

Inference for VARs Identified with Sign Restrictions

Eleonora Granziera*
Bank of Finland

Hyungsik Roger Moon
*University of Southern California
and Yonsei University*

Frank Schorfheide
*University of Pennsylvania,
CEPR, NBER, and PIER*

September 15, 2017

*Correspondence: E. Granziera: Bank of Finland, Snellmaninaukio, 00101 Helsinki, Finland. E-mail: eleonora.granziera@bof.fi. H.R. Moon: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: moonr@usc.edu. F. Schorfheide: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. E-mail: schorf@ssc.upenn.edu. We thank James Stock, Fabio Canova, and Eric Renault as well as participants at various conferences and seminars for helpful comments. We also thank Mihye Lee for her contributions to the first draft of this paper and Minchul Shin for research assistance. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grants SES 1061725 and 1424843. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Finland. The Online Technical Appendix as well as data and software to replicate the empirical analysis are available at <https://web.sas.upenn.edu/schorf>.

Abstract

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

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

1 Introduction

During the three decades following Sims (1980) “Macroeconomics and Reality,” structural vector autoregressions (SVARs) have become an important tool in empirical macroeconomics. They have been used for macroeconomic forecasting and policy analysis, as well as to investigate the sources of business cycle fluctuations and to provide a benchmark against which modern dynamic macroeconomic theories can be evaluated. The most controversial step in the specification of a structural VAR is the mapping between reduced form one-step-ahead forecast errors and orthogonalized, economically interpretable structural innovations. Traditionally, SVARs have been constructed by imposing sufficiently many restrictions such that the relationship between structural innovations and forecast errors is one-to-one.

Over the past decade, starting with Faust (1998), Canova and De Nicolò (2002), and Uhlig (2005), empirical researchers have used more agnostic approaches that generate bounds on structural impulse response functions by restricting the sign of certain responses. We refer to this class of models as sign-restricted SVARs. They have been employed, for instance, to measure the effects of monetary policy shocks (e.g., Faust (1998), Canova and De Nicolò (2002), Uhlig (2005)), technology shocks (e.g., Dedola and Neri (2007), Peersman and Straub (2009)), government spending shocks (Mountford and Uhlig (2009), Pappa (2009)), oil price shocks (e.g., Baumeister and Peersman (2013), Kilian and Murphy (2012)), and financial shocks (e.g., Hristov, Hülsewig, and Wollmershäuser (2012), Gambetti and Musso (2017)).

Because impulse responses in sign-restricted SVARs can only be restricted to a bounded set, they belong to the class of set-identified econometric models, using the terminology of Manski (2003).¹ With the exception of Faust, Rogers, Swanson, and Wright (2003) and Faust, Swanson, and Wright (2004) (since the two papers are methodologically equivalent, we are using the abbreviation FRSW to refer to both of them), researchers have exclusively reported Bayesian credible bands for sign-restricted VARs, and a general method for constructing uniformly asymptotically valid frequentist confidence intervals was absent from the literature when the first draft of this paper was written; see Moon, Schorfheide, Granziera, and Lee (2011). As shown in detail in Moon and Schorfheide (2012), the large-sample numerical equivalence of frequentist confidence sets and Bayesian credible

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

sets breaks down in set-identified models, which means that Bayesian credible bands may not be interpreted as approximate frequentist confidence bands.

The goal of this paper is to provide researchers with an easy-to-use tool to construct valid frequentist confidence bands for impulse responses and other measures of the dynamic effects of structural shocks (e.g., variance decompositions) of sign-restricted SVARs. The specific contributions are the following: First, we formulate the problem of analyzing set-identified sign-restricted SVAR models in a moment-inequality-based minimum distance framework. Second, we find an easily interpretable sufficient condition for the non-emptiness of the identified set of sign-restricted structural impulse responses and we propose a consistent estimator of the identified set that is straightforward to compute. Third, using our minimum distance framework, we formally analyze Bonferroni confidence sets.² Fourth, we provide step-by-step recipes for practitioners on how to compute these confidence sets.

At an abstract level, our inference problem is characterized by a vector of point-identified reduced-form parameters ϕ , a vector of structural parameters (impulse responses or variance decompositions) θ , and a vector of nuisance parameters q . The sign restrictions generate an identified set for q , $F^q(\phi)$. Conditional on q and ϕ the vector θ is point identified, but because q is set-identified, so is θ and we denote its identified set by $F^\theta(\phi)$. To obtain a confidence set for θ we pursue a Bonferroni approach: we construct a confidence interval for the set-identified nuisance parameter q and then take the union of standard Wald confidence sets for θ that are generated conditional on all q in the first-stage confidence set. The Bonferroni inequality is used to ensure the desired coverage probability of the resulting confidence set for θ . We also show that the plug-in estimator $F^\theta(\hat{\phi})$ delivers a consistent estimate of the identified set for θ , denoted by $F^\theta(\phi)$.

In the first draft of this paper, we used a projection approach instead; see Moon, Schorfheide, Granziera, and Lee (2011).³ We constructed a joint confidence interval for the set-identified pair (q, θ) and projected it onto the θ ordinate. Here θ is the response of a particular variable at a particular horizon to a particular shock. In order to generate point-wise confidence bands for impulse response functions we repeated the computations for different definitions of θ . In subsequent

²The 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 confidence bands that are valid from a frequentist perspective with econometric tools to compute such bands.

³The projection approach for subvector inference has been further developed by Kaido, Molinari, and Stoye (2016) and Bugni, Canay, and Shi (2017).

research we compared the projection approach and the Bonferroni approach and found that there is no clear ranking of the two types of confidence sets. However, the Bonferroni approach has a clear computational advantage: the irregular confidence set for the nuisance parameter q only has to be computed once. Conditional on $F^q(\phi)$ one can easily generate standard confidence sets for impulse responses of different variables at different horizons, of vectors of responses, and of variance decompositions and then take unions over q . Thus, we decided to focus on the Bonferroni set in the current version of the paper.

The Bonferroni approach has a long history in the time series literature. For instance, Cavanagh, Elliott, and Stock (1995) and Campbell and Yogo (2006) use it to eliminate nuisance parameters that characterize the persistence of error terms or regressors. In the context of structural VARs, the Bonferroni approach has been used by FRSW. However, FRSW's setup is quite different from ours. In their framework the set identification of q arises from a rank deficiency in equality restrictions, which depend on estimated parameters. While FRSW restrict q further by imposing inequality conditions, these inequality conditions do not depend on estimated parameters. Our analysis, on the other hand, focuses on inequality restrictions for q that may or may not be binding and do depend on estimated parameters. This generalization is essential to cover the wide range of empirical applications referenced above. In the Monte Carlo analysis we explore ideas by Campbell and Yogo (2006) and McCloskey (2017) to tighten the Bonferroni sets.

Building upon recent advances in the moment-inequality literature in microeconometrics, in particular, Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010), we provide an asymptotic analysis of the Bonferroni approach to constructing confidence sets for the dynamic effects of shocks in sign-restricted SVARs. Because the number of linearly independent moment conditions is a function of the nuisance parameter q , we need to modify some of the existing microeconomic theory. As is common in the literature, we use a point-wise testing procedure to obtain a confidence set for q . We use Andrews and Soares' (2010a) moment selection procedure to tighten the critical values for the point-wise testing procedures.⁴ We adapt their theory to account for the q -dependent rank of the set of moment inequalities in our model and prove that the proposed confidence sets are asymptotically valid in a uniform sense. Our results on the non-emptiness of identified sets for impulse responses

⁴A recent survey of the moment-inequality literature is provided by Canay and Shaikh (2017). Alternative procedures include Andrews and Barwick (2012) and Romano, Shaikh, and Wolf (2014). Because of a potentially large number of inequality restrictions, the refinements to the moment selection proposed in Andrews and Barwick (2012) did not seem practical.

complement the equality-restriction-based VAR identification results reported in Rubio-Ramirez, Waggoner, and Zha (2010).

Since the working paper versions of this paper have been written, a series of alternative approaches for the construction of frequentist confidence bands of sign-restricted SVARs have been proposed. Gafarov, Meier, and Montiel Olea (2016b) developed an alternative projection-based approach that starts from a Wald confidence ellipsoid for the reduced-form VAR parameters and takes unions of the identified sets $F^\theta(\phi)$. Because it relies on the regular behavior of the estimator of the reduced-form coefficients, this method also has the interpretation of delivering Bayesian credible sets for the identified set $F^\theta(\phi)$. The downside is that it is quite conservative. Gafarov, Meier, and Montiel Olea (2016a) propose a δ -method confidence interval for sign-restricted SVARs which relies on a closed-form characterization of the endpoints of the identified set. While the resulting intervals are less conservative, their drawbacks are that they are only pointwise, but not uniformly consistent and that they can only be applied to scalar θ 's. Finally, Giacomini and Kitagawa (2015) construct robust Bayesian credible sets for impulse response functions in set-identified SVARs that have good frequentist properties.

The remainder of the paper is organized as follows. Section 2 develops the notation used in this paper and provides a simple example of a sign-restricted SVAR. We describe how set-identification arises in this model and sketch the Bonferroni approach to the construction of confidence intervals for the dynamic effects of structural shocks. Section 3 is geared toward practitioners and discusses the step-by-step implementation and computational aspects of the proposed inference method in the context of a general SVAR. Technical assumptions and large sample results are presented in Section 4. Some extensions are discussed in Section 5. To illustrate the proposed methods, we conduct a Monte Carlo study in Section 6 and generate confidence bands for output, inflation, interest rate, and money responses to a monetary policy shock in an empirical application in Section 7. Finally, Section 8 concludes. Proofs and detailed derivations as well as further information about the Monte Carlo experiments and the empirical analysis are relegated to a supplemental Online Appendix.

We use the following notation throughout the remainder of the paper: $\mathcal{I}\{x \geq a\}$ is the indicator function that is one if $x \geq a$ and zero otherwise. $0_{n \times m}$ is an $n \times m$ matrix of zeros and I_n is the $n \times n$ identity matrix. We use $[AB]_{(i)}$ to denote the i 'th row of the matrix $(A \cdot B)$ and $[AC]_i$ to denote the i 'th element of the vector $(A \cdot C)$. \otimes is the Kronecker product, $vec(\cdot)$ stacks the columns of a matrix, and $vech(\cdot)$ vectorizes the lower triangular part of a square matrix. We use

$\text{diag}(A_1, \dots, A_k)$ to denote a quasi-diagonal matrix with submatrices A_1, \dots, A_k on its diagonal and zeros elsewhere. If A is an $n \times m$ matrix, then $\|A\|_W = \sqrt{\text{tr}[WA'A]}$. In the special case of a vector, our definition implies that $\|A\|_W = \sqrt{A'WA}$. If the weight matrix is the identity matrix, we omit the subscript. We write $x \gg 0$ to mean that all elements of the vector x are strictly greater than zero; we write $x > 0$ to mean that all elements of x are greater than or equal to zero but not all equal to zero, that is, $x \neq 0$; finally, we write $x \geq 0$ to mean that all elements of x are greater than or equal to zero. We use \propto to indicate proportionality and “ \xrightarrow{p} ” and “ \implies ” to indicate convergence in probability and convergence in distribution, respectively, as $T \rightarrow \infty$. A multivariate normal distribution is denoted by $N(\mu, \Sigma)$. We use χ_m^2 to denote a χ^2 distribution with m degrees of freedom.

2 General Setup and Illustrative Example

Throughout this paper we consider an n -dimensional VAR with p lags, which takes the form

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad \mathbb{E}[u_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[u_t u_t' | \mathcal{F}_{t-1}] = \Sigma_u. \quad (1)$$

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

$$u_t = A_\epsilon \epsilon_t = \Sigma_{tr} \Omega_\epsilon \epsilon_t, \quad \mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}] = I_n, \quad (2)$$

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

$$y_t = \sum_{h=0}^{\infty} C_h(A_1, \dots, A_p) \Sigma_{tr} \Omega_\epsilon \epsilon_{t-h}. \quad (3)$$

We assume that the object of interest is the propagation of the first shock, $\epsilon_{1,t}$, and denote the first column of the matrix Ω_ϵ by q , where q is a unit-length vector. The domain of q is the n -dimensional unit sphere $\mathbb{S}^n = \{q \in \mathbb{R}^n \mid \|q\| = 1\}$. In Section 2.1 we discuss how imposing sign restrictions on some impulse responses generates set identification of the dynamic effects of $\epsilon_{1,t}$. Section 2.2 provides an illustration in the context of a bivariate VAR(0). Section 2.3 introduces some

important notation. We present the construction of a Bonferroni confidence set for the dynamic effects of $\epsilon_{1,t}$ in Section 2.4. The Bonferroni set is based on a confidence set for q , which is described in Section 2.5.

2.1 Sign Restrictions and Set Identification

The SVAR identification problem arises because the one-step-ahead forecast error covariance matrix Σ_u is invariant to the orthogonal matrix Ω_ϵ , which implies that Ω_ϵ and its first column q are not identifiable from the data. Point identification could be achieved by selecting a particular q as a function of $(A_1, \dots, A_p, \Sigma_{tr})$. The recent SVAR literature has pursued a more agnostic approach and restricted the set of admissible q by a collection of sign restrictions on impulse responses.⁵ The impulse response of variable $y_{i,t}$ to $\epsilon_{1,t}$ at horizon h is given by

$$IRF(i, h | \epsilon_{1,t} = 1) = [C_h(A_1, \dots, A_p) \Sigma_{tr}]_{(i, \cdot)} q. \quad (4)$$

We define the $n \times 1$ vector

$$\phi_r = [C_h(A_1, \dots, A_p) \Sigma_{tr}]'_{(i, \cdot)}$$

as the responses of a variable i at horizon h to the vector of reduced-form innovations u_t and summarize the sign-restricted impulse responses as

$$\Phi'_q q \geq 0, \quad \text{where} \quad \Phi_q = [\phi_1, \dots, \phi_r] \quad (5)$$

and Φ_q is a $n \times r$ matrix. The vectors ϕ_j are possibly multiplied by -1 to restrict the response to be weakly negative rather than weakly positive. Moreover, the notation in (5) is general enough to accommodate sign restrictions on cumulative impulse responses over \bar{h} periods, obtained from $\sum_{h=0}^{\bar{h}} C_h(A_1, \dots, A_p) \Sigma_{tr} q$.

The object of inference is a k -dimensional parameter defined as

$$\theta = f(\Phi_\theta(A_1, \dots, A_p, \Sigma_{tr}), q) \in \Theta \subset \mathbb{R}^k. \quad (6)$$

The parameter set Θ is chosen to be consistent with potential sign restrictions for elements of θ implied by (5). Our leading example of $f(\cdot)$ is a vector of impulse responses, which can be expressed as a linear function of the reduced-form impulse responses

$$f(\Phi_\theta(A_1, \dots, A_p, \Sigma_{tr}), q) = \Phi'_\theta q, \quad (7)$$

⁵We assume that these sign restrictions do not encode equality restrictions (e.g., by representing $a = 0$ as $a \leq 0$ and $a \geq 0$.) The extension to models that combine sign-restrictions and equality restrictions is deferred to Section 5.

where the definition of Φ_θ is similar to the definition of Φ_q in (5). In addition to impulse responses, researchers often report variance decompositions. For instance, the fraction of the one-step-ahead forecast error variance of variable $y_{1,t}$ explained by shock $\epsilon_{1,t}$ is given by

$$f(\Phi_\theta(A_1, \dots, A_p, \Sigma_{tr}), q) = \frac{\iota_1' \Sigma_{tr} q q' \Sigma_{tr}' \iota_1}{\iota_1' \Sigma_{tr} \Sigma_{tr}' \iota_1},$$

where $\iota_1 = [1, 0, \dots, 0]'$ is a $n \times 1$ vector.

While Φ_q and Φ_θ can be consistently estimated, the vector q as well as the object of interest θ are only set identified. We use $F^q(\Phi_q)$ and $F^\theta(\Phi_q, \Phi_\theta)$ to denote the identified sets of q and θ , respectively. Formally, they are defined as

$$F^q(\Phi_q) = \{q \in \mathbb{S}^n \mid \Phi_q' q \geq 0\} \quad (8)$$

$$F^\theta(\Phi_q, \Phi_\theta) = \{\theta \in \Theta \mid \exists q \in F^q(\Phi_q) \text{ s.t. } \theta = f(\Phi_\theta, q)\}. \quad (9)$$

The goal is to construct a confidence set for θ . As an intermediate step in a Bonferroni approach we will also construct a confidence set for q .

2.2 Identified Sets $F^q(\cdot)$ and $F^\theta(\cdot)$ in a Bivariate VAR(0)

For concreteness consider the following example. Suppose the vector y_t is composed of inflation and output growth and the one-step-ahead forecast errors are linear functions of structural demand and supply shocks, stacked in the vector $\epsilon_t = [\epsilon_{D,t}, \epsilon_{S,t}]'$. In order to obtain bounds for the effects of a demand shock, we impose the sign restriction that, contemporaneously, a demand shock moves prices and output in the same direction and the normalization restriction that a positive demand shock increases prices:

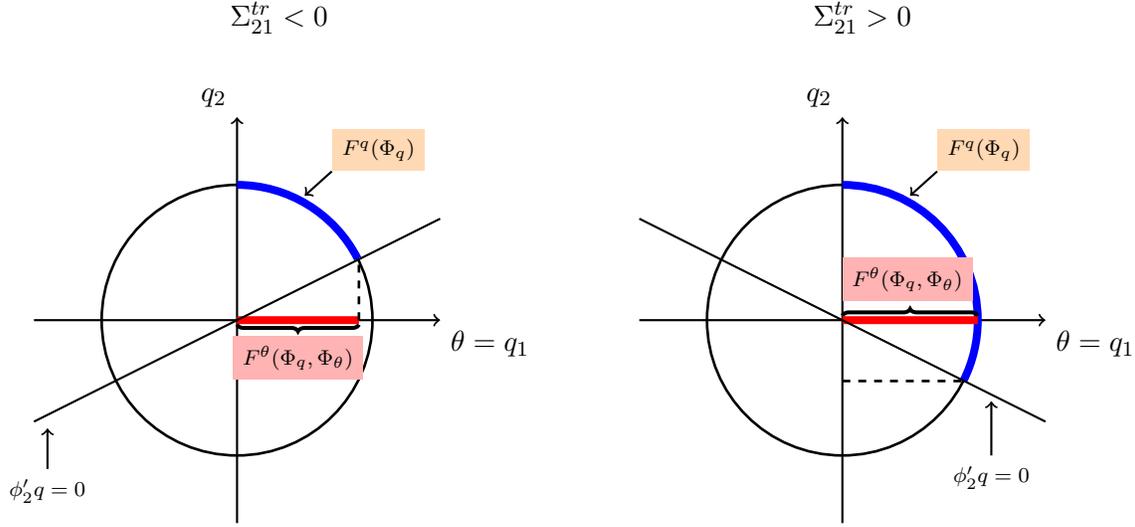
$$\phi_1' q = [\Sigma_{11}^{tr} \quad 0] \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \geq 0, \quad \phi_2' q = [\Sigma_{21}^{tr} \quad \Sigma_{22}^{tr}] \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \geq 0, \quad \Phi_q = [\phi_1, \phi_2]. \quad (10)$$

Suppose that the object of interest, θ , is the contemporaneous inflation response to a demand shock $\epsilon_{D,t}$:

$$\theta = \phi_1' q, \quad \Phi_\theta = \phi_1. \quad (11)$$

Figure 1 provides an illustration of the two inequality constraints and the resulting identified sets. To simplify the graphical illustration, we assume that $\Sigma_{11}^{tr} = 1$ which implies $\theta = q_1$. Because

Figure 1: Identified Sets for VAR(0)



the Cholesky factorization of the covariance matrix Σ_u is normalized such that Σ_{11}^{tr} and Σ_{22}^{tr} are nonnegative, (10) implies that

$$q_1 \geq 0 \quad \text{and} \quad q_2 \geq - \left(\frac{\Sigma_{21}^{tr}}{\Sigma_{22}^{tr}} \right) q_1. \quad (12)$$

The x -axis of both panels corresponds to $\theta = q_1$ and the y -axis represents q_2 . Each panel depicts a unit circle as well as the locus $\phi'_2 q = 0$. In the left panel $\Sigma_{21}^{tr} < 0$, whereas in the right panel $\Sigma_{21}^{tr} > 0$. The identified set $F^q(\Phi_q)$ is given by the arc that ranges from the intersection of the unit circle with the y -axis to the intersection with $\phi'_2 q = 0$. The identified set $F^\theta(\Phi_q, \Phi_\theta)$ is given by the projection of the arc onto the x -axis. The identified set is a singleton only if $\Sigma_{22}^{tr} = 0$, which means that the one-step-ahead forecast error covariance matrix is singular. We will rule out this case because in practice it is not empirically relevant.

2.3 Rank Reductions and Notation

In order to develop a notation for the general inference problem we need to accommodate two types of rank reductions. First, consider the matrix Φ_q for the bivariate VAR(0):

$$\Phi_q = \begin{bmatrix} \Sigma_{11}^{tr} & \Sigma_{21}^{tr} \\ 0 & \Sigma_{22}^{tr} \end{bmatrix}.$$

To construct a confidence set for q , we will replace the unknown Φ_q by a sample estimate $\hat{\Phi}_q$ and start from the high-level assumption that $\hat{\Phi}_q$ has a normal limit distribution. Because of the zeros in the Cholesky factorization of Σ_u (and possibly other restrictions imposed on the reduced-form parameters), the covariance matrix of $\hat{\Phi}_q$ has a rank-reduction. We circumvent this issues as follows. Let

$$\bar{S}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

be the selection matrix that deletes the zero elements in $\text{vec}(\Phi_q)$ and such that

$$\phi_q = \bar{S}' \text{vec}(\Phi_q) \quad \text{and} \quad \bar{S} \phi_q = \text{vec}(\Phi_q).$$

Then, we can express the sign restrictions as

$$\tilde{S}(q) \phi_q = (I \otimes q') \bar{S} \phi_q \geq 0. \quad (13)$$

The second type of rank reduction arises as follows. After eliminating $\Sigma_{12}^{tr} = 0$, we obtain $\phi_q = [\Sigma_{11}^{tr}, \Sigma_{21}^{tr}, \Sigma_{22}^{tr}]'$ in the bivariate VAR(0). In turn, the matrix $\tilde{S}(q)$ takes the form:

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

The first row of $\tilde{S}(q)$ becomes zero if $q = [q_1, q_2]' = [0, 1]$. The rank deficiency arises because S eliminates the second column of the matrix $(I \otimes q')$, which for $q_1 = 0$ contains the only non-zero entry in the first row of $\tilde{S}(q)$. As a consequence the covariance matrix of $\tilde{S}(q)\hat{\phi}_q$ is singular at $q = [0, 1]'$ and cannot be inverted to form a weight matrix for the estimation of q . In order to eliminate the rows of zeros in $\tilde{S}(q)$, we introduce the selection matrix $V(q)$ and define

$$S(q) = V(q)\tilde{S}(q). \quad (14)$$

The row dimension of $S(q)$ is $r(q)$. By construction, the matrix $S(q)$ has the full row rank for all q .

Inference about q will be based on the objective function

$$G(q; \phi_q, W(\cdot)) = \min_{\mu \geq 0} \|S(q)\phi_q - V(q)\mu\|_{W(q)}^2. \quad (15)$$

The vector μ captures the slackness in the inequalities generated by the sign restrictions in (13). $W(q)$ is a symmetric and positive-definite weight matrix with a dimension that adjusts to the dimension of $V(q)$:

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

The matrix $\tilde{W}(q)$ is a weight matrix that conforms with $\tilde{S}(q)\phi_q$. An important example of a $\tilde{W}(q)$ is the inverse of the asymptotic covariance matrix of $\sqrt{T}\tilde{S}(q)(\hat{\phi}_q - \phi_q)$. It can be verified that

$$q \in F^q(\phi_q) \quad \text{if and only if} \quad G(q; \phi_q, W(\cdot)) = 0, \quad (16)$$

where we now write $F^q(\phi_q)$ instead of $F(\Phi_q)$.

2.4 A Bonferroni Confidence Interval for θ

Let S'_θ be the selection matrix that deletes the zeros other pre-determined elements in $\text{vec}(\Phi_\theta)$ and define $\phi_\theta = S'_\theta \text{vec}(\Phi_\theta)$. Moreover, let $\phi = [\phi'_q, \phi'_\theta]'$. We will often write $F^\theta(\phi)$ to abbreviate $F^\theta(\Phi_q, \Phi_\theta)$. The goal is to obtain a confidence interval $CS^\theta(\hat{\phi})$ that satisfies the condition

$$\liminf_{T \rightarrow \infty} \inf_{\rho \in \mathcal{R}} \inf_{\theta \in F^\theta(\phi(\rho))} P_\rho \{ \theta \in CS^\theta(\hat{\phi}) \} \geq 1 - \alpha. \quad (17)$$

The vector ϕ may be a subvector of a larger reduced-form parameter vector ρ with domain \mathcal{R} , that characterizes the distribution of the data y_1, \dots, y_T , which is why we write $\phi(\rho)$. For instance, in a Gaussian VAR ρ comprises the elements of A_1, \dots, A_p , and Σ_u ; see (1). The parameter θ does not appear as an index of the probability distribution P , because conditional on ϕ the parameter θ does not affect the distribution of the estimator of the reduced-form parameters $\hat{\phi}$. The confidence interval is indexed by $\hat{\phi}$ because it is a sufficient statistic in our setup.

As mentioned in the Introduction, in our application it is quite natural to use a Bonferroni approach to compute an asymptotic $1 - \alpha$ confidence set for θ . Under the Bonferroni approach, one first constructs a confidence set for q . This is a “non-standard” object because q is a set-identified parameter. Conditional on q , however, inference for θ becomes “standard,” because θ is point-identified. Let $\alpha = \alpha_1 + \alpha_2$. Bonferroni confidence steps can be obtained in three steps:

1. Construct a $1 - \alpha_1$ confidence interval $CS^q(\hat{\phi}_q)$ for q with the property that

$$\liminf_{T \rightarrow \infty} \inf_{\rho \in \mathcal{R}} \inf_{q \in F^q(\phi_q)} P_\rho \{ q \in CS^q(\hat{\phi}_q) \} \geq 1 - \alpha_1. \quad (18)$$

2. Generate a confidence set for θ conditional on q . This is a “regular” problem because conditional on q , the vector θ is point identified. For instance, if θ is scalar and defined as $\theta = \phi'_\theta q$, then one can use a Wald confidence interval of the form:

$$CS^\theta_q(\hat{\phi}_\theta) = \Theta \cap \left[\hat{\phi}'_\theta q - z_{\alpha_2/2} \hat{\sigma}_\theta, \hat{\phi}'_\theta q + z_{\alpha_2/2} \hat{\sigma}_\theta \right]. \quad (19)$$

Here $z_{\alpha_2/2}$ is the two-sided α_2 critical value associated with the $N(0, 1)$ distribution and $\hat{\sigma}_{\hat{\theta}}$ is a standard error estimate for $\hat{\theta} = \hat{\phi}'_q q$. The intersection of the symmetric confidence interval for θ can be used, for instance, to truncate the symmetric interval at zero, if θ is a response that is assumed to be non-negative.

3. Construct the confidence set for θ by taking the following union of $CS_q^\theta(\hat{\phi}_\theta)$ sets:

$$CS^\theta(\hat{\phi}) = \bigcup_{q \in CS^q(\hat{\phi}_q)} CS_q^\theta(\hat{\phi}_\theta). \quad (20)$$

2.5 A Confidence Set for q

The main contribution of this paper is to adapt an inference procedure from the moment inequality literature to obtain a confidence set for q in the first step of the Bonferroni procedure. The confidence set is generated as a level set based on the sample analogue of the objective function in (15):

$$CS^q(\hat{\phi}_q) = \left\{ q \in \mathbb{S}^n \mid G(q; \hat{\phi}_q, W(\cdot)) \leq c^{\alpha_1}(q) \right\}. \quad (21)$$

Here $c^{\alpha_1}(q)$ is a critical value that guarantees that the confidence set satisfies (18). In the remainder of this subsection, we outline the derivation of the critical value $c^{\alpha_1}(q)$ for the bivariate VAR(0) example.

For illustrative purposes, suppose that the estimates of the reduced-form parameters have an exact standard normal distribution:

$$\sqrt{T}(\hat{\phi}_q - \phi_q) \sim N(0, I_3).$$

We parameterize the slackness in the inequality restrictions as

$$q_1 \phi_{q,1} = \tilde{\mu}_1, \quad q_1 \phi_{q,1} + q_2 \phi_{q,2} = \tilde{\mu}_2$$

and use the weight matrix that standardizes the distribution of $\tilde{S}(q)\hat{\phi}_q$:

$$W(q) = T \begin{bmatrix} \frac{1}{q_1^2} & 0 \\ 0 & 1 \end{bmatrix}.$$

In turn, we can express the sample analogue of the objective function in (15) as

$$\begin{aligned} G(q; \hat{\phi}_q, W(\cdot)) &= \min_{\mu \geq 0} \left[\sqrt{T}(\hat{\phi}_{q,1} - \phi_{q,1}) + \sqrt{T}(\mu_1 - \tilde{\mu}_1)/|q_1| \right]^2 \mathcal{I}\{q \neq [0, 1]'\} \\ &\quad + \left[\sqrt{T}(\hat{\phi}_{q,2} - \phi_{q,2})q_1 + \sqrt{T}(\hat{\phi}_{q,3} - \phi_{q,3})q_2 + \sqrt{T}(\mu_2 - \tilde{\mu}_2) \right]^2 \\ &= \min_{\nu \geq -\sqrt{T}\tilde{\mu}} \left[Z_1 - \nu_1/|q_1| \right]^2 \mathcal{I}\{q \neq [0, 1]'\} + [Z_2 - \nu_2]^2, \end{aligned}$$

where $\mu = [\mu_1, \mu_2]'$, $\nu = [\nu_1, \nu_2]'$, and Z_1 and Z_2 are two independent $N(0, 1)$ random variables.

A conservative upper bound on the sample objective function can be obtained by assuming that both inequalities are binding, that is, $\tilde{\mu}_1 = \tilde{\mu}_2 = 0$:

$$G(q; \hat{\phi}_q, W(\cdot)) \leq Z_1^2 \mathcal{I}\{Z_1 \leq 0\} + Z_2^2 \mathcal{I}\{Z_2 \leq 0\}.$$

Critical values for the distribution of the bound can be easily obtained by simulation. A sharper bound and a smaller critical value that leads to a smaller confidence set can be obtained by realizing that at most one inequality is binding. Thus, we will use the moment selection approach of Andrews and Soares (2010) to eliminate non-binding moment conditions constructing critical values for $G(q; \hat{\phi}_q, W(\cdot))$. This means that in our illustrative example, the critical value can be essentially reduced to the $100(1 - \alpha_1)$ quantile of the distribution of $Z_1^2 \mathcal{I}\{Z_1 \leq 0\}$.

3 Implementation

This section focuses on the implementation of the proposed inference methods. A formal statement of assumptions and a rigorous analysis of the large sample properties of the confidence set will follow in Section 4. The remainder of this section is organized as follows. Section 3.1 briefly discusses the estimation of ϕ . Section 3.2 describes how we construct the confidence set for q . The calculation of confidence sets for θ given q is reviewed in Section 3.3. Finally, we provide some additional details for the computation of confidence bands for impulse responses in Section 3.4. Throughout this section we assume that the impulse responses are not restricted through equality conditions (e.g., the restriction that certain responses have to be zero). Extensions of our approach to a setting in which some identifying information is extracted from equality conditions are straightforward but notationally cumbersome and discussed in Section 5.

3.1 Estimating the Reduced-Form Coefficients ϕ

We start from the assumption that $\hat{\phi}_q$ and $\hat{\phi}_\theta$ have Gaussian limit distributions and that the asymptotic covariance matrices can be estimated consistently:

$$\begin{aligned} \sqrt{T}(\hat{\phi}_q - \phi_q) &\implies N(0, \Lambda_{qq}) \quad \text{and} \quad \hat{\Lambda}_{qq} \xrightarrow{p} \Lambda_{qq} > 0 \\ \sqrt{T}(\hat{\phi}_\theta - \phi_\theta) &\implies N(0, \Lambda_{\theta\theta}) \quad \text{and} \quad \hat{\Lambda}_{\theta\theta} \xrightarrow{p} \Lambda_{\theta\theta} > 0. \end{aligned} \tag{22}$$

This assumption requires that all roots of the characteristic polynomial associated with the difference equation (1) lie outside of the unit circle. Throughout the paper, we are ruling out the presence of unit roots and are assuming that y_t is trend stationary. Moreover, we proceed under the assumption that Λ_{qq} and $\Lambda_{\theta\theta}$ are full rank. Because we previously eliminated the $n(n-1)/2$ zero elements of the lower triangular matrix Σ_{tr} from the vectors ϕ_r , $r = 0, 1, \dots, R$, the rank condition does not impose serious constraints on the applicability of our analysis.⁶

The VAR coefficient matrices A_1, \dots, A_p and Σ_u can be estimated by OLS. An estimate of Σ_{tr} is obtained by applying the Cholesky decomposition to $\hat{\Sigma}_u$. We then evaluate the functions $\Phi_q(\cdot)$ and $\Phi_\theta(\cdot)$ at $\hat{A}_1, \dots, \hat{A}_p, \hat{\Sigma}_{tr}$ to obtain $\hat{\phi}_q$ and $\hat{\phi}_\theta$. We obtain $\hat{\Lambda}_{qq}$ and $\hat{\Lambda}_{\theta\theta}$ by using a parametric bootstrap procedure: conditional on $\hat{A}_1, \dots, \hat{A}_p, \hat{\Sigma}_u$ we simulate n_Λ bootstrap samples $Y_{1:T}^*$ from the VAR in (1). From each bootstrap sample we compute $\hat{\phi}_q^*$. Finally we compute the bootstrap sample covariance matrix of $\hat{\phi}_q^*$ and scale it appropriately to obtain $\hat{\Lambda}_{qq}$. The same approach is used to compute $\hat{\Lambda}_{\theta\theta}$.

3.2 Confidence Set for q

The confidence interval for q is obtained by verifying whether

$$G(q; \hat{\phi}_q, W(\cdot)) \leq c^{\alpha_1}(q)$$

for $q \in \mathcal{Q}$. This requires the selection of a grid \mathcal{Q} and the evaluation of the critical-value function $c^{\alpha_1}(q)$.

⁶Consider a 4-variable VAR(4) and suppose that the responses of 3 of the 4 variables are restricted upon impact and for the subsequent 3 periods. Moreover, the object of interest is the response of the fourth variable at horizon $h = 9$. The number of reduced-form coefficients is $4 \cdot 16 + 10 = 74$. In order to construct the sign-restricted responses as well as θ , the number of elements in the vector ϕ is bounded by $4 \cdot 3 \cdot 4 + 4 = 52$, which means that the covariance matrix Λ is non-singular.

Generating a Grid for q . We generate $n_{\mathcal{Q}}$ grid points for $q \in \mathbb{S}^n$ from a distribution that is uniform under rotations using a well-known result by James (1954). Let $Z^{(j)}$, $j = 1, \dots, n_{\mathcal{Q}}$, be a sequence of $n \times 1$ vectors of $iidN(0, I_n)$ random vectors and define $q^{(j)} = Z^{(j)} / \|Z^{(j)}\|$. Then, $q^{(j)}$ is uniformly distributed on the unit sphere \mathbb{S}^n . We define the grid as $\mathcal{Q} = \{q^{(1)}, \dots, q^{(n_{\mathcal{Q}})}\}$.

Weight Matrix for Sample Objective Function. The weight matrix $\hat{W}_q(q)$ is obtained as follows. We denote the asymptotic covariance matrix of $\sqrt{T}S(q)(\hat{\phi}_q - \phi_q)$ as $\Sigma(q) = S(q)\Lambda_{qq}S(q)'$. A consistent estimator is given by

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

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

$$\hat{W}(q) = T\hat{D}^{-1/2}(q)\hat{B}(q)\hat{D}^{-1/2}(q) \quad (23)$$

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

$$G(q; \hat{\phi}_q, \hat{W}(\cdot)) = \min_{\mu \geq 0} T \left\| \hat{D}^{-1/2}(q)S(q)\hat{\phi}_q - \hat{D}^{-1/2}(q)V(q)\mu \right\|_{\hat{B}(q)}^2$$

The objective function $G(q; \hat{\phi}_q, \hat{W}(\cdot))$ has the same structure as the objective functions considered in the literature on moment inequality models, e.g., Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010). The main difference of our set up for the VAR application is the dimension of $S(q)\hat{\phi}_q$, $\hat{D}(q)$ and $\hat{B}(q)$ varies with q and the limit $\hat{D}(q)$ as a function of q can be singular.

Critical Values. In order to obtain the critical value function $c^{\alpha_1}(q)$ in we apply the moment selection approach of Andrews and Soares (2010). The moment selection tries to eliminate clearly non-binding inequality conditions in the weak limit of the objective function $G(q; \hat{\phi}_q, \hat{W}(\cdot))$ and compute the required critical value. An estimate of the slackness in inequality condition $j = 1, \dots, r(q)$ is provided by

$$\hat{\xi}_{j,T}(q) = \hat{D}_{jj}^{-1/2}(q)[S(q)]_{(j,\cdot)}\sqrt{T}\hat{\phi}_q. \quad (24)$$

Inequality condition j is deemed non-binding if

$$\hat{\xi}_{j,T}(q) \geq \kappa_T, \quad (25)$$

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

$$\hat{r}_2(q) = \sum_{j=1}^{r(q)} \mathcal{I}\{\hat{\xi}_{j,T}(q) \geq \kappa_T\} \quad \text{and} \quad \hat{r}_1(q) = r(q) - \hat{r}_2(q), \quad (26)$$

respectively.

Define the $(\hat{r}_1(q) \times r(q))$ selection matrix $M_{\hat{\xi}}(q)$ that deletes rows of $\hat{D}^{-1/2}(q)S(q)\hat{\phi}_q$ that correspond to non-binding inequality conditions in the sense of (25). Moreover, let m be the dimension of the vector ϕ_q ,

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

Define the random function

$$\bar{\mathcal{G}}(q; \hat{B}(q), M_{\hat{\xi}}(q)) = \min_{\nu \geq 0} \left\| M_{\hat{\xi}}(q)\hat{A}'(q)Z_m - \nu \right\|_{M_{\hat{\xi}}(q)\hat{B}(q)M'_{\hat{\xi}}(q)}^2, \quad (27)$$

where ν is a $\hat{r}_1(q) \times 1$ vector. We adopt the convention that $\bar{\mathcal{G}}(q; \hat{B}(q), M_{\hat{\xi}}(q)) = 0$ if $\hat{r}_1(q) = 0$. The critical value $c^{\alpha_1}(q)$ is defined as

$$c^{\alpha_1}(q) = 1 - \alpha_1 \text{ quantile of } \bar{\mathcal{G}}(q; \hat{B}(q), M_{\hat{\xi}}(q)) \quad (28)$$

and can be obtained from a simulation approximation of the limit objective function.⁷ If $\hat{B}(q) = I$, then the evaluation of $\bar{\mathcal{G}}(\cdot)$ is fast because the quadratic programming problem has the following closed-form solution:

$$\bar{\mathcal{G}}(q; \hat{B}(q), M_{\hat{\xi}}(q)) = \sum_{j=1}^{\hat{r}_1(q)} [M_{\hat{\xi}}(q)\hat{A}'(q)Z_m]_j^2 \mathcal{I}\{[M_{\hat{\xi}}(q)\hat{A}'(q)Z_m]_j < 0\}.$$

3.3 Confidence Set for θ Conditional on q

Conditional on q , the dynamic effects of the shock $\epsilon_{1,t}$ on y_t are point-identified and the inference about impulse responses and variance decompositions becomes regular. Methods on how to construct confidence intervals for these objects date back to Runkle (1987), who proposed to use either the δ -method in combination with numerical derivatives of the mapping from reduced-form VAR coefficients into IRFs and variance decompositions or to use a residual-based bootstrap. Lütkepohl

⁷For $j = 1, \dots, n_Z$ generate random vectors $Z_m^{(j)}$ and compute $\bar{\mathcal{G}}^{(j)}(q; \hat{B}(q), M_{\hat{\xi}}(q))$. Then compute the $1 - \alpha_1$ percentile of the empirical distribution of the simulated limit objective functions.

(1990) derived asymptotic distributions based on analytical derivatives for the δ -method and Mitnik and Zadrozny (1993) provided extensions to VARMA models. A recent survey of the literature on frequentist inference for IRFs and variance decompositions in point-identified settings is provided by Kilian and Lütkepohl (2017). Any of these methods can be embedded into our Bonferroni approach.

3.4 Special Case: Confidence Bands for IRFs

Confidence bands for impulse responses in the VAR literature predominantly depict pointwise confidence intervals, which means that we can express the scalar parameter θ as $\tilde{S}_\theta(q)\phi_\theta$, where ϕ_θ summarizes the reduced-form impulse responses that are necessary to generate the structural response θ and $\tilde{S}_\theta(q)$ is defined similarly as $\tilde{S}(q)$ in Section 2.3. For this important special case one can show that $F^\theta(\phi_q, \phi_\theta)$ is convex and bounded, which simplifies computations and reporting of results.

Lemma 1 *Suppose that $F^q(\phi_q)$ is non-empty and not a singleton. Moreover $\theta = S_\theta(q)\phi_\theta$ and $k = \dim(\theta) = 1$. Then $F^\theta(\phi_q, \phi_\theta)$ is convex and bounded.*

Approximating $F^q(\hat{\phi}_q)$ and $F^\theta(\hat{\phi})$. An estimate of the identified set for q can be obtained from

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

Thus, for every $q \in \mathcal{Q}$ one checks whether $\tilde{S}(q)\hat{\phi}_q \geq 0$ and retains the q 's for which the condition is satisfied. Denote the elements of $\hat{F}^q(\hat{\phi}_q)$ by $q^{(j)}$, $j = 1, \dots, n_q$, where $n_q \leq n_{\mathcal{Q}}$. Compute $\theta^{(j)} = \hat{\phi}'_\theta q^{(j)}$. We show in the Online Appendix that $F^\theta(\hat{\phi})$ is a bounded interval. Thus, we define the interval

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

Computing $CS^q(\hat{\phi}_q)$ and $CS^\theta(\hat{\phi})$. The computation of $CS^q(\hat{\phi}_q)$ follows the steps outlined in Section 3.2. Note that by construction $\hat{F}^q(\hat{\phi}_q) \subseteq CS^q(\hat{\phi}_q)$. Denote the elements of $CS^q(\hat{\phi}_q)$ by $q^{(j)}$, $j = 1, \dots, n_q$. Then, for each $q^{(j)}$, compute the Wald interval with bounds

$$\theta_l^{(j)} = \hat{\phi}'_\theta q^{(j)} - z_{\alpha_2/2} \sqrt{q^{(j)'} \hat{\Lambda}_{\theta\theta} q^{(j)} / T} \quad \text{and} \quad \theta_u^{(j)} = \hat{\phi}'_\theta q^{(j)} + z_{\alpha_2/2} \sqrt{q^{(j)'} \hat{\Lambda}_{\theta\theta} q^{(j)} / T},$$

and let

$$CS^\theta(\hat{\phi}) = \Theta \bigcap [(\min_{j=1, \dots, n_q} \theta_l^{(j)}), (\max_{j=1, \dots, n_q} \theta_u^{(j)})].$$

The intersection with Θ can be used to restrict the confidence interval to values of θ that are consistent with the assumed sign restriction. While $CS^q(\hat{\phi}_q)$ has to be calculated only once, the computations for $CS^\theta(\hat{\phi})$ have to be repeated for every response $\theta = \partial y_{i,t+h}/\partial \epsilon_{1,t}$ of interest. Here i potentially ranges from $i = 1, \dots, n$ and $h = 0, 1, \dots, h_{max}$.

4 Large Sample Analysis

This section formally establishes the consistency of the plug-in estimators $F^q(\hat{\phi}_q)$ and $F^\theta(\hat{\phi})$ and the asymptotic validity of the confidence sets $CS^q(\hat{\phi}_q)$ and $CS^\theta(\hat{\phi})$. As mentioned in Section 2.4, the vectors ϕ_q and ϕ may generally not be sufficient to characterize the sampling distribution of data and estimators. Thus, we again will use ρ to characterize the distribution of the data under the reduced-form VAR model (1). The statements about uniform asymptotic coverage probabilities will be made for $\rho \in \mathcal{R}$. Some of the regularity conditions will be required to hold for a slightly larger, δ -inflated open set

$$\mathcal{R}^\delta = \{\tilde{\rho} \in \bar{\mathcal{R}} \mid \exists \rho \in \mathcal{R} \text{ s.t. } \|\tilde{\rho} - \rho\| < \delta\}, \quad (29)$$

where $\bar{\mathcal{R}} \supset \mathcal{R}$ and $\delta > 0$.⁸ Asymptotic inference for q is discussed in Section 4.1 and Section 4.2 considers inference for θ .

4.1 Asymptotic Inference for q

We begin by stating some high-level assumptions.

Assumption 1 *There exists a compact reduced-form parameter set \mathcal{R} and a δ -inflated superset $\mathcal{R}^\delta \subset \bar{\mathcal{R}}$ defined in (29) such that:*

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

$$\Phi_q \lambda = 0.$$

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

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

⁸For instance, suppose ρ is an autocorrelation parameter for an AR(1) model. We could define $\mathcal{R} = [0, 0.999]$, $\mathcal{R}^\delta = [0, 1)$ for $\delta = 0.001$, and $\bar{\mathcal{R}} = [0, 1]$.

- (iv) For each $\rho \in \mathcal{R}$ the matrix $\Lambda_{qq}(\rho)$ is continuous, positive definite, and there exists a full-rank positive-definite matrix Λ_{min} such that $\Lambda_{qq}(\rho) - \Lambda_{min} \geq 0$ for all $\rho \in \mathcal{R}$.
- (v) There exists an estimator $\hat{\Lambda}_{qq}$ of $\Lambda_{qq}(\rho_T)$ such that $\|\hat{\Lambda}_{qq} - \Lambda_{qq}(\rho_T)\| \xrightarrow{p} 0$ for any converging sequence $\{\rho_T\}$.

Condition (i) of Assumption 1 states that the convex cone generated by the columns of the reduced-form impulse response matrix Φ_q does not contain the zero vector. This assumption is sufficient to ensure that the identified set $F^q(\phi_q(\rho))$ is non-empty and that the plug-in estimator $F^q(\hat{\phi}_q)$ is consistent whenever $\hat{\phi}_q \xrightarrow{p} \phi_q$ (see Theorem 1 below). Assumption 1(i) rules out, for instance, that equality conditions are coded as pairs of inequalities, and, more generally, that linear combinations of inequalities constrain impulse responses to be equal to zero. We discuss in Section 5 how our framework can be extended to allow for a mixture of inequality and equality restrictions on impulse responses.

Condition (i) is typically not satisfied for all values of the reduced-form parameter $\rho \in \bar{\mathcal{R}}$, which is why we only require it to hold on the set $\mathcal{R}^\delta \subset \bar{\mathcal{R}}$. For instance, consider a VAR(1) generalization of the bivariate VAR(0) in Section 2.2 with autoregressive coefficient matrix A_1 . As before, suppose y_t is composed of inflation and output growth and the investigator imposes the sign restriction that in response to a (positive) demand shock inflation and output responses are both non-negative upon impact and one period after impact. In this case

$$\Phi_q' = \begin{bmatrix} \Sigma_{tr} \\ A_1 \Sigma_{tr} \end{bmatrix}.$$

If $A_1 = \text{diag}(\rho_1, \rho_2)$ and $\rho_1, \rho_2 < 0$, then Condition (i) is violated. Conditional on these reduced-form parameters, the identified set is empty. Assumption 1 excludes these values of ρ from \mathcal{R}^δ . From a practitioner's perspective, an empty confidence set $CS^q(\hat{\phi}_q)$ provides evidence that the imposed sign restrictions are inconsistent with the estimated reduced-form parameters.

The continuity in Condition (ii) is with respect to the Euclidean norm. While ρ could in principle be infinite-dimensional if the distribution of the error terms is treated nonparametrically, the function $\phi_q(\cdot)$ only depends on the finite-dimensional subvector of ρ that contains the reduced-form parameters $A_1, \dots, A_p, \Sigma_u$; see Equation (4). Conditions 1(iii) and (v) require that $\hat{\phi}_T$ and $\hat{\Lambda}_{qq}$ converge uniformly for $\rho \in \mathcal{R}$. The uniform convergence of $\hat{\phi}_q$ to a Gaussian limit distribution also requires a restriction of the domain of ρ because it breaks down at the boundary of the stationary region in the VAR parameter space. For instance, in the context of an AR(1) model $y_t = \rho_T y_{t-1} + u_t$

with autoregressive coefficient $\rho_T = 1 - c/T$, an estimator of an impulse response at horizon $h = 1$, that is, $\phi_q(\rho) = \rho$, behaves according to

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

Uniform convergence to a Gaussian limit distribution can be achieved if \mathcal{R} is restricted to the interval $[-1 + \epsilon, 1 - \epsilon]$ for some $\epsilon > 0$.⁹ From a practitioner's perspective we are essentially assuming that the researcher has applied some stationarity-inducing transformations, e.g., transformed prices into inflation rates. Because some authors, e.g., Uhlig (2005), prefer to specify VARs in terms of variables that exhibit (near) non-stationary dynamics, our Monte Carlo experiments in Section 6 include designs in which the roots of the vector autoregressive lag polynomial are close to the unit circle.¹⁰

Our first theorem establishes that the identified set $F^q(\phi_q)$ is non-empty and not a singleton, that is, the dynamic effects of $\epsilon_{1,t}$ are set-identified instead of point-identified. This result can be deduced from Assumption 1(i) using Gordan's Alternative Theorem (see, for instance, Border (2007)).

Theorem 1 *Suppose Assumption 1(i) is satisfied. Then, the identified set $F^q(\phi_q(\rho))$ is non-empty and is not a singleton for all $\rho \in \mathcal{R}^\delta$.*

The second theorem focuses on asymptotic inference. The first part establishes the consistency of the plug-in estimator $F^q(\hat{\phi}_q)$. The consistency is stated in terms of the Hausdorff distance. We denote the Hausdorff distance between two sets A and B by $d_H(A, B)$.¹¹ The consistency relies on the compactness of $F^q(\phi_q)$ and the continuity of the correspondences with respect to ϕ_q . Unlike in some of the models studied by Chernozhukov, Hong, and Tamer (2007), it is not necessary to inflate the set $F^q(\hat{\phi}_q)$ by $\epsilon_T \downarrow 0$ to achieve consistency.¹² The second part of Theorem 2 establishes the asymptotic validity of the confidence set $CS^q(\hat{\phi}_q)$. A formal proof of the Theorem is provided in the

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

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

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

¹²A result similar to ours in a general GMM setting is provided by Yildiz (2012). We prove the result directly based on Assumption 1.

Online Appendix. The proof of the second part closely follows the proof of Theorem 1 in Andrews and Soares (2010). However, a number of non-trivial modifications are required to account for the potential rank reduction of $\tilde{S}(q)$ as a function of q .

Theorem 2 *Suppose that Assumption 1 is satisfied. (i) If, in addition, the domain of ϕ_q is compact and $\hat{\phi}_q \xrightarrow{p} \phi_q$, then, $d_H(F^q(\hat{\phi}_q), F^q(\phi_q)) \xrightarrow{p} 0$. (ii) Suppose that $0 < \alpha < 1/2$. The confidence set $CS^q(\hat{\phi}_q)$, defined in (21), is an asymptotically valid confidence set for q :*

$$\liminf_{T \rightarrow \infty} \inf_{\rho \in \mathcal{R}} \inf_{q \in F^q(\phi_q(\rho))} P_\rho \{q \in CS^q(\hat{\phi}_q)\} \geq 1 - \alpha.$$

4.2 Asymptotic Inference for θ

As discussed in Section 3.3, we do not provide any new results on confidence intervals for impulse responses or variance decompositions *conditional on* the vector q . For these intervals, we rely on the existing literature. We use $CS_q^\theta(\hat{\phi}_\theta)$ to denote a confidence set for θ conditional on q . The following assumption is required for asymptotic inference about the parameter θ .

Assumption 2 *(i) The function $\theta = f(\Phi_\theta, q)$ is continuous in both its arguments. (ii) Suppose $CS_q^\theta(\hat{\phi}_\theta)$ satisfies:*

$$\liminf_T \inf_{\rho \in \mathcal{R}} \inf_{(\theta, q) \in F^{\theta, q}(\phi(\rho))} P_\rho \left\{ \theta \in CS_q^\theta(\hat{\phi}_\theta) \right\} \geq 1 - \alpha_2,$$

where $F^{\theta, q}(\phi) = \{\theta \in \Theta, q \in \mathbb{S}_n \mid q \in F^q(\phi_q), \theta = f(\Phi_\theta, q)\}$.

The first condition of Assumption 2(i) is quite weak and the two leading examples of θ in Section 2.1 satisfy this condition. The second condition of Assumption 2(ii) is a high-level condition that is needed for the asymptotic validity of the Bonferroni confidence set of θ . The condition requires that the pointwise confidence set $CS_q^\theta(\hat{\phi}_\theta)$ in q is uniformly valid. In the leading examples of θ , an impulse response of the form $\theta = \Phi'_\theta q$, the conditional confidence set CS_q^θ in (19) satisfies the uniformity condition because the asymptotic normality of $\hat{\phi}_\theta$ in (22) holds uniformly in ϕ . Combining the results of Theorem 2(ii) with Assumption 2 leads to the following theorem:

Theorem 3 *Suppose that Assumption 1 is satisfied. (i) If, in addition, the domain of ϕ is compact and $\hat{\phi} \xrightarrow{p} \phi$, then, $d_H(F^\theta(\hat{\phi}), F^\theta(\phi)) \xrightarrow{p} 0$. (ii) Suppose that $0 < \alpha < 1/2$ and Assumption 2 is satisfied. The confidence set $CS^\theta(\hat{\phi})$, defined in (20), is an asymptotically valid confidence set for θ :*

$$\liminf_{T \rightarrow \infty} \inf_{\rho \in \mathcal{R}} \inf_{\theta \in F^\theta(\phi(\rho))} P_\rho \{ \theta \in CS^\theta(\hat{\phi}) \} \geq 1 - \alpha.$$

5 Extensions

We now discuss three extensions to the construction of $CS^q(\hat{\phi}_q)$: (i) models that use both sign restrictions and zero restrictions to identify structural impulse responses, (ii) the identification of multiple shocks, (iii) and the use of bootstrapped critical values instead of simulated asymptotic critical values.

Sign Restrictions Combined with Equality Restrictions. Assumption 1(i) rules out that opposing sign restrictions are used to represent equality restrictions on impulse responses. Nonetheless, it is straightforward to sharpen the identified set by combining sign restrictions with more traditional exclusion restrictions. In some applications, the restriction that certain responses are zero on impact (zero restrictions) can be translated into a domain restriction for q that does not depend on any other reduced-form parameters. For instance, in the empirical analysis in Section 7.2 we will replace an unrestricted 4×1 vector $q \in \mathbb{S}^n$ by the restricted vector $q = [0_{1 \times 2}, q_2']'$, where q_2 is a 2×1 vector with $\|q_2\| = 1$. In this case the previously developed methods can be applied without any modification.

If the equality restrictions imposed on the impulse responses lead to restrictions on q that depend on some of the reduced-form parameters, then they can be accommodated by generalizing the objective function $G(q; \phi, \tilde{W})$ in (15) as follows. Define

$$\tilde{G}(q; \phi, \tilde{W}) = \min_{\mu \geq 0} \left\| \begin{pmatrix} S_{eq}(q)\phi \\ S(q)\phi - V(q)\mu \end{pmatrix} \right\|_{\tilde{W}(q)}^2, \quad (30)$$

where $S_{eq}(q)\phi$ corresponds to the responses that are restricted to be zero. Following the arguments in Andrews and Soares (2010), it is straightforward albeit tedious to extend the proof of Theorem 2 to a mixture of equality and inequality conditions.¹³ The extension closely resembles the proof of Theorem 2(i) in the working paper version Moon, Schorfheide, and Granziera (2013) for the projection-based confidence set, which also involves a mix of equality and inequality conditions. From a practitioners perspective, the only other modification that is required, is to replace the limit objective function $\bar{G}(\cdot)$ in (27) that is used to simulate the critical value $c^{\alpha_1}(q)$ by the limit expression of $\tilde{G}(q; \phi, \tilde{W})$ in (30).

¹³If we denote the matrix of zero-restricted orthogonalized responses by $\Phi_{q,eq}$, then the generalization of Assumption 1(i) is: there do not exist vectors $\lambda > 0$ and $\lambda_{eq} \geq 0$ such that $\Phi_q \lambda + \Phi_{q,eq} \lambda_{eq} = 0$. The generalized analysis would use Motzkin's Transposition Theorem; see Border (2007).

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

$$\tilde{S}(\Omega)\phi_q \geq 0$$

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

Bootstrapped Critical Values Instead of Asymptotic Critical Values. Our simulated critical values rely on the Gaussian limit distribution of $\sqrt{T}\hat{D}^{-1/2}(q)S(q)(\hat{\phi}_q - \phi_q)$, which is reflected in the vector $\hat{A}'(q)Z_m$ in the random function $\bar{G}(\cdot)$ in (27). Alternatively, the critical values could be constructed by replacing draws from $\hat{A}'(q)Z_m$ with draws from the bootstrap approximation of $\sqrt{T}\hat{D}^{-1/2}(q)S(q)(\hat{\phi}_q - \phi_q)$. Bootstrap procedures for VAR impulse response functions are discussed, for instance, in Kilian (1998) and Kilian and Lütkepohl (2017).

6 Monte Carlo Illustrations

In this section we conduct three Monte Carlo experiments to illustrate the properties of our proposed confidence sets. In these experiments θ is a scalar impulse response. During preliminary computations we noticed that the results for $\hat{B}(q) = I$ and $\hat{B}(q) = \hat{\Omega}^{-1}(q)$ were very similar. Thus, we decided to subsequently report results for $\hat{B}(q) = I$ because in this case the critical values can be computed much faster. We will drop the $\hat{\phi}$ arguments from the confidence sets and report coverage probabilities and average lengths for CS^q and CS^θ . Each Monte Carlo experiment involves the steps summarized in Table 1, which are repeated $n_{sim} = 5,000$ times.

The three experiments differ with respect to the data generating process (DGP). Experiment 1 (Section 6.1) is based on the bivariate VAR(0) model in Section 2.2. Experiment 2 (Section 6.2) features a bivariate VAR(1). The simulation designs for the Experiments 1 and 2 are obtained by fitting a VAR(0) to data on U.S. inflation and GDP growth and fitting first-order VARs to inflation

Table 1: Steps of Monte Carlo Experiments

1. Generate a sample of size T from the data-generating process.
2. Compute $\hat{\phi}_q$, $\hat{\phi}_\theta$, and the bounds of $F^\theta(\hat{\phi}_q, \hat{\phi}_\theta)$.
3. Compute $\hat{\Lambda}_{qq}$ and $\hat{\Lambda}_{\theta\theta}$ using a parametric bootstrap approach.
4. Compute the $1 - \alpha_1$ confidence set CS^q .
5. For each definition of θ , compute the $1 - \alpha_2$ confidence sets CS^θ .
6. For each definition of θ , compute the $1 - (\alpha_1 + \alpha_2)$ confidence sets CS^θ .

Table 2: Monte Carlo Design

	Experiment 1		Experiment 2	
	Design 1	Design 2	Design 3	Design 4
	VAR(0)	VAR(1)	VAR(1)	VAR(1)
Σ_{11}^{tr}	0.597	0.295	0.283	0.210
Σ_{21}^{tr}	-0.205	-0.092	-0.081	-0.043
Σ_{22}^{tr}	0.812	0.795	0.817	0.542
$A_{1,11}$		0.873	0.806	0.450
$A_{1,12}$		0.003	0.032	0.014
$A_{1,21}$		-0.229	-0.278	0.060
$A_{1,22}$		0.230	0.985	0.953
$\lambda_1(A_1)$		0.871	$0.89 - 0.03i$	0.955
$\lambda_2(A_1)$		0.231	$0.89 + 0.03i$	0.498

Notes: Designs are obtained by estimating a VAR(0) or VAR(1) of the form $y_t = A_0 + A_1 y_{t-1} + u_t$, $\mathbb{E}[u_t u_t'] = \Sigma_{tr} \Sigma_{tr}'$ using OLS. $\lambda_i(A_1)$ is the i 'th eigenvalue of A_1 . $y_{1,t}$ is the log difference of the U.S. GDP deflator, scaled by 100 to convert into percentages. $y_{2,t}$ is either the log difference of U.S. GDP or deviations of log GDP from a linear trend, scaled by 100. Design 1: inflation and GDP growth, 1964:I to 2006: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.

and either output growth or linearly detrended log GDP. Finally, Experiment 3 (Section 6.3) mimics the four-variable VAR(2) fitted to U.S. data on output, inflation, interest rates, and money balances in the empirical analysis of Section 7.

Table 3: Experiments 1 and 2: Single-Horizon Sign Restrictions

Experiment 1		Experiment 2						
Design 1		Design 2		Design 3		Design 4		
Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	
$F^q(\phi_q)$	$\frac{42}{100}\pi$		$\frac{36}{100}\pi$		$\frac{47}{100}\pi$		$\frac{51}{100}\pi$	
$F^\theta(\phi)$	0.579		0.233		0.226		0.094	
Sample Size $T = 100$								
CS^q	0.938	$\frac{47}{100}\pi$	0.936	$\frac{81}{100}\pi$	0.932	$\frac{57}{100}\pi$	0.940	$\frac{67}{100}\pi$
CS^θ	0.980	0.671	0.979	0.295	0.934	0.265	0.942	0.128
CS^{ϕ_q}	0.879		0.865		0.865		0.871	
Sample Size $T = 500$								
CS^q	0.930	$\frac{44}{100}\pi$	0.936	$\frac{44}{100}\pi$	0.932	$\frac{51}{100}\pi$	0.936	$\frac{56}{100}\pi$
CS^θ	0.990	0.622	0.991	0.265	0.963	0.244	0.958	0.110
CS^{ϕ_q}	0.909		0.894		0.901		0.904	

Notes: *Length* refers to the average length of the confidence intervals across Monte Carlo repetitions. For $F^q(\phi_q)$ and CS^q we report the arc length, see Figure 1. We let $\alpha_1 = \alpha_2 = 0.05$, which implies that the nominal coverage probabilities are 95% for CS^q and 90% for CS^θ and CS^{ϕ_q} . The confidence interval for ϕ_q has a nominal coverage probability of 90%.

6.1 Experiment 1

Design. The parameterization of the DGP $y_t \sim iidN(0, \Sigma_u)$ is provided in Table 2 in the column labeled *Design 1*. We define θ as the response of $y_{1,t}$ to $\epsilon_{1,t}$. Because $\Sigma_{21}^{tr} < 0$ in our design, the geometry of the Monte Carlo design corresponds to the left panel of Figure 1. Thus, the upper bounds (in polar coordinates) of F^q and CS^q are $\pi/2$ and the lower bounds of F^θ and CS^θ are zero, respectively. The identified set for θ is $F^\theta(\phi_{q,0}, \phi_{\theta,0}) = [0, 0.578]$. Below, we report coverage probabilities for the lower bound of F^q and the upper bound of F^θ because they are the least favorable parameter values in the respective identified sets. We consider sample sizes of $T = 100$ and $T = 500$. The grid \mathcal{Q} for q is obtained as follows: q is transformed into polar coordinates $[\cos(\varphi), \sin(\varphi)]'$ and we choose $n_{\mathcal{Q}} = 315$ equally spaced grid points for φ on the interval $(-\pi/2, \pi/2]$. The number of bootstrap repetitions to obtain $\hat{\Lambda}_{qq}$ and $\hat{\Lambda}_{\theta\theta}$ is $n_{\Lambda} = 1,000$ and the number of simulations to obtain the critical value and $c^{\alpha_1}(q)$ is $n_Z = 500$. Further details on the implementation are provided in the Online Appendix.

Results. Detailed results for the frequentist confidence intervals are summarized in Table 3. Recall that the nominal coverage probability for θ is 90%. For $T = 100$ the actual coverage probability for the Bonferroni sets is 0.98. As we increase the sample size to $T = 500$, the length of the confidence intervals shrinks, while the actual coverage probabilities increases to 0.99. It is instructive to also examine the coverage probabilities of CS^q and the Wald confidence set for $\phi_q = \text{vech}(\Sigma_{tr})$, which we denote by CS^{ϕ_q} . The coverage probability for the reduced-form parameter vector ϕ_q is 88% for $T = 100$ and approaches its nominal value of 90% as the sample size is increased to $T = 500$. This increase in coverage probability for ϕ_q mirrors the increase in coverage probability for θ . The Bonferroni intervals are computed based on $\alpha_1 = \alpha_2 = 0.05$, which implies that the nominal coverage probability of CS^q is 95%. The actual coverage probabilities for the nuisance parameter vector q are slightly smaller, namely, around 93%. Overall, the Bonferroni-type marginalization generates conservative confidence intervals for θ .

6.2 Experiment 2

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

$$y_t = A_1 y_{t-1} + u_t, \quad u_t \sim iidN(0, \Sigma_u).$$

The choices for A_1 and Σ_u are summarized in Table 2 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 A_1 are 0.871 and 0.231. *Design 4* is the most persistent with eigenvalues 0.955 and 0.498. We focus on responses at horizon $h = 1$, which can be obtained from $\phi_q = \text{vec}((A_1 \Sigma_{tr})')$. The structural parameter of interest, θ , is defined as $\partial y_{1,t+1} / \partial \epsilon_{1,t}$. As in Experiment 1, we compute coverage probabilities for the lower bound of $F^q(\phi_q)$ and the upper bound of $F^\theta(\phi)$. The grid \mathcal{Q} for q is obtained as follows: q is transformed into polar coordinates $[\cos(\varphi), \sin(\varphi)]'$ and we choose $n_{\mathcal{Q}} = 629$ equally spaced grid points for φ on the interval $(-\pi, \pi]$. The remaining aspects of the design are the same as in Experiment 1.

Sign Restrictions over a Single Horizon. We impose the sign restrictions that $\partial y_{1,t+1} / \partial \epsilon_{1,t}$ and $\partial y_{2,t+1} / \partial \epsilon_{1,t}$ are non-negative:

$$\phi_{q,1}q_1 + \phi_{q,2}q_2 \geq 0 \quad \text{and} \quad \phi_{q,3}q_1 + \phi_{q,4}q_2 \geq 0.$$

Table 4: Experiment 2: Multiple-Horizon Sign Restrictions, Sample Size $T = 100$

	Design 2			Design 3			Design 4		
	Coverage	Length	Bind.Ineq	Coverage	Length	Bind.Ineq	Coverage	Length	Bind.Ineq
Restrictions: $h = 0, 1$									
$F^q(\phi_q)$		$\frac{35}{100}\pi$			$\frac{43}{100}\pi$			$\frac{47}{100}\pi$	
CS^q	0.953	$\frac{48}{100}\pi$	1.29	0.976	$\frac{50}{100}\pi$	1.97	0.953	$\frac{53}{100}\pi$	2.06
$F^\theta(\phi)$		0.265			0.277			0.209	
CS^θ	0.982	0.307		0.977	0.317		0.949	0.236	
Restrictions: $h = 0, \dots, 4$									
$F^q(\phi_q)$		$\frac{6}{1000}\pi$			$\frac{37}{100}\pi$			$\frac{47}{100}\pi$	
CS^q	0.985	$\frac{40}{100}\pi$	7.78	0.989	$\frac{49}{100}\pi$	4.48	0.979	$\frac{56}{100}\pi$	7.83
$F^\theta(\phi)$		0.006			0.261			0.208	
CS^θ	1.000	0.277		0.993	0.315		0.949	0.235	

Notes: *Length* either refers to the length of the population identified set or the average length of the confidence intervals across Monte Carlo repetitions. For $F^q(\phi_q)$ and CS^q we report the arc length, see Figure 1. *Bind.Ineq* is the average number of inequalities considered binding by the Andrews and Soares (2010) moment selection procedure. We let $\alpha_1 = \alpha_2 = 0.05$, which implies that the nominal coverage probabilities are 95% for CS^q and 90% for CS^θ .

For now we do not impose sign restrictions on the responses at impact or at horizons greater than $h = 1$. The geometry of the identified sets $F^q(\phi_q)$ and $F^\theta(\phi_q, \phi_\theta)$ and its projections is similar to the geometry depicted in Figure 1. The main difference is that the second boundary of the identified set is given by the solution of $q_1\phi_{q,3} + q_2\phi_{q,4} = 0$ and $\|q\| = 1$, instead of $q = [0, 1]'$. Overall, the results for Experiment 2 reported in Table 3 are qualitatively similar to those for *Design 1*. The actual coverage probabilities of the confidence sets for q are around 0.94 and therefore close to the nominal coverage probability of $1 - \alpha_1 = 0.95$. The θ sets, on the other hand, are conservative. Their coverage probabilities range from 0.942 to 0.991, thereby exceeding the nominal level of 0.9.

Sign Restrictions over Multiple Horizons. As before, we define θ as the contemporaneous impact of the shock on $y_{1,t}$: $\theta = \partial y_{1,t} / \partial \epsilon_{1,t}$. However, we now restrict the signs of the impulse responses $\partial y_{i,t+h} / \partial \epsilon_{1,t} \geq 0$ for both variables $i = 1, 2$ over multiple periods: $h = 0, 1, \dots, H$. This increases the number of inequality conditions. Monte Carlo results are presented in Table 4. The effect of adding sign restrictions differs across the three designs. In *Design 2* the lengths of the identified sets $F^q(\cdot)$ and $F^\theta(\cdot)$ shrink drastically: from 0.35π and 0.265 for $H = 1$ to 0.006π and 0.007 , respectively, for $H = 4$. Under *Design 4* the sizes of the two identified sets remain constant as H is increased from 1 to 4. *Design 2* is an intermediate case. Restricting impulse responses at

multiple horizons essentially adds rays to Figure 1. The location of the new rays relative to the $H = 1$ rays determines whether the identified sets shrink or not.

For *Design 2* and *3* the length of the confidence intervals for q and θ are decreasing in the number of inequality restrictions, but they do not shrink as quickly as the length of the identified sets. Simultaneously, the actual coverage probability for q increases with the number of sign restrictions. In *Design 2* the coverage probability for $H = 1$ is 0.953, which is close to the nominal coverage probability of $1 - \alpha_1 = 0.95$. For $H = 4$ the coverage probability increases to 0.985. While in population for any H only one inequality is binding¹⁴ at the boundary of the identified set for q , the average number of moment conditions deemed binding by the Andrews and Soares (2010) selection rule rises from 1.29 to 7.78 in *Design 2*. Recall that in order to guarantee a uniform asymptotic coverage probability, the selection rule has to classify too many rather than too few moment conditions as binding. This inflates the critical value as well as the coverage probability and makes the q confidence set more conservative. We observe a similar pattern for *Design 3*. The Bonferroni sets for θ are generally conservative across all designs and maximum horizons H . Under *Design 2* and *3* the actual coverage probabilities tend to increase as more restrictions are added. Nonetheless the average length decreases. This decrease is most pronounced for *Design 2*. Here the length shrinks from 0.307 ($H = 1$) to 0.277 ($H = 4$). Under *Design 4* the average length of the confidence interval and the coverage probability stay essentially constant as we vary the number of restrictions.

6.3 Experiment 3

Design. Finally, we consider a four-variable VAR(2) that mimics the model for per capita GDP (in deviations from a linear trend), inflation, the federal funds rate, and real money balances used in the empirical application in Section 7:

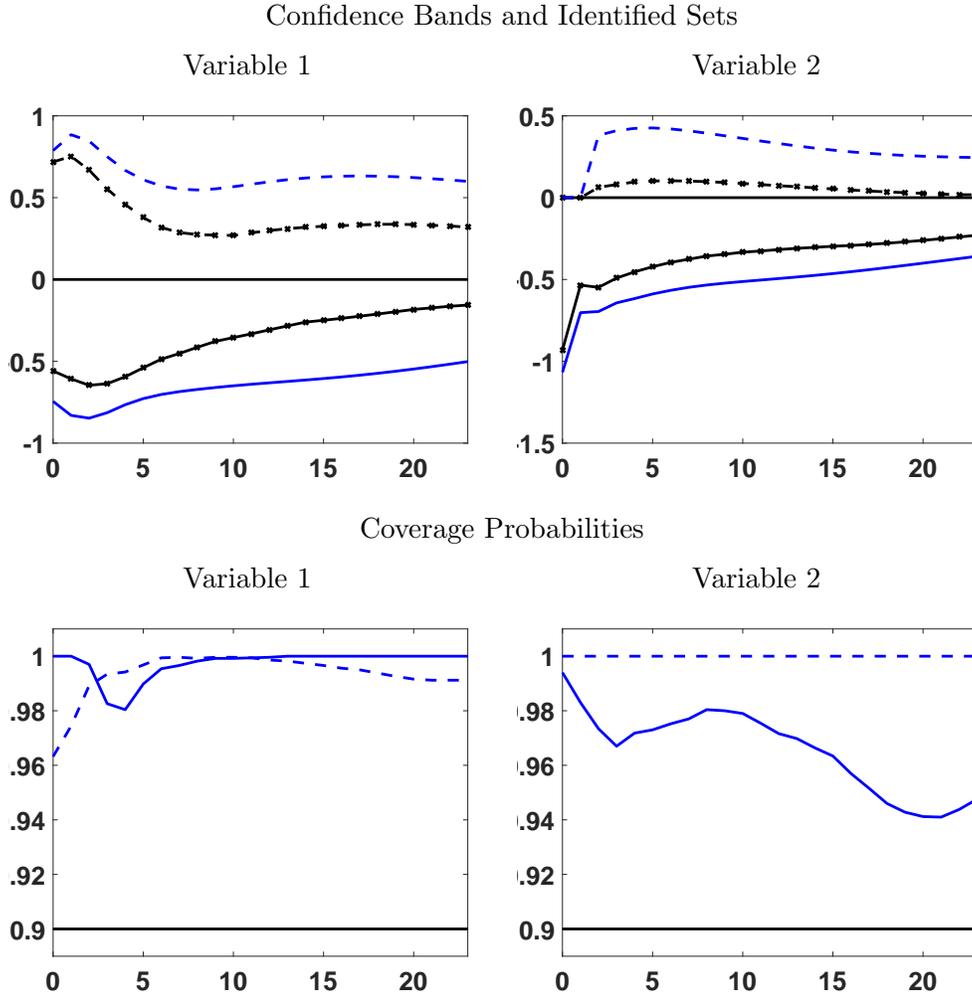
$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + u_t \quad \text{with } u_t \sim N(0, \Sigma). \quad (31)$$

The reduced-form parameters are set equal to the empirical point estimates (reported in the Online Appendix). As in the application below, we consider the following sign restrictions:

$$\frac{\partial y_{2,t+h}}{\partial \epsilon_{1,t}} \leq 0, \quad \frac{\partial y_{3,t+h}}{\partial \epsilon_{1,t}} \geq 0, \quad \frac{\partial y_{4,t+h}}{\partial \epsilon_{M,t}} \leq 0, \quad h = 0, 1.$$

¹⁴An inspection of Figure 1 suggests that if more than one inequality condition is binding then it must be the case that the rays corresponding to the binding inequality conditions are identical.

Figure 2: Impulse Responses Bands and Coverage Probabilities



Notes: Top panels: population identified set $F^\theta(\phi)$ (black with crosses) and averaged upper (dashed) and lower (solid) bounds of pointwise confidence sets (blue). Bottom panels: actual coverage probabilities for 90% confidence sets at the lower (solid) and upper (dashed) bounds of the pointwise identified sets. $\alpha_1 = \alpha_2 = 0.05$.

The sample size for the simulated data sets is $T = 170$. We set the number of (randomly generated from a uniform distribution on the hypersphere) grid points for \mathcal{Q} to $n_{\mathcal{Q}} = 20,000$. The number of bootstrap repetitions to obtain $\hat{\Lambda}_{qq}$ and $\hat{\Lambda}_{\theta\theta}$ is $n_{\Lambda} = 1,000$ and the number of simulations to obtain the critical value $c^{\alpha_1}(q)$ is $n_Z = 1,000$. As in the previous experiments, we focus on $\hat{B} = I$. The implementation of the computations for $CS^\theta(I)$ follows the description in Section 3.

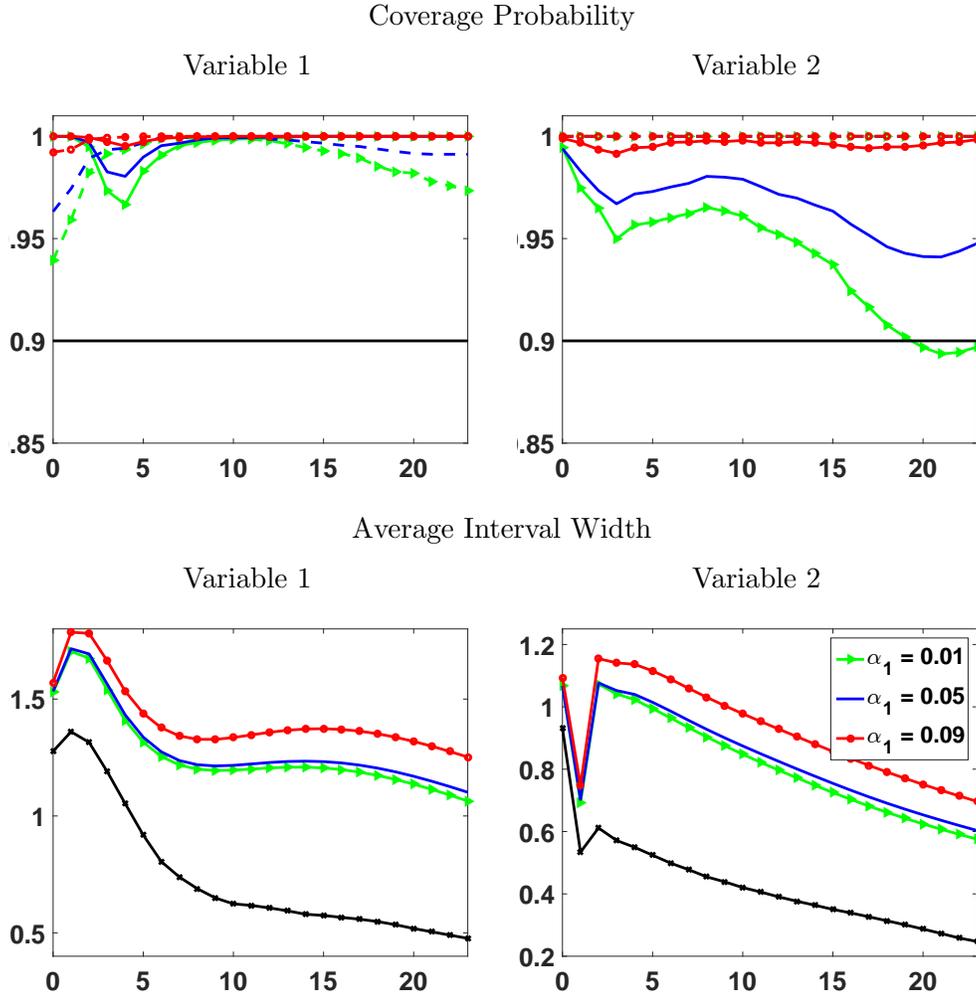
Benchmark Results. Baseline results for $\alpha_1 = \alpha_2 = 0.05$ are plotted in Figure 2. The top panels depict the upper bounds and the lower bounds for pointwise identified sets and confidence

sets for the responses of Variables 1 and Variables 2 at horizons $h = 0, 1, \dots, 23$. The bounds of the confidence sets are averaged across Monte Carlo repetitions. An economic interpretation of the responses will be provided in Section 7. For now we focus on the widths of the confidence bands relative to the widths of the population identified sets and the coverage probabilities, which are depicted in the bottom panels. The identified sets have a considerable width that leaves the sign of the response of $y_{1,t}$ undetermined. The confidence bands are noticeably wider than the identified sets, which is a reflection of the sampling uncertainty associated with the estimators of the reduced-form parameters. The coverage probabilities in the bottom panels are computed for the upper bounds (dashed) and lower bounds (solid) of the pointwise identified sets. As in Experiments 1 and 2, the actual coverage probability is substantially larger than the nominal coverage probability of 90% (indicated by the solid horizontal line), making the confidence bands conservative.

Adjusting α_1 , Keeping $\alpha = 0.1$ Fixed. The baseline choice of $\alpha_1 = 0.05$ has been arbitrary. Thus, it is worthwhile to explore what happens if we decrease or increase α_1 , which determines the size of CS^q . Figure 1 provides some intuition for the potential outcomes of this experiment. In the left panel of the figure (labeled $\Sigma_{21}^{tr} < 0$) the upper bound for the identified set $F^\theta(\cdot)$ is determined by the lower bound of $F^q(\cdot)$. Thus, increasing α_1 and thereby decreasing the size of $F^q(\cdot)$ can potentially sharpen the confidence set for θ , provided that the decrease in $F^q(\cdot)$ exceeds the increase in the conditional confidence set CS_q^θ . Alternatively, if $\Sigma_{21}^{tr} > 0$ (depicted in the right panel of Figure 1), the upper bound of $F^\theta(\cdot)$ is determined by a value of q that lies strictly in the interior of $F^q(\cdot)$. Thus, in order to shrink the confidence set for θ one should lower α_1 and raise α_2 so that CS_q^θ shrinks.

Figure 3 depicts the coverage probabilities and the average interval width for three levels of α_1 : $\alpha_1 = 0.05$ which we used to generate the baseline results in Figure 2, $\alpha_1 = 0.01$, and $\alpha_1 = 0.09$. It turns out that in this particular Monte Carlo design it is advantageous to reduce α_1 . Setting $\alpha_1 = 0.01$ reduces the coverage probability of the intervals and shrinks the width of the intervals. However, the effect is modest at best. Relative to the overall width of the identified sets and the baseline confidence bands, the reduction is very small. The actual coverage probability remains above 95% except for the long horizon responses of $y_{2,t}$ which fall slightly below the nominal level of $\alpha = 0.9$ for $h \geq 18$. As we saw in Table 3, the confidence set for the reduced-form VAR parameters can have an actual coverage probability that is less than its nominal coverage probability, which in turn tightens the confidence interval for θ .

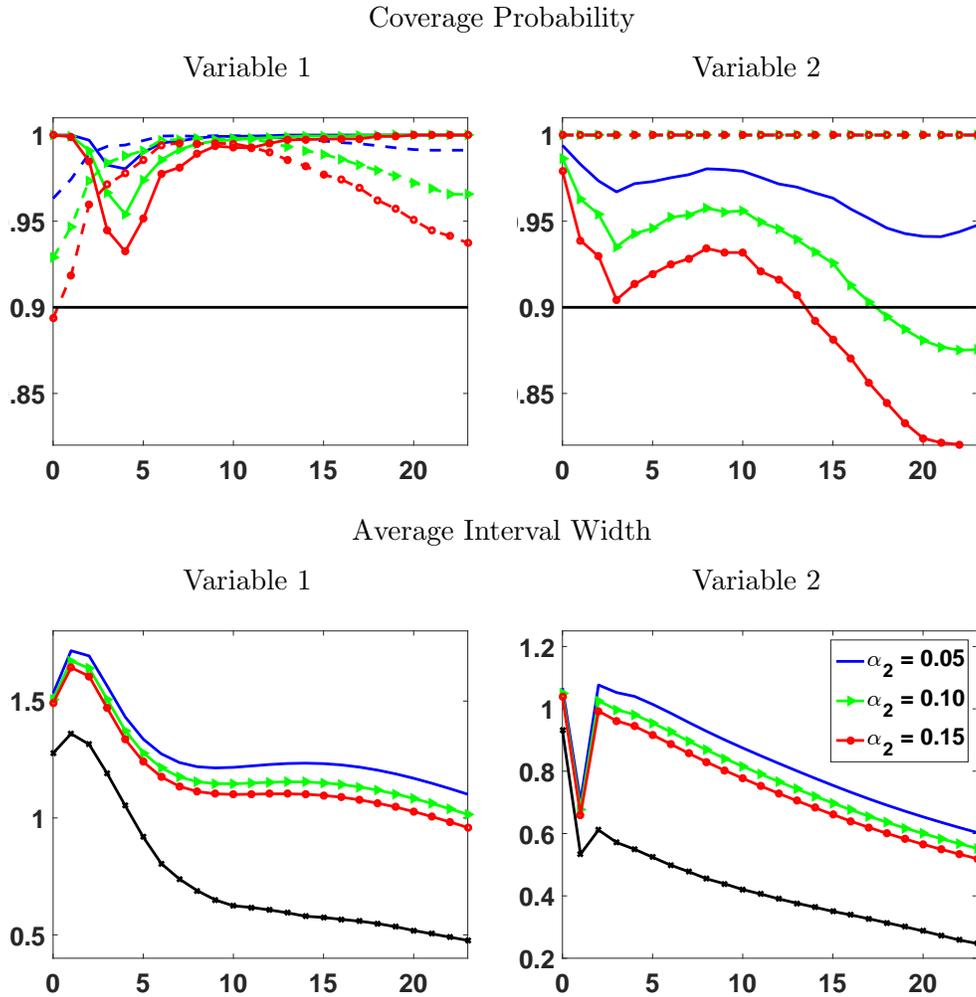
Figure 3: The Effect of Varying α_1 , Fixed $\alpha = 0.10$



Notes: Top panels: actual coverage probabilities for 90% confidence sets. Dashed lines refer to upper bounds and solid lines to lower bounds of pointwise identified sets; $\alpha_1 = 0.01$ is green, $\alpha_1 = 0.05$ (baseline) is blue, and $\alpha_1 = 0.09$ is red. Bottom panels: average width of confidence bands (green, blue, red) and width of population identified set (black).

Adjusting α , Keeping α_1 Fixed. Several authors devised methods to overcome the conservativeness of Bonferroni confidence intervals by raising the nominal level α to target a desired actual coverage probability $1 - \alpha_* = 0.90$, say. Examples of this approach are Campbell and Yogo (2006) and, most recently, McCloskey (2017). The former paper reduces the size of the first-stage confidence interval by raising α_1 , keeping the size of the second-stage intervals, CS_q^θ in our notation constant. The latter paper proposes to reduce the size of the second-stage intervals, keeping α_1

Figure 4: The Effect of Varying α_2 , Fixed $\alpha_1 = 0.05$



Notes: Top panels: actual coverage probabilities for 90% confidence sets. Dashed lines refer to upper limits and solid lines to lower limits of pointwise identified sets; $\alpha_2 = 0.05$ (baseline) is blue, $\alpha_2 = 0.10$ is green, and $\alpha_2 = 0.15$ is red. Bottom panels: average width of confidence bands (green, blue, red) and width of population identified set (black).

constant. In view of the results depicted in Figure 3, we informally follow McColskey’s (2017) approach by increasing α_2 from 0.05 (baseline) to 0.10 and 0.15, respectively. While, in principle, we could choose a different α_2 for each variable and each horizon, the results depicted in Figure 4 are generated using the same α_2 for each response. As expected, the actual coverage probability falls as we increase α_2 (and thereby $\alpha = \alpha_1 + \alpha_2$). For the $y_{1,t}$ responses the coverage probabilities remain above 90%, while for the long-horizon responses of $y_{2,t}$ the coverage probability drops substantially below 90% for $h \geq 15$ and $\alpha_2 = 0.10$. The attainable reduction in the width of the confidence band

is larger than in the case of fixed α , but it remains small relative to the overall width of the bands.

7 Empirical Illustration

We now apply the previously developed methods to a four-variable VAR. The vector of observables consists of per capita real GDP (in deviations from a linear trend), inflation, the federal funds rate, and real money balances. We use quarterly U.S. data from 1965:I to 2006:IV, excluding the 2007-09 recession and the subsequent period of zero nominal interest rates. A detailed description of the data set is provided in the Online Appendix. All VARs are estimated with $p = 2$ lags, which is the preferred lag length according to BIC. We will consider two set-identification schemes for monetary policy shocks. The first scheme involves only sign restrictions (Section 7.1), whereas the second identification is based on a combination of equality and sign restrictions (Section 7.2). As in Monte Carlo Experiment 3, we set $n_\Lambda = 1,000$, $n_Q = 20,000$, and $n_Z = 1,000$.

In addition to computing Bonferroni confidence bands, we also generate pointwise Bayesian credible intervals for the impulse responses, which have been widely used in empirical research. The Bayesian credible sets reported subsequently are based on the VAR(p) given in (1) with Gaussian innovations $u_t \sim iidN(0, \Sigma_u)$. Let $A = [A_1, \dots, A_p]'$ and define the unnormalized vector \tilde{q} such that $q = \tilde{q}/\|\tilde{q}\|$. If $\tilde{q} \sim N(0, I_n)$, then q is uniformly distributed on the hypersphere. Following Uhlig (2005), we use an improper prior of the form

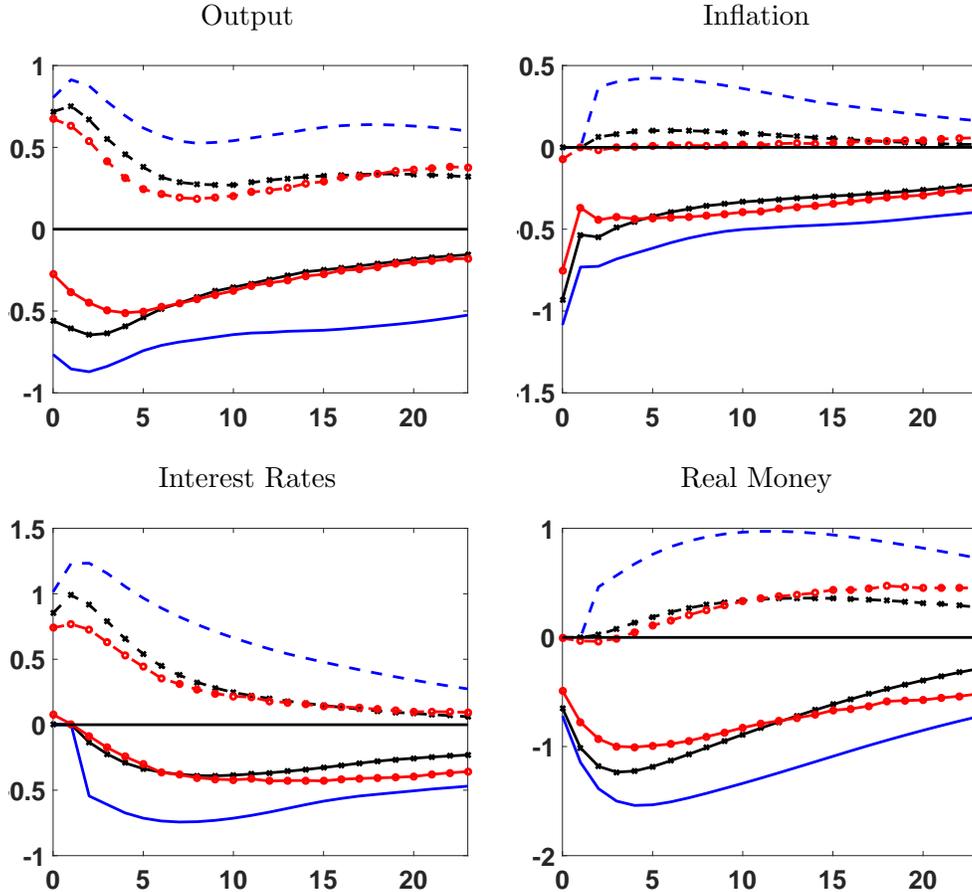
$$p(A, \Sigma, \tilde{q}) \propto |\Sigma|^{-(n+1)/2} \exp\{-\tilde{q}'\tilde{q}/2\} \mathcal{I} \left\{ \frac{\tilde{q}}{\|\tilde{q}\|} \in F^q(\phi(A, \Sigma)) \right\}. \quad (32)$$

We use the acceptance sampler described in Uhlig (2005) to generate 50,000 draws from the posterior distribution of (A, Σ, \tilde{q}) . These draws are then converted into impulse responses and credible sets are computed from the impulse response draws.

7.1 Pure Sign Restrictions

In order to make inference about the effects of a contractionary monetary policy shock, we use the following sign restrictions to bound the identified set: in periods $h = 0, 1$ (i) the interest rate response is weakly positive; (ii) the inflation response is weakly negative; and (iii) real money balances do not rise above their steady-state level. These sign restrictions were also used in Monte Carlo Experiment 3 in Section 6.3.

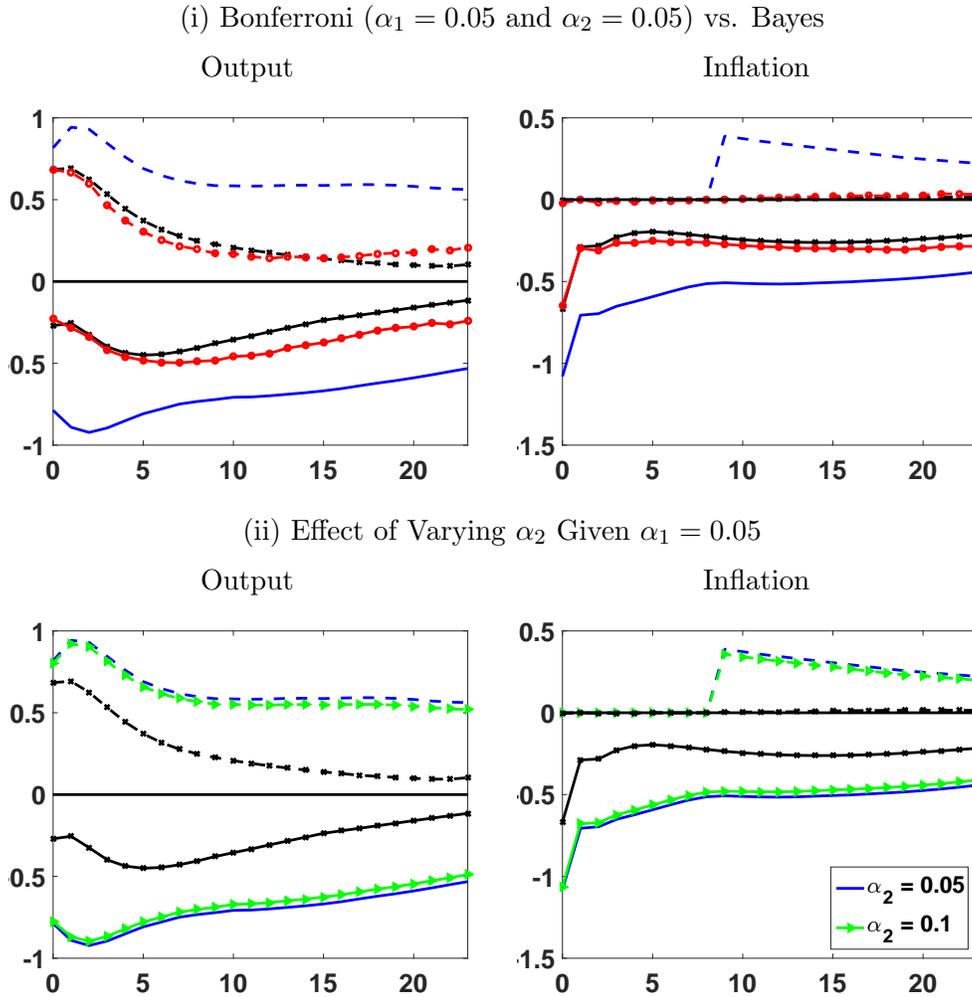
Figure 5: Pure Sign Restrictions over Horizons $h = 0, 1$



Notes: The figure depicts 90% Bonferroni confidence bands $CS^\theta(I)$ (blue) with $\alpha_1 = \alpha_2 = 0.05$; 90% Bayesian credible bands (red with circles); and the estimated sets $F^\theta(\hat{\phi})$ (black with crosses).

Figure 5 depicts three bands: (pointwise) 90% Bonferroni confidence intervals (using a diagonal weight matrix) $CS^\theta(I)$, estimated sets $F^\theta(\hat{\phi})$, and (pointwise) 90% Bayesian credible sets. The two most notable features of the bands are that the frequentist confidence bands (solid) are substantially wider than the Bayesian credible bands (short dashes) and that the Bayesian credible bands approximately coincide with the estimated set $F^\theta(\hat{\phi})$. As explained in detail in Moon and Schorfheide (2012), in a large sample, i.e., a sample in which uncertainty about ϕ is small compared to the size of $F^\theta(\hat{\phi})$, the Bayesian intervals lie inside the estimated sets $F^\theta(\hat{\phi})$ because in the limit essentially all of the probability mass is concentrated on $F^\theta(\hat{\phi})$ and a 90% credible interval is always a subset of the support of the posterior distribution. The frequentist interval, on the other hand, has to extend beyond the boundaries of $F^\theta(\hat{\phi})$ because it has to have, say, 90% coverage probability

Figure 6: Pure Sign Restrictions over Horizon $h = 0, 1, \dots, 8$



Notes: Top panels: 90% Bonferroni confidence bands $CS^\theta(I)$ (blue); 90% Bayesian credible bands (red with circles); and the estimated sets $F^\theta(\hat{\phi}_q, \hat{\phi}_\theta)$ (black with crosses). Bottom panels: Bonferroni confidence bands $CS^\theta(I)$ (blue) with $\alpha_2 = 0.05$; Bonferroni confidence bands $CS^\theta(I)$ (green with triangles) $\alpha_2 = 0.10$; and the estimated sets $F^\theta(\hat{\phi})$ (black with crosses).

for every element of the identified set $F^\theta(\phi)$, including the boundary points. From a substantive perspective, the use of sign restrictions leaves the direction of the output response undetermined.

The top panels of Figure 6 show output and inflation responses obtained by requiring the sign-restrictions hold for periods $h = 0, 1, \dots, 8$, keeping $\alpha_1 = \alpha_2 = 0.05$. This modification increases the number of inequality restrictions from 6 to 27. As the number of sign restrictions increases, the width of the identified sets decreases. As suggested by the Monte Carlo simulations,

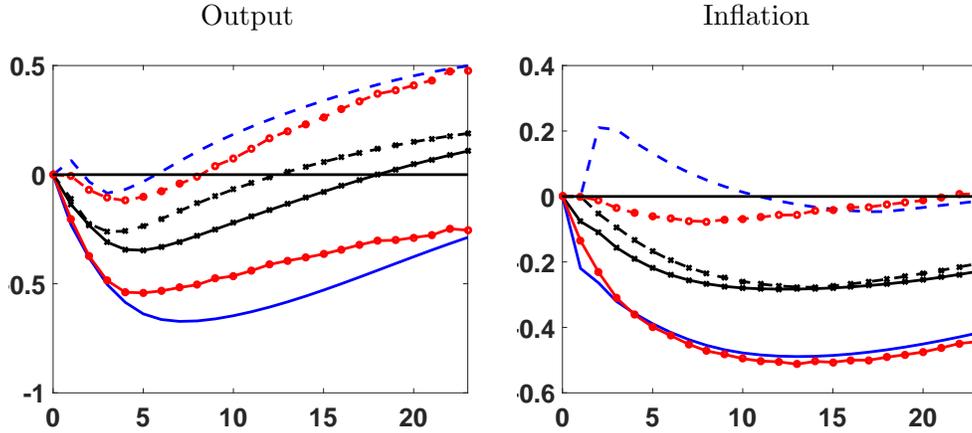
the width of the Bonferroni bands also decreases. The bottom panels of Figure 6 show the effect of raising α_2 from 0.05 to 0.10. According to the simulations in Section 6.3, this decrease in the nominal coverage probability brings the actual coverage probability closer to the desired coverage probability of 90%. As a result the width of the confidence bands shrinks, but not by much. In fact, in percentage terms, the width reduction is very small.

7.2 Combining Sign Restrictions and Zero Restrictions

A commonly used identification assumption for monetary policy shocks is that private-sector variables such as output and inflation cannot respond to changes in the federal funds rate within the period; see, for instance, Christiano, Eichenbaum, and Evans (1999). Because 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 reduce the dimension of the vector q as follows: $q = [0, 0, \cos \varphi, \sin \varphi]'$, where $\varphi \in [0, 2\pi]$. This is more efficient than adding two equality conditions to the set of inequality conditions; see Section 5. The zero restriction on the instantaneous inflation response replaces the sign restriction used in Section 7.1. We maintain the other sign restrictions used previously, that is, the interest rate responses for $h = 0$ and $h = 1$ are weakly positive and the inflation response in period $h = 1$ as well as the real money balance responses in periods $h = 0$ and $h = 1$ are weakly negative.

Impulse response bands are depicted in Figure 7. A comparison of $F^\theta(\hat{\phi})$ in Figures 5 and 7 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 coverage 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.

Figure 7: Combining Zero and Sign Restrictions over Horizons $h = 0, 1$



Notes: The figure depicts 90% Bonferroni confidence bands $CS^\theta(I)$ (blue) with $\alpha_1 = \alpha_2 = 0.05$; 90% Bayesian credible bands (red with circles); and the estimated sets $F^\theta(\hat{\phi})$ (black with crosses).

8 Conclusion

With the exception of FRSW, the coverage bands for impulse responses of sign-restricted SVARs that have been reported in the literature thus far were only meaningful from a Bayesian perspective. The main contribution of our paper is to develop an easy-to-use frequentist method based on the Bonferroni approach to construct confidence intervals for impulse responses and other measures of the dynamic effects of structural shocks in VARs that are set-identified based on sign restrictions. In the first stage, a confidence set for the vector of weights q on the reduced-form impulse responses is obtained by inverting a point-wise hypothesis test for the moment inequalities implied by the sign restrictions. We employ the Andrews and Soares (2010) moment selection procedure to obtain critical values for this test that are not diluted by non-binding inequality conditions. Our empirical application illustrates that in set-identified VARs, frequentist confidence bands can be substantially wider than Bayesian credible bands. As a by-product, we establish the consistency of the plug-in estimator $F^\theta(\hat{\phi})$ of the identified set of impulse responses. $F^\theta(\hat{\phi})$ is also useful from a Bayesian perspective. Because in a Bayesian analysis, the prior distribution of the impulse response functions conditional on the reduced-form parameters does not get updated, it is useful to report the identified set and the prior conditional on some estimate of ϕ , say, the posterior mean, so that the audience can judge whether the conditional prior distribution is highly concentrated in a particular area of the identified set.

References

- ANDREWS, D. W. K., AND P. J. BARWICK (2012): “Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,” *Econometrica*, 80, 2805–2826.
- ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): “Validity of Subsampling and ‘Plug-in Asymptotics’ Inference for Parameters Defined by Moment Inequalities,” *Econometric Theory*, 25, 669–709.
- ANDREWS, D. W. K., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78, 119–157.
- ARUOBA, B., AND F. SCHORFHEIDE (2011): “Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs,” *American Economic Journal: Macroeconomics*, 3, 60–90.
- BAUMEISTER, C., AND G. PEERSMAN (2013): “Time-Varying Effects of Oil Supply Shocks on the U.S. Economy,” *American Economic Journal: Macroeconomics*, 5(4), 1–28.
- BORDER, K. C. (2007): “Alternative Linear Inequalities,” Manuscript, California Institute of Technology.
- (2010): “Introduction to Correspondences,” Manuscript, California Institute of Technology.
- BUGNI, F. A., I. A. CANAY, AND X. SHI (2017): “Inference For Subvectors And Other Functions Of Partially Identified Parameters In Moment Inequality Models,” *Quantitative Economics*, 8(1), 1–38.
- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient tests of stock return predictability,” *Journal of Financial Economics*, 81, 27–60.
- CANAY, I. A., AND A. M. SHAIKH (2017): “Practical and Theoretical Advances in Inference for Partially-Identified Models,” in *Proceedings of the 11th World Congress of the Econometric Society*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, vol. forthcoming. Cambridge University Press, Cambridge.
- CANOVA, F., AND G. DE NICOLO (2002): “Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7,” *Journal of Monetary Economics*, 49, 1131–1159.

- CAVANAGH, C., G. ELLIOTT, AND J. STOCK (1995): “Inference in Models with Nearly Integrated Regressors,” *Econometric Theory*, 11, 1131–1147.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75, 1243–1284.
- CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999): “Monetary Policy Shocks: What Have we Learned and to What End?,” in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1a, pp. 65–148. North Holland.
- DEDOLA, L., AND S. NERI (2007): “What Does a Technology Shock Do? A VAR Analysis with Model-Based Sign Restrictions,” *Journal of Monetary Economics*, 54, 512–549.
- FAUST, J. (1998): “The Robustness of Identified VAR Conclusions about Money,” *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.
- FAUST, J., J. ROGERS, E. SWANSON, AND J. WRIGHT (2003): “Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data,” *Journal of the European Economic Association*, 1, 1031–1057.
- FAUST, J., E. SWANSON, AND J. WRIGHT (2004): “Identifying VARs Based on High Frequency Futures Data,” *Journal of Monetary Economics*, 51, 1107–1131.
- GAFAROV, B., M. MEIER, AND J. L. MONTIEL OLEA (2016a): “Delta-Method Inference for a Class of Set-Identified SVARs,” *Manuscript, Columbia University*, Manuscript.
- (2016b): “Projection Inference for Set-Identified SVARs,” *Manuscript, Columbia University*, Manuscript.
- GAMBETTI, L., AND A. MUSSO (2017): “Loan Supply Shocks and Business Cycle,” *Journal of Applied Econometrics*, 32, 764–782.
- GIACOMINI, R., AND T. KITAGAWA (2015): “Robust Inference About Partially-Identified SVARs,” *Manuscript, UCL*.
- GIRAITIS, L., AND P. C. PHILLIPS (2004): “Uniform Limit Theory for Stationary Autoregression,” Cowles Foundation Discussion Paper 1475.

- HRISTOV, N., O. HÜLSEWIG, AND T. T. WOLLMERSHÄUSER (2012): “Loan Supply Shocks during the financial crisis: Evidence for the euro area,” *Journal of International Money and Finance*, 31, 569–592.
- KAIDO, H., F. MOLINARI, AND J. STOYE (2016): “Confidence Intervals for Projections of Partially Identified Parameters,” *Manuscript, Cornell University*.
- KILIAN, L. (1998): “Small-Sample Confidence Intervals for Impulse Response Functions,” *Review of Economics and Statistics*, 80, 218–230.
- KILIAN, L., AND H. LÜTKEPOHL (2017): *Structural Vector Autoregressive Analysis*. Cambridge University Press.
- KILIAN, L., AND D. MURPHY (2012): “Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models,” *Journal of the European Economic Association*, 10, 1166–1188.
- LÜTKEPOHL, H. (1990): “Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models,” *Review of Economic and Statistics*, 72(1), 116–125.
- MANSKI, C. (2003): *Partial Identification of Probability Distributions*. Springer Verlag, New York.
- MCCLOSKEY, A. (2017): “Bonferroni-based size-correction for nonstandard testing problems,” *Journal of Econometrics*, forthcoming.
- MIKUSHEVA, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75, 1411–1452.
- MITTNIK, S., AND P. A. ZADROZNY (1993): “Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Models,” *Econometrica*, 61(4), 857–870.
- MOON, H. R., AND F. SCHORFHEIDE (2012): “Bayesian and frequentist inference in partially identified models,” *Econometrica*, 80(755–782).
- MOON, H. R., F. SCHORFHEIDE, AND E. GRANZIERA (2013): “Inference for VARs Identified with Sign Restrictions,” *Manuscript, University of Pennsylvania*.

- MOON, H. R., F. SCHORFHEIDE, E. GRANZIERA, AND M. LEE (2011): “Inference for VARs Identified with Sign Restrictions,” *NBER Working Paper*, 17140.
- MOUNTFORD, A., AND H. UHLIG (2009): “What are the Effects of Fiscal Policy Shocks?,” *Journal of Applied Econometrics*, 24, 960–992.
- PAPPA, E. (2009): “The Effects of Fiscal Shocks on Employment and the Real Wage,” *International Economic Review*, 50, 217–244.
- PEERSMAN, G. (2005): “What Caused the Early Millennium Slowdown? Evidence Based on Vector Autoregressions,” *Journal of Applied Econometrics*, 20, 185–207.
- PEERSMAN, G., AND R. STRAUB (2009): “Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR,” *International Economic Review*, 50, 727–750.
- ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): “A practical two-step method for testing moment inequalities,” *Econometrica*, 82, 1979–2002.
- ROSEN, A. (2008): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” *Journal of Econometrics*, 146, 107–117.
- RUBIO-RAMIREZ, J. F., D. WAGGONER, AND T. ZHA (2010): “Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference,” *Review of Economic Studies*, 77, 665–696.
- RUNKLE, D. E. (1987): “Vector Autoregressions and Reality,” *Journal of Business & Economic Statistics*, 5(4), 437–442.
- SIMS, C. A. (1980): “Macroeconomics and Reality,” *Econometrica*, 48, 1–48.
- UHLIG, H. (2005): “What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure,” *Journal of Monetary Economics*, 52, 381–419.
- YILDIZ, N. (2012): “Consistency of Plug-in estimators of upper contour and level set,” *Econometric Theory*, 28, 309–327.

Online Technical Appendix

This Online Appendix accompanies the paper “Inference for VARs Identified with Sign Restrictions” by E. Granziera, H.R. Moon, and F. Schorfheide. In Section A we provide proofs for the theoretical results in Section 4 of the main paper. Additional technical Lemmas are stated and proved in Section B. Section C provides analytical derivations for the Monte Carlo experiment presented in Section 6 of the main text. Section D contains additional information about the empirical analysis.

A Proofs of Main Results

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

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

instead of

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

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

A.1 Proof of Lemma 1

To simplify the notation we omit tildes and write $S_\theta(q)$, and $S(q)$ instead of $\tilde{S}_\theta(q)$ and $\tilde{S}(q)$.

Convexity: Suppose $\theta_i \in F^\theta(\phi_q, \phi_\theta)$, $i = 1, 2$, and $\theta_1 < \theta_2$. Then there exist q_i with $\|q_i\| = 1$ and $\mu_i \geq 0$ such that

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

We distinguish two cases: $q_1 \neq -q_2$ and $q_1 = -q_2$.

Case (i): Suppose that $q_1 \neq -q_2$. We now verify that for any $\lambda \in [0, 1]$ $\theta = \lambda\theta_1 + (1 - \lambda)\theta_2 \in F^\theta(\cdot)$. 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) implies that

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

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

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

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

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

The first equality follows from the linearity of $S(q)$, the second equality is implied by (A.1), and the inequality follows from $\mu_i \geq 0$. Thus, $\theta \in F^\theta(\phi_q, \phi_\theta)$.

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 $S(q_3)\phi \geq 0$. Let $\theta_3 = S_\theta(q_3)\phi$. By construction, $\theta_3 \in F^\theta(\cdot)$. If $\theta_3 = \theta_1$ ($\theta_3 = \theta_2$) we simply replace q_1 (q_2) by q_3 and follow the steps outlined for Case (i). If $\theta_1 < \theta_3 < \theta_2$, then the Case (i) argument implies that any θ in the intervals $[\theta_1, \theta_3]$ and $[\theta_3, \theta_2]$ and thereby any $\theta = \lambda\theta_1 + (1 - \lambda)\theta_2$ is included in the identified set. Finally, if $\theta_3 < \theta_1$ ($\theta_2 < \theta_3$), we deduce from Case (i) that the interval $[\theta_3, \theta_2]$ ($[\theta_1, \theta_3]$) is included in the identified set.

Boundedness: We shall prove a slightly more general result. Throughout the proof we omit tildes. Suppose that $\tilde{\theta} \in F^\theta(\phi_q, \phi_\theta)$. Since $F^\theta(\phi_q, \phi_\theta)$ is a multiple-value set, we assume without loss of generality that $\tilde{\theta} > 0$. So, the sign restriction $\theta \geq 0$ is satisfied if it exists. Define

$$G^\theta(\theta; \phi_q, \phi_\theta) = \min_{q=\|1\|, \mu \geq 0} \|S_\theta(q)\phi_\theta - \theta\|^2 + \|S(q)\phi_q - \mu\|^2$$

such that $G^\theta(\theta; \phi_q, \phi_\theta) = 0$ if and only if $\theta \in F^\theta(\phi_q, \phi_\theta)$. We now show by contradiction that $F^\theta(\phi_q, \phi_\theta)$ has an upper bound.

Suppose, to the contrary, that no such upper bound exists. This guarantees the existence of a series $a_n > 0$ with $a_n \uparrow \infty$ such that $a_n\tilde{\theta}_n \in F^\theta(\phi_q, \phi_\theta)$ for each n . Consider the bound

$$G^\theta(a_n\tilde{\theta}; \phi_q, \phi_\theta) \geq \min_{q=\|1\|} \|S_\theta(q)\phi_\theta - a_n\tilde{\theta}\|^2.$$

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

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

which contradicts the requirement $G^\theta(\theta; \phi_q, \phi_\theta) = 0$. 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. \square

A.2 Proof of Theorem 1

Recall the definition $F^q(\Phi_q) = \{q \in \mathbb{S}^n \mid \Phi'_q q \geq 0\}$. Thus, $\{q \in \mathbb{S}^n \mid \Phi'_q q \gg 0, \} \subset F^q(\phi)$. The statement of the theorem follows once we have shown that there exists a non-empty, non-singleton, n -dimensional subset \mathbb{Q} of \mathbb{S}^n , such that $\Phi_q q \gg 0$ if $q \in \mathbb{Q}$.

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

$$\Phi'_q x \gg 0.$$

Let

$$\epsilon = \Phi'_q x \gg 0.$$

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

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

$$\Phi'_q = \underbrace{U}_{r \times r} \underbrace{D}_{r \times n} \underbrace{V'}_{n \times n} = \underbrace{U_1}_{r \times d} \underbrace{D_1}_{d \times d} \underbrace{V'_1}_{d \times n},$$

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

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

In addition, let

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

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

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

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

In turn,

$$0 = c_1 \Phi_q' x_1 + \dots + c_d \Phi_q' x_d = c_1 \epsilon_1 + \dots + c_d \epsilon_d.$$

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

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

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

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

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

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

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

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

A.3 Proof of Theorem 2, Part (i)

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

The proof of the theorem exploits the continuity of $F^q(\phi)$ with respect to ϕ . The statement of the theorem is a consequence of Lemma 2, Lemma 3, and the continuous mapping theorem. \square

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

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

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

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

Lemma 3 *Suppose that Assumption 1(i) is satisfied. Then,*

- (i) $F^q(\Phi_q)$ is compact for all Φ_q ;
- (ii) $F^q(\Phi_q)$ is continuous at $\Phi_{q,0}$.

Proof of Lemma 3: For notational simplicity, we omit the subscript notation q and write Φ_q as Φ . Let $\mathbb{S}^n = \{q \in \mathbb{R}^n : \|q\| = 1\}$ be the unit sphere in \mathbb{R}^n . Recall from Theorem 1 that $F^q(\Phi)$ is nonempty.

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

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

Closedness: Consider any sequence $q_j \in F^q(\Phi)$ such that $q_j \rightarrow q_0$, where $\|q_j\| = 1$ and $\|q_0\| = 1$. Then, $0 \leq \Phi'q_j \rightarrow \Phi'q_0$, so that it should be $\Phi'q_0 \geq 0$. This implies that $q_0 \in F^q(\Phi)$, as required for closedness.

Part (ii): We show $F^q(\Phi)$ is upper hemi-continuous (UHC) and lower hemi-continuous (LHC) at Φ_0 .

UHC: Since $F^q(\Phi)$ is non-empty and compact-valued, the UHC of $F^q(\Phi)$ at Φ_0 follows if we show that for every sequence $\Phi_j \rightarrow \Phi_0$ and $q_j \in F^q(\Phi_j)$, there exists a subsequence q_{j_i} of q_j such that $q_{j_i} \rightarrow q_0 \in F^q(\Phi_0)$. (See Border (2010) Proposition 20). Since $\{q_j\} \subset \mathbb{S}^n$ and \mathbb{S}^n is compact, we can choose a convergent subsequence q_{j_i} such that $q_{j_i} \rightarrow q_0$. Then, $0 \leq \Phi'_{j_i} q_{j_i} \rightarrow \Phi'_0 q_0$, and it follows that $\Phi'_0 q_0 \geq 0$. This implies that $q_0 \in F^q(\Phi_0)$, as required.

LHC: $F^q(\Phi)$ is LHC at Φ_0 if and only if for any sequence $\{\Phi_j\}$ with $\Phi_j \rightarrow \Phi_0$ and $q_0 \in F^q(\Phi_0)$, there exists a sequence $q_j \in F^q(\Phi_j)$ with $q_j \rightarrow q_0$. We re-order and partition the matrix Φ_0 to $\Phi_0 = [\Phi_1, \Phi_2]$, where $\Phi'_1 q_0 = 0$ and $\Phi'_2 q_0 \gg 0$.

For a matrix A , we denote the l^{th} column of A as $(A)_l$. By Gordan's Alternative Theorem, see Border (2007), Assumption 1(i) implies that there exists a $\xi^* \in \mathbb{R}^m$ such that

$$\Phi'_1 \xi^* \gg 0.$$

Let

$$\xi = \frac{1}{\min_l (\Phi_1)'_l \xi^*} \xi^*$$

such that for all l

$$(\Phi_1)'_l \xi > 1.$$

Set $\epsilon_{j,l} = \|(\Phi_j - \Phi_0)_l\|$ and $\epsilon_j = \max_l \{\epsilon_{j,l}\}$, and define

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

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

Case (i): Suppose $(\Phi_0)'_l q_0 = 0$. Then, when j is large so that $(\Phi_0)'_l \xi - 1 \geq \epsilon_j \|\xi\|$, we have

$$\begin{aligned} (\Phi_j)'_l q_j &= (\Phi_{r,j} - \Phi_{r,0})'_l q_j + (\Phi_0)'_l q_j \\ &= \frac{1}{\|q_0 + \epsilon_j \xi\|} \{(\Phi_j - \Phi_0)'_l q_0 + \epsilon_j (\Phi_j - \Phi_0)'_l \xi + (\Phi_0)'_l q_0 + \epsilon_j (\Phi_0)'_l \xi\} \\ &\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} \{-\|(\Phi_j - \Phi_0)_l\| \|q_0\| - \epsilon_j \|(\Phi_j - \Phi_0)_l\| \|\xi\| + \epsilon_j (\Phi_0)'_l \xi\} \\ &\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} (-\epsilon_j - \epsilon_j^2 \|\xi\| + \epsilon_j (\Phi_0)'_l \xi) \\ &= \frac{1}{\|q_0 + \epsilon_j \xi\|} \epsilon_j ((\Phi_0)'_l \xi - 1 - \epsilon_j \|\xi\|) \\ &\geq 0. \end{aligned}$$

Case (ii): Suppose $(\Phi_0)'_l q_0 > 0$. Then, since $\|(\Phi_0)_l\| \leq M$ (compact parameter set), we have

$$\begin{aligned}
(\Phi_j)'_l q_j &= (\Phi_j - \Phi_0)'_l q_j + (\Phi_0)'_l q_j \\
&= \frac{1}{\|q_0 + \epsilon_j \xi\|} \{(\Phi_j - \Phi_0)'_l q_0 + \epsilon_j (\Phi_j - \Phi_0)'_l \xi + (\Phi_0)'_l q_0 + \epsilon_j (\Phi_0)'_l \xi\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} \{-\|(\Phi_j - \Phi_0)_l\| \|q_0\| - \epsilon_j \|(\Phi_j - \Phi_0)_l\| \|\xi\| + (\Phi_0)'_l q_0 - \epsilon_j \|(\Phi_0)_l\| \|\xi\|\} \\
&\geq \frac{1}{\|q_0 + \epsilon_j \xi\|} ((\Phi_0)'_l q_0 - \epsilon_j - \epsilon_j^2 M - \epsilon_j M^2) \\
&\geq 0,
\end{aligned}$$

when j is large. The last inequality holds since $(\Phi_0)'_l q_0 > 0$.

From these, we can deduce that

$$\Phi'_j q_j \geq 0.$$

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

$$q_j \rightarrow q_0.$$

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

A.4 Proof of Theorem 2, Part (ii)

We closely follow the proofs of Theorem 1 and Lemma 2 of Andrews and Soares (2010). The main modification is to accommodate the reduced rank possibility of $\Sigma(q)$ and $D(q)$. The proof makes use of various lemmas that are stated and proved in Section B below. To simplify the notation we eliminate ρ from the formulas and index the probability distribution by $\phi \in \mathcal{P}$ instead of $\rho \in \mathcal{R}$. We also skip the subscription notation q and write, for example, $\phi_q, \hat{\phi}_q, \Lambda_{qq}, D_q$ as $\phi, \hat{\phi}, \Lambda, D$, respectively. Thus we write

$$\inf_{\phi \in \mathcal{P}} \inf_{q \in F^q(\phi)} P_\phi \{\theta \in CS^q(\hat{\phi})\} \quad \text{instead of} \quad \inf_{\rho \in \mathcal{R}} \inf_{q \in F^q(\phi(\rho))} P_\rho \{q \in CS^q(\hat{\phi}_q)\}.$$

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

We need to show

$$\liminf_T \inf_{\phi \in \mathcal{P}} \inf_{q \in F^q(\phi)} P_\phi \{q \in CS^q\} \geq 1 - \alpha. \quad (\text{A.2})$$

Let

$$AsyCP = \liminf_T \inf_{\phi \in \mathcal{P}} \inf_{q \in F^q(\phi)} P_\phi \{q \in CS^q\}.$$

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

$$AsyCP = \liminf_T P_{\phi_T} \{q_T \in CS^q\}.$$

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

$$AsyCP = \lim_{T'} P_{\phi_{T'}} \{q_{T'} \in CS^q\}.$$

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

$$\lim_{T''} P_{\phi_{T''}} \{q_{T''} \in CS^q\} \geq 1 - \alpha. \quad (A.3)$$

Then, the desired result (A.2) follows and the proof of the theorem is complete.

Define

$$\mu(q, \phi) = S(q) \phi$$

and decompose

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

Moreover, let

$$A(q) = L'S'(q)D^{-1/2}(q).$$

To simplify the notation we suppress the dependence of matrices on ϕ . The matrix $\Omega(q) = A'(q)A(q)$ is a correlation matrix and $D^{1/2}(q)$ is a diagonal matrix of standard deviations. Also, recall that $W(q) = D^{1/2}(q)B(q)D^{1/2}(q)$, where either $B(q) = \Omega^{-1}(q)$ or $B(q) = I$. The proof is completed in three steps.

Step 1: Choosing the subsequence T'' . We choose a subsequence T'' from T' along which the subsequent conditions are satisfied. This is done sequentially by choosing a subsequence that satisfies criterion (i), and then, step-by-step choosing subsequences of the subsequences to satisfy the next criterion until all five conditions are satisfied:

- (i) $\phi_{T''} \rightarrow \phi$.
- (ii) $r(q_{T''}) = r$, $V(q_{T''}) = V$ for all T'' .
- (iii) For $j = 1, \dots, r$, the slackness (recall that $\mu_j = [S(q)\phi_q]_j \mathcal{I}\{[S(q)\phi_q]_j \geq 0\}$) in inequality j converges to

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

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

(iv) The sequence $A(q_{T''})$ has a full rank limit, denoted by A .

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

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

Condition (ii) is satisfied since $r(q_{T'})$ and $V(q_{T'})$ are sequences that take only a finite number of discrete values. Condition (iii) is satisfied because the range of the sequences of interest is $[0, \infty]$ and by a similar argument used in the proof of Theorem 1 of Andrews and Soares (2010b). Roughly speaking, in Case (iii)-(a) the slackness is small and the selection criterion regards the inequality asymptotically as binding. In Case (iii)-(c) the slackness is large and the selection criterion regards the inequality as non-binding and (iii)-(b) is an intermediate case. Condition (iv) is satisfied according to Lemma B 2. If Condition (iv) is satisfied, then the following conditions also hold ((vii) is a consequence of Lemma B 2):

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

We now reorder the rows of $S(q_{T''})$ such that $\pi_j = 0$ for rows $j = 1, \dots, r_1$ and $\pi_j > 0$ for rows $j = r_1 + 1, \dots, r$. Along the sequence T'' , the last $r_2 = r - r_1$ restrictions correspond to non-binding moment inequalities. In the subsequent steps we show that the inequality selection procedure used in the critical-value computation in (28) asymptotically underestimates r_2 (and thereby overestimates r_1), which makes the critical value asymptotically conservative to achieve the uniform coverage requirement.

Step 2: Constructing an upper bound for the critical value $c^\alpha(q)$ in (28). For notational simplicity we use sequence notation $\{T\}$ for the subsubsequence $\{T''\}$ in Step 1. Recall the definitions

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

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

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

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

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

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

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

Finally, define the vector π_* with elements

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

To characterize the critical values, define the objective function

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

Note that the notation in the proof is slightly different from the notation in the main text. In (27) of the main text we defined $\bar{\mathcal{G}}(q; \hat{B}(q), M_{\hat{\xi}}(q))$, which corresponds to $\bar{\mathcal{G}}(q_T; \hat{A}(\cdot), \hat{B}(\cdot), \hat{\varphi}(\cdot))$ in this proof. We let

$$c_T^\alpha(A(\cdot), B(\cdot), \varphi(\cdot)) = (1 - \alpha) \text{ quantile of } \bar{\mathcal{G}}(q_T; A(\cdot), B(\cdot), \varphi(\cdot)). \quad (\text{A.4})$$

To cover the special case $r = r_2 > 0$, i.e., all the inequality conditions are non-binding, we adopt the convention that

$$c_T^\alpha(A(\cdot), B(\cdot), \varphi(\cdot)) = 0 \quad (\text{A.5})$$

if $\varphi(q_T) = \hat{\varphi}_T^*(q_T)$ or $\varphi(q_T) = \pi^*$. The critical value $c^\alpha(q)$ in (28) in the main text can be expressed as

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

Notice by definition that

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

This implies that

$$c_T^\alpha(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T^*(q_T)) \leq c_T^\alpha(\hat{A}(q_T), \hat{B}(q_T), \hat{\varphi}_T(q_T)). \quad (\text{A.6})$$

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

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

We consider two different cases: (i) some inequalities are “binding”, i.e., $r_1 > 0$; (ii) all inequalities are “non-binding”, i.e., $r_1 = 0$.

Step 3(i). Suppose that $r_1 > 0$. By Lemma B 1 and (A.6), we have

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

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

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

The last inequality holds because $h \geq \pi^*$. (This is true because $\pi_j = 0$ implies that $h_j < \infty$ and $\pi_j^* = 0$, while $\pi_j > 0$ implies that $h_j = \pi_j^* = \infty$.) According to Lemma B 3,

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

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

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

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

$c_T^\alpha(A, B, \pi^*) = 0$. Then, by Lemma B 1 and (A.6), we have

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

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

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

A.5 Proof of Theorem 3, Part (i)

The proof of the theorem exploits the continuity of $F^q(\phi_q)$ with respect to ϕ_q . The statement of the theorem is a consequence of Lemma 2, Lemma 4, and the continuous mapping theorem. \square

Lemma 4 *Suppose that Assumptions 1(i) and 2(ii) are satisfied. Then,*

- (i) $F^\theta(\Phi_q, \Phi_\theta)$ is compact for all Φ_q ;
- (ii) $F^\theta(\Phi_q, \Phi_\theta)$ is continuous at $(\Phi_{q,0}, \Phi_{\theta,0})$.

Proof of Lemma 4, Part (i). Since $F^\theta(\Phi_q, \Phi_\theta) \subset \mathbb{R}^k$, for the required result, we show that $F^\theta(\Phi_q, \Phi_\theta)$ is closed and bounded.

Boundedness: The set $\{\theta = f(\Phi_\theta, q) : \|q\| = 1\}$ is compact because $f(\cdot)$ is continuous in both of its arguments by Assumption 2(ii) and the domain of q , \mathbb{S}^n , is compact. Since $F^\theta(\Phi_q, \Phi_\theta) \subset \{\theta = f(\Phi_\theta, q) : \|q\| = 1\}$, we deduce that $F^\theta(\Phi_q, \Phi_\theta)$ is bounded.

Closedness: Consider any sequence $\theta_j \in F^\theta(\Phi_q, \Phi_\theta)$, $j = 1, 2, \dots$, such that $\theta_j \rightarrow \theta_0$. We show that $\theta_0 \in F^\theta(\Phi_q, \Phi_\theta)$, that is we need to find a $q_0 \in F^q(\Phi_q)$ such that $\theta_0 = f(\Phi_\theta, q_0)$. Then, the desired result follows.

For $\theta_j \in F^\theta(\Phi_q, \Phi_\theta)$, by definition we can choose a $q_j \in F^q(\Phi_q)$ such that $\theta_j = f(\Phi_\theta, q_j)$. Since $\{q_j\} \in \mathbb{S}^n$ and \mathbb{S}^n is compact, we can choose a convergent subsequence q_{j_i} such that $q_{j_i} \rightarrow q_0$. Then it follows from the continuity of $f(\cdot)$ that $f(\Phi_\theta, q_{j_i}) \rightarrow f(\Phi_\theta, q_0)$. Since the subsequence θ_{j_i} also converges to θ_0 , we have $f(\Phi_\theta, q_0) = \theta_0$. By definition of $F^\theta(\Phi_q, \Phi_\theta)$, then, we have $\theta_0 \in F^\theta(\Phi_q, \Phi_\theta)$, as required for closedness.

Part (ii). According to Assumption 2(i) the function $f(\Phi_\theta, q)$ is continuous. Then, the product correspondence

$$\tilde{F}^q(\Phi_q, \Phi_\theta) = F^q(\Phi_q) \times \Phi_\theta$$

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

$$F^\theta(\Phi_q, \Phi_\theta) = \bigcup_{q \times \Phi_\theta \in \tilde{F}^q(\Phi_q, \Phi_\theta)} f(\Phi_\theta, q).$$

Since both $f(\cdot)$ and $\tilde{F}^q(\cdot)$ are continuous, by Theorem 12.23 of Aliprantis and Border (2006), $F^\theta(\Phi_q, \Phi_\theta)$ is continuous. \square

A.6 Proof of Theorem 3, Part (ii)

Let $F^{\theta, q}(\phi_q, \phi_\theta) = \{\theta \in \Theta, q \in \mathcal{Q} : q \in F^q(\phi_q), \theta = f(\Phi_\theta, q)\}$. Notice that

$$\begin{aligned} & \liminf_T \inf_{\phi \in \mathcal{P}} \inf_{\theta \in F^\theta(\phi_q, \phi_\theta)} P_\phi \left\{ \theta \in CS^\theta(\hat{\phi}_q, \hat{\phi}_\theta) \right\} \\ & \geq \liminf_T \inf_{\phi \in \mathcal{P}} \inf_{(\theta, q) \in F^{\theta, q}(\phi_q, \phi_\theta)} P_\phi \left\{ q \in CS^q(\hat{\phi}_q), \theta \in CS_q^\theta(\hat{\phi}_\theta) \right\} \\ & \geq \liminf_T \inf_{\phi \in \mathcal{P}} \inf_{q \in F^q(\phi_q)} P_\phi \left\{ q \in CS^q(\hat{\phi}_q) \right\} + \liminf_T \inf_{\phi \in \mathcal{P}} \inf_{(\theta, q) \in F^{\theta, q}(\phi_q, \phi_\theta)} P_\phi \left\{ \theta \in CS_q^\theta(\hat{\phi}_\theta) \right\} - 1. \end{aligned}$$

Recall that $\theta = f(\Phi_\theta, q)$. According to Theorem 2(ii)

$$\liminf_T \inf_{\phi \in \mathcal{P}} \inf_{q \in F^q(\phi_q)} P_\phi \left\{ q \in CS^q(\hat{\phi}_q) \right\} \geq 1 - \alpha_1 \quad (\text{A.7})$$

and

$$\liminf_T \inf_{\phi \in \mathcal{P}} \inf_{(\theta, q) \in F^{\theta, q}(\phi_q, \phi_\theta)} P_\phi \left\{ \theta \in CS_q^\theta(\hat{\phi}_\theta) \right\} \geq 1 - \alpha_2. \quad (\text{A.8})$$

holds according to Assumption 2. \square

B Additional Technical Lemmas

Throughout this section we use the following notation. When A is a matrix, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the largest and the smallest eigenvalues of A , respectively. We denote A_k as the k^{th} column vector of A ; A^j as the j^{th} row vector of A ; and A_{jk} as the $(j, k)^{\text{th}}$ element of A . Throughout the proofs, we sometimes omit the ϕ_T argument from the asymptotic covariance matrix $\Lambda = LL'$ and

use the notation Λ_T , $\hat{\Lambda}_T$, L_T , and \hat{L}_T for simplicity. We also often omit the q_T argument for some of the matrices that depend on q_T and simply write, say, S_T , D_T , \hat{D}_T , A_T , \hat{A}_T , Ω_T , and $\hat{\Omega}_T$ for short.

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

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

Proof of Lemma B 1. According to condition (ii) in the proof of Theorem 2(ii) $V(q_T) = V$ for all T . If $V = 0$, i.e., $S(q_T) = 0$ for all T , it is trivial to deduce the required result because by definition

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

Now suppose that $V \neq 0$. Notice that $S_T \phi_T \geq 0$ and $D_T^{-1/2}$ and $\hat{D}_T^{-1/2}$ are well defined since S_T is a full (row) rank matrix and $\Lambda_T, \hat{\Lambda}_T > 0$. We now consider the two cases (i) $B_T = \hat{B}_T = I$ and (ii) $B_T = \Omega_T^{-1}$ and $\hat{B}_T = \hat{\Omega}_T^{-1}$ separately.

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

$$\begin{aligned} G\left(q_T; \hat{\phi}, \hat{W}(\cdot)\right) &= \min_{\mu \geq 0} T \left\| \hat{D}_T^{-1/2} S_T \hat{\phi} - \hat{D}_T^{-1/2} V_T \mu \right\|^2 \\ &= \min_{v \geq -\sqrt{T} \hat{D}_T^{-1/2} \mu(q_T, \phi_T)} \left\| \hat{D}_T^{-1/2} S_T \sqrt{T} (\hat{\phi} - \phi_T) - v \right\|^2 \end{aligned}$$

and

$$G\left(q; \hat{\phi}, W(\cdot)\right) = \min_{v \geq -\sqrt{T} D_T^{-1/2} \mu(q, \phi)} \left\| D_T^{-1/2} S_T \sqrt{T} (\hat{\phi} - \phi_T) - v \right\|^2,$$

where $\mu(q_T, \phi_T) = S_T \phi_T$. Define

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

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

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

The last equality holds because $\hat{A}_T - A_T = o_p(1)$, $A_T = O(1)$, $\hat{L}_T^{-1} - L_T^{-1} = o_p(1)$, $L_T^{-1} = O(1)$, $\sqrt{T}(\hat{\phi} - \phi_T) = O_p(1)$, and $\hat{B}_T \xrightarrow{p} B > 0$ according to Lemma B 2 and Assumption 1(v-vi).

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

$$\begin{aligned}
G\left(q_T; \hat{\phi}, \hat{W}(\cdot)\right) &= \min_{\mu \geq 0} T \left\| S_T \hat{\phi} - V_T \mu \right\|_{\hat{\Sigma}_T^{-1}}^2 \\
&= \min_{v \geq -\sqrt{T} \mu(q_T, \phi_T)} \left\| S_T \sqrt{T} \left(\hat{\phi} - \phi_T\right) - v \right\|_{\hat{\Sigma}_T^{-1}}^2
\end{aligned}$$

and

$$G\left(\theta, q; \hat{\phi}, W(\cdot)\right) = \min_{v \geq -\sqrt{T} \mu(q_T, \phi_T)} \left\| S_T \sqrt{T} \left(\hat{\phi} - \phi_T\right) - v \right\|_{\Sigma_T^{-1}}^2,$$

where $\mu(q_T, \phi_T) = S_T \phi_T$. Define

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

Then,

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

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

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

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

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

The last line holds because

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

Moreover,

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

This completes the proof for Case (ii). \square

Lemma B 2 *Suppose that a converging sequence $\{\phi_T, q_T\}$ satisfies the rank condition $r(q_T) = r > 0$ and $V(q_T)$ is a non-zero constant selection matrix for all T . Then, there exists a subsequence $\{T'\} \subset \{T\}$ such that along the subsequence, we have (i)*

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

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

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

Proof of Lemma B 2. Part (i): Recall that $S_T = V_T \tilde{S}_T$. The rank reduction of \tilde{S}_T is caused only by zero rows (see Section 2.3). Moreover, according to condition (ii) in the proof of Theorem 2(ii) the non-zero row selection matrix is V_T constant over T . Thus, we can construct an index set \mathcal{J} of non-zero rows of \tilde{S}_T . By construction, the size of \mathcal{J} is l and

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

In turn we obtain

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

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

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

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

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

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

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

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

$$AL^{-1} = \lim_{T' \rightarrow \infty} \left[\frac{\tilde{S}_{T'}^j}{\|\tilde{S}_{T'}^j L_{T'}\|} \right]_{j \in \mathcal{J}} \quad (\text{A.9})$$

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

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

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

for all T' . We will now show that

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

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

Write

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

We begin with the following bound:

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

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

According to Assumption 1(vi). $\hat{L}_{T'} \xrightarrow{p} L$. Moreover, we deduce from (A.9) and $A^j L^{-1} \neq 0$ that

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

Therefore,

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

Similarly, we obtain the bound

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

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

$$I = o_p(1).$$

Term II can be bounded as follows:

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

and so

$$II = o_p(1).$$

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

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

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

where the critical value function $c_T^\alpha(\cdot)$ is defined in (A.4) and (A.5).

Proof of Lemma B 3. The proof is very similar to that of Lemma 2(a) in Andrews and Soares (2010b) and we provide a sketch. The proof proceeds in three steps. First, show

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

Second, show

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

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

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

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

Notice that if $\pi_j = 0$, then $\xi_T(q_T) < \kappa_T$ as $T \rightarrow \infty$ and by using an argument similar to the one used in the proof of Lemma B 2(ii), we have $\hat{\xi}_T(q_T) < \kappa_T$ in probability as $T \rightarrow \infty$. Therefore, $\text{plim } \hat{\varphi}_{j,T}^*(q_T) = \text{plim } \hat{\varphi}_{j,T}(q_T) = 0 = \pi_j^*$ with probability one. On the other hand, if $\pi_j > 0$, then $\hat{\varphi}_{j,T}^*(q_T) = \infty = \pi_j^*$. Therefore, $\hat{\varphi}_T^*(q_T) \xrightarrow{p} \pi^*$.

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

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

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

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

C Description of Monte Carlo Experiments

C.1 Experiment 1: Bivariate VAR(0)

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

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

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

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

Estimating the Reduced-Form Parameters

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

Computing the Confidence Intervals

- *Step 1: Construct a $(1 - \alpha_1)$ confidence set for q .* The following computations are executed for each $q \in \mathcal{Q}$. As before, it is convenient to express q in terms of the angle φ and generate \mathcal{Q} by equally spaced grid points on the interval $[-\pi, \pi]$. Recall the definition of $\xi_{1,T}(q)$ and $\xi_{2,T}(q)$ in (24).

– If $\varphi \in \{-\frac{\pi}{2}, \frac{\pi}{2}\}$, the objective function is given by

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

- * If $\xi_{2,T}(q) < \kappa_T$, the inequality condition is considered binding and the critical value $c^{\alpha_1}(q)$ is the $(1 - \alpha_1)$ quantile of a squared truncated normal $Z^2 \mathcal{I} \{Z \geq 0\}$.
- * If $\xi_{2,T}(q) \geq \kappa_T$, the inequality condition is considered non-binding and $c^{\alpha_1}(q) = 0$.

– If $\varphi \notin \{-\frac{\pi}{2}, \frac{\pi}{2}\}$, the objective function is given by

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

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

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

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

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

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

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

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

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

- *Step 3: Construct the $1 - \alpha$ Bonferroni set for θ :* Compute the minimum of the lower bounds of CS_q^θ and the maximum of the upper bounds of CS_q^θ for $q \in CS^q$.

C.2 Experiment 2: Bivariate VAR(1)

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

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

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

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

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

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

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

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

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

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

Bonferroni Approach

- *Step 1: Construct a $1 - \alpha_1$ confidence set for q .*

– The objective function is

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

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

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

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

- *Step 2: Construct the $(1 - \alpha_2)$ confidence set for θ conditional on q .* Follow the same steps as in Experiment 1.

- *Step 3: Construct the $1 - \alpha$ Bonferroni set for θ .* Follow the same steps as in the Experiment 1.

C.3 Experiment 3: Four-Variable VAR(2)

Design. The coefficient matrices for the DGP are given by

$$A'_1 = \begin{bmatrix} 1.001 & -0.100 & 0.302 & -0.085 \\ 0.065 & 0.585 & 0.089 & -0.055 \\ 0.126 & 0.284 & 1.072 & -0.073 \\ 0.233 & 0.141 & 0.056 & 1.522 \end{bmatrix}, \quad A'_2 = \begin{bmatrix} -0.080 & 0.119 & -0.269 & 0.078 \\ -0.056 & 0.262 & 0.065 & 0.013 \\ -0.223 & -0.222 & -0.178 & 0.070 \\ -0.230 & -0.097 & -0.069 & -0.538 \end{bmatrix}$$

$$c = \begin{bmatrix} 0.626 \\ 0.175 \\ 0.064 \\ 0.204 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 0.542 & -0.124 & 0.199 & 0.095 \\ -0.124 & 1.164 & 0.129 & -0.369 \\ 0.199 & 0.129 & 0.912 & -0.263 \\ 0.095 & -0.369 & -0.263 & 0.549 \end{bmatrix}.$$

D Further Details on the Empirical Analysis

The construction of the data set follows Aruoba and Schorfheide (2011). Unless otherwise noted, the data are obtained from the FRED2 database maintained by the Federal Reserve Bank of St. Louis. Per capita output is defined as real GDP (GDPC96) divided by the civilian non-institutionalized population (CNP16OV). The population series is provided at a monthly frequency and converted to quarterly frequency by simple averaging. We take the natural log of per capita output and extract a deterministic trend by OLS regression over the period 1959:I to 2006:IV. The deviations from the linear trend are scaled by 100 to convert them into percentages. Inflation is defined as the log difference of the GDP deflator (GDPDEF), scaled by 400 to obtain annualized percentage rates. Our measure of nominal interest rates corresponds to the federal funds rate (FEDFUNDS), which is provided at monthly frequency and converted to quarterly frequency by simple averaging. We use the sweep-adjusted M2S series provided by Cynamon, Dutkowsky and Jones (2006). This series is recorded at monthly frequency without seasonal adjustments. The EViews default version of the X12 filter is applied to remove seasonal variation. The M2S series is divided by quarterly nominal GDP to obtain inverse velocity. We then remove a linear trend from log inverse velocity and scale the deviations from trend by 100. Since our VAR is expressed in terms of real money balances, we take the sum of log inverse velocity and real GDP. Finally, we restrict our quarterly observations to the period from 1965:I to 2005:I.

Additional References

- Aliprantis, Charalambos D. and Kim C. Border (2006): *Infinite-Dimensional Analysis*, 3rd Edition, Springer Verlag, New York.
- Andrews, Donald and Patrick Guggenberger (2009): “Validity of Subsampling and ‘Plug-in Asymptotics’ Inference for Parameters Defined by Moment Inequalities,” *Econometric Theory*, **25**, 669-709.
- Andrews, Donald and Gustavo Soares (2010): “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.

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

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

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