represented by $M_\theta \theta \geq 0$ and $S_R(q)\phi \geq 0$. To impose the exclusion restriction we could add another constraint to (8) of the form $(S_{eq} \otimes q')\phi = 0$, where S_{eq} is a selection matrix that chooses the relevant elements to ϕ that are needed to construct the contemporaneous response of output. If the zero restrictions are imposed on long-run effects of the structural shock as in Blanchard and Quah (1989), then ϕ has to include $R_\infty^v = \sum_{h=0}^\infty R_h^v$. As we will illustrate in

Section 6, rather than adding equality constraints to the objective function, it is sometimes easier to constrain certain elements of the vector q directly and then minimize $G_T(\theta, q, \mu; \hat{\phi})$ with respect to q over a restricted unit hypersphere.

Identifying Multiple Shocks. Some authors use sign-restricted SVARs to identify multiple shocks simultaneously. For instance, Peersman (2005) considers a $n = 4$ dimensional VAR, composed of oil price inflation, output growth, consumer price inflation, and nominal interest rates. He uses sign restrictions to identify an oil price shock, aggregate demand and supply shocks, and a monetary policy shock. To identify n shocks the unit vector q has to be replaced by an orthogonal matrix and the restrictions will take the form

$$S_\theta(\Omega)\phi = \theta \quad \text{and} \quad S_R(\Omega)\phi \geq 0$$

for suitably defined functions $S_\theta(\Omega)$ and $S_R(\Omega)$. While all our results easily generalize to multiple shocks (just replace q by Ω in the equations in Sections 4.1 and 4.2), the implementation becomes more tedious. The objective functions $G_T^{(1)}(\theta, \Omega, \mu; \hat{\phi})$ and $G_T^{(2)}(\theta, \Omega, \mu; \hat{\phi})$ have now to be minimized over the domain of Ω rather than the unit hypersphere. Such a minimization could be implemented by expressing Ω as a product of Given matrices. For $n = 3$ this product takes the form

$$\Omega(\varphi_1, \varphi_2, \varphi_3) = \begin{bmatrix} \cos \varphi_1 & \sin \varphi_1 & 0 \\ -\sin \varphi_1 & \cos \varphi_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \varphi_2 & 0 & \sin \varphi_2 \\ 0 & 1 & 0 \\ -\sin \varphi_2 & 0 & \cos \varphi_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_3 & \sin \varphi_3 \\ 0 & -\sin \varphi_3 & \cos \varphi_3 \end{bmatrix}.$$

Notice, however, that $\Omega(\varphi_1, \varphi_2, \varphi_3)$ does not generate all orthogonal matrices of dimension 3. It only generates matrices with determinant one.

Variance Decompositions and Dynamic Correlations. Faust (1998) was not interested the impulse responses to a monetary policy shock. Instead his goal was to measure the fraction of the variance of output, explained by monetary policy shocks. Canova and Nicolò (2002) did not restrict the sign of impulse responses. Instead they restricted the sign of dynamic correlations generated by structural shocks to attain partial identification. Both variance decompositions and dynamic correlations involve objects of the form

$$\sum_{h=0}^H C_h \Sigma_{tr} q q' \Sigma_{tr}' C_{h+j}'$$

where H is potentially infinite. Handling variance decompositions and dynamic correlations requires a non-trivial extension of our framework, because our analysis exploited the fact that we could express the restrictions $S_\theta(q)\phi = \theta$ and $S_R(q)\phi \geq 0$ as linear functions of ϕ .

Nonstationary VARs. Some authors, e.g. Uhlig (2005), specify the VAR in terms of variables that exhibit (near) nonstationary dynamics, such as the log level of GDP, or the log levels of consumer or commodity price indices. We assumed that $\hat{\phi}$ has a Gaussian limit distribution. This assumption is violated in VARs with non-stationary endogenous variables, see Phillips (1996). An extension of our analysis to VARs with unit roots or cointegration restrictions is beyond the scope of this paper.

Error Bands versus Point Estimates. In addition to error bands authors in practice often report median or mean response functions for sign-restricted VARs. While these median or mean responses are well defined in a Bayesian framework – as mean or median of the posterior distribution, which asymptotically concentrates on the identified set – they are not meaningful objects in a frequentist framework. If θ is scalar and hence interval identified one could construct a point estimator from a minimax decision problem:

$$\hat{\theta} = \operatorname{argmin}_{\tilde{\theta} \in \Theta(\phi)} \max_{\theta \in \Theta(\phi)} L(\tilde{\theta} - \theta)$$

If the error loss function is symmetric and $\Theta(\phi)$ is an interval, then it is optimal to choose the mid point of the interval. Thus, in some of our illustrations we will plot the mid-point of $\Theta(\hat{\phi})$. A general analysis of minimax decision problems in interval identified models is provided by Song (2009).

5 Monte Carlo Illustrations

In this section we provide two Monte Carlo illustrations of our proposed confidence sets for impulse responses from sign-restricted SVARs. The first illustration is a continuation of the simple example considered in the preceding sections. For the second illustration we introduce some autoregressive dynamics to examine the effect of serial correlation on the estimation of the reduced form parameters as well as the impulse responses. The simulation designs are obtained by fitting a VAR(0) to data on U.S. inflation and GDP growth (Section 5.1) and fitting VAR(1)'s to data inflation and either output growth or

linearly detrended log GDP (Section 5.2). We also provide a comparison between frequentist confidence sets and Bayesian credible sets.

5.1 The Simple Example, Continued

In this section we complete the simple example discussed in the previous sections with a numerical illustration. The parameterization of the data generating process is provided in Table 1 in the column labelled *Design 1*. The identified set for the response of $y_{1,t}$ to $\epsilon_{1,t}$ is $\Theta(\phi_0) = [0, 0.578]$. We consider 90% credible sets and confidence sets, that is, $\tau = 0.1$. The critical values used in the construction of $CS_{(1),\tau}^\theta$ and $CS_{(2),\tau}^\theta$ are 6.25 and 3.82, respectively. The Monte Carlo simulations are implemented based on the following algorithm.

Algorithm: We repeat the following steps n_{sim} times:

1. Generate a sample of size T from the data generating process.
2. We use an improper prior $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$ and a prior for q that is uniform over the one-dimensional subspaces of \mathbb{R}^2 spanned by q , restricted to the set of q 's that are consistent with the sign restrictions. Based on this prior we generate draws from the posterior distribution of Σ and q (see Section 3), which we convert into the desired impulse response. From these draws we construct a Bayesian 90% credible set.
3. We let $\hat{\phi}$ be the maximum likelihood estimator of ϕ and compute $\Theta(\hat{\phi})$.
4. We use a parametric bootstrap to approximate the sampling distribution of $\hat{\phi}$, that is, conditional on the estimate $\hat{\Sigma}_{tr}$ we generate bootstrap samples from $y_t^* = \hat{\Sigma}_{tr} v_t^*$ where $v_t^* \sim N(0, I_2)$. For each bootstrap sample we compute $\hat{\phi}^*$ and estimate $\hat{\Lambda}$ as the covariance matrix of $\hat{\phi}^*$ across bootstrap samples.
5. We compute $CS_{(1),\tau}^\theta$ and $CS_{(2),\tau}^\theta$ using the algorithm described in Section 4.1.

The results for $n_{sim} = 10$ and $T = 100$ are plotted in Figure 1. The x -axis denotes the iteration of the simulation algorithm. The lower bound of $\Theta(\phi_0)$, $\Theta(\hat{\phi})$, and of the frequentist confidence intervals is zero. Hence, we focus our discussion on the upper bound. We re-ordered the simulations according to the upper bound of $\Theta(\hat{\phi})$. In about half of the simulations the upper bound of $\Theta(\hat{\phi})$ exceeds that of $\Theta(\phi_0)$. In 8 out of 10 repetitions

each value $\theta \in \Theta(\phi)$ is contained in the frequentist confidence set $CS_{(2),\tau}^\theta$. By construction, $CS_{(2),\tau}^\theta \subset CS_{(1),\tau}^\theta$. The upper bound of the Bayesian credible sets essentially coincides with the upper bound of $\Theta(\hat{\phi})$. This is consistent with the derivations at the end of Section 3, in which we show for the simple example that the posterior density is increasing in θ , and conditional on the reduced form parameters, peaks at the upper bound of $\Theta(\phi)$. The lower bounds of the Bayesian intervals are strictly greater than zero, which means that for our relatively large sample the Bayesian credible intervals lie inside $\Theta(\hat{\phi})$, a point emphasized in Moon and Schorfheide (2009).

If we increase the number of repetitions to $n_{sim} = 1000$, then the boundary point of the identified set is covered by the frequentist interval in 93% of the repetitions and only in 40% of the repetitions by the Bayesian interval. Detailed results for the frequentist confidence interval are summarized in Table 2. At a sample size of $T = 5,000$ the actual coverage probability of CS^ϕ equals the nominal coverage probability of 90%. Since the nuisance-parameter free critical values that we use to construct the impulse response confidence intervals are based on an upper bound of the criterion function, the resulting confidence intervals are conservative. For $T = 5,000$, the actual coverage probability of $CS_{(2)}^\theta$ for the upper bound of $\Theta(\phi)$ is 95% instead of 90%. For sample smaller sizes of $T = 100$ and $T = 500$ the coverage probability of CS^ϕ is 82% and 94% and thus deviates from the desired nominal size. Since $CS_{(2)}^\theta$ is conservative, a fairly low coverage probability for ϕ at $T = 100$ still translates into a confidence interval for θ that exceeds the nominal coverage probability. The average length of $CS_{(2)}^\theta$ shrinks from 0.64 for $T = 100$ to 0.59 for $T = 5,000$. Thus, as the sample size increases, the length of the confidence interval approaches the length of the identified set. As a comparison, the 90% Bayesian credible sets have an average length of 0.5, which is less than the length of the identified set. From a frequentist perspective, the Bayesian intervals have a coverage probability of about 45%.

5.2 A Bivariate VAR(1)

We now add first-order autoregressive terms to the simulation design to introduce persistence in the endogenous variables. Our choices for Φ_1 and Σ are summarized in Table 1 under the headings *Design 2*, *Design 3*, and *Design 4*. The designs differ with respect to the persistence of the vector autoregressive process. *Design 2* is the least persistent. The eigenvalues of Φ_1 are 0.871 and 0.231. *Design 4* is the most persistent with eigenvalues 0.955

and 0.498. We focus on responses at horizon $h = 1$, which can be obtained from $R_1^u = \Phi \Sigma_{tr}$. The structural parameter of interest, θ , is defined as $\partial y_{1,t+1} / \partial \epsilon_{1,t}$ and we impose the sign restrictions that both θ as well as $\partial y_{2,t+1} / \partial \epsilon_{1,t}$ are non-negative. To simplify the computations, in particular the minimization of the objective function $G_T^{(2)}(\theta, q, \mu; \hat{\phi})$, we do not impose sign restrictions on the responses a impact or at horizons greater than $h = 1$. We use the algorithm described in Section 5.1 to implement the Monte Carlo simulation. A simplified representation for the objective function $Q_T^{(2)}(\theta)$ can be found in the Appendix.

The simulation results for confidence interval with a nominal coverage of 90% are summarized in Table 2. We consider sample sizes $T = 100$ and $T = 500$. As for the VAR(0), the confidence intervals for the impulse response are generally conservative. The actual coverage probabilities range from 92% to 98% and reflect the somewhat distorted coverage probabilities of CS^ϕ , which range from 83% to 95%. However, due to the conservativeness of the critical values that are used to construct $CS_{(2)}^\theta$, its actual coverage probability never falls below 90%. The average length of the confidence sets under *Design 2* and *Design 3* drops by about 10% as the sample size is increased from 100 to 500 observations. For *Design 4* the reduction is slightly larger than 20%. For $T = 500$ the confidence intervals are about 10% longer than the identified sets.

6 Empirical Illustration

We now apply the previously developed methods to a four variable VAR. The vector of observables consists of real GDP, inflation, a nominal interest rate, and real money balances. We will consider two partial identification schemes for monetary policy shocks and compare the typically computed Bayesian credible sets with the proposed frequentist error bands.

6.1 Data

Unless otherwise noted, the data are obtained from the FRED2 database maintained by the Federal Reserve Bank of St. Louis. Per capita output is defined as real GDP (GDPC96) divided by civilian non-institutionalized population (CNP16OV). The population series is provided at a monthly frequency and converted to quarterly frequency by simple averaging. We take the natural log of per capita output and extract a deterministic trend by OLS

regression over the period 1959:I to 2006:IV. The deviations from the linear trend are scaled by 100 to convert them into percentages. Inflation is defined as the log difference of the GDP deflator (GDPDEF), scaled by 400 to obtain annualized percentage rates. Our measure of nominal interest rates corresponds to the Federal Funds Rate (FEDFUNDS), which is provided at monthly frequency and converted to quarterly frequency by simple averaging. We use the sweep-adjusted M2S series provided by Cynamon, Dutkowsky and Jones (2006). This series is recorded at monthly frequency without seasonal adjustments. The EVIEWS default version of the X12 filter is applied to remove seasonal variation. The M2S series is divided by quarterly nominal GDP to obtain inverse velocity. We then remove a linear trend from log inverse velocity and scale the deviations from trend by 100. Since our VAR is expressed in terms of real money balances rather we take the sum of log inverse velocity and real GDP. Finally, we restrict our quarterly observations to the period from 1965:I to 2005:I. All VAR's are estimated with $p = 2$ lags.

6.2 Exclusion and Sign Restrictions

A commonly used identification assumption for monetary policy shocks is that private sector variables such as output and inflation cannot respond within the period, see for instance Boivin and Giannoni (2006). Since the initial impact of the monetary policy shock is given by $\Sigma_{tr}q$ and output and inflation appear before interest rates and real money balances in the vector y_t , the identification condition implies that the first two elements of the vector q have to be equal to zero. Thus, we can express $q = [0, 0, \cos \varphi, \sin \varphi]'$, where $\varphi \in [0, 2\pi]$. In addition to the exclusion restrictions, we consider the following sign restrictions to bound the responses to a contractionary monetary policy shock: (i) the inflation response in period $h = 1$ is non-positive; (ii) the interest rate responses for $h = 0$ and $h = 1$ are non-negative; (iii) real money balances don't rise upon impact of the monetary policy shock and stay below steady state in period $h = 1$.

The Bayesian analysis is conducted under an improper prior $p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2}$, truncated according to (13). Posterior inference is implemented with the acceptance sampler described in Uhlig (2005). Posterior draws of Φ , Σ , and q are then converted into impulse responses and pointwise 90% credible intervals are generated by constructing the shortest intervals that contain 90% of the posterior draws.

To implement the frequentist inference we make the following modifications to the algorithms described in Section 4.1. First, $\min_{\|q\|=1, \mu \geq 0} G_T(\tilde{\theta}, q, \mu; \hat{\phi})$ is computed in two steps. (i) We use a standard quadratic programming routine to construct the concentrated objective function

$$\underline{G}_T(\tilde{\theta}, q; \hat{\phi}) = \min_{\mu \geq 0} G_T(\tilde{\theta}, q, \mu; \hat{\phi}).$$

Since the domain of q is a (restricted) unit hypersphere, we express the unit length vector in spherical rather than Cartesian coordinates, as indicated above. (ii) We then use a constrained optimization routine to minimize $\underline{G}_T(\tilde{\theta}, q(\varphi))$. Notice that the minimum does not need to be unique. Moreover, it suffices to check whether the objective function is less than the critical value, which means that the minimization can be terminated as soon as $\underline{G}_T(\tilde{\theta}, q(\varphi); \hat{\phi})$ is less than the critical value.

Second, rather than evaluating $Q_T(\theta)$ over a fixed grid of θ values, we use the boundaries of the Bayesian credible interval for θ as the starting points. If $Q_T(\theta)$ is less (greater) than the critical value for the upper bound of the Bayesian credible interval, we increase (decrease) θ until we find a value for which the objective function exceeds (falls below) the critical value. This yields the upper bound of the frequentist interval. The lower bound is found accordingly, by starting from the lower bound of the Bayesian interval.

Results are depicted in Figures 2. Due to the two zero restrictions, the identified sets for the (pointwise) impulse responses, computed conditional on the maximum likelihood estimates of the reduced form parameters, are fairly small. In fact, at long horizons the impulse responses are essentially point identified. The Bayesian error bands extend well beyond the identified set, indicating that posterior uncertainty about Φ and Σ dominates uncertainty about θ conditional on Φ and Σ . The frequentist error bands tend to be wider than the Bayesian error bands at short horizons, when the identified set is relatively wide. This observation is qualitatively consistent with the large sample results obtained in Moon and Schorfheide (2009).

According to the posterior mean (not shown in the Figure) interest rates rise by about 40 basis points (bp) in response to a one-standard deviation contractionary monetary policy shock. Output falls for a year until it is roughly 40 bp below trend and then rises again. The inflation response is more persistent and reaches a trough after 4 years. At that point it is about 30 bp below its steady state level. Real money balances drop by 1 percent upon impact and then revert back to steady state. While according to the Bayesian error bands

the responses of output and inflation are negative in the first three years, the signs are ambiguous in view of the frequentist confidence intervals.

6.3 Pure Sign Restrictions

We now proceed by removing the exclusion restrictions used in the previous analysis and imposing only sign restrictions. At horizons $h = 0$ and $h = 1$ the inflation and real money balance responses are non-positive and the interest rate responses to a contractionary monetary policy shock are non-negative. Figure 3 contrasts the identified sets (conditional on $\hat{\Phi}$ and $\hat{\Sigma}$) obtained under the pure sign restriction (*Sign*) approach and the combination of exclusion and sign restrictions (*ExclSign*). Under the less restrictive identification assumptions, the bounds for the impulse responses are substantially wider, in particular for horizons less than 4 years. Since the *ExclSign* identification restricts the unit length vector q to $[0, 0, \cos \varphi, \sin \varphi]$, whereas the *Sign* Scheme leaves the domain of q unrestricted, the resulting identified sets are nested.² While the *ExclSign* identified sets imply that output falls below trend in the first four years after a contractionary monetary policy shock, the sign of the output response becomes ambiguous under the *Sign* identification scheme.

Figure 4 depicts frequentist and Bayesian error bands for the *Sign* approach. For convenience, we also overlay the identified sets. Except for horizons greater than six years, the Bayesian 90% credible intervals and the boundaries of the identified sets more or less coincide. The frequentist error bands are two to three times as wide as the Bayesian error bands. Since the identified sets are much wider than under the *ExclSign* approach, the difference between Bayesian credible sets and frequentist confidence sets is now much more pronounced. As explained in detail in Moon and Schorfheide (2009), in a large sample the Bayesian intervals lie inside the identified set because in the limit the entire probability mass is concentrated on the identified set and a 90% credible interval is always a subset of the support of the posterior distribution. The frequentist interval, on the other hand, has to extend beyond the boundaries of the identified set, because it has to have, say, 90% coverage probability for every element of the identified set, including the boundary points.

²The fact that for the $h = 0$ response the *Sign* identified set excludes zero is due to a small numerical inaccuracy.

7 Conclusion

This paper develops a method to construct error bands for impulse responses in a VAR that is identified based on sign-restrictions. The error bands that have been reported in the literature thus far, were only meaningful from a Bayesian perspective. We show that our error bands delimit valid frequentist confidence intervals. The impulse response confidence intervals are constructed through a point-wise testing procedure, which is a technique that is widely used in models that are either weakly or only partially identified. While our intervals are conservative, the proposed procedure is easy to use in practice, since it relies on asymptotic critical values that are nuisance parameter free.

As a by-product we also provide a procedure to compute the identified set of impulse responses conditional on the reduced form parameters. Since in a Bayesian analysis, the prior distribution of the impulse response functions conditional on the reduced form parameters does not get updated, it is informative to report the identified set conditional on some estimate, say the posterior mean, of Φ and Σ so that the audience can judge whether the conditional prior distribution is highly concentrated in a particular area of the identified set.

References

- Anderson, T.W. and Herman 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, Donald and Patrick Guggenberger (2009): “Validity of Subsampling and ‘Plug-in Asymptotics’ Inference for Parameters defined by Moment Inequalities,” *Econometric Theory*, **25**, 669-709.
- Baumeister, Christiane and Geert Peersman (2008): “Time-Varying Effects of Oil Supply Shocks on the U.S. Economy,” *Manuscript*, Ghent University.
- Blanchard, Olivier, and Danny Quah (1989): “The Dynamic Effects of Aggregate Demand and Supply Disturbances,” *American Economic Review*, **79**, 655-673.

- Boivin, Jean, and Marc Giannoni (2006): "Has Monetary Policy Become More Effective?" *Review of Economic and Statistics*, **88**, 445-462.
- Canova, Fabio and Gianni De Nicolò (2002): "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7," *Journal of Monetary Economics*, **49**, 1131-59.
- Chernozhukov, Victor, Han Hong, and Elie Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, **75**, 1243-1284.
- Cynamon, Barry, Donald Dutkowsky, and Barry Jones (2006): "Redefining the Monetary Aggregates: A Clean Sweep," *Eastern Economic Journal*, **32**, 661-672.
- Dedola, Luca and Stefano Neri (2007): "What Does a Technology Shock Do? A VAR Analysis with Model-Based Sign Restrictions," *Journal of Monetary Economics*, **54**, 512-549.
- Dufour, Jean-Marie (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, **65**, 1365-1387.
- Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Conference Series on Public Policy*, **49**, 207-244.
- James, A. T. (1954): "Normal Multivariate Analysis and the Orthogonal Group," *Annals of Mathematical Statistics*, **25**, 40-75.
- Kadane, Joseph B. (1974): "The Role of Identification in Bayesian Theory," in S.E. Fienberg and A. Zellner (eds.) *Studies in Bayesian Econometrics and Statistics*, 175-191. North Holland, Amsterdam.
- Kilian, Lutz and Dan Murphy (2009): "Why Agnostic Sign Restrictions are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Manuscript*, University of Michigan.
- Manski, Charles (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.
- Manski, Charles and Elie Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, **70**, 519-547.

- Moon, Hyungsik Roger and Frank Schorfheide (2009): "Bayesian and Frequentist Inference in Partially Identified Models," *NBER Working Paper*, **14882**.
- Mountford, Andrew and Harald Uhlig (2008): "What are the Effects of Fiscal Policy Shocks?," *NBER Working Paper*, **14551**.
- Pappa, Evi (2009): "The Effects of Fiscal Shocks on Employment and the Real Wage," *International Economic Review*, **50**, 217-244.
- Perlman, Michael (1969): "One-Sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, **40**, 549-567.
- Peersman, Geert and Roland Straub (2009): "Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR," *International Economic Review*, **50**, 727-750.
- Phillips, Peter (1998): "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VAR's," *Journal of Econometrics*, **83**, 21-56.
- Poirier, Dale (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, **14**, 483-509.
- Rosen, Adam (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, **146**, 107-117.
- Song, Kyungchul (2009): "Point Decisions for Interval-Identified Parameters," *Manuscript*, University of Pennsylvania.
- Staiger, Douglas and James Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, **65**, 557-586.
- Uhlig, Harald (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, **52**, 381-419.

A Proofs

Proof of Lemma 1: (i) Show that $CS_{U,\tau}^\theta \subseteq CS_{(1),\tau}^\theta$. Suppose $\theta_* \in CS_{U,\tau}^\theta$. Thus, there exists a $\phi_* \in CS_\tau^\phi$ such that $\theta_* \in \Theta(\phi_*)$. So, $T\|\hat{\phi} - \phi_*\|_{\hat{\Lambda}^{-1}} \leq c_\tau(\chi_m^2)$. This implies that there exist a q_* with $\|q_*\| = 1$ and a $\mu_* \geq 0$ such that $S(q_*)\phi_* = [\theta'_*, \mu'_*]'$. In turn, $G_T^{(1)}(\theta_*, q_*, \mu_*; \hat{\phi}) \leq c_\tau(\chi_m^2)$ and $\min_{\|q\| \leq 1, \mu \geq 0} G_T^{(1)}(\theta_*, q, \mu; \hat{\phi}) \leq c_\tau(\chi_m^2)$. This shows that $\theta_* \in CS_{(1),\tau}^\theta$.

(ii) Show that $CS_{(1),\tau}^\theta \subseteq CS_{U,\tau}^\theta$. Suppose that $\theta_* \in CS_{(1),\tau}^\theta$ and $\theta_* \notin CS_{U,\tau}^\theta$. Thus, $\theta_* \in (CS_{U,\tau}^\theta)^c = \bigcap_{\phi \in CS_\tau^\phi} (\Theta(\phi))^c$. Hence, for all $\phi \in CS_\tau^\phi$, $\theta_* \notin \Theta(\phi)$. In turn, for any q with $\|q\| = 1$ and any $\phi \in CS_\tau^\phi$, either $(S_\theta \otimes q')\phi \neq \theta_*$ or $(S_R \otimes q')\phi < 0$. Thus, to satisfy the constraints in (19) for θ_* it has to be that (the constrained) $\left(\operatorname{argmin}_{\phi \in \mathcal{P}} T\|\hat{\phi} - \phi\|_{\hat{\Lambda}^{-1}}^2\right) \notin CS_\tau^\phi$, which implies that $\min_{\|q\| \leq 1, \mu \geq 0} G_T^{(1)}(\theta_*, q, \mu; \hat{\phi}) > c_\tau(\chi_m^2)$. Therefore, $\theta_* \notin CS_{(1),\tau}^\theta$, which contradicts our initial assumption. \square

Proof of Theorem 1: We will distinguish two cases: (i) $S(q)$ has full row rank $k + r_2$; (ii) $S(q)$ has reduced row rank $l < k + r_2$. If no ambiguity arises, we shall omit the (q) argument from matrices that depend on q .

Case (i): In the main text we started from

$$\min_{\|q\|=1, \mu \geq 0} G_T^{(2)}(\tilde{\theta}, q, \mu; \hat{\phi}) \leq \min_{\nu \geq 0} \left\| S(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) - M_\nu \nu \right\|_{W_T(\tilde{q})}^2 = \bar{Q}_T^{(2)}(\tilde{q}).$$

We then derived the bound

$$\begin{aligned} \bar{Q}_T^{(2)}(q) &\leq \|S_1\sqrt{T}(\hat{\phi} - \phi)\|_{\Omega_{11}}^2 + \min_{\nu_2 \geq 0} \|(S_2 - \Omega_{21}\Omega_{11}^{-1}S_1)\sqrt{T}(\hat{\phi} - \phi) - \nu_2\|_{(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}}^2 \\ &= \hat{\zeta}'P_{A_1}\hat{\zeta} + \min_{\nu_2 \geq 0} \|(A_2'(I - P_{A_1})\hat{\zeta} - \nu_2)\|_{A_2'(I - P_{A_1})A_2}^2, \end{aligned}$$

where $\Omega = S\hat{\Lambda}S'$, $S' = [S'_1, S'_2]$, S_2 is $1 \times m$, $\Omega_{ij} = S_i\hat{\Lambda}S'_j$, $\hat{\Lambda} = LL'$, $\hat{\zeta} = L^{-1}\sqrt{T}(\hat{\phi} - \phi)$, $A_1 = L'S'_1$, $A_2 = L'S'_2$, and $P_{A_1} = A_1(A_1'A_1)^{-1}A_1'$. Here P_{A_1} is the matrix that projects onto the space spanned by the $k + r_2 - 1$ columns of A_1 . Let $V_{A_1}D_{k+r_2-1}V_{A_1}'$ be a singular value decomposition of P_{A_1} . Here V_{A_1} is a $m \times m$ orthonormal matrix and D_{k+r_2-1} is a $m \times m$ diagonal matrix. The first $k + r_2 - 1$ diagonal elements of D_{k+r_2-1} are equal to one and the last diagonal element is equal to zero. Using this factorization, we can write:

$$\bar{Q}_T^{(2)}(q) \leq \hat{\zeta}'V_{A_1}D_{k+r_2-1}V_{A_1}'\hat{\zeta} + \min_{\nu_2 \geq 0} \|(A_2'(I - P_{A_1})\hat{\zeta} - \nu_2)\|_{A_2'(I - P_{A_1})A_2}^2. \quad (34)$$

Case (ii): We begin with the singular value decomposition $S = VDU' = V_1 D_{11} U_1'$, where $V = [V_1, V_2]$ and $U = [U_1, U_2]$ are orthonormal matrices and D_{11} is a diagonal $l \times l$ matrix with the non-zero singular values of S . The dimensions of V_1 and U_1' are $(k + r_2) \times l$ and $l \times m$, respectively. As before, we factorize $\hat{\Lambda} = LL'$ and let $\hat{\zeta} = \sqrt{T}L^{-1}(\hat{\phi} - \phi)$. We define $B = L'U_1$, and $P_B = B(B'B)^{-1}B'$. Then

$$\begin{aligned} \bar{Q}_T^{(2)}(q) &= \min_{\nu \geq 0} \|V_1 D_{11} U_1' L \hat{\zeta} - M_\nu \nu\|_{V_1 D_{11}^{-1} (U_1' L L' U_1)^{-1} D_{11}^{-1} V_1'}^2 \\ &\leq \|V_1 D_{11} U_1' L \hat{\zeta}\|_{V_1 D_{11}^{-1} (U_1' L L' U_1)^{-1} D_{11}^{-1} V_1'}^2 \\ &= \hat{\zeta}' B (B'B)^{-1} B' \hat{\zeta} \\ &= \hat{\zeta}' P_B \hat{\zeta}. \end{aligned}$$

Here P_B is the matrix that projects $\hat{\zeta}$ onto the space spanned by the $l < k + r_2$ columns of B . Let $V_B D_{l(B)} V_B'$ be a singular value decomposition of P_B , using the same notation as above. We can bound P_B as follows:

$$P_B = V_B D_{l(B)} V_B' \leq V_B D_{k+r_2-1} V_B'.$$

Thus,

$$\bar{Q}_T^{(2)}(q) \leq \hat{\zeta}' V_B D_{k+r_2-1} V_B' \hat{\zeta}. \quad (35)$$

Combining (34) and (35) yields

$$\bar{Q}_T^{(2)}(q) \leq \begin{cases} \hat{\zeta}' V_{A_1} D_{k+r_2-1} V_{A_1}' \hat{\zeta} & \text{if } \text{rank}(S(q)) = k + r_2 \\ + \min_{\nu_2 \geq 0} \|A_2'(I - P_{A_1})\hat{\zeta} - \nu_2\|_{A_2'(I - P_{A_1})A_2} & \\ \hat{\zeta}' V_B D_{k+r_2-1} V_B' \hat{\zeta} & \text{otherwise} \end{cases}$$

Suppose we now construct a probability space with random variables ζ and ξ_T such that $\hat{\zeta}_T \sim \zeta + \xi_T$, where $\zeta \sim N(0, I)$ and $\xi_T = o_p(1)$. Then,

$$\zeta' V_{A_1} D_{k+r_2-1} V_{A_1}' \zeta \sim \chi_{k+r_2-1}^2 \quad \text{and} \quad \zeta' V_B D_{k+r_2-1} V_B' \zeta \sim \chi_{k+r_2-1}^2,$$

regardless of A_1 and B . Moreover, since V_B (V_{A_1}) is orthogonal, we deduce that

$$\begin{aligned} \|\xi_T' V_B D_{k+r_2-1} V_B \zeta\|^2 &\leq \|\xi_T\| \cdot \|V_B D_{k+r_2-1} V_B \zeta\| \leq \|\xi_T\| \cdot \|\zeta\| = o_p(1) \\ \|\xi_T' V_B D_{k+r_2-1} V_B \xi_T\|^2 &\leq \|\xi_T\| \cdot \|V_B D_{k+r_2-1} V_B \xi_T\| \leq \|\xi_T\|^2 = o_p(1). \end{aligned}$$

The $o_p(1)$ bounds do not depend on V_B (V_{A_1}). A similar argument can be used to establish that the convergence

$$\min_{\nu_2 \geq 0} \|A'_2(I - P_{A_1})\hat{\zeta} - \nu_2\|_{(A'_2(I - P_{A_1})A_2)^{-1}} \implies \mathcal{I}\{Z \geq 0\}Z^2 \quad Z \sim N(0, 1)$$

is uniform in A_1 and A_2 . Since $P_{A_1}\hat{\zeta}$ and $A'_2(I - P_{A_1})\hat{\zeta}$ are asymptotically independent, the statement of the theorem follows. \square

B Derivations for Bivariate VAR(1)

We consider a bivariate ($n = 2$) VAR(1) of the form $y_t = \Phi y_{t-1} + u_t$ and focus on the response at horizon $h = 1$, which will be constructed from $R_1^v = \Phi \Sigma_t r$. Hence, let $\phi = \text{vec}((R_1^v)')$. The object of interest is $\theta = \partial y_{1,t+1} / \partial \epsilon_{1,t}$ and we impose the sign restriction that both θ as well as $\partial y_{2,t+1} / \partial \epsilon_{1,t}$ are non-negative. Let $q = [q_1, q_2]'$. Then

$$\begin{aligned} S_R(q) &= \begin{bmatrix} 0 & 0 & q_1 & q_2 \end{bmatrix} \\ S_\theta(q) &= \begin{bmatrix} q_1 & q_2 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Notice that in this example $S(q) = [S_\theta'(q), S_R'(q)]'$ is of full row rank for all values of q . For any positive-definite weight matrix W the identified set is given by

$$\Theta(\phi) = \left\{ \theta \geq 0 \mid \min_{\|q\|=1, \mu \geq 0} G(\theta, q, \mu; \phi) = 0 \right\},$$

where

$$G(\theta, q, \mu; \phi) = \left\| \begin{bmatrix} S_\theta(q)\phi - \theta \\ S_R(q)\phi - \mu \end{bmatrix} \right\|_W^2.$$

Since $S(q)$ is always of full row rank, there is no need to conduct a singular value decomposition. Let $\hat{\Lambda}$ be the bootstrap estimator of the covariance matrix of $\sqrt{T}(\hat{\phi} - \phi)$. Then our inference is based on the weight matrix $W(q) = (S(q)\hat{\Lambda}S'(q))^{-1}$ and we let

$$G_T(\theta, q, \mu; \hat{\phi}) = T \left\| \begin{bmatrix} S_\theta(q)\hat{\phi} - \theta \\ S_R(q)\hat{\phi} - \mu \end{bmatrix} \right\|_{W(q)}^2.$$

In order to economize on notation, we will drop the q argument from the matrices S_θ , S_R , S , and W .

In this simple illustrative example, it is fairly straightforward to solve $\min_{\mu \geq 0} G_T(\theta, q, \mu; \hat{\phi})$ analytically. Decompose

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} 1 & W_{12}W_{22}^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} W_{11.22} & 0 \\ 0 & W_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ W_{22}^{-1}W_{21} & 1 \end{bmatrix},$$

where

$$W_{11.22} = W_{11} - W_{12}W_{22}^{-1}W_{21}.$$

Thus, we can rewrite

$$G_T(\theta, q, \mu; \phi) = TW_{11.22}(\theta - S_\theta \hat{\phi})^2 + TW_{22} \left(\mu - \left[S_R \hat{\phi} - W_{12} W_{22}^{-1} (\theta - S_\theta \hat{\phi}) \right] \right)^2.$$

Now let

$$\begin{aligned} \hat{\mu}(\theta, q) &= \operatorname{argmin}_{\mu \geq 0} G_T(\theta, q, \mu; \phi) \\ &= \begin{cases} S_R \hat{\phi} - W_{12} W_{22}^{-1} (\theta - S_\theta \hat{\phi}) & \text{if } S_R \hat{\phi} - W_{12} W_{22}^{-1} (\theta - S_\theta \hat{\phi}) \geq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Thus,

$$\begin{aligned} &G_T(\theta, q, \hat{\mu}(\theta, q); \phi) \\ &= TW_{11.22}(\theta - S_\theta \hat{\phi})^2 \\ &\quad + \begin{cases} 0 & \text{if } S_R \hat{\phi} - W_{12} W_{22}^{-1} (\theta - S_\theta \hat{\phi}) \geq 0 \\ TW_{22} \left(S_R \hat{\phi} - W_{12} W_{22}^{-1} (\theta - S_\theta \hat{\phi}) \right)^2 & \text{otherwise} \end{cases} \end{aligned}$$

Finally one can parameterize the unit length vector q as $q(\alpha) = [\cos \alpha, \sin \alpha]'$ such that the objective function $Q_T^{(2)}(\theta)$ can be computed as

$$Q_T^{(2)}(\theta) = \min_{\alpha \in [-\pi, \pi]} G_T(\theta, q(\alpha), \hat{\mu}(\theta, q(\alpha)); \phi).$$

Table 1: Monte Carlo Design

	Design 1	Design 2	Design 3	Design 4
	VAR(0)	VAR(1)	VAR(1)	VAR(1)
Σ_{11}	0.356	0.087	0.080	0.044
Σ_{21}	-0.122	-0.027	-0.023	-0.009
Σ_{22}	0.701	0.640	0.674	0.296
Φ_{11}		0.873	0.806	0.450
Φ_{12}		0.003	0.032	0.014
Φ_{21}		-0.229	-0.278	0.060
Φ_{22}		0.230	0.985	0.953
$\lambda_1(\Phi_1)$		0.871	$0.89 - 0.03i$	0.955
$\lambda_2(\Phi_1)$		0.231	$0.89 + 0.03i$	0.498

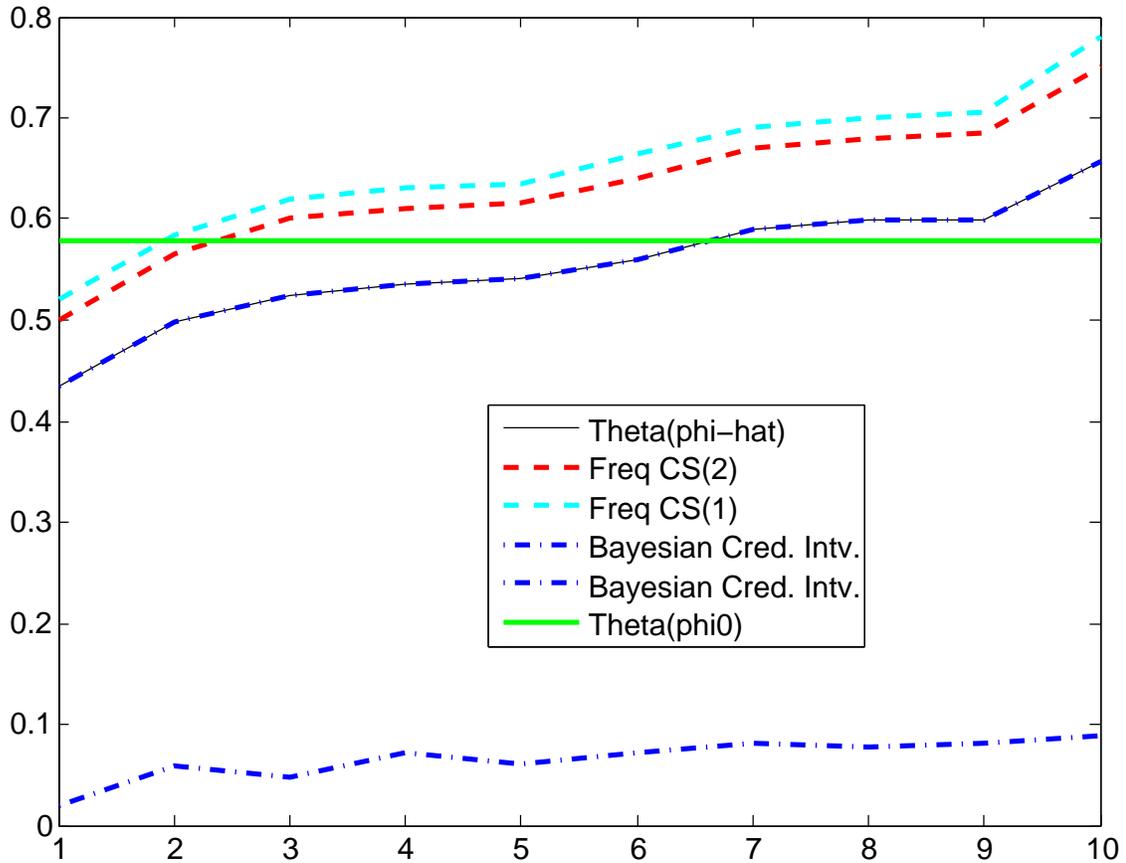
Notes: Designs are obtained by estimating a VAR(0) or VAR(1) of the form $y_t = \Phi_0 + \Phi_1 y_{t-1} + u_t$, $\mathbb{E}[u_t u_t'] = \Sigma$. We use OLS estimates, Φ entries refer to elements of Φ_1 . $\lambda_i(\Phi_1)$ is the i 'th eigenvalue of Φ_1 . $y_{1,t}$ is log difference of U.S. GDP deflator, scaled by 100 to convert into percentages. $y_{2,t}$ is either log difference of U.S. GDP or deviations of log GDP from a linear trend, scaled by 100. Design 1: inflation and GDP growth, 1964:I to 2004:IV. Design 2: inflation and output deviations from trend, 1964:I to 2006:IV. Design 3: inflation and output growth, 1964:I to 2006:IV. Design 4: inflation and output deviations from trend, 1983:I to 1006:IV.

Table 2: Monte Carlo Results for 90% Nominal Coverage Probability

		Design 1			Design 2		Design 3		Design 4	
		Sample Size			Sample Size		Sample Size		Sample Size	
		100	500	5,000	100	500	100	500	100	500
CS^ϕ	Coverage	82	94	90	83	90	80	95	86	86
$CS_{(2)}^\theta$	Coverage	93	99	95	95	96	92	96	98	96
$CS_{(2)}^\theta$	Length	0.64	0.61	0.59	0.31	0.28	0.27	0.25	0.16	0.11
$\Theta(\phi)$	Length	0.58	0.58	0.58	0.26	0.26	0.23	0.23	0.09	0.09

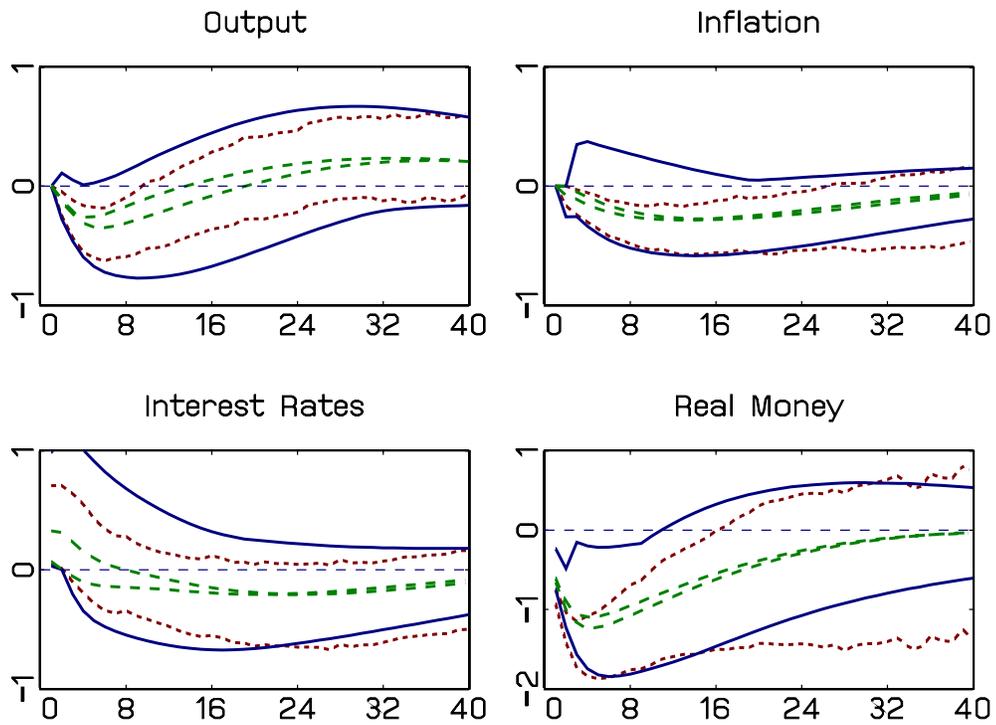
Notes: The coverage probability entries are measured in percent. Length refers to the average length of the confidence interval or credible set across Monte Carlo repetitions.

Figure 1: Frequentist and Bayesian Interval Estimates of θ (*Design 1*), $T = 100$



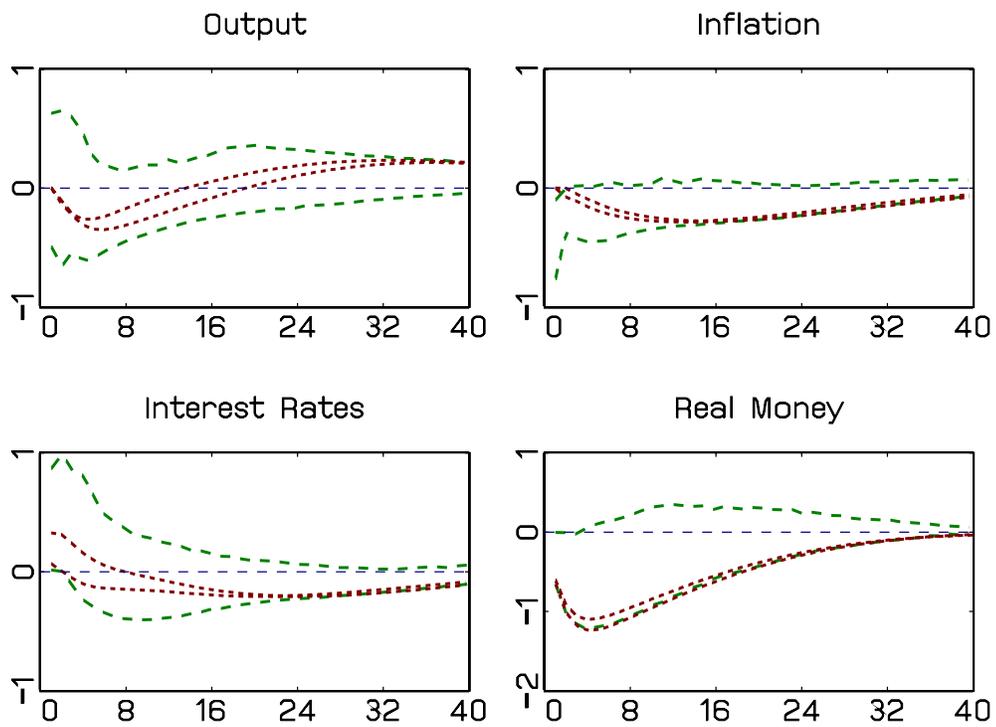
Notes: The graph depicts results from 10 replications of the Monte Carlo exercise described in Section 5.1. The replications (x -axis) are sorted with respect to the upper bound of $\Theta(\hat{\phi})$. The figure depicts the upper bounds of the identified set $\Theta(\phi_0)$, the estimated identified set $\Theta(\hat{\phi})$, the frequentist confidence sets $CS_{(1)}^\theta$ and $CS_{(2)}^\theta$, as well as the boundaries of Bayesian credible intervals. The upper bound of the Bayesian credible set essentially coincides with the upper bound of $\Theta(\hat{\phi})$.

Figure 2: Impulse Responses Based on Exclusion and Sign Restrictions



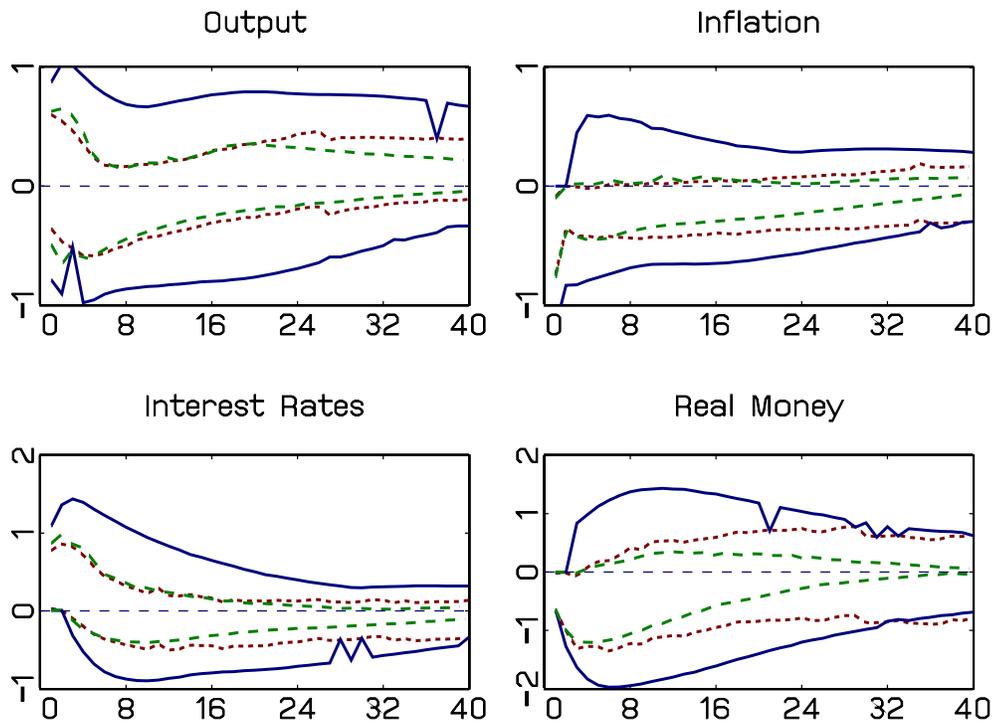
Notes: The figure depicts 90% Bayesian credible intervals (red, short dashes); 90% frequentist confidence sets (blue, solid); and identified sets conditional on the maximum likelihood estimates $\hat{\Phi}$ and $\hat{\Sigma}$ (green, long dashes).

Figure 3: Identified Sets



Notes: The figure depicts identified sets conditional on the maximum likelihood estimates $\hat{\Phi}$ and $\hat{\Sigma}$ for the estimation with exclusion and sign restrictions (red, short dashes); and pure sign restrictions (green, long dashes).

Figure 4: Impulse Responses Based on Pure Sign Restrictions



Notes: The figure depicts 90% Bayesian credible intervals (red, short dashes); 90% frequentist confidence sets (blue, solid); and identified sets conditional on the maximum likelihood estimates $\hat{\Phi}$ and $\hat{\Sigma}$ (green, long dashes).