Inference for VARs Identified with Sign Restrictions

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania, CEPR, and NBER

Eleonora Granziera
Bank of Canada

January 17, 2013

*Correspondence: H.R. Moon: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: hyungsikmoon@gmail.edu. F. Schorfheide: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. E-mail: schorf@ssc.upenn.edu. E. Granziera: Bank of Canada, 234 Rue Wellington, Ottawa (Ontario) K1A 0G9, Canada. E-mail: egranziera@bankofcanada.ca. We thank Jim Stock (co-editor), four anonymous referees, Fabio Canova and Eric Renault as well as seminar participants at the 2009 Canadian Econometric Study Group Meeting, the 2010 California Econometric Conference, the 2010 Econometric Society World Congress, the 2010 NBER Summer Institute, Boston University, Harvard-MIT, Queens University, University College London, the University of Pennsylvania, and Yale University. We also thank Mihye Lee for her contributions to the first draft of this paper and Minchul Shin for research assistance. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grants SES 0617803 and 1061725. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada. The Online Technical Appendix as well as data and software to replicate the empirical analysis are available at *http://www.ssc.upenn.edu/˜schorf.*

Abstract

There is a fast growing literature that set-identifies structural vector autoregressions (SVARs) by imposing sign restrictions on the responses of a subset of the endogenous variables to a particular structural shock (sign-restricted SVARs). Most methods that have been used to construct error bands for impulse responses of sign-restricted SVARs are justified only from a Bayesian perspective. This paper demonstrates how to formulate the inference problem for sign-restricted SVARs within a moment-inequality framework. In particular, it develops methods of constructing error bands for impulse response functions of sign-restricted SVARs that are valid from a frequentist perspective. The paper also provides a comparison of frequentist and Bayesian error bands in the context of an empirical application - the former can be substantially wider than the latter. (JEL: C1, C32)

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

1 Introduction

During the three decades following Sims's (1980) "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, as well as 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, economically interpretable structural innovations. Traditionally, SVARs have been constructed by imposing sufficiently many restrictions such that the relationship between structural innovations and forecast errors is one-to-one.

Over the past decade, starting with Faust (1998), Canova and De 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 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 De Nicolo, 2002; Uhlig, 2005), technology shocks (Dedola and Neri, 2007; Peersman and Straub, 2009), government spending shocks (Mountford and Uhlig, 2009; Pappa, 2009), and oil price shocks (Baumeister and Peersman, 2008; Kilian and Murphy, 2012).

Since impulse responses in sign-restricted SVARs can only be bounded, they belong to the class of set-identified econometric models, using the terminology of Manski (2003).[1] With the exception of Faust, Rogers, Swanson, and Wright (2003) and Faust, Swanson, and Wright (2004) (since the two papers are methodologically equivalent, we are using the abbreviation FRSW to refer to both of them), researchers have exclusively reported Bayesian error bands for sign-restricted VARs, and a general method for constructing uniformly asymptotically valid frequentist confidence intervals is absent from the literature. As shown in detail in Moon and Schorfheide (2012), the large-sample numerical equivalence of frequentist confidence sets and Bayesian credible sets breaks down in set-identified models, which means that Bayesian error bands may not be interpreted as approximate frequentist error bands.

[1] The microeconometrics literature uses the terms *set* and *partially* identified model interchangeably. In the VAR literature a *partially* identified structural VAR is one in which the researcher tries to identify only a subset of the structural shocks. To avoid confusion we shall use the term *set* identified because we are focusing on models in which impulse responses can only be bounded.

The goal of this paper is to provide researchers with tools to construct valid frequentist confidence bands for impulse responses of sign-restricted SVARs. The specific contributions are the following: First, we formulate the problem of analyzing set-identified sign-restricted SVAR models in a moment-inequality-based minimum distance framework. Second, we find an easily interpretable sufficient condition for the non-emptiness of the identified set of sign-restricted structural impulse responses and we propose a consistent estimator of the identified set that is straightforward to compute. Third, using our minimum distance framework, we formally analyze two methods of constructing error bands that delimit valid frequentist confidence intervals.[2] Fourth, we provide step-by step recipes for practitioners on how to compute these intervals.

At an abstract level, our inference problem is characterized by a vector of point-identified reduced-form parameters ϕ, a vector of structural parameters (impulse responses) θ, and a vector of nuisance parameters q. The sign restrictions generate an identified set for θ and q: $F^{\theta,q}(\phi)$. Moreover, conditional on q and ϕ the vector θ is point identified. To obtain a confidence set for θ we consider two approaches. A projection-based confidence set can be obtained by projecting a joint confidence interval for $(\theta, q) \in F^{\theta,q}(\phi)$ onto the θ ordinate. Alternatively, one can construct a confidence interval for the set-identified nuisance parameter q and then take the union of standard Wald confidence sets for θ that are generated conditional on all q in the first-stage confidence set. The Bonferroni inequality can be used to ensure the desired coverage probability of the resulting confidence set for θ. Finally, we show that the plug-in estimator $F^{\theta}(\hat{\phi})$ delivers a consistent estimate of the identified set for θ, denoted by $F^{\theta}(\phi)$.

The projection (also sometimes called Scheffé) method and the Bonferroni approach have a long history in the time series literature. For instance, Dufour (1990) and Cavanagh, Elliott, and Stock (1995) use it to eliminate nuisance parameters that characterize the persistence of error terms or regressors. In the context of structural VARs, the Bonferroni approach has been used by FRSW.[3] However, FRSW's setup is quite different from ours. In their framework the set identification of q arises from a rank deficiency in equality restrictions, which depend on estimated parameters. While FRSW restrict q further by imposing inequality conditions, these inequality conditions do not

[2] The contribution of this paper is meant to be positive. We do not criticize the use of Bayesian inference methods as long as it is understood that their output needs to be interpreted from a Bayesian perspective. We provide applied researchers who are interested in impulse response error bands that are valid from a frequentist perspective with econometric tools to compute such error bands.

[3] An earlier version of this paper focused solely on the projection method. We thank James Stock and an anonymous referee for pointing out the FRSW reference and prompting us to also analyze the Bonferroni approach.

depend on estimated parameters. Our analysis, on the other hand, focuses on inequality restrictions for q that may or may not be binding and do depend on estimated parameters. This generalization is essential to cover the wide range of empirical applications referenced above.

Building upon recent advances in the moment-inequality literature in microeconometrics, in particular, Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010a) we provide a complete asymptotic analysis of both the projection and the Bonferroni error bands, which is new in the sign-restricted SVAR literature. We use point-wise testing procedures to obtain the confidence sets for (θ, q) in the projection approach and for q in the Bonferroni procedure. We use Andrews and Soares's (2010a) moment selection procedure to tighten the critical values for the point-wise testing procedures. Since the number of linearly independent moment conditions is a function of the nuisance parameter q, the theorems in Andrews and Soares (2010a) are not directly applicable. We adapt their arguments to suit our model and prove that the proposed confidence sets are asymptotically valid in a uniform sense. Our results on the non-emptiness of identified sets for impulse responses complement the equality-restriction-based VAR identification results reported in Rubio-Ramirez, Waggoner, and Zha (2010).

The remainder of the paper is organized as follows. Section 2 provides a simple example of a sign-restricted SVAR. We describe how set identification arises in this model and sketch the projection and the Bonferroni approach to the construction of confidence intervals for impulse responses. Section 3 generalizes the model specification and introduces additional notation. The computation of the proposed confidence intervals is discussed in detail in Section 4. This section is geared toward practitioners. Technical assumptions and large sample results are presented in Section 5. Some extensions of the methods are discussed in Section 6. To illustrate the proposed methods, we conduct a small Monte Carlo study in Section 7 and generate error bands for output, inflation, interest rate, and money responses to a monetary policy shock in an empirical application in Section 8. Finally, Section 9 concludes. The proof of the main theorem presented in Section 5 is presented in the Appendix to this paper. All other derivations and proofs are relegated to a supplemental Online Appendix.

We use the following notation throughout the remainder of the paper: $\mathcal{I}\{x \geq a\}$ is the indicator function that is one if $x \geq a$ and zero otherwise. $0_{n\times m}$ is an $n \times m$ matrix of zeros and I_n is the $n \times n$ identity matrix. $\otimes$ is the Kronecker product, $vec(\cdot)$ stacks the columns of a matrix, and $vech(\cdot)$ vectorizes the lower triangular part of a square matrix. We use $diag(A_1, \ldots, A_k)$ to denote a quasi-diagonal matrix with submatrices A_1, ..., A_k on its diagonal and zeros elsewhere. If A is an $n \times m$ matrix, then $\|A\|_W = \sqrt{tr[WA'A]}$. 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. We write $x \gg 0$ to mean that all elements of the vector x are strictly greater than zero; we write $x > 0$ to mean that all elements of x are greater than or equal to zero but not all equal to zero, that is, $x \neq 0$; finally, we write $x \geq 0$ to mean that all elements of x are greater than or equal to zero. We use $\propto$ to indicate proportionality. "$\xrightarrow{p}$" A multivariate normal distribution is denoted by $N(\mu, \Sigma)$. We use χ^2_m to denote a χ^2 distribution with m degrees of freedom.

2 An Illustrative Example

Consider a bivariate VAR of lag order zero with Gaussian innovations: $y_t = u_t$ where $u_t \sim iidN(0, \Sigma_u)$. Suppose the vector y_t is composed of inflation and output growth and the one-step-ahead forecast errors are linear functions of "structural" demand and supply shocks, stacked in the vector $\epsilon_t = [\epsilon_{D,t}, \epsilon_{S,t}]'$. Thus, $u_t = \Sigma_{tr}\Omega_\epsilon\epsilon_t$, where $\epsilon_t \sim iidN(0, I)$, Σ_{tr} is the lower triangular Cholesky factor of Σ_u with elements Σ^{tr}_{ij} and Ω_ϵ is an orthogonal matrix. The object of interest, θ, is the inflation response to a demand shock $\epsilon_{D,t}$. Let the unit-length vector q denote the first column of Ω_ϵ. Using the notation

$$\phi = [\phi_1, \phi_2, \phi_3]' = [\Sigma^{tr}_{11}, \Sigma^{tr}_{21}, \Sigma^{tr}_{22}]' \quad \text{and} \quad q = [q_1, q_2]',$$

one can express the inflation response as

$$\theta = q_1\phi_1. \tag{1}$$

Here ϕ is a vector of reduced-form parameters that is consistently estimable. Suppose that in order to set-identify θ we impose the sign restriction that a demand shock moves prices and output in the same direction and the normalization restriction that a positive demand shock increases prices:

$$q_1\phi_1 \geq 0, \quad q_1\phi_2 + q_2\phi_3 \geq 0. \tag{2}$$

The goal of this paper is to construct a confidence interval for θ. We begin with a heuristic analysis of this stylized VAR(0) example. Section 2.1 discusses the identified set for (θ, q) and its projections onto the θ and q ordinates and Section 2.2 develops the projection-based and Bonferroni confidence intervals for θ.

2.1 Identified Sets

The identified set for (θ, q) is the set of values that satisfies (1) and (2) conditional on ϕ. Since the Cholesky factorization of the covariance matrix Σ_u is normalized such that ϕ_1 and ϕ_3 are nonnegative, (2) implies that

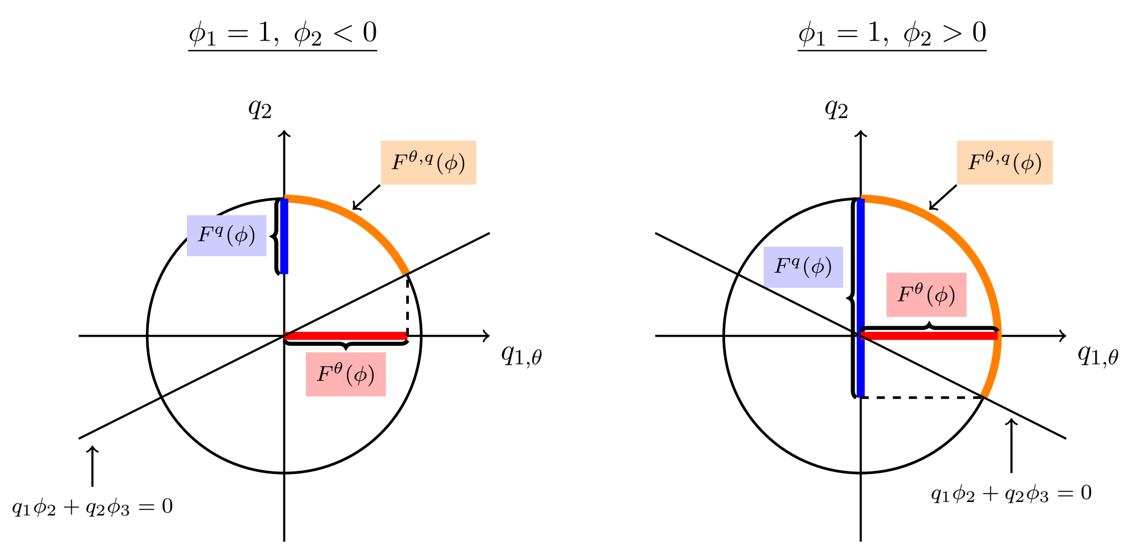
$$q_1 \geq 0 \quad \text{and} \quad q_2 \geq -\frac{\phi_2}{\phi_3} q_1. \tag{3}$$

Figure 1 provides an illustration of the constraint. Each panel depicts a unit circle as well as the locus $q_1\phi_2 + q_2\phi_3 = 0$. In the left panel $\phi_2 < 0$, whereas in the right panel $\phi_2 > 0$. Under the assumption that $\phi_1 = 1$ we obtain that $\theta = q_1$ and the x-axis in the two panels also represents the parameter of interest θ. We denote the identified set for (θ, q) by $F^{\theta,q}(\phi)$. $F^{\theta,q}(\phi)$ is given by the arc that ranges from the intersection of the unit circle with the y-axis to the intersection with $q_1\phi_2 + q_2\phi_3 = 0$. The identified set can be represented as the argmin of the following non-negative objective function:

$$G(\theta, q; \phi, W) = \min_{\mu \geq 0} \left\| \begin{bmatrix} q_1\phi_1 - \theta \\ q_1\phi_2 + q_2\phi_3 - \mu \end{bmatrix} \right\|_W^2, \tag{4}$$

where W is a symmetric positive-definite weight matrix. It is straightforward to verify that $(\theta, q) \in F^{\theta,q}(\phi)$ if and only if $G(\theta, q; \phi, W) = 0$.

Figure 1: Identified Sets for VAR(0)



Since the goal is to make inference about the impulse response θ, an important object in the subsequent analysis is the projection of $F^{\theta,q}(\phi)$ onto the θ-ordinate. If $\phi_2 \geq 0$, the projection is straightforward because the second inequality in (3) provides no constraint on q_1, which can take any value on the unit interval. Thus, the identified set ranges from 0 to ϕ_1. If $\phi_2 < 0$, then the inequality constrains q_1 to be less than $\sqrt{\phi_3^2/(\phi_2^2+\phi_3^2)}$. Overall, we obtain

$$F^{\theta}(\phi) = \left[0,\ \phi_1 \max\left\{\mathcal{I}\{\phi_2 \geq 0\},\ \sqrt{\frac{\phi_3^2}{\phi_2^2+\phi_3^2}}\right\}\right]. \tag{5}$$

The non-negative function

$$G^{\theta}(\theta;\phi,W) = \min_{\|q\|=1}\ G(\theta,q;\phi,W), \tag{6}$$

has the property that $\theta \in F^{\theta}(\phi)$ if and only if $G^{\theta}(\theta;\phi,W)=0$. We can also project the identified set onto the q-ordinate to obtain $F^{q}(\phi)$. Since in our example one of the inequality constraints is placed directly on θ, $F^{q}(\phi)$ can be represented as the argmin of the function

$$G^{q}(q;\phi,W) = \min_{\theta \geq 0}\ G(\theta,q;\phi,W). \tag{7}$$

The constrained minimization with respect to θ ensures that the inequality constraint $q_1\phi_1 \geq 0$ is satisfied.

2.2 Projection-Based and Bonferroni Confidence Sets

The starting point for the inference procedures considered in this paper is the distribution of the estimator $\hat{\phi}$ of the reduced-form VAR parameters. Rather than approximating the distribution of $\hat{\phi}$ implied by $y_t \sim iidN(0,\Sigma_u)$, we proceed under the assumption that $\sqrt{T}(\hat{\phi}-\phi) \sim N(0,I)$ for all $\phi \in \mathcal{P}$ to illustrate the inferential problem. The goal is to obtain a confidence interval $CS^{\theta}(\hat{\phi})$ that satisfies the condition

$$\inf_{\phi \in \mathcal{P}}\ \inf_{\theta \in F^{\theta}(\phi)}\ P_{\phi}\{\theta \in CS^{\theta}(\hat{\phi})\} \geq 1-\alpha. \tag{8}$$

Note that the parameter θ does not appear as an index of the probability distribution P, because conditional on ϕ the parameter θ does not affect the distribution of the estimator of the reduced-form parameters $\hat{\phi}$. The confidence interval is indexed by $\hat{\phi}$ because it is a sufficient statistic in our setup. Both the projection and the Bonferroni approach amount to the marginalization of a joint confidence set for (θ,q).[4] In the context of structural VARs, the Bonferroni idea was first used in

[4]Following the approach of Imbens and Manski (2004), one could alternatively try to construct a confidence interval based on the distribution of an estimator for the boundary of $F^{\theta}(\phi)$. However, in a typical VAR application,

work by FRSW, though not at the same level of generality as in this paper. In particular, in FRSW the inequality restrictions on q did not depend on estimated parameters.

To construct the confidence sets we use the function $G(\theta, q; \phi, W)$ in (4) evaluated at $\phi = \hat{\phi}$. In this example, we choose the weight matrix such that the random variables $q_1\hat{\phi}_1$ and $q_1\hat{\phi}_2 + q_2\hat{\phi}_3$ are transformed into independent standard normals:

$$W^*(q) = T \left[\begin{array}{cc} 1/q_1^2 & 0 \\ 0 & 1 \end{array} \right].$$

Since the upper-right element of the weight matrix is infinity if $q = [0, 1]'$, we separate this case using indicator function notation:[5]

$$\begin{aligned} G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big) &= \min_{\mu \geq 0} \frac{T}{q_1^2} \|q_1\hat{\phi}_1 - \theta\|^2 \mathcal{I}\{q_1 \neq 0\} + T\|q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu\|^2 \\ &\quad + T\mathcal{I}\{q_1 = 0 \text{ and } \theta \neq 0\}. \end{aligned} \tag{9}$$

Recall that $q = [0, 1]'$ generates the lower bound $\theta_l = 0$ of $F^\theta(\phi)$. Since there is no uncertainty with respect to this lower bound, we subsequently drop the indicator functions from the representation $G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big)$ to simplify the notation.

Projection Approach. A joint confidence set of (θ, q) can be obtained as a level set constructed from the objective function $G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big)$:

$$CS^{\theta,q} = \left\{ \theta, q \;\middle|\; \theta \geq 0,\ \|q\| = 1,\ G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big) - c_P^\alpha \leq 0 \right\}. \tag{10}$$

Here c_P^α is a cut-off level or critical value such that the confidence set satisfies the coverage probability condition in (8). For now we assume that the critical value is constant. The projection of this confidence set onto the θ-ordinate is obtained by minimizing $G\big(\theta, q; \hat{\phi}, W(\cdot)\big)$ with respect to q:

$$CS_P^\theta = \left\{ \theta \;\middle|\; \theta \geq 0,\ \min_{\|q\|=1} \left(G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big) - c_P^\alpha \right) \leq 0 \right\}. \tag{11}$$

A critical value c_P^α can be obtained from the analysis of the sampling distribution of the objective function $G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big)$. For each $(\theta, q) \in F^{\theta,q}(\phi)$ we have $\theta = q_1\phi_1$ and there exists a $\tilde{\mu} \geq 0$ such

the characterization of the endpoints would involve a complicated constrained optimization problem; adjustments along the lines of Stoye (2009) would be necessary to ensure that the confidence interval is valid uniformly if the size of the identified set is small compared to the sampling variation in $\hat{\phi}$; and extensions to θ's that have a dimension that is greater than one are very difficult. For these reasons, we leave such an approach for future research.

[5]If $q_1 = 0$, the variance of $q_1\hat{\phi}_1$ is zero. This q-dependent singularity distinguishes the VAR application from the standard moment inequality setup assumed in the econometrics literature, e.g., Andrews and Soares (2010a).

that $\tilde{\mu} = q_1\phi_2 + q_2\phi_3 \geq 0$. Let $\nu = \sqrt{T}(\mu - \tilde{\mu})$. The assumption $\sqrt{T}(\hat{\phi} - \phi) \sim N(0, I)$ implies that

$$\begin{bmatrix} q_1\sqrt{T}(\hat{\phi}_1 - \phi_1) \\ q_1\sqrt{T}(\hat{\phi}_2 - \phi_2) + q_2\sqrt{T}(\hat{\phi}_3 - \phi_3) \end{bmatrix} = \begin{bmatrix} q_1 Z_1 \\ Z_2 \end{bmatrix}, \tag{12}$$

where Z_1 and Z_2 are independent $N(0,1)$ random variables. Thus, we can rewrite (9) to obtain

$$G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big) = \min_{\nu \geq -\sqrt{T}\tilde{\mu}} \big(Z_1^2 + (Z_2 - \nu)^2\big). \tag{13}$$

Notice that the distribution of $G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big)$ depends on the slackness $\tilde{\mu}$ in the inequality condition (scaled by $\sqrt{T}$). If the slackness is large, the minimization with respect to ν eliminates the second term in (13) from the objective function with high probability. A conservative bound for the objective function is obtained by assuming that the slackness in the inequality condition is zero:

$$G\big(\theta, q; \hat{\phi}, W^*(\cdot)\big) \leq \min_{\nu \geq 0} \; Z_1^2 + (Z_2 - \nu)^2 = Z_1^2 + Z_2^2 \mathcal{I}\{Z_2 \leq 0\}. \tag{14}$$

Choosing c_P^α as the $100(1-\alpha)$ percentile of the distribution of $Z_1^2 + Z_2^2\mathcal{I}\{Z_2 \geq 0\}$ leads to a uniformly valid projection-based confidence set that satisfies (8).

Notice from Figure 1 that if $\phi_2 > 0$ the inequality $q_1\phi_2 + q_2\phi_3$ is not binding at the boundary of the identified set. In turn, the distribution of the $G(\cdot)$ objective function is asymptotically $Z_1^2 < Z_1^2 + Z_2^2\mathcal{I}\{Z_2 \geq 0\}$. In Section 4 we will use the moment-selection approach of Andrews and Soares (2010a) to eliminate non-binding moment conditions before we construct critical values for the objective function. This reduces the size of the confidence set.

Bonferroni Approach. The Bonferroni approach of constructing a $1-\alpha$ confidence set for θ consists of three steps. Let $\alpha = \alpha_1 + \alpha_2$. First, construct a $1-\alpha_1$ confidence interval CS_B^q for q with the property that

$$\inf_{\phi \in \mathcal{P}} \; \inf_{q \in F^q(\phi)} \; P_\phi\{q \in CS_B^q\} \geq 1 - \alpha_1. \tag{15}$$

Second, generate a confidence interval for θ conditional on q. Since $\theta = q_1\phi$ is point-identified conditional on q, the second step is straightforward:

$$CS_{B,q}^\theta = \mathbb{R}^+ \bigcap \Big[q_1(\hat{\phi}_1 - z_{\alpha_2/2}/\sqrt{T}),\; q_1(\hat{\phi}_1 + z_{\alpha_2/2}/\sqrt{T})\Big]. \tag{16}$$

Here $z_{\alpha_2/2}$ is the two-sided α_2 critical value associated with the $N(0,1)$ distribution. Since our model implies that $q_1 \geq 0$ we truncate the confidence interval at zero. Third, obtain the confidence set for θ by taking the following union of $CS_{B,q}^\theta$ sets:

$$CS_B^\theta = \bigcup_{q \in CS_B^q} CS_{B,q}^\theta. \tag{17}$$

The Bonferroni inequality implies that the coverage probability of CS_B^θ is at least $1-\alpha$.

To obtain a confidence set for q in the first step we follow the same approach as for the construction of $CS^{\theta,q}$ above. For any $q \in F^q(\phi)$ there exists a $\tilde{\theta} \geq 0$ and $\tilde{\mu} \geq 0$ such that

$$q_1\phi_1 = \tilde{\theta} \quad \text{and} \quad q_1\phi_2 + q_2\phi_3 = \tilde{\mu}.$$

Letting $\eta = \sqrt{T}(\theta - \tilde{\theta})$ and $\nu = \sqrt{T}(\mu - \tilde{\mu})$ we obtain

$$\begin{aligned} G^q\big(q;\hat{\phi}, W^*(\cdot)\big) &= \min_{\eta \geq -\sqrt{T}\tilde{\theta},\, \nu \geq -\sqrt{T}\tilde{\mu}} (Z_1 - \eta/|q_1|)^2 + (Z_2 - \nu)^2 \\ &\leq Z_1^2 \mathcal{I}\{Z_1 \leq 0\} + Z_2^2 \mathcal{I}\{Z_2 \leq 0\}. \end{aligned} \tag{18}$$

Let $c_B^{\alpha_1}$ be the $100(1-\alpha_1)$ percentile of the distribution of $Z_1^2\mathcal{I}\{Z_1 \geq 0\} + Z_2^2\mathcal{I}\{Z_2 \geq 0\}$, then a $1-\alpha_1$ confidence set for q can be obtained as level set

$$CS_B^q = \left\{ q \;\middle|\; \|q\| = 1 \text{ and } G^q\big(q;\hat{\phi}, W^*(\cdot)\big) \leq c_B^{\alpha_1} \right\}. \tag{19}$$

As for the projection-based confidence interval, in Section 4 we will use the Andrews and Soares (2010a) moment-selection approach to eliminate non-binding moment conditions when constructing critical values for the objective function $G^q(\cdot)$. In this VAR(0) example, at most one inequality condition is binding at the boundary of $F^q(\phi)$ so that the critical value can be reduced to the $100(1-\alpha_1)$ quantile of the distribution of $Z_1^2\mathcal{I}\{Z_1 \leq 0\}$.

We show in the Online Appendix that even in this simple example there is no clear ranking of the projection-based and the Bonferroni confidence interval if a moment-selection procedure is used to eliminate non-binding inequality conditions. If $\phi_2 > 0$, the uncertainty about the boundary of $F^q(\phi)$ is not relevant for the least favorable q that determines the upper bound of CS_B^θ. Moreover, the moment-selection procedure detects that the inequality condition $\phi_2 q_1 + \phi_3 q_2 \geq 0$ is non-binding. As a consequence, for any $\alpha_1 > 0$ the projection interval is shorter than the Bonferroni interval. If $\phi_2 < 0$, then this ranking may get reversed, depending on ϕ and α_1. Now the uncertainty about q is relevant for the upper bound of the Bonferroni confidence interval for θ. Moreover, the inequality moment condition $\phi_2 q_1 + \phi_3 q_2 \geq 0$ is binding at the boundary of $F^\theta(\phi)$, which raises the critical value for the projection-based confidence set. In Section 7 we provide some Monte Carlo evidence on the relative length of the two confidence sets.

3 General Setup and Notation

We now generalize the setup to an n-dimensional VAR with p lags, which takes the form

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad \mathbb{E}[u_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[u_t u_t'|\mathcal{F}_{t-1}] = \Sigma_u. \tag{20}$$

Here y_t is an $n \times 1$ vector and the information set $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ is composed of the lags of y_t's. Constants and deterministic trend terms are omitted because they are irrelevant for the subsequent discussion. The one-step-ahead forecast errors u_t are linear functions of a vector of fundamental innovations ϵ_t:

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega_\epsilon \epsilon_t, \quad \mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t'|\mathcal{F}_{t-1}] = I, \tag{21}$$

where Σ_{tr} is the lower triangular Cholesky factor of Σ_u and Ω_ϵ is an arbitrary orthogonal matrix. Assuming that the lag polynomial associated with the VAR in (20) is invertible, one can express y_t as the following infinite-order vector moving average (VMA) process:

$$y_t = \sum_{h=0}^{\infty} C_h(\Phi_1, \ldots, \Phi_p) u_{t-h} = \sum_{h=0}^{\infty} C_h(\Phi_1, \ldots, \Phi_p) \Sigma_{tr} v_{t-h}, \tag{22}$$

where $\mathbb{E}[v_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[v_t v_t'|\mathcal{F}_{t-1}] = I$. Each column of the matrices of the moving average representation can be interpreted as impulse responses to the orthogonalized innovations $v_{i,t}$:

$$\mathbb{E}[y_{t+h}|v_{i,t} = 1, \mathcal{F}_t] - \mathbb{E}[y_{t+h}|\mathcal{F}_t] = \frac{\partial y_{t+h}}{\partial v_{i,t}} = (C_h \Sigma_{tr})_{.i}, \quad h = 0, 1, 2, \ldots. \tag{23}$$

Here $(A)_{.i}$ denotes the i'th column of matrix A.

The object of interest is the $\tilde{k} \times 1$ vector θ of impulse responses to a structural shock. Let R^v be the $\tilde{l} \times n$ matrix of non-redundant reduced-form impulse responses that are needed to construct θ and the sign-restricted structural responses.[6] This matrix is composed of rows of the $C_h \Sigma_{tr}$ matrices in (23). Throughout this paper we are focusing on the identification of a single structural shock that is obtained by post-multiplying R^v with a unit length vector q, which can be interpreted as a column of the Ω_ϵ matrix in (21).

We now apply two operations to $R^v q$. First, let $M^{S,1}$ be an $\tilde{l} \times \tilde{l}$ diagonal matrix with diagonal entries of 1 or -1. This matrix can be used to change the sign of impulse responses. Second, let $M^{S,2}$ be a full-rank $\tilde{l} \times \tilde{l}$ matrix that re-orders the rows of $R^v q$, such that the $\tilde{k}$ impulse responses

[6] We assume that these sign restrictions do not encode equality restrictions (e.g., by representing $a = 0$ as $a \leq 0$ and $a \geq 0$.) The extension to models that combine sign-restrictions and equality restrictions is deferred to Section 6.

collected in the vector θ appear on top. Overall, the structural impulse responses can be expressed as

$$M^{S,2}M^{S,1}R^v q = (M^{S,2}M^{S,1} \otimes q')vec(R^{v'}).$$

We now introduce the $m \times \tilde{l}n$ selection matrix S'_ϕ that eliminates elements from the vector $vec(R^{v'})$ that are known to be zero. These zeros arise, for example, if R^v contains rows of the instantaneous response matrix Σ_{tr}. We stack the non-trivial elements of $vec(R^{v'})$ in the $m \times 1$ vector of reduced-form parameters $\phi = S'_\phi vec(R^{v'})$, which can be consistently estimated from the data. Since $S_\phi S'_\phi = I$ we can write the structural responses as

$$\tilde{S}(q)\phi = (M^{S,2}M^{S,1} \otimes q')S_\phi \phi, \quad \text{where} \quad \tilde{S}(q) = \begin{bmatrix} \tilde{S}_\theta(q) \\ \tilde{S}_R(q) \end{bmatrix}. \tag{24}$$

The matrix $\tilde{S}(q)$ is partitioned such that $\theta = \tilde{S}_\theta(q)\phi$ and the $\tilde{r} \times m$ matrix $\tilde{S}_R(q)$ generates the impulse responses that are sign restricted (but not part of θ). Let M_θ be the matrix that selects the elements of θ that are sign restricted (and possibly multiplies them by -1), then the VAR identified with sign restrictions can be represented as

$$\tilde{S}_\theta(q)\phi = \theta, \quad M_\theta \tilde{S}_\theta(q)\phi \geq 0, \quad \text{and} \quad \tilde{S}_R(q)\phi \geq 0. \tag{25}$$

In the context of the VAR(0) model presented in Section 2 these three conditions are

$$\begin{bmatrix} q_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \theta, \quad 1 \cdot \begin{bmatrix} q_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} \geq 0, \quad \text{and} \quad \begin{bmatrix} 0 & q_1 & q_2 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} \geq 0.$$

Throughout this paper we focus on the case in which the inequality conditions in (25) do not uniquely identify q and θ.

In our model we can define three types of identified sets conditional on ϕ: an identified set for the pair (θ, q), denoted by $F^{\theta,q}(\phi)$; an identified set for θ, denoted by $F^\theta(\phi)$; and an identified set for q, denoted by $F^q(\phi)$. To do so, it is useful to introduce the non-negative function

$$\tilde{G}(\theta, q; \phi, \tilde{W}) = \min_{\mu \geq 0} \left\| \tilde{S}(q)\phi - \begin{pmatrix} \theta \\ \mu \end{pmatrix} \right\|^2_{\tilde{W}}, \tag{26}$$

where $\tilde{W}$ is a positive-definite matrix and μ regulates the slackness of the inequalities from the sign restrictions on the responses not included in θ. It is straightforward to verify that

$$(\theta, q) \in F^{\theta,q}(\phi) \quad \text{if and only if} \quad \tilde{G}(\theta, q; \phi, \tilde{W}) = 0 \quad \text{and} \quad M_\theta \theta \geq 0. \tag{27}$$

We then can obtain $F^{\theta}(\phi)$, that is the set of impulse responses θ that are consistent with a particular ϕ, and $F^{q}(\phi)$, the set of unit-length vectors that generates responses that satisfy the sign restrictions, as projections by finding the roots of

$$\tilde{G}^{\theta}(\theta;\phi,\tilde{W}) = \min_{\|q\|=1} \tilde{G}(\theta,q;\phi,\tilde{W}), \quad \tilde{G}^{q}(q;\phi,\tilde{W}) = \min_{\theta:\, M_{\theta}\theta\geq 0} \tilde{G}(\theta,q;\phi,\tilde{W}). \tag{28}$$

We show in the Online Appendix that $F^{\theta}(\phi)$ is a bounded interval for $\tilde{k}=1$. Thus, despite the lack of point-identification, it is guaranteed that the response of the endogenous variables y_t to a one-standard-deviation structural shock is bounded. Moreover, the pointwise responses in this model can be characterized by two numbers: the lower bound and the upper bound of the identified interval.

4 Computing Confidence Sets for Impulse Responses

This section describes how to compute confidence sets for θ based on the projection and the Bonferroni approach. These confidence sets generalize (11) and (17). A formal statement of assumptions and a rigorous analysis of the large sample properties of the confidence set will follow in Section 5. The remainder of this section is organized as follows. Section 4.1 briefly discusses the estimation of ϕ. Section 4.2 describes how to deal with rank reductions in $\tilde{S}(q)$ as a function of q and Section 4.3 introduces a straightforward plug-in estimator of the identified set $F^{\theta}(\phi)$. The construction of the projection-based and the Bonferroni confidence sets is presented in Sections 4.4 and 4.5, respectively. Computational details are provided in Section 4.6. Throughout this section we assume that the impulse responses are not restricted through equality conditions (e.g., the restriction that certain responses have to be zero). Extensions of our approach to a setting in which some identifiying information is extracted from equality conditions are straightforward but notationally cumbersome and discussed in Section 6.

4.1 Estimating the Reduced-Form Responses ϕ

The first step of the analysis is to estimate the $m\times 1$ reduced-form parameter vector ϕ and its asymptotic variance-covariance matrix, denoted by Λ. 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 and that the asymptotic covariance matrix can be estimated consistently:

$$\sqrt{T}(\hat{\phi} - \phi) \Longrightarrow N\big(0, \Lambda\big) \quad \text{and} \quad \hat{\Lambda} \stackrel{p}{\longrightarrow} \Lambda \geq 0. \tag{29}$$

This assumption requires that all roots of the characteristic polynomial associated with the difference equation (20) lie outside of the unit circle. Throughout the paper, we are ruling out the presence of unit roots and are assuming that y_t is trend stationary. Moreover, we proceed under the assumption that Λ is full rank. Since we previously eliminated the $n(n-1)/2$ zero elements of the lower triangular matrix Σ_{tr} from the vector ϕ, the rank condition does not impose serious constraints on the applicability of our analysis.[7]

4.2 Dealing with the Rank Reduction in the Jacobian Matrix $\tilde{S}(q)$

The objective function $\tilde{G}(\theta, q; \phi, \tilde{W})$ in (26) depends on the matrix-valued function $\tilde{S}(q)$. Although this function is continuous in q, its row rank tends to be discontinuous. Since, among others, we will consider a weight function that is based on the inverse covariance matrix of $\sqrt{T}\tilde{S}(q)(\hat{\phi} - \phi)$ we need to pay special attention to the potential rank reduction. To fix ideas, recall the example of Section 2, which leads to

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

The first row of $\tilde{S}(q)$ becomes zero if $q = [q_1, q_2]' = [0, 1]$. The singularity arises because S_ϕ eliminates the second column of the matrix $(I \otimes q')$, which for $q_1 = 0$ contains the only non-zero entry in the first row of $\tilde{S}(q)$.

Recall the definition $\tilde{S}(q) = (M^{S,2}M^{S,1} \otimes q')S_\phi$ in (24). $M^{S,1}$ is a diagonal matrix with diagonal elements 1 or -1 and $M^{S,2}$ is a matrix that re-orders the rows of $M^{S,1}$. Thus, the rows of $(M^{S,2}M^{S,1} \otimes q')$ are orthogonal to each other. Since the matrix S_ϕ eliminates columns of $(M^{S,2}M^{S,1} \otimes q')$ and thereby columns of q', we deduce that a row rank reduction of $\tilde{S}(q)$ can only

[7] Consider a 4-variable VAR(4) and suppose that the responses of 3 of the 4 variables are restricted upon impact and for the subsequent 3 periods. Moreover, the object of interest is the response of the fourth variable at horizon $h = 9$. The number of reduced-form coefficients is $4 \cdot 16 + 10 = 74$. In order to construct the sign-restricted responses as well as θ, the number of elements in the vector ϕ is bounded by $4 \cdot 3 \cdot 4 + 4 = 52$, which means that the covariance matrix Λ is non-singular.

arise through one or more rows of zero. In order to eliminate these rows of zeros in $\tilde{S}(q)$, we introduce the selection matrices $V(q)$, $V_\theta(q)$, and $V_R(q)$ and define the matrices

$$S(q) = V(q)\tilde{S}(q),\ S_\theta(q) = V_\theta(q)\tilde{S}_\theta(q),\ S_R(q) = V_R(q)\tilde{S}_R(q). \tag{30}$$

The row dimensions of the three matrices are $l(q)$, $k(q)$, and $r(q)$, respectively. By construction, the three matrices have full row rank.

Subsequently we consider weight matrices that depend on q to standardize either the joint or the marginal limit distributions of the elements of $\sqrt{T}\tilde{S}(q)(\hat{\phi}-\phi)$. Thus, we define $W(q)$ such that

$$\tilde{W}(q) = V'(q)W(q)V(q)$$

Using the above notation this leads to

$$G(\theta, q; \phi, W(\cdot)) = \min_{\mu \geq 0} \left\| S(q)\phi - V(q)\begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{W(q)} + T\sum_{j=1}^{\tilde{k}} \mathcal{I}\{\tilde{S}_{j,\theta}(q) = 0 \text{ and } \theta_j \neq 0\}, \tag{31}$$

where the penalty accounts for the fact that if $(\theta, q) \in F^{\theta,q}(\phi)$ then the θ_j associated with deleted rows of $\tilde{S}_\theta(q)$ has to be zero (see Equation (9)). The factor T is essentially arbitrary and could be replaced by any number that exceeds the critical value used in the construction of the confidence set below. We will now use the sample analogue of $G(\theta, q; \phi, W(\cdot))$ to construct our confidence sets.

4.3 A Plug-In Estimator of $F^\theta(\phi)$

While the ultimate goal is to construct confidence intervals for θ, it is useful to also have an estimator of the identified set $F^\theta(\phi)$ available. Such an estimator can be easily obtained using a plug-in approach that replaces ϕ by $\hat{\phi}$ in (25) and the objective function $G(\theta, q; \phi, W(\cdot))$. Let

$$\begin{aligned} F^\theta(\hat{\phi}) &= \left\{ \theta \;\middle|\; M_\theta\theta \geq 0 \text{ and } \exists\, \|q\| = 1 \text{ s.t. } \tilde{S}_\theta(q)\hat{\phi} = \theta,\ \tilde{S}_R(q)\hat{\phi} \geq 0 \right\} \\ &= \left\{ \theta \;\middle|\; M_\theta\theta \geq 0 \text{ and } \left(\min_{\|q\|=1} G\big(\theta, q; \hat{\phi}, W(\cdot)\big) \right) = 0 \right\}, \end{aligned} \tag{32}$$

where $W(\cdot)$ is a symmetric and positive definite weight matrix that possibly is a function of q (see above). Plug-in estimators of $F^{\theta,q}(\phi)$ and $F^q(\phi)$ can be obtained in a similar manner.

4.4 Projection-Based Confidence Set

The joint confidence set for θ and q is constructed as the following generalized level set:

$$CS^{\theta,q} = \left\{ \theta,\, q \;\middle|\; \|q\| = 1,\; M_\theta \theta \geq 0,\; \text{and}\; G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big) - c_P^\alpha(q) \leq 0 \right\}. \tag{33}$$

We use a critical value $c_P^\alpha(q)$ that is a function of q (but not a function of θ). The projection of the joint confidence set $CS^{\theta,q}$ onto Θ takes the form

$$CS_P^\theta = \left\{ \theta \;\middle|\; M_\theta \theta \geq 0,\; \text{and}\; \min_{\|q\|=1} \Big(G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big) - c_P^\alpha(q) \Big) \leq 0 \right\}. \tag{34}$$

The computation of the confidence set is implemented by pointwise evaluation of the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big)$ and the critical value function $c_P^\alpha(q)$. For each value (θ, q) on a suitably chosen grid, we determine whether the grid point is included in the confidence set or not.

Sample Objective Function. The estimator $\hat{\phi}$ in the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big)$ is an estimator that satisfies (29). The weight matrix $\hat{W}(q)$ is obtained as follows. Let $\Sigma(q)$ be the asymptotic covariance matrix of $\sqrt{T} S(q)(\hat{\phi} - \phi)$. A consistent estimator is given by

$$\hat{\Sigma}(q) = S(q)\hat{\Lambda}S'(q) = \hat{D}^{1/2}(q)\hat{\Omega}(q)\hat{D}^{1/2}(q),$$

where $\hat{\Omega}(q)$ is the correlation matrix associated with $\hat{\Sigma}(q)$ and $\hat{D}^{1/2}(q)$ is a diagonal matrix of standard deviations. We then let

$$\hat{W}(q) = T\hat{D}^{-1/2}(q)\hat{B}(q)\hat{D}^{-1/2}(q) \tag{35}$$

and focus on two particular choices of $\hat{B}(q)$: $\hat{B}(q) = \hat{\Omega}^{-1}(q)$ and $\hat{B}(q) = I$. The corresponding weight matrices are denoted by $\hat{W}^*(q)$ and $\hat{W}^D(q)$, respectively. Overall, this leads to the sample objective function

$$\begin{aligned} G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big) &= \min_{\tilde{\mu} \geq 0} T \left\| \hat{D}^{-1/2}(q)S(q)\hat{\phi} - \hat{D}^{-1/2}(q)V(q)\begin{pmatrix}\theta \\ \tilde{\mu}\end{pmatrix} \right\|^2_{\hat{B}(q)} \\ &\quad + T\sum_{j=1}^{\tilde{k}} \mathcal{I}\big\{\tilde{S}_{j,\theta}(q) = 0 \text{ and } \theta_j \neq 0\big\}. \end{aligned} \tag{36}$$

The objective function $G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big)$ has the same structure as the objective functions considered in the literature on moment inequality models, e.g., Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010a). The two main differences in the VAR application are that q is a nuisance parameter and the object of the interest is a marginal confidence set, not a joint confidence set, and that the dimension of $S(q)\hat{\phi}$, $\hat{D}(q)$ and $\hat{B}(q)$ varies with q and the limit $\hat{D}(q)$ as a function of q can be singular.

Critical Values. In order to obtain the critical value function $c_P^\alpha(q)$ in (33) and (34), we apply the moment selection approach of Andrews and Soares (2010a). The moment selection tries to eliminate clearly non-binding inequality conditions from the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big)$ before approximating its distribution. Let $\hat{D}_R^{-1/2}$ be the submatrix of $\hat{D}$ that is conformable with the submatrix $S_R(q)$. Note from (36) that the minimization with respect to $\tilde{\mu}$ annihilates elements of $\hat{D}_R^{-1/2}(q)S_R(q)\hat{\phi}$ that are positive with probability (close to) one. In turn, these elements can be excluded from the computation of the critical value. An estimate of the slackness in inequality condition $j = 1, \dots, r(q)$ is provided by

$$\hat{\xi}_{j,T}(q) = \hat{D}_{jj,R}^{-1/2}(q)S_R(q)\sqrt{T}\hat{\phi}. \tag{37}$$

Inequality condition j is deemed non-binding if

$$\hat{\xi}_{j,T}(q) \geq \kappa_T, \tag{38}$$

where κ_T is a diverging sequence, e.g., $\kappa_T = 1.96\ln(\ln T)$. Thus, estimates of the number of non-binding and binding moment inequality constraints are given by

$$\hat{r}_2(q) = \sum_{j=1}^{r(q)} \mathcal{I}\{\hat{\xi}_{j,T}(q) \geq \kappa_T\} \quad \text{and} \quad \hat{r}_1(q) = r(q) - \hat{r}_2(q), \tag{39}$$

respectively.

Define the $(k(q)+\hat{r}_1(q)) \times (k(q)+r(q))$ selection matrix $M_{\hat{\xi}}(q)$ that deletes rows of $\hat{D}^{-1/2}(q)S(q)\hat{\phi}$ that correspond to non-binding inequality conditions in the sense of (38). Moreover, let

$$Z_m \sim N(0, I_m) \quad \text{and} \quad \hat{A}'(q) = \hat{D}^{-1/2}(q)S(q)\hat{L}, \quad \text{where} \quad \hat{\Lambda} = \hat{L}\hat{L}'$$

and define the random function

$$\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big) = \min_{\nu \geq 0} \left\| M_{\hat{\xi}}(q)\hat{A}'(q)Z_m - \begin{pmatrix} 0_{k(q)} \\ \nu \end{pmatrix} \right\|^2_{M_{\hat{\xi}}(q)\hat{B}(q)M'_{\hat{\xi}}(q)}, \tag{40}$$

where ν is a $\hat{r}_1(q) \times 1$ vector. Note that $k(q)$ or $\hat{r}_1(q)$ could be zero, in which case $\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big)$ simplifies. We adopt the convention that $\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big) = 0$ if $k(q) = \hat{r}_1(q) = 0$. The critical value $c_P^\alpha(q)$ is defined as

$$c_P^\alpha(q) = 1 - \alpha \text{ quantile of } \bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big) \tag{41}$$

and can be simulated by generating draws of Z_m and evaluating $\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big)$. The evaluation simplifies considerably if $\hat{B}(q) = I$ because the quadratic programming problem has the trivial solution

$$\begin{aligned} \bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big) &= \sum_{j=1}^{k(q)} [M_{\hat{\xi}}(q)\hat{A}'(q)Z_m]_j^2 \\ &\quad + \sum_{j=k(q)+1}^{k(q)+\hat{r}_1(q)} [M_{\hat{\xi}}(q)\hat{A}'(q)Z_m]_j^2 \mathcal{I}\big\{[M_{\hat{\xi}}(q)\hat{A}'(q)Z_m]_j < 0\big\}. \end{aligned} \tag{42}$$

4.5 Bonferroni Confidence Set

The Bonferroni approach consists of three steps. In the first step a $(1-\alpha_1)$ confidence set for q is constructed. In the second step, a $1-\alpha_2$ confidence set for θ is constructed conditional on q. The final step consists of taking a union of the conditional confidence sets to obtain a confidence set for θ with coverage probability of at least $1-\alpha$, where $\alpha = \alpha_1 + \alpha_2$.

Step 1. The confidence set for q can essentially be obtained by projecting the joint confidence set $CS_P^{\theta,q}$ in (33) onto the q-ordinate. More specifically, we assume without loss of generality that $\tilde{S}_R(q)$ is re-defined such that $\tilde{S}_R(q)\phi \geq 0$ summarizes *all* the sign restrictions that the structural impulse responses have to satisfy. Then we let

$$G^q\big(q; \hat{\phi}, \hat{W}(\cdot)\big) = \min_{\tilde{\mu} \geq 0} \; T \left\| \hat{D}_R^{-1/2}(q) S_R(q)\hat{\phi} - \hat{D}_R^{-1/2}(q) V_R(q)\tilde{\mu} \right\|^2_{\hat{B}_R(q)}. \tag{43}$$

The objective function $G^q(\cdot)$ closely resembles $G(\cdot)$ in (36) with the exception that θ disappeared and all matrices are replaced by submatrices that correspond to the sign restrictions $S_R(q)\phi \geq 0$. The confidence interval for q is given by the generalized level set

$$CS_B^q = \left\{ q \;\middle|\; \Big(G^q\big(q; \hat{\phi}, \hat{W}(\cdot)\big) - c_B^{\alpha}(q) \Big) \leq 0 \right\}. \tag{44}$$

As in Section 4.4, the critical value is based on Andrews and Soares' (2010a) moment selection approach. Using the selection criterion (38), we define the $\hat{r}_1(q) \times r(q)$ selection matrix $M_{R,\hat{\xi}}(q)$ that deletes rows of $\hat{D}_R^{-1/2}(q) S_R(q)\hat{\phi}$ that correspond to non-binding inequality conditions. Moreover, let

$$Z_m \sim N(0, I_m) \quad \text{and} \quad \hat{A}'_R(q) = \hat{D}_R^{-1/2}(q) S_R(q) \hat{L}, \quad \text{where} \quad \hat{\Lambda} = \hat{L}\hat{L}'$$

and define the random function

$$\bar{\mathcal{G}}^q\big(q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big) = \min_{\nu \geq 0} \; \left\| M_{R,\hat{\xi}}(q)\hat{A}'_R(q) Z_m - \nu \right\|^2_{M_{R,\hat{\xi}}(q)\hat{B}_R(q)M'_{R,\hat{\xi}}(q)}, \tag{45}$$

where ν is an $\hat{r}_1(q) \times 1$ vector. As before, we adopt the convention that $\bar{\mathcal{G}}\big(\theta, q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big) = 0$ if $\hat{r}_1(q) = 0$. The critical value $c_B^{\alpha_1}(q)$ is defined as

$$c_B^{\alpha_1}(q) = 1 - \alpha_1 \text{ quantile of } \bar{\mathcal{G}}^q\big(q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big). \tag{46}$$

Step 2. We now construct a confidence interval for θ conditional on q. Since θ is point-identified conditional on q, we can use a standard Wald confidence set that is based on

$$\sqrt{T}(S_\theta(q)\hat{\phi} - \theta) = \sqrt{T}S_\theta(q)(\hat{\phi} - \phi) \Longrightarrow N\big(0, S_\theta(q)\Lambda S_\theta'(q)\big). \tag{47}$$

If we define $\hat{\Sigma}_\theta = S_\theta(q)\hat{\Lambda}S_\theta'(q)$, then the Wald set is given by

$$CS_{B,q}^\theta = \left\{\theta \;\middle|\; M_\theta\theta \geq 0 \text{ and } \|\sqrt{T}(S_\theta(q)\hat{\phi} - \theta)\|^2_{\hat{\Sigma}_\theta^{-1}} \leq c_{B,q}^{\alpha_2}\right\}, \tag{48}$$

where

$$c_{B,q}^{\alpha_2} = 1 - \alpha_2 \text{ quantile of } \chi^2_{k(q)}. \tag{49}$$

Step 3. The Bonferroni confidence set for θ is obtained by taking the union

$$CS_B^\theta = \bigcup_{q \in CS_B^q} CS_{B,q}^\theta. \tag{50}$$

4.6 Computational Issues

Since error bands in the VAR literature predominantly depict pointwise confidence sets, we focus on the computation of $F^\theta(\hat{\phi})$ as well as the confidence sets CS_P^θ and CS_B^θ for $\tilde{k} = 1$. We now describe how the computations are implemented for the empirical analysis presented in Section 8. First, the reduced-form parameter vector ϕ is estimated. Second, a grid for q is generated. Third, we approximate the sets $F^q(\hat{\phi})$ and $F^\theta(\hat{\phi})$. Finally, we describe how to compute the Bonferroni and the projection-based confidence sets. If the goal is to construct pointwise error bands for impulse responses, then the computations for the confidence sets have to be repeated for every response $\partial y_{i,t+h}/\partial \epsilon_{1,t}$ of interest. Here i potentially ranges from $i = 1, \dots, n$ and $h = 0, 1, \dots, h_{max}$.

Estimating the Reduced-Form Parameters. The VAR coefficient matrices $\Phi_1, \dots, \Phi_p$ and Σ_u can be estimated by OLS. An estimate of Σ_{tr} is obtained by applying the Cholesky decomposition to $\hat{\Sigma}_u$. We obtain $\hat{\Lambda}$ by using a parametric bootstrap procedure. Conditional on $\hat{\Phi}_1, \dots, \hat{\Phi}_p, \hat{\Sigma}_u$ we simulate bootstrap samples $Y_{1:T}^*$ from the VAR in (20). From each bootstrap sample we compute

$\hat{\phi}^*$. Finally we compute the bootstrap sample covariance matrix of $\hat{\phi}^*$ and scale it appropriately to obtain $\hat{\Lambda}$.

Generating a Grid for q. The vector q takes values on an n-dimensional hypersphere. We generate the grid points for q from a distribution that is uniform on the hypersphere using a well-known result by James (1954). Let $Z_{(s)}$, $s = 1, \ldots, s_{max}$, be a sequence of $n \times 1$ vectors of $iidN(0,1)$ random variables and define $q_{(s)} = Z_{(s)}/\|Z_{(s)}\|$. Then, q_s is uniformly distributed on the unit-hypersphere defined by $\|q\| = 1$. Thus, we define the grid as $\mathcal{Q} = \{q_{(1)}, \ldots, q_{(s_{max})}\}$.

Approximating $F^q(\hat{\phi})$. Suppose that $\tilde{S}_R(q)$ is defined such that the full set of sign restrictions is given by $\tilde{S}_R(q)\phi \geq 0$. Then,

$$F^q(\hat{\phi}) \approx \hat{F}^q(\hat{\phi}) = \big\{q \in \mathcal{Q} \mid \tilde{S}_R(q)\hat{\phi} \geq 0\big\}.$$

Thus, for every $q \in \mathcal{Q}$ one checks whether $\tilde{S}_R(q)\hat{\phi} \geq 0$ and retains the q's for which the condition is satisfied.

Approximating $F^\theta(\hat{\phi})$. Denote the elements of $\hat{F}^q(\hat{\phi})$ by $q^{(j)}$, $j = 1, \ldots, n_q$. Compute $\theta^{(j)} = \tilde{S}_\theta(q^{(j)})\hat{\phi}$. We show in the Online Appendix that $F^\theta(\hat{\phi})$ is a bounded interval. Thus, we define the interval

$$\hat{F}^\theta(\hat{\phi}) = \big[(\min_{j=1,\ldots,n_q} \theta^{(j)}),\ (\max_{j=1,\ldots,n_q} \theta^{(j)})\big].$$

Bonferroni Interval. To obtain CS_B^q in Step 1 we verify for each $q \in \mathcal{Q}$ whether $G^q\big(q; \hat{\phi}, \hat{W}\big) - c_B^{\alpha_1}(q) < 0$. $\hat{F}^q(\hat{\phi}) \subset CS_B^q$ because $G^q\big(q; \hat{\phi}, \hat{W}\big) = 0$ for each $q \in \hat{F}^q(\hat{\phi})$.[8] For each value of $q \in \mathcal{Q}$ the critical value $c_B^{\alpha_1}(q)$ can be obtained from a simulation approximation of the limit objective function $\mathcal{G}^q\big(q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big)$ in (45). More specifically, for $j = 1, \ldots, n_z$ generate random vectors $Z_m^{(j)}$ and compute $\mathcal{G}^{q(j)}\big(q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big)$. The critical value can be approximated by the $1 - \alpha_1$ percentile of the empirical distribution of $\mathcal{G}^{q(j)}\big(q; \hat{B}_R(q), M_{R,\hat{\xi}}(q)\big)$. If $\hat{B}_R(q) = I$ then the evaluation of $\mathcal{G}^q(\cdot)$ is fast because (42) provides a closed-form solution for the quadratic programming problem. If $\hat{B}_R(q) = \hat{\Omega}_R^{-1}$, then the quadratic programming problem has to be solved numerically, which considerably slows down the computation of the critical values.

The computation of $CS_{B,q}^\theta$ in (48) in Step 2 is straightforward. The interval CS_B^θ in Step 3 is obtained by taking the minimum with respect to $q \in \mathcal{Q}$ of the left endpoints of $CS_{B,q}^\theta$ and the maximum of the right endpoints of $CS_{B,q}^\theta$. Note that CS_B^q in Step 1 only has to be computed once,

[8]In general, the evaluation of $G^q(\cdot)$ involves solving a standard quadratic programming problem for which efficient algorithms are readily available in software packages such as MATLAB, GAUSS, or R.

whereas Steps 2 and 3 have to be repeated for each choice of θ if the goal is to obtain pointwise confidence intervals for multiple impulse responses.

Projection Interval. The computation of CS_P^θ requires pointwise testing for $\theta \in \Theta$. By construction, $F^\theta(\hat{\phi}) \subset CS_P^\theta$. Thus, we compute the confidence interval for θ by expanding the boundaries of $\hat{F}^\theta(\hat{\phi})$ in a stepwise fashion. Our description focuses on the computation of the upper bound. Set $\theta_{(0)}$ equal to the upper bound of $\hat{F}^\theta(\hat{\phi})$ and choose the step size δ_θ as a fraction of the length of $\hat{F}^\theta(\hat{\phi})$. For $j = 0, 1, \ldots$, let $\theta_{(j)} = \theta_{(j-1)} + \delta_\theta$ and evaluate the criterion function

$$\min_{q \in \mathcal{Q}} \left(G\big(\theta_{(j)}, q; \hat{\phi}, \hat{W}(\cdot)\big) - c_P^\alpha(q) \right).$$

We terminate the iterations at j^* and set the upper bound of CS_P^θ to $\theta_{(j^*-1)}$ when the criterion function becomes bigger than zero. The projection is carried out by minimization over the grid $\mathcal{Q}$. The critical values $c_P^\alpha(q)$ are obtained via simulation of $\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big)$ defined in (40) as in Step 2 of the computation of the Bonferroni interval.

Simulating critical values based on the non-diagonal weight matrix $\hat{B}(q) = \hat{\Omega}^{-1}(q)$ is computationally costly. To expedite the computations one could use the boundaries of a confidence interval based on $\hat{B}(q) = I$ as starting points and then contract or expand these boundaries. Such a two-step procedure speeds up the computations if the difference between the confidence intervals based on the two weight matrices is small. If the goal is to obtain pointwise confidence intervals for multiple impulse responses, the computations have to be repeated for each choice of θ.

5 Large Sample Results

This section formally establishes that the plug-in estimators $F^\theta(\hat{\phi})$ and $F^q(\hat{\phi})$ are consistent for the identified sets $F^\theta(\phi)$ and $F^q(\phi)$ and that CS_P^θ and CS_B^θ are asymptotically valid confidence sets. In general the vector ϕ is not sufficient to characterize the distribution of the data and hence the distribution of an estimator $\hat{\phi}$. For this reason we introduce some additional notation. Let ρ be the parameter vector with domain $\bar{\mathcal{R}}$ that characterizes the reduced-form VAR model (20) and let $\mathcal{R} \subset \bar{\mathcal{R}}$. For instance, if the innovations u_t are Gaussian, then $\rho = [vec(\Phi_1), \ldots, vec(\Phi_p), vech(\Sigma_u)]'$ and $\bar{\mathcal{R}} = \mathbb{R}^{n^2p+n(n+1)/2}$ is the $n(np)+n(n+1)/2$-dimensional Euclidian space. For reasons explained below we also define the δ-inflated open set

$$\mathcal{R}^\delta = \big\{ \tilde{\rho} \in \bar{\mathcal{R}} \;\big|\; \exists \rho \in \mathcal{R} \text{ s.t.} \|\tilde{\rho} - \rho\| < \delta \big\}. \tag{51}$$

The vector ϕ can be expressed as a smooth function of ρ and we express the validity condition for a confidence set for θ as

$$\liminf_T \inf_{\rho \in \mathcal{R},\, \theta \in F^\theta(\phi(\rho))} P_\rho\{\theta \in CS^\theta\} \geq 1 - \alpha.$$

We are now in a position to state our high-level assumptions.

Assumption 1 *There exists a compact reduced-form parameter set $\mathcal{R}$ and a δ-inflated superset $\mathcal{R}^\delta \subset \bar{\mathcal{R}}$ defined in (51) such that:*

(i) For every $\rho \in \mathcal{R}^\delta$, there does not exist an $\tilde{l} \times 1$ vector $\lambda > 0$ such that

$$\lambda'\big(M^{S,2} M^{S,1} R^v\big) = 0.$$

(ii) $\phi(\rho)$ is continuously differentiable for all $\rho \in \mathcal{R}^\delta$.

(iii) There exists an estimator $\hat{\phi}_T$ of $\phi(\rho_T)$ and a matrix $\Lambda^{-1/2}(\rho_T)$ such that for each sequence $\{\rho_T\} \in \mathcal{R}$ (a) $\|\hat{\phi}_T - \phi(\rho_T)\| \xrightarrow{p} 0$; (b) $\sqrt{T}\Lambda^{-1/2}(\rho_T)(\hat{\phi}_T - \phi(\rho_T)) \Longrightarrow N(0, I)$.

(iv) For each $\rho \in \mathcal{R}$ the matrix $\Lambda(\rho)$ is continuous, positive definite, and there exists a full-rank positive-definite matrix Λ_{min} such that $\Lambda(\rho) - \Lambda_{min} \geq 0$ for all $\rho \in \mathcal{R}$.

(v) There exists an estimator $\hat{\Lambda}_T$ of $\Lambda(\rho_T)$ such that $\|\hat{\Lambda}_T - \Lambda(\rho_T)\| \xrightarrow{p} 0$ for any converging sequence $\{\rho_T\}$.

Condition (i) of Assumption 1 states that the convex cone generated by the rows of the reduced-form impulse response matrix $\lambda'\big(M^{S,2} M^{S,1} R^v\big) > 0$ does not contain the zero vector. This assumption is sufficient to ensure that the identified sets $F^\theta(\phi(\rho))$ and $F^q(\phi(\rho))$ are non-empty and that the plug-in estimators $F^\theta(\hat{\phi}_T)$ and $F^q(\hat{\phi}_T)$ are consistent whenever $\hat{\phi}_T \xrightarrow{p} \phi_0$ (see Theorem 1 below). Assumption 1(i) rules out, for instance, that equality conditions are coded as pairs of inequalities, and, more generally, that linear combinations of inequalities constrain impulse responses to be equal to zero. We discuss in Section 6 how our framework can be extended to allow for a mixture of inequality and equality restrictions on impulse responses.

Condition (i) is typically not satisfied for all values of the reduced-form parameter $\rho \in \bar{\mathcal{R}}$, which is why we only require it to hold on the set $\mathcal{R}^\delta \subset \bar{\mathcal{R}}$. Consider a VAR(1) generalization of the example in Section 2 with autoregressive coefficient matrix Φ_1. As before, suppose y_t is composed of inflation and output growth and the investigator imposes the sign restriction that in response to a (positive) demand shock inflation and output responses are both non-negative upon impact and one period after impact. In this case

$$\left(M^{S,2}M^{S,1}R^{v}\right)=\begin{bmatrix}\Sigma_{tr}\\ \Phi_1\Sigma_{tr}\end{bmatrix}.$$

If $\Phi_1 = diag(\rho_1, \rho_2)$ and $\rho_1, \rho_2 < 0$, then Condition (i) is violated. Conditional on these reduced-form parameters, the identified set is empty. Assumption 1 excludes these values of ρ from $\mathcal{R}^{\delta}$. From a practitioner's perspective, empty confidence sets CS_B^{θ} and CS_P^{θ} provide evidence that the imposed sign restrictions are inconsistent with the estimated reduced-form parameters.

Conditions 1(iii) and (iv) require that $\hat{\phi}_T$ and $\hat{\Lambda}_T$ converge uniformly for $\rho \in \mathcal{R}$. The uniform convergence of $\hat{\phi}_T$ to a Gaussian limit distribution also requires a restriction of the domain of ρ because it breaks down at the boundary of the stationary region in the VAR parameter space. For instance, in the context of an AR(1) model $y_t = \rho_T y_{t-1} + u_t$ with autoregressive coefficient $\rho_T = 1 - c/T$, an estimator of an impulse response at horizon $h = 1$, that is, $\phi(\rho) = \rho$, behaves according to

$$\sqrt{T}\left(1-\rho_T^2\right)^{-1/2}(\hat{\phi}_T-\rho_T)=\frac{\frac{1}{T}\sum y_{t-1}u_t}{\sqrt{c(2-c/T)\frac{1}{T^2}\sum y_{t-1}^2}}\not\Longrightarrow N(0,1).$$

Uniform convergence to a Gaussian limit distribution can be achieved if $\mathcal{R}$ is restricted to the interval $[-1+\epsilon,\ 1-\epsilon]$ for some $\epsilon > 0$.[9] From a practitioner's perspective we are essentially assuming that the researcher has applied some stationarity-inducing transformations, e.g., transformed prices into inflation rates. However, since some authors, e.g., Uhlig (2005), prefer to specify VARs in terms of variables that exhibit (near) non-stationary dynamics, our Monte Carlo experiments in Section 7 include designs in which the roots of the vector autoregressive lag polynomial are close to the unit circle.[10]

Our first theorem establishes that the identified sets are non-empty and that they are not singletons, that is, the impulse responses are set-identified. This result can be deduced from Assumption 1(i) using Gordan's Alternative Theorem (see, for instance, Border (2007)). Moreover, the theorem establishes the consistency of the plug-in estimators $F^{\theta}(\hat{\phi})$ and $F^{q}(\hat{\phi})$. The consistency is stated in terms of the Haussdorf distance. We denote the Hausdorff distance between two sets

[9]See Giraitis and Phillips (2004) for a more general discussion.

[10]An extension of our analysis to VARs with unit roots or cointegration restrictions is beyond the scope of this paper. The construction of uniformly valid confidence intervals for reduced-form parameters in itself is a very challenging task; see Mikusheva (2007).

A and B by $d_H(A, B)$.[11] The consistency relies on the compactness of $F^q(\phi)$ and $F^\theta(\phi)$ and the continuity of the correspondences with respect to ϕ. Unlike in some of the models studied by Chernozhukov, Hong, and Tamer (2007), it is not necessary to inflate the sets $F^\theta(\hat{\phi})$ and $F^q(\hat{\phi})$ by $\epsilon_T \downarrow 0$ to achieve consistency. A formal proof of Theorem 1 is provided in the Appendix.

Theorem 1 *Suppose Assumption 1(i) is satisfied.*

(i) For all $\rho \in \mathbb{R}^\delta$ the identified sets $F^\theta(\phi(\rho))$ and $F^q(\phi(\rho))$ are non-empty and they are not singletons.

If, in addition, the domain of ϕ is compact and $\hat{\phi} \xrightarrow{p} \phi_0$, then,

(ii) $d_H\big(F^\theta(\hat{\phi}), F^\theta(\phi_0)\big) \xrightarrow{p} 0$ and

(iii) $d_H\big(F^q(\hat{\phi}), F^q(\phi_0)\big) \xrightarrow{p} 0$.

Our second theorem establishes the validity of the proposed confidence sets. A formal proof is provided in the Appendix and closely follows the proof of Theorem 1 in Andrews and Soares (2010a). However, a number of non-trivial modifications are required to account for the potential rank reduction of $\tilde{S}(q)$ as a function of q.

Theorem 2 *Suppose that Assumption 1 is satisfied. Let $0 < \alpha < 1/2$.*

(i) The confidence set CS_P^θ, defined in (34), is an asymptotically valid confidence set for θ:

$$\liminf_T \inf_{\rho \in \mathcal{R},\, \theta \in F^\theta(\phi(\rho))} P_\rho\{\theta \in CS_P^\theta\} \geq 1 - \alpha.$$

(ii) The confidence set CS_B^θ, defined in (50), is an asymptotically valid confidence set for θ:

$$\liminf_T \inf_{\rho \in \mathcal{R},\, \theta \in F^\theta(\phi(\rho))} P_\rho\{\theta \in CS_B^\theta\} \geq 1 - \alpha.$$

6 Extensions

We now discuss extensions to cumulative impulse response functions, models that use both equality and inequality restrictions to identify structural impulse responses, identification of multiple shocks, and variance decompositions and dynamic correlations.

[11]Formally, the Hausdorff distance is defined as $d(A, B) = \max\ \{d(A|B),\ d(B|A)\}$, where $d(A|B) = \sup_{a \in A}\ d(a, B)$ and $d(a, B) = \inf_{b \in B}\ \|a - b\|$. We set $d(A, B) = \infty$ if either A or B is empty.

Cumulative Impulse Responses. As in Section 2, consider a bivariate VAR composed of inflation and output growth, but now with non-trivial dynamics. 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 and 1. This case can be handled by redefining ϕ to include the cumulative responses.

Sign Restrictions Combined with Zero Restrictions. Assumption 1(i) rules out that opposing sign restrictions are used to represent equality restrictions on impulse responses. Nonetheless, it is straightforward to sharpen the identified set by combining sign restrictions with more traditional exclusion restrictions. Zero restrictions, e.g., on the impact or the long-run effect of a shock, could be imposed in one of two ways. In some applications, e.g., the one considered in Section 8, the zero restrictions easily translate into domain restrictions for q. Alternatively, one can modify the function $\tilde{G}(\theta, q; \phi, \tilde{W})$ in (26) as follows:

$$\tilde{G}(\theta, q; \phi, \tilde{W}) = \min_{\mu \geq 0} \left\| \begin{pmatrix} \begin{bmatrix} \tilde{S}_\theta(q) \\ \tilde{S}_{eq}(q) \end{bmatrix} \phi - \begin{bmatrix} \theta \\ 0 \end{bmatrix} \\ \tilde{S}_R(q)\phi - \mu \end{pmatrix} \right\|^2_{\tilde{W}}, \qquad (52)$$

where $\tilde{S}_{eq}(q)\phi$ corresponds to the responses that are restricted to be zero.[12] The projection-based confidence set is based on a pointwise testing procedure that conditions on $\theta \in F^\theta(\phi)$. Thus, the inclusion of the equality restrictions essentially amounts to augmenting $\tilde{S}_\theta(q)$ by $\tilde{S}_{eq}(q)$ and θ by a vector of zeros. For the Bonferroni confidence set, one needs to combine the inequality conditions $\tilde{S}_R(q)\phi \geq 0$ with the equality moment conditions $\tilde{S}_{eq}(q)\phi = 0$ and generalize the criterion function $G^q\big(q; \hat{\phi}, \hat{W}(\cdot)\big)$ in (43) accordingly. Theorem 2 can be extended to cover the objective function (52).

Identifying Multiple Shocks. Some authors use sign-restricted SVARs to identify multiple shocks simultaneously. For instance, Peersman (2005) considers an $n = 4$ 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

$$\tilde{S}_\theta(\Omega)\phi = \theta \quad \text{and} \quad \tilde{S}_R(\Omega)\phi \geq 0$$

[12] If we denote the matrix of zero-restricted orthogonalized responses by R^v_{eq} then the generalization of Assumption 1(i) is: there do not exist vectors $\lambda > 0$ and $\lambda_{eq} \geq 0$ such that $\lambda'(M^{S,2}M^{S,1}R^v) + \lambda'_{eq}R^v_{eq} = 0$. The generalized analysis would use Motzkin's Transposition Theorem; see Border (2007).

for suitably defined functions $\tilde{S}_\theta(\Omega)$ and $\tilde{S}_R(\Omega)$. While all our results easily generalize to multiple shocks (just replace q by Ω), the implementation becomes computationally more difficult because the grid for the $n-1$ dimensional vector q has to be replaced by a grid for orthogonal matrix Ω, which has $n(n-1)/2$ degrees of freedom.

Bootstrapped Critical Values Instead of Asymptotic Critical Values. Our simulated critical values rely on the Gaussian limit distribution of $\sqrt{T}\hat{D}^{-1/2}(q)S(q)(\hat{\phi}-\phi)$, which is reflected in the vector $\hat{A}'(q)Z_m$ in the random function $\bar{\mathcal{G}}(\cdot)$ in (45). Alternatively, the critical values could be constructed by replacing draws from $\hat{A}'(q)Z_m$ with draws from the bootstrap approximation of $\sqrt{T}\hat{D}^{-1/2}(q)S(q)(\hat{\phi}-\phi)$. Bootstrap procedures for VAR impulse response functions are discussed, for instance, in Kilian (1998).

7 Monte Carlo Illustrations

In this section we conduct some Monte Carlo experiments to illustrate the properties of our proposed confidence sets. We consider the projection-based and the Bonferroni confidence sets introduced in Sections 4.4 and 4.5 with weight matrices based on $\hat{B}(q)=I$ and $\hat{B}(q)=\hat{\Omega}^{-1}(q)$. These sets are denoted by $CS_P^\theta(I)$, $CS_P^\theta(\hat{\Omega}^{-1})$, $CS_B^\theta(I)$, and $CS_B^\theta(\hat{\Omega}^{-1})$, respectively. As part of the calculations for the Bonferroni confidence intervals, we also obtain confidence sets for q: $CS^q(\hat{\Omega}^{-1})$ and $CS^q(I)$. Finally, we compute coverage probabilities for the Wald confidence set CS^ϕ for the reduced-form parameter vector ϕ.

The first experiment is based on the VAR(0) model analyzed in Section 2. For the second set of experiments we introduce 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, summarized in Table 1, are obtained by fitting a VAR(0) to data on U.S. inflation and GDP growth (Section 7.1) and fitting VAR(1)s to inflation and either output growth or linearly detrended log GDP (Section 7.2). While we initially impose the sign restriction for only one horizon, we subsequently impose restrictions over multiple horizons (Section 7.3). Each Monte Carlo experiment involves the steps summarized in Table 2, which are repeated $n_{sim}=1,000$ times.

The computations of the frequentist confidence intervals are implemented as described in Section 4.6. Further details are provided in the Online Appendix. We consider sample sizes of $T=100$ and $T=500$. The grid $\mathcal{Q}$ for q is obtained as follows: q is transformed into polar coordinates

Table 1: Monte Carlo Design

	Design 1 VAR(0)	Design 2 VAR(1)	Design 3 VAR(1)	Design 4 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 $y_t = \Phi_0 + \Phi_1 y_{t-1} + u_t$, $I\!E[u_t u_t'] = \Sigma_u$. We use OLS estimates, Φ entries refer to elements of Φ_1 and Σ_{ij} entries refer to the (non-redundant) elements of Σ_u. $\lambda_i(\Phi_1)$ is the i'th eigenvalue of Φ_1. $y_{1,t}$ is the log difference of the U.S. GDP deflator, scaled by 100 to convert into percentages. $y_{2,t}$ is either the 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 2006:IV.

$[\cos(\varphi), \sin(\varphi)]'$ and we choose $s_{max} = 629$ equally spaced grid points for φ on the interval $(-\pi, \pi]$. The number of bootstrap repetitions to obtain $\hat{\Lambda}$ is $B = 1000$ and the number of simulations to obtain the critical values $c_P^{\alpha}(q)$ and $c_B^{\alpha_1}(q)$ is $n_z = 500$. To construct the projection-based confidence set we expand $\hat{F}^{\theta}(\hat{\phi})$ with steps of size $\delta_\theta = 0.001$. We report the coverage probability for the least favorable parameter in the identified set, i.e., the parameter with the lowest coverage probability. Moreover, we report the average length of the confidence intervals.

Table 2: Steps of Monte Carlo Experiments

1.	Generate a sample of size T from the data-generating process.
2.	Compute $\hat{\phi}$ and the bounds of $F^{\theta}(\hat{\phi})$.
3.	Compute $\hat{\Lambda}$ using a parametric bootstrap approach.
4.	Compute the bounds of the 90% frequentist confidence intervals $CS^{\theta}_{P}(I)$, $CS^{\theta}_{P}(\hat{\Omega}^{-1})$, $CS^{\theta}_{B}(I)$, and $CS^{\theta}_{P}(\hat{\Omega}^{-1})$.

7.1 Experiment 1

The parameterization of the data-generating process (DGP) $y_t \sim iidN(0, \Sigma_u)$ is provided in Table 1 in the column labeled *Design 1*. We define θ as the response of $y_{1,t}$ to $\epsilon_{1,t}$. According to our simulation design, $\phi_2 < 0$ and the identified set for θ is $F^{\theta}(\phi_0) = [0, 0.578]$. The geometry of the Monte Carlo design, illustrated in Figure 1, implies that the lower bounds of $F^{\theta}(\hat{\phi})$, $CS^{\theta}_{P}(\cdot)$, and $CS^{\theta}_{B}(\cdot)$ are zero. As a consequence, we focus on the upper bounds of the confidence sets and the frequency with which they exceed (and hence cover) the least favorable parameter value in the identified set: $\theta = 0.578$.

Detailed results for the frequentist confidence intervals are summarized in Table 3. The results obtained under $\hat{B} = \hat{\Omega}^{-1}$ and $\hat{B} = I$ are almost identical for the projection-based confidence intervals and exactly identical for the Bonferroni confidence sets. Recall that the nominal coverage probability for θ is 90%. For $T = 100$ the actual coverage probability for the projection-based sets is about 93% and for the Bonferroni sets it is 0.97. Accordingly, the Bonferroni intervals are on average slightly longer than the projection-based intervals. As we increase the sample size to $T = 500$ the length of the confidence intervals shrinks, while the relative ranking between the Bonferroni and the projection-based intervals remain unchanged. The actual coverage probabilities increase compared to $T = 100$.

It is instructive to also examine the coverage probabilities of CS^{ϕ} and CS^{q}. The coverage probability for the reduced-form parameter vector ϕ is 86% for $T = 100$ and reaches its nominal value of 90% as the sample size is increased to $T = 500$. This increase in coverage probability for ϕ mirrors the increase in coverage probability for θ. The Bonferroni intervals are computed based on $\alpha_1 = \alpha_2 = 0.05$, which implies that the nominal coverage probability of CS^{q} is 95%. The actual coverage probabilities for the nuisance parameter vector q are slightly smaller, namely, around

Table 3: Monte Carlo Results (I): 90% Nominal Coverage Probability, Sign Restrictions at Horizon $h = 1$

	Design 1		Design 2		Design 3		Design 4	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
$F^{\theta}(\phi)$		0.578		0.232		0.226		0.094
$F^{q}(\phi)$		$\frac{42}{100}\pi$		$\frac{35}{100}\pi$		$\frac{47}{100}\pi$		$\frac{50}{100}\pi$
Sample Size $T = 100$								
$CS_P^{\theta}(\hat{\Omega}^{-1})$	0.929	0.651	0.926	0.281	0.928	0.264	0.959	0.129
$CS_P^{\theta}(I)$	0.946	0.656	0.954	0.285	0.934	0.265	0.960	0.129
$CS_B^{\theta}(\hat{\Omega}^{-1})$	0.972	0.668	0.988	0.296	0.945	0.267	0.963	0.129
$CS_B^{\theta}(I)$	0.972	0.668	0.988	0.296	0.945	0.267	0.963	0.129
$CS_B^{q}(\hat{\Omega}^{-1})$	0.932	$\frac{47}{100}\pi$	0.941	$\frac{81}{100}\pi$	0.925	$\frac{57}{100}\pi$	0.942	$\frac{66}{100}\pi$
$CS_B^{q}(I)$	0.932	$\frac{47}{100}\pi$	0.941	$\frac{81}{100}\pi$	0.925	$\frac{57}{100}\pi$	0.942	$\frac{66}{100}\pi$
CS^{ϕ}	0.860		0.858		0.879		0.876	
Sample Size $T = 500$								
$CS_P^{\theta}(\hat{\Omega}^{-1})$	0.965	0.615	0.964	0.257	0.958	0.244	0.959	0.111
$CS_P^{\theta}(I)$	0.971	0.617	0.976	0.259	0.958	0.244	0.959	0.111
$CS_B^{\theta}(\hat{\Omega}^{-1})$	0.990	0.622	0.996	0.266	0.969	0.245	0.959	0.111
$CS_B^{\theta}(I)$	0.990	0.622	0.996	0.266	0.969	0.245	0.959	0.111
$CS_B^{q}(\hat{\Omega}^{-1})$	0.931	$\frac{44}{100}\pi$	0.950	$\frac{44}{100}\pi$	0.939	$\frac{51}{100}\pi$	0.938	$\frac{56}{100}\pi$
$CS_B^{q}(I)$	0.931	$\frac{44}{100}\pi$	0.950	$\frac{44}{100}\pi$	0.939	$\frac{51}{100}\pi$	0.938	$\frac{56}{100}\pi$
CS^{ϕ}	0.903		0.903		0.892		0.901	

Notes: Length refers to the average length of the confidence intervals across Monte Carlo repetitions. For $F^{q}(\phi)$ and CS_B^{q} we report the arc length, see illustration; in Figure 1. For the Bonferroni confidence intervals we let $\alpha_1 = \alpha_2 = 0.05$, which implies that the nominal coverage probability of CS_B^{q} is 95%.

93%. Overall, the projection of the joint confidence set for (θ, q) as well as the Bonferroni-type marginalization generates conservative confidence intervals for θ.

7.2 Experiment 2

We now add first-order autoregressive terms to the simulation design to introduce persistence in the endogenous variables:

$$y_t = \Phi_1 y_{t-1} + u_t, \quad u_t \sim iidN(0, \Sigma_u).$$

The choices for Φ_1 and Σ_u 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 $\phi = \text{vec}\big((\Phi\Sigma_{tr})'\big)$. 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:

$$\theta = q_1\phi_1 + q_2\phi_2 \geq 0 \quad \text{and} \quad q_1\phi_3 + q_2\phi_4 \geq 0.$$

To simplify the computations, in particular the evaluation and minimization of the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big)$, for now we do not impose sign restrictions on the responses at impact or at horizons greater than $h = 1$. The geometry of the identified set $F^\theta(\phi, q)$ and its projections is similar to the geometry depicted in Figure 1. The main difference is that the second boundary of the identified set is given by the solution of $q_1\phi_1 + q_2\phi_2 = 0$ and $\|q\| = 1$, instead of $q = [0, 1]'$. Overall, the results are qualitatively similar to those for *Design 1*. The actual coverage probabilities of the confidence set for θ range from 0.926 to 0.996 and thereby exceed the nominal coverage level. For *Designs 2* and *3* the projection-based confidence sets are shorter than the Bonferroni sets and their coverage probabilities are closer to 90%. For *Design 4* the two types of confidence intervals essentially have the same average length.

7.3 Experiment 3

Finally, we consider sign restrictions imposed over multiple horizons using the VAR(1) designs of Section 7.2. As in Experiment 1, we define θ as the contemporaneous impact of the shock on $y_{1,t}$: $\theta = \partial y_{1,t}/\partial \epsilon_{1,t}$. However, unlike before, the sign restrictions $\partial y_{i,t+h}/\partial \epsilon_{1,t} \geq 0$ for both variables $i = 1, 2$ are imposed on multiple horizons: $j = 0, 1, \ldots, H$. This increases the number of inequality conditions. Monte Carlo results are presented in Table 4. For *Design 2* and *3* the identified set $F^\theta(\phi)$ shrinks considerably as the number of inequality conditions is increased, whereas the length stays constant under *Design 4*. We focus on $\hat{B} = I$ because the quadratic programming problem underlying the simulation of critical values has a trivial solution that can be computed quickly. The most interesting aspect of the results is the relative ranking of the projection-based and the Bonferroni confidence intervals. While the projection-based interval tends to be shorter if the number of sign restrictions is small, the ranking reverses under the more persistent designs 3 and 4. If the number of horizons over which the sign restrictions are imposed exceeds $H = 2$ for *Design 3* and $H = 1$ for *Design 4*, then the average length of the Bonferroni interval becomes shorter than the average length of the projection-based interval.

A heuristic explanation for the conservativeness of the projection interval is as follows. Suppose that $V(q) = I$, $B = I$, and none of the sign restrictions involves θ directly. Then the relationship between the objective functions used to construct $CS_P^\theta(I)$ and $CS_B^q(I)$ – see Equations (36) and (43) – is given by

$$G\big(\theta, q; \hat{\phi}, \hat{W}(\cdot)\big) = T\big(\hat{D}_\theta^{-1/2}(S_\theta(q)\hat{\phi} - \theta)\big)^2 + G^q\big(q; \hat{\phi}, \hat{W}(\cdot)\big),$$

where $\hat{D}_\theta^{-1/2}$ is the submatrix of $\hat{D}^{-1/2}$ that corresponds to θ. In the Monte Carlo simulations, the confidence set $CS_B^q(I)$ contains values $\tilde{q}$ such that $G^q\big(\tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big)$ is close to zero, yet most of the moment inequalities are considered to be binding by the moment selection criterion. Therefore, the critical value $c_B^{\alpha_1}(\tilde{q})$ associated with the limit distribution of $G^q\big(\tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big)$ is large, as it involves the $1 - \alpha_1$ quantile of a sum of approximately 2(H+1) squared truncated normal random variables (see Equation (42)). Conditional on $\tilde{q}$ the confidence interval for θ, which is a scalar, is given by values of θ such that $T\big(\hat{D}_\theta^{-1/2}(S_\theta(q)\hat{\phi} - \theta)\big)^2$ is less than the critical value of a χ^2_1 random variable.

Now consider the projection-based confidence set. Since $G\big(\theta, \tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big) \geq G^q\big(\tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big)$ the critical value $c_P^\alpha(\tilde{q})$ is of similar magnitude as $c_B^{\alpha_1}(\tilde{q})$. As a consequence, a value of θ that is several standard deviations away from $S_\theta(\tilde{q})\hat{\phi}$ is required for $G\big(\theta, \tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big)$ to exceed the critical value $c_P^\alpha(\tilde{q})$, because $G^q\big(\tilde{q}; \hat{\phi}, \hat{W}(\cdot)\big)$ is close to zero. Thus, the projection interval may contain θ values that are not contained in the Bonferroni interval. Our simulations as well as the subsequent empirical application suggest using the Bonferroni interval if the number of inequalities is large.

Table 4: Monte Carlo Results (II): 90% Nominal Coverage Probability, Sample Size $T = 100$

	Design 2		Design 3		Design 4	
	Coverage	Length	Coverage	Length	Coverage	Length
			Sign Restrictions $h = 0, 1$			
$F^{\theta}(\phi)$		0.265		0.277		0.209
$CS^{\theta}_{P}(I)$	0.951	0.312	0.948	0.313	0.936	0.234
$CS^{\theta}_{B}(I)$	0.980	0.325	0.965	0.317	0.953	0.235
			Sign Restrictions $h = 0, 1, 2$			
$F^{\theta}(\phi)$		0.137		0.272		0.209
$CS^{\theta}_{P}(I)$	0.962	0.289	0.966	0.316	0.970	0.242
$CS^{\theta}_{B}(I)$	0.976	0.309	0.970	0.317	0.948	0.236
			Sign Restrictions $h = 0, \ldots, 3$			
$F^{\theta}(\phi)$		0.038		0.267		0.209
$CS^{\theta}_{P}(I)$	0.998	0.264	0.991	0.318	0.986	0.249
$CS^{\theta}_{B}(I)$	0.999	0.288	0.991	0.316	0.949	0.236
			Sign Restrictions $h = 0, \ldots, 4$			
$F^{\theta}(\phi)$		0.007		0.262		0.209
$CS^{\theta}_{P}(I)$	0.999	0.248	0.993	0.320	0.994	0.256
$CS^{\theta}_{B}(I)$	1.000	0.271	0.995	0.315	0.949	0.236

Notes: Length refers to the average length of the confidence intervals across Monte Carlo repetitions.

8 Empirical Illustration

We now apply the previously developed methods to a four-variable VAR. The vector of observables consists of per capita real GDP (in deviations from a linear trend), inflation, the federal funds rate, and real money balances. We use quarterly U.S. data from 1965:I to 2006:IV, excluding the most recent recession. A detailed description of the data set is provided in the Online Appendix. All VARs are estimated with $p = 2$ lags. We will consider two set-identification schemes for monetary policy shocks. The first scheme involves only sign restrictions (Section 8.1), whereas the second identification is based on a combination of equality and sign restrictions (Section 8.2).

In addition to computing projection-based and Bonferroni error bands, we also generate pointwise Bayesian credible intervals for the impulse responses, which have been widely used in empirical research. The Bayesian credible sets reported subsequently are based on the VAR(p) given in (20) with Gaussian innovations $u_t \sim iidN(0, \Sigma_u)$. Let $\Phi = [\Phi_1, \ldots, \Phi_p]'$ and define the unnormalized vector $\tilde{q}$ such that $q = \tilde{q}/\|\tilde{q}\|$. If $\tilde{q} \sim N(0, I_n)$, then q is uniformly distributed on the hypersphere. Following Uhlig (2005), we use an improper prior of the form

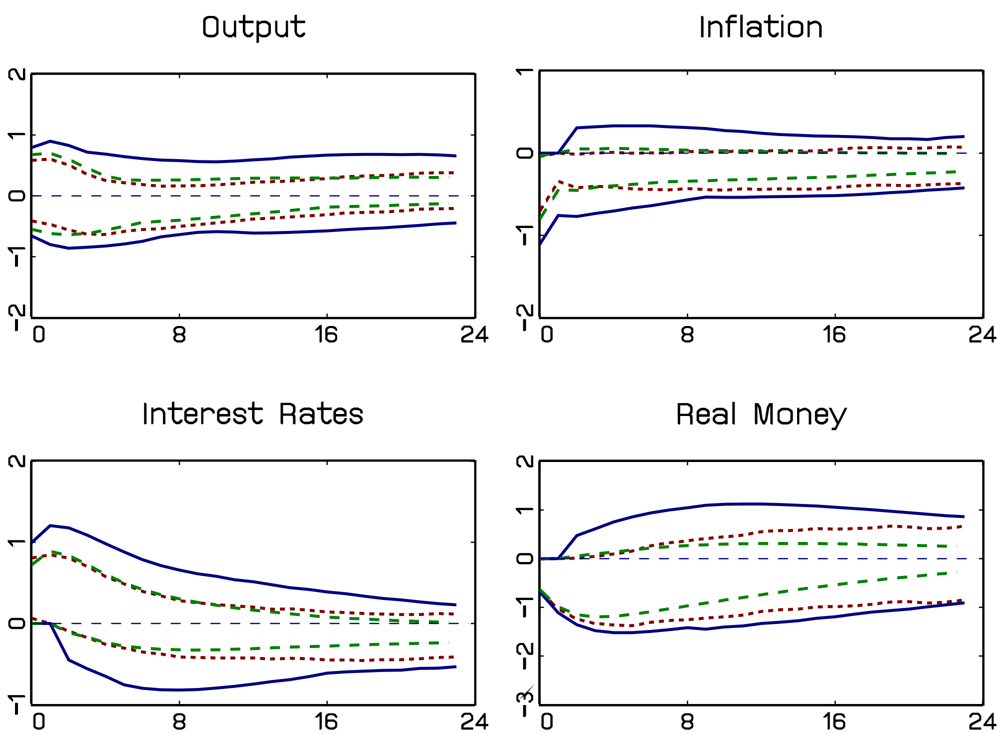
$$p(\Phi, \Sigma, \tilde{q}) \propto |\Sigma|^{-(n+1)/2} \exp\{-\tilde{q}'\tilde{q}/2\} \mathcal{I}\left\{ \frac{\tilde{q}}{\|\tilde{q}\|} \in F^q\big(\phi(\Phi, \Sigma)\big) \right\}. \tag{53}$$

Draws from the posterior distribution of $(\Phi, \Sigma, \tilde{q})$ can be easily generated with the acceptance sampler described in Uhlig (2005). These draws can then be converted into impulse responses and credible sets can be computed from the impulse response draws.

8.1 Pure Sign Restrictions

In order to make inference about the effects of a contractionary monetary policy shock, we use the following sign restrictions to bound the identified set: in periods $h = 0, 1$ (i) the inflation response is non-positive; (ii) the interest rate response is non-negative; and (iii) real money balances do not rise above their steady-state level. Figure 2 depicts three bands: (pointwise) projection-based 90% frequentist confidence intervals (using a diagonal weight matrix) $CS_P^\theta(I)$, estimated sets $F^\theta(\hat{\phi})$, and (pointwise) 90% Bayesian credible sets. The two most notable features of the error bands are that the frequentist error bands (solid) are substantially wider than the Bayesian error bands (short dashes) and that the Bayesian error bands approximately coincide with the estimated set $\hat{F}^\theta(\hat{\phi})$. As explained in detail in Moon and Schorfheide (2012), in a large sample (a sample in which uncertainty about ϕ is small compared with the size of $F^\theta(\phi)$) the Bayesian intervals lie inside the estimated set $F^\theta(\hat{\phi})$ because in the limit essentially all of the probability mass is concentrated on $F^\theta(\hat{\phi})$ 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 $F^\theta(\hat{\phi})$ because it has to have, say, 90% coverage probability for every element of the identified set $F^\theta(\phi)$, including the boundary points. From a substantive perspective, the use of sign restrictions leaves the direction of the output response undetermined.

Figure 2: Impulse Responses Based on Pure Sign Restrictions

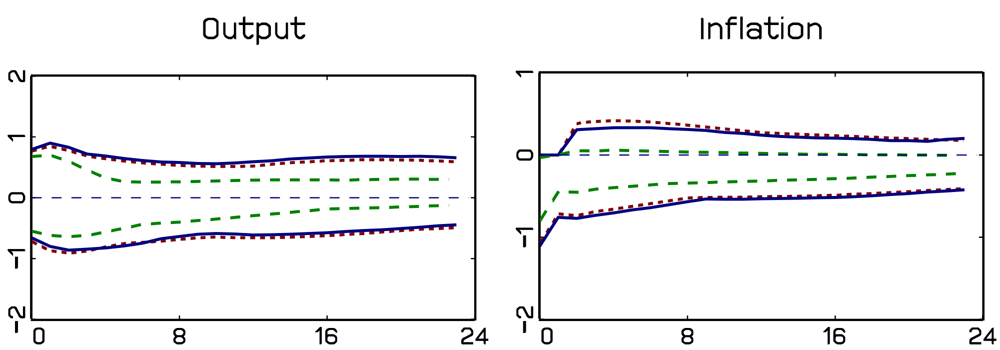


Notes: The figure depicts projection-based 90% frequentist confidence sets $CS_P^\theta(I)$ (blue, solid); 90% Bayesian credible intervals (red, short dashes); and estimated sets $F^\theta(\hat{\phi})$ (green, long dashes).

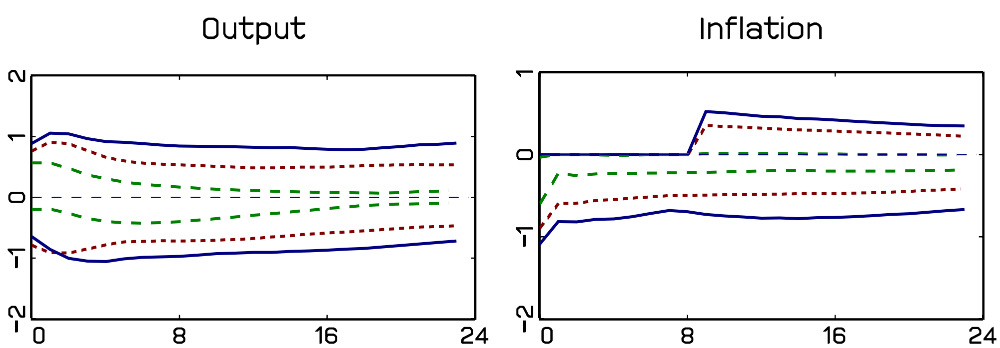
Figure 3 compares the projection-based confidence sets $CS_P^\theta(I)$ of Section 4.4 with the Bonferroni sets $CS_B^\theta(I)$ of Section 4.5. In the top panels of the figure the sign restrictions are imposed over the horizons $h = 0$ and $h = 1$. Here the projection-based and Bonferroni error bands are very similar. In the bottom panels of Figure 3, the sign restrictions are imposed at horizons $h = 0, 1, \ldots, 8$, which increases the number of inequality restrictions from 6 to 27. As the number of sign restrictions increases, the width of the identified sets decreases. Somewhat paradoxically, the width of the projection-based error bands seems to increase. Despite the elimination of non-binding moment inequalities, the projection-based intervals seem to become more conservative. As suggested by the Monte Carlo simulations, the width of the Bonferroni bands decreases noticeably and conditional on the observed data $CS_B^\theta(I)$ is now smaller than $CS_P^\theta(I)$.

Figure 3: Projection-Based versus Bonferroni Error Bands

Sign Restrictions Imposed over Horizons $h = 0, 1$



Sign Restrictions Imposed over Horizons $h = 0, 1, \ldots, 8$



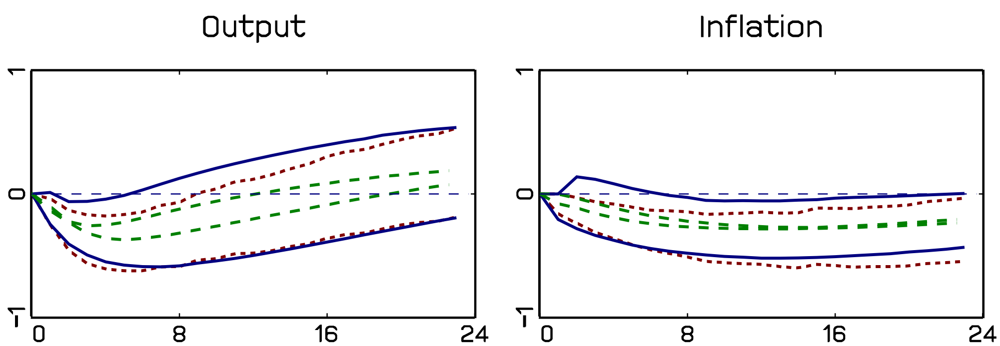
Notes: The figure depicts 90% projection-based confidence sets $CS_P^\theta(I)$ (blue, solid); Bonferroni confidence sets $CS_B^\theta(I)$ (red, short dashes); and estimated sets $F^\theta(\hat{\phi})$ (green, long dashes).

8.2 Combining Sign Restrictions and Zero Restrictions

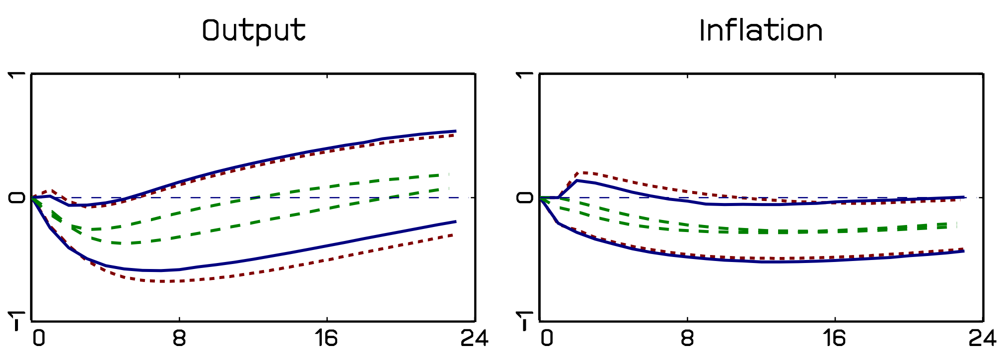
A commonly used identification assumption for monetary policy shocks is that private-sector variables such as output and inflation cannot respond to changes in the federal funds rate within the period; see, for instance, Christiano, Eichenbaum, and Evans (1999). Since the initial impact of the monetary policy shock is given by $\Sigma_{tr} q$ and we ordered the elements of y_t such that output and inflation appear before interest rates and real money balances, 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 $\varphi \in [0, 2\pi]$. The zero restriction on the instantaneous inflation response replaces the sign restriction used in Section 8.1. We maintain the other sign restrictions used previously, that is, the inflation response in period $h = 1$ as well as the real money balance responses in periods $h = 0$ and $h = 1$ are non-positive and the interest rate responses for $h = 0$ and $h = 1$ are non-negative.

Figure 4: Combining Zero and Sign Restrictions

Projection-Based Frequentist Error Bands vs. Bayesian Error Bands



Projection-Based versus Bonferroni Error Bands



Notes: The top panel depicts 90% projection-based frequentist confidence sets $CS_P^{\theta}(I)$ (blue, solid); 90% Bayesian credible intervals (red, short dashes); and estimated sets $F^{\theta}(\hat{\phi})$ (green, long dashes). The bottom panel depicts projection-based confidence sets $CS_P^{\theta}(I)$ (blue, solid); Bonferroni confidence sets $CS_B^{\theta}(I)$ (red, short dashes); and estimated sets $F^{\theta}(\hat{\phi})$ (green, long dashes).

Impulse response bands are depicted in Figure 4. The first panel compares the projection-based frequentist bands $CS_P^{\theta}(I)$, Bayesian credible bands, and the estimated sets $F^{\theta}(\hat{\phi})$. A comparison of $F^{\theta}(\hat{\phi})$ in Figures 3 and 4 indicates that the use of zero restrictions reduces the size of the identified set drastically. For instance, if the zero restrictions are imposed, the inflation response is essentially point identified for horizons exceeding 8 quarters. As a consequence, for output as well as medium- and long-run inflation responses, the width of the frequentist and Bayesian error bands is now much more similar than under the pure-sign-restriction scenario. However, some differences remain with respect to the short-run inflation response. For the first two years, the frequentist intervals cover both positive and negative inflation responses, whereas the Bayesian credible intervals suggest that the inflation response is negative. With the zero restrictions imposed, the direction of the output response is no longer ambiguous – it is negative over the first two years. The bottom panel of Figure 4 provides a comparison of the projection-based and the Bonferroni confidence sets, which are fairly similar in width in this application.

9 Conclusion

With the exception of FRSW, the error bands for impulse responses of sign-restricted SVARs that have been reported in the literature thus far were only meaningful from a Bayesian perspective. The main contribution of our paper is to develop general frequentist methods to construct error bands for impulse responses in VARs that are set-identified based on sign restrictions. We consider projection-based confidence intervals and Bonferroni intervals. Both types of intervals use point-wise testing procedures. We employ the Andrews and Soares (2010a) moment selection procedure to obtain critical values that are not diluted by non-binding inequality conditions. As a by-product, we also establish the consistency of the plug-in estimator $F^{\theta}(\hat{\phi})$ of the identified set of impulse responses. Our empirical application illustrates that in set-identified VARs, frequentist error bands can be substantially wider than Bayesian error bands. The plug-in estimator $F^{\theta}(\hat{\phi})$ is also useful from a Bayesian perspective. 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 useful to report the identified set and the prior conditional on some estimate of ϕ, say, the posterior mean, so that the audience can judge whether the conditional prior distribution is highly concentrated in a particular area of the identified set.

References

Aliprantis, Charalambos D. and Kim C. Border (2006): *Infinite-Dimensional Analysis*, 3rd Edition, Springer Verlag, New York.

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.

Andrews, Donald and Gustavo Soares (2010a): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, **78**, 119-157.

Andrews, Donald and Gustavo Soares (2010b): "Supplement to 'Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,' " *Econometrica Supplementary Material*.

Aruoba, Boragan and Frank Schorfheide (2011): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, **3**, 60-90.

Baumeister, Christiane and Geert Peersman (2008): "Time-Varying Effects of Oil Supply Shocks on the U.S. Economy," *Manuscript*, Ghent University.

Border, Kim C. (2007): "Alternative Linear Inequalities," *Manuscript*, California Institute of Technology.

Border, Kim C. (2010): "Introduction to Correspondences," *Manuscript*, California Institute of Technology.

Canova, Fabio and Gianni De Nicolo (2002): "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7," *Journal of Monetary Economics*, **49**, 1131-59.

Cavanagh, Christopher, Graham Elliott, and James Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, **11**, 1131-1147.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, **75**, 1243-1284.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (1999): "Monetary Policy Shocks: What Have we Learned and to What End?" In: John Taylor and Michael Woodford (eds): *Handbook of Macroeconomics*, Vol 1a, North Holland, 65-148.

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 (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, **58**, 475-494.

Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Conference Series on Public Policy*, **49**, 207-244.

Faust, Jon, John Rogers, Eric Swanson, and Jonathan Wright (2003): "Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data," *Journal of the European Economic Association*, **1**, 1031-1057.

Faust, Jon, Eric Swanson, and Jonathan Wright (2004): "Identifying VARs Based on High Frequency Futures Data," *Journal of Monetary Economics*, **51**, 1107-1131.

Giraitis, Liudas and Peter C.B. Phillips (2004): "Uniform Limit Theory for Stationary Autoregression," *Cowles Foundation Discussion Paper*, **1475**.

Imbens, Guido and Charles Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, **72**, 1845-1857.

James, Alan. T. (1954): "Normal Multivariate Analysis and the Orthogonal Group," *Annals of Mathematical Statistics*, **25**, 40-75.

Kilian, Lutz (1998): "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, **80**, 218-230.

Kilian, Lutz and Daniel 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.

Manski, Charles (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.

Mikusheva, Anna (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, **75**, 1411-1452.

Moon, Hyungsik Roger and Frank Schorfheide (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, **80**, 755-782.

Mountford, Andrew and Harald Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?," *Journal of Applied Econometrics*, **24**, 960-992.

Pappa, Evi (2009): "The Effects of Fiscal Shocks on Employment and the Real Wage," *International Economic Review*, **50**, 217-244.

Peersman, Gert (2005): "What Caused the Early Millennium Slowdown? Evidence Based on Vector Autoregressions," *Journal of Applied Econometrics*, **20**, 185-207.

Peersman, Geert and Roland Straub (2009): "Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR," *International Economic Review*, **50**, 727-750.

Rosen, Adam (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, **146**, 107-117.

Rubio-Ramirez, Juan F., Daniel Waggoner, and Tao Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, **77**, 665-696.

Sims, Christopher A. (1980): "Macroeconomics and Reality," *Econometrica*, **48**, 1-48.

Stoye, Jörg (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, **77**, 1299-1315.

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 Proof of Main Theorems

A.1 Consistency of Set Estimator

To economize on notation we absorb the transformation matrices $M^{S,1}$ and $M^{S,2}$ into the definition of R^v and drop the v superscript. Accordingly, the definition of θ and the sign restrictions can be

$$R_\theta q = \theta \quad \text{and} \quad R_r q \geq 0, \tag{54}$$

where R is the matrix of the reduced-form impulse responses, R_θ is the matrix that stacks the rows of R associated with θ, and R_r is a matrix that stacks all the rows of R associated with all the sign restrictions, including the sign restrictions on θ. The parameter vector ϕ is obtained by deleting zero elements from $vec(R')$. The identified sets can be expressed as

$$\begin{aligned} F^\theta(R) &= \{R_\theta q \,|\, \|q\| = 1 \text{ and } R_r q \geq 0\} \\ F^q(R) &= \{\|q\| = 1 \,|\, R_r q \geq 0\}. \end{aligned}$$

The proof of Theorem 1 is based on three Lemmas, which are stated below.

Proof of Theorem 1:

Part (i)-(a): Since $\{q \mid R_r q \gg 0, \|q\| = 1\} \subset F^q(\phi)$ we deduce from Lemma 1 that $F^q(\phi)$ is non-empty and not a singleton.

Part (i)-(b): We need to show that the set

$$F^\theta(\phi) = \{R_\theta q \mid R_r q \geq 0, \|q\| = 1\}$$

is non-empty and not a singleton. Non-emptiness follows immediately from Part (i)-(a). Suppose, to the contrary, that $F^\theta(\phi)$ is a singleton. This implies that

$$F^q(\phi) = \{q \mid R_r q \geq 0, \|q\| = 1\} \subset \{q \mid R_\theta q = \theta, \|q\| = 1\}.$$

Notice that $\{q \mid R_\theta q = \theta, \|q\| = 1\}$ is a subset of the unit sphere $\mathcal{S}^n$ with dimension less or equal to $(n-1)$. Thus, the dimension of $F^q(\phi)$ has to be less than n. However, this contradicts Lemma 1, which states that the dimension of

$$\{q \mid R_r q \gg 0, \|q\| = 1\} \subset F^q(\phi)$$

is equal to n.

Part (ii): The proof exploits the continuity of $F^{\theta}(\phi)$ and $F^{q}(\phi)$ with respect to ϕ. The statement of the theorem is a consequence of Lemma 2, Lemma 3, and the continuous mapping theorem. □

Recall the definition of the Hausdorff distance: $d(A,B) = \max\ \{d(A|B),\ d(B|A)\}$, where $d(A|B) = \sup_{a\in A}\ d(a,B)$ and $d(a,B) = \inf_{b\in B}\ \|a-b\|$ We set $d(A,B)=\infty$ if either A or B is empty. For any $\varepsilon > 0$, define an open ball around set $A \subset \mathbb{R}^K$ as $\mathbb{B}(A,\varepsilon) = \left\{b \in \mathbb{R}^K : d(b|A) < \varepsilon\right\}$.

Lemma 1 *Let $\mathbb{S}^n$ denote the n-dimensional unit sphere. Suppose that Assumption 1(i) holds. Then there exists an n-dimensional subset $\mathbb{Q}$ of $\mathbb{S}^n$ such that $R_r q \gg 0$ if $q \in \mathbb{Q}$.*

Proof of Lemma 1: Suppose R_r is an $l \times n$ matrix. According to Gordan's Alternative Theorem – see, for instance, Border (2007) – exactly one of the two alternatives holds: (a) there exists an $x \in \mathbb{R}^n$ satisfying $R_r x \gg 0$; or (b) there exists a $z > 0$ satisfying $z' R_r = 0$. Assumption 1(i) rules out alternative (b). Thus, there exists an x such that

$$R_r x \gg 0.$$

Let

$$\epsilon = R_r x \gg 0.$$

Moreover, let $\mathcal{C}(R_r)$ and $\mathcal{N}(R_r)$ be the column space and the null space of matrix R_r, respectively. Since $\epsilon \gg 0$ we can find an open neighborhood around ϵ, denoted by $\mathbb{B}(\epsilon,\delta)$, such that $\mathbb{B}(\epsilon,\delta) \subset \mathbb{R}^l_{++}$, where $\mathbb{R}^l_{++}$ is the subset of $\mathbb{R}^l$ with strictly positive ordinates. Since $\mathcal{C}(R_r)$ is a d-dimensional subspace of $\mathbb{R}^l$ and $\mathbb{B}(\epsilon,\delta)$ is an l-dimensional subset of $\mathbb{R}^l$, we can choose d linearly independent vectors, $\epsilon_1,\ldots,\epsilon_d$, from $\mathbb{B}(\epsilon,\delta) \cap \mathcal{C}(R_r)$. Since $\epsilon_k \in \mathcal{C}(R_r)$, $k=1,\ldots,d$, we can find $n \times 1$ vectors x_k such that $R_r x_k = \epsilon_k$.

Formally, the system of equations can be solved as follows. Consider a singular value decomposition of R_r:

$$R_r = \underbrace{U}_{l\times l}\underbrace{D}_{l\times n}\underbrace{V'}_{n\times n} = \underbrace{U_1}_{l\times d}\underbrace{D_1}_{d\times d}\underbrace{V_1'}_{d\times n},$$

where U and V are orthogonal matrices with partitions $U=[U_1,U_2]$ and $V=[V_1,V_2]$, and D_1 is a diagonal submatrix matrix of D with non-zero entries on the diagonal. If we express $\epsilon_k = U_1\lambda_k$, then we can define

$$x_k = V_1 D_1^{-1}\lambda_k, \quad k=1,\ldots,d.$$

In addition, let

$$x_{d+i} = V_2 \iota_i, \quad i=1,\ldots,n-d,$$

where V_2 is of dimension $n \times (n-d)$ and ι_i is a $(n-d) \times 1$ vector with the i'th element equal to one and all other elements equal to zero. Since V_2 is composed of columns of an orthogonal matrix and ι_i selects the i'th column, $\{x_{d+1}, \ldots, x_n\}$ are by construction linearly independent. Since $V_1'V_2 = 0$ $\{x_{d+1}, \ldots, x_n\}$ are orthogonal to $\{x_1, \ldots, x_d\}$.

Moreover, $\{x_1, \ldots, x_d\}$ are also linearly independent. Suppose, to the contrary, that they are not. Then, there exist constants $c_1, \ldots, c_d$ such that at least two of them are different from zero and

$$0 = c_1 x_1 + \ldots + c_d x_d.$$

In turn,

$$0 = c_1 R_r x_1 + \ldots + c_d R_r x_d = c_1 \epsilon_1 + \ldots + c_d \epsilon_d.$$

Since $\epsilon_1, \ldots, \epsilon_d$ are by construction linearly independent, it must be the case that $c_1 = \ldots = c_d$ and we have a contradiction. Thus, we deduce that $\{x_1, \ldots, x_n\}$ are linearly independent.

Now define the following n-dimensional subset of $\mathbb{R}^n$:

$$\Xi = \left\{ \chi \in \mathbb{R}^n \mid \chi = \sum_{k=1}^{n} c_k x_k, \ [c_1, \ldots, c_n]' \in \mathbb{R}^n, \text{ and } c_1, \ldots, c_d > 0 \right\}.$$

By construction, for each $\chi \in \Xi$ we have

$$R_r \chi = \sum_{k=1}^{d} c_k R_r x_k + \sum_{k=d+1}^{n} c_k U_1 D_1 V_1' V_2 \iota_{k-d} = \sum_{k=1}^{d} c_k R_r x_k = \sum_{k=1}^{d} c_k \epsilon_k \gg 0.$$

Note that the second sum is zero because $V_1'V_2 = 0$. Finally, define

$$\mathbb{Q} = \left\{ \frac{\chi}{\|\chi\|} \,\middle|\, \chi \in \Xi \right\}.$$

Then, $\mathbb{Q}$ is an n-dimensional subset of the hypersphere $\mathbb{S}^n$ such that $R_r q \gg 0$ if $q \in \mathbb{Q}$. □

Lemma 2 *Suppose that $F(\phi)$ is a non-empty compact-valued continuous correspondence. Then, $\phi \longrightarrow \phi^*$ implies that $d(F(\phi), F(\phi^*)) \longrightarrow 0$.*

Proof of Lemma 2: Consider any $\varepsilon > 0$. For the required result, we show that there exists $\delta > 0$ such that $\|\phi - \phi^*\| < \delta$ implies that (i) $d(F(\phi) | F(\phi^*)) < \varepsilon$ and (ii) $d(F(\phi^*) | F(\phi)) < \varepsilon$.

Part (i): Consider an open neighborhood $\mathbb{B}(F(\phi^*), \epsilon)$ of $F(\phi^*)$. Since $F(\phi)$ is upper hemi-continuous, we can find $\delta > 0$ such that for any $\phi \in \mathbb{B}(\phi^*, \delta)$, $F(\phi) \subset \mathbb{B}(F(\phi^*), \epsilon)$. This implies that if $\|\phi - \phi^*\| < \delta$, then $d(F(\phi)|F(\phi^*)) = \sup_{\theta \in F(\phi)} d(\theta | F(\phi^*)) < \epsilon$, which proves Part (i).

Part (ii): Suppose that (ii) does not hold. Then, there exists an $\epsilon > 0$ such that for any $\delta > 0$, it is possible that $\|\phi - \phi^*\| < \delta$ but $d(F(\phi^*)|F(\phi)) = \sup_{\theta \in F(\phi^*)} d(\theta|F(\phi)) \geq 3\epsilon$. Since $F(\phi^*)$ is compact and $d(\theta|F(\phi))$ is continuous in θ, there exists $\theta^* \in F(\phi^*)$ such that $\theta^* \in \arg\max_{\theta \in F(\phi^*)} d(\theta|F(\phi))$. Then, since $d(\theta^*, F(\phi)) \geq 3\epsilon$, we have $\varnothing = \mathbb{B}(\theta^*, \epsilon) \cap \mathbb{B}(F(\phi), \epsilon) \supset \mathbb{B}(\theta^*, \epsilon) \cap F(\phi)$. This implies that if Part (ii) does not hold, then we can find an open set $\mathbb{B}(\theta^*, \epsilon)$ with $\mathbb{B}(\theta^*, \epsilon) \cap F(\phi^*) \neq \varnothing$ (since $\theta^* \in F(\phi^*)$ and $F(\phi^*)$ is non-empty), but for any open set $\mathbb{B}(\phi^*, \delta)$, $\phi \in \mathbb{B}(\phi^*, \delta)$ (that is, $\|\phi - \phi^*\| < \delta$) implies that $\mathbb{B}(\theta^*, \epsilon) \cap F(\phi) = \varnothing$. This contracts to the assumption that $F(\phi)$ is lower hemi-continuous at ϕ^*. $\square$

Lemma 3 *Assume that $F^\theta(R^v)$ is non-empty for all R^v in a neighborhood of R_0^v. Then,*

(i) $F^q(R^v)$ is compact for all R^v;

(ii) $F^\theta(R^v)$ is compact for all R^v;

(iii) $F^q(R^v)$ is continuous at R_0^v;

(iv) $F^\theta(R^v)$ is continuous at R_0^v.

Proof of Lemma 3: For notational simplicity, we omit the superscript notation v and write R^v as R. Let $\mathbb{S}^m = \{q \in \mathbb{R}^m : \|q\| = 1\}$ be the unit sphere in $\mathbb{R}^m$. Recall from Theorem 1(i) that $F^q(\phi)$ and $F^\theta(\phi)$ are nonempty.

Part (i): We show that $F^q(R)$ is bounded and closed.

Boundedness: It is straightforward since $F^q(R) \subset \mathbb{S}^m$.

Closedness: Consider any sequence $q_j \in F^q(R)$ such that $q_j \to q_0$. Then, $0 \leq R_r q_j \longrightarrow R_r q_0$, so that it should be $R_r q_0 \geq 0$. This implies that $q_0 \in F^q(R)$, as required for closedness.

Part (ii): Since $F^\theta(R) \subset \mathbb{R}^{\tilde{k}}$, for the required result, we show that $F^\theta(R)$ is bounded and closed.

Boundedness: Notice that the set $\{R_\theta q : \|q\| = 1\}$ is compact because $R_\theta q$ is continuous in q and the domain of q, $\mathbb{S}^m$, is compact. Since $F^\theta(R) \subset \{R_\theta q : \|q\| = 1\}$, $F^\theta(R)$ is bounded.

Closedness: Consider any sequence $\theta_j \in F^\theta(R)$ such that $\theta_j \longrightarrow \theta_0$. We show that $\theta_0 \in F^\theta(R)$, that is, we need to find a q_0 such that $\theta_0 = R_\theta q_0$ and $R_r q_0 \geq 0$. Then, the desired result follows. For $\theta_j \in F^\theta(R)$, by definition we can choose q_j such that $R_\theta q_j = \theta_j$ and $R_r q_j \geq 0$. Since $\{q_j\} \subset \mathbb{S}^m$ and $\mathbb{S}^m$ is compact, we can choose a convergent subsequence q_{j_i} such that $q_{j_i} \longrightarrow q_0$. Then, it follows that $0 \leq R_r q_{j_i} \longrightarrow R_r q_0$, so that $R_r q_0 \geq 0$. Also, $R_\theta q_{j_i} = \theta_{j_i} \longrightarrow R_\theta q_0$. Since θ_{j_i} should also converge to θ_0, we have $R_\theta q_0 = \theta_0$. By definition of $F^\theta(R)$, then, we have $\theta \in F^\theta(R)$, as required for closedness.

Part (iii): We show $F^q(R)$ is upper hemi-continuous (UHC) and lower hemi-continuous (LHC) at R_0.

UHC: Since $F^q(R)$ is non-empty and compact-valued, the UHC of $F^q(R)$ at R^0 follows if we show that for every sequence $R_j \longrightarrow R_0$ and $q_j \in F^q(R_j)$, there exists a subsequence q_{j_i} of q_j such that $q_{j_i} \longrightarrow q_0 \in F^q(R_0)$. (See Border (2010) Proposition 20). Since $\{q_j\} \subset \mathbb{S}^m$ and $\mathbb{S}^m$ is compact, we can choose a convergent subsequence q_{j_i} such that $q_{j_i} \longrightarrow q_0$. Then, $0 \leq R_{r,j_i} q_{j_i} \longrightarrow R_0 q_0$, and it follows that $R_0 q_0 \geq 0$. This implies that $q_0 \in F^q(R_0)$, as required.

LHC: $F^q(R)$ is LHC at R^0 if and only if for any sequence $\{R_j\}$ with $R_j \longrightarrow R_0$ and $q_0 \in F^q(R_0)$, there exists a sequence $q_j \in F^q(R_j)$ with $q_j \longrightarrow q_0$. For a matrix A, we denote the l^{th} row of A as $(A)^l$. We partition the matrix $R_{r,0}$ to $R_{r,0} = \left[R'_{r,1}, R'_{r,2}\right]'$, where $R_{r,1} q_0 = 0$ and $R_{r,2} q_0 > 0$.

Let $R^l_{r,1}$ be the lth row of $R_{r,1}$. By Gordan's Alternative Theorem (see Border, 2007), Assumption 1(i) implies that there exists a $\xi^* \in \mathbb{R}^m$ such that

$$R_{r,1}\xi^* \gg 0.$$

Let

$$\xi = \frac{1}{\min_l \ R^l_{r,1}\xi^*}\xi^*$$

such that for all l

$$R^l_{r,1}\xi \geq 1.$$

Set $\epsilon_{j,l} = \left\|(R_j - R_0)^l\right\|$ and $\epsilon_j = \max_l \{\epsilon_{j,l}\}$, and define

$$q_j = \frac{q_0 + \epsilon_j \xi}{\|q_0 + \epsilon_j \xi\|}.$$

Notice that q_j is well defined when ϵ_j is small enough because $q_0 \in \mathbb{S}^m$, the unit sphere of dimension n, and as a result, $q_0 \neq \epsilon_j \xi$ when ϵ_j is small and ξ is fixed.

Case (i): Suppose $(R_{r,0})^l q_0 = 0$. Then, when j is large so that $(R_{r,0})^l \xi - 1 \geq \epsilon_j \|\xi\|$, we have

$$\begin{aligned}
(R_{r,j})^l q_j &= (R_{r,j} - R_{r,0})^l q_j + (R_{r,0})^l q_j \\
&= \frac{1}{\|q_0 + \epsilon_j \xi\|}\left\{(R_{r,j} - R_{r,0})^l q_0 + \epsilon_j (R_{r,j} - R_{r,0})^l \xi + (R_{r,0})^l q^0 + \epsilon_j (R_{r,0})^l \xi\right\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|}\left\{-\left\|(R_{r,j} - R_{r,0})^l\right\| \|q_0\| - \epsilon_j \left\|(R_{r,j} - R_{r,0})^l\right\| \|\xi\| + \epsilon_n (R_{r,0})^l \xi\right\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|}\left(-\epsilon_j - \epsilon_j^2 \|\xi\| + \epsilon_j (R_{r,0})^l \xi\right) \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|}\epsilon_j \left((R_{r,0})^l \xi - 1 - \epsilon_j \|\xi\|\right) \\
&\geq 0.
\end{aligned}$$

Case (ii): Suppose $(R_{r,0})^l q_0 > 0$. Then, since $\left\|(R_{r,0})^l\right\| \leq M$ (compact parameter set), we have

$$
\begin{aligned}
(R_{r,j})^l q_j &= (R_{r,j} - R_{r,0})^l q_j + (R_{r,0})^l q_j \\
&= \frac{1}{\|q_0 + \epsilon_j \xi\|} \left\{ (R_{r,j} - R_{r,0})^l q_0 + \epsilon_j (R_{r,j} - R_{r,0})^l \xi + (R_{r,0})^l q^0 + \epsilon_j (R_{r,0})^l \xi \right\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} \left\{ - \left\|(R_{r,j} - R_{r,0})^l\right\| \|q_0\| - \epsilon_j \left\|(R_{r,j} - R_{r,0})^l\right\| \|\xi\| + (R_{r,0})^l q_0 - \epsilon_j \left\|(R_{r,0})^l\right\| \|\xi\| \right\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} \left((R_{r,0})^l q_0 - \epsilon_j - \epsilon_j^2 M - \epsilon_j M^2 \right) \\
&\geq 0,
\end{aligned}
$$

when j is large. The last inequality holds since $(R_{r,0})^l q_0 > 0$.

From these, we can deduce that

$$R_{r,j} q_j \geq 0.$$

Also, since $\epsilon_j \longrightarrow 0$, we have

$$q_j \longrightarrow q_0.$$

Then, we have all the required results for the LHC.

Part (iv): Define function $f(q, R_\theta) = R_\theta q$. Then, it is continuous in (q, R_θ). Define a projection $F_{proj}(R) = R_\theta$. Then, the product correspondence

$$\tilde{F}^q(R) = F^q(R) \times F_{proj}(R)$$

is continuous by Proposition 34 of Border (2010). Notice that the correspondence $F^\theta(R)$ is a composite of f and $\tilde{F}^q$ as

$$F^\theta(R) = \left(f \circ \tilde{F}^q\right)(R) = \bigcup_{(q \times R_\theta) \in \tilde{F}^q(R)} f(q, R_\theta)$$

Since both f and $\tilde{F}^q$ are continuous, by Theorem 17.23 of Aliprantis and Border (2006), $F^\theta(R)$ is continuous. $\square$

A.2 Coverage Probability of the Confidence Set

This section provides a proof for Theorem 2(i). The proof makes use of various lemmas that are stated and proved in the Online Appendix that accompanies this paper. The proof closely follows the proofs provided in Andrews and Soares (2010b). However, several modifications are needed to account for the potential row rank reduction of the matrix $\tilde{S}(q)$.

To simplify the notation in the proofs we eliminate ρ from the formulas and index the probability distribution by $\phi \in \mathcal{P}$ instead of $\rho \in \mathcal{R}$. Thus we write

$$\inf_{\phi \in \mathcal{P}} \inf_{\theta \in F^{\theta}(\phi)} P_{\phi}\{\theta \in CS^{\theta}(\hat{\phi})\}$$

instead of

$$\inf_{\rho \in \mathcal{R}} \inf_{\theta \in F^{\theta}(\phi(\rho))} P_{\rho}\{\theta \in CS^{\theta}(\hat{\phi})\}.$$

Reduced-form parameter sequences ρ_T and $\phi(\rho_T)$ are simply abbreviated by ϕ_T.

Proof of Theorem 2(i): We closely follow the proofs of Theorem 1 and Lemma 2 of Andrews and Soares (2010b). The main modification is to accommodate the reduced rank possibility of $\Sigma(q)$ and $D(q)$. According to Lemma B 4, the desired result of the theorem follows if we show that $CS^{\theta,q}$ is a valid $1-\alpha$ confidence set. That is, we need to show

$$\liminf_{T} \inf_{\phi \in \mathcal{P}} \inf_{(\theta,q) \in F^{\theta,q}(\phi)} P_{\phi}\left\{(\theta,q) \in CS^{\theta,q}\right\} \geq 1-\alpha. \tag{55}$$

Let

$$AsyCP = \liminf_{T} \inf_{\phi \in \mathcal{P}} \inf_{(\theta,q) \in F^{\theta,q}(\phi)} P_{\phi}\left\{(\theta,q) \in CS^{\theta,q}\right\}.$$

Then, there exists sequences $\{\phi_T, \theta_T, q_T\}$ such that $(\theta_T, q_T) \in F^{\theta,q}(\phi_T)$ and

$$AsyCP = \liminf_{T} P_{\phi_T}\left\{(\theta_T, q_T) \in CS^{\theta,q}\right\}.$$

Furthermore, there exists a subsequence of $\{T\}$, $\{T'\} \subset \{T\}$, such that

$$AsyCP = \lim_{T'} P_{\phi_{T'}}\left\{(\theta_{T'}, q_{T'}) \in CS^{\theta,q}\right\}.$$

In what follows we show that there exists a subsubsequence, say $\{T''\} \subset \{T'\}$, such that

$$\lim_{T''} P_{\phi_{T''}}\left\{\left(\theta_{T''}, q_{T''}\right) \in CS^{\theta,q}\right\} \geq 1-\alpha. \tag{56}$$

Then, the desired result (55) for the theorem follows and the proof is complete.

Define $\mu(q,\phi) = S_R(q)\phi$. Recall the definitions

$$\Sigma(q) = S(q)\Lambda S(q)' = S(q)LL'S(q)' = D^{1/2}(q)\Omega(q)D^{1/2} \quad \text{and} \quad A(q) = L'S'(q)D^{-1/2}(q).$$

To simplify the notation we suppress the dependence of matrices on ϕ. The matrix $\Omega(q) = A'(q)A(q)$ is a correlation matrix and $D^{1/2}(q)$ is a diagonal matrix of standard deviation that can be partitioned into $D^{1/2}(q) = diag\left(D_{\theta}^{1/2}(q), D_R^{1/2}(q)\right)$, where the partitions conform with

$S(q) = [S'_\theta(q), S'_R(q)]'$. Also, recall that $W(q) = D^{1/2}(q) B(q) D^{1/2}(q)$, where either $B(q) = \Omega^{-1}(q)$ or $B(q) = I$. The proof is completed in three steps.

Step 1: Choosing the subsequence T''. We can choose a subsequence T'' from T' along which the following conditions are satisfied:

(i) $\phi_{T''} \longrightarrow \phi$.

(ii) $k(q_{T''}) = k$, $r(q_{T''}) = r$, $l(q_{T''}) = l$, $V(q_{T''}) = V$ for all T''.

(iii) If $r \geq 1$, then for $j = 1, ..., r$, the slackness in inequality j converges to

$$\begin{aligned} \sqrt{T''}\mu_j(q_{T''}, \phi_{T''}) &\longrightarrow h_j \\ \kappa_{T''}^{-1} D_{jj,R}^{-1/2}(q_{T''}) \sqrt{T''}\mu_j(q_{T''}, \phi_{T''}) &\longrightarrow \pi_j \end{aligned}$$

such that one of the following is true: (a) $h_j < \infty$ and $\pi_j = 0$; (b) $h_j = \infty$ and $\pi_j < \infty$; (c) $h_j = \infty$ and $\pi_j = \infty$.

(iv) If $l > 0$, the sequence $A(q_{T''})$ has a full rank limit, denoted by A.

Condition (i) is satisfied since the reduced-form parameter set $\mathcal{R}$ is assumed to be compact (Assumption 1). If condition (i) holds, then we have

(v) The convergence $\phi_{T''} \longrightarrow \phi$ implies that $\Lambda(\phi_{T''}) \longrightarrow \Lambda(\phi)$ since $\Lambda(\phi)$ is continuous by Assumption 1(v). Also, $\hat{\Lambda}(\hat{\phi}_{T''}) \xrightarrow{p} \Lambda$ by Assumption 1(vi).

Condition (ii) is satisfied since $k(q_{T'})$, $r(q_{T'})$, $l(q_{T'})$, and $V(q_{T'})$ are sequences that take only a finite number of discrete values. Condition (iii) is satisfied because the range of the sequences of interest is $[0, \infty]$ and by a similar argument used in the proof of Theorem 1 of Andrews and Soares (2010b). Roughly speaking, in Case (iii)-(a) the slackness is small and the selection criterion regards the inequality asymptotically as binding. In Case (iii)-(c) the slackness is large and the selection criterion regards the inequality as non-binding and (iii)-(b) is an intermediate case. Condition (iv) is satisfied by Lemma B 3. If Condition (iv) is satisfied, then the following conditions also hold ((vii) is a consequence of Lemma 3):

(vi) $\Omega(q_{T''}) \longrightarrow A'A > 0$ and $B(q_{T''}) \longrightarrow B > 0$, where $B = (A'A)^{-1}$ if $B(q) = \Omega^{-1}(q)$ and $B = I$ if $B(q) = I$.

(vii) $\hat{\Omega}(q_{T''}) \xrightarrow{p} A'A > 0$ and $\hat{B}(q_{T''}) \xrightarrow{p} B > 0$, where $B = (A'A)^{-1}$ if $\hat{B}(q) = \hat{\Omega}^{-1}(q)$ and $B = I$ if $\hat{B}(q) = I$.

We now reorder the rows of $S(q_{T''})$ such that $\pi_j = 0$ for rows $j = 1, \ldots, r_1$ and $\pi_j > 0$ for rows $j = r_1 + 1, \ldots, r$. Under this ordering, the inequality selection procedure will eventually eliminate the last $r_2 = r - r_1$ rows of $S(q_{T''})$.

Step 2: Constructing an upper bound for the critical value $c_P^\alpha(q)$ in (41). For notational simplicity we use sequence notation $\{T\}$ for the subsubsequence $\{T''\}$ in Step 1. Recall the definitions

$$\xi_{j,T}(q_T) = D_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T,\hat{\phi}) \quad \text{and} \quad \hat{\xi}_{j,T}(q_T) = \hat{D}_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T,\hat{\phi}).$$

Let $\varphi_T(q_T)$ and $\hat{\varphi}_T$ be vectors with elements

$$\varphi_{j,T}(q_T) = \begin{cases} \infty & \text{if } \xi_{j,T}(q_T) \geq \kappa_T \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\varphi}_{j,T}(q_T) = \begin{cases} \infty & \text{if } \hat{\xi}_{j,T}(q_T) \geq \kappa_T \\ 0 & \text{otherwise} \end{cases},$$

respectively. Moreover, define $\varphi_T^*(q_T)$ and $\hat{\varphi}_T^*(q_T)$ with elements

$$\varphi_{j,T}^*(q_T) = \begin{cases} \varphi_{j,T}(q_T) & \text{if } \pi_j = 0 \\ \infty & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\varphi}_{j,T}^*(q_T) = \begin{cases} \hat{\varphi}_{j,T}(q_T) & \text{if } \pi_j = 0 \\ \infty & \text{otherwise} \end{cases},$$

where, according to Case (iii) in Step 1,

$$\pi_j = \lim \ \kappa_T^{-1} D_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T,\phi_T).$$

Finally, define the vector π_* with elements

$$\pi_j^* = \begin{cases} 0 & \text{if } \pi_j = 0 \\ \infty & \text{otherwise} \end{cases}.$$

To characterize the critical values, define the objective function

$$\bar{\mathcal{G}}\big(q_T; A(\cdot), B(\cdot), \varphi(\cdot)\big) = \min_{v \geq -\varphi(q_T)} \ \left\| A\left(q_T\right)' Z_m - M_v v \right\|_{B(q_T)}^2.$$

Here we set $M_v = 0$ if $S_R\left(q_T\right) = 0$ (*i.e.*, $r = 0$). Note that the notation in the proof is slightly different from the notation in the main text. In (40) of the main text we defined $\bar{\mathcal{G}}\big(q; \hat{B}(q), M_{\hat{\xi}}(q)\big)$, which corresponds to $\bar{\mathcal{G}}\big(q_T; \hat{A}(\cdot), \hat{B}(\cdot), \hat{\varphi}(\cdot)\big)$ in this proof. We let

$$c_T^\alpha\big(A\left(\cdot\right), B\left(\cdot\right), \varphi(\cdot)\big) = (1-\alpha) \text{ quantile of } \bar{\mathcal{G}}\left(\theta_T, q_T; A\left(\cdot\right), B\left(\cdot\right), \varphi(\cdot)\right). \tag{57}$$

To cover the special case $l = r_2 > 0$, i.e., the dimension of $A'Z_m$ equals the dimension of ν and all the inequality conditions are non-binding, we adopt the convention that

$$c_T^\alpha\left(A\left(\cdot\right), B\left(\cdot\right), \varphi(\cdot)\right) = 0 \tag{58}$$

if $\varphi(q_T) = \hat{\varphi}_T^*(q_T)$ or $\varphi(q_T) = \pi_*$. The critical value $c_P^\alpha(q)$ in (41) in the main text can be expressed as

$$c_P^\alpha (q_T) = c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T (q_T) \right).$$

Notice by definition that

$$\hat{\varphi}_T^*(q_T) \geq \hat{\varphi}_T(q_T).$$

This implies that

$$c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T^* (q_T) \right) \leq c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T (q_T) \right). \tag{59}$$

Step 3: Establish the asymptotic coverage probability. Along the sequence defined in Step 1 we will show the desired result

$$AsyCP = \lim_T P_{\phi_T} \left\{ (\theta_T, q_T) \in CS^{\theta,q} \right\} \geq 1 - \alpha.$$

We consider three different cases: (i) $l > 0$ and $l > r_2$; (ii) $l = r_2 > 0$ (*i.e.*, $k = r_2 = 0$); and (iii) $l = 0$. Before we proceed, notice that since $(\theta_{T'}, q_{T'}) \in F^{\theta,q} (\phi_{T'})$, we have $\theta_{T'} = \tilde{S}_\theta (q_{T'}) \phi_{T'}$ for all T' and the penalty term in (43) is equal to zero.

Step 3(i). Suppose that $l > 0$ and $l > r_2$. By Lemma B 2 and (59), we have

$$\begin{aligned} AsyCP &= \lim_T P_{\phi_T} \left\{ (\theta_T, q_T) \in CS^{\theta,q} \right\} \\ &= \lim_T P_{\phi_T} \left\{ G \left(\theta_T, q_T; \hat{\phi}, \hat{W} (\cdot) \right) \leq c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T (q_T) \right) \right\} \\ &= \lim_T P_{\phi_T} \left\{ G \left(\theta_T, q_T; \hat{\phi}, W (\cdot) \right) + o_p (1) \leq c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T (q_T) \right) \right\} \\ &\geq \lim_T P_{\phi_T} \left\{ G \left(\theta_T, q_T; \hat{\phi}, W (\cdot) \right) + o_p (1) \leq c_T^\alpha \left(\hat{A} (q_T) , \hat{B} (q_T) , \hat{\varphi}_T^* (q_T) \right) \right\}. \end{aligned}$$

By using an argument similar to that used in showing (A.10) of Andrews and Guggenberger (2009), it can be shown that

$$\begin{aligned} & G \left(\theta_T, q_T; \hat{\phi}, W (\cdot) \right) + o_p (1) \\ &= \min_{v \geq -D_R^{-1/2}(q_T)\sqrt{T}\mu(q_T,\phi_T)} \left\| A(q_T)' L^{-1} \sqrt{T} \left(\hat{\phi} - \phi_T \right) - M_v v \right\|_{B(q_T)}^2 + o_p (1) \\ &\Longrightarrow \min_{v \geq -h} \left\| A' Z_m - M_v v \right\|_B \\ &\leq \min_{v \geq -\pi^*} \left\| A' Z_m - M_v v \right\|_B . \end{aligned}$$

The last inequality holds because $h \geq \pi^*$. (This is true because $\pi_j = 0$ implies that $h_j < \infty$ and $\pi_j^* = 0$, while $\pi_j > 0$ implies that $h_j = \pi_j^* = \infty$.) According to Lemma B 5,

$$c_T^\alpha \left(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T^*(q_T) \right) \xrightarrow{p} c_T^\alpha (A, B, \pi^*).$$

Since $l > 0$ and $l > r_2$, $c_T^\alpha (A, B, \pi^*) > 0$. Also, the distribution function of $\min_{v \geq -\pi^*} \|A' Z_m - M_v v\|_B$ is continuous near the $(1-\alpha)^{th}$ quantile. (See page 6 of Andrews and Soares (2010b).) Then, we have the required result:

$$\begin{aligned} AsyCP &= \lim_T P_{\phi_T} \left\{ (\theta_T, q_T) \in CS^{\theta,q} \right\} \\ &\geq \lim_T P_{\phi_T} \left\{ G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) \leq c_T^\alpha \left(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T^*(q_T) \right) + o_p(1) \right\} \\ &\geq P\left\{ \min_{v \geq -\pi^*} \|A' Z_m - M_v v\|_B \leq c_T^\alpha (A, B, \pi^*) \right\} \\ &= 1 - \alpha. \end{aligned}$$

Step 3(ii). Suppose that $l = r_2$. In this case, we have $k = 0$ and $r_1 = 0$. Also, $h_j = \infty$ and $\pi_j > 0$ for all $j = 1, ..., r$. Then, we have $\hat{\varphi}_T^*(q_T) = \pi = \infty$ for all T. Recall the definitions that $c_T^\alpha \left(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T^*(q_T) \right) = c_T^\alpha (A, B, \pi^*)) = 0$. Then, by Lemma B 2 and (59), we have

$$\begin{aligned} AsyCP &= \lim_T P_{\phi_T} \left\{ (\theta_T, q_T) \in CS^{\theta,q} \right\} \\ &= \lim_T P_{\phi_T} \left\{ G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) \leq c_T^\alpha \left(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T(q_T) \right) \right\} \\ &= \lim_T P_{\phi_T} \left\{ G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) + o_p(1) \leq c_T^\alpha \left(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T(q_T) \right) \right\} \\ &\geq \lim_T P_{\phi_T} \left\{ G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) + o_p(1) \leq c_T^\alpha (A, B, \pi^*) = 0 \right\}. \end{aligned}$$

By using the same argument used in (S1.23) on page 7 of Andrews and Soares (2010b), we can deduce that

$$\lim_T P_{\phi_T} \left\{ G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) + o_p(1) \leq 0 \right\} \geq 1 - \alpha.$$

Step 3(iii). When $l = 0$, the objective function is zero by definition and the required result holds trivially. □

Online Technical Appendix

This Online Appendix accompanies the paper "Inference for VARs Identified with Sign Restrictions" by H.R. Moon, F. Schorfheide, and E. Granziera. The appendix has three sections. In Section A we provide proofs of Theorem 2(ii) and the lemmas stated in the main text. Section B states and proves lemmas that are needed to prove Theorem 2 in the main text. Finally, Section C provides analytical derivations for the Monte Carlo experiment presented in Section 7 of the main text.

To simplify the notation in some of the proofs we eliminate ρ from the formulas and index the probability distribution by $\phi \in \mathcal{P}$ instead of $\rho \in \mathcal{R}$. Thus we write

$$\inf_{\phi \in \mathcal{P}} \inf_{\theta \in F^{\theta}(\phi)} P_{\phi}\{\theta \in CS^{\theta}(\hat{\phi})\}$$

instead of

$$\inf_{\rho \in \mathcal{R}} \inf_{\theta \in F^{\theta}(\phi(\rho))} P_{\rho}\{\theta \in CS^{\theta}(\hat{\phi})\}.$$

Reduced-form parameter sequences ρ_T and $\phi(\rho_T)$ are simply abbreviated by ϕ_T.

A Proof of Theorem 2(ii)

Proof of Theorem 2(ii): Notice that

$$\begin{aligned}
&\liminf_{T} \inf_{\phi} \inf_{(\theta,q) \in F^{\theta,q}(\phi)} P_{\phi}\left\{\theta \in CS_B^{\theta}\right\} \\
&\geq \liminf_{T} \inf_{\phi} \inf_{(\theta,q) \in F^{\theta,q}(\phi)} P_{\phi}\left\{q \in CS_B^{q},\ \theta \in CS_{B,q}^{\theta}\right\} \\
&\geq \liminf_{T} \inf_{\phi} \inf_{q \in F^{q}(\phi)} P_{\phi}\left\{q \in CS_B^{q}\right\} + \liminf_{T} \inf_{\phi} \inf_{q \in F^{q}(\phi)} P_{\phi}\left\{\theta \in CS_{B,q}^{\theta}\right\} - 1.
\end{aligned}$$

Recall that $\theta = \tilde{S}_{\theta}(q)\phi$. The desired result for the theorem follows if we show

$$\liminf_{T} \inf_{\phi} \inf_{q \in F^{q}(\phi)} P_{\phi}\left\{q \in CS_B^{q}\right\} \geq 1 - \alpha_1 \tag{A.1}$$

and

$$\liminf_{T} \inf_{\phi} \inf_{q \in F^{q}(\phi)} P_{\phi}\left\{\tilde{S}_{\theta}(q)\,\phi \in CS_{B,q}^{\theta}\right\} \geq 1 - \alpha_2. \tag{A.2}$$

(A.1) follows from Theorem 2(i) as the special case of $k(q) = 0$. $(A.2)$ can also be obtained as a special case of Theorem 2(i) by setting $r(q) = 0$. If θ is a scalar parameter, then the standard Wald-interval could be used for $CS_{B,q}^{\theta}$. $\square$

B Additional Technical Lemmas

Throughout this section we use the following notation. When A is a matrix, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the largest and the smallest eigenvalues of A, respectively. We denote A_k as the k^{th} column vector of A; A^j as the j^{th} row vector of A; and A_{jk} as the $(j,k)^{th}$ element of A. Throughout the proofs, we sometimes omit the ϕ_T argument from the asymptotic covariance matrix $\Lambda = LL'$ and use the notation Λ_T, $\hat{\Lambda}_T$, L_T, and $\hat{L}_T$ for simplicity. We also often omit the q_T argument for some of the matrices that depend on q_T and simply write, say, S_T, D_T, $\hat{D}_T$, A_T, $\hat{A}_T$, Ω_T, and $\hat{\Omega}_T$ for short.

Lemma B 1 *Suppose that Assumptions 1(i) and (ii) are satisfied and $\tilde{k}=1$. Then $F^{\theta}(\phi)$ is convex and bounded.*

Proof of Lemma B 1. To simplify the notation in the proof we omit tildes and write $S_\theta(q)$, $S_R(q)$ instead of $\tilde{S}_\theta(q)$, $\tilde{S}_R(q)$. Moreover, unlike in the main text we assume that $S_R(q)$ is defined such that it also captures inequality constraints directly imposed on θ. *Convexity:* Suppose $\theta_i \in \Theta(\phi)$, $i=1,2$, and $\theta_1 < \theta_2$. Then there exist q_i with $\|q_i\|=1$ and $\mu_i \geq 0$ such that

$$S_\theta(q_i)\phi - \theta_i = 0, \quad S_R(q_i)\phi - \mu_i = 0. \tag{B.1}$$

Convexity-Case (i): Suppose that $q_1 \neq -q_2$. We now verify that for any $\lambda \in [0,1]$ $\theta = \lambda\theta_1 + (1-\lambda)(\theta_2) \in \Theta(\phi)$. For $\tau \in [0,1]$ define

$$q(\tau) = \frac{\tau q_1 + (1-\tau)q_2}{\|\tau q_1 + (1-\tau)q_2\|}, \quad H(\tau) = S_\theta(q(\tau))\phi - \theta.$$

The linearity of $S_\theta(q)$ with respect to q and (B.1) implies that

$$\begin{aligned} H(\tau) &= \frac{\tau S_\theta(q_1)\phi}{\|\tau q_1 + (1-\tau)q_2\|} + \frac{(1-\tau)S_\theta(q_2)\phi}{\|\tau q_1 + (1-\tau)q_2\|} - \lambda\theta_1 - (1-\lambda)\theta_2 \\ &= \frac{\tau \theta_1}{\|\tau q_1 + (1-\tau)q_2\|} + \frac{(1-\tau)\theta_2}{\|\tau q_1 + (1-\tau)q_2\|} - \lambda\theta_1 - (1-\lambda)\theta_2. \end{aligned}$$

Using $\|q_i\|=1$ we obtain

$$\begin{aligned} H(0) &= \theta_2 - \lambda\theta_1 - (1-\lambda)\theta_2 = \lambda(\theta_2-\theta_1) \geq 0 \\ H(1) &= \theta_1 - \lambda\theta_1 - (1-\lambda)\theta_2 = -(1-\lambda)(\theta_2-\theta_1) \leq 0 \end{aligned}$$

Since $H(\tau)$ is continuous we deduce that there exists a τ^* such that $H(\tau^*) = 0$. Now consider

$$\begin{aligned} S_R(q(\tau^*)) &= \frac{\tau^* S_R(q_1)\phi}{\|\tau^* q_1 + (1-\tau^*)q_2\|} + \frac{(1-\tau^*)S_R(q_2)\phi}{\|\tau^* q_1 + (1-\tau^*)q_2\|} \\ &= \frac{\tau^* \mu_1}{\|\tau^* q_1 + (1-\tau^*)q_2\|} + \frac{(1-\tau^*)\mu_2}{\|\tau^* q_1 + (1-\tau^*)q_2\|} \\ &\geq 0. \end{aligned}$$

The first equality follows from the linearity of $S_R(q)$, the second equality is implied by (B.1), and the inequality follows from $\mu_i \geq 0$. Thus, $\theta \in \Theta(\phi)$.

Convexity-Case (ii): Suppose that $q_1 = -q_2$. The linearity of $S_\theta(q)$ implies that $\theta_1 = -\theta_2$. By assumption there exists a $q_3 \neq q_1, -q_1$ with the property that $S_R(q_3)\phi \geq 0$. Let $\theta_3 = S_\theta(q_3)\phi$. By construction, $\theta_3 \in \Theta(\phi)$. If $\theta_3 = \theta_1$ ($\theta_3 = \theta_2$) we simply replace q_1 (q_2) by q_3 and follow the steps outlined for Case (i). If $\theta_1 < \theta_3 < \theta_2$, then the Case (i) argument implies that any θ in the intervals $[\theta_1, \theta_3]$ and $[\theta_3, \theta_2]$ and thereby any $\theta = \lambda\theta_1 + (1-\lambda)\theta_2$ is included in the identified set. Finally, if $\theta_3 < \theta_1$ ($\theta_2 < \theta_3$), we deduce from Case (i) that the interval $[\theta_3, \theta_2]$ ($[\theta_1, \theta_3]$) is included in the identified set.

Boundedness: We shall prove a slightly more general result. Throughout the proof we omit tildes and set the weight matrix W equal to the identity matrix. Recall the definition

$$G^\theta(\theta; \phi, I) = \min_{\|q\|=1} G(\theta, q; \phi, I)$$

and

$$F^\theta(\phi) = \left\{ \theta \,\middle|\, M_\theta \theta \geq 0 \text{ and } G^\theta(\theta; \phi, I) = 0 \right\}.$$

Suppose that $\tilde{\theta} \in F^\theta(\phi)$. Since $F^\theta(\phi)$ is a multiple-value set, we assume without loss of generality that $\tilde{\theta} > 0$. So, the sign restriction $\theta \geq 0$ is satisfied if it exists. Now we show by contradiction that $F^\theta(\phi)$ has an upper bound. Suppose, to the contrary, that no such upper bound exists. This guarantees the existence of a series $a_n > 0$ with $a_n \uparrow \infty$ such that $a_n \tilde{\theta}_n \in F^\theta(\phi)$ for each n. Thus,

$$G^\theta\big(a_n \tilde{\theta}; \phi, I\big) = 0. \tag{B.2}$$

By definition we have

$$\begin{aligned} G(a_n\tilde{\theta}; \phi, I) &= \min_{q=\|1\|,\, \mu\geq 0} \|S_\theta(q)\phi - a_n\tilde{\theta}\|^2 + \|S_R(q)\phi - \mu\|^2 \\ &\geq \min_{q=\|1\|} \|S_\theta(q)\phi - a_n\tilde{\theta}\|^2 \end{aligned}$$

Since $\|S_\theta(q)\phi\|$ is a continuous function of q for fixed ϕ and the set of q is a compact unit sphere, there exists a finite constant M such that $\|S_\theta(q)\phi\| < M$. From this we deduce that

$$\min_{q=\|1\|} \|S_\theta(q)\phi - a_n\tilde{\theta}\|^2 \longrightarrow \infty,$$

which contradicts (B.2). The existence of a lower bound can be established by considering a sequence $-a_n$. Moreover, $\theta < 0$ can be handled by a straightforward modification of the argument. □

Lemma B 2 *Suppose that Assumption 1 is satisfied. Consider the sequence* $\{(\phi_T, \theta_T, q_T)\}$ *with* $(\theta_T, q_T) \in F^{\theta,q}(\phi_T)$ *that satisfies conditions (i) and (iv) in the proof of Theorem 2(i). Then,*

$$G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) - G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) = o_p(1).$$

Proof of Lemma B 2. According to condition (ii) in the proof of Theorem 2(i) $V(q_T) = V = diag(V_\theta, V_R)$ for all T. If $V = 0$, i.e., $S(q_T) = 0$ for all T, it is trivial to deduce the required result because by definition

$$G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) = G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) = 0.$$

Now suppose that $V \neq 0$. Since $(\theta_T, q_T) \in F^{\theta,q}(\phi_T)$, the penalty term in (36) is zero. Also, $S_{\theta,T}\phi_T = V_{\theta,T}\theta_T$ and $S_{R,T}\phi_T \geq 0$. Notice that $D_T^{-1/2}$ and $\hat{D}_T^{-1/2}$ are well defined since S_T is a full (row) rank matrix and $\Lambda_T, \hat{\Lambda}_T > 0$. We now consider the two cases (i) $B_T = \hat{B}_T = I$ and (ii) $B_T = \Omega_T^{-1}$ and $\hat{B}_T = \hat{\Omega}_T^{-1}$ separately.

Case (i): $B_T = \hat{B}_T = I$. Write

$$\begin{aligned}
& G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) \\
&= \min_{\mu \geq 0} T \left\| \hat{D}_T^{-1/2} S_T \hat{\phi} - \hat{D}_T^{-1/2} \begin{pmatrix} V_{\theta,T}\theta_T \\ V_{R,T}\mu \end{pmatrix} \right\|^2 \\
&= \begin{cases} \min_{v \geq -\sqrt{T}\hat{D}_{R,T}^{-1/2}\mu(q_T,\phi_T)} \left\| \hat{D}_T^{-1/2} S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2 & \text{if } V_R \neq 0 \\ \left\| \hat{D}_T^{-1/2} S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) \right\|^2 & \text{if } V_R = 0 \end{cases}
\end{aligned}$$

and

$$G\left(\theta,q;\hat{\phi},W\left(\cdot\right)\right) = \begin{cases} \min_{v\geq -\sqrt{T}D_{R,T}^{-1/2}\mu(q,\phi)} \left\| D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2 & \text{if } V_R \neq 0 \\ \left\| D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right)\right\|^2 & \text{if } V_R = 0. \end{cases},$$

where $\mu\left(q_T,\phi_T\right) = S_{R,T}\phi_T$.

If $V_R \neq 0$, define

$$\begin{aligned} v_T(\hat{\Lambda}_T) &= \text{argmin}_{v\geq -\sqrt{T}\hat{D}_{R,T}^{-1/2}\mu(q_T,\phi_T)} \left\| \hat{D}_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2 \\ v_T\left(\Lambda_T\right) &= \text{argmin}_{v\geq -\sqrt{T}D_{R,T}^{-1/2}\mu(q,\phi)} \left\| D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2. \end{aligned}$$

Recall that $A' = D^{-1/2}SL$ and therefore $D^{-1/2}S = A'L^{-1}$. Then,

$$\begin{aligned} & G\left(\theta_T,q_T;\hat{\phi},\hat{W}\left(\cdot\right)\right) - G\left(\theta_T,q_T;\hat{\phi},W\left(\cdot\right)\right) \\ \leq & \left\| \hat{D}_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - \begin{pmatrix} 0 \\ v_T\left(\Lambda_T\right) \end{pmatrix} \right\|^2 - \left\| D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - \begin{pmatrix} 0 \\ v_T\left(\Lambda_T\right) \end{pmatrix} \right\|^2 \\ \leq & \left\| \hat{D}_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right)\right\|^2 \\ \leq & \left\| \left(\hat{A}_T - A_T\right)'\hat{L}_T^{-1}\sqrt{T}\left(\hat{\phi}-\phi_T\right)\right\| + \left\| A_T'\left(\hat{L}_T^{-1}-L_T^{-1}\right)\sqrt{T}\left(\hat{\phi}-\phi_T\right)\right\| \\ = & o_p\left(1\right). \end{aligned}$$

The last equality holds because $\hat{A}_T - A_T = o_p\left(1\right)$, $A_T = O\left(1\right)$, $\hat{L}_T^{-1} - L_T^{-1} = o_p\left(1\right)$, $L_T^{-1} = O\left(1\right)$, $\sqrt{T}\left(\hat{\phi}-\phi_T\right) = O_p\left(1\right)$, and $\hat{B}_T \stackrel{p}{\longrightarrow} B > 0$ according to Lemma 3 and Assumption 1(v-vi). If $V_R = 0$, the required result follows directly from

$$\begin{aligned} & G\left(\theta_T,q_T;\hat{\phi},\hat{W}\left(\cdot\right)\right) - G\left(\theta_T,q_T;\hat{\phi},W\left(\cdot\right)\right) \\ \leq & \left\| \hat{D}_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right) - D_T^{-1/2}S_T\sqrt{T}\left(\hat{\phi}-\phi_T\right)\right\|^2 = o_p\left(1\right). \end{aligned}$$

Using an argument similar to the one above, we can also show that

$$G\left(\theta_T,q_T;\hat{\phi},W\left(\cdot\right)\right) - G\left(\theta_T,q_T;\hat{\phi},\hat{W}\left(\cdot\right)\right) \leq o_p\left(1\right),$$

which proves the lemma for Case (i).

Case (ii): $B_T = \Omega_T^{-1}$ and $\hat{B}_T = \hat{\Omega}_T^{-1}$. In this case, we can write

$$
\begin{aligned}
& G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) \\
&= \min_{\mu \geq 0} T \left\| S_T\hat{\phi} - \begin{pmatrix} V_{\theta,T}\theta_T \\ V_{R,T}\mu \end{pmatrix} \right\|^2_{\hat{\Sigma}_T^{-1}} \\
&= \begin{cases} \min_{v \geq -\sqrt{T}\mu(q_T,\phi_T)} \left\| S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2_{\hat{\Sigma}_T^{-1}} & \text{if } V_R \neq 0 \\ \left\| S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right)\right\|^2_{\hat{\Sigma}_T^{-1}} & \text{if } V_R = 0 \end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
& G\left(\theta, q; \hat{\phi}, W(\cdot)\right) \\
&= \begin{cases} \min_{v \geq -\sqrt{T}\mu(q_T,\phi_T)} \left\| S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2_{\Sigma_T^{-1}} & \text{if } V_R \neq 0 \\ \left\| D_T^{-1/2} S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right)\right\|^2_{\Sigma_T^{-1}} & \text{if } V_R = 0 \end{cases},
\end{aligned}
$$

where $\mu(q_T, \phi_T) = S_{R,T}\phi_T$.

If $V_R \neq 0$, define

$$
\begin{aligned}
v_T(\hat{\Lambda}_T) &= \text{argmin}_{v \geq -\sqrt{T}\mu(q_T,\phi_T)} \left\| S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2_{\hat{\Sigma}_T^{-1}} \\
v_T(\Lambda_T) &= \text{argmin}_{v \geq -\sqrt{T}\mu(q_T,\phi_T)} \left\| S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2_{\Sigma_T^{-1}}.
\end{aligned}
$$

Then,

$$\begin{aligned}
& G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) - G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) \\
\leq & \left\| S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v_T(\Lambda_T) \end{pmatrix} \right\|^2_{\hat{\Sigma}_T^{-1}} - \left\| S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v_T(\Lambda_T) \end{pmatrix} \right\|^2_{\Sigma_T^{-1}} \\
= & \left[S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v_T(\Lambda_T) \end{pmatrix} \right]' \Sigma_T^{-1/2} \left[\Sigma_T^{1/2} \hat{\Sigma}_T^{-1} \Sigma_T^{1/2} - I_l \right] \\
& \times \Sigma_T^{-1/2} \left[S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v_T(\Lambda_T) \end{pmatrix} \right] \\
\leq & \left\| S_T \sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ v_T(\Lambda_T) \end{pmatrix} \right\|^2_{\Sigma_T^{-1}} \left\| \Sigma_T^{1/2} \hat{\Sigma}_T^{-1} \Sigma_T^{1/2} - I_l \right\| \\
= & I \times II, \text{ say.}
\end{aligned}$$

For term I, notice that since $\mu(q_T, \phi_T) \geq 0$, we have

$$I = \min_{v \geq -\sqrt{T}\mu(q_T, \phi_T)} \left\| S_T \sqrt{T}\left(\hat{\phi} - \phi\right) - \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2_{\Sigma_T^{-1}} \leq \left\| A_T' L_T^{-1} \sqrt{T}\left(\hat{\phi} - \phi_T\right) \right\|^2_{\Omega_T^{-1}} = O_p(1).$$

The last equality holds since $A_T' L_T^{-1} \sqrt{T}\left(\hat{\phi} - \phi_T\right) = O_p(1)$ and $\Omega_T^{-1} = (A_T' A_T)^{-1} \longrightarrow (A'A) > 0$ by condition (vi) in the proof of Theorem 2(i). Since $\Sigma_T = S_T L_T L_T' S_T'$, term II can be bounded as follows:

$$\begin{aligned}
II &= \left\| \Sigma_T^{1/2} \left(\hat{\Sigma}_T^{-1} - \Sigma_T^{-1} \right) \Sigma_T^{1/2} \right\| \\
&= \left\| \Sigma_T^{-1/2} \left(\Sigma_T - \hat{\Sigma}_T \right) \hat{\Sigma}_T^{-1/2} \hat{\Sigma}_T^{-1/2} \Sigma_T^{1/2} \right\| \\
&= \left\| \Sigma_T^{-1/2} S_T \left(\Lambda_T - \hat{\Lambda}_T \right) S_T' \hat{\Sigma}_T^{-1/2} \hat{\Sigma}_T^{-1/2} \Sigma_T^{1/2} \right\| \\
&= \left\| \left(\Sigma_T^{-1/2} S_T L_T \right) \left(L_T' \hat{L}_T'^{-1} - L_T^{-1} \hat{L}_T \right) \left(\hat{L}_T S_T' \hat{\Sigma}_T^{-1/2} \right) \hat{\Sigma}_T^{-1/2} \Sigma_T^{1/2} \right\| \\
&\leq \left\| \Sigma_T^{-1/2} S_T L_T \right\| \left\| L_T' \hat{L}_T'^{-1} - L_T^{-1} \hat{L}_T \right\| \left\| \hat{L}_T S_T' \hat{\Sigma}_T^{-1/2} \right\| \left\| \hat{\Sigma}_T^{-1/2} \Sigma_T^{1/2} \right\| \\
&= O(1)\, o_p(1)\, O_p(1)\, O_p(1).
\end{aligned}$$

The last line holds because

$$\begin{aligned}
\left\| \Sigma_T^{-1/2} S_T L_T \right\|^2 &= tr\left(L_T' S_T' \left(S_T L_T L_T' S_T' \right)^{-1} S_T L_T \right) = l \\
\left\| \hat{L}_T S_T' \hat{\Sigma}_T^{-1/2} \right\| &= tr\left(\hat{L}_T' S_T' \left(S_T \hat{L}_T \hat{L}_T' S_T' \right)^{-1} S_T \hat{L}_T \right) = l \\
\left\| L_T' \hat{L}_T'^{-1} - L_T^{-1} \hat{L}_T \right\| &= o_p(1) \text{ under Assumption 1(vi).}
\end{aligned}$$

Moreover,

$$\begin{aligned}\left\|\hat{\Sigma}_T^{-1/2}\Sigma_T^{1/2}\right\|^2 &= \left\|\hat{\Sigma}_T^{-1/2}\left(S_T L_T\right)\left[L_T' S_T'\left(S_T L_T L_T S_T'\right)^{-1}\right]\Sigma_T^{1/2}\right\|^2 \\ &= \left\|\hat{\Sigma}_T^{-1/2}\left(S_T \hat{L}_T\right)\left(\hat{L}_T^{-1} L_T\right)\left(L_T' S_T' \Sigma_T^{-1/2}\right)\right\|^2 \\ &\leq \left\|\hat{\Sigma}_T^{-1/2} S_T \hat{L}_T\right\|^2 \left\|\hat{L}_T^{-1} L_T\right\|^2 \left\|L_T' S_T' \Sigma_T^{-1/2}\right\|^2 \\ &= O_p(1)\, O_p(1)\, O(1) = O_p(1)\,.\end{aligned}$$

Similarly, if $V_R = 0$, we can show that

$$\begin{aligned}& G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) - G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) \\ \leq\ & \left\|S_T\sqrt{T}\left(\hat{\phi} - \phi_T\right) - \begin{pmatrix} 0 \\ \upsilon_T(\Lambda_T)\end{pmatrix}\right\|^2_{\Sigma_T^{-1}} \left\|\Sigma_T^{1/2}\hat{\Sigma}_T^{-1}\Sigma_T^{1/2} - I_l\right\| \\ =\ & o_p(1)\,.\end{aligned}$$

Using an argument similar to the one above, we can also show that

$$G\left(\theta_T, q_T; \hat{\phi}, W(\cdot)\right) - G\left(\theta_T, q_T; \hat{\phi}, \hat{W}(\cdot)\right) = o_p(1)\,,$$

which proves the lemma for Case (ii). □

Lemma B 3 *Suppose that a converging sequence $\{\phi_T, \theta_T, q_T\}$ satisfies the rank condition $l(q_T) = l > 0$ and $V(q_T)$ is a non-zero constant selection matrix for all T. Then, there exists a subsequence $\{T'\} \subset \{T\}$ such that along the subsequence, we have (i)*

$$D^{-1/2}(q_{T'})\, S(q_{T'})\, L(\phi_{T'}) \longrightarrow A,$$

where A is a full rank matrix, and (ii)

$$\hat{D}^{-1/2}(q_{T'})\, S(q_{T'})\, \hat{L}(\hat{\phi}_{T'}) = D^{-1/2}(q_{T'})\, S(q_{T'})\, L(\phi_{T'}) + o_p(1)\,.$$

Proof of Lemma B 3. Part (i): Recall that $S_T = V_T \tilde{S}_T$. We established in Section 4.2 that the rank reduction of $\tilde{S}_T$ is caused only by zero rows. Moreover, according to condition (ii) in the proof of Theorem 2(i) the non-zero row selection matrix is V_T constant over T. Thus, we can construct an index set $\mathcal{J}$ of non-zero rows of $\tilde{S}_T$. By construction, the size of $\mathcal{J}$ is l and

$$S_T = \left[\tilde{S}_T^j\right]_{j\in\mathcal{J}}.$$

In turn we obtain

$$D_T^{-1/2} S_T L_T = D_T^{-1/2} \big[\tilde{S}_T^j L_T\big]_{j \in \mathcal{J}}.$$

Recall from the definition of L and D that (omitting the T subscripts)

$$SLL'S' = D^{1/2} \Omega D^{1/2} \quad \text{and} \quad D^{-1/2} SLL'S' D^{-1/2} = \Omega,$$

where Ω is a correlation matrix with ones on its diagonal. Thus, $D_{ii}^{-1/2}$ normalizes the length of the i'th row of the matrix (SL) to one. Therefore,

$$D_{T'}^{-1/2} S_{T'} L_{T'} = \left[\frac{\tilde{S}_T^j L_T}{\|\tilde{S}_T^j L_T\|} \right]_{j \in \mathcal{J}} = \left[\frac{\tilde{S}_T^j}{\left\|\tilde{S}_T^j L_T\right\|} \right]_{j \in \mathcal{J}} L_T.$$

By construction, $\tilde{S}_T^j \neq 0$ for all T and $j \in \mathcal{J}$. Since $L_T > 0$, it follows that $\tilde{S}_T^j L_T \neq 0$ for all T and $j \in \mathcal{J}$. In turn, $\|\tilde{S}_T^j L_T\| > 0$ for all T and $j \in \mathcal{J}$ and $\tilde{S}_T^j L_T / \|\tilde{S}_T^j L_T\|$ is well defined for all T and $j \in \mathcal{J}$. Notice that $\left\{ \tilde{S}_T^j L_T / \|\tilde{S}_T^j L_T\| \right\}_T$ is a sequence on a unit sphere, which is compact. We can then choose a subsequence $\{T'\}$ such that $\tilde{S}_{T'}^j L_{T'} / \|\tilde{S}_{T'}^j L_{T'}\|$ converges for all $j \in \mathcal{J}$. Thus, we can write

$$D_{T'}^{-1/2} S_{T'} L_{T'} = \left[\frac{\tilde{S}_{T'}^j}{\left\|\tilde{S}_{T'}^j L_{T'}\right\|} \right]_{j \in \mathcal{J}} L_{T'} \longrightarrow A.$$

To obtain the desired result, it remains to be shown that A is full rank. Since $L_{T'}^{-1} \longrightarrow L^{-1} > 0$, it suffices to show that the limit

$$AL^{-1} = \lim_{T' \longrightarrow \infty} \left[\frac{\tilde{S}_{T'}^j}{\left\|\tilde{S}_{T'}^j L_{T'}\right\|} \right]_{j \in \mathcal{J}} \tag{B.3}$$

has full rank. Recall that $\tilde{S}(q) = \left(M^{S,2} M^{S,1} \otimes q'\right) S_\phi$. By definition of $M^{S,2} M^{S,1}$ and S_ϕ, the non-zero rows of $\tilde{S}_{T'}^j$ are orthogonal to each other because $\left\{ \tilde{S}_{T'}^j \right\}_{j \in \mathcal{J}}$ is composed of rows $(+/-) \left(I^j \otimes q\right) S_\phi$ that are orthogonal to each other. This implies that each row of the limit AL^{-1} is non-zero and orthogonal, which delivers the required result.

Part (ii): Consider the subsequence $\{T'\}$ in the proof of Part (i). Since $\hat{L}_{T'} > 0$ and $\tilde{S}_{T'}^j \neq 0$ for all T',

$$\|\tilde{S}_{T'}^j \hat{L}_{T'}\| > 0$$

for all T'. We will now show that

$$\frac{\tilde{S}_{T'}^j \hat{L}_{T'}}{\left\|\tilde{S}_{T'}^j \hat{L}_{T'}\right\|} = \frac{\tilde{S}_{T'}^j L_{T'}}{\left\|\tilde{S}_{T'}^j L_{T'}\right\|} + o_p(1)$$

for all $j \in \mathcal{J}$. Since it could be the case that $\|\tilde{S}^j_{T'} L_{T'}\| \longrightarrow 0$, we provide a detailed argument. Write

$$\frac{\tilde{S}^j_{T'}\hat{L}_{T'}}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} - \frac{\tilde{S}^j_{T'}L_{T'}}{\left\|\tilde{S}^j_{T'}L_{T'}\right\|} = \frac{\tilde{S}^j_{T'}L_{T'}}{\left\|\tilde{S}^j_{T'}L_{T'}\right\|}\left(\frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} - 1\right) + \frac{\tilde{S}^j_{T'}\left(\hat{L}_{T'} - L_{T'}\right)}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} = I + II, \text{ say.}$$

We begin with the following bound:

$$\begin{aligned}
\frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} - 1 &= \frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\| - \left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&= \frac{\left\|\tilde{S}^j_{T'}\hat{L}_{T'} - \tilde{S}^j_{T'}\left(\hat{L}_{T'} - L_{T'}\right)\right\| - \left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&\le \frac{\left\|\tilde{S}^j_{T'}\left(\hat{L}_{T'} - L_{T'}\right)\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&\le \frac{\left\|\tilde{S}^j_{T'}\right\|\left\|\hat{L}_{T'} - L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&= \frac{\left\|\hat{L}_{T'} - L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\| / \left\|\tilde{S}^j_{T'}\right\|}.
\end{aligned}$$

The last equality holds because $\|\tilde{S}^j_{T'}\| > 0$ for all T'.

According to Assumption 1(vi). $\hat{L}_{T'} \xrightarrow{p} L$. Moreover, we deduce from (B.3) and $A^j L^{-1} \neq 0$ that

$$\begin{aligned}
0 \quad &< \quad \frac{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\right\|} \le \frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\| + \left\|\tilde{S}^j_{T'}\left(\hat{L}_{T'} - L_{T'}\right)\right\|}{\left\|\tilde{S}^j_{T'}\right\|} \\
&\le \quad \frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\right\|} + \left\|\hat{L}_{T'} - L_{T'}\right\| \\
&\xrightarrow{p} \quad \frac{1}{\|A^j L^{-1}\|} > 0.
\end{aligned}$$

Therefore,

$$0 \le \frac{\left\|\hat{L}_{T'} - L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\| / \left\|\tilde{S}^j_{T'}\right\|} \le o_p(1)\left\|A^j L^{-1}\right\| = o_p(1).$$

Similarly, we obtain the bound

$$\begin{aligned}
1-\frac{\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} &= \frac{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|-\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&= \frac{\left\|\tilde{S}^j_{T'}L_{T'}+\tilde{S}^j_{T'}\left(\hat{L}_{T'}-L_{T'}\right)\right\|-\left\|\tilde{S}^j_{T'}L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \\
&\leq \frac{\left\|\tilde{S}^j_{T'}\left(\hat{L}_{T'}-L_{T'}\right)\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \xrightarrow{p} 0.
\end{aligned}$$

Since $\tilde{S}^j_{T'}L_{T'}/\|\tilde{S}^j_{T'}L_{T'}| = O(1)$, we have established that term I vanishes asymptotically:

$$I = o_p(1).$$

Term II can be bounded as follows:

$$\|II\| = \frac{\left\|\tilde{S}^j_{T'}\left(\hat{L}_{T'}-L_{T'}\right)\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \leq \frac{\left\|\tilde{S}^j_{T'}\right\|\left\|\hat{L}_{T'}-L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|} \leq \frac{\left\|\hat{L}_{T'}-L_{T'}\right\|}{\left\|\tilde{S}^j_{T'}\hat{L}_{T'}\right\|/\left\|\tilde{S}^j_{T'}\right\|} \xrightarrow{p} 0,$$

and so

$$II = o_p(1).$$

Combining the two $o_p(1)$ results completes the proof of Part (ii). $\square$

Lemma B 4 *If $CS^{\theta,q}$ in (33) is a valid $1-\alpha$ confidence set, then CS^{θ}_P in (34) is a valid $1-\alpha$ confidence set.*

Proof of Lemma B 4: The lemma follows since $(\theta, q) \in CS^{\theta,q}$ if and only if

$$M_\theta\theta \geq 0 \quad \text{and} \quad G(\theta, q; \hat{\phi}, \hat{W}) - c^\alpha_P(q),$$

where $\|q\| = 1$. Thus,

$$M_\theta\theta \geq 0 \quad \text{and} \quad \min_{\tilde{q}=\|1\|}\left[G(\theta, \tilde{q}; \hat{\phi}, \hat{W}) - c^\alpha_P(\tilde{q})\right] \leq 0$$

and therefore $\theta \in CS^\theta_P$. In turn,

$$\begin{aligned}
1-\alpha &\leq \liminf_T \inf_{\phi\in\mathcal{P}} \inf_{(\theta,q)\in F^{\theta,q}(\phi)} P_\phi\left\{(\theta,q)\in CS^{\theta,q}\right\} \\
&\leq \liminf_T \inf_{\phi\in\mathcal{P}} \inf_{\theta\in F^{\theta}(\phi)} P_\phi\left\{\theta\in CS^{\theta}_P\right\}. \quad \square
\end{aligned}$$

Lemma B 5 *Suppose Assumption 1 is satisfied. Consider Case (i) in Step 3 of the proof of Theorem 2(i). Along the $\{T\}$ sequence defined in Step 1 of the proof of Theorem 2(i),*

$$c_T^{\alpha}\left(\hat{A}\left(q_T\right), \hat{B}\left(q_T\right), \hat{\varphi}_T^*(q_T)\right) \to_p c_T^{\alpha}\left(A, B, \pi^*\right),$$

where the critical value function $c_T^{\alpha}(\cdot)$ is defined in (57) and (58).

Proof of Lemma B 5. The proof is very similar to that of Lemma 2(a) in Andrews and Soares (2010b) and we provide a sketch. The proof proceeds in three steps. First, show

$$\left(\hat{\xi}_T, \hat{A}(q_T), \hat{B}\left(q_T\right)\right) \xrightarrow{p} (\pi, A, B) \quad \text{and} \quad \hat{\varphi}_T^*(q_T) \xrightarrow{p} \pi^*.$$

Second, show

$$\mathbb{P}\left\{\min_{v \geq -\hat{\varphi}_T^*(q_T)} \left\|(\hat{A}(q_T)' Z_m - M_v v\right\|^2_{\hat{B}(q_T)} \leq x\right\} \xrightarrow{p} \mathbb{P}\left\{\min_{v \geq -\pi^*}\left\{\left\|A' Z_m - M_v v\right\|_B^2\right\} \leq x\right\}.$$

Third, deduce $c_T^{\alpha}\left(\hat{A}\left(q_T\right), \hat{B}\left(q_T\right), \hat{\varphi}_T^*(q_T)\right) \xrightarrow{p} c_T^{\alpha}\left(\hat{A}\left(q_T\right), \hat{B}\left(q_T\right), \pi^*\right)$, as required for the lemma.

Proof of Step 1: By the choice of the sequence $\{T\}$ and the limit result in Step 1 of the proof of Theorem 2(i)

$$\left(\hat{\xi}_T, \hat{A}(q_T), \hat{B}\left(q_T\right)\right) \xrightarrow{p} (\pi, A, B).$$

Notice that if $\pi_j = 0$, then $\xi_T(q_T) < \kappa_T$ as $T \longrightarrow \infty$ and by using an argument similar to the one used in the proof of Lemma 3(ii), we have $\hat{\xi}_T(q_T) < \kappa_T$ in probability as $T \longrightarrow \infty$. Therefore, plim $\hat{\varphi}_{j,T}^*(q_T) =$ plim $\hat{\varphi}_{j,T}(q_T) = 0 = \pi_j^*$ with probability one. On the other hand, if $\pi_j > 0$, then $\hat{\varphi}_{j,T}^*(q_T) = \infty = \pi_j^*$. Therefore, $\hat{\varphi}_T^*(q_T) \xrightarrow{p} \pi^*$.

Proof of Step 2: The desired result can be obtained by the same argument used in the proof of (S1.17) of Andrews and Soares (2010b).

Proof of Step 3: It is immediate from Step 2 and the fact that the distribution of

$$\min_{v \geq -\pi^*} \left\|A' Z_m - M_v v\right\|_B^2$$

is continuous if $k \geq 1$, and continuous near the $(1-\alpha)'s$ quantile, where $\alpha < 1/2$, if $k = 0$. $\square$

C Description of Monte Carlo Experiments

C.1 VAR(0) Experiment

Computations for the Monte Carlo experiment with the VAR(0) model, e.g., Design 1 in Table 1 of the main article: $y_t = u_t$, $u_t \sim N(0, \Sigma)$. The "true" identified set is given by $F^{\theta}(\phi) = \left[0, \max\left\{\mathcal{I}\{\phi_2 \geq 0\}, \sqrt{\frac{\phi_3^2}{\phi_2^2+\phi_3^2}}\right\}\right]$ where $\phi = [\phi_1, \phi_2, \phi_3]' = \left[\Sigma_{11}^{tr}, \Sigma_{21}^{tr}, \Sigma_{22}^{tr}\right]'$ and Σ_{ij}^{tr} are the elements of Σ_{tr}, the lower triangular matrix from the Cholesky decomposition of Σ.

It is convenient to reparameterize q in spherical coordinates: $q = q(\varphi) = \left[\cos(\varphi) \;\; \sin(\varphi)\right]'$. However, for brevity we typically write q, omitting the φ argument. We generate a grid $\mathcal{Q}$ for q by dividing the domain of φ, $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, into equally sized partitions of length δ_{φ}. As discussed in the main text, since $\phi_1 = \Sigma_{11}^{tr} > 0$ the inequality restriction $\theta = q_1\phi_1 \geq 0$ implies that $q_1 \geq 0$. Thus, it suffices to conduct the grid search with respect to φ over the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

The following steps are repeated N_{sim} times. The results reported in the main text are averages across these repetitions. We report the average length of the confidence intervals and compute the coverage probability as the fraction of times for which the upper bound of $F^{\theta}(\phi)$ is contained in the confidence interval. The upper bound of the identified set determines the lower bound of the coverage probability.

Generating Data: Generate a sample of length T of data from the VAR(0) using the parameters reported in Table 1.

Estimating the Reduced-Form Parameters

- Compute the sample covariance $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})'(y_t - \bar{y})$ where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$. Denote by $\hat{\Sigma}_{tr}$ the lower triangular matrix from the Cholesky decomposition of $\hat{\Sigma}$. Then $\hat{\phi} = [\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3]' = [\hat{\Sigma}_{11}^{tr}, \hat{\Sigma}_{21}^{tr}, \hat{\Sigma}_{22}^{tr}]'$, where $\hat{\Sigma}_{ij}^{tr}$ are the elements of $\hat{\Sigma}_{tr}$.
- Estimate Λ, the asymptotic variance covariance matrix of $\hat{\phi}$, using a parametric bootstrap:
  - Generate bootstrap samples $b = 1, \ldots, B$ of length T from $y_t^{(b)} = u_t^{(b)}$ where $u_t^{(b)} \sim N(0, \hat{\Sigma})$.
  - For each bootstrap sample, estimate $\hat{\Sigma}^{(b)}$ and compute $\hat{\phi}^{(b)} = [\hat{\phi}_1^{(b)}, \hat{\phi}_2^{(b)}, \hat{\phi}_3^{(b)}]'$.
  - Let $\hat{\Lambda} = \frac{1}{B}\sum_{b=1}^{B}[\sqrt{T}(\hat{\phi}^{(b)} - \hat{\phi})][\sqrt{T}(\hat{\phi}^{(b)} - \hat{\phi})]'$ with factorization $\hat{\Lambda} = \hat{L}\hat{L}'$.

Projection Approach

- Compute $F^{\theta}\left(\hat{\phi}\right)=\left[0,\max\left\{\mathcal{I}\left\{\hat{\phi}_2\geq 0\right\},\sqrt{\frac{\hat{\phi}_3^2}{\hat{\phi}_2^2+\hat{\phi}_3^2}}\right\}\right]$;
- Starting from the upper bound of $F^{\theta}(\hat{\phi})$, evaluate $\min_{\|q\|=1}\ \big(G\big(\theta_j,q;\hat{\phi},\hat{W}(\cdot)\big)-c_P^{\alpha}(q)\big)$ to determine whether $\theta_{(j)}$ belongs to

$$CS_P^{\theta}=\left\{\theta\mid\theta\geq 0\text{ and }\min_{\|q\|=1}\Big(G\big(\theta,q;\hat{\phi},\hat{W}\left(\cdot\right)\big)-c_P^{\alpha}\left(q\right)\Big)\leq 0\right\}.$$

  Let $\theta_{(j)}=\theta_{(j-1)}+\delta_\theta$, setting $\delta_\theta=0.001$.

- *Evaluation of the Objective Function*

  - Given θ evaluate the objective function for every $q\in\mathcal{Q}$:

$$G\left(\theta,q;\hat{\phi},\hat{W}\left(\cdot\right)\right)=\min_{\mu\geq 0}T\left\|\hat{D}^{-1/2}\left(q\right)V(q)\tilde{S}\left(q\right)\hat{\phi}-\hat{D}^{-1/2}\left(q\right)V\left(q\right)\begin{pmatrix}\theta\\ \mu\end{pmatrix}\right\|^2_{\hat{B}(q)}+T\mathcal{I}\left\{q_1=0\text{ and }\theta\neq 0\right\}\quad\text{(C.1)}$$

    where

$$\tilde{S}(q)=\begin{bmatrix}q_1&0&0\\ 0&q_1&q_2\end{bmatrix},\quad\hat{\Sigma}(q)=V(q)\tilde{S}(q)\hat{\Lambda}\tilde{S}'(q)V'(q),\quad\hat{D}(q)=diag(\hat{\Sigma}(q)),$$

    and $V(q(\varphi))$ deletes rows of zeros from $\tilde{S}(q)$:

$$V(q(\varphi))=\begin{cases}[0\ 1]&\text{if }\varphi\in\left\{-\frac{\pi}{2},\ \frac{\pi}{2}\right\}\\ \begin{bmatrix}1&0\\ 0&1\end{bmatrix}&\text{otherwise}\end{cases}$$

    We consider two choices of the matrix $\hat{B}(q)$: $\hat{B}(q)=I$ or $\hat{B}(q)=\hat{\Omega}(q)^{-1}$, where $\hat{\Omega}(q)=\hat{D}^{-1/2}(q)\hat{\Sigma}(q)\hat{D}^{-1/2}(q)$.

  - If $\varphi\in\left\{-\frac{\pi}{2},\frac{\pi}{2}\right\}$, then

$$\hat{\Sigma}\left(q\right)=\begin{bmatrix}0&q_1&q_2\end{bmatrix}\hat{\Lambda}\begin{bmatrix}0\\ q_1\\ q_2\end{bmatrix}=\hat{\Sigma}_{22}\left(q\right),\quad\hat{D}\left(q\right)=1/\hat{\Sigma}_{22}\left(q\right)$$

and therefore $\hat{B}(q) = 1$. Thus, the objective function (C.1) simplifies to

$$G\left(\theta, q; \hat{\phi}, \hat{W}(\cdot)\right) = \min_{\mu \geq 0} \frac{T}{\hat{\Sigma}_{22}(q)} \left(q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu\right)^2.$$

The value μ_* that solves the minimization problem is:

$$\mu_* = \begin{cases} q_1\hat{\phi}_2 + q_2\hat{\phi}_3 & \text{if } q_1\hat{\phi}_2 + q_2\hat{\phi}_3 \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

– If $\varphi \notin \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}$, then it is convenient to rewrite the objective function $G(\cdot)$. Recall that $V(q) = I_2$ and $S(q) = \tilde{S}(q)$. Let $A(q) = \hat{D}^{-1/2}S(q)\phi$ with row elements A_1 and A_2. Let $\bar{\theta} = \hat{D}_{11}^{-1/2}\theta$ and $\bar{\mu} = \hat{D}_{22}^{-1/2}\mu$. Moreover, notice that

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

where

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

Thus, we can write the objective function as (omitting the indicator function)

$$\begin{aligned} G\left(\theta, q; \hat{\phi}, \hat{W}(\cdot)\right) &= \min_{\bar{\mu} \geq 0} \Bigg[T(\hat{B}_{11} - \hat{B}_{12}\hat{B}_{22}^{-1}\hat{B}_{21})\left(A_1(q) - \bar{\theta}\right)^2 \\ &\quad + T\hat{B}_{22}\left(A_2(q) + \frac{\hat{B}_{21}(q)}{\hat{B}_{22}(q)}\left(A_1(q) - \bar{\theta}\right) - \bar{\mu}\right)^2 \Bigg]. \end{aligned}$$

The value μ_* that solves the minimization problem is:

$$\mu_* = \begin{cases} A_2(q) + \frac{\hat{B}_{21}(q)}{\hat{B}_{22}(q)}\left(A_1(q) - \bar{\theta}\right) & \text{if } A_2(q) + \frac{\hat{B}_{21}(q)}{\hat{B}_{22}(q)}\left(A_1(q) - \bar{\theta}\right) \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad \text{(C.2)}$$

- *Simulating the Critical Values.* It is convenient to define the standardized responses

$$\begin{aligned} \xi_{1,T}(q) &= \left(\frac{T}{\hat{\Sigma}_{11}(q)}\right)^{1/2} \begin{bmatrix} q_1 & 0 & 0 \end{bmatrix}' \hat{\phi} = \left(\frac{T}{\hat{\Sigma}_{11}(q)}\right)^{1/2} \left(q_1\hat{\phi}_1\right) \quad \text{(C.3)} \\ \xi_{2,T}(q) &= \left(\frac{T}{\hat{\Sigma}_{22}(q)}\right)^{1/2} \begin{bmatrix} 0 & q_1 & q_2 \end{bmatrix}' \hat{\phi} = \left(\frac{T}{\hat{\Sigma}_{22}(q)}\right)^{1/2} \left(q_1\hat{\phi}_2 + q_2\hat{\phi}_3\right) \end{aligned}$$

Define the threshold $\kappa_T = 1.96\ln(\ln(T))$.

– If $\varphi \in \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}$, then:

  * if $\xi_{2,T}(q) < \kappa_T$, the inequality condition is considered binding and $c_P^{\alpha}(q)$ is the $(1-\alpha)\, th$ quantile of a squared truncated normal $Z^2\mathcal{I}\{Z \geq 0\}$;
  * if $\xi_{2,T}(q) \geq \kappa_T$, the inequality condition is considered non-binding and $c_P^{\alpha}(q) = 0$.
- If $\varphi \notin \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}$, then:
  * if $\xi_{2,T}(q) < \kappa_T$, the inequality is considered binding and the critical value is obtained by simulation: for $j = 1, \ldots, n_z$ draw from $Z_3^{(j)} \sim N(0, I_3)$ and obtain $c_P^{\alpha}(q)$ as the $1-\alpha$ quantile of

$$\bar{\mathcal{G}}^{(j)}\left(q; \hat{B}(q)\right) = \min_{\nu \geq 0} \left\| \hat{D}^{-1/2}(q)\, S(q)\, \hat{L} Z_3^{(j)} - \begin{pmatrix} 0 \\ \nu \end{pmatrix} \right\|^2_{\hat{B}(q)}.$$

    The optimal ν for this minimization problem takes the same form as μ_* in (C.2).
  * If $\xi_{2,T}(q) \geq \kappa_T$, the inequality condition is considered non-binding and $c_P^{\alpha}(q)$ is the $(1-\alpha)$ quantile of a $\chi^2_{(1)}$ distribution.

Bonferroni Approach

- *Step 1: Construct a* $(1-\alpha_1)$ *confidence set for* q. The following computations are executed for each $q \in \mathcal{Q}$. As before, it is convenient to express q in terms of the angle φ and generate $\mathcal{Q}$ by equally spaced grid points on the interval $[-\pi, \pi]$. Recall the definition of $\xi_{1,T}(q)$ and $\xi_{2,T}(q)$ in (C.3).
  - If $\varphi \in \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}$, the objective function is given by

$$G^q\left(q; \hat{\phi}, \hat{W}\right) = \min_{\mu \geq 0} \frac{T}{\hat{\Sigma}_{22}(q)} \left(q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu\right)^2.$$

    * If $\xi_{2,T}(q) < \kappa_T$, the inequality condition is considered binding and the critical value $c_B^{\alpha_1}(q)$ is the $(1-\alpha_1)$ quantile of a squared truncated normal $Z^2\mathcal{I}\{Z \geq 0\}$.
    * If $\xi_{2,T}(q) \geq \kappa_T$, the inequality condition is considered non-binding and $c_B^{\alpha_1}(q) = 0$.
  - If $\varphi \notin \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}$, the objective function is given by

$$G^q\left(q; \hat{\phi}, \hat{W}\right) = \min_{\mu_1 \geq 0, \mu_2 \geq 0} T \left\| \hat{D}^{-1/2}(q) \begin{bmatrix} q_1\hat{\phi}_1 - \mu_1 \\ q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu_2 \end{bmatrix} \right\|^2_{\hat{B}(q)}.$$

  * If $\xi_{1,T}(q) < \kappa_T$ and $\xi_{2,T}(q) < \kappa_T$, both inequality conditions are considered binding. For $j = 1, \ldots, n_z$ draw $Z_3^{(j)}$ from $N(0, I_3)$. The critical value is the $(1-\alpha_1)$ quantile of

$$\bar{\mathcal{G}}^{(j)}\left(q; \hat{B}(q)\right) = \min_{\nu \geq 0} \left\| \hat{D}^{-1/2}(q) S(q) \hat{L} Z_3^{(j)} - \nu \right\|^2_{\hat{B}(q)}.$$

    The minimization can be executed with a numerical routine that solves quadratic programming problems.

  * If $\xi_{1,T}(q) < \kappa_T$ and $\xi_{2,T}(q) \geq \kappa_T$ or if $\xi_{1,T}(q) \geq \kappa_T$ and $\xi_{2,T}(q) < \kappa_T$, i.e., only one inequality condition is considered binding, then $c_B^{\alpha_1}(q)$ is the $(1-\alpha_1)\,th$ quantile of a squared truncated normal $Z^2 \mathcal{I}\{Z \geq 0\}$.

  * if $\xi_{1,T}(q) \geq \kappa_T$ and $\xi_{2,T}(q) \geq \kappa_T$, then no inequality condition is considered binding and $c_B^{\alpha_1}(q) = 0$.

  – Let $CS_B^q = \left\{ q \in \mathcal{Q} \mid \left(G^q\left(q; \hat{\phi}, \hat{W}\right) - c_B^{\alpha_1}(q) \right) \leq 0 \right\}$.

- *Step 2: Construct a $(1-\alpha_2)$ confidence set for θ conditional on q:*

$$CS_{B,q}^{\theta} = \left[\max\left\{ 0, q_1\hat{\phi}_1 - z_{\alpha_2/2}\sqrt{q_1^2 \hat{\Lambda}_{11}/T} \right\},\; q_1\hat{\phi}_1 + z_{\alpha_2/2}\sqrt{q_1^2 \hat{\Lambda}_{11}/T} \right],$$

  where $z_{\alpha_2/2}$ is the $(1-\alpha_2/2)$ quantile of a $N(0,1)$ distribution and $\hat{\Lambda}_{11}$ is the $(1,1)$ element of the matrix $\hat{\Lambda}$.

- *Step 3: Construct the $1-\alpha$ Bonferroni set for θ*: Compute the minimum of the lower bounds of $CS_{B,q}^{\theta}$ and the maximum of the upper bounds of $CS_{B,q}^{\theta}$ for $q \in CS_B^q$.

C.2 VAR(1) Experiments

The computations are very similar to the computations for the VAR(0) experiment. Thus, we focus on highlighting the differences. The model takes the form (Designs 2 to 4 in Table 1 of the main article): $y_t = \Phi y_{t-1} + u_t$, where $u_t \sim N(0, \Sigma)$. Let Σ_{tr} denote the lower-triangular Cholesky factor of Σ. The reduced-form parameters are given by

$$\phi = \text{vec}\big((\Phi\Sigma_{tr})'\big) = \big[\phi_1, \phi_2, \phi_3, \phi_4\big]' = \big[\Phi_{11}\Sigma_{11}^{tr} + \Phi_{12}\Sigma_{21}^{tr}, \Phi_{12}\Sigma_{22}^{tr}, \Phi_{21}\Sigma_{11}^{tr} + \Phi_{22}\Sigma_{21}^{tr}, \Phi_{22}\Sigma_{22}^{tr}\big]',$$

where Σ_{ij}^{tr} are the elements of Σ_{tr}. Under our three Monte Carlo designs the identified set $F^q(\phi)$ has a geometry similar to that of the identified set for the VAR(0) design, depicted in Figure 1 of the main article. Roughly speaking, it is an arc located in the Northeast section of the unit circle. Under the parameterization of the data-generating processes (DGPs), the top-left endpoint of $F^q(\phi)$ is given by the solution of

$$q_{1,l}^2 = \frac{1}{1+(\phi_1/\phi_2)^2},$$

whereas the bottom-right endpoint of $F^q(\phi)$ is given by the solution of

$$q_{1,r}^2 = \frac{1}{1+(\phi_3/\phi_4)^2}.$$

The structural parameter of interest is $\theta = q_1\phi_1 + q_2\phi_2$. For our Monte Carlo designs the lower bound of the identified set $F^\theta(\phi)$ is determined by $\theta_l = q_{1,l}\phi_1 + q_{2,l}\phi_2$. The upper bound is $\theta_u = q_{1,r}\phi_1 + q_{2,r}\phi_2$ if $q_{2,r} > 0$; or is $\theta_u = q_{1,r}\phi_1 + q_{2,r}\phi_2$ otherwise.

As for the VAR(0) experiment, minimizations with respect to q are carried out using a grid $q \in \mathcal{Q}$, where $q = \begin{bmatrix} \cos(\varphi) & \sin(\varphi) \end{bmatrix}'$ and φ takes values on an equally spaced grid over $[-\pi, \pi]$ with spacing δ_φ.

Generating Data: The DGP is now given by $y_t = \Phi y_{t-1} + u_t$.

Estimating the Reduced-Form Parameters: Follow the same steps as in the VAR(0) experiment.

Projection Approach: Follow the same steps as in the VAR(0) experiment, with the following modifications:

- $\hat{F}^\theta\left(\hat{\phi}\right)$ is obtained as follows:
  - Compute $\hat{F}^q\left(\hat{\phi}\right) = \left\{ q^{(j)} \in \mathcal{Q} \mid \begin{bmatrix} q_1^{(j)}\hat{\phi}_1 + q_2^{(j)}\hat{\phi}_2 \\ q_1^{(j)}\hat{\phi}_3 + q_2^{(j)}\hat{\phi}_4 \end{bmatrix} \geq 0 \right\}$.
  - For each $q^{(j)} \in \hat{F}^q\left(\hat{\phi}\right)$ let $\theta^{(j)} = q_1^{(j)}\hat{\phi}_1 + q_2^{(j)}\hat{\phi}_2$.
  - Then $\hat{F}^\theta\left(\hat{\phi}\right) = \left[\left(\min_{j=1,..,n_q} \theta^{(j)}\right), \left(\max_{j=1,..,n_q} \theta^{(j)}\right)\right]$.
- Progressively expand the boundaries of $\hat{F}^\theta(\hat{\phi})$ by $\delta_\theta = 0.001$ and determine whether

$$\min_{\|q\|=1} \left(G\left(\theta, q; \hat{\phi}, \hat{W}(\cdot)\right) - c_P^\alpha(q) \right) \leq 0.$$

- *Evaluation of Objective Function*: The objective function for the VAR(1) experiment takes the form

$$\begin{aligned} G\left(\theta, q; \hat{\phi}, \hat{W}(\cdot)\right) &= \min_{\mu \geq 0} T \left\| \hat{D}^{-1/2}(q) S(q) \hat{\phi} - \hat{D}^{-1/2}(q) \begin{pmatrix} \theta \\ \mu \end{pmatrix} \right\|^2_{\hat{B}(q)} \\ &\quad + T\mathcal{I}\left\{ q_1\hat{\phi}_1 + q_2\hat{\phi}_2 = 0 \text{ and } \theta \neq 0 \right\}, \end{aligned}$$

where $S(q) = \begin{bmatrix} q_1 & q_2 & 0 & 0 \\ 0 & 0 & q_1 & q_2 \end{bmatrix}$. In our design the elements of ϕ and $\hat{\phi}$ are non-zero, thus $V(q) = I$ for all q.

- *Simulating the Critical Values.* Define the standardized responses

$$\begin{aligned} \xi_{1,T}(q) &= \left(\frac{T}{\hat{\Sigma}_{11}(q)}\right)^{1/2} \begin{bmatrix} q_1 & q_2 & 0 & 0 \end{bmatrix}' \hat{\phi} = \left(\frac{T}{\hat{\Sigma}_{11}(q)}\right)^{1/2} \left(q_1\hat{\phi}_1 + q_2\hat{\phi}_2\right) \qquad \text{(C.4)} \\ \xi_{2,T}(q) &= \left(\frac{T}{\hat{\Sigma}_{22}(q)}\right)^{1/2} \begin{bmatrix} 0 & 0 & q_1 & q_2 \end{bmatrix}' \hat{\phi} = \left(\frac{T}{\hat{\Sigma}_{22}(q)}\right)^{1/2} \left(q_1\hat{\phi}_3 + q_2\hat{\phi}_4\right) \end{aligned}$$

  Also, define the threshold $\kappa_T = 1.96 \ln(\ln(T))$.

  - If $\xi_{2,T}(q) < \kappa_T$, the inequality condition is considered binding and the critical value is obtained through simulation: for $j = 1, \ldots, n_z$ draw from $Z_4^{(j)} \sim N(0, I_4)$ and obtain $c_P^{\alpha}(q)$ as the $1-\alpha$ quantile

$$\bar{\mathcal{G}}^{(j)}\left(q; \hat{B}(q)\right) = \min_{\nu \geq 0} T \left\| \hat{D}^{-1/2}(q) S(q) \hat{L} Z_4^{(j)} - \begin{pmatrix} 0 \\ \nu \end{pmatrix} \right\|^2_{\hat{B}(q)}$$

    The optimal value of ν in this minimization problem has the same form as μ_* in (C.2).
  - If $\xi_{2,T}(q) \geq \kappa_T$ the inequality condition is considered non-binding and $c_P^{\alpha}(q)$ is $(1-\alpha)$ quantile of the $\chi^2_{(1)}$ distribution.

Bonferroni Approach

- *Step 1: Construct a* $1 - \alpha_1$ *confidence set for* q.
  - The objective function is

$$G^q\left(q; \hat{\phi}, \hat{W}\right) = \min_{\mu_1 \geq 0, \mu_2 \geq 0} T \left\| \hat{D}^{-1/2}(q) \begin{bmatrix} q_1\hat{\phi}_1 + q_2\hat{\phi}_2 - \mu_1 \\ q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu_2 \end{bmatrix} \right\|^2_{\hat{B}(q)}$$

- If $\xi_{1,T}(q) < \kappa_T$ and $\xi_{2,T}(q) < \kappa_T$, both inequality conditions are considered binding. For $j = 1, \ldots, n_z$ draw $Z_4^{(j)}$ from $N(0, I_4)$. The critical value is the $(1-\alpha_1)$ quantile of

$$\bar{\mathcal{G}}^{(j)}\left(q; \hat{B}(q)\right) = \min_{\nu \geq 0} T \left\| \hat{D}^{-1/2}(q) S(q) \hat{L} Z_4^{(j)} - \nu \right\|^2_{\hat{B}(q)}$$

- If $\xi_{1,T}(q) < \kappa_T$ and $\xi_{2,T}(q) \geq \kappa_T$ or if $\xi_{1,T}(q) \geq \kappa_T$ and $\xi_{2,T}(q) < \kappa_T$, i.e., only one inequality condition is considered binding, then $c_B^{\alpha_1}(q)$ is the $(1-\alpha_1)$ quantile of a squared truncated normal $Z^2 \mathcal{I}\{Z \geq 0\}$.
- If $\xi_{1,T}(q) \geq \kappa_T$ and $\xi_{2,T}(q) \geq \kappa_T$, then no inequality condition is considered binding and $c_B^{\alpha_1}(q) = 0$.

- *Step 2: Construct the* $(1-\alpha_2)$ *confidence set for* θ *conditional on* q. Follow the same steps as in the VAR(0) experiment.
- *Step 3: Construct the* $1-\alpha$ *Bonferroni set for* θ. Follow the same steps as in the VAR(0) experiment.

D Further Details on the Empirical Analysis

D.1 Data

The construction of the data set follows Aruoba and Schorfheide (2011). 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 the 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, 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.

E Further Discussion of the VAR(0) Example in Section 2

For the simple VAR(0) example in Section 2, we can provide a more detailed comparison of the two proposed confidence intervals.

Case 1: $\phi_2 > 0$. Write the endpoint of the projection-based confidence set as $\hat{\phi} + \delta/\sqrt{T}$. This endpoint has to satisfy the relationship

$$c_P^\alpha = \min_{q=\|1\|} \min_{\mu \geq 0} T\|\hat{\phi}_1 - (\hat{\phi}_1 + \delta/\sqrt{T})/|q_1|\|^2 + T\|q_1\hat{\phi}_2 + q_2\hat{\phi}_3 - \mu\|^2 \tag{D.1}$$

The right panel of Figure 1 illustrates that with probability approaching one for any q_1, there exists a q_2 such that the inequality $q_1\hat{\phi}_2 + q_2\hat{\phi}_3 \geq 0$ is satisfied. Thus, the minimization with respect to q and μ annihilates the second term in (D.1) and we obtain the condition (w.p.a. 1)

$$c_P^\alpha = \min_{0<q_1\leq 1} T\|\hat{\phi}_1 - (\hat{\phi}_1 + \delta/\sqrt{T})/|q_1|\|^2 = \delta^2. \tag{D.2}$$

The last equality holds because the minimum with respect to q_1 is attained at the boundary. This leaves us with

$$CS_P^\theta = \left[0,\, \hat{\phi} + \sqrt{c_P^\alpha}/\sqrt{T}\right].$$

The moment selection detects that at the boundary of the identified set the inequality $q_1\phi_2 + q_2\phi_3 \geq 0$ is not binding. Thus, the slackness $\tilde{\mu}$ is strictly positive. In turn, the distribution of the $G(\cdot)$ function is asymptotically $Z_1^2 < Z_1^2 + Z_2^2\mathcal{I}\{Z_2 \leq 0\}$. Thus, $\sqrt{c_P^\alpha} = z_{\alpha/2}$.

The right panel of Figure 1 is also useful for understanding the Bonferroni interval. If $\phi_2 > 0$ the uncertainty about the endpoint of $F^q(\phi)$ is asymptotically not important because w.p.a. 1 the least favorable value for the construction of the θ interval, $q = [1, 0]'$, is always included. Thus, the Bonferroni interval is given by

$$CS_B^\theta = \left[0,\, \hat{\phi} + z_{\alpha_2/2}/\sqrt{T}\right].$$

Since $\alpha_2 \leq \alpha$, the Bonferroni interval is more conservative for any $\alpha_1 > 0$.

Both the projection-based confidence interval and the Bonferroni interval are conservative compared to the interval that one would obtain following Imbens and Manski's (2004) approach of using the distribution of the boundary estimator. The lower endpoint of $F^\theta(\phi)$ in (5) is $\theta_l = 0$ for all values of ϕ. Let $\theta_u(\phi)$ denote the upper bound. A confidence interval $CS^\theta(\hat{\phi})$ can in principle be obtained by constructing a one-sided confidence interval for θ_u based on the sampling distribution of the endpoint estimator

$$\hat{\theta}_u = \begin{cases} \hat{\phi}_1 & \text{if } \phi_2 \geq 0 \\ \hat{\phi}_1 \sqrt{\frac{\hat{\phi}_3^2}{\hat{\phi}_2^2 + \hat{\phi}_3^2}} & \text{if } \phi_2 < 0 \end{cases},$$

which leads to $CS^\theta_{IM} = \big[0,\, \hat{\phi} + z_\alpha/\sqrt{T}\big]$. Provided the length of the identified set is sufficiently large, the Imbens-Manski interval is based on a one-sided critical value.

Case 2: $\phi_2 < 0$**.** It is more difficult to compare the length of the projection-based and the Bonferroni confidence interval. Let θ^*_B and q^*_B denote the endpoint of the Bonferroni interval and the associated value of q. At this endpoint

$$z^2_{\alpha_2/2} = T\big\|\hat{\phi}_1 - \theta^*_B/|q^*_{1,B}|\big\|^2$$

and

$$c^q_B(\alpha_1) = \max_{\theta \geq 0,\, \mu \geq 0} T\big\|\hat{\phi}_1 - \theta/|q^*_{1,B}|\big\|^2 + T\big\|q^*_{1,B}\hat{\phi}_2 + q^*_{2,B}\hat{\phi}_3 - \mu\big\|^2 = \max_{\mu \geq 0} T\big\|q^*_{1,B}\hat{\phi}_2 + q^*_{2,B}\hat{\phi}_3 - \mu\big\|^2.$$

The second equality holds because $q^*_{1,B}\hat{\phi}_1$ is non-negative (see Figure 1). Thus, the endpoint of the Bonferroni interval satisfies

$$G\big(\theta^*_B, q^*_B; \hat{\phi}, W^*(\cdot)\big) = z^2_{\alpha_2/2} + c^{\alpha_1}_{B,q} \tag{D.3}$$

In addition, the following two inequalities hold

$$\begin{aligned} G\big(\theta^*_B, q^*_B; \hat{\phi}, W^*(\cdot)\big) &\geq \min_{\|q\|=1} G\big(\theta^*_B, q; \hat{\phi}, W^*(\cdot)\big) \\ z^2_{\alpha_2/2} + c^q_B(\alpha_1) &\geq c^\alpha_P, \end{aligned}$$

which provides no clear ranking of the projection-based and Bonferroni interval.

Additional References

Aruoba, Boragan and Frank Schorfheide (2011): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, **3**, 60-90.

Cynamon, Barry, Donald Dutkowsky, and Barry Jones (2006): "Redefining the Monetary Aggregates: A Clean Sweep," *Eastern Economic Journal*, **32**, 661-672.