WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 11-20

INFERENCE FOR VARS IDENTIFIED WITH SIGN RESTRICTIONS

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide
University of Pennsylvania, CEPR, NBER, and
Visiting Scholar, Federal Reserve Bank of Philadelphia

Eleonora Granziera
Bank of Canada

Mihye Lee
University of Southern California

June 6, 2011

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

Inference for VARs Identified with Sign Restrictions

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania,
CEPR, NBER, and Visiting Scholar
Federal Reserve Bank of Philadelphia

Eleonora Granziera
Bank of Canada

Mihye Lee
University of Southern California

June 6, 2011

*Correspondence: H.R. Moon, M. Lee: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: moonr@usc.edu, mihyelee@usc.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 Fabio Canova and Eric Renault as well as seminar participants at the 2010 California Econometric Conference, the 2009 Canadian Econometric Study Group Meeting, Boston University, the 2010 Econometric Society World Congress, Harvard-MIT, the 2010 NBER Summer Institute, Queens University, the University of Pennsylvania, and Yale University. We also thank Hilda Guay and Minchul Shin for proofreading. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 0617803. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or the Bank of Canada. This paper is available free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/. 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 partially 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). To date, the methods that have been used are only justified from a Bayesian perspective. This paper develops methods of constructing error bands for impulse response functions of sign-restricted SVARs that are valid from a frequentist perspective. We also provide a comparison of frequentist and Bayesian error bands in the context of an empirical application - the former can be twice as wide as the latter. (JEL: C1, C32)

KEY WORDS: Bayesian Inference, Frequentist Inference, Partially 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, interpretable, structural innovations. Most SVARs in the literature have been constructed by sufficiently imposing many restrictions such that the relationship between structural innovations and forecast errors is one-to-one. However, in the past decade, starting with Faust (1998), Canova and 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 will refer to this class of models as sign-restricted SVARs. They have been employed, for instance, to measure the effects of monetary policy shocks (Faust, 1998; Canova and De Nicolo, 2002; Uhlig, 2005), technology shocks (Dedola and Neri, 2007; Peersman and Straub, 2009), government spending shocks (Mountford and Uhlig, 2008; Pappa, 2009), and oil price shocks (Baumeister and Peersman, 2008; Kilian and Murphy, 2009).

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

The main goal of the paper is to provide methods of constructing error bands that delimit valid frequentist confidence intervals.[1] In this regard, the paper makes several specific contributions. First, we demonstrate how to formulate the frequentist inference problem for impulse responses of sign-restricted SVARs as a minimum distance problem. Second, we provide a general characterization of the identified sets associated with impulse responses as well as an efficient way of computing them. For the case of scalar point-wise responses, we show that the identified set is a bounded interval. Third, building on recent work in microeconometrics by Chernozhukov, Hong, and Tamer (2007, henceforth CHT), Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010a), we propose three different methods to obtain frequentist error bands. We prove that all three methods generate asymptotically valid confidence sets. They differ with respect to conservativeness and computational burden - the most conservative error bands are the fastest to compute. Fourth, in an empirical application we compare the proposed frequentist error bands to Bayesian error bands for the effects of a monetary policy shock in a four-variable VAR. In our application, the frequentist bands are up to twice as wide as the Bayesian bands, which is consistent with the large sample results obtained in Moon and Schorfheide (2009).

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, discuss the commonly used Bayesian inference in this model, outline our procedure to construct frequentist error bands, and discuss more precisely in what dimension we modify and extend inference methods developed in the microeconometric literature. Section 3 generalizes the setup and introduces additional notation. In Section 4 we develop our frequentist inference procedures and provide three types of error bands that differ with respect to conservativeness and computational burden. The section also provides a detailed discussion of how to compute the error bands efficiently. To illustrate our methods, we conduct a small Monte Carlo study in Section 5 and generate error bands for output, inflation, interest rate, and money responses to a monetary policy shock in an empirical application in Section 6. Finally, Section 7 concludes. Proofs for the two theorems stated in Section 4 are presented in the appendix of this paper. Proofs of the lemmas that appear in the main text as well as additional lemmas used in the proofs of Theorems 1 and 2 are relegated to a supplemental Online Appendix. This Online Appendix also contains detailed derivations for the Monte Carlo study presented in Section 5.

[1] 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.

We use the following notation throughout the remainder of the paper: "$\stackrel{p}{\longrightarrow}$" and "$\Longrightarrow$" denote convergence in probability and distribution, respectively. "$\equiv$" signifies distributional equivalence. $\mathcal{I}\{x \geq a\}$ is the indicator function that is one if $x \geq a$ and zero otherwise. We use $\text{sgn}(x)$ to denote the sign of x and $\propto$ to indicate proportionality. $0_{n \times m}$ is a $n \times m$ matrix of zeros and I_n is the $n \times n$ identity matrix. $\otimes$ is the Kronecker product, $vec(\cdot)$ stacks the columns of a matrix, and $tr[\cdot]$ is the trace operator. 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 a $n \times m$ vector, 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. A p-variate normal distribution is denoted by $N_p(\mu, \Sigma)$. A $p \times q$ matrix X is matrix-variate normal $MN_{p \times q}(M, Q \otimes P)$ if $vec(X) \sim N_{pq}(vec(M), Q \otimes P)$. A $q \times q$ matrix Σ has the Inverted Wishart $IW_q(S, \nu)$ distribution if $p(\Sigma|S, \nu) \propto |\Sigma|^{-(\nu+q+1)/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\}$. If $X|\Sigma \sim MN_{p \times q}(M, \Sigma \otimes P)$ and $\Sigma \sim IW_q(S, \nu)$, we say that $(X, \Sigma) \sim MNIW(M, P, S, \nu)$. If there is no ambiguity about the dimension of the random vectors and matrices, we drop the subscripts that signify dimensions. We use χ^2_m to denote a χ^2 distribution with m degrees of freedom.

2 A Simple Example

Consider a VAR in which the vector y_t of endogenous variables is composed of inflation and output growth. For simplicity, we shall assume that the VAR has lag order of zero, that is, $y_t = u_t$ where $u_t \sim iidN(0, \Sigma_u)$. Moreover, it is assumed that the one-step ahead forecast errors are linear functions of "structural" demand and supply shocks, stacked in the vector $\epsilon_t = [\epsilon_{D,t}, \epsilon_{S,t}]'$. Let Σ_{tr} be the lower triangular Cholesky factor of Σ_u with elements Σ^{tr}_{ij} and Ω_ϵ an arbitrary orthogonal matrix. Thus, $y_t = \Sigma_{tr}\Omega_\epsilon\epsilon_t$, where $\epsilon_t \sim iidN(0, I)$. The covariance matrix of y_t is by construction invariant to Ω_ϵ. In order to restrict the set of admissible Ω_ϵ's, one can impose the sign restrictions that a demand shock moves prices and output in the same direction and the normalization that a positive demand shock increases prices.

Without loss of generality, suppose that the first element of ϵ_t is the structural demand shock and the first column of Ω_ϵ is given by the 2×1 vector q. The object of interest, θ, is the inflation response. Using the notation

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

one can express the inflation response to a demand shock and the sign restrictions on inflation and output respectively as

$$\theta = q_1\phi_1 \geq 0, \quad q_1\phi_2 + q_2\phi_3 \geq 0. \tag{1}$$

We shall refer to θ as structural parameter and ϕ as reduced form parameter. The latter is consistently estimable from the data. Since $\phi_1 \geq 0$, we deduce that $q_1 = \cos(\varphi) \geq 0$ and $\varphi \in [-\pi/2, \pi/2]$. The second inequality in (1) implies

$$q_2 \geq -\frac{\phi_2}{\phi_3} q_1. \tag{2}$$

For $\phi_2 \leq 0$, this inequality is always satisfied. Using the unit length restriction on q, it can also be verified that for $\phi_2 < 0$, the inequality (2) is satisfied whenever $q_1^2 \leq \phi_3^2/(\phi_2^2 + \phi_3^2)$. Thus, conditional on ϕ, the structural parameter $\theta = q_1\phi_1$ has to lie in the following set:

$$\Theta(\phi) = \left[0,\ \phi_1 \max\left\{\mathcal{I}\{\phi_2 \geq 0\},\ \sqrt{\frac{\phi_3^2}{\phi_2^2 + \phi_3^2}}\right\}\right]. \tag{3}$$

$\Theta(\phi)$ is called the identified set. Since this set is not a singleton, θ is only partially identified.

In applied work with sign-restricted structural VARs, researchers have used Bayesian inference to obtain error bands for impulse response functions. Bayesian inference in this setting amounts to specifying a joint prior distribution on (ϕ, q), which can be factorized as $p(\phi, q) = p(\phi)p(q|\phi)$. Researchers typically use a prior that is uniform with respect to φ, which implies that $q(\varphi)$ is uniformly distributed on the hypersphere. This prior is then truncated to ensure that the sign restrictions are satisfied. In the context of our example, we obtain

$$\varphi|\phi \sim \mathcal{U}\left[-\text{sgn}(\phi_2)\arccos\left(\sqrt{\frac{\phi_3^2}{\phi_2^2 + \phi_3^2}}\right),\ \frac{\pi}{2}\right].$$

For instance, if $\phi_2 \leq 0$, then the change of variables $\theta = \phi_1 \cos\varphi$ implies that

$$p(\theta|\phi) \propto \frac{\mathcal{I}\{\theta \in \Theta(\phi)\}}{\sqrt{1 - (\theta/\phi_1)^2}}. \tag{4}$$

Thus, due to the change of variables, θ is not uniformly distributed on the identified set. Since θ conditional on ϕ does not enter the likelihood function, the posterior distribution of θ can be expressed as the mixture

$$p(\theta|Y) = \int p(\theta|\phi)p(\phi|Y)d\phi, \tag{5}$$

where $p(\phi|Y)$ is the posterior of the reduced form parameter. As the sample size increases, the posterior distribution of ϕ concentrates around the maximum likelihood estimate $\hat{\phi}$ of the reduced form parameter. Moon and Schorfheide (2009) show that posterior credible sets for θ can be approximated by prior credible sets obtained from $p(\theta|\hat{\phi})$, which lie inside of the (estimated) identified set $\Theta(\hat{\phi})$.

The objective of this paper is to construct frequentist confidence sets for θ. The basic idea is the following. Define the nonnegative function

$$Q(\theta;\phi,W) = \min_{\varphi\in[-\pi/2,\pi/2],\,\mu\geq 0} \left\| \begin{bmatrix} (\cos\varphi)\phi_1 - \theta \\ (\cos\varphi)\phi_2 + (\sin\varphi)\phi_3 - \mu \end{bmatrix} \right\|_W^2, \tag{6}$$

where W is a positive-definite weight matrix. Based on (1) it is straightforward to verify that $\theta \in \Theta(\phi)$ if and only if $Q(\theta;\phi,W) = 0$. We now replace ϕ by the maximum likelihood estimator $\hat{\phi}$ and W by a data-based weight matrix $\hat{W}$. A confidence set can then be obtained as a level set associated with the sample analog of $Q(\theta;\phi,W)$:

$$CS^\theta = \left\{ \theta \,\middle|\, \theta \geq 0 \text{ and } Q(\theta;\hat{\phi},\hat{W}) \leq c \right\}, \tag{7}$$

where c is a critical value that ensures that the confidence set has the desired coverage probability, at least asymptotically, for every $\theta \in \Theta(\phi)$. Notice that for any critical value $c > 0$, the confidence set has the property that $\Theta(\hat{\phi}) \subset CS^\theta$. Thus, as explained in detail in Moon and Schorfheide (2009), the frequentist confidence set will be asymptotically larger than the Bayesian credible set. The computation of the confidence set essentially amounts to checking the inequality $Q(\theta;\hat{\phi},\hat{W}) \leq c$ for values of θ on a suitably chosen grid.

The use of a point-wise testing procedure to construct confidence sets dates back to work by Anderson and Rubin (1949) and is widely employed to implement identification-robust inference. It has been used in the weak-instrument literature, e.g., Dufour (1997) and Staiger and Stock (1997), and starting with CHT also in the literature on set-identified econometric models. While (6) resembles one of the popular objective functions used in the literature on moment inequality models, two important differences exist. First, the data enter the objective function only through the reduced form parameter estimator $\hat{\phi}$. In this regard, $Q(\theta;\hat{\phi},\hat{W})$ is similar to an objective function for a minimum-distance estimator, and the starting point of our analysis will be an assumption about the limit distribution of $\hat{\phi}$. Second, an important difference between $Q(\theta;\hat{\phi},\hat{W})$ and the objective functions studied in the moment-inequality model literature, e.g., Rosen (2008) and Andrews and Guggenberger (2009), is the presence of the nuisance parameter $q(\varphi)$ in (6). We shall consider two approaches to construct confidence sets: a profile objective function in which $q(\varphi)$ is concentrated out and a projection of a joint confidence set for θ and q on the domain of θ. We use insights from Rosen (2008) to construct somewhat conservative critical values for a nuisance parameter-free bound of the profile objective function. In the construction of joint (θ, q) confidence sets, we apply Andrews and Soares's (2010a) moment selection procedure. Using techniques developed in Andrews and Guggenberger (2009) and Andrews and Soares (2010a), we prove that our confidence sets are asymptotically valid in a uniform sense. The moment selection procedure yields less conservative error bands but is computationally more involved. Our paper provides empirical researchers with a menu of choices in regard to computational burden and conservativeness.

3 General Setup and Notation

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

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad \mathbb{E}[u_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[u_t u_t'|\mathcal{F}_{t-1}] = \Sigma_u, \tag{8}$$

where the information set $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ is composed of the lags of y_t's. Deterministic trend terms are omitted because they are irrelevant for the subsequent discussion. (8) is a reduced form representation of the VAR because the u_t's are simply one-step-ahead forecast errors and do not have a specific economic interpretation. As in Section 2, it is assumed that the one-step-ahead forecast errors are functions of a vector of fundamental innovations ϵ_t, for instance, composed of innovations in aggregate technology, preferences, or monetary policy:

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega_\epsilon \epsilon_t, \quad \mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t'|\mathcal{F}_{t-1}] = I, \tag{9}$$

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 (8) is invertible, one can express y_t as the following infinite-order vector moving average (VMA) process:

$$y_t = \sum_{h=0}^{\infty} C_h u_{t-h} = \sum_{j=0}^{\infty} C_h \Sigma_{tr} v_{t-h}, \tag{10}$$

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

$$R_h^v = I\!E[y_{t+h}|v_t = I, \mathcal{F}_t] - I\!E[y_{t+h}|\mathcal{F}_t] = \frac{\partial y_{t+h}}{\partial v_t'} = C_h \Sigma_{tr}, \quad h = 0, \ldots, H-1. \tag{11}$$

The goal is to construct a confidence set for a $\tilde{k} \times 1$ vector θ of impulse responses to a structural shock. Since such confidence sets are often presented as pointwise error bands, the case of $\tilde{k} = 1$ is of particular importance. Partial identification is achieved by restricting the signs of the responses of a subset of the endogenous variables at particular horizons. Let $\mathcal{H}$ denote the set that collects the impulse response horizons $h = 0, \ldots, H - 1$ and define the $nH \times n$ matrix $R^v_{\mathcal{H}}$ that stacks the matrices R^v_h for $h \in \mathcal{H}$. The responses to the orthogonolized shocks v_t can be converted into responses to structural shocks ϵ_t as follows:

$$R^{\epsilon}_{\mathcal{H}} = R^v_{\mathcal{H}} \Omega_{\epsilon}. \tag{12}$$

Rather than examining responses to the full vector of structural shocks, ϵ_t, we will focus on responses to one particular shock. Without loss of generality, we assume that the shock of interest is $\epsilon_{1,t}$ and will denote the first column of Ω_ϵ by the $n \times 1$ unit length vector $q \in \mathbb{Q}$.

Let $\tilde{S}_\theta$ be a $\tilde{k} \times nH$ matrix that selects and potentially transforms the structural impulse responses $R^v_{\mathcal{H}} q$ into the object of interest, θ:

$$\theta = \tilde{S}_\theta R^v_{\mathcal{H}} q. \tag{13}$$

Likewise, $\tilde{S}^*_R$ is an $\tilde{r} \times nH$ matrix that extracts the sign-restricted impulse responses such that the full set of sign restrictions can be expressed as

$$\tilde{S}^*_R R^v_{\mathcal{H}} q \geq 0. \tag{14}$$

Depending on whether a response is restricted to be nonnegative or nonpositive, the corresponding entries of $\tilde{S}^*_R$ are either 1 or -1. The sign restrictions summarized with $\tilde{S}^*_R$ either affect responses that are contained in θ, or they affect responses not contained in θ. We partition $\tilde{S}^*_R$ into an $\tilde{r}_1 \times \tilde{k}$ matrix M_θ that selects the elements of θ that are sign restricted and an $\tilde{r}_2 \times nH$ matrix $\tilde{S}_R$ that selects the sign-restricted responses not contained in θ:

$$M_\theta \theta = M_\theta \tilde{S}_\theta R^v_{\mathcal{H}} q \geq 0, \quad \tilde{S}_R R^v_{\mathcal{H}} q \geq 0. \tag{15}$$

Finally, the $m \times 1$ vector ϕ is defined as

$$\phi = S'_\phi vec((R^\nu_{\mathcal{H}})'), \tag{16}$$

where S'_ϕ is an $m \times n^2 H$ selection matrix that deletes zero responses in $R^\nu_{\mathcal{H}}$ arising from the lower triangular structure of $R^\nu_0 = \Sigma_{tr}$. Transpose and vectorize (13), (14) as well as the second set of sign restrictions in (15). Since S'_ϕ in (16) eliminates zeros from $R^\nu_{\mathcal{H}}$, one can write $vec((R^\nu_{\mathcal{H}})') = S_\phi \phi$. In turn, we can define the functions $\tilde{S}_\theta(q)$, $\tilde{S}^*_R(q)$, $\tilde{S}_R(q)$, and $\tilde{S}(q) = [\tilde{S}_\theta(q)', \tilde{S}_R(q)']'$ such that

$$\tilde{S}_\theta(q) = (\tilde{S}_\theta \otimes q')S_\phi, \quad \tilde{S}^*_R(q) = (\tilde{S}^*_R \otimes q')S_\phi, \quad \tilde{S}_R(q) = (\tilde{S}_R \otimes q')S_\phi. \tag{17}$$

We will use $k(q)$, $r(q)$, $r_2(q)$, and $l(q)$ to denote the row ranks of $\tilde{S}_\theta(q)$, $\tilde{S}^*_R(q)$, $\tilde{S}_R(q)$, and $\tilde{S}(q)$, respectively. The following assumption states that we are focusing on the case of set-identified impulse responses in this paper.

Assumption 1 *Given a set of orthogonolized responses ϕ, there exist q_1, q_2 such that $q_1 \neq q_2$, $q_1 \neq -q_2$ and $\tilde{S}^*_R(q_i)\phi \geq 0$, $i = 1, 2$ (set identification).*

Low-level conditions for identification of structural VARs can be found in Rubio-Ramirez, Waggoner, and Zha (2010). The identified set $\Theta(\phi)$ is defined as the set of impulse responses θ that are consistent with a particular ϕ. A general characterization of the identified set can be obtained as follows. Define the nonnegative function $Q(\theta; \phi, W)$

$$Q(\theta; \phi, W) = \min_{\|q\|=1,\ \mu \geq 0} \left\| \tilde{S}(q)\phi - \begin{pmatrix} \theta \\ \mu \end{pmatrix} \right\|^2_W, \tag{18}$$

where 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

$$\theta \in \Theta(\phi) \quad \text{if and only if} \quad Q(\theta; \phi, W) = 0 \quad \text{and} \quad M_\theta \theta \geq 0. \tag{19}$$

The following lemma states that for $\tilde{k} = 1$ the identified set $\Theta(\phi)$ is a bounded interval.

Lemma 1 *Suppose Assumption 1 is satisfied and $\tilde{k} = 1$. Then $\Theta(\phi)$ is convex and bounded.*

Lemma 1 is comforting from a practitioner's perspective. 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 lemma guarantees that scalar responses in this model can be characterized by two numbers: the lower bound and the upper bound of the identified interval.

4 Frequentist Inference

Inference for θ is conducted in two steps. First, an estimator $\hat{\phi}$ of a vector of reduced form parameters is constructed. Second, we conduct inference on θ conditional on the estimator $\hat{\phi}$ using a sample analog of the objective function $Q(\theta; \phi, W)$ in (18). Since empirical researchers typically depict error bands for impulse response functions that delimit point-wise credible or confidence intervals, in most applications θ is a scalar, that is $k = 1$, representing the response of variable $i \in \{1, \dots, n\}$ to a shock $j \in \{1, \dots, n\}$ at horizon h.

Rather than placing low-level restrictions on the VAR coefficient matrices Φ and Σ, as well as the distribution of the reduced-form innovations u_t, and deriving the distribution of $\hat{\phi}$, we directly assume that $\hat{\phi}$ has a Gaussian limit distribution. It is noteworthy that this assumption requires that all roots of the characteristic polynomial associated with the difference equation (8) lie outside of the unit circle. Hence, we are ruling out the presence of unit roots and are implicitly assuming that y_t is trend stationary.

Assumption 2 *(i) There exists an estimator $\hat{\phi}$ of the $m \times 1$ vector ϕ such that $\hat{\phi} \xrightarrow{p} \phi$ and $\sqrt{T}(\hat{\phi} - \phi) \Longrightarrow \mathcal{N}(0, \Lambda(\phi))$ uniformly for $\phi \in \mathcal{P}$. (ii) The matrix $\Lambda(\phi)$ is positive definite and there exists a full-rank matrix Λ_{min} such that $\Lambda(\phi) \geq \Lambda_{min}$ for all $\phi \in \mathcal{P}$. (iii) There exists an estimator $\hat{\Lambda} \xrightarrow{p} \Lambda(\phi)$ uniformly for $\phi \in \mathcal{P}$.*

For notational simplicity, the dependence of Λ on ϕ is suppressed, unless this dependence plays a crucial role. The assumption that Λ is full rank does not impose serious constraints on the applicability of our analysis. Recall that we previously eliminated the $n(n-1)/2$ zero elements of the lower triangular matrix R_0^v by appropriately defining ϕ. Furthermore, in practice it is possible to delete all elements of ϕ that are associated with columns of zeros in the matrix $\tilde{S}(q)$. These are elements that do not enter the construction of the structural responses that appear in θ or that are sign restricted.[2]

[2]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 $h = 9$. Following the definitions in Section 3, the dimension of $m = 10 \cdot 4^2 - 6 = 154$, whereas the number of reduced form VAR coefficients is $4 \cdot 16 + 10 = 74$, which suggests that – abstracting from the effect of nonlinearities – the covariance matrix Λ is rank deficient. However, in order to construct the sign-restricted responses as well as θ, one can easily reduce the dimension of ϕ to $4 \cdot 3 \cdot 4 + 4 = 52$ and obtain a nonsingular covariance matrix Λ.

Three methods of constructing confidence sets for θ are considered subsequently. All of the methods have to deal with the unit-length nuisance parameter q. The first method (Section 4.2) concentrates out q from the objective function that is used to construct the confidence interval. The other two methods (Section 4.3) are based on a projection of a joint confidence interval for θ and q. Section 4.4 provides computational details and step-by-step guidance for the implementation. Finally, we discuss some extensions and limitations in Section 4.5. However, before we can proceed, we need to state an assumption about the matrix $\tilde{S}(q) = [\tilde{S}_\theta(q)', \tilde{S}_R(q)']'$.

4.1 An Assumption About $\tilde{S}(q)$

While $\tilde{S}_R^*(q)$ was useful to state our assumption about set identification (Assumption 1), the most important object for the subsequent development of inference methods for θ is the matrix-valued function $\tilde{S}(q)$. Although this function is continuous in q, its row rank tends to be discontinuous. Since we will use a weight function W that is based on the inverse covariance matrix of $\sqrt{T}\tilde{S}(q)(\hat{\phi} - \phi)$ to construct the objective function $Q(\theta; \hat{\phi}, W)$, we have to invert the matrix $\tilde{S}(q)\Lambda\tilde{S}'(q)$ and need to pay special attention to the potential rank reduction. To fix ideas, recall the example of Section 2, which leads to

$$\tilde{S}(q) = (I \otimes q')S_\phi = \begin{bmatrix} q_1 & q_2 & 0 & 0 \\ 0 & 0 & q_1 & q_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_1 & q_2 \end{bmatrix}$$

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 nonzero entry in the first row of $\tilde{S}(q)$.

We will now state an assumption to guarantee that row rank reductions of $\tilde{S}(q)$ only arise from rows of zeros. The assumption states that $\tilde{S}(q)$ can be obtained from a matrix $\bar{S}(q)$ through a series of transformations. The starting point is the matrix

$$\bar{S}(q) = (I_{nH} \otimes q')S_\phi. \tag{20}$$

Equation (16) and the definition of S_ϕ imply that $\bar{S}(q)\phi$ generates the structural impulse responses of the n variables for horizons $h = 0, \ldots, H-1$. Now consider the following five transformations of $\bar{S}(q)$:

1. $M^{S,1}$: Multiplication of rows by (-1) to change the sign of the impulse response.

2. $M^{S,2}$: Quasi-lower triangular transformations of the form

$$\begin{bmatrix} M^{S,2}_{\Sigma\Sigma} & 0 \\ M^{S,2}_{B\Sigma} & M^{S,2}_{BB} \end{bmatrix},$$

   where $M_{\Sigma\Sigma}$ is an $n \times n$ full rank lower triangular matrix and the submatrix $M^{S,2}_{BB}$ has full row rank.

3. $M^{S,3}$: Re-ordering of rows.

4. $M^{S,4}$: Deletion of rows.

5. $M^{S,5}$: Deletion of columns of zeros.

The first transformation switches signs of impulse responses, e.g., to turn a nonnegativity constraint into a nonpositivity constraint. The second transformation can be used to generate cumulative impulse responses, e.g., convert responses of inflation rates into responses of the log-level of inflation, or to transform impulse responses, e.g., turn responses of log nominal output and log prices into responses of real output. The second transformation can also be used to constrain (backward) differences of impulse responses in order to, say, impose monotonicity restrictions. The third transformation reorders the impulse responses, e.g., to ensure that those responses that enter the vector θ appear first. The fourth transformation can be used to eliminate those responses that are neither the object of interest, i.e., contained in θ, nor sign restricted. Finally, the last transformation can be used to reduce the dimension of ϕ by eliminating unused reduced form impulse responses as discussed above.[3] Notice that the identity matrix is a special case of all five transformations. In turn, we can state our assumption about $\tilde{S}(q)$.

Assumption 3 *The matrix* $\tilde{S}(q) = [\tilde{S}_\theta(q)', \tilde{S}_R(q)']'$ *can be expressed as*

$$\tilde{S}(q) = \left(\prod_{k=1}^{4} M^{S,k} \right) \bar{S}(q) M^{S,5}.$$

[3]In order to economize on the notation, we do not distinguish between the original vector ϕ introduced in (16) and the vector that obtains by removing the elements that correspond to the columns of $\bar{S}(q)$ that get eliminated by $M^{S,5}$.

By construction, the rows of $\bar{S}(q)$ are orthogonal. Since q lies on the unit-hypersphere, we can deduce that a rank reduction of $\bar{S}(q)$ can only occur if one or more of the first $n-1$ rows of $\bar{S}(q)$ are zero. These rows contain elements of the form q_1, $[q_1, q_2]'$, ..., $[q_1, \ldots, q_{n-1}]'$. One can verify that the transformations $M^{S,k}$, $k = 1, \ldots, 5$ preserve the property that rank reductions are due to rows of zeros, which lead to the following lemma.

Lemma 2 *Suppose Assumption 3 is satisfied. Then for a particular value of q, a row rank reduction of $\tilde{S}(q)$ arises only through one or more rows of zeros.*

In order to eliminate the rows of zeros in $\tilde{S}(q)$, we introduce the selection matrices $V(q)$, $V_\theta(q)$, and $V_R(q)$ to define the matrices

$$S(q) = V(q)\tilde{S}(q),\ S_\theta(q) = V_\theta(q)\tilde{S}_\theta(q),\ S_R(q) = V_R(q)\tilde{S}_R(q). \tag{21}$$

The row dimensions of the three matrices are $l(q)$, $k(q)$, and $r_2(q)$, respectively. By construction, the three matrices have full row rank.

4.2 Profile Objective Function with Fixed Critical Value

We replace ϕ and W in the objective function $Q(\theta; \phi, W)$, defined in (18) by $\hat{\phi}$ and a weight matrix $\hat{W}(q)$ that is allowed to depend on the sample and on q. Thus,

$$Q\big(\theta; \hat{\phi}, \hat{W}(\cdot)\big) = \min_{\|q\|=1,\, \mu\geq 0} \left\| \tilde{S}(q)\hat{\phi} - \begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{\hat{W}(q)}. \tag{22}$$

The function $Q\big(\theta, \hat{\phi}, \hat{W}(\cdot)\big)$ is a profile objective function in the sense that the nuisance parameter q has been concentrated out. As a weight matrix, we use the inverse of the asymptotic covariance matrix of $\sqrt{T}\tilde{S}(q)(\hat{\phi} - \phi)$. In order to account for the potentially defined row rank of $\tilde{S}(q)$, we let

$$\hat{W}^*(q) = TV'(q)\Big(V(q)\tilde{S}(q)\hat{\Lambda}\tilde{S}'(q)V'(q)\Big)^{-1}V(q), \tag{23}$$

where $\hat{\Lambda}$ is a consistent estimator of Λ (see Assumption 2). Using the definition $\hat{\Sigma}(q) = S(q)\hat{\Lambda}S'(q)$, the objective function can be rewritten as

$$\begin{aligned} Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) &= \min_{\|q\|=1,\, \mu\geq 0} \Bigg(T\left\| V(q)\tilde{S}(q)\hat{\phi} - V(q)\begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{\hat{\Sigma}^{-1}(q)} \\ &\quad + T\sum_{j=1}^{\tilde{k}} \mathcal{I}\big\{\tilde{S}_{j,\theta}(q) = 0 \text{ and } \theta_j \neq 0\big\}\Bigg). \end{aligned} \tag{24}$$

The penalty term ensures that the objective function takes on a large value if θ elements that correspond to rows of zeros in $\tilde{S}_\theta(q)$ are different from zero.[4] 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.

The confidence set for θ is constructed as a level set of the profile objective function:

$$CS^\theta_{(1)} = \left\{ \theta \;\middle|\; M_\theta \theta \geq 0 \text{ and } Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq c^{(1)}_{(\tilde{k},\tilde{r}_2)} \right\}. \tag{25}$$

The critical value, denoted by $c^{(1)}_{(\tilde{k},\tilde{r}_2)}$, depends on the number of elements in θ, $\tilde{k}$, and the number of sign restrictions, $\tilde{r}_2$. It is given by

$$c^{(1)}_{(\tilde{k},\tilde{r}_2)} = 1-\tau \text{ quantile of } \begin{cases} \chi^2_{\tilde{k}+\tilde{r}_2-1} + Z^2\mathcal{I}\{Z \leq 0\} & \text{if } \tilde{r}_2 \geq 1 \\ \chi^2_{\tilde{k}} & \text{otherwise} \end{cases}, \tag{26}$$

where $Z \sim N(0,1)$ and is independent of the chi-square random variable. We follow the convention that $\chi^2_0 = 0$. Notice that the critical value $c^{(1)}_{(\tilde{k},\tilde{r}_2)}$ used in the construction of the confidence set does not depend on θ. The following theorem states that $CS^\theta_{(1)}$ is an asymptotically valid confidence set.

Theorem 1 *Suppose that Assumption 1 is satisfied for $\phi \in \mathcal{P}$ and Assumptions 2 and 3 are satisfied. Then the confidence set $CS^\theta_{(1)}$ defined in (25) is an asymptotically valid confidence set for θ:*

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

A formal proof of Theorem 1 is provided in the Appendix. The proof is based on an upper bound of the concentrated objective function with a nuisance parameter-free limit distribution. The gist of the argument is the following. Consider a particular $\theta \in \Theta(\phi)$. By definition of the identified set there exists a $\tilde{q}$ and $\tilde{\mu}$ such that $\tilde{S}_\theta(\tilde{q})\phi = \theta$ and $\tilde{S}_R(\tilde{q})\phi = \tilde{\mu}$. Using the definition in (21), replace $V(q)\tilde{S}(q)$ by $S(q)$ to write the objective function as:

$$Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq \min_{\mu \geq 0} \left\| S(\tilde{q})\sqrt{T}(\hat{\phi}-\phi) - V(q)\begin{pmatrix} 0 \\ \sqrt{T}(\mu-\tilde{\mu}) \end{pmatrix} \right\|^2_{\hat{\Sigma}^{-1}(\tilde{q})}. \tag{27}$$

Notice that for $\theta \in \Theta(\phi)$ the penalty term that appears in (24) has to be zero. Now let $\nu = \sqrt{T}(\mu-\tilde{\mu})$ and $M_\nu = [0_{k\times r_2}, I_{r_2}]'$. Under this reparameterization, the inequality $\mu \geq 0$ becomes $\nu \geq -\sqrt{T}\tilde{\mu}$. If

[4] In a nutshell, we are rewriting an objective function of the form $T\frac{1}{q^2}(q\hat{\phi}-\theta)^2 = T(\hat{\phi}-\theta/q)^2$ as $T\big(\mathcal{I}\{q \neq 0\}(\hat{\phi}-\theta/q)^2 + \mathcal{I}\{q=0 \text{ and } \theta \neq 0\}\big)$ to account for the case $q = 0$.

we replace $\tilde{\mu}$ on the right-hand side of the reparameterized inequality by 0, we obtain the following bound

$$Q\big(\theta, \hat{\phi}, \hat{W}^*(\cdot)\big) \leq \min_{\nu \geq 0} \left\| S(\tilde{q})\sqrt{T}(\hat{\phi} - \phi) - V(\tilde{q}) M_\nu \nu \right\|^2_{\hat{\Sigma}^{-1}(\tilde{q})} = \bar{Q}\big(\tilde{q}; \hat{\phi}, \hat{W}^*(\cdot)\big). \tag{28}$$

It turns out that the weight matrix $\hat{W}^*(\cdot)$, which depends on the the estimated covariance matrix $\hat{\Lambda}$, can be replaced by $W^*(\cdot)$, which is constructed from the population covariance matrix Λ. We will now sketch the analysis of $\bar{Q}\big(\tilde{q}; \hat{\phi}, W^*(\cdot)\big)$ for the cases $l(\tilde{q}) = \tilde{k} + \tilde{r}_2$ and $l(\tilde{q}) < \tilde{k} + \tilde{r}_2$, abstracting from uniformity issues.

Suppose that $l(\tilde{q}) = \tilde{k} + \tilde{r}_2$ and $\tilde{k} \geq 1$, $\tilde{r}_2 \geq 1$. Thus, $\tilde{S}(\tilde{q})$ has full row rank and $V(\tilde{q}) = I$ and $S(\tilde{q}) = \tilde{S}(\tilde{q})$. Partition $S'(\tilde{q}) = [S_1'(\tilde{q}), S_2'(\tilde{q})]'$, where $S_2(\tilde{q})$ is the last row of $S(\tilde{q})$. Denote the conforming partitions of $\Sigma(\tilde{q})$ by $\Sigma_{ij}(\tilde{q}) = S_i(\tilde{q})\Lambda S_j'(\tilde{q})$. Moreover, factorize $\Lambda = LL'$ and let ν_2 be the last element of the vector $M_\nu \nu$. Then, omitting the $\tilde{q}$ argument, we obtain

$$\begin{aligned}
\bar{Q}(\tilde{q}; \hat{\phi}, W^*) \quad &\leq \quad \left\| S_1 \sqrt{T}(\hat{\phi} - \phi) \right\|^2_{\Sigma_{11}^{-1}} \\
&\quad + \min_{\nu_2 \geq 0} \left\| (S_2 - \Sigma_{21}\Sigma_{11}^{-1} S_1)\sqrt{T}(\hat{\phi} - \phi) - \nu_2 \right\|^2_{(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}} \\
&= \quad \hat{\zeta}' P_{A_1} \hat{\zeta} + \min_{\nu_2 \geq 0} \left\| A_2'(I - P_{A_1})\hat{\zeta} - \nu_2 \right\|^2_{(A_2'(I - P_{A_1})A_2)^{-1}} \\
&\Longrightarrow \quad \chi^2_{\tilde{k} + \tilde{r}_2 - 1} + \mathcal{I}\{Z \leq 0\} Z^2, \quad Z \sim N(0,1).
\end{aligned} \tag{29}$$

The inequality is obtained by setting all but the very last element of the vector $M_\nu \nu$ equal to zero. The second expression on the right-hand side is obtained by defining $\hat{\zeta} = L^{-1}\sqrt{T}(\hat{\phi} - \phi)$, $A_i = L' S_i' D_i^{-1/2}$, and $P_{A_i} = A_i(A_i' A_i)^{-1} A_i'$. Here $D_i^{1/2}$ is the diagonal matrix of standard deviations associated with the covariance matrix Σ_{ii}. Since $\hat{\zeta}$ converges in distribution to an $m \times 1$ vector of standard normals and P_{A_1} is a projection onto a $\tilde{k} + \tilde{r}_2 - 1$-dimensional subspace, we obtain the convergence of $\hat{\zeta}' P_{A_1} \hat{\zeta}$ to a $\chi^2_{\tilde{k} + \tilde{r}_2 - 1}$. The solution to the minimization problem is

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

which generates the term $\mathcal{I}\{Z \leq 0\} Z^2$ in the limit distribution. Since $P_{A_1}\hat{\zeta}$ and $A_2'(I - P_{A_1})\hat{\zeta}$ are asympotically uncorrelated, Z is independent of the χ^2 term in the characterization of the asymptotic distribution.

Alternatively, if $l(\tilde{q}) < \tilde{k} + \tilde{r}_2$, there is no need to partition $S(\tilde{q})$. By setting $\nu = 0$, one obtains:

$$\bar{Q}(\tilde{q}; \hat{\phi}, W^*) \leq \left\| S\sqrt{T}(\hat{\phi} - \phi) \right\|^2_{\Sigma^{-1}} \Longrightarrow \chi^2_{l(\tilde{q}) - 1}. \tag{30}$$

Since in this case $l(\tilde{q}) \leq \tilde{k} + \tilde{r}_2 - 1$, the critical value associated with (29) remains valid, though it is conservative. The limit distribution in (29) arises commonly in multivariate generalizations of one-sided hypothesis problems, e.g., Perlman (1969), and its quantiles are used by Rosen (2008) to construct contour confidence sets for moment inequality models. (26) provides a convenient characterization of the critical value associated with this limit distribution.

4.3 Projection Approach with Moment Selection

As an alternative to the fixed-critical value approach based on the profile objective function, we consider a projection-based confidence set obtained from an objective function that does not concentrate out q. Let

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

We maintain the choice of weight matrix $\hat{W}^*$ in (23) and define a joint confidence set for θ and q as the generalized level set:

$$CS^{\theta,q}_{(2)} = \left\{ \theta,\, q \;\middle|\; \|q\| = 1,\; M_\theta \theta \geq 0, \text{ and } G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq c_{(2)}(q, \theta) \right\}. \tag{32}$$

The critical value $c_{(2)}(\cdot)$ is potentially a function of both q and θ. However, it turns out that conditional on q the distribution of $G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big)$ does not depend on θ for $\theta \in \Theta(\phi)$ because θ does not enter the inequality conditions and therefore does not affect the slackness μ. Also, when $\theta \in \Theta(\phi)$, the penalty term that appears in (31) has to be zero. Subsequently, the θ-argument is dropped from the critical value function and the projection of the joint confidence set $CS^{\theta,q}_{(2)}$ onto Θ takes the form

$$CS^{\theta}_{(2)} = \left\{ \theta \;\middle|\; M_\theta \theta \geq 0, \text{ and } \min_{\|q\|=1} \Big(G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big) - c_{(2)}(q) \Big) \leq 0 \right\}. \tag{33}$$

Since $Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) = \min_{\|q\|=1} G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big)$, we obtain the following lemma.

Lemma 3 *Consider the two confidence sets defined in (25) and (33). Suppose that $c_{(2)}(q) \leq c_{(1)}$ for all q, then $CS^{\theta}_{(2)} \subseteq CS^{\theta}_{(1)}$.*

Lemma 3 implies that the projection-based confidence set is potentially smaller than the confidence set constructed from the profile objective function. However, the disadvantage of the projection-based approach is that the calculation of q-dependent critical values might require cumbersome simulations. We will discuss this trade-off in the context of the empirical application. If one defines $\vartheta = [\theta', q]$, then 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., CHT, Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010a). The three main differences in the VAR application are that q is a nuisance parameter, that the objective function $G(\cdot)$ corresponds to a minimum-distance rather than a GMM problem, and that the q-dependent weight matrix in the objective function $G(\cdot)$ could be singular for some values of $q's$. In the remainder of this subsection it is discussed how the moment selection approach of Andrews and Soares (2010a) can be applied to obtain critical value functions $c_{(2)}(q)$.

For each $\theta \in \Theta(\phi)$, we define the set $\mathbb{Q}(\theta, \phi) = \big\{q \mid \|q\| = 1,\, \tilde{S}_\theta(q)\phi = \theta,\, \tilde{S}_R(q)\phi \geq 0\big\}$ as well as the function $\tilde{\mu}(q, \phi) = \tilde{S}_R(q)\phi$. Let $\hat{\Sigma}(q) = S(q)\hat{\Lambda}S(q)'$ and decompose the covariance matrix into

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

where $\hat{\Omega}$ is a correlation matrix and $\hat{D}^{1/2}$ is a diagonal matrix of standard deviations. Let $M_\mu = [0_{\tilde{k}\times\tilde{r}_2}, I_{\tilde{r}_2}]'$ and consider a $\theta \in \Theta(\phi)$ and a $q \in \mathbb{Q}(\theta, \phi)$. Then, using the fact that the penalty term is zero for $\theta \in \Theta(\phi)$ and $q \in \mathbb{Q}(\theta, \phi)$, we can rewrite (31):

$$\begin{aligned} & G\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big) \\ &= \min_{\tilde{\mu}\geq 0} \left\| \hat{D}^{-1/2}S(q)\sqrt{T}(\hat{\phi} - \phi) - \hat{D}^{-1/2}V(q)M_\mu\sqrt{T}(\tilde{\mu} - \tilde{\mu}(q, \phi)) \right\|^2_{\hat{\Omega}^{-1}(q)}. \end{aligned} \tag{34}$$

Recall that $V(q)$ eliminates elements of the vector $M_\mu\tilde{\mu}$, corresponding to rows of zeros in $\tilde{S}(q)$. Moreover, the matrix $\hat{D}^{-1/2}$ is diagonal. Thus, we define the $r_2(q) \times 1$ vectors μ and $\mu(q, \phi)$ as transformation of $\tilde{\mu}$ and $\tilde{\mu}(q, \phi)$ in which the elements corresponding to rows of zeros in $\tilde{S}_R(q)$ have been eliminated. Let $\hat{D}_R$ be the submatrix of $\hat{D}$ that is conformable with the $S_R(q)$ partition of $S(q)$ and define the $r_2(q) \times 1$ vector $\nu = \sqrt{T}\hat{D}_R^{-1/2}(\mu - \mu(q, \phi))$. Moreover, define the matrix M_ν by deleting unnecessary columns from the matrix $V(q)M_\mu$ to make it conformable with ν.[5] In turn, the objective function can be expressed as

$$G\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big) = \min_{\nu \geq -\sqrt{T}\hat{D}_R^{-1/2}\mu(q,\phi)} \left\| \hat{D}^{-1/2}S(q)\sqrt{T}(\hat{\phi} - \phi) - M_\nu\nu \right\|^2_{\hat{\Omega}^{-1}(q)}. \tag{35}$$

[5]When $S(q) = S_\theta(q)$, one can set $M_\nu = 0$.

The moment selection approach of Andrews and Soares (2010a) amounts to raising the lower bounds for ν_j to zero if the slackness ($\sqrt{T}\hat{D}_{jj,R}^{-1/2}\mu_j(q,\phi)$) is small and setting them to infinity if the slackness is large.

Define the standardized slackness in inequality moment condition $j = 1,\ldots,r_2(q)$ as

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

A moment condition is deemed nonbinding if $\hat{\xi}_{j,T}(q)$ exceeds the threshold κ_T, where κ_T is a diverging sequence, e.g., $\kappa_T = 1.96\ln(\ln T)$. Thus, estimates of the number of nonbinding and binding moment conditions are given by

$$\hat{r}_{22}(q) = \sum_{j=1}^{r_2(q)} \mathcal{I}\{\hat{\xi}_{j,T}(q) \geq \kappa_T\} \quad \text{and} \quad \hat{r}_{21}(q) = r_2(q) - \hat{r}_{22}(q). \tag{36}$$

Recall that $l(q)$ is the row rank of matrix $S(q)$ and $k(q) = l(q) - r_2(q)$. Now define the $r_2(q)$ vector $\hat{\varphi}_T(q)$ with elements

$$\hat{\varphi}_{j,T}(q) = \begin{cases} \infty & \text{if } \hat{\xi}_{j,T}(q) \geq \kappa_T \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

Using this notation, we obtain the following approximate upper bound for $G\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big)$:

$$\bar{G}\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big) = \min_{\nu \geq -\hat{\varphi}_T(q)} \left\| \hat{D}^{-1/2}S(q)\sqrt{T}(\hat{\phi} - \phi) - M_\nu \nu \right\|^2_{\hat{\Omega}^{-1}(q)}. \tag{38}$$

Thus, whenever $\hat{\varphi}_{j,T}(q) = 0$, the lower bound for ν_j is raised to zero, and whenever $\hat{\varphi}_{j,T}(q) = \infty$, the constraint on ν_j is eliminated. Let $\hat{A}(q) = \hat{L}'S'(q)\hat{D}^{-1/2}$ and define the submatrices $\hat{A}_b$, M_{ν_b}, ν_b, and $\hat{\Omega}_b$ by deleting the rows and columns corresponding to $\hat{\varphi}_{j,T} = \infty$ (nonbinding moment conditions), then the distribution of the upper bound can be approximated by

$$\bar{\mathcal{G}}\big(\theta, q; \hat{\Omega}_b\big) = \min_{\nu_b \geq 0} \left\| \hat{A}_b'(q)Z_m - M_{\nu_b}\nu_b \right\|^2_{(\hat{A}_b'(q)\hat{A}_b(q))^{-1}}, \quad Z_m \sim N(0, I_m). \tag{39}$$

This representation highlights that the moment selection approach reduces the degrees of freedom in the distribution of the bounding function.

Confidence sets for θ of the form (33) can now be constructed by using a fixed critical value or a simulated critical value:

$$\begin{aligned} \hat{c}_{(21)}(q) &= c^{(1)}_{(k(q),\hat{r}_{21}(q))} \\ \hat{c}_{(22)}(q) &= 1-\tau \text{ quantile of } \bar{\mathcal{G}}\big(\theta, q; \hat{\Omega}_b\big). \end{aligned} \tag{40}$$

The first critical value is obtained from an argument similar to the one in Section 4.2 and can be easily computed based on (26). Notice that either $k(q)$ or $\hat{r}_{21}(q)$ could be equal to zero. If both are zero, then $\bar{G}\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big) = 0$ and the critical value $c_{(0,0)}^{(1)} = 0$, which implies $(\theta, q) \in CS_{(2)}^{\theta,q}$ as required. The second critical value, $\hat{c}_{(22)}(q)$, is potentially less conservative than the first, but requires the simulation of a stochastic quadratic programming problem.

Theorem 2 *Suppose that Assumption 1 is satisfied for $\phi \in \mathcal{P}$ and Assumptions 2 and 3 are satisfied. Then the confidence set $CS_{(2)}^{\theta}$, defined in (33) with one of the two critical values in (40), is an asymptotically valid confidence set for θ:*

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

where $0 < \tau < 1/2$.

A formal proof of Theorem 2 is provided in the Appendix and closely follows the proof of Theorem 1 in Andrews and Soares (2010a). However, a number of modifications are required to account for the potential rank reduction of $\tilde{S}(q)$.

To construct a confidence set for θ, one could consider various alternatives by choosing different objective functions and different moment selection rules, and by employing different approximation methods for the critical values. We leave it as a future research topic to consider the alternative approaches and compare them with the methods in the paper. For more details on the potential alternative approaches, readers can refer to Andrews and Soares (2010a) and the references therein.

4.4 Implementation

The computation of frequentist error bands based on the confidence sets constructed in Sections 4.2 and 4.3 involves the discretization of the impulse response domain and the point-wise inversion of test statistics based on potentially simulated critical values. Since error bands in the VAR literature predominantly depict point-wise confidence sets, we focus on the computation of $\Theta(\hat{\phi})$ as well as the confidence sets $CS_{(1)}^{\theta}$ and $CS_{(2)}^{\theta}$ for $k = 1$. This computation has to be repeated for every response $\partial y_{i,t+h}/\partial \epsilon_{1,t}$ of interest. Here i potentially ranges from $i = 1, \ldots, n$ and $h = 0, 1, \ldots, h_{max}$. In order to compute the confidence intervals, we start from a preliminary interval and then expand or contract the boundaries of this preliminary interval until we have found the boundaries of $\Theta(\hat{\phi})$

or $CS^{\theta}_{(i)}$. According to Lemma 1 it is guaranteed that the set $\Theta(\hat{\phi})$ is a bounded interval. The following lemma states that the confidence sets $CS^{\theta}_{(i)}$ are also bounded.[6]

Lemma 4 *Suppose that Assumption 1 is satisfied and $\tilde{k} = 1$. Then $CS^{\theta}_{(1)}$ and $CS^{\theta}_{(2)}$ defined in (25) and (33) are bounded.*

The computation of the confidence sets involves minimizations with respect to q, where q is restricted to lie on the unit hypersphere. We start by randomly generating a set of q's as follows: let $Z_{(s)}$, $s = 1, \ldots, s_{max}$, be a sequence of $n \times 1$ vectors of standard normal random variables and define $q_{(s)} = Z_{(s)}/\|Z_{(s)}\|$. It is well known, e.g., James (1954), that q_s is uniformly distributed on the unit-hypersphere defined by $\|q\| = 1$. Now define the grid $\mathcal{Q} = \{q_{(1)}, \ldots, q_{(s_{max})}\}$.

Computing a Preliminary Interval $\mathcal{I}_\theta$ for $\Theta(\hat{\phi})$. A preliminary interval can be obtained as follows:

1. Compute the estimator $\hat{\phi}$.
2. For each $q_{(s)} \in \mathcal{Q}$ determine whether the responses $\theta_{(s)} = \tilde{S}_\theta(q_{(s)})\hat{\phi}$ and $\tilde{S}_\theta(q_{(s)})\hat{\phi}$ satisfy the sign restrictions.
3. Define the boundaries of $\mathcal{I}_\theta$ as the min and the max of the $\theta_{(s)}$ responses that do satisfy the sign restrictions.

Computing the Boundaries of $\Theta(\hat{\phi})$. By construction, $\mathcal{I}_\theta \subseteq \Theta(\hat{\phi})$. Thus, in order to find the boundaries of $\Theta(\hat{\phi})$, one can raise (lower) the upper (lower) bound of $\mathcal{I}_\theta$ in a step-wise fashion, where the step size δ_θ can be chosen as a fraction of the length of $\mathcal{I}_\theta$. Thus, assuming that $\theta_{j-1} \in \Theta(\hat{\phi})$, iteration j in the construction of the upper bound for $\Theta(\hat{\phi})$ takes the following form:

1. Let $\theta_{(j)} = \theta_{(j-1)} + \delta_\theta$.
2. Compute $Q(\theta_{(j)}; \hat{\phi}, W)$.
3. If $Q(\theta_{(j)}; \hat{\phi}, W) = 0$, then proceed to iteration $j+1$. If $Q(\theta_{(j)}; \hat{\phi}, W) > 0$, then terminate the iterations and set the upper bound of $\Theta(\hat{\phi})$ to $\theta_{(j-1)}$.

[6] Since the weight matrix $W^*(\cdot)$ used in the construction of the confidence sets is a function of q, the proof of Lemma 4 is not general enough to establish the convexity of the confidence interval.

The lower bound of $\mathcal{I}_\theta$ can be found in a similar manner. The objective function $Q(\theta_{(j)}; \hat{\phi}, W)$ is given in (18). While at this point W could be any positive-definite weight matrix, for the empirical analysis in Section 6 we set $W = \hat{W}(q)$ and replace the threshold of zero by $\epsilon = c^{(1)}_{\tilde{k},\tilde{r}_2}/10^6$. Recall from Section 4.3 that we can express the objective function $Q(\theta; \hat{\phi}, \hat{W}^*)$ as

$$Q(\theta; \hat{\phi}, \hat{W}^*) = \min_{\|q\|=1} \; G\big(\theta; q; \hat{\phi}, \hat{W}^*\big),$$

where

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

A parametric bootstrap procedure is used to estimate the covariance matrix $\hat{\Lambda}$, which enters the weight matrix. A standard quadratic programming procedure can be used to evaluate the function $G\big(\theta; q; \hat{\phi}, \hat{W}^*\big)$. The minimization of $G\big(\,\cdot\,\big)$ with respect to q is carried out in two steps. First, we conduct a grid search over $q \in \mathcal{Q}$. As soon as we find a value q_* such that $G\big(\theta; q_*; \hat{\phi}, \hat{W}^*\big) < \epsilon$, the grid search at iteration j can be terminated. In this case we reorder $\mathcal{Q}$ such that q_* appears first. If none of the $q_{(s)} \in \mathcal{Q}$ satisfies the condition $G\big(\theta; q; \hat{\phi}, \hat{W}^*\big) < \epsilon$, we use the minimizing $q_{(s)}$ as a starting value for a gradient-based minimization of the $G(\cdot)$ function. To conduct the gradient-based minimization, q is transformed into spherical coordinates.

Computing a Preliminary Interval $\mathcal{C}_\theta$ for $CS^\theta_{(i)}$. If the sampling variability of $\hat{\phi}$ is small compared with the size of the estimated interval $\Theta(\hat{\phi})$, then $\Theta(\hat{\phi})$ is a reasonable choice as a preliminary interval for the $CS^\theta_{(i)}$ confidence sets. If the sampling variability of $\hat{\phi}$ is relatively large compared with the length of $\Theta(\hat{\phi})$, then it might be preferable to start from a Bayesian credible set for θ. Posterior computations under a conjugate MNIW (see the definition in Section 1) prior distribution for the reduced form parameters and a prior distribution for q that is uniform (on the unit hypersphere that is truncated to ensure that the sign restrictions are satisfied) are fairly straightforward and described in detail in Uhlig (2005).

Computing the Boundaries of $CS^\theta_{(i)}$. The boundaries of $CS^\theta_{(i)}$ can be obtained by expanding or contracting the boundaries of the preliminary interval $\mathcal{C}_\theta$ in a step-wise manner. We have the following ordering of the sets:

$$\Theta(\hat{\phi}) \subseteq CS^\theta_{(2)}(\hat{c}_{22}) \subseteq CS^\theta_{(2)}(\hat{c}_{21}) \subseteq CS^\theta_{(1)},$$

where $\hat{c}_{22}$ is the simulated critical value and $\hat{c}_{21}$ is the more conservative fixed critical value described in Section 4.3. The computational strategy is similar to the one used for $\Theta(\hat{\phi})$. Two differences are noteworthy. First, we use grid search only over $\mathcal{Q}$ to carry out minimizations with respect to q. For all practical purposes, this grid search seemed to be sufficient in the empirical application. Second, the computational time for the three confidence intervals decreases drastically with conservativeness. In applications with large numbers of inequality constraints, the simulated critical value $\hat{c}_{22}$ might take a very long time to compute. While we did not attempt to measure CPU time carefully, it turned out that for the empirical analysis in Section 6.2, the computation of the $CS^{\theta}_{(1)}$ error bands took a few hours, whereas the computation of the $CS^{\theta}_{(2)}$ error bands with simulated critical values based on restrictions over horizons $h = 0, \ldots, 8$ took several days on a multi-processor computer. Thus, we strongly recommend to start the empirical analysis by computing $CS^{\theta}_{(1)}$ first. This interval can then subsequently be refined by switching to the projection-based approach.

4.5 Discussion

We will subsequently provide a brief discussion of extensions and limitations of the results obtained in Sections 4.2 and 4.3.

Profile versus Projection Approach. According to Lemma 3, the confidence set constructed based on the projection approach (in conjunction with the selection of potentially binding moments) weakly dominates the confidence set obtained with the profile approach. The main appeal of the profile approach is that it is faster to compute $CS^{\theta}_{(1)}$ than $CS^{\theta}_{(2)}$. While the evaluation of $\hat{r}_{21}$ is fairly straightforward, the calculation of the critical value $c_{(2)}(\tilde{q})$ is very time consuming.

Union of Identified Sets. A conceptually straightforward approach of constructing a valid confidence set for partially identified parameters is to take the union of identified sets $\Theta(\phi)$ over all values of ϕ in a $1-\tau$ confidence set CS^{ϕ}_{τ}. According to Assumption 2, one can obtain an asymptotically valid confidence set for the reduced form parameter as follows:

$$CS^{\phi} = \left\{ \phi \in \mathcal{P} \;\middle|\; T\|\hat{\phi} - \phi\|^2_{\hat{\Lambda}^{-1}} \leq c(\chi^2_m) \right\},$$

where $c(\chi^2_m)$ is the $1-\tau$ quantile of a χ^2 distribution with m degrees of freedom. Then,

$$CS^{\theta}_{U} = \bigcup_{\phi \in CS^{\phi}} \Theta(\phi). \tag{41}$$

This confidence set is valid because

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

The following lemma provides a convenient representation for CS^{θ}_{U}.

Lemma 5 *Suppose Assumption 2 is satisfied. Then the confidence set constructed by taking unions of the identified-sets, CS_U^θ, defined in (41) can be represented as*

$$CS_U^\theta = \left\{ \theta \;\middle|\; M_\theta \theta \geq 0 \text{ and } Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq c(\chi^2_m) \right\}, \tag{42}$$

where $\hat{W}^*(\cdot)$ is defined according to (23).

Lemma 5 implies that $CS_{(1)}^\theta$ defined in (25) and CS_U^θ are identical except for the critical value that is used to construct the level set. Since $\tilde{k} + \tilde{r}_2 < m$, the confidence set constructed by taking unions of the identified-sets is more conservative than our proposed confidence sets based on the profile and the projection approach.

Cumulative Impulse Responses. As in Section 2, consider a bivariate VAR composed of inflation and output growth, but now with nontrivial 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 nonnegative in periods 0 and 1. This case can be handled by defining the cumulative responses. It can be verified that the cumulation of responses is a special case of the quasi-triangular transformation $M^{S,2}$. Thus, Assumption 3 is satisfied.

Sign Restrictions Combined with Zero Restrictions. In our framework it is straightforward to sharpen the identified set by combining sign restrictions with more traditional exclusion restrictions. Zero restrictions, e.g., on the impact or long-run effect of the shock of interest on the variables y_t could be imposed in one of two ways. First, in applications in which the zero restrictions are imposed upon the impact effect of the shock and the variables y_t are ordered appropriately, the zero restrictions easily translate into domain restrictions for q. For instance, in the application presented in Section 6 below, we consider a four-variable VAR and can impose the equality restrictions of interest by setting the first two elements of q equal to zero. Second and more generally, one can modify the function $Q(\theta; \phi, W)$ in (18) as follows:

$$Q(\theta; \phi, W) = \min_{\|q\|=1,\, \mu \geq 0} \left\| \begin{pmatrix} \begin{bmatrix} \tilde{S}_\theta(q) \\ \tilde{S}_{eq}(q) \end{bmatrix} \phi - \begin{bmatrix} \theta \\ 0 \end{bmatrix} \\ \tilde{S}_R(q)\phi - \mu \end{pmatrix} \right\|_W^2 ,$$

where $\tilde{S}_{eq}(q)\phi$ corresponds to the responses that are restricted to be zero. Since the construction of our confidence sets is based on a point-wise testing procedure that conditions on $\theta \in \Theta(\phi)$, 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. Thus, the results in Theorems 1 and 2 are directly applicable.

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

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

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 in the equations in Section 4.2), the implementation becomes computationally more difficult because the objective function $G(\theta, \Omega; \hat{\phi}, \hat{W}^*(\cdot))$ now has to be minimized over the domain of Ω rather than the unit hypersphere.

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

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

where H is potentially infinite. Thus, our linear function $\tilde{S}(q)\phi$ would have to be replaced by a nonlinear function of the form $\tilde{S}(q, \phi)$, which in turn needs to be approximated with a first-order Taylor expansion. Since most of the recent empirical literature focuses on impulse responses, we do not pursue the extension to nonlinear functions $\tilde{S}(q, \phi)$ in this paper.

Nonstationary VARs. Some authors, e.g., Uhlig (2005), specify the VAR in terms of variables that exhibit (near) nonstationary dynamics, such as the log level of GDP, or the log levels of consumer or commodity price indices. We assumed that $\hat{\phi}$ has a Gaussian limit distribution. This assumption is violated in VARs with nonstationary endogenous variables, see Phillips (1998). However, an extension of our analysis to VARs with unit roots or cointegration restrictions is beyond the scope of this paper because even the construction of a uniformly valid confidence interval for ϕ in such an environment is very challenging, e.g., Mikusheva (2007).

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

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

If the error loss function is symmetric and $\Theta(\phi)$ is an interval, then it is optimal to choose the mid point of the interval. A general analysis of minimax decision problems in interval identified models is provided by Song (2009).

5 Monte Carlo Illustrations

In this section we conduct two Monte Carlo experiments to illustrate the properties of our proposed confidence sets. The first experiment is based on the simple example of Section 2. For the second experiment we introduce some autoregressive dynamics to examine the effect of serial correlation on the estimation of the reduced form parameters as well as the impulse responses. The simulation designs, summarized in Table 1, are obtained by fitting a VAR(0) to data on U.S. inflation and GDP growth (Section 5.1) and fitting VAR(1)s to inflation and either output growth or linearly detrended log GDP (Section 5.2).

We also provide a comparison between frequentist confidence sets and Bayesian credible sets. The Bayesian credible sets are based on a Gaussian VAR that can be written as a linear regression model of the form

$$y_t' = x_t'\Phi + u_t', \quad u_t \sim \mathcal{N}(0,\Sigma). \tag{43}$$

Here $x_t' = [y_{t-1}', \ldots, y_{t-p}']$ and $\Phi = [\Phi_1, \ldots, \Phi_p]'$. The matrices Φ and Σ collect the reduced-form parameters of the VAR. We now introduce an 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

$$p(\Phi,\Sigma,\tilde{q}) \propto |\Sigma|^{-(n+1)/2} \exp\{-\tilde{q}'\tilde{q}/2\}\mathcal{I}\{(\Phi,\Sigma,\tilde{q}) \in \mathcal{S}\}. \tag{44}$$

$\mathcal{S}$ denotes the set of triplets $(\Phi, \Sigma, \tilde{q})$ such that the impulse responses of the corresponding structural VAR satisfy the sign restrictions:

$$\mathcal{S} = \left\{ (\Phi, \Sigma, \tilde{q}) \;\middle|\; \exists\, \theta \in \Theta \text{ s.t. } \tilde{S}_{\theta}(\tilde{q}/\|\tilde{q}\|)\phi(\Phi, \Sigma) = \theta,\ \tilde{S}^{*}_{R}(\tilde{q}/\|\tilde{q}\|)\phi(\Phi, \Sigma) \geq 0 \right\}.$$

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.

5.1 Experiment 1

The first Monte Carlo experiment is based on the simple example discussed in Section 2: $y_t = u_t$ where $u_t \sim iidN(0, \Sigma_u)$. The parameterization of the data generating process is provided in Table 1 in the column labeled *Design 1.* The goal is to obtain a confidence set for the response of $y_{1,t}$ to $\epsilon_{1,t}$, denoted by θ. According to our simulation design, the identified set for θ is $\Theta(\phi_0) = [0, 0.578]$. The objective function for the construction of the confidence set is given by (6), where the generic weight matrix W is replaced by the optimal weight matrix $W^*(q)$ in (23). The computation of $W^*(q)$ requires an estimate of the asymptotic covariance matrix Λ. To obtain this estimate we use a parametric bootstrap to approximate the sampling distribution of $\hat{\phi}$.[7] Since $r_2 = 1$ and the objective function incorporates only one inequality condition - which is binding at the boundary of the confidence interval - we only compute the profile-objective-function based confidence set $CS^{\theta}_{(1)}$ with critical value $c^{(1)}_{1,1} = 3.82$. The lower bound of the confidence set is equal to zero, and we find its upper bound by step-wise expansion of the upper bound of $\Theta(\phi_0)$. To conduct the Monte Carlo experiment, the following steps are repeated n_{sim} times:

1. Generate a sample of size T from the data generating process.
2. Compute $\hat{\phi}$, $\hat{\Lambda}$, and the upper bound of $\Theta(\hat{\phi})$.
3. Compute the upper bound of the 90% frequentist confidence interval $CS^{\theta}_{(1)}$.
4. Compute a 90% Bayesian credible set for θ based on the Bayesian VAR given by (43) and (44).

[7] Conditional on the estimate $\hat{\Sigma}_{tr}$ we generate bootstrap samples from $y^*_t = \hat{\Sigma}_{tr} v^*_t$ where $v^*_t \sim N(0, I_2)$. For each bootstrap sample, we compute $\hat{\phi}^*$ and estimate $\hat{\Lambda}$ as the covariance matrix of $\hat{\phi}^*$ across bootstrap samples.

The results for $n_{sim} = 10$ and $T = 100$ are plotted in Figure 1. The $x - axis$ denotes the iteration of the simulation algorithm. Since the lower bounds of $\Theta(\phi_0)$, $\Theta(\hat{\phi})$, and the frequentist confidence intervals are zero, our discussion focuses on the upper bound. We reordered the simulations according to the upper bound of $\Theta(\hat{\phi})$. In about half of the simulations, the upper bound of $\Theta(\hat{\phi})$ exceeds that of $\Theta(\phi_0)$. In 9 out of 10 repetitions, each value $\theta \in \Theta(\phi)$ is contained in the frequentist confidence set $CS^{\theta}_{(1)}$. The upper bound of the Bayesian credible sets essentially coincides with the upper bound of $\Theta(\hat{\phi})$. This is consistent with the formula for $p(\theta|\phi)$ in (4), which implies that in this stylized model the prior density is increasing in θ, and conditional on the reduced form parameters, peaks at the upper bound of $\Theta(\phi)$. The lower bounds of the Bayesian intervals are strictly greater than zero, which means that for our relatively large sample the Bayesian credible intervals lie inside $\Theta(\hat{\phi})$, a point emphasized in Moon and Schorfheide (2009).

If we increase the number of repetitions to $n_{sim} = 1,000$, then the upper bound of the identified set is covered by the frequentist interval in 93% of the repetitions and by the Bayesian interval only in 40% of the repetitions. Detailed results for the frequentist confidence interval are summarized in Table 2. At a sample size of $T = 5,000$, the actual coverage probability of the reduced-form parameter confidence set CS^{ϕ} equals the nominal coverage probability of 90%. The actual coverage probability of $CS^{\theta}_{(1)}$ for the upper bound of $\Theta(\phi)$, on the other hand, is 95% instead of 90%. Since the critical value that we use to construct the impulse response confidence interval $CS^{\theta}_{(1)}$ is based on an upper bound of the criterion function in which we replace the argmin with respect to q by a specific value $\tilde{q}$, see (28), the resulting confidence interval is conservative. In this particular design with $\tilde{k} = 1$ and $\tilde{r}_2 = 1$, there is no gain from the moment-selection approach and the simulation of q-specific critical values, because at the upper of the identified set the one and only moment inequality is binding. Thus, $CS^{\theta}_{(1)} = CS^{\theta}_{(2)}$ and the discrepancy between actual and nominal coverage probability can also be interpreted as conservativeness induced by the projection approach.

For smaller sample sizes of $T = 100$ and $T = 500$, the coverage probability of CS^{ϕ} is 82% and 94% and thus deviates from the desired nominal size. Since $CS^{\theta}_{(1)}$ is conservative, a fairly low coverage probability for ϕ at $T = 100$ still translates into a confidence interval for θ that exceeds the nominal coverage probability. The average length of $CS^{\theta}_{(2)}$ shrinks from 0.64 for $T = 100$ to 0.59 for $T = 5,000$. Thus, as the sample size increases, the length of the confidence interval approaches the length of the identified set. As a comparison, the 90% Bayesian credible sets have an average length of 0.5, which is less than the length of the identified set. From a frequentist perspective, the

Bayesian intervals have a coverage probability of about 45%.

5.2 Experiment 2

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

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

Our 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 $R_1^v = \Phi\Sigma_{tr}$. The structural parameter of interest, θ, is defined as $\partial y_{1,t+1}/\partial \epsilon_{1,t}$ and we impose the sign restrictions that both θ as well as $\partial y_{2,t+1}/\partial \epsilon_{1,t}$ are nonnegative. To simplify the computations, in particular the evaluation and minimization of the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big)$, we do not impose sign restrictions on the responses at impact or at horizons greater than $h = 1$. We follow the steps outlined in Section 5.1 to implement the Monte Carlo experiment. A simplified representation for the objective function $G\big(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)\big)$ can be found in the Appendix.

The simulation results for confidence intervals with a nominal coverage of 90% are summarized in Table 2. We consider sample sizes $T = 100$ and $T = 500$. As for the VAR(0), the confidence intervals for the impulse response are generally conservative. The coverage probabilities reported in the table refer to the upper endpoint of the identified interval $\Theta(\phi_0)$. The actual coverage probabilities range from 92% to 98% and reflect the somewhat distorted coverage probabilities of CS^{ϕ}, which range from 83% to 95%. However, due to the conservativeness of the critical values that are used to construct $CS^{\theta}_{(1)}$, its actual coverage probability never falls below 90%. The average length of the confidence sets under *Design 2* and *Design 3* drops by about 10% as the sample size is increased from 100 to 500 observations. For *Design 4* the reduction is slightly larger than 20%. For $T = 500$ the confidence intervals are about 10% longer than the identified sets.

6 Empirical Illustration

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

6.1 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 civilian noninstitutionalized 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. All VAR's are estimated with $p = 2$ lags.

6.2 Pure Sign Restrictions

In order to make inference about the effects of a contractionary monetary policy shock, the following sign restrictions are used to bound the identified set: in periods $h = 0, 1$ (i) the inflation response is nonpositive; (ii) the interest rate response is nonnegative; (iii) real money balances do not rise above their steady state level. Figure 2 depicts three bands: (point-wise) 90% frequentist confidence intervals, estimated sets $\Theta(\hat{\phi})$, and (point-wise) 90% Bayesian credible sets. The confidence intervals are the ones obtained from the projection-based approach, using the Andrews and Soares (2010a) moment selection with simulated critical values, denoted by $CS^{\theta}_{(2)}(\hat{c}_{22})$ in Section 4. 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 $\Theta(\hat{\phi})$. As explained in detail in Moon and Schorfheide (2009), in a large sample (a sample in which uncertainty about ϕ is small compared with the size of $\Theta(\phi)$) the Bayesian intervals lie inside the estimated set $\Theta(\hat{\phi})$ because in the limit essentially all the probability mass is concentrated on $\Theta(\hat{\phi})$ 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 $\Theta(\hat{\phi})$ because it has to have, say, 90% coverage probability for every element of the identified set $\Theta(\phi)$, including the boundary points. From a substantive perspective, the use of sign restrictions leaves the direction of the output response undetermined.

Figure 3 compares the profile-objective-function-based confidence set $CS^{\theta}_{(1)}$ of Section 4.2 with the two projection-based sets $CS^{\theta}_{(2)}(\hat{c}_{21})$ and $CS^{\theta}_{(2)}(\hat{c}_{22})$ of Section 4.2. As discussed previously, the three sets are nested, which is also apparent from the figure. In the top panel, the sign restrictions are imposed over the horizons $h = 0$ and $h = 1$. It turns out that the error bands are very similar because the moment selection procedure only eliminates very few inequalities, or, in other words, most of the inequalities seem to be binding at the boundary of the confidence sets. In the bottom panel 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. While the gain from eliminating nonbinding moment conditions in itself is relatively small, i.e., $CS^{\theta}_{(1)}$ and $CS^{\theta}_{(2)}(\hat{c}_{21})$ are quite similar, replacing the conservative critical value $\hat{c}_{21}$ by the simulated critical value $\hat{c}_{22}$ generates substantially smaller bands. A comparison of the $CS^{\theta}_{(2)}(\hat{c}_{21})$ intervals for output in the top and the bottom panel of Figure 3 suggests that expanding the horizon over which the sign restrictions are imposed increases the uncertainty. However, the widening of the error bands is due to an increase in the conservativeness of the confidence intervals. It turns out that the use of simulated critical values corrects the paradoxical feature: as we expand the horizon, the $CS^{\theta}_{(2)}(\hat{c}_{22})$ bands indeed shrink.

6.3 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 within the period, see for instance Boivin and Giannoni (2006). Since the initial impact of the monetary policy shock is given by $\Sigma_{tr} q$ and 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 6.2. 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 nonpositive and the interest rate responses for $h = 0$ and $h = 1$ are nonnegative.

Impulse response bands are depicted in Figure 4. The first panel compares the frequentist bands $CS^{\theta}_{(2)}(\hat{c}_{22})$, Bayesian credible bands, and the estimated sets $\Theta(\hat{\phi})$. A comparison of $\Theta(\hat{\phi})$ in Figures 2 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 three different frequentist confidence intervals considered in Sections 4.2 and 4.3. It turns out that the two projection-based confidence intervals are essentially indistinguishable, which means that there is no gain from simulating the critical values. However, the projection-based intervals are noticeably narrower than the profile-objective-function based bands. It turned out that the sign restrictions imposed on the real money balance response (not shown in the figure) are not binding. Thus, the moment selection procedure is able to produce somewhat sharper inference.

7 Conclusion

This paper develops methods to construct error bands for impulse responses in VARs that are identified based on sign restrictions. The error bands that have been reported in the literature thus far were only meaningful from a Bayesian perspective. Our empirical application illustrates that in partially identified VARs, frequentist error bands can be substantially wider than Bayesian error bands. The impulse response confidence intervals are constructed through a point-wise testing procedure, which is a technique that is widely used in models that are either weakly or only partially identified. We consider three different intervals. The most conservative one is based on a profile-objective function that concentrates out the nuisance parameter q, which maps the orthogonolized shocks into the structural shock of interest. The advantage of this confidence set is that it is easy to compute because it relies on an asymptotic critical value that is nuisance parameter free. The sharpest confidence intervals are obtained by a projection-approach that combines the Andrews and Soares (2010a) moment selection procedure with simulated critical values that depend on q. Its disadvantage is that it takes a long time to compute because each simulated critical value requires the solution of thousands of quadratic programming problems. As a by-product, we also provide a procedure to compute the set $\Theta(\hat{\phi})$ for impulse responses conditional on the estimated reduced form parameters. Since in a Bayesian analysis, the prior distribution of the impulse response functions conditional on the reduced form parameters does not get updated, it is useful to report the identified set conditional on some estimate, say, the posterior mean of Φ and Σ so that the audience can judge whether the conditional prior distribution is highly concentrated in a particular area of the identified set.

References

Anderson, T.W. and Herman Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, **20**, 46-63.

Andrews, Donald and Patrick Guggenberger (2009): "Validity of Subsampling and 'Plug-in Asymptotics' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, **25**, 669-709.

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.

Blanchard, Olivier, and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, **79**, 655-673.

Boivin, Jean, and Marc Giannoni (2006): "Has Monetary Policy Become More Effective?" *Review of Economic and Statistics*, **88**, 445-462.

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

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

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

Dedola, Luca and Stefano Neri (2007): "What Does a Technology Shock Do? A VAR Analysis with Model-Based Sign Restrictions," *Journal of Monetary Economics*, **54**, 512-549.

Dufour, Jean-Marie (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, **65**, 1365-1387.

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

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

Kilian, Lutz and Dan Murphy (2009): "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Manuscript*, University of Michigan.

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

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

Moon, Hyungsik Roger and Frank Schorfheide (2009): "Bayesian and Frequentist Inference in Partially Identified Models," *NBER Working Paper*, **14882**.

Mountford, Andrew and Harald Uhlig (2008): "What are the Effects of Fiscal Policy Shocks?," *NBER Working Paper*, **14551**.

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

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.

Perlman, Michael (1969): "One-Sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, **40**, 549-567.

Phillips, Peter (1998): "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VAR's," *Journal of Econometrics*, **83**, 21-56.

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.

Song, Kyungchul (2009): "Point Decisions for Interval-Identified Parameters," *Manuscript*, University of Pennsylvania.

Staiger, Douglas and James Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, **65**, 557-586.

Uhlig, Harald (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, **52**, 381-419.

A Proofs of Main Theorems

This section provides proofs for Theorems 1 and 2. The proofs make use of various lemmas that are stated and proved in the Online Appendix that accompanies this paper. The proof of Theorem 2 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 $\tilde{S}(q)$.

Proof of Theorem 1: Define $CP_T(\phi,\theta) = P_\phi\{\theta \in CS^\theta_{(1)}\}$ and

$$AsyCP = \lim\,\inf_T \inf_{\phi\in\mathcal{P},\,\theta\in\Theta(\phi)} CP_T(\phi,\theta).$$

Then there exists a sequence $\{\phi_T,\theta_T\}$ such that $\theta_T \in \Theta(\phi_T)$ and

$$AsyCP = \lim\,\inf_T CP_T(\phi_T,\theta_T).$$

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

$$AsyCP = \lim_{T'} CP_T(\phi_{T'},\theta_{T'}).$$

Recall the definition $\Sigma(q) = S(q)\Lambda S(q)'$. We will use the decompositions $\Lambda(\phi) = L(\phi)L'(\phi)$ and $\Sigma(q) = D^{1/2}(q)\Omega(q)D^{1/2}(q)$, where $D^{1/2}(q)$ is a diagonal matrix of standard deviations and $\Omega(q)$ is a correlation matrix. Without loss of generality we can choose a further subsequence $\{T''\} \subseteq \{T'\}$ such that the following conditions are satisfied: (i) $\phi_{T''} \longrightarrow \phi$ and $\Lambda(\phi_{T''}) \longrightarrow \Lambda(\phi)$. (ii) $r_2(q_{T''}) = r_2$, $k(q_{T''}) = k$, and $l(q_{T''}) = k + r_2$. (iii) $A(q_{T''}) = [D^{-1/2}(q_{T''})S(q_{T''})L(q_{T''})]' \longrightarrow A$ and $\Omega(q_{T''}) \longrightarrow A'A > 0$. Condition (ii) can be satisfied because the row dimensions are integer valued. Condition (iii) is a consequence of Lemma B 5. Along the T'' sequence, the rank of $S(q_{T''})$ stays constant.

Before proceeding with the proof, it is instructive to consider the following example. Suppose $q = [q_1, q_2]'$,

$$S(q) = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_1 & q_2 \end{bmatrix}, \quad L(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma(q) = \begin{bmatrix} q_1^2 & \phi q_1^2 \\ \cdot & (1+\phi^2)q_1^2 + q_2^2 \end{bmatrix}.$$

Thus,

$$D^{1/2}(q) = \begin{bmatrix} q_1 & 0 \\ 0 & \sqrt{q_1^2(1+\phi^2)+q_2^2} \end{bmatrix}, \quad \Omega(q) = \begin{bmatrix} 1 & \frac{\phi q_1}{\sqrt{q_1^2(1+\phi^2)+q_2^2}} \\ \cdot & 1 \end{bmatrix}.$$

Now consider a sequence of the form (dropping the primes) $q_T = [c/T, 1 - c/T]'$, $T = 1, 2, \ldots$, with $\lim_{T\longrightarrow\infty}\ q_T = \bar{q} = [0,1]'$. Notice that although the row rank of $S(\bar{q})$ is one and $\Sigma(q_T)$ converges to the reduced rank and noninvertible matrix $\Sigma(\bar{q})$, the rank of $S(q_T)$ is $l(q_T) = 2$ for every T and the following limits are well defined:

$$\lim_{T\longrightarrow\infty} D^{-1/2}(q_T)S(q_T)L(\phi_T) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} L(\bar{\phi}) = I_2 = A', \quad \lim_{T\longrightarrow\infty} \Omega(q_T) = I_2 > 0.$$

For notational convenience from now on we denote $\{T''\}$ as $\{T\}$. For each ϕ_T and $\theta_T \in \Theta(\phi_T)$, there exist q_T and μ_T such that $\|q_T\| = 1$, $\tilde{S}_\theta(q_T)\phi_T = \theta_T$ and $\tilde{S}_R(q_T)\phi_T = \mu_T$. Recall that for $\theta \in \Theta(\phi)$ the penalty term that appears in (24) has to be zero. The starting point for the remainder of the proof is the bounding function obtained in (28):

$$\begin{aligned} \bar{Q}\big(\tilde{q}; \hat{\phi}, \hat{W}^*(\cdot)\big) &= \min_{\nu\geq 0} \left\| S(q_T)\sqrt{T}(\hat{\phi} - \phi_T) - V(q_T)M_\nu \nu \right\|^2_{\hat{\Sigma}^{-1}(q_T)} \\ &= \min_{\nu\geq 0} \left\| S(q_T)\sqrt{T}(\hat{\phi} - \phi_T) - V(q_T)M_\nu \nu \right\|^2_{\Sigma^{-1}(q_T)} + o_p(1) \\ &= \min_{\nu\geq 0} \left\| D^{-1/2}(q_T)S(q_T)L(\phi_T)\hat{\zeta}_T - V(q_T)M_\nu \nu \right\|^2_{\Omega^{-1}(q_T)} + o_p(1). \end{aligned} \tag{45}$$

According to Lemma B 1 we can replace $\hat{W}^*(\cdot)$ by $W^*(\cdot)$, which leads to the second equality. The final equality is based on the definition $\hat{\zeta}_T = \sqrt{T}L^{-1}(\phi_T)(\hat{\phi} - \phi_T)$.

For brevity, let $S_T = S(q_T)$, $L_T = L(\phi_T)$, $\Omega_T = \Omega(q_T)$, $D_T = D(q_T)$, and $A_T = A(q_T)$. We consider the following three cases: (i) $r_2 \geq 1$, (ii) $k = 1$ and $r_2 = 0$, and (iii) $l = k = r_2 = 0$.

Case (i): $r_2 \geq 1$. Partition $S_T' = [S_{1,T}',\ S_{2,T}]'$, where $S_{2,T}$ is the last row of S_T. Since $r_2 \geq 1$, it is guaranteed that $S_{2,T}$ is nonempty. (When $k = 0$, we set $S_T = S_{2,T}$ and there is no partition.)

Denote the conforming partitions of Σ_T by $\Sigma_{ij,T} = S_{i,T}\Lambda_T S_{j,T}'$. Moreover, let Ω_T be composed of $\Omega_{ij,T} = D_{i,T}^{-1/2} S_{i,T}\Lambda S_{j,T}' D_{j,T}^{-1/2}$ and $A_{i,T} = L_T' S_{i,T}' D_{i,T}^{-1/2}$, $i, j = 1, 2$. Finally, define the projection matrix $P_{A_{i,T}} = A_{i,T}(A_{i,T}'A_{i,T})^{-1}A_{i,T}'$ and denote the last element of $V(q_T)M_\nu \nu$ by ν_2. Then, using a similar argument as in the main text

$$\begin{aligned} &\min_{\nu\geq 0} \left\| D_T^{-1/2} S_T L_T \hat{\zeta}_T - V(q_T)M_\nu \nu \right\|^2_{\Omega_T^{-1}} \\ &\leq \left\| A_{1,T}'\hat{\zeta}_T \right\|^2_{(A_{1,T}'A_{1,T})^{-1}} + \min_{\nu_2\geq 0} \left\| A_{2,T}'(I - P_{A_{1,T}})\hat{\zeta}_T - \nu_2 \right\|^2_{(A_{2,T}'(I-P_{A_{2,T}})A_{2,T})^{-1}} \\ &\Longrightarrow \chi^2_{k+r_2-1} + \mathcal{I}\{Z \leq 0\}Z^2. \end{aligned} \tag{46}$$

The inequality is obtained by setting all but the very last element of the vector $V(q_T)M_\nu\nu$ equal to zero. Due to the projection, $Z \sim N(0,1)$ is independent of the χ^2_{l-1} random variable in the limit distribution. Let $c^{(1)}_l$ be the $1-\tau$ critical value associated with the limit distribution $\chi^2_{l-1} + \mathcal{I}\{Z \leq 0\}Z^2$. Since $l = k + r_2 \leq \tilde{k} + \tilde{r}_2$, we obtain $c^{(1)}_{(k,r_2)} \leq c^{(1)}_{(\tilde{k},\tilde{r}_2)}$ and can deduce from (45) and (46) that

$$\begin{aligned} AsyCP &= \lim_T \; P_{\phi_T}\left\{\left[\min_{\nu\geq 0}\left\|D_T^{-1/2}S_T L_T\hat{\zeta}_T - V(q_T)M_\nu\nu\right\|^2_{\Omega_T^{-1}}\right] \leq c^{(1)}_{\tilde{k}+\tilde{r}_2}\right\} \\ &\geq \lim_T \; P_{\phi_T}\left\{\left[\left\|A'_{1,T}\hat{\zeta}_T\right\|^2_{(A'_{1,T}A_{1,T})^{-1}}\right.\right. \\ &\quad \left.\left. + \min_{\nu_2\geq 0}\left\|A'_{2,T}(I-P_{A_{1,T}})\hat{\zeta}_T - \nu_2\right\|^2_{(A'_{2,T}(I-P_{A_{2,T}})A_{2,T})^{-1}}\right] \leq c^{(1)}_l\right\} \\ &= 1-\tau. \end{aligned}$$

Case (ii): $k = 1$ and $r_2 = 0$. In this case, $V(q_T)M_\nu = 0$ and

$$\min_{\nu\geq 0}\left\|D_T^{-1/2}S_T L_T\hat{\zeta}_T - V(q_T)M_\nu\nu\right\|^2_{\Omega_T^{-1}} = \left\|D_T^{-1/2}S_T L_T\hat{\zeta}_T\right\|^2_{\Omega_T^{-1}} \Longrightarrow \chi^2_1.$$

Since $c^{(1)}_{(1,0)} \leq c^{(1)}_{(\tilde{k},\tilde{r}_2)}$, we have

$$\begin{aligned} AsyCP &= \lim_T \; P_{\phi_T}\left\{\left[\min_{\nu\geq 0}\left\|D_T^{-1/2}S_T L_T\hat{\zeta}_T - \nu\right\|^2_{\Omega_T^{-1}}\right] \leq c^{(1)}_{(\tilde{k},\tilde{r}_2)}\right\} \\ &\geq \lim_T \; P_{\phi_T}\left\{\left[\min_{\nu\geq 0}\left\|D_T^{-1/2}S_T L_T\hat{\zeta}_T - \nu\right\|^2_{\Omega_T^{-1}}\right] \leq c^{(1)}_{(0,1)}\right\} \\ &= 1-\tau, \end{aligned}$$

as required for Case (ii).

Case (iii): $l = k = r_2 = 0$. In this case the objective function is zero for $\theta_T \in \Theta(\phi_T)$ and it is guaranteed that θ_T is included in the confidence set. This completes the proof of the theorem. □

Proof of Theorem 2: According to Lemma B 6 it suffices to show that $CS^{\theta,q}_{(2)}$ is a valid $1-\tau$ confidence set. Denote $CP_T(\phi,\theta,q) = P_\phi\{\theta \in CS^{\theta,q}_{(2)}\}$ and

$$AsyCP = \liminf_T \inf_{\phi\in\Phi} \inf_{\theta\in\Theta(\phi)} \inf_{q\in\mathbb{Q}(\theta,\phi)} CP_T(\phi,\theta,q).$$

Then, there exist sequences $\{\phi_T, \theta_T, q_T\}$ such that $\theta_T \in \Theta(\phi_T)$, $q_T \in \mathbb{Q}(\theta_T,\phi_T)$, and

$$AsyCP = \liminf_T CP_T(\phi_T,\theta_T,q_T).$$

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

$$AsyCP = \lim_{T'} CP_{T'}(\phi_{T'}, \theta_{T'}, q_{T'}).$$

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

$$\lim_{T''} CP_{T''}(\phi_{T''}, \theta_{T''}, q_{T''}) \geq 1 - \tau, \tag{47}$$

which proves the theorem.

Since $q_{T'} \in \mathbb{Q}\left(\theta_{T'}, \phi_{T'}\right)$, we have $\theta_{T'} = S_\theta\left(q_{T'}\right)\phi_{T'}$. Moreover, the penalty term that appears in (24) is equal to zero. Define $\mu\left(q_{T'}, \phi_{T'}\right) = S_R\left(q_{T'}\right)\phi_{T'}$. Then, by Lemma B 2, we have

$$G\big(\theta_{T'}, q_{T'}; \hat{\phi}, \hat{W}^*(\cdot)\big) = G\big(\theta_{T'}, q_{T'}; \hat{\phi}, W^*(\cdot)\big) + o_p(1),$$

where

$$\begin{aligned} & G\big(\theta_{T'}, q_{T'}; \hat{\phi}, W^*(\cdot)\big) \\ = & \min_{v \geq -\sqrt{T}\mu(q_{T'}, \phi_{T'})} \left\| S\left(q_{T'}\right)\sqrt{T'}\left(\hat{\phi} - \phi_{T'}\right) - M_v v \right\|^2_{\Sigma^{-1}(q_{T'})} \\ = & \min_{v \geq -\sqrt{T} D_R^{-1/2}\mu(q_{T'}, \phi_{T'})} \left\| D^{-1/2} S\left(q_{T'}\right)\sqrt{T'}\left(\hat{\phi} - \phi_{T'}\right) - M_v v \right\|^2_{\Omega^{-1}(q_{T'})}. \end{aligned}$$

Here we used the notation that $\Sigma(q) = S(q)\Lambda S(q)'$ with the factorization $\Sigma(q) = D^{1/2}(q)\Omega(q)D^{1/2}(q)$, where Ω is a correlation matrix and $D^{1/2}$ is a diagonal matrix of standard deviations that can be partitioned into $diag(D_\theta^{1/2}, D_R^{1/2})$. The partitions conform with $S(q) = [S'_\theta(q), S'_R(q)]'$.

The subsequence T'' is chosen such that the following conditions are satisfied: (i) $\phi_{T''} \longrightarrow \phi$ and $\Lambda(\phi_{T''}) \longrightarrow \Lambda$. (ii) $k(q_{T''}) = k$, $r_2(q_{T''}) = r_2$, $l(q_{T''}) = l$. (iii) For $j = 1, \ldots, r_2$ the slackness in inequality j converges to

$$\sqrt{T''}\mu_j(q''_T, \phi''_T) \quad \longrightarrow \quad h_j \tag{48}$$

$$\kappa_{T''}^{-1} D_{jj,R}^{-1/2}(q_{T''})\sqrt{T''}\mu_j(q''_T, \phi''_T) \quad \longrightarrow \quad \pi_j. \tag{49}$$

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 $\pi_j = \infty$. Roughly speaking, in case (a) the slackness is small and the selection criterion regards the inequality asymptotically as binding. In case (c) the slackness is large and the selection criterion regards the inequality as nonbinding. (iv) $[D^{-1/2}(q_{T''})S(q_{T''})L(\phi_{T''})]' \longrightarrow A$ and the correlation matrix $\Omega(q_{T''}) \longrightarrow A'A > 0$. Condition (ii) can be satisfied because the row dimensions are integer-valued. Lemma B 5 guarantees the existence of the full rank matrix A in condition (iv).

We now reorder the rows of $S(q_{T''})$ such that $\pi_j = 0$ for rows $j = 1, \ldots, r_{21}$ and $\pi_j > 0$ for rows $j = r_{21}+1, \ldots, r_2$. Under this ordering, the moment-selection procedure will eventually eliminate the last $r_{22} = r_2 - r_{21}$ rows of $S(q_{T''})$. For notational convenience, instead of using the subsubsequence notation T'', we shall use T from now on. Moreover, we denote $\mu_T = \mu(q_T, \phi_T)$, $L_T = L(\phi_T)$, $D_T = D(q_T)$, and $\Omega_T = \Omega(q_T)$.

In what follows we distinguish three cases: (i) $l > 0$ and $l > r_{22}$. When $l > r_{22}$, along the sequence q_T, either the selection criterion picks up at least one inequality as binding (i.e., $r_{21} \geq 1$), or $S_\theta(q)$ is not zero (i.e., $k \geq 1$). (ii) $l = r_{22}$, which implies that $k = 0$ and $r_{12} = 0$. (iii) Finally, we consider $l = 0$.

Case (i): $l > 0$, and $l > r_{22}$. We consider the fixed asymptotic critical value approach and the simulated critical value approach separately.

Fixed Asymptotic Critical Value Approach: Conformable to the dimensions r_{21} and r_{22}, partition

$$\mu_T = [\mu_{1,T}',\ \mu_{2,T}']', \quad \nu = [\nu_1',\ \nu_2']', \quad S_{R,T} = [S_{1,R,T}',\ S_{2,R,T}']'.$$

Moreover, let

$$S_{1,T} = [S_{\theta,T}',\ S_{1,R,T}']', \quad \text{and} \quad S_{2,T} = S_{2,R,T}.$$

and partition

$$\Sigma_T = \begin{bmatrix} \Sigma_{11,T} & \Sigma_{12,T} \\ \Sigma_{12,T} & \Sigma_{22,T} \end{bmatrix}, \quad \Omega_T = \begin{bmatrix} \Omega_{11,T} & \Omega_{12,T} \\ \Omega_{12,T} & \Omega_{22,T} \end{bmatrix}.$$

Notice that when $r_{21} = 0$, then $S_{1,T} = S_{\theta,T}$. If $r_{22} = 0$, then the $S_{2,T}$ partition is empty. According to Lemma B 7 we can use the following approximation:

$$G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) = G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) + o_p(1),$$

where

$$G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) = \min_{\nu_1 \geq -\sqrt{T}\mu_{1,T}} \left\| S_{1,T}\sqrt{T}(\hat{\phi} - \phi_T) - M_{\nu_1}\nu_1 \right\|^2_{\Sigma_{11,T}^{-1}}.$$

Using an argument similar to the one used in the proof of Theorem 1, we deduce

$$\lim_T \ P_{\phi_T}\left\{ G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \leq c^{(1)}_{(k, r_{21})} \right\} \geq 1 - \tau. \tag{50}$$

Then along the sequence q_T, we have

$$
\begin{aligned}
AsyCP &= \lim_T \; CP_T(\phi_T, \theta_T, q_T) \\
&= \lim_T \; P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))} \right\} \\
&\ge \lim_T \; P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))}, \; c^{(1)}_{(k,r_{21})} \le c^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))} \right\} \\
&\ge \lim_T \; P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k,r_{21})}, \; c^{(1)}_{(k,r_{21})} \le c^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))} \right\} \\
&\ge \lim_T \; P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k,r_{21})} \right\} + \left[1 - \lim_T \; P_{\phi_T} \left\{ c^{(1)}_{(k,r_{21})} \le c^{(1)}_{(k(q_T),\hat{r}_{21})} \right\} \right] \\
&= \lim_T \; P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k,r_{21})} \right\} \\
&= \lim_T \; P_{\phi_T} \left\{ G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \le c^{(1)}_{(k,r_{21})} \right\} \\
&\ge 1 - \tau.
\end{aligned}
$$

In the preceding inequalities the critical value based on the estimated number of potentially binding moment conditions, $c^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))}$, is replaced by the critical value that depends on the number of binding moment conditions along the q_T sequence, $c^{(1)}_{(k,r_{21})}$. The fifth line holds since $P(A \cap B) = P(A) - P(A \cap B^c) \ge P(A) - P(B^c)$ and the sixth line is a consequence of Lemma B 8. Finally, the last equality follows from (50). Then, we have the required result for (47).

Simulated Critical Value Approach: We start with the simple case when $r_2 = 0$. In this case, since $M_\nu = 0$, we have

$$
G\big(\theta, q; \hat{\phi}, W_T^*(\cdot)\big) = \left\| \hat{D}^{-1/2} S(q) \sqrt{T} (\hat{\phi} - \phi) \right\|^2_{\hat{\Omega}^{-1}(q)} \Rightarrow \chi^2_k. \tag{51}
$$

and $\hat{c}_{(22)}(q)$ is the $(1-\tau)$ quantile of $\bar{\mathcal{G}}\big(\theta, q; \hat{\Omega}_b\big)$, where $\bar{\mathcal{G}}\big(\theta, q; \hat{\Omega}_b\big) = \left\| \hat{A}' Z_m \right\|^2_{(\hat{A}'\hat{A})^{-1}} \sim \chi^2_k$, as required.

Now suppose that $r_2 \ge 1$. Given q and the slackness measure $\xi_T(q)$, define

$$
\varphi_{j,T}(q) = \begin{cases} \infty & \text{if } \xi_{j,T}(q) \ge \kappa_T \\ 0 & \text{otherwise} \end{cases} \tag{52}
$$

for $j = 1, \ldots, r_2(q)$. Let $\varphi_T(q) = [\varphi_{1,T}(q), \ldots, \varphi_{r_2(q)}(q)]'$, and define $\hat{\varphi}_T(q)$ by replacing $\xi_{j,T}(q)$ in (52) with $\hat{\xi}_{j,T}$. Recall that the critical value $\hat{c}_{(22)}(q)$ in Section 4.3 (using slightly different notation) was defined as

$$
\hat{c}_{(22)}(q) = 1 - \tau \text{ quantile of } \left(\min_{\nu \ge -\hat{\varphi}_T(q)} \left\| \hat{A}'(q) Z_m - M_\nu \nu \right\|^2_{(\hat{A}'(q)\hat{A}(q))^{-1}} \right).
$$

We will proceed by defining two additional critical values, denoted by $\hat{c}^*_{(22)}(q)$ and $c^*_{(22)}$. Let

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

for $j = 1, \ldots, r_2(q)$ and collect the individual elements in the vector $\varphi^*_T(q)$. Similarly, define $\hat{\varphi}^*_{j,T}(q)$ and $\hat{\varphi}^*_T(q)$. Notice that by construction $\hat{\varphi}^*_T(q) \geq \hat{\varphi}_T(q)$. Define $\hat{c}^*_{(22)}$ as

$$\hat{c}^*_{(22)}(q) = 1 - \tau \text{ quantile of } \left(\min_{\nu \geq -\hat{\varphi}^*_T(q)} \left\| \hat{A}'(q) Z_m - M_\nu \nu \right\|^2_{(\hat{A}'(q)\hat{A})^{-1}(q)} \right). \tag{53}$$

By construction $\hat{c}^*_{(22)}(q) \leq \hat{c}_{(22)}(q)$. Now consider a sequence $\{\phi_T, \theta_T, q_T\}$ satisfying the above convergence assumptions and define

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

Moreover, collect the π^*_j elements in the vector π^* and define the third critical value $c^*_{(22)}$ as

$$c^*_{(22)} = 1 - \tau \text{ quantile of } \left(\min_{\nu \geq -\pi_*} \left\| A' Z_m - M_\nu \nu \right\|^2_{(A'A)^{-1}} \right). \tag{54}$$

Along the sequence $\{T\}$

$$\begin{aligned} AsyCP &= \lim_T \, CP_T(\phi_T, \theta_T, q_T) \\ &= \lim_T \, P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq \hat{c}_{22}(q_T) \right\} \\ &= \lim_T \, P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \leq \hat{c}_{22}(q_T) \right\} \\ &\geq \lim_T \, P_{\phi_T} \left\{ G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \leq \hat{c}^*_{22}(q_T) \right\}. \end{aligned}$$

The second equality is a consequence of Lemma B 2. The inequality follows from $\hat{c}^*_{(22)} \leq \hat{c}_{(22)}$. By using a similar argument as in Andrews and Guggenberger (2009), it can be shown that

$$\begin{aligned} G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) &= \min_{\nu \geq -D^{-1/2}_{R,T}\sqrt{T}\mu_T} \left\| D_T^{-1/2} S_T L \sqrt{T} L^{-1}(\hat{\phi} - \phi_T) - M_\nu \nu \right\|^2_{\Omega_T^{-1}} \\ &\Longrightarrow \min_{\nu \geq -h} \left\| A' Z_m - M_\nu \nu \right\|^2_{(A'A)^{-1}} \\ &\leq \min_{\nu \geq -\pi^*} \left\| A' Z_m - M_\nu \nu \right\|^2_{(A'A)^{-1}}. \end{aligned}$$

The last inequality holds because along the {T} sequence $h \geq \pi^*$. Recall that $\pi_j = 0$ implies $h_j < \infty$ and $\pi^*_j = 0$. $\pi_j > 0$, on the other hand, implies that $h_j = \pi^*_j = \infty$.

According to Lemma B 9 $\hat{c}^*_{22}(q_T) \xrightarrow{p} c^*_{(22)}$ and we obtain

$$\begin{aligned} AsyCP &\geq P_{\phi_T}\left\{G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \leq \hat{c}^*_{22}(q_T)\right\} \\ &\longrightarrow P\left\{\min_{\nu \geq -h} \left\|A'Z_m - M_\nu \nu\right\|^2_{(A'A)^{-1}} \leq c^*_{(22)}\right\} \\ &\geq P\left\{\min_{\nu \geq -\pi^*} \left\|A'Z_m - M_\nu \nu\right\|^2_{(A'A)^{-1}} \leq c^*_{(22)}\right\} \\ &= 1-\tau. \end{aligned}$$

In the limit provided in the second line we have replaced $\hat{c}^*_{22}(q_T)$ by c^*_{22} using the fact that $c^*_{(22)} > 0$ and that the distribution function of

$$\min_{\nu \geq -\pi_*} \left\|A'Z_m - M_\nu \nu\right\|^2_{(A'A)^{-1}} \tag{55}$$

is continuous near the $1-\tau$'th quantile. The inequality $c^*_{(22)} > 0$ and the continuity near the $1-\tau$'th quantile of the distribution in (55) can be established as follows. Since it is assumed that $l > r_{22}$ and $r_2 \geq 1$, we have $k + r_{21} \geq 1$. Also, it is assumed that $\tau < 1/2$. If $k \geq 1$, then the distribution of (55) is continuous with support $\mathbb{R}_+$. If $k = 0$ and $r_{21} \geq 1$, the distribution of (55) is continuous around the $1-\tau$'th quantile, since there exists at least one zero element in π^* and $\tau < 1/2$. Therefore, we have the desired result for the case of $r_2 \geq 1$.

Combining the results for $r_2 = 0$ and $r_2 \geq 1$, two cases, $r_2 \geq 1$ and $r_2 = 0$, we have established the validity of the simulated critical values for Case (i).

Case (ii): $l = r_{22}$. In this case, we have $k = 0$ and $r_{21} = 0$. Also, $h_j \to \infty$ and $\pi_j > \infty$ for all $j = 1, ..., r_2$. When $\hat{c}_{(2)}(q) = \hat{c}_{(21)}(q)$ (the fixed asymptotic critical value approach), by Lemma B 8 we have $\hat{c}_{(21)}(q_T) = \hat{c}^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))} \geq c^{(1)}_{(k,r_{21})}$ with probability approaching one. Also, by definition we have $c^{(1)}_{(k,r_{21})} = c^{(1)}_{(0,0)} = 0$. Then, as in the Case (i) of the fixed asymptotic critical value approach, we have

$$\begin{aligned} AsyCP &= \lim_T CP_T\left(\phi_T, \theta_T, q_T\right) \\ &= \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, \hat{W}^*\left(\cdot\right)\right) \leq \hat{c}^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))}\right\} \\ &= \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq \hat{c}^{(1)}_{(k(q_T),\hat{r}_{21}(q_T))}\right\} \quad \text{by Lemma B 2} \\ &\geq \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq c^{(1)}_{(k,r_{21})} = 0\right\}. \end{aligned}$$

When $\hat{c}_{(2)}\left(q\right) = \hat{c}_{(22)}\left(q\right)$ (the simulated critical value approach), since $\pi_j > \infty$ for all $j = 1, ..., r_2$, we have $\hat{\varphi}^*_T\left(q_T\right) = \infty$ for all T and $\pi^* = \infty$. Set $\hat{c}^*_{(22)}\left(q_T\right) = 0$ for all T and $c^*_{(22)} = 0$.

Then,

$$\begin{aligned}
AsyCP &= \lim_T CP_T\left(\phi_T, \theta_T, q_T\right) \\
&= \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, \hat{W}^*\left(\cdot\right)\right) \leq \hat{c}_{(22)}\left(q_T\right)\right\} \\
&= \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq \hat{c}_{(22)}\left(q_T\right)\right\} \text{ by Lemma B 2} \\
&\geq \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq \hat{c}^*_{(22)}\left(q_T\right)\right\} \\
&= \lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq c^*_{(22)} = 0\right\},
\end{aligned}$$

where the inequality holds since $\hat{c}_{(22)}\left(q_T\right) \leq \hat{c}^*_{(22)}\left(q_T\right)$.

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

$$\lim_T P_{\phi_T}\left\{G\left(\theta_T, q_T; \hat{\phi}, W^*\left(\cdot\right)\right) \leq 0\right\} \geq 1 - \tau,$$

as desired for Case (ii).

Case (iii): $l = 0$ In this case the objective function is zero, which means that θ_T is included in any confidence set with a nonnegative critical value. □

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

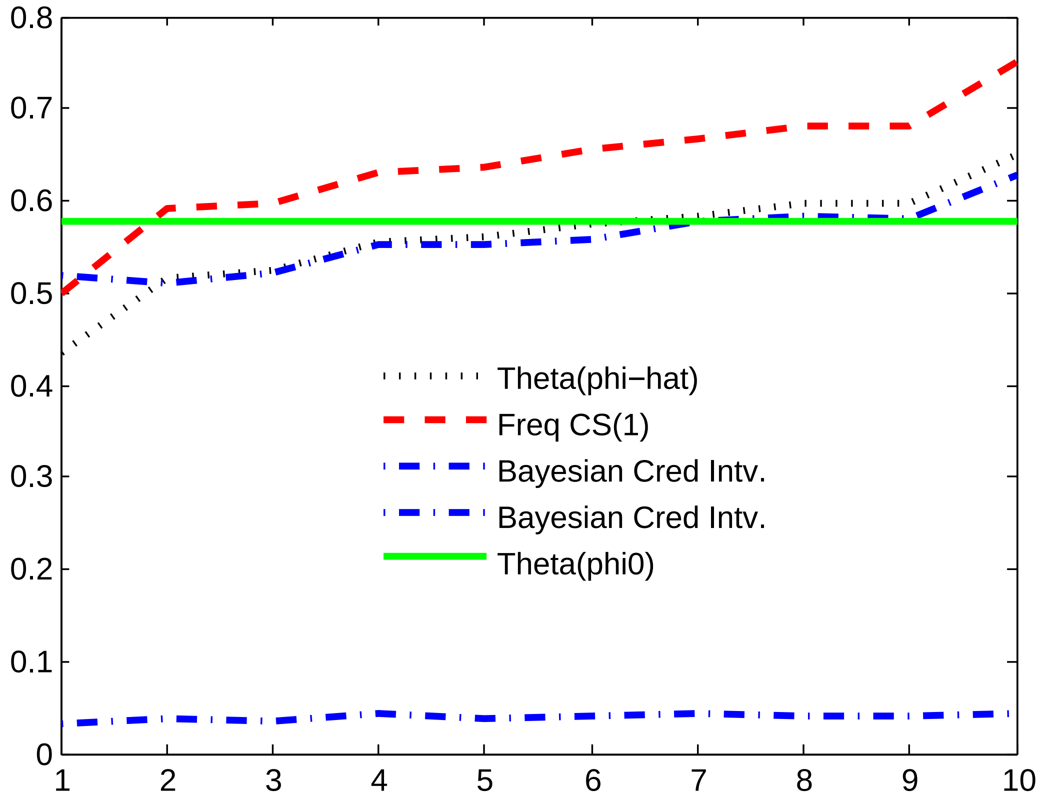
Notes: Designs are obtained by estimating a VAR(0) or VAR(1) of the form $y_t = \Phi_0 + \Phi_1 y_{t-1} + u_t$, $\mathbb{E}[u_t u_t'] = \Sigma_u$. We use OLS estimates, Φ entries refer to elements of Φ_1 and Σ_{ij} entries refer to the (nonredundant) elements of Σ_u. $\lambda_i(\Phi_1)$ is the i'th eigenvalue of Φ_1. $y_{1,t}$ is the log difference of U.S. GDP deflator, scaled by 100 to convert into percentages. $y_{2,t}$ is either the log difference of the U.S. GDP or deviations of the 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.

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

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

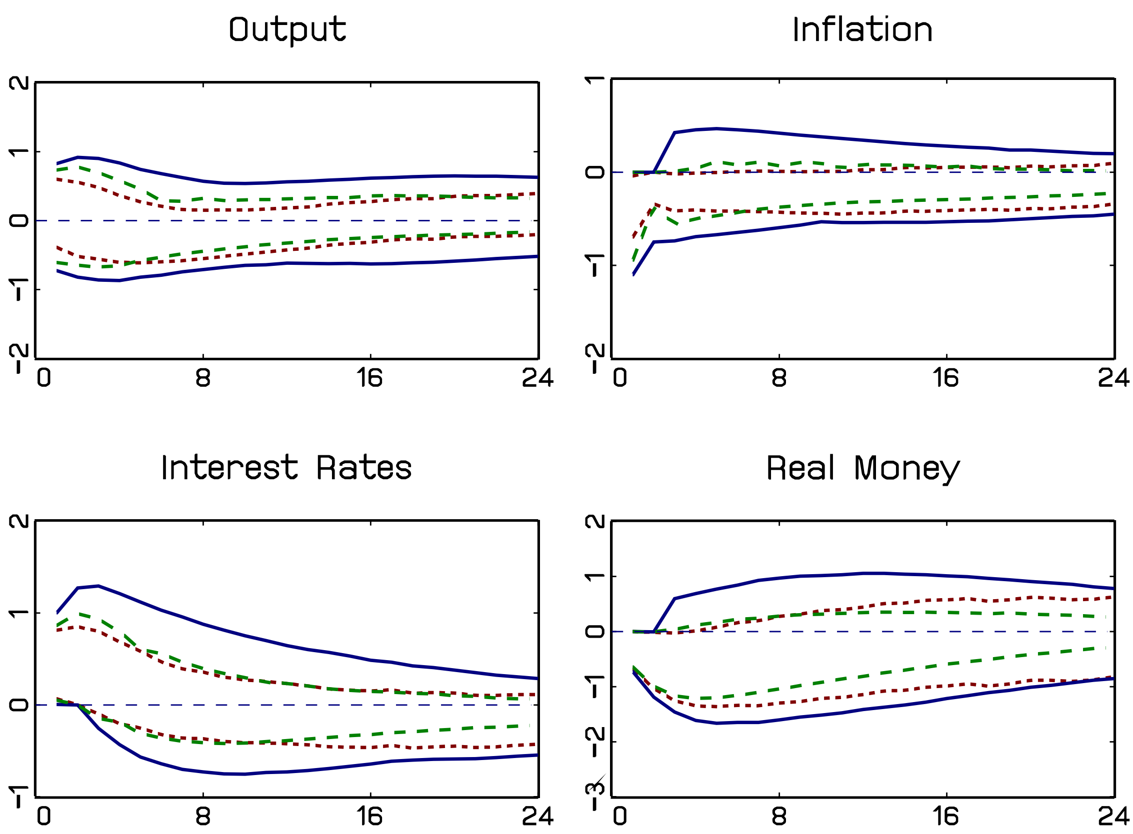
Notes: The coverage probability entries are measured in percent. Length refers to the average length of the confidence interval or credible set across Monte Carlo repetitions.

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



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

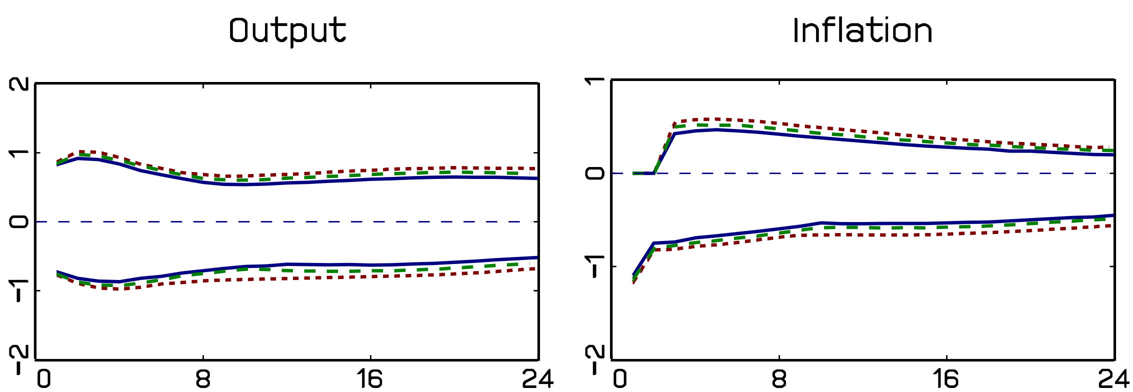
Figure 2: Impulse Responses Based on Pure Sign Restrictions



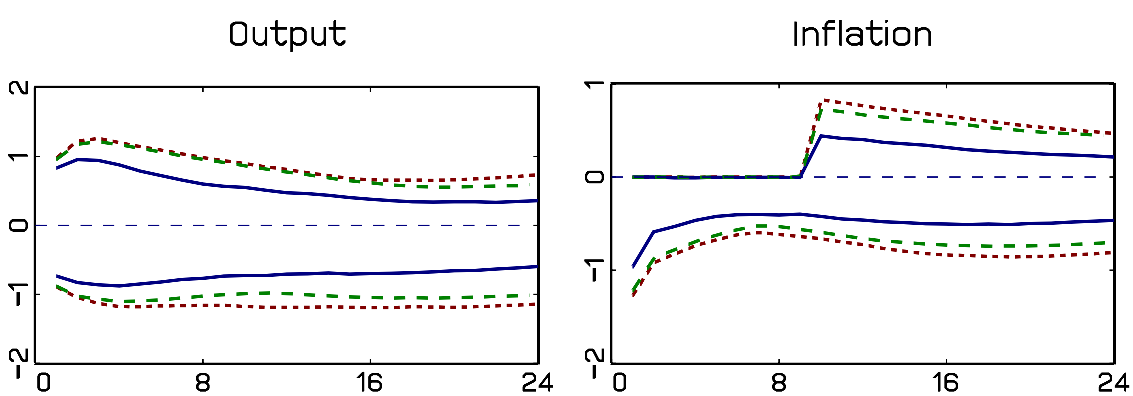
Notes: The figure depicts 90% frequentist confidence sets $CS^{\theta}_{(2)}(\hat{c}_{22})$ (blue, solid); estimated sets $\Theta(\hat{\phi})$ (green, long dashes); and 90% Bayesian credible intervals (red, short dashes).

Figure 3: Comparison of Frequentist Error Bands

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



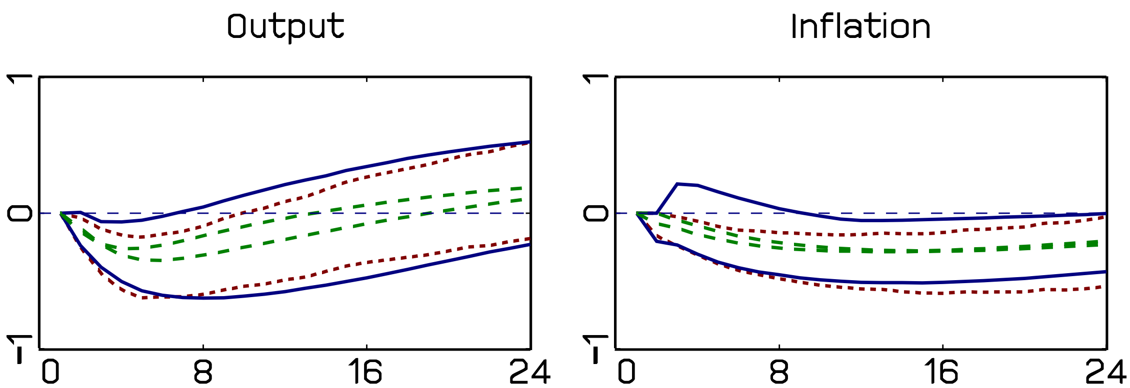
Sign Restrictions Imposed over Horizons $h = 0, 1, \ldots, 8$



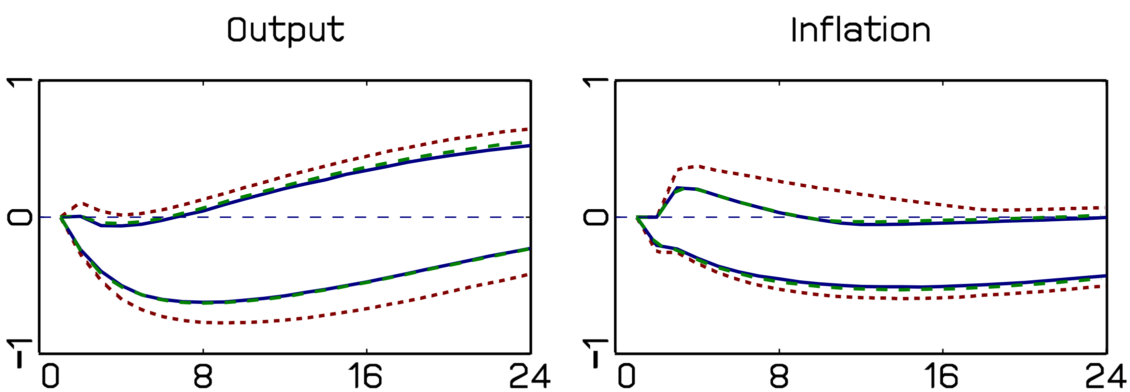
Notes: The figure depicts 90% frequentist confidence sets: $CS^{\theta}_{(1)}$ (red, short dashes), $CS^{\theta}_{(2)}(\hat{c}_{21})$ (green, long dashes), and $CS^{\theta}_{(2)}(\hat{c}_{22})$ (blue, solid).

Figure 4: Combining Zero and Sign Restrictions

Frequentist Error Bands vs. $\Theta(\hat{\phi})$ vs. Bayesian Error Bands



Comparison of Frequentist Error Bands



Notes: The top panel depicts 90% frequentist confidence sets $CS^{\theta}_{(2)}(\hat{c}_{22})$ (blue, solid); estimated sets $\Theta(\hat{\phi})$ (green, long dashes); and 90% Bayesian credible intervals (red, short dashes). The bottom panel depicts 90% frequentist confidence sets: $CS^{\theta}_{(1)}$ (red, short dashes), $CS^{\theta}_{(2)}(\hat{c}_{21})$ (green, long dashes), and $CS^{\theta}_{(2)}(\hat{c}_{22})$ (blue, solid).

Online Technical Appendix

This Online Appendix accompanies the paper "Inference for VARs Identified with Sign Restrictions" by H.R. Moon, F. Schorfheide, E. Granziera, and M. Lee. The Appendix has three sections. In Section A we provide proofs of the lemmas stated in the main text. Section B states and proves lemmas that are needed to prove Theorems 1 and 2 in the main text. Finally, Section C provides analytical derivations for the Monte Carlo experiment presented in Section 5 of the main text.

A Proofs of Lemmas stated in the Main Text

Proof of Lemma 1. To simplify the notation in the proof we omit tildes and write $S_\theta(q)$, $S_R(q)$ instead of $\tilde{S}_\theta(q)$, $\tilde{S}_R(q)$. *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

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

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 (A.1) imply 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 (A.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. Let

$$\mathcal{S}_c^\theta = \left\{ \theta \;\middle|\; Q(\theta; \phi, I) \leq c \right\}$$

For $c = 0$ $\mathcal{S}_c^\theta = \Theta(\phi)$. Suppose that $\tilde{\theta} \in \mathcal{S}_c^\theta$. We will assume that $\tilde{\theta} > 0$ and show by contradiction that $\mathcal{S}_c^\theta$ must have an upper bound. Consider a sequence $a_n > 0$ with $a_n \uparrow \infty$ with the property that $a_n\tilde{\theta} \in \mathcal{S}_c^\theta$ for each n. The unboundedness of $\Theta(\phi)$ guarantees the existence of such a series. Then

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

The definition of $S_\theta(q)$ implies that

$$\|S_\theta(q)\| = \sqrt{tr[(S_\theta \otimes q')S_\phi S_\phi'(S_\theta \otimes q')']} = \sqrt{tr[(S_\theta S_\theta')} = 1.$$

Since ϕ and $\tilde{\theta}$ are fixed, we deduce that as $n \longrightarrow \infty$

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

Thus, $Q(a_n\tilde{\theta}; \phi, I) > c$ eventually, which contradicts the assumption that $a_n\tilde{\theta} \in S_c^\theta$ for all n. 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. □

Proof of Lemma 4: In the proof of Lemma 1 we showed that sets of the form

$$\mathcal{S}_c^\theta = \left\{ \theta \;\middle|\; Q(\theta; \phi, I) \leq c \right\} \tag{A.2}$$

are bounded. The finite-sample confidence sets take the form

$$CS^{\theta}_{(i)} = \left\{ \theta \;\middle|\; Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) \leq c_{(i)} \right\}.$$

Recall that

$$Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) = \min_{\|q\|=1, \mu \geq 0} T \left\| S(q)\hat{\phi} - V(q)\begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{\hat{\Sigma}^{-1}(q)},$$

where

$$\hat{\Sigma}^{-1} = (S(q)\hat{\Lambda}S'(q))^{-1}.$$

Now let

$$\hat{W}_{min}(q) = \frac{1}{\lambda_{max}(\hat{\Lambda})\lambda_{max}(S(q)S(q)')} I.$$

By construction $\hat{W}^*(q) \geq \hat{W}_{min}(q) > 0$ for all q and

$$Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) \geq Q\big(\theta; \hat{\phi}, \hat{W}_{min}(\cdot)\big) = \frac{1}{\lambda_{max}(\hat{\Lambda})\lambda_{max}(S(q)S(q)')} Q\big(\theta; \hat{\phi}, I\big).$$

The statement of the lemma follows from setting $c = c_i \lambda_{max}(\hat{\Lambda})\lambda_{max}(S(q)S(q)')$ in (A.2). $\square$

Proof of Lemma 5: We need to verify that the confidence set constructed by taking unions of the identified sets can be represented according to (42). Let

$$CS^*_U = \bigcup_{\phi \in CS^{\phi}} \Theta(\phi),$$

where

$$CS^{\phi} = \left\{ \phi \in \mathcal{P} \;\middle|\; T\|\hat{\phi} - \phi\|^2_{\hat{\Lambda}^{-1}} \leq c(\chi^2_m) \right\}.$$

(i) Show that $CS^*_U \subseteq CS^{\theta}_U$. Suppose $\theta \in CS^*_U$. Thus, there exists a $\phi \in CS^{\phi}$ such that $\theta \in \Theta(\phi)$. So, $T\|\hat{\phi} - \phi\|^2_{\hat{\Lambda}^{-1}} \leq c(\chi^2_m)$. This implies that there exist a q_* with $\|q_*\| = 1$ and a $\mu_* \geq 0$ such that $\tilde{S}(q_*)\phi = [\theta', \mu_*']'$. In turn,

$$\begin{aligned} Q(\theta; \hat{\phi}, \hat{W}^*) &= \min_{\|q\|=1,\, \mu \geq 0} \left\| \tilde{S}(q)\hat{\phi} - \begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{\hat{W}^*(q)} \\ &\leq \left\| \tilde{S}(q_*)\hat{\phi} - \begin{pmatrix}\theta \\ \mu_*\end{pmatrix} \right\|^2_{\hat{W}^*(q_*)} \\ &= \left\| \tilde{S}(q_*)(\hat{\phi} - \phi) \right\|^2_{\hat{W}^*(q_*)} \\ &= T(\hat{\phi} - \phi)'S'(S\hat{\Lambda}S')^{-1}S(\hat{\phi} - \phi) \\ &\leq T\|\hat{\phi} - \phi\|^2_{\hat{\Lambda}^{-1}} \leq c(\chi^2_m). \end{aligned}$$

The third equality uses the definition of $\hat{W}^*$ in (23) and $S(q) = V(q)\tilde{S}(q)$. The second inequality can be obtained as follows. Factorize $\Lambda = \hat{L}\hat{L}'$ and define $\hat{A} = \hat{L}'S'$ as well as the projection matrix $P_{\hat{A}} = \hat{A}(\hat{A}'\hat{A})^{-1}\hat{A}'$. Then,

$$\begin{aligned} T(\hat{\phi}-\phi)'S'(S\hat{\Lambda}S')^{-1}S(\hat{\phi}-\phi) &= T(\hat{\phi}-\phi)'(\hat{L}')^{-1}\hat{L}'S'(S\hat{L}\hat{L}'S')^{-1}S\hat{L}\hat{L}^{-1}(\hat{\phi}-\phi) \\ &= T(\hat{\phi}-\phi)'(\hat{L}')^{-1}P_{\hat{A}}\hat{L}^{-1}(\hat{\phi}-\phi) \\ &\leq T(\hat{\phi}-\phi)'(\hat{L}')^{-1}\hat{L}^{-1}(\hat{\phi}-\phi) \\ &= T\|\hat{\phi}-\phi\|^2_{\hat{\Lambda}^{-1}} \end{aligned}$$

Thus, we deduce that $\theta \in CS^{\theta}_U$.

(ii) Show that $CS^{\theta}_U \subseteq CS^*_U$. Suppose to the contrary that there exists a $\theta \in CS^{\theta}_U$ and $\theta \notin CS^*_U$. Thus, $\theta \in (CS^*_U)^c = \bigcap_{\phi \in CS^{\phi}} (\Theta(\phi))^c$. Hence, for any ϕ such that $\theta \in \Theta(\phi)$, it has to be the case that $\phi \notin CS^{\phi}$. Define $\psi(\mu) = [\theta', \mu']'$. Thus,

$$\begin{aligned} c(\chi^2_m) &< \left(\min_{\phi \in \mathcal{P},\, \|q\|=1,\, \mu \geq 0} T\|\hat{\phi}-\phi\|^2_{\hat{\Lambda}^{-1}} \quad \text{s.t.} \quad 0 = \tilde{S}(q)\phi - \psi(\mu) \right) \\ &= \min_{\|q\|=1,\, \mu\geq 0} \left(\min_{\phi \in \mathcal{P}} T\|\hat{\phi}-\phi\|^2_{\hat{\Lambda}^{-1}} \quad \text{s.t.} \quad 0 = \tilde{S}(q)\phi - \psi(\mu) \right) \end{aligned} \tag{A.3}$$

Since the matrix $V(b)$ eliminates rows of zeros from $\tilde{S}(q)$, we can rewrite the constraint in (A.3) as

$$0 = V(q)\tilde{S}(q) - V(q)\psi(\mu) = S(q) - V(q)\psi(\mu).$$

Conditional on q and μ the Lagrangian associated with the minimization over ϕ is given by

$$\mathcal{L} = T\|\hat{\phi}-\phi\|^2_{\hat{\Lambda}^{-1}} - \lambda'(S(q)\phi - V(q)\psi(\mu)).$$

The first-order conditions are

$$0 = T\hat{\Lambda}^{-1}(\phi_* - \hat{\phi}) - S'\lambda_*, \quad 0 = S(\phi_* - \hat{\phi}) + S\hat{\phi} - V\psi(\mu).$$

Solving for ϕ_* yields

$$\phi_* = \hat{\phi} + \hat{\Lambda}S'(S\hat{\Lambda}S')^{-1}(S\hat{\phi} - V\psi).$$

Thus, using the definition $\hat{\Sigma}(q) = S(q)\hat{\Lambda}S'(q)$ we can express the constrained minimization in (A.3) as

$$\min_{\|q\|=1,\, \mu\geq 0} T \left\| S(q)\hat{\phi} - V(q)\begin{pmatrix}\theta \\ \mu\end{pmatrix} \right\|^2_{\hat{\Sigma}^{-1}(q)}. \tag{A.4}$$

Thus, $\min_{\|q\|=1,\, \mu\geq 0}\ Q(\theta; \hat{\phi}, \hat{W}^*) > c(\chi^2_m)$, which contradicts the initial assumption that $\theta \in CS^{\theta}_U$. □

B Technical Lemmas Used in the Proofs of the Main Results

Lemma B 1 *Suppose that Assumptions 1 to 3 are satisfied. The sample estimate $\hat{\Lambda}$ that enters $\hat{W}^*(\cdot)$ in the bounding function $\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big)$, defined in (28), can be replaced by the population covariance matrix Λ:*

$$\left|\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big) - \bar{Q}\big(q;\hat{\phi},W^*(\cdot)\big)\right| = o_p(1)$$

uniformly in (ϕ,q).

Proof of Lemma B 1: Notice that the penalty term in (24) cancels and thus is omitted from the subsequent calculations. Let

$$\begin{aligned} v(\hat{\Lambda}) &= \arg\min_{v\geq 0} \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v\right\|^2_{\hat{\Sigma}^{-1}(q)} \\ v(\Lambda) &= \arg\min_{v\geq 0} \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v\right\|^2_{\Sigma^{-1}(q)}. \end{aligned}$$

First we show that uniformly in (ϕ,q)

$$\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big) - \bar{Q}\big(q;\hat{\phi},W^*(\cdot)\big) \leq o_p(1). \tag{B.1}$$

To do so, consider the following inequalities: Notice that

$$\begin{aligned} &\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big) - \bar{Q}\big(q;\hat{\phi},W^*(\cdot)\big) \\ &= \min_{v\geq 0} \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v\right\|^2_{\hat{\Sigma}^{-1}(q)} - \min_{v\geq 0} \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v\right\|^2_{\Sigma^{-1}(q)} \\ &= \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\hat{\Lambda})\right\|^2_{\hat{\Sigma}^{-1}(q)} - \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\right\|^2_{\Sigma^{-1}(q)} \\ &\leq \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\right\|^2_{\hat{\Sigma}^{-1}(q)} - \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\right\|^2_{\Sigma^{-1}(q)} \\ &= \Big[S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\Big]' \Sigma^{-1/2}(q) \\ &\quad \times \Big[\Sigma^{1/2}(q)\hat{\Sigma}^{-1}(q)\Sigma^{1/2}(q) - I_{l(q)}\Big]\Sigma^{-1/2}(q)\Big[S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\Big] \\ &\leq \left\| S(q)\sqrt{T}(\hat{\phi}-\phi) - V(q)M_v v(\Lambda)\right\|^2_{\Sigma^{-1}(q)} \left\|\Sigma^{1/2}(q)\hat{\Sigma}^{-1}(q)\Sigma^{1/2}(q) - I_{l(q)}\right\| \\ &= I \times II, \text{ say.} \end{aligned}$$

The first inequality is obtained by replacing the minimizer $\nu(\hat{\Lambda})$ with the inferior value $\nu(\Lambda)$. The third equality follows from writing out the two norms and rearranging the weight matrix differential $\hat{\Sigma}^{-1} - \Sigma^{-1}$.

We now bound the terms I and II. Recall that $\Sigma = S\Lambda S'$ and $\Lambda = LL'$. For term I we obtain

$$\begin{aligned} I &= \min_{v \geq 0} \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - V(q)M_v v \right\|^2_{\Sigma^{-1}(q)} \\ &\leq \left\| S(q)LL^{-1}\sqrt{T}(\hat{\phi} - \phi) \right\|^2_{\Sigma^{-1}(q)} \\ &= [L^{-1}\sqrt{T}(\hat{\phi} - \phi)]' P_{L'S'(q)} [L^{-1}\sqrt{T}(\hat{\phi} - \phi)] \\ &\leq \left\| L^{-1}\sqrt{T}(\hat{\phi} - \phi) \right\|^2 \lambda_{\max}(P_{L'S'(q)}) \\ &= \left\| L^{-1}\sqrt{T}(\hat{\phi} - \phi) \right\|^2 \\ &= O_p(1) \end{aligned}$$

uniformly in (ϕ, q). Here $P_{L'S'(q)}$ is the matrix that projects onto the column space of $L'S'(q)$. The second term can be bounded as follows:

$$\begin{aligned} II &= \left\| \Sigma^{1/2}(q)(\hat{\Sigma}^{-1}(q) - \Sigma^{-1}(q))\Sigma^{1/2}(q) \right\| \\ &= \left\| \Sigma^{-1/2}(q)(\Sigma(q) - \hat{\Sigma}(q))\hat{\Sigma}^{-1}(q)\Sigma^{1/2}(q) \right\| \\ &= \left\| \Sigma^{-1/2}(q)S(q)(\Lambda - \hat{\Lambda})S(q)'\hat{\Sigma}^{-1/2}(q)\hat{\Sigma}^{-1/2}(q)\Sigma^{1/2}(q) \right\| \\ &= \left\| (\Sigma^{-1/2}(q)S(q)L)(L'(\hat{L}')^{-1} - L^{-1}\hat{L})\hat{L}'S(q)'\hat{\Sigma}^{-1/2}(q)\hat{\Sigma}^{-1/2}(q)\Sigma^{1/2}(q) \right\| \\ &\leq \left\| \Sigma^{-1/2}(q)S(q)L \right\| \left\| L'(\hat{L}')^{-1} - L^{-1}\hat{L} \right\| \left\| \hat{L}'S(q)'\hat{\Sigma}^{-1/2}(q) \right\| \left\| \hat{\Sigma}^{-1/2}(q)\Sigma^{1/2}(q) \right\| \\ &= II_1 \times II_2 \times II_3 \times II_4, \text{ say.} \end{aligned}$$

Bounds for the four terms can be obtained as follows. First,

$$II_1^2 = \left\| \Sigma^{-1/2}(q)\, S(q)\, L \right\|^2 = \mathrm{tr}\left(L'S(q)' \left(S(q)\, LL'S(q)' \right)^{-1} S(q)\, L \right) = l(q) \leq \tilde{k} + \tilde{r}_2. \tag{B.2}$$

Similarly, the second term can be bounded by

$$II_2^2 \leq \left\| \hat{L}'S(q)'\, \hat{\Sigma}^{-1/2}(q) \right\|^2 = l(q) \leq \tilde{k} + \tilde{r}_2. \tag{B.3}$$

Third, since $\hat{\Lambda} - \Lambda(\phi) = o_p(1)$ uniformly in ϕ and $\Lambda(\phi) > \Lambda_{\min} > 0$, for some positive definite matrix $\Lambda_{\min}$, we have

$$II_3 = \left\| L'(\hat{L}')^{-1} - L^{-1}\hat{L} \right\| = o_p(1) \tag{B.4}$$

uniformly in (ϕ, q) . Finally, we obtain

$$\begin{aligned} II_4^2 &= \left\|\hat{\Sigma}^{-1/2}(q)\left(S(q)L\right)\left(L'S(q)'\left(S(q)LL'S(q)'\right)^{-1}\right)\Sigma^{1/2}(q)\right\|^2 \\ &= \left\|\hat{\Sigma}^{-1/2}(q)\left(S(q)L\right)\left(L'S(q)'\Sigma^{-1}(q)\right)\Sigma^{1/2}(q)\right\|^2 \\ &= \left\|\hat{\Sigma}^{-1/2}(q)\left(S(q)L\right)\left(L'S(q)'\right)\Sigma^{-1/2}(q)\right\|^2 \\ &\leq \left\|\hat{\Sigma}^{-1/2}(q)(S(q)\hat{L})\right\|^2\left\|(\hat{L}^{-1}L)\right\|^2\left\|(L'S(q)')\Sigma^{-1/2}(q)\right\|^2 \\ &\leq (\tilde{k}+\tilde{r}_2)^2\left\|(\hat{L}^{-1}L)\right\|^2 \text{ by } (B.2) \text{ and } (B.3) \\ &\leq (\tilde{k}+\tilde{r}_2)^2 O_p(1) \end{aligned} \tag{B.5}$$

where $O_p(1)$ is uniform in (ϕ, q) . From the bounds $(B.2), (B.3), (B.4), (B.5)$, we have

$$II = o_p(1)$$

uniformly in (ϕ, q) . The desired result in (B.1) is obtained by combining I and II. In a similar manner it can be shown that

$$\bar{Q}\big(q;\hat{\phi},W^*(\cdot)\big) - \bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big) \leq o_p(1). \quad \square$$

The sample estimate $\hat{\Lambda}$ that enters $\hat{W}^*(\cdot)$ in the bounding function $\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big)$ can be replaced by the population covariance matrix Λ:

$$\left|\bar{Q}\big(q;\hat{\phi},\hat{W}^*(\cdot)\big) - \bar{Q}\big(q;\hat{\phi},W^*(\cdot)\big)\right| = o_p(1)$$

uniformly in (ϕ, q).

Lemma B 2 *Suppose that Assumptions 1 to 3 are satisfied. The sample estimate $\hat{\Lambda}$ that enters $\hat{W}^*(\cdot)$ in the objective function $G\big(\theta,q;\hat{\phi},\hat{W}^*(\cdot)\big)$, defined in (31), can be replaced by the population covariance matrix Λ:*

$$\left|G\big(\theta,q;\hat{\phi},\hat{W}^*(\cdot)\big) - G\big(\theta,q;\hat{\phi},W^*(\cdot)\big)\right| = o_p(1)$$

uniformly in (ϕ,θ,q) for $\phi\in\mathcal{P}$, $\theta\in\Theta(\phi)$, and $q\in\mathbb{Q}(\theta,q)$.

Proof of Lemma B 2: Notice that the penalty term in (24) cancels and thus is omitted from the subsequent calculations. By definition

$$\begin{aligned} G(\theta,q;\hat{\phi},\hat{W}^*(\cdot)) &= \min_{v\geq-\sqrt{T}\hat{D}_R^{-1/2}\mu(q,\phi)} \left\|\hat{D}^{-1/2}S(q)\sqrt{T}(\hat{\phi}-\phi) - M_v v\right\|^2_{\hat{\Omega}^{-1}(q)} \\ &= \min_{v\geq-\sqrt{T}\mu(q,\phi)} \left\|S(q)\sqrt{T}(\hat{\phi}-\phi) - M_v v\right\|^2_{\hat{\Sigma}^{-1}(q)}. \end{aligned}$$

Moreover, define

$$\begin{aligned} v(\hat{\Lambda}) &= \arg \min_{v \geq -\sqrt{T}\mu(q,\phi)} \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v \right\|^2_{\hat{\Sigma}^{-1}(q)} \\ v(\Lambda) &= \arg \min_{v \geq -\sqrt{T}\mu(q,\phi)} \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v \right\|^2_{\Sigma^{-1}(q)}. \end{aligned}$$

By using similar arguments as in the proof of Lemma B 1, we have

$$\begin{aligned} & G(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)) - G(\theta, q; \hat{\phi}, W^*(\cdot)) \\ \leq\ & \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v(\Lambda) \right\|^2_{\hat{\Sigma}^{-1}(q)} - \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v(\Lambda) \right\|^2_{\Sigma^{-1}(q)} \\ =\ & \left\{ S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v(\Lambda) \right\}' \Sigma^{-1/2}(q) \\ & \times \left\{ \Sigma^{1/2}(q)\hat{\Sigma}^{-1}(q)\Sigma^{1/2}(q) - I_{l(q)} \right\} \Sigma^{-1/2}(q) \left\{ S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v(\Lambda) \right\} \\ \leq\ & \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v(\Lambda) \right\|^2_{\Sigma^{-1}(q)} \left\| \Sigma^{1/2}(q)\hat{\Sigma}^{-1}(q)\Sigma^{1/2}(q) - I_{l(q)} \right\| \\ =\ & I \times II, \text{ say.} \end{aligned}$$

Notice that

$$\begin{aligned} I &= \min_{v \geq -\sqrt{T}\mu(q,\phi)} \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) - M_v v \right\|^2_{\Sigma^{-1}(q)} \\ &\leq \left\| S(q)\sqrt{T}(\hat{\phi} - \phi) \right\|^2_{\Sigma^{-1}(q)} \text{ since } \mu(q,\phi) \geq 0 \\ &\leq O_p(1) \end{aligned}$$

uniformly in (ϕ, θ, q) for $\phi \in \mathcal{P}$, $\theta \in \Theta(\phi)$, and $q \in \mathbb{Q}(\theta, q)$. The first inequality follows from $\mu(q,\phi) \geq 0$. The second inequality can be verified using the same steps as in the proof of Lemma B 1. Similarly, we can follow the proof of Lemma B 1 to establish that $II = o_p(1)$ uniformly in (ϕ, θ, q) for $\phi \in \mathcal{P}$, $\theta \in \Theta(\phi)$, and $q \in \mathbb{Q}(\theta, q)$. This leads to the required result. The bound

$$G(\theta, q; \hat{\phi}, W^*(\cdot)) - G(\theta, q; \hat{\phi}, \hat{W}^*(\cdot)) \leq o_p(1)$$

can be established in a similar fashion and the statement of the lemma follows. □

Lemma B 3 *Suppose Assumption 3 is satisfied. Denote* $\bar{S}^{M,0}(q) = \bar{S}(q)$ *and* $\bar{S}^{M,k}(q) = M^{S,k}\bar{S}^{M,k-1}(q)$ *for* $k = 1, \ldots, 4$. *Finally, let* $\tilde{S}(q) = \bar{S}^{M,4}(q)M^{S,5}$. *Suppose that* q_T *is a converging sequence. Then, there exists a subsequence* $q_{T'}$ *and index sets* $\mathcal{J}_k$ *that are constant over* T' *such that for all* $k = 0, 1, ..., 5$, *(i)* $\left\| \bar{S}_j^{M,k}(q_{T'}) \right\| > 0$ *if and only if* $j \in \mathcal{J}_k$ *and (ii)*

$$\left[\frac{\bar{S}_j^{M,k}(q_{T'})}{\left\| \bar{S}_j^{M,k}(q_{T'}) \right\|} \right]_{j \in \mathcal{J}_k} \longrightarrow \left[\bar{S}_{j,*}^{M,k} \right]_{j \in \mathcal{J}_k},$$

where $\left[\bar{S}^{M,k}_{j,*}\right]_{j\in\mathcal{J}_k}$ *has full row rank.*

Proof of Lemma B 3: Part (i): Since the singularity of $\bar{S}^{M,k}(q_T)$ is caused only by zero rows, we can choose a subsequence of $\{T\}$ such that the rank of $\bar{S}^{M,k}(q_T)$ is the same along the subsequence. Then, we can choose a further subsequence such that the index of the nonzero rows are the same. We set this subsubsequence as $\{T''\}$, and the index set contains nonzero rows along this sequence as $\mathcal{J}_k$.

Part (ii): Along the subsequence chosen in Part (i), $\frac{\bar{S}^{M,k}_j(q_T)}{\|\bar{S}^{M,k}_j(q_T)\|}$ is well defined for all $j\in\mathcal{J}_k$, $k=0,...,5$. Since $\left\{\frac{\bar{S}^{M,k}_j(q_{T''})}{\|\bar{S}^{M,k}_j(q_{T''})\|}\right\}\subset\mathbb{S}^m$, the unit sphere in $\mathbb{R}^m$, which is a compact set, we can choose a further subsequence, denoted by $\{T'\}$ such that

$$\frac{\bar{S}^{M,k}_j(q_{T'})}{\left\|\bar{S}^{M,k}_j(q_{T'})\right\|}\longrightarrow\bar{S}^{M,k}_{j,*}.$$

For the required result in Part (ii), we show that the row vectors $\bar{S}^{M,k}_{j,*}$ over $j\in\mathcal{J}_k$ are linearly independent for $k=0,...,5$. In what follows we show this required result for the cases $k=0,1,2$. The cases of $k=3,4$ follow immediately from the case $k=2$ because $\mathcal{J}_2=\mathcal{J}_3\supset\mathcal{J}_4$ and $\left\{\bar{S}^{M,2}_{j,*}:j\in\mathcal{J}_2\right\}=\left\{\bar{S}^{M,3}_{j,*}:j\in\mathcal{J}_3\right\}\supset\left\{\bar{S}^{M,4}_{j,*}:j\in\mathcal{J}_4\right\}$ by the definition of the $M^{S,3}$ and $M^{S,4}$. The case of $k=5$ follows because $M^{S,5}$ deletes only the zero columns.

Before we start the proof, notice that since $M^{S,1}$ is a full rank diagonal matrix, $M^{S,2}$ is the quasi-lower triagular structure of $M^{S,2}$ with full rank, and both $M^{S,1}$ and $M^{S,2}$ do not depend on q, we have $\mathcal{J}_0=\mathcal{J}_1=\mathcal{J}_2$.

Case $k=0$*:* First notice that the row vectors in $\left\{\bar{S}_j(q_T):j\in\mathcal{J}_0\right\}$ are orthogonal to each other and so are the row vectors in $\left\{\bar{S}^{M,0}_{j,*}:j\in\mathcal{J}_0\right\}$. Therefore, the row vectors in $\left\{\bar{S}^{M,0}_{j,*}:j\in\mathcal{J}_0\right\}$ are linearly independent.

Case $k=1$*:* Notice that $M^{S,1}$ is a full rank diagonal matrix that does not depend on q_T. Therefore, $\begin{bmatrix}\vdots\\ \bar{S}^{M,1}_{j,*}\\ \vdots\end{bmatrix}=\begin{bmatrix}\vdots\\ \left[M^{S,1}\bar{S}^{M,0}_*\right]_j\\ \vdots\end{bmatrix}$ has full row rank, and we have the required result for the case $k=1$.

Case $k=2$*:* Similarly, $M^{S,1}$ is a full rank matrix that does not depend on q_T, it follows that

$\begin{bmatrix} \vdots \\ \bar{S}^{M,2}_{j,*} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \left[M^{S,2}\bar{S}^{M,1}_{*}\right]_j \\ \vdots \end{bmatrix}$ has full row rank, and we have the required result for the case $k = 2$. □

Lemma B 4 *Suppose Assumption 3 is satisfied. For a converging sequence $\{q_T\}$ such that $V(q_T)$ is constant and the rank of $S(q_T)$ equals to l for all T, there exists a subsequence $\{T'\} \subseteq \{T\}$ such that the matrix defined as*

$$\lim_T \begin{bmatrix} \frac{S_1(q_{T'})}{\|S_1(q_{T'})\|} \\ \vdots \\ \frac{S_l(q_{T'})}{\|S_l(q_{T'})\|} \end{bmatrix}$$

has full row rank l.

Proof of Lemma B 4: Along the sequence q_T, $V(q_T)$ is a constant matrix such that each row of $V(q_T)$ has only one nonzero element, which is one. Choose the subsequence $\{T'\}$ in Lemma B 3. Then,

$$\left[\frac{S_j(q_{T'})}{\|S_j(q_{T'})\|}\right]_{j=1,\ldots,l} = \left[\frac{\tilde{S}_j(q_{T'})}{\left\|\tilde{S}_j(q_{T'})\right\|}\right]_{j\in\mathcal{J}_5},$$

where $\mathcal{J}_5$ is defined in Lemma B 3. The required result follows by Lemma B 3. □

Lemma B 5 *Suppose Assumptions 1 to 3 are satisfied. For a converging sequence $\{\phi_T, \theta_T, q_T\}$ that satisfies the rank condition $l(q_T) = l$ for all T, there exists a subsequence $\{T''\} \subseteq \{T\}$*

$$\begin{aligned} [D^{-1/2}(q_{T''})S(q_{T''})L(\phi_{T''})]' &\longrightarrow A \\ \Omega(q_{T''}) &\longrightarrow A'A, \end{aligned}$$

where A is a full rank matrix.

Proof of Lemma B 5: For notational convenience we denote $\Lambda(\phi_T) = \Lambda_T$ and $L(\phi_T) = L_T$. Consider the spectral decomposition $\Lambda_T = U_T diag(\lambda_{1,T}, \ldots, \lambda_{m,T})U_T'$. Since $\Lambda_T > \Lambda_{min} > 0$, the smallest eigenvalue $\lambda_{min,T} > \delta > 0$. Then,

$$\Sigma_T = S(q_T)U_T diag\left(\lambda_{1,T}, \ldots, \lambda_{m,T}\right) U_T' S'(q_T)$$

and the diagonal elements of Σ_T are given by

$$D_{jj}(q_T) = \lambda_{j,T} S_j(q_T)(U_T U_T') S_j'(q_T) = \lambda_{j,T} \|S_j(q_T)\|^2 > 0 \quad \forall j, T.$$

Now we can express

$$D^{-1/2}(q_T) S(q_T) L_T = diag\left(\lambda_{1,T}^{-1/2}, \dots, \lambda_{l,T}^{-1/2}\right) \begin{bmatrix} \frac{S_1(q_T)}{\|S_1(q_T)\|} \\ \vdots \\ \frac{S_l(q_T)}{\|S_l(q_T)\|} \end{bmatrix}.$$

Since $\lambda_{j,T}^{-1/2} > 0$ for all j and $S_j(q_T)/\|S_j(q_T)\|$ exists on the unit hypersphere there exists a subsequence $\{T'\}$ such that

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

According to Lemma B 4, we can construct a further subsequence $\{T''\}$ along which the matrix

$$\begin{bmatrix} \frac{S_1(q_{T''})}{\|S_1(q_{T''})\|} \\ \vdots \\ \frac{S_l(q_{T''})}{\|S_l(q_{T''})\|} \end{bmatrix}$$

has full row rank. Since the limit of the matrix $diag\left(\lambda_{1,T}^{-1/2}, \dots, \lambda_{l,T}^{-1/2}\right)$ is full rank, the limit matrices A and $A'A$ are also full rank. $\square$

Lemma B 6 *If $CS^{\theta,q}_{(2)}$ in (32) is a valid $1-\tau$ confidence set, then $CS^{\theta}_{(2)}$ in (33) is a valid $1-\tau$ confidence set.*

Proof of Lemma B 6: The lemma follows since $(\theta, q) \in CS^{\theta,q}_{(2)}$ if and only if

$$M_\theta \theta \geq 0 \quad \text{and} \quad G(\theta, q; \hat{\phi}, \hat{W}^*) \leq c_{(2)}(q),$$

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

$$M_\theta \theta \geq 0 \quad \text{and} \quad \min_{\tilde{q} = \|1\|} \left[G(\theta, \tilde{q}; \hat{\phi}, \hat{W}^*) - c_{(2)}(\tilde{q}) \right] \leq 0$$

and therefore $\theta \in CS^{\theta}_{(2)}$. In turn,

$$\begin{aligned} 1 - \tau &\leq \liminf_T \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} \inf_{q \in \mathbb{Q}(\theta,\phi)} P_\phi \left\{ (\theta, q) \in CS^{\theta,q}_{(2)} \right\} \\ &\leq \liminf_T \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_\phi \left\{ \theta \in CS^{\theta}_{(2)} \right\}. \quad \square \end{aligned}$$

Lemma B 7 *Suppose Assumptions 1 to 3 are satisfied. Then under Case (i) considered in the proof of Theorem 2, the contribution of the moment conditions deemed to be nonbinding to the objective function $G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big)$ is asymptotically negligible:*

$$G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) = G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) + o_p(1),$$

Proof of Lemma B 7: Recall the definition

$$\begin{aligned} G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) &= \min_{\nu_1 \geq -\sqrt{T}\mu_{1,T}} \left\| S_{1,T}\sqrt{T}(\hat{\phi} - \phi_T) - M_{\nu_1}\nu_1 \right\|^2_{\Sigma^{-1}_{11,T}} \\ &= \min_{\nu_1 \geq -\sqrt{T} D^{-1/2}_{R,1,T}\mu_{1,T}} \left\| D^{-1/2}_{1,T} S_{1,T} L_T L_T^{-1} \sqrt{T}(\hat{\phi} - \phi_T) - M_{\nu_1}\nu_1 \right\|^2_{\Omega^{-1}_{11,T}}, \end{aligned}$$

where $D_{1,T} = diag(D_{\theta,T},\, D_{R,1,T})$. Now define $h_{1,T} = \sqrt{T} D^{-1/2}_{R,1,T}\mu_{1,T}$ and $\zeta_T = L_T^{-1}\sqrt{T}(\hat{\phi} - \phi_T)$. Thus, we can write

$$G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) = \min_{\nu_1 \geq -h_{1,T}} \left\| D^{-1/2}_{1,T} S_{1,T} L_T \zeta_T - M_{\nu_1}\nu_1 \right\|^2_{\Omega^{-1}_{11,T}}$$

and define

$$\nu^*_{1,T} = \text{argmin}_{\nu_1 \geq -h_{1,T}} \left\| D^{-1/2}_{1,T} S_{1,T} L_T \zeta_T - M_{\nu_1}\nu_1 \right\|^2_{\Omega^{-1}_{11,T}}.$$

Using the definitions $h_{2,T} = \sqrt{T} D^{-1/2}_{R,2,T}\mu_{2,T}$ and $h_T = [h_{1,T},\, h_{2,T}]'$, we can write

$$G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) = \min_{\nu \geq -h_T} \left\| D_T^{-1/2} S_T L_T \zeta_T - M_\nu \nu \right\|^2_{\Omega_T^{-1}}.$$

The $G(\cdot)$ function can be decomposed as follows:

$$\begin{aligned} G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) &= G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \\ &\quad + \min_{\nu_2 \geq -h_{2,T}} \Big\| \big(D^{-1/2}_{2,T} S_{2,T} L_T - \Omega_{21,T}\Omega^{-1}_{11,T} D^{-1/2}_{1,T} S_{1,T} L_T\big)\zeta_T \\ &\quad - \big(\nu_2 - \Omega_{21,T}\Omega^{-1}_{11,T} M_{\nu_1}\nu^*_{1,T}\big) \Big\|^2_{\Omega^{-1}_{2.11,T}}, \end{aligned}$$

where $\Omega_{2.11,T} = \Omega_{22,T} - \Omega_{21,T}\Omega^{-1}_{11,T}\Omega_{12,T}$. Now denote

$$\zeta^*_T = \big(D^{-1/2}_{2,T} S_{2,T} L_T - \Omega_{21,T}\Omega^{-1}_{11,T} D^{-1/2}_{1,T} S_{1,T} L_T\big)\zeta_T + \Omega_{21,T}\Omega^{-1}_{11,T} M_{\nu_1}\nu^*_{1,T}.$$

For any $\eta > 0$ it follows that

$$P\left\{ \left| G\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) - G_1\big(\theta_T, q_T; \hat{\phi}, W^*(\cdot)\big) \right| \leq \eta \right\} \geq P\{\zeta^*_T \geq -h_{2,T}\}.$$

In the remainder of the proof we will show that $\liminf_T \; P\{\zeta_T^* \geq -h_{2,T}\} \geq 1-\epsilon$ for any $\epsilon > 0$, which implies the desired result.

We proceed by showing that $\nu_{1,T}^*$ is stochastically bounded. Notice that

$$\left\|D_{1,T}^{-1/2} S_{1,T} L_T \zeta_T\right\| \leq \|D_{1,T}^{-1/2} S_{1,T} L_T\| \cdot \left\|\zeta_T\right\| = O(1)O_p(1) = O_p(1).$$

Since our subsequence is constructed such that $\Omega_T \longrightarrow A'A > 0$ and Ω_T is a sequence of correlation matrices, we deduce that

$$\nu_{1,T}^* = O_p(1). \tag{B.6}$$

In turn, $\Omega_{21,T}\Omega_{11,T}^{-1} M_{\nu_1} \nu_{1,T}^* = O_p(1)$. Now consider

$$\begin{aligned} &\left\|\left(D_{2,T}^{-1/2} S_{2,T} L_T - \Omega_{21,T}\Omega_{11,T}^{-1} D_{1,T}^{-1/2} S_{1,T} L_T\right)\zeta_T\right\| \qquad \text{(B.7)} \\ \leq\; & \left(\left\|D_{2,T}^{-1/2} S_{2,T} L_T\right\| + \|\Omega_{21,T}\| \cdot \left\|\Omega_{11,T}^{-1}\right\| \cdot \left\|D_{1,T}^{-1/2} S_{1,T} L_T\right\|\right) \|\zeta_T\| \\ =\; & O(1)O_p(1) = O_p(1) \end{aligned}$$

Combining the $O_p(1)$ results in B.6 and (B.7), we can deduce that

$$\zeta_T^* = O_p(1).$$

Thus, for any $\epsilon > 0$, there exists a constant $M > 0$ such that

$$\liminf_T \; P\big\{\zeta_T^* \in [-M,\, M]^{r_{22}}\big\} \geq 1-\epsilon.$$

Here $[-M,\, M]^{r_{22}}$ denotes the Cartesian power of the interval $[-M,\, M]$. Since $h_{2,T} \longrightarrow \infty$ there exists a T^* such that

$$P\{\zeta_T^* \geq -h_{2,T}\} \geq P\left\{\zeta_T^* \in [-M,\, M]^{r_{22}}\right\}, \quad \text{for all} \quad T > T^*,$$

which completes the proof. □

Lemma B 8 *Suppose Assumptions 1 to 3 are satisfied. Consider the Case (i) in the proof of Theorem 2. Along the (ϕ_T, θ_T, q_T) sequence, the critical value based on the estimated number of potentially binding moment conditions is more conservative in the following sense:*

$$c^{(1)}_{k(q_T)+\hat{r}_{21}(q_T)} \geq c^{(1)}_{k+r_{21}}.$$

in probability approaching one.

Proof of Lemma B 8: From Section 4.3 recall the definition

$$\hat{r}_{21}(q_T) = \sum_{j=1}^{r_2} \mathcal{I}\left\{\hat{\xi}_{j,T}(q_T) \leq \kappa_T\right\}, \quad \text{where} \quad \hat{\xi}_{j,T}(q_T) = \hat{D}_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T, \hat{\phi}).$$

Then,

$$\begin{aligned}
\kappa_T^{-1}\hat{\xi}_{j,T}(q_T) &= \kappa_T^{-1} D_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T, \phi_T) \\
&\quad + \left[\hat{D}_{jj,R}^{-1/2}(q_T) D_{jj,R}^{1/2}(q_T) - 1\right] \kappa_T^{-1} D_{jj,R}^{-1/2}(q_T)\sqrt{T}\mu_j(q_T, \phi_T) \\
&\quad + \kappa_T^{-1}\hat{D}_{jj,R}^{-1/2}(q_T)\sqrt{T}[\mu_j(q_T, \hat{\phi}) - \mu_j(q_T, \phi_T)] \\
&= I + II \times I + III, \text{ say.}
\end{aligned}$$

Term I: by definition we obtain

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

Term II: can be bounded as follows:

$$\begin{aligned}
\left|\hat{D}_{jj,R}^{-1}(q_T) D_{jj,R}(q_T) - 1\right| &= \frac{\left|D_{jj,R}(q_T) - \hat{D}_{jj,R}(q_T)\right|}{\hat{D}_{jj,R}(q_T)} \\
&\leq \frac{\left\|\hat{\Lambda} - \Lambda(\phi_T)\right\|}{\lambda_{\min}(\hat{\Lambda})} = o_p(1).
\end{aligned}$$

The $o_p(1)$ statement follows because $\|\hat{\Lambda} - \Lambda(\phi_T)\| \xrightarrow{p} 0$ and $\Lambda(\phi_T) > 0$.

Term III: Let $S_{j,R}(q_T)$ be the j^{th} row of $S_R(q_T)$. Then, since

$$\begin{aligned}
&\left|\hat{D}_{jj,R}^{-1/2}(q_T)\sqrt{T}(\mu_j(q_T, \hat{\phi}) - \mu_j(q_T, \phi_T))\right| \\
&= \left|\frac{S_{j,R}(q_T)\hat{L}}{\left\|S_{j,R}(q_T)\hat{L}\right\|}\hat{L}^{-1}\sqrt{T}(\hat{\phi} - \phi_T)\right| \leq \left\|\hat{L}^{-1}\sqrt{T}(\hat{\phi} - \phi_T)\right\| = O_p(1),
\end{aligned}$$

we have

$$III = O_p(\kappa_T^{-1}) = o_p(1).$$

Combining the results, we deduce that if $\pi_j < \infty$,

$$\kappa_T^{-1}\hat{\xi}_{j,T}(q_T) = \kappa_T^{-1}\hat{D}_{jj,R}^{-1/2}(q_T)\mu_j(q_T, \hat{\phi}) \longrightarrow_p \pi_j. \tag{B.8}$$

In particular, if $\pi_j = 0$, then

$$\mathcal{I}\left\{\hat{\xi}_{j,T}(q_T) \leq \kappa_T\right\} \xrightarrow{p} 1,$$

which leads to the desired result:

$$\text{plim}_T\ \hat{r}_{21}(q_T) = \text{plim}_T \sum_{j=1}^{r_2} \mathcal{I}\left\{\hat{\xi}_{j,T}(q_T) \le \kappa_T\right\} \ge \sum_{j=1}^{r_2} 1\left\{\pi_j = 0\right\} = r_{21}. \quad \square$$

Lemma B 9 *Suppose Assumptions 1 to 3 are satisfied. Consider the Case (i) in the proof of Theorem 2. Along the $\{T\}$ sequence, $\hat{c}^*_{(22)}(q_T) \xrightarrow{p} c^*_{(22)}$. The two critical values are defined in (53) and (54).*

Proof of Lemma B 9: The proof proceeds in three steps. First, show

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

Second, show

$$\begin{aligned} & P\left\{\min_{v \ge -\hat{\varphi}^*_T(q_T)} \left\|(\hat{D}^{-1/2}(q_T)S(q_T)\hat{L})Z_m - M_v v\right\|^2_{\hat{\Omega}^{-1}(q_T)} \le x\right\} \\ \xrightarrow{p} \quad & P\left\{\min_{v \ge -\pi^*} \left\{\left\|A'Z_m - M_v v\right\|^2_{(A'A)^{-1}}\right\} \le x\right\}. \end{aligned}$$

Third, deduce $\hat{c}^*_{(22)}(q_T) \xrightarrow{p} c^*_{(22)}$, as required for Part (b).

Proof of Step 1: By the choice of the sequence $\{T\}$ and the limit result in (B.8) and $\hat{\Lambda} \xrightarrow{p} \Lambda$,

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

Notice that if $\pi_j = 0$, then $\hat{\xi}_T(q_T) < \kappa_T$ and $\hat{\varphi}^*_{j,T}(q_T) = \hat{\varphi}_{j,T}(q_T) = 0 = \pi^*_j$ with probability one. On the other hand, if $\pi_j > 0$, then $\hat{\varphi}^*_{j,T}(q_T) = \infty = \pi^*_j$.

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

$$\min_{v \ge -\pi^*} \left\{\left\|A'Z_m - M_v v\right\|^2_{(A'A)^{-1}}\right\}$$

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

C Derivations for Bivariate VAR(1)

Consider a bivariate ($n = 2$) VAR(1) of the form $y_t = \Phi_1 y_{t-1} + u_t$ and focus on the response at horizon $h = 1$, which can be constructed from $R_1^v = \Phi\Sigma_{tr}$. Hence, let $\phi = vec((R_1^v)')$. The object of interest is $\theta = \partial y_{1,t+1}/\partial\epsilon_{1,t}$, and we impose the sign restriction that both θ as well as $\partial y_{2,t+1}/\partial\epsilon_{1,t}$ are nonnegative. Let $q = [q_1, q_2]'$. Then

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

Notice that in this example $\tilde{S}(q) = [\tilde{S}_\theta'(q), \tilde{S}_R'(q)]'$ is of full row rank for all values of q. Thus, we can set $V(q) = I$, replace $\tilde{S}(q)$ by $S(q)$, and write the objective function as

$$Q\big(\theta;\phi,W(\cdot)\big) = \min_{\|q\|=1} \; G\big(\theta,q;\phi,W(\cdot)\big)$$

where

$$G\big(\theta,q;\phi,W(\cdot)\big) = \min_{\mu\geq 0} \left\| S(q)\phi - \begin{pmatrix}\theta\\ \mu\end{pmatrix}\right\|_W^2 .$$

An analytical expression for $G\big(\theta,q;\phi,W(\cdot)\big)$ can be obtained as follows. Decompose

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

where

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

Thus, we can write

$$\left\| S(q)\phi - \begin{pmatrix}\theta\\ \mu\end{pmatrix}\right\|_W^2 = W_{11.22}(\theta - S_\theta\phi)^2 + W_{22}\Big(\mu - \Big[S_R\phi - W_{12}W_{22}^{-1}(\theta - S_\theta\phi)\Big]\Big)^2.$$

Now let

$$\begin{aligned} \hat{\mu}(\theta,q) &= \text{argmin}_{\mu\geq 0} \left\| S_(q)\phi - \begin{pmatrix}\theta\\ \mu\end{pmatrix}\right\|_W^2 \\ &= \begin{cases} S_R\phi - W_{12}W_{22}^{-1}(\theta - S_\theta\phi) & \text{if } S_R\phi - W_{12}W_{22}^{-1}(\theta - S_\theta\phi) \geq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Thus,

$$\begin{aligned} & G(\theta, q; \phi, W) \\ &= W_{11.22}(\theta - S_\theta \phi)^2 \\ &\quad + \begin{cases} 0 & \text{if } S_R\phi - W_{12}W_{22}^{-1}(\theta - S_\theta\phi) \geq 0 \\ W_{22}\Big(S_R\phi - W_{12}W_{22}^{-1}(\theta - S_\theta\phi)\Big)^2 & \text{otherwise} \end{cases} \end{aligned}$$

Let $\hat{\Lambda}$ be the bootstrap estimator of the covariance matrix of $\sqrt{T}(\hat{\phi} - \phi)$. The weight matrix $\hat{W}_T^*$ is given by $\hat{W}^*(q) = T(S(q)\hat{\Lambda}S'(q))^{-1}$, and the unit length vector q can be parameterized in spherical coordinates as $q(\alpha) = [\cos\alpha,\ \sin\alpha]'$. Thus the objective function for the construction of the confidence set is given by

$$Q\big(\theta; \hat{\phi}, \hat{W}^*(\cdot)\big) = \min_{\alpha \in [-\pi, \pi]} G\big(\theta, q(\alpha); \hat{\phi}, \hat{W}^*(\cdot)\big).$$