cemmap

centre for microdata methods and practice

Robust Bayesian inference for set-identified models

Raffaella Giacomini
Toru Kitagawa

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP61/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Robust Bayesian Inference for Set-Identified Models *

Raffaella Giacomini† and Toru Kitagawa‡

This draft: October 2018

Abstract

This paper reconciles the asymptotic disagreement between Bayesian and frequentist inference in set-identified models by adopting a multiple-prior (robust) Bayesian approach. We propose new tools for Bayesian inference in set-identified models. We show that these tools have a well-defined posterior interpretation in finite samples and are asymptotically valid from the frequentist perspective. The main idea is to construct a prior class that removes the source of the disagreement: the need to specify an unrevisable prior. The corresponding class of posteriors can be summarized by reporting the 'posterior lower and upper probabilities' of a given event and/or the 'set of posterior means' and the associated 'robust credible region'. We show that the set of posterior means is a consistent estimator of the true identified set and the robust credible region has the correct frequentist asymptotic coverage for the true identified set if it is convex. Otherwise, the method can be interpreted as providing posterior inference about the convex hull of the identified set. For impulse-response analysis in set-identified Structural Vector Autoregressions, the new tools can be used to overcome or quantify the sensitivity of standard Bayesian inference to the choice of an unrevisable prior.

Keywords: multiple priors, identified set, credible region, consistency, asymptotic coverage, identifying restrictions, impulse-response analysis.

*This paper merges and extends two previously circulated (and now retired) working papers: Giacomini, R. and T. Kitagawa (2015): 'Robust Inference about Partially Identified SVARs' and Kitagawa, T. (2012): 'Estimation and Inference for Set-Identified Parameters using Posterior Lower Probabilities'. We would like to thank Matthew Read for outstanding research assistance and Alessio Volpicella for providing useful computational insights. We also thank Gary Chamberlain, Jean-Pierre Florens, Eleonora Granziera, Frank Kleibergen, Sophocles Mavroeidis, Ulrich Mueller, Andriy Norets, Joris Pinkse, Frank Schorfheide, and several seminar and conference participants for their valuable comments. Both authors gratefully acknowledge financial support from ERC grants (numbers 536284 and 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001).

†University College London, Department of Economics. Email: r.giacomini@ucl.ac.uk

‡University College London, Department of Economics. Email: t.kitagawa@ucl.ac.uk

1 Introduction

It is well known that the asymptotic equivalence between Bayesian and frequentist inference breaks down in set-identified models. First, the sensitivity of Bayesian inference to the choice of the prior does not vanish asymptotically, unlike in the point identified case (Poirier (1998)). Second, any prior choice can lead to 'overly informative' inference, in the sense that Bayesian interval estimates asymptotically lie inside the true identified set (Moon and Schorfheide (2012)). This paper reconciles this disagreement between Bayesian and frequentist inference by adopting a multiple-prior robust Bayesian approach.

In a set-identified structural model the prior for the model's parameter can be decomposed into two components: the prior for the reduced-form parameter, which is revised by the data; and the prior for the structural parameter given the reduced-form parameter, which cannot be revised by data. Our robust Bayesian approach removes the need to specify the prior for the structural parameter given the reduced-form parameter, which is the component of the prior that is responsible for the asymptotic disagreement between Bayesian and frequentist inference. This is accomplished by constructing a class of priors that shares a single prior for the reduced-form parameter but allows for arbitrary conditional priors for (or ambiguous beliefs about) the structural parameter given the reduced-form parameter. By applying Bayes' rule to each prior in this class, we obtain a class of posteriors and show that it can be used to perform posterior sensitivity analysis and to conduct inference about the identified set.

We propose summarizing the information contained in the class of posteriors by considering the 'posterior lower and upper probabilities' of an event of interest and/or by reporting two sets: the 'set of posterior means (or quantiles)' in the class of posteriors and the 'robust credible region'. These can all be expressed in terms of the (single) posterior of the reduced-form parameter, so they can be obtained numerically if it is possible to draw the reduced-form parameter randomly from this posterior.

We show that, if the true identified set is convex, the set of posterior means converges asymptotically to the true identified set and the robust credible region attains the desired frequentist coverage for the true identified set asymptotically (in a pointwise sense). If the true identified set is not convex, the method provides posterior inference about the convex hull of the identified set.

The paper further proposes diagnostic tools that measure the plausibility of the identifying restrictions, the information contained in the identifying restrictions, and the information introduced by the unrevisable prior that is required by a standard Bayesian approach.

The second part of the paper presents a detailed illustration of the method in the context of impulse-response analysis in Structural Vector Autoregressions (SVARs) that are set-identified due to under-identifying zero and/or sign restrictions (Faust (1998); Canova and Nicolo (2002); Uhlig (2005); Mountford and Uhlig (2009), among others). As is typical in this literature, we focus on pointwise inference about individual impulse responses. A scalar object of interest facilitates computing the set of posterior means and the robust credible region, since the posterior of an interval can be reduced to the posterior of a two-dimensional object (the upper and lower bounds).[1]

Most empirical applications of set-identified SVARs adopt standard Bayesian inference and select a 'non-informative' – but unrevisable – prior for the 'rotation matrix' which transforms reduced-form shocks into structural shocks.[2]. Baumeister and Hamilton (2015) strongly caution against this approach and show that it may result in spuriously informative posterior inference. Our method overcomes this drawback by removing the need to specify a single prior for the rotation matrix.

We give primitive conditions that ensure frequentist validity of the robust Bayesian method in the context of SVARs. The conditions are mild or easy to verify, and cover a wide range of applications in practice. In particular, the results on the types of equality and/or sign restrictions that give rise to a convex identified set with continuous and differentiable endpoints are new to the literature and may be of separate interest regardless of whether one favours a Bayesian or a frequentist approach.

We provide an algorithm for implementing the procedure, which in practice consists of adding an optimization step to the algorithms already used in the literature, such as those of Uhlig (2005) and Arias et al. (2018) (we provide a Matlab toolbox that implements the method and automatically checks many of the conditions for its validity).

Our practical suggestion in empirical applications is to report the posterior lower (or upper) probability of an event and/or the set of posterior means and the robust credible region, as an alternative or addition to the output that is reported in a standard Bayesian setting. Reporting the outputs from both approaches, together with the diagnostic tools, can provide useful information to help empirical researchers separate the information contained in the data and in the imposed identifying restrictions from that introduced by choosing a particular unrevisable prior.

As a concrete example of how to interpret the robust Bayesian output in an SVAR application, the finding that the posterior lower probability of the event 'the impulse response is negative' equals, say, 60%, means that the posterior probability of a negative impulse response is at least 60%, regardless of the choice of unrevisable prior for the rotation matrix. The set of posterior means can be interpreted as an estimate of the impulse-response identified set. The robust credible region is an interval for the impulse-response such that the posterior probability assigned to it is greater than or equal to, say, 90%, regardless of the prior for the rotation matrix.

The empirical illustration applies the method to a standard monetary SVAR that imposes various combinations of the equality and sign restrictions that are typically used in the literature. The findings illustrate that common sign restrictions alone have little identifying power, which

[1]Extending the analysis to the vector case would in principle be possible, but would be challenging in terms of both visualization and computation. This is also true in point-identified SVARs (see the discussion in Inoue and Kilian (2013)).

[2]Gafarov et al. (2018) and Granziera et al. (2018) are notable exceptions that consider a frequentist setting.

means that standard Bayesian inference is largely driven by the choice of the unrevisable prior for the rotation matrix. The addition of even a single zero restriction tightens the estimated identified set considerably, makes standard Bayesian inference less sensitive to the choice of prior for the rotation matrix and can lead to informative inference about the sign of the output response to a monetary policy shock.

This paper is related to several literatures in econometrics and statistics.

Robust Bayesian analysis in statistics has been considered by DeRobertis and Hartigan (1981), Berger and Berliner (1986), Wasserman (1989, 1990), Wasserman and Kadane (1990) and Berger (1994). In econometrics, pioneering contributions using multiple priors are Chamberlain and Leamer (1976) and Leamer (1982), who obtain the bounds for the posterior mean of regression coefficients when a prior varies over a certain class. No previous studies explicitly consider set-identified models, but rather focus on point identified models, and view the approach as a way to measure the global sensitivity of the posterior to the choice of prior (as an alternative to a full Bayesian analysis requiring the specification of a hyperprior over the priors in the class).

In econometrics, there is a large literature on estimation and inference in set-identified models from the frequentist perspective, including Horowitz and Manski (2000), Imbens and Manski (2004), Chernozhukov et al. (2007), Stoye (2009), Romano and Shaikh (2010), to list a few. See Canay and Shaikh (2017) for a survey of the literature. There is also a growing literature on Bayesian inference for partially identified models. Some propose posterior inference based on a single prior irrespective of the posterior sensitivity introduced by the lack of identification (Epstein and Seo (2014); Baumeister and Hamilton (2015); Gustafson (2015)). Our paper does not intend to provide any normative argument as to whether one should adopt a single prior or multiple priors in set-identified models. Our main goal is to offer new tools for inference in set-identified models, and to show that these tools have a well-defined posterior interpretation in finite samples and yield asymptotically valid frequentist inference. In parallel work, Norets and Tang (2014) and Kline and Tamer (2016) consider Bayesian inference about the identified set. Norets and Tang (2014) focus on the specific setting of dynamic discrete choice models and Kline and Tamer (2016) consider a non-standard environment where a well-defined Bayesian analysis is not possible because the likelihood for the structural parameters of interest is not available (e.g., incomplete structural models or models defined by moment inequalities). In contrast, we consider a general setting where a likelihood is available and we obtain well-defined (robust) Bayesian inferential statements by introducing the notion of ambiguity over the identified set through multiple priors. Liao and Jiang (2010), Wan (2013), and Chen et al. (forthcoming) propose using Bayesian Markov Chain Monte Carlo methods to overcome some computational challenges of the frequentist approach to inference about the identified set (e.g., for the criterion-function approach considered in Chernozhukov et al. (2007)).

Some of the technical aspects of this paper relate to the literature on random sets (Beresteanu and Molinari (2008); Beresteanu et al. (2012); Galichon and Henry (2009); Molchanov and Molinari (2018)), since the set of posterior means can be viewed as the Aumann expectation of the random identified set. The main difference is that in our case the source of randomness for the identified set is the posterior uncertainty about the reduced-form parameter, not the sampling distribution of the observations.

The remainder of the paper is organized as follows. Section 2 considers the general setting of set identification and introduces the multiple-prior robust Bayesian approach. Section 3 analyzes the asymptotic properties of the method. Section 4 illustrates the application to SVARs. Section 5 discusses the numerical implementation of the method. Sections 4 and 5 are self-contained, so a reader who is interested in SVARs can focus on these two sections. Section 6 contains the empirical application and Section 7 concludes. The proofs are in Appendix A and Appendix B contains additional results on convexity of the impulse-response identified set.

2 Set Identification and Robust Bayesian Inference

2.1 Notation and Definitions

This section describes the general framework of set-identified structural models. In particular, it introduces the definitions of structural parameter θ, reduced-form parameter ϕ and parameter of interest η that are used throughout the paper.

Let $(\mathbf{Y}, \mathcal{Y})$ and $(\Theta, \mathcal{A})$ be measurable spaces of a sample $Y \in \mathbf{Y}$ and a parameter vector $\theta \in \Theta$, respectively. The general framework in the paper allows for both a parametric model with $\Theta = \mathcal{R}^d$, $d < \infty$, and a non-parametric model with Θ a separable Banach space. Assume that the conditional distribution of Y given θ exists and has a probability density $p(y|\theta)$ at every $\theta \in \Theta$ with respect to a σ-finite measure on $(\mathbf{Y}, \mathcal{Y})$, where $y \in \mathbf{Y}$ indicates sampled data.

Set identification of θ arises when multiple values of θ are observationally equivalent, so that for θ and $\theta' \neq \theta$, $p(y|\theta) = p(y|\theta')$ for every $y \in \mathbf{Y}$ (Rothenberg (1971), Drèze (1974), and Kadane (1974)). Observational equivalence can be represented by a many-to-one function $g : (\Theta, \mathcal{A}) \to (\Phi, \mathcal{B})$, such that $g(\theta) = g(\theta')$ if and only if $p(y|\theta) = p(y|\theta')$ for all $y \in \mathbf{Y}$. This relationship partitions the parameter space Θ into equivalent classes, in each of which the likelihood of θ is "flat" irrespective of observations, and $\phi = g(\theta)$ maps each of the equivalent classes to a point in a parameter space Φ. In the language of structural models in econometrics (Hurwicz (1950), and Koopmans and Reiersol (1950)), $\phi = g(\theta)$ is reduced-form parameter that indexes the distribution of data. The reduced-form parameter carries all the information for the structural parameter θ through the value of the likelihood function, in the sense that there exists a $\mathcal{B}$-measurable function $\hat{p}(y|\cdot)$ such that $p(y|\theta) = \hat{p}(y|g(\theta))$ holds $\forall y \in \mathbf{Y}$ and $\theta \in \Theta$.[3]

[3]In Bayesian statistics, $\phi = g(\theta)$ is referred to as the (minimal) sufficient parameters that satisfy conditional independence $Y \perp \theta|\phi$ (Barankin (1960); Dawid (1979); Florens and Mouchart (1977); Picci (1977); Florens et al.

Let the parameter of interest $\eta \in \mathcal{H}$ be a finite-dimensional subvector or a transformation of θ, $\eta = h(\theta)$ with $h : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$, $\mathcal{H} \subset \mathcal{R}^k$, $k < \infty$. The identified sets of θ and η are defined as follows.

Definition 1 (Identified Sets of θ and η). *(i) The identified set of θ is the inverse image of $g(\cdot)$: $IS_\theta(\phi) = \{\theta \in \Theta : g(\theta) = \phi\}$, where $IS_\theta(\phi)$ and $IS_\theta(\phi')$ for $\phi \neq \phi'$ are disjoint and $\{IS_\theta(\phi) : \phi \in \Phi\}$ constitutes a partition of Θ.*

(ii) The identified set of $\eta = h(\theta)$ is a set-valued map $IS_\eta : \Phi \rightrightarrows \mathcal{H}$ defined by the projection of $IS_\theta(\phi)$ onto $\mathcal{H}$ through $h(\cdot)$, $IS_\eta(\phi) \equiv \{h(\theta) : \theta \in IS_\theta(\phi)\}$.

(iii) The parameter $\eta = h(\theta)$ is point-identified at ϕ if $IS_\eta(\phi)$ is a singleton, and η is set-identified at ϕ if $IS_\eta(\phi)$ is not a singleton.

We define the identified set for θ in terms of the likelihood-based definition of observational equivalence of θ. As a result, $IS_\theta(\phi)$ and $IS_\eta(\phi)$ are ensured to give their sharp identification regions at every distribution of data indexed by ϕ. In some structural models, including SVARs, the space of the reduced-form parameter Φ on which the reduced-form likelihood is well-defined can be larger than the space of the reduced-form parameter generated from the structure ($g(\Theta)$); that is, the model is observationally restrictive in the sense of Koopmans and Reiersol (1950). In this case, the model is falsifiable, and $IS_\theta(\phi)$ can be empty for some $\phi \in \Phi$.

2.2 Multiple Priors

In this section we discuss how set identification induces unrevisable prior knowledge and we introduce the use of multiple priors.

Let π_θ be a prior (distribution) for θ and π_ϕ be the corresponding prior for ϕ, obtained as the marginal probability measure on $(\Phi, \mathcal{B})$ induced by π_θ and $g(\cdot)$:

$$\pi_\phi(B) = \pi_\theta(IS_\theta(B)) \quad \text{for all } B \in \mathcal{B}. \tag{2.1}$$

Since the likelihood for θ is flat on $IS_\theta(\phi)$ for any Y, conditional independence $\theta \perp Y|\phi$ holds. The posterior of θ, $\pi_{\theta|Y}$, is accordingly obtained as

$$\pi_{\theta|Y}(A) = \int_\Phi \pi_{\theta|\phi}(A|\phi) d\pi_{\phi|Y}(\phi), \quad A \in \mathcal{A}, \tag{2.2}$$

where $\pi_{\theta|\phi}$ is the conditional distribution of θ given ϕ, and $\pi_{\phi|Y}$ is the posterior of ϕ.

Expression (2.2) shows that the prior of the reduced-form parameter, π_ϕ, can be updated by the data, whereas the conditional prior of θ given ϕ is never updated because the likelihood is flat on $IS_\theta(\phi) \subset \Theta$ for any realization of the sample. In this sense, one can interpret π_ϕ as the *revisable prior knowledge* and the conditional priors, $\left\{\pi_{\theta|\phi}(\cdot|\phi) : \phi \in \Phi\right\}$, as the *unrevisable prior knowledge*.

(1990)).

In a standard Bayesian setting the posterior uncertainty about θ is summarized by a single probability distribution. This requires specifying a single prior for θ, which necessarily induces a single conditional prior $\pi_{\theta|\phi}$. If one could justify this choice of conditional prior, the standard Bayesian updating formula (2.2) would yield a valid posterior for θ. A challenging situation arises if a credible conditional prior is not readily available. In this case, a researcher who is aware that $\pi_{\theta|\phi}$ is never updated by the data might worry about the influence that a potentially arbitrary choice can have on posterior inference.

The robust Bayesian analysis in this paper focuses on this situation, and removes the need to specify a single conditional prior by introducing ambiguity for $\pi_{\theta|\phi}$ in the form of multiple priors.

Definition 2 (Multiple-Prior Class). *Given a unique π_ϕ supported only on $g(\Theta)$, the class of conditional priors for θ given ϕ is:*

$$\Pi_{\theta|\phi} = \left\{\pi_{\theta|\phi} : \pi_{\theta|\phi}(IS_\theta(\phi)) = 1, \pi_\phi - \textit{almost surely}\right\}. \tag{2.3}$$

$\Pi_{\theta|\phi}$ consists of arbitrary conditional priors as long as they assign probability one to the identified set of θ. $\Pi_{\theta|\phi}$ induces a class of proper priors for θ, $\Pi_\theta \equiv \{\pi_\theta = \int \pi_{\theta|\phi} d\pi_\phi : \pi_{\theta|\phi} \in \Pi_{\theta|\phi}\}$, which consists of all priors for θ whose marginal distribution for ϕ coincides with the prespecified π_ϕ. Our proposal requires a researcher to specify a single prior only for the reduced-form parameter ϕ, but it otherwise leaves the conditional prior $\pi_{\theta|\phi}$ unspecified.[4]

In this paper we shall not discuss how to select the prior π_ϕ for the reduced-form parameter, and treat it as given. As the influence of this prior choice on posterior inference disappears asymptotically, any sensitivity issues in this respect would potentially only concern small samples.

2.3 Posterior Lower and Upper Probabilities

This section discusses how to summarize the posterior information when the robust Bayesian prior input is given by $(\Pi_{\theta|\phi}, \pi_\phi)$.

Applying Bayes' rule to each prior in the class Π_θ generates the class of posteriors for θ. Transforming each member of the class gives the class of posteriors for the parameter of interest η:

$$\Pi_{\eta|Y} \equiv \left\{\pi_{\eta|Y}(\cdot) = \int_\Phi \pi_{\theta|\phi}(h(\theta) \in \cdot) d\pi_{\phi|Y} : \pi_{\theta|\phi} \in \Pi_{\theta|\phi}\right\}. \tag{2.4}$$

We propose to summarize this posterior class by the *posterior lower probability* $\pi_{\eta|Y*}(\cdot) : \mathcal{D} \rightarrow$

[4] The reduced-form parameter ϕ is defined by examining the entire model $\{p(y|\theta) : y \in \mathcal{Y}, \theta \in \Theta\}$, so the prior class is, by construction, model dependent. This distinguishes the approach here from the robust Bayesian analysis of, e.g., Berger (1985), where a prior class represents the researcher's subjective assessment of her imprecise prior knowledge.

$[0,1]$ and the *posterior upper probability* $\pi^*_{\eta|Y}(\cdot):\mathcal{D}\to[0,1]$, defined as

$$\pi_{\eta|Y*}(D) \equiv \inf_{\pi_{\eta|Y}\in\Pi_{\eta|Y}} \pi_{\eta|Y}(D),$$
$$\pi^*_{\eta|Y}(D) \equiv \sup_{\pi_{\eta|Y}\in\Pi_{\eta|Y}} \pi_{\eta|Y}(D).$$

Note the conjugate property, $\pi_{\eta|Y*}(D) = 1-\pi^*_{\eta|Y}(D^c)$, so it suffices to focus on one of them. The lower and upper probabilities provide the set of posterior beliefs that are valid irrespective of the choice of unrevisable prior. When $\{\eta\in D\}$ specifies a hypothesis of interest, $\pi_{\eta|Y*}(D)$ can be interpreted as saying that 'the posterior credibility for $\{\eta\in D\}$ is at least equal to $\pi_{\eta|Y*}(D)$, no matter which unrevisable prior one assumes'. These quantities are useful for conducting global sensitivity analysis with respect to a prior that cannot be revised by the data. Furthermore, if one agrees that the ultimate goal of partial identification analysis is to establish a 'domain of consensus' (Manski (2007)) among assumptions that the data are silent about, the posterior lower and upper probabilities constructed upon arbitrary unrevisable prior knowledge are natural quantities to focus on when considering partial identification from the Bayesian perspective.

In order to derive an analytical expression for $\pi_{\eta|Y*}(\cdot)$, we assume the following regularity conditions.

Assumption 1. *(i) The prior of* ϕ, π_ϕ, *is proper, absolutely continuous with respect to a* σ*-finite measure on* $(\Phi,\mathcal{B})$, *and* $\pi_\phi(g(\Theta))=1$, *i.e.,* $IS_\theta(\phi)$ *and* $IS_\eta(\phi)$ *are nonempty,* π_ϕ*-a.s.*

(ii) The mapping between θ *and* ϕ, $g:(\Theta,\mathcal{A})\to(\Phi,\mathcal{B})$, *is measurable and its inverse image* $IS_\theta(\phi)$ *is a closed set in* Θ, π_ϕ*-almost every* ϕ.

(iii) The mapping between θ *and* η, $h:(\Theta,\mathcal{A})\to(\mathcal{H},\mathcal{D})$, *is measurable and* $IS_\eta(\phi)=h(IS_\theta(\phi))$ *is a closed set in* $\mathcal{H}$, π_ϕ*-almost every* ϕ.

Assumption 1 (i) guarantees that the identified set $IS_\eta(\phi)$ can be viewed as a random set defined on the probability space both a priori $(\Phi,\mathcal{B},\pi_\phi)$ and a posteriori $\left(\Phi,\mathcal{B},\pi_{\phi|Y}\right)$, which we exploit in the proof of Theorem 1 below. As we discuss in Section 5, the numerical implementation of our method only requires the posterior to be proper, so in practical terms it allows an improper prior with support larger than $g(\Theta)$. The assumption is then imposed in the numerical procedure by only retaining draws that give a non-empty identified set. Assumptions 1 (ii) and 1 (iii) are mild conditions ensuring that $IS_\theta(\phi)$ and $IS_\eta(\phi)$ are random closed sets satisfying a certain measurability requirement. The closedness of $IS_\theta(\phi)$ and $IS_\eta(\phi)$ is implied, for instance, by continuity of $g(\cdot)$ and $h(\cdot)$.

The next proposition expresses the posterior lower and upper probabilities for the parameter of interest in terms of the posterior of ϕ. This proposition provides the basis for the numerical approximation of lower and upper probabilities, which only requires the ability to compute the identified set at values of ϕ randomly drawn from its posterior.

Theorem 1 *Under Assumption 1, for $D \in \mathcal{D}$,*

$$\begin{aligned} \pi_{\eta|Y*}(D) &= \pi_{\phi|Y}(\{\phi : IS_\eta(\phi) \subset D\}), \\ \pi^*_{\eta|Y}(D) &= \pi_{\phi|Y}(\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\}). \end{aligned}$$

The expression for $\pi_{\eta|Y*}(D)$ shows that the lower probability on D is the probability that the (random) identified set $IS_\eta(\phi)$ is contained in subset D in terms of the posterior probability of ϕ. The upper probability is the posterior probability that the set $IS_\eta(\phi)$ hits subset D. Setting $\eta = \theta$ gives the posterior lower and upper probabilities for θ in terms of the containment and hitting probabilities of $IS_\theta(\phi)$. In standard Bayesian inference, the posterior of θ is transformed into a posterior for $\eta = h(\theta)$ by integrating the posterior probability measure of θ for η, while here it corresponds to projecting random sets $IS_\theta(\phi)$ onto $\mathcal{H}$ via $\eta = h(\cdot)$. This highlights the difference between standard Bayesian analysis and robust Bayesian analysis based on the lower probability. As remarked in the proof of Theorem 1, for each $D \in \mathcal{D}$, the set of posterior probabilities $\{\pi_{\eta|Y}(D) : \pi_{\eta|Y} \in \Pi_{\eta|Y}\}$ coincides with the connected intervals $[\pi_{\eta|Y*}(D), \pi^*_{\eta|Y}(D)]$, implying that any posterior probability in this set can be attained by some posterior in $\Pi_{\eta|Y}$.

It is well known in the robust statistics literature (e.g., Huber (1973)) that the lower probability of a set of probability measures is in general a monotone nonadditive measure (capacity). The posterior lower and upper probabilities in this paper coincide with the construction of the posterior lower and upper probabilities of Wasserman (1990) when it is applied to our prior class. An important distinction from Wasserman's analysis is that our posterior lower probability is guaranteed to be an ∞-order monotone capacity (a containment functional of random sets), which simplifies investigating its analytical properties and implementing the method in practice.[5]

2.4 Set of Posterior Means and Quantiles.

The posterior lower and upper probabilities shown in Theorem 1 summarize the set of posterior probabilities for an arbitrary event of interest D. To summarize the information in the posterior class without specifying D, we propose to report the set of posterior means of η.

The next proposition shows that the set of posterior means of η is equivalent to the Aumann expectation of the convex hull of the identified set.

Theorem 2 *Suppose Assumption 1 holds and the random set $IS_\eta(\phi) \subset \mathcal{H}$, $\phi \sim \pi_{\phi|Y}$, is L_1-integrable with respect to $\pi_{\phi|Y}$ in the sense that $E_{\phi|Y}\left(\sup_{\eta \in IS_\eta(\phi)} \|\eta\|\right) < \infty$. Let $co(IS_\eta(\phi))$ be*

[5] Wasserman (1990, p.463) posed an open question asking which class of priors can ensure that the posterior lower probability is a containment functional of random sets. Theorem 1 provides an answer to this open question in the case that the model is set identified.

the convex hull of $IS_\eta(\phi)$[6] *and let* $E^A_{\phi|Y}(\cdot)$ *denote the Aumann expectation of a random set with underlying probability measure* $\pi_{\phi|Y}$.[7] *Then, the set of posterior means is convex and equals the Aumann expectation of the convex hull of the identified set:*

$$\left\{E_{\eta|Y}(\eta) : \pi_{\eta|Y} \in \Pi_{\eta|Y}\right\} = E^A_{\phi|Y}[co(IS_\eta(\phi))]. \tag{2.5}$$

Let $s(IS_\eta(\phi), q) \equiv \sup_{\eta \in IS_\eta(\phi)} \eta' q$, $q \in \mathcal{S}^{k-1}$, be the support function of identified set $IS_\eta(\phi) \subset \mathcal{R}^k$, where $\mathcal{S}^{k-1}$ is the unit sphere in $\mathcal{R}^k$. It is known that the Aumann expectation of $co(IS_\eta(\phi))$ satisfies $s(E^A_{\phi|Y}[co(IS_\eta(\phi))], \cdot) = E_{\phi|Y}[s(IS_\eta(\phi), \cdot)]$ (see, e.g., Theorem 1.26 in Chap. 2 of Molchanov (2005)) and a support function one-to-one corresponds to the closed convex set. Hence, the analytical characterization shown in Theorem 2 suggests that the set of posterior means can be computed by approximating $E_{\phi|Y}[s(IS_\eta(\phi), \cdot)]$ using the draws of $IS_\eta(\phi), \phi \sim \pi_{\phi|Y}$ and mapping back the approximated average support function to obtain the set of posterior means $E^A_{\phi|Y}[co(IS_\eta(\phi))]$.

In case of scalar η, the set of posterior means has the particularly simple form $E^A_{\phi|Y}[co(IS_\eta(\phi))] = \left[E_{\phi|Y}(\ell(\phi)), E_{\phi|Y}(u(\phi))\right]$, where $\ell(\phi) = \inf\{\eta : \eta \in IS_\eta(\phi)\}$ and $u(\phi) = \sup\{\eta : \eta \in IS_\eta(\phi)\}$ are the lower and upper bounds of $IS_\eta(\phi)$. Thus, in applications where it is feasible to compute $\ell(\phi)$ and $u(\phi)$, we can approximate $E_{\phi|Y}(\ell(\phi))$ and $E_{\phi|Y}(u(\phi))$ by using a random sample of ϕ drawn from $\pi_{\phi|Y}$.

In case of scalar η, the set of posterior τ-th quantiles of η is also simple to compute. We apply Theorem 1 with $D = (-\infty, t]$, $-\infty < t < \infty$, to obtain the set of the posterior cumulative distribution functions (CDF) of η for each t. Inverting the upper and lower bounds of this set at $\tau \in (0, 1)$ gives the set of posterior τ-th quantiles of η.

2.5 Robust Credible Region

This section introduces the robust Bayesian counterpart of the highest posterior density region that is typically reported in standard Bayesian inference.

For $\alpha \in (0, 1)$, consider a subset $C_\alpha \subset \mathcal{H}$ such that the posterior lower probability $\pi_{\eta|Y*}(C_\alpha)$ is greater than or equal to α:

$$\pi_{\eta|Y*}(C_\alpha) = \pi_{\phi|Y}(IS_\eta(\phi) \subset C_\alpha)) \geq \alpha. \tag{2.6}$$

C_α is interpreted as "a set on which the posterior credibility of η is at least α, *no matter which posterior is chosen within the class*". Dropping the italicized part from this statement yields the

[6] $co(IS_\eta) : \Phi \rightrightarrows \mathcal{H}$ is viewed as a closed random set defined on the probability space $\left(\Phi, \mathcal{B}, \pi_{\phi|Y}\right)$

[7] Let $X : \Phi \rightrightarrows \mathcal{H}$ be a closed random set defined on the probability space $\left(\Phi, \mathcal{B}, \pi_{\phi|Y}\right)$, and $\xi(\phi) : \Phi \to \mathcal{H}$ be its measurable selection, i.e., $\xi(\phi) \in X(\phi)$, $\pi_{\phi|Y}$-a.s. Let $S^1(X)$ be the class of integrable measurable selections, $S^1(X) = \left\{\xi : \xi(\phi) \in X(\phi),\ \pi_{\phi|Y}\text{-a.s., } E_{\phi|Y}(\|\xi\|) < \infty\right\}$. The Aumann expectation of X is defined as (Aumann (1965)) $E^A_{\phi|Y}(X) \equiv \left\{E_{\phi|Y}(\xi) : \xi \in S^1(X)\right\}$.

usual interpretation of the posterior credible region, so this definition seems like a natural extension to our robust Bayesian setting. We refer to C_α satisfying (2.6) as a *robust credible region with credibility* α.

As in the standard Bayesian case, there are multiple ways to construct C_α satisfying (2.6). We propose resolving this multiplicity by choosing C_α such that it has the smallest volume in terms of the Lebesgue measure:

$$\begin{aligned} C_\alpha^* &\in \arg\min_{C \in \mathcal{C}} Leb(C) \\ \text{s.t.} \quad & \pi_{\phi|Y}(IS_\eta(\phi) \subset C)) \geq \alpha, \end{aligned} \tag{2.7}$$

where $Leb(C)$ is the volume of C in terms of the Lebesgue measure and $\mathcal{C}$ is a family of subsets in $\mathcal{H}$.[8] We refer to C_α^* defined in this way as a *smallest robust credible region* with credibility α.[9] The credible regions for the identified set proposed in Moon and Schorfheide (2011), Norets and Tang (2014), and Kline and Tamer (2016) satisfy (2.6), so they are robust credible regions in our definition. However, these works do not consider the volume-optimized credible region (2.7).[10]

Obtaining C_α^* is challenging if η is a vector and no restriction is placed on the class $\mathcal{C}$ in (2.7). Proposition 1 below shows that for scalar η this can be overcome by constraining $\mathcal{C}$ to be the class of closed connected intervals. C_α^* can then be computed by solving a simple optimization problem.

Proposition 1 (Smallest Robust Credible Region for Scalar η). *Let η be scalar and let $d : \mathcal{H} \times \mathcal{D} \to \mathcal{R}_+$ measure the distance from $\eta_c \in \mathcal{H}$ to the set $IS_\eta(\phi)$ by*

$$d(\eta_c, IS_\eta(\phi)) \equiv \sup_{\eta \in IS_\eta(\phi)} \{\|\eta_c - \eta\|\}.$$

For each $\eta_c \in \mathcal{H}$, let $r_\alpha(\eta_c)$ be the α-th quantile of the distribution of $d(\eta_c, IS_\eta(\phi))$ induced by the posterior distribution of ϕ, i.e.,

$$r_\alpha(\eta_c) \equiv \inf\left\{r : \pi_{\phi|Y}\left(\{\phi : d(\eta_c, IS_\eta(\phi)) \leq r\}\right) \geq \alpha\right\}.$$

Then, C_α^* is a closed interval centered at $\eta_c^* = \arg\min_{\eta_c \in \mathcal{H}} r_\alpha(\eta_c)$ with radius $r_\alpha^* = r_\alpha(\eta_c^*)$.

---

[8] In case that $IS_\eta(\phi)$ lies in a k'-dimensional manifold of $\mathcal{R}^k$, $k' < k$, $\pi_{\phi|Y}$-a.s., we modify the Lebesgue measure on $\mathcal{R}^k$ in this optimization to that of $\mathcal{R}^{k'}$ so that this "volume' minimization problem can have a well-defined solution.

[9] Focusing on the smallest set estimate has a decision-theoretic justification; C_α^* can be supported as a solution to the following posterior minimax problem:

$$C_\alpha^* \in \arg\min_{C \in \mathcal{C}} \left[\sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \int L(\eta, C)\, d\pi_{\eta|Y}\right]$$

with a loss function that penalizes volume and non-coverage, $L(\eta, C) = Leb(C) + b(\alpha)[1 - 1_C(\eta)]$, where $b(\alpha)$ is a positive constant that depends on the credibility level α, and $1_C(\cdot)$ is the indicator function for $\{\eta \in C\}$.

[10] Moon and Schorfheide (2011) and Norets and Tang (2014) propose credible regions for the identified set by taking the union of $IS_\eta(\phi)$ over ϕ in its Bayesian credible region.

2.6 Diagnostic Tools

2.6.1 Plausibility of Identifying Restrictions

For observationally restrictive models (i.e., $g(\Theta)$ is a proper subset of Φ), quantifying posterior information for assessing the set-identifying restrictions can be of interest. To do so, we start with a prior of ϕ that supports the entire Φ, which we denote by $\tilde{\pi}_\phi$. Trimming the support of $\tilde{\pi}_\phi$ on $g(\Theta) = \{\phi : IS_\theta(\phi) \neq \emptyset\}$ gives π_ϕ satisfying Assumption 1 (i). We update $\tilde{\pi}_\phi$ to obtain the posterior of ϕ with extended domain $\tilde{\pi}_{\phi|Y}$.

Since emptiness of the identified set can refute the imposed identifying restrictions, their plausibility can be measured by the posterior probability that the identified set is non-empty, $\tilde{\pi}_{\phi|Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$.[11] Note that this measure depends only on the posterior of the reduced-form parameter, so it is free from the issue of posterior sensitivity due to set identification. By reporting the posterior plausibility of the identifying restrictions and the set of posterior means conditional on $\{IS_\eta(\phi) \neq \emptyset\}$, we can separate inferential statements about the validity of the identifying restrictions from inferential statements about the parameter of interest, which is difficult to do from a frequentist perspective (see the discussion in Sims and Zha (1999)).

2.6.2 Informativeness of Identifying Restrictions and of Priors

The strength of identifying restrictions can be measured by comparing the set of posterior means relative to that of a model that does not impose these restrictions but is otherwise identical. For instance, suppose the object of interest η is a scalar. Let M_s be the set-identified model imposing the identifying restrictions whose strength is to be measured and M_l be the model that relaxes the restrictions. For identification of η, the identifying power of the restrictions imposed in M_s but not in M_l can be measured by:

$$\begin{aligned}&\text{Informativeness of restrictions imposed in model } M_s \text{ but not in } M_l\\ =&1 - \frac{\text{width of set of posterior means of } \eta \text{ in model } M_s}{\text{width of set of posterior means of } \eta \text{ in model } M_l}.\end{aligned} \tag{2.8}$$

This measure captures by how much (in terms of the fraction) the restrictions in model M_s reduce the width of the set of posterior means of η compared to the model M_l.

The amount of information in the posterior provided by the choice of a single unrevisable prior (i.e., the choice of a conditional prior $\pi_{\theta|\phi}$, as used in standard Bayesian inference) can be measured in a similar way. In this case, we compare the width of a robust credible region C_α satisfying (2.6)

[11] An alternative measure is the prior-posterior odds of the nonemptiness of the identified set, $\frac{\tilde{\pi}_{\phi|Y}\left(\{\phi : IS_\eta(\phi)\neq\emptyset\}\right)}{\tilde{\pi}_{\phi}\left(\{\phi : IS_\eta(\phi)\neq\emptyset\}\right)}$. A value greater than one indicates that the data support the plausibility of the imposed restrictions.

relative to the width of the standard Bayesian credible region obtained from the single prior:

$$\begin{aligned} & \text{Informativeness of the choice of prior} \\ = \quad & 1 - \frac{\text{width of a Bayesian credible region of } \eta \text{ with credibility } \alpha}{\text{width of a robust credible region of } \eta \text{ with credibility } \alpha}. \end{aligned} \tag{2.9}$$

This measure captures by what fraction the credible region of η is tightened by choosing a particular unrevisable prior $\pi_{\theta|\phi}$.

3 Asymptotic Properties

The set of posterior means or quantiles and the robust credible region introduced in Section 2 have well-defined (robust) Bayesian interpretations in finite samples and they are useful for conducting Bayesian sensitivity analysis to the choice of an unrevisable prior.

To examine whether these robust Bayesian quantities are useful from the frequentist perspective, this section analyzes their asymptotic frequentist properties. We show two main results. First, the set of posterior means can be viewed as an estimator of the identified set that converges to the true identified set asymptotically when the true identified set is convex. Otherwise, the set of posterior means converges to the convex hull of the true identified set. Second, the robust credible region has the correct asymptotic coverage for the true identified set. These results show that introducing ambiguity for nonidentified parameters induces asymptotic equivalence between (robust) Bayesian and frequentist inference in set-identified models. An implication of this finding is that the proposed robust Bayesian analysis can also appeal to frequentists.

In this section we let $\phi_0 \in \Phi$ denote the true value of the reduced-form parameter and $Y^T = (y_1, \ldots, y_T)$ denote a sample of size T generated from $p(Y^T|\phi_0)$.

3.1 Consistency of the Set of Posterior Means

Assume the following conditions:

Assumption 2. *(i) $IS_\eta(\phi_0)$ is bounded, and the identified set map $IS_\eta : \Phi \rightrightarrows H$ is continuous at $\phi = \phi_0$.*

(ii) The posterior of ϕ is consistent for ϕ_0, $p(Y^\infty|\phi_0)$-a.s.[12]

(iii) $IS_\eta(\phi)$ is L_2-integrable with respect to $\pi_{\phi|Y^T}$, $E_{\phi|Y^T}\left(\sup_{\eta \in IS_\eta(\phi)} \|\eta\|^2\right) < \infty$, $p\left(Y^T|\phi_0\right)$-a.s., for all $T = 1, 2, 3, \ldots$.

[12] Posterior consistency of ϕ means that $\lim_{T\to\infty} \pi_{\phi|Y^T}(G) = 1$ for every G open neighborhood of ϕ_0 and for almost every sampling sequence following $p(Y^\infty|\phi_0)$. For a finite-dimensional ϕ, posterior consistency is implied by higher-level conditions for the likelihood of ϕ. We do not list these here for the sake of brevity, and refer to Section 7.4 of Schervish (1995) for details.

Assumption 2 imposes mild conditions. Assumption 2 (i) requires that the identified set of η is a continuous correspondence at the true value ϕ_0. In the case of scalar η with convex identified set $IS_\eta(\phi) = [\ell(\phi), u(\phi)]$, this means that $\ell(\phi)$ and $u(\phi)$ are continuous at ϕ_0. Assumption 2 (ii) requires that Bayesian estimation of the reduced-form parameter is a standard estimation problem in the sense that almost-sure posterior consistency holds. Assumption 2 (iii) strengthens 2 (i) by assuming that $IS_\eta(\phi)$ is $\pi_{\phi|Y^T}$-almost surely compact-valued and its radius has finite posterior variance. In the scalar case, Assumption 2 (iii) requires that $\ell(\phi)$ and $u(\phi)$ have finite posterior variances.

Theorem 3 (Consistency). *Suppose Assumption 1 holds.*

(i) Under Assumption 2 (i) and (ii), $\lim_{T\to\infty} \pi_{\phi|Y^T}\left(\{\phi : d_H\left(IS_\eta(\phi), IS_\eta(\phi_0)\right) > \epsilon\}\right) = 0$, $p(Y^\infty|\phi_0)$*-a.s., where* $d_H(\cdot,\cdot)$ *is the Hausdorff distance.*

(ii) Under Assumption 2, the set of posterior means almost surely converges to the convex hull of the true identified set, i.e.,

$$\lim_{T\to\infty} d_H\left(E^A_{\phi|Y^T}\left[co\left(IS_\eta(\phi)\right)\right], co\left(IS_\eta(\phi_0)\right)\right) \to 0,\ \ p\left(Y^\infty|\phi_0\right)\text{-a.s.}$$

The first claim of Theorem 3 states that the identified set $IS_\eta(\phi)$, viewed as a random set induced by the posterior of ϕ, converges in posterior probability to the true identified set $IS_\eta(\phi_0)$ in the Hausdorff metric. This claim only relies on continuity of the identified set correspondence and does not rely on Assumption 2 (iii) or on convexity of the identified set. The second claim of the theorem provides a justification for using (a numerical approximation of) the set of posterior means as a consistent estimator of the convex hull of the identified set. The theorem implies that the set of posterior means converges to the true identified set if this set is convex.

3.2 Asymptotic Coverage Properties of the Robust Credible Region

We first state a set of conditions under which the robust credible region asymptotically attains correct frequentist coverage for the true identified set $IS_\eta(\phi_0)$.

Assumption 3. *(i) The identified set* $IS_\eta(\phi)$ *is* π_ϕ*-almost surely closed and bounded, and* $IS_\eta(\phi_0)$ *is closed and bounded.*

(ii) The robust credible region C_α *belongs to the class of closed and convex sets* $\mathcal{C}$ *in* $\mathcal{R}^k$*.*

Assumption 3 (i) is a weak requirement in practical applications. We allow the identified set $IS_\eta(\phi)$ to be nonconvex, while Assumption 3 (ii) constrains the robust credible region to be closed and convex. Under convexity of C_α, $IS_\eta(\phi) \subset C_\alpha$ holds if and only if $co\left(IS_\eta(\phi)\right) \subset C_\alpha$ holds, so that the inclusion of the identified set by C_α is equivalent to the dominance of their support functions, $s(IS_\eta(\phi), q) = s(co(IS_\eta(\phi)), q) \leq s(C_\alpha, q)$ for all $q \in \mathcal{S}^{k-1}$ (see, e.g., Corollary 13.1.1 in Rockafellar (1970)). This fact enables us to characterize a set of conditions for correct asymptotic coverage of C_α in terms of the limiting probability law of the support functions, which has been studied in the literature on frequentist inference for the identified set (e.g., Beresteanu and Molinari (2008); Bontemps et al. (2012); Kaido (2016)).

Assumption 4. *Let $C\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ be the set of continuous functions from the k-dimensional unit sphere $\mathcal{S}^{k-1}$ to $\mathcal{R}$. For a sequence $a_T \to \infty$ as $T \to \infty$, define stochastic processes in $C\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ indexed by $q \in \mathcal{S}^{k-1}$,*

$$\begin{aligned} X_{\phi|Y^T}(q) &\equiv a_T\left[s(IS_\eta(\phi), q) - s(IS_\eta(\hat{\phi}), q)\right], \\ X_{Y^T|\phi_0}(q) &\equiv a_T\left[s(IS_\eta(\phi_0), q) - s(IS_\eta(\hat{\phi}), q)\right], \end{aligned}$$

where the probability law of $X_{\phi|Y^T}$ is induced by $\pi_{\phi|Y^T}$, $T = 1, 2, \ldots$, and the probability law of $X_{Y^T|\phi_0}$ is induced by the sampling process $p_{Y^T|\phi_0}$, $T = 1, 2, \ldots$. The following conditions hold:

(i) $X_{\phi|Y^T} \rightsquigarrow X$ as $T \to \infty$ for $p_{Y^\infty|\phi_0}$-almost every sampling sequence, where $\rightsquigarrow$ denotes weak convergence.

(ii) $X_{Y^T|\phi_0} \rightsquigarrow Z$ as $T \to \infty$, and $Z \sim X$.

(iii) $\Pr\left(X(\cdot) \leq c(\cdot)\right)$ is continuous in $c \in C\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ with respect to the supremum metric, and $\Pr\left(X = c\right) = 0$ for any nonrandom function $c \in C\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$.

(iv) Let C_α be a robust credible region satisfying $\alpha \leq \pi_{\phi|Y^T}\left(IS_\eta(\phi) \subset C_\alpha\right) \leq 1-\epsilon$ for some $\epsilon > 0$ for all $T = 1, 2, \ldots$. The stochastic process in $C\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$, $\hat{c}_T(\cdot) \equiv a_T\left[s(C_\alpha, \cdot) - s\left(IS_\eta(\hat{\phi}), \cdot\right)\right]$, converges in $p_{Y^T|\phi_0}$-probability to $c \in \mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ as $T \to \infty$.

Assumption 4 (i) states that the posterior distribution of the support function of the identified set $IS_\eta(\phi)$, centered at the support function of $IS_\eta(\hat{\phi})$ and scaled by a_T, converges weakly to the stochastic process X. The weak convergence of the scaled support function to the tight Gaussian process on $\mathcal{S}^{k-1}$ holds with $a_T = \sqrt{T}$, for instance, if the central limit theorem for random sets applies; see, e.g, Molchanov (2005) and Beresteanu and Molinari (2008). Our Assumption 4 (i) is a Bayesian analogue to the frequentist central limit theorem for the support functions. Assumption 4 (ii) states that from the viewpoint of the support function, the difference between $IS_\eta(\hat{\phi})$ and the true identified set scaled by the same factor a_T converges in distribution to the stochastic process Z, and the probability law of Z coincides with the probability law of X.[13] Since the distribution of X is defined conditional on a sampling sequence while Z is unconditional, the agreement of the distributions of X and Z implies that the dependence of the posterior distribution of $X_{\phi|Y^T}$ on the

[13]The stochastic process X is induced by the large sample posterior distribution, while Z is induced by the large sample sampling distribution. We therefore use different notations for them.

sample Y^T vanishes as $T \to \infty$.[14] As shown in Beresteanu and Molinari (2008) and Kaido and Santos (2014), practical examples have the limiting process Z as a mean zero tight Gaussian process in $\mathcal{C}(\mathcal{S}^{k-1}, \mathcal{R})$. Assumption 4 (iii) means that the limiting process X is continuously distributed and non-degenerate in the stated sense, which holds true if X follows a nondegenerate Gaussian process. In addition to the convexity requirement of Assumption 3 (ii), Assumption 4 (iv) requires C_α to be bounded and to lie in a neighborhood of $IS_\eta(\hat{\phi})$ shrinking at rate $1/a_T$.

Theorem 4 (Asymptotic Coverage). *Under Assumptions 3 and 4, C_α, $\alpha \in (0,1)$, is an asymptotically valid frequentist confidence set for the true identified set $IS_\eta(\phi_0)$ with asymptotic coverage probability at least α.*

$$\liminf_{T\to\infty} P_{Y^T|\phi_0}\left(IS_\eta(\phi_0) \subset C_\alpha\right) \geq \alpha.$$

If in Assumption 4 (iv), C_α satisfies $\pi_{\phi|Y^T}\left(IS_\eta(\phi) \subset C_\alpha\right) = \alpha$, $p_{Y^T|\phi_0}$-a.s., for all $T \geq 1$, C_α asymptotically attains the exact coverage probability,

$$\lim_{T\to\infty} P_{Y^T|\phi_0}\left(IS_\eta(\phi_0) \subset C_\alpha\right) = \alpha.$$

Remarks: First, unlike in Imbens and Manski (2004) and Stoye (2009), the frequentist coverage statement of C_α is for the true identified set rather than for the true value of the parameter of interest. Therefore, when η is a scalar with nonsingleton $IS_\eta(\phi_0)$, C_α will be asymptotically wider than the frequentist (connected) confidence interval for η.

Second, Theorem 4 considers pointwise asymptotic coverage rather than asymptotic uniform coverage over the sampling processes ϕ_0. The frequentist literature has stressed the importance of the uniform coverage property (e.g., Andrews and Guggenberger (2009); Stoye (2009); Romano and Shaikh (2010); Andrews and Soares (2010). Examining whether the robust posterior credible region can attain a uniformly valid coverage probability for the identified set is beyond the scope of this paper and is left for future research.

Third, C_α proposed in Moon and Schorfheide (2011) and Norets and Tang (2014) can attain asymptotically correct coverage under a different set of assumptions (Assumptions 1 and 5 (i) in this paper). Although these assumptions may be easier to check than Assumption 4, their credible region is generally conservative. In contrast, Theorem 4 shows that if C_α is constructed to satisfy (2.6) with equality (e.g., it is the smallest robust credible region C_α^*), the asymptotic coverage probability is exact. Theorem 5 in Kline and Tamer (2016) shows a similar conclusion to Theorem 4 under the conditions that the Bernstein-von Mises property holds for estimation of ϕ and that $a_T(\hat{\phi} - \phi_0)$ and $\hat{c}_T(\cdot)$ are asymptotically independent. Our Assumption 4 (iv) implies

[14] If the support function $s\left(IS_\eta(\phi), u\right)$ is not differentiable, but only directionally differentiable at some u, the asymptotic distribution of $X_{\phi|Y^T}(u)$ generally depends on $\hat{\phi}$ even asymptotically, leading to a violation of Assumption 4 (i). See Kitagawa et al. (2017) for details on the asymptotic posterior of directionally differentiable function.

the asymptotic independence condition of Kline and Tamer (2016) by assuming $\hat{c}_T$ converges to a constant. Theorem 4, on the other hand, assumes the Bernstein-von Mises property in terms of the support functions of the identified set rather than the underlying reduced-form parameters.

Assumption 4 consists of rather high-level assumptions, some of which could be difficult to check when η is a multi-dimensional object. In cases of scalar η with finite-dimensional ϕ, we can obtain a set of sufficient conditions for Assumption 4 (i) - (iii) that are simple to verify in empirical applications, e.g., the set-identified SVARs considered in Section 4.

Assumption 5. *Let the reduced-form parameter ϕ be finite-dimensional, and the parameter of interest η be a scalar. Denote the convex hull of the identified set by $co\,(IS_\eta(\phi)) = [\ell(\phi), u(\phi)]$.*

(i) The maximum likelihood estimator $\hat{\phi}$ is strongly consistent for ϕ_0, and the posterior of ϕ and the sampling distribution of the maximum likelihood estimator $\hat{\phi}$ are $\sqrt{T}$-asymptotically normal with an identical covariance matrix;

$$\begin{aligned} \sqrt{T}\left(\phi-\hat{\phi}\right)|Y^T &\rightsquigarrow \mathcal{N}\left(0,\Sigma_\phi\right), \text{ as } T\to\infty,\ p_{Y^\infty|\phi_0}\text{-a.s., and} \\ \sqrt{T}\left(\hat{\phi}-\phi_0\right)|\phi_0 &\rightsquigarrow \mathcal{N}\left(0,\Sigma_\phi\right), \text{ as } T\to\infty. \end{aligned}$$

(ii) $\ell(\phi)$ and $u(\phi)$ are continuously differentiable in an open neighborhood of ϕ_0, and their derivatives are nonzero at ϕ_0.

Assumption 5 (i) implies that likelihood-based estimation of ϕ satisfies the Bernstein–von Mises property in the sense of Theorem 7.101 in Schervish (1995). It holds when the likelihood function and the prior for ϕ satisfy the following properties: (a) regularity of the likelihood of ϕ as shown in Schervish (1995, Section 7.4) and (b) π_ϕ puts a positive probability on every open neighborhood of ϕ_0 and the density of π_ϕ is smooth at ϕ_0. Additionally imposing Assumption 5 (ii) implies applicability of the delta method to $\ell(\cdot)$ and $u(\cdot)$, which implies Assumption 4 (i) - (iii) for scalar η. In addition, it can be shown that the shortest robust credible region in (2.7) satisfies Assumption 4 (iv). Hence, C_α^* is an asymptotically valid frequentist confidence set for the true identified set with asymptotic coverage probability exactly equal to α.

Proposition 2. *Suppose Assumptions 3 and 5 hold. Assumption 4 (i) - (iii) holds true and the smallest robust credible region C_α^* defined in (2.7) satisfies Assumption 4 (iv). Hence, by Theorem 2, C_α^* is an asymptotically valid frequentist confidence set for $IS_\eta(\phi_0)$ with exact coverage,*

$$\lim_{T\to\infty} P_{Y^T|\phi_0}\left(IS_\eta\left(\phi_0\right)\subset C_\alpha^*\right)=\alpha.$$

Lemma 1 of Kline and Tamer (2016) obtains a similar result for a robust credible region different from our smallest credible region C_α^*; theirs takes the form $C_\alpha = [\ell(\hat{\phi}) - c_\alpha/\sqrt{T}, u(\hat{\phi}) + c_\alpha/\sqrt{T}]$, where c_α is chosen to satisfy (2.6) with equality.

4 Robust Bayesian Inference in SVARs

In this section we illustrate in detail the application of the proposed method to impulse-response analysis in set-identified SVARs. This section is self-contained. Consider an SVAR(p):

$$A_0 y_t = a + \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t \quad \text{for } t = 1, \dots, T, \tag{4.1}$$

where y_t is an $n \times 1$ vector and ϵ_t is an $n \times 1$ vector white noise process, normally distributed with mean zero and variance the identity matrix I_n. The initial conditions $y_1, \dots, y_p$ are given. We follow Christiano et al. (1999) and assume that one always imposes sign normalization restrictions by letting the diagonal elements of A_0 be nonnegative. This amounts to interpreting a unit positive change in a structural shock as a one standard-deviation positive shock to the corresponding endogenous variable.

The reduced-form VAR(p) representation of the model is

$$y_t = b + \sum_{j=1}^{p} B_j y_{t-j} + u_t, \tag{4.2}$$

where $b = A_0^{-1} a$, $B_j = A_0^{-1} A_j$, $u_t = A_0^{-1} \epsilon_t$, and $E(u_t u_t') \equiv \Sigma = A_0^{-1} \left(A_0^{-1}\right)'$. Since the value of the Gaussian likelihood for the SVAR (4.1) depends on $(A_0, a, A_1, \dots, A_p)$ only through $(b, B_1, \dots, B_p, \Sigma)$, we can set the reduced-form parameter to $\phi = (vec(B)', vech(\Sigma)')' \in \Phi \subset \mathcal{R}^{n+n^2p} \times \mathcal{R}^{n(n+1)/2}$, where $B = [b, B_1, \dots, B_p]$. We restrict the domain Φ to the set of ϕ's such that the reduced-form VAR(p) model can be inverted into a VMA(∞) model.

For convenience in representing and computing the identified-set correspondence of the impulse responses, we reparameterize the model and let $\theta = \left(\phi', vec(Q)'\right)' \in \Phi \times vec(\mathcal{O}(n))$, where Q is an $n \times n$ orthonormal 'rotation' matrix and $\mathcal{O}(n)$ is the set of $n \times n$ orthonormal matrices. Note that θ is a transformation of $(A_0, a, A_1, \dots, A_p)$ via $B = A_0^{-1} [a, A_1, \dots, A_p]$, $\Sigma = A_0^{-1} \left(A_0^{-1}\right)'$, and $Q = \Sigma_{tr}^{-1} A_0^{-1}$, where Σ_{tr} denotes the lower-triangular Cholesky factor with nonnegative diagonal elements of Σ. Also note that this transformation is invertible whenever Σ is nonsingular; $A_0 = Q' \Sigma_{tr}^{-1}$ and $[a, A_1, \dots, A_p] = Q' \Sigma_{tr}^{-1} B$.

Translating the sign normalization restrictions $diag(A_0) \geq 0$ into constraints on θ gives the space of structural parameters as $\Theta = \left\{ \left(\phi', vec(Q)'\right)' \in \Phi \times vec(\mathcal{O}(n)) : diag\left(Q' \Sigma_{tr}^{-1}\right) \geq 0 \right\}$. Individually, the sign normalization restrictions can be written as linear inequalities

$$\left(\sigma^i\right)' q_i \geq 0 \quad \text{for all } i = 1, \dots, n, \tag{4.3}$$

where $\left[\sigma^1, \sigma^2, \dots, \sigma^n\right]$ are the column vectors of Σ_{tr}^{-1} and $[q_1, q_2, \dots, q_n]$ are the column vectors of Q.

Assuming the lag polynomial $\left(I_n - \sum_{j=1}^{p} B_j L^p\right)$ is invertible (which is implied by the domain restriction on Φ) the VMA(∞) representation of the model is:

$$\begin{aligned} y_t &= c + \sum_{j=0}^{\infty} C_j u_{t-j} \\ &= c + \sum_{j=0}^{\infty} C_j \Sigma_{tr} Q \epsilon_{t-j}, \end{aligned} \tag{4.4}$$

where C_j is the j-th coefficient matrix of $\left(I_n - \sum_{j=1}^{p} B_j L^j\right)^{-1}$.

We denote the h-th horizon impulse response by the $n \times n$ matrix IR^h, $h = 0, 1, 2, \ldots$

$$IR^h = C_h \Sigma_{tr} Q, \tag{4.5}$$

the long-run impulse-response matrix by

$$IR^\infty = \lim_{h \to \infty} IR^h = \left(I_n - \sum_{j=1}^{p} B_j\right)^{-1} \Sigma_{tr} Q, \tag{4.6}$$

and the long-run cumulative impulse-response matrix by

$$CIR^\infty = \sum_{h=0}^{\infty} IR^h = \left(\sum_{h=0}^{\infty} C_h\right) \Sigma_{tr} Q. \tag{4.7}$$

The scalar parameter of interest η is a single impulse-response, i.e., the (i,j)-element of IR^h, which can be expressed as

$$\eta = IR^h_{ij} \equiv e_i' C_h \Sigma_{tr} Q e_j \equiv c_{ih}'(\phi)\, q_j = \eta(\phi, Q), \tag{4.8}$$

where e_i is the i-th column vector of the identity matrix I_n and $c_{ih}'(\phi)$ is the i-th row vector of $C_h \Sigma_{tr}$. Note that the analysis developed below for the impulse responses can be easily extended to the structural parameters A_0 and $[A_1, \ldots, A_p]$, since the (i,j)-th element of A_l can be obtained as $e_j' \left(\Sigma_{tr}^{-1} B_l\right)' q_i$, with $B_0 = I_n$.

4.1 Set Identification in SVARs

Set identification in an SVAR arises when knowledge of the reduced-form parameter ϕ does not pin down a unique A_0. Since any $A_0 = Q' \Sigma_{tr}^{-1}$ satisfies $\Sigma = (A_0' A_0)^{-1}$, in the absence of identifying restrictions $\left\{A_0 = Q' \Sigma_{tr}^{-1} : Q \in \mathcal{O}(n)\right\}$ is the identified set of A_0's, i.e., the set of A_0's that are consistent with ϕ (Uhlig (2005), Proposition A.1). Imposing identifying restrictions can be viewed as restricting the set of feasible $Q's$ to lie in a subspace $\mathcal{Q}$ of $\mathcal{O}(n)$, so that the identified set of A_0 is $\left\{A_0 = Q' \Sigma_{tr}^{-1} : Q \in \mathcal{Q}\right\}$ and the corresponding identified set of η is:

$$IS_\eta(\phi) = \left\{\eta(\phi, Q) : Q \in \mathcal{Q}\right\}. \tag{4.9}$$

In the following we characterize the subspace $\mathcal{Q}$ under common identifying restrictions.

4.2 Identifying Restrictions

4.2.1 Under-identifying Zero Restrictions

Examples of under-identifying zero restrictions typically used in the literature are restrictions on some off-diagonal elements of A_0, on the lagged coefficients $\{A_l : l = 1, \ldots p\}$, on contemporaneous impulse responses $IR^0 = A_0^{-1}$, and on the long-run responses IR^∞ in (4.6) or CIR^∞ in (4.7).

All these restrictions can be viewed as linear constraints on the columns of Q. For example:

$$\begin{aligned}
((i,j)\text{-th element of } A_0) &= 0 \iff \left(\Sigma_{tr}^{-1} e_j\right)' q_i = 0, \\
((i,j)\text{-th element of } A_l) &= 0 \iff \left(\Sigma_{tr}^{-1} B_l e_j\right)' q_i = 0, \\
\left((i,j)\text{-th element of } A_0^{-1}\right) &= 0 \iff \left(e_i' \Sigma_{tr}\right) q_j = 0, \\
((i,j)\text{-th element of } CIR^\infty) &= 0 \iff \left[e_i' \sum_{h=0}^{\infty} C_h(B)\, \Sigma_{tr}\right] q_j = 0.
\end{aligned} \tag{4.10}$$

We can thus represent a collection of zero restrictions in the general form:

$$F(\phi, Q) \equiv \begin{pmatrix} F_1(\phi)\, q_1 \\ F_2(\phi)\, q_2 \\ \vdots \\ F_n(\phi)\, q_n \end{pmatrix} = \mathbf{0}, \tag{4.11}$$

where $F_i(\phi)$ is an $f_i \times n$ matrix. Each row in $F_i(\phi)$ corresponds to the coefficient vector of a zero restriction that constrains q_i as in (4.10), and $F_i(\phi)$ stacks all the coefficient vectors that multiply q_i into a matrix. Hence, f_i is the number of zero restrictions constraining q_i. If the zero restrictions do not constrain q_i, $F_i(\phi)$ does not exist and $f_i = 0$.

In order to implement our method, one must first order the variables in the model.

Definition 3 (Ordering of Variables). *Order the variables in the SVAR so that the number of zero restrictions f_i imposed on the i-th column of Q (i.e., the rows of $F_i(\phi)$ in (4.11)) satisfy $f_1 \geq f_2 \geq \cdots \geq f_n \geq 0$. In case of ties, if the impulse response of interest is that to the j-th structural shock, order the j-th variable first. That is, set $j = 1$ when no other column vector has a larger number of restrictions than q_j. If $j \geq 2$, then order the variables so that $f_{j-1} > f_j$.*[15]

Rubio-Ramirez et. al. (2010) show that, under regularity assumptions, a necessary and sufficient condition for point identification is that $f_i = n - i$ for all $i = 1, \ldots, n$. Here we consider restrictions that make the SVAR set-identified because

$$f_i \leq n - i \quad \text{for all } i = 1, \ldots, n, \tag{4.12}$$

[15] Our assumption for the ordering of the variables pins down a unique j, while it does not necessarily yield a unique ordering for the other variables if some of them admit the same number of constraints. However, the condition for the convexity of the identified set for the impulse responses to the j-th structural shock that we provide in Appendix B is not affected by the ordering of the other variables as long as the f_i's are in decreasing order.

with strict inequality for at least one $i \in \{1, \dots, n\}$.[16]

The following example illustrates how to order the variables in order to satisfy Definition 3.

Example 1. *Consider a SVAR for* $(\pi_t, y_t, m_t, i_t)'$, *where* π_t *is inflation,* y_t *is (detrended) real GDP,* m_t *is the (detrended) real money stock and* i_t *is the nominal interest rate. Consider the following under-identifying restrictions imposed on* A_0^{-1},

$$\begin{pmatrix} u_{\pi,t} \\ u_{y,t} \\ u_{m,t} \\ u_{i,t} \end{pmatrix} = \begin{pmatrix} a^{11} & a^{12} & 0 & 0 \\ a^{21} & a^{22} & 0 & 0 \\ a^{31} & a^{32} & a^{33} & a^{34} \\ a^{41} & a^{42} & a^{43} & a^{44} \end{pmatrix} \begin{pmatrix} \epsilon_{\pi,t} \\ \epsilon_{y,t} \\ \epsilon_{m,t} \\ \epsilon_{i,t} \end{pmatrix}. \tag{4.13}$$

Let the objects of interest be the impulse responses to $\epsilon_{i,t}$ *(a monetary policy shock). Let* $[q^{\pi}, q^{y}, q^{m}, q^{i}]$ *be a* 4×4 *orthonormal matrix. By (4.10), the imposed restrictions imply two restrictions on* q^m *and two restrictions on* q^i. *An ordering consistent with Definition 3 is* $(i_t, m_t, \pi_t, y_t)'$, *and the corresponding numbers of restrictions are* $(f_1, f_2, f_3, f_4) = (2, 2, 0, 0)$ *with* $j = 1$. *The restrictions in this example satisfy (4.12). If instead the objects of interest are the impulse responses to* $\epsilon_{y,t}$ *(interpreted as a demand shock), order the variables as* (i_t, m_t, y_t, π_t) *and let* $j = 3$.

4.2.2 Sign Restrictions

Sign restrictions could be considered alone or in addition to zero restrictions. If there are zero restrictions, we maintain the order of the variables as in Definition 3. If there are only sign restrictions, we order first the variable whose structural shock is of interest, i.e., $j = 1$. Suppose there are sign restrictions on the responses to the j-th structural shock. Sign restrictions are linear constraints on the columns of Q: $S_{hj}(\phi)\, q_j \geq \mathbf{0}$,[17] where $S_{hj}(\phi) \equiv D_{hj} C_h(B)\, \Sigma_{tr}$ is an $s_{hj} \times n$ matrix, and D_{hj} is an $s_{hj} \times n$ matrix that selects the sign-restricted responses from the $n \times 1$ impulse-response vector $C_h(B)\, \Sigma_{tr} q_j$. The nonzero elements of D_{hj} equal 1 or -1 depending on whether the corresponding impulse responses are positive or negative.

Stacking $S_{hj}(\phi)$ over multiple horizons gives the set of sign restrictions on the responses to the j-th shock as

$$S_j(\phi)\, q_j \geq \mathbf{0}, \tag{4.14}$$

[16] The class of under-identified models considered here does not exhaust the universe of all possible non-identified SVARs, since there exist models that do not satisfy (4.12), but for which the structural parameter is not globally identified for some values of the reduced-form parameter. For instance, in the example in Section 4.4 of Rubio-Ramírez et al. (2010), with $n = 3$ and $f_1 = f_2 = f_3 = 1$, the structural parameter is locally, but not globally, identified. For another example, the zero restrictions in page 77 of Christiano et al. (1999) correspond to a case with $n = 3$ and $f_1 = f_2 = f_3 = 1$ where even local identification fails. These cases are ruled out by condition (4.12).

[17] In this section, for a vector $y = (y_1, \dots, y_m)'$, $y \geq \mathbf{0}$ means $y_i \geq 0$ for all $i = 1, \dots, m$, and $y > \mathbf{0}$ means $y_i \geq 0$ for all $i = 1, \dots, m$ and $y_i > 0$ for some $i \in \{1, \dots, m\}$.

where $S_j(\phi)$ is a $\left(\sum_{h=0}^{\bar{h}} s_{hj}\right) \times n$ matrix $S_j(\phi) = \left[S_{0j}(\phi)', \ldots, S_{\bar{h}j}(\phi)\right]'$, where $0 \leq \bar{h} \leq \infty$ is the maximal horizon in the impulse-response analysis. If there are no sign restrictions on the $\tilde{h}$-th horizon responses, $\tilde{h} \in \{0, \ldots, \bar{h}\}$, $s_{\tilde{h}j} = 0$ and $S_{\tilde{h}j}(\phi)$ is not present in $S_j(\phi)$.

Denote by $\mathcal{I}_S \subset \{1, 2, \ldots, n\}$ the set of indices such that $j \in \mathcal{I}_S$ if some of the impulse responses to the j-th structural shock are sign-constrained. The set of all the sign restrictions can be expressed as $S_j(\phi) q_j \geq \mathbf{0}$, for $j \in \mathcal{I}_S$, or, as a shorthand notation, as

$$S(\phi, Q) \geq \mathbf{0}. \tag{4.15}$$

Note that the sign restrictions do not have to be limited to the impulse responses. Since $A_0' = \left(\Sigma_{tr}^{-1}\right)' Q$ and $A_l' = B_l' \left(\Sigma_{tr}^{-1}\right)' Q$, $l = 1, \ldots, p$, any sign restrictions on the j-th row of A_0 or A_l take the form of linear inequalities for q_j, so they could be appended to $S_j(\phi)$ in (4.14).

4.3 The Impulse-Response Identified Set

The identified set for the impulse response in the presence of under-identifying zero restrictions and sign restrictions is given by:

$$IS_\eta(\phi|F, S) = \{\eta(\phi, Q) : Q \in \mathcal{Q}(\phi|F, S)\}, \tag{4.16}$$

where $\mathcal{Q}(\phi|F, S)$ is the set of Q's that jointly satisfy the sign restrictions (4.15), the zero restrictions (4.11) and the sign normalizations (4.3),

$$\mathcal{Q}(\phi|F, S) = \left\{Q \in \mathcal{O}(n) : S(\phi, Q) \geq \mathbf{0},\ F(\phi, Q) = \mathbf{0},\ diag\left(Q'\Sigma_{tr}^{-1}\right) \geq \mathbf{0}\right\}. \tag{4.17}$$

Proposition 3 below shows that, unlike when there are only zero restrictions, when there are sign restrictions $\mathcal{Q}(\phi|F, S)$ can be empty, in which case the identified set of η is defined as an empty set.

4.4 Multiple Priors in SVARs

Let $\tilde{\pi}_\phi$ be a prior for the reduced-form parameter. Since the identifying restrictions can be observationally restrictive, we ensure that the prior for ϕ is consistent with Assumption 1 (i) by trimming the support of $\tilde{\pi}_\phi$ as

$$\pi_\phi \equiv \frac{\tilde{\pi}_\phi 1\{\mathcal{Q}(\phi|F, S) \neq \emptyset\}}{\tilde{\pi}_\phi(\{\mathcal{Q}(\phi|F, S) \neq \emptyset\})}, \tag{4.18}$$

where $\{\phi \in \Phi : \mathcal{Q}(\phi|F, S) \neq \emptyset\}$ is the set of reduced-form parameters that yields nonempty identified sets for any structural parameters or the impulse responses.

A joint prior for $\theta = (\phi, Q) \in \Phi \times \mathcal{O}(n)$ that has ϕ-marginal π_ϕ can be expressed as $\pi_\theta = \pi_{Q|\phi}\pi_\phi$, where $\pi_{Q|\phi}$ is supported only on $\mathcal{Q}(\phi|F, S)$. Since $(A_0, A_1, \ldots, A_p)$ and η are functions of $\theta = (\phi, Q)$,

π_θ induces a unique prior for the structural parameters and the impulse responses. Conversely, a prior for $(A_0, A_1, \ldots, A_p)$ that incorporates the sign normalizations induces a unique prior for π_θ. While the prior for ϕ is updated by the data, the conditional prior $\pi_{Q|\phi}$ is not updated.

Under point identification the restrictions pin down a unique Q (i.e., $\mathcal{Q}(\phi|F,S)$ is a singleton), in which case $\pi_{Q|\phi}$ is degenerate and gives a point mass at such Q. Specifying π_ϕ thus suffices to induce a single posterior for the structural parameters and for the impulse responses. In contrast, in the set-identified case where $\mathcal{Q}(\phi|F,S)$ is non-singleton for ϕ's with a positive measure, specifying only π_ϕ cannot yield a single posterior. To obtain a single posterior one would need to specify $\pi_{Q|\phi}$, which is supported only on $\mathcal{Q}(\phi|F,S)$ at each $\phi \in \Phi$. This is the standard Bayesian approach adopted by the vast majority of the empirical literature using set-identified SVARs (e.g., Uhlig (2005)), and its potential pitfalls have been discussed by Baumeister and Hamilton (2015).[18]

The robust Bayesian procedure in this paper does not require specifying a prior for $\pi_{Q|\phi}$, but considers the class of all priors $\pi_{Q|\phi}$. In the current SVAR application, the set of priors introduced in Definition 2 can be expressed as

$$\Pi_{Q|\phi} = \left\{\pi_{Q|\phi} : \pi_{Q|\phi}\left(\mathcal{Q}(\phi|F,S)\right) = 1,\ \pi_\phi\text{-almost surely}\right\}. \tag{4.19}$$

Combining $\Pi_{Q|\phi}$ with the posterior for ϕ generates the class of posteriors for $\theta = (\phi, Q)$,

$$\Pi_{\theta|Y} = \left\{\pi_{\theta|Y} = \pi_{Q|\phi}\pi_{\phi|Y} : \pi_{Q|\phi} \in \Pi_{Q|\phi}\right\}. \tag{4.20}$$

Marginalizing these posteriors to the impulse response η yields the class of posteriors (2.4). In the current notation for SVARs,

$$\Pi_{\eta|Y} \equiv \left\{\pi_{\eta|Y}(\cdot) = \int \pi_{Q|Y}\left(\eta(\phi,Q) \in \cdot\right) d\pi_{\phi|Y} : \pi_{Q|Y} \in \Pi_{Q|Y}\right\}. \tag{4.21}$$

4.5 Set of Posterior Means and Robust Credible Region

Applying Theorem 2 to the impulse response (scalar), we obtain the set of posterior means:

$$\left[\int_\Phi \ell(\phi) d\pi_{\phi|Y}, \int_\Phi u(\phi)\, d\pi_{\phi|Y}\right],$$

where $\ell(\phi) = \inf\left\{\eta(\phi,Q) : Q \in \mathcal{Q}(\phi|F,S)\right\}$ and $u(\phi) = \sup\left\{\eta(\phi,Q) : Q \in \mathcal{Q}(\phi|F,S)\right\}$. Section 5 discusses computation of $\ell(\phi)$ and $u(\phi)$.

The smallest robust credible region with credibility α for the impulse response can be computed using draws of $[\ell(\phi), u(\phi)]$, $\phi \sim \pi_{\phi|Y}$ and applying Proposition 1. It is interpreted as the shortest interval estimate for the impulse response η, such that the posterior probability put on the interval is greater than or equal to α uniformly over the posteriors in the class (4.21).

[18]Since (ϕ, Q) and $(A_0, A_1, \ldots, A_p)$ are one-to-one (under the sign normalizations), the difficulty of specifying a prior for $\pi_{Q|\phi}$ can be equivalently stated as the difficulty of specifying a joint prior for all structural parameters that is compatible with π_ϕ.

4.6 Verifying the Assumptions for Frequentist Validity

To validate the frequentist interpretation of the set of posterior means, this section examines convexity, continuity, and differentiability of the identified set map $IS_\eta(\phi|F,S)$ for the impulse response. By Theorems 2 and 3 (ii), convexity and continuity of $IS_\eta(\phi|F,S)$ as a function of ϕ allow us to interpret the set of posterior means as a consistent estimator of the true identified set. In addition, if $[\ell(\phi), u(\phi)]$ is differentiable in ϕ (Assumption 5 (ii)), Proposition 5 guarantees that the robust credible region is an asymptotically valid confidence set for the true identified set.

4.6.1 Convexity

The next proposition shows conditions for the convexity of the impulse-response identified set. See Appendix B for additional analytical results, examples of convex and nonconvex impulse-response identified sets, and further discussion.

Proposition 3 (Convexity). *Let the object of interest be $\eta = c'_{ih}(\phi)\, q_{j^*}$, the impulse response to the j^*-th structural shock, $i \in \{1,2,\dots,n\}$, $h \in \{0,1,2,\dots\}$, where the variables are ordered according to Definition 3.*

(I) Suppose there are only zero restrictions of the form (4.11). Assume $f_i \leq n-i$ for all $i = 1,\dots,n$. Then, for every i and h, and almost every $\phi \in \Phi$, the identified set of η is non-empty and bounded, and it is convex if any of the following mutually exclusive conditions holds:

(i) $j^* = 1$ and $f_1 < n-1$.

(ii) $j^* \geq 2$, and $f_i < n-i$ for all $i = 1,\dots,j^*-1$.

(iii) $j^* \geq 2$ and there exists $1 \leq i^* \leq j^*-1$ such that $f_i < n-i$ for all $i = i^*+1,\dots,j^*$ and $[q_1,\dots,q_{i^*}]$ is exactly identified, meaning that, for almost every $\phi \in \Phi$, the constraints

$$\begin{pmatrix} F_1(\phi)\, q_1 \\ F_2(\phi)\, q_2 \\ \vdots \\ F_{i^*}(\phi)\, q_{i^*} \end{pmatrix} = \mathbf{0}$$

and the sign-normalizations $\left(\sigma^i\right)' q_i \geq 0$, $i = 1,\dots,i^*$, pin down a unique $[q_1,\dots,q_{i^*}]$.[19]

(II) Consider the case with both zero and sign restrictions, and suppose that sign restrictions are placed only on the responses to the j^*-th structural shock, i.e., $\mathcal{I}_S = \{j^*\}$.

[19] If $rank(F_i(\phi)) = f_i$ for all $i = 1,\dots,i^*$, and for almost every $\phi \in \Phi$, a necessary condition for exact identification of $[q_1,\dots,q_{i^*}]$ is that $f_i = n-i$ for all $i = 1,2,\dots,i^*$. One can check if the condition is also sufficient by checking that Algorithm 1 of Rubio-Ramírez et al. (2010) (reported as Algorithm 3 in Appendix B) yields a unique set of orthonormal vectors $[q_1,\dots,q_{i^*}]$ for every ϕ randomly drawn from a prior supporting the entire Φ.

(iv) Suppose the zero restrictions satisfy one of conditions (i) and (ii) in the current proposition. If there exists a unit-length vector $q \in \mathcal{R}^n$ such that

$$F_{j^*}(\phi)\, q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi) \\ (\sigma^{j^*})' \end{pmatrix} q > \mathbf{0}, \tag{4.22}$$

then the identified set of η, $IS_\eta(\phi|F,S)$, is nonempty and convex for every i and h.

(v) Suppose that the zero restrictions satisfy condition (iii) in the current proposition. Let $[q_1(\phi), \ldots, q_{i^*}(\phi)]$ be the first i^*-th orthonormal vectors that are exactly identified (see condition (iii)). If there exists a unit length vector $q \in \mathcal{R}^n$ such that

$$\begin{pmatrix} F_{j^*}(\phi) \\ q_1'(\phi) \\ \vdots \\ q_{i^*}'(\phi) \end{pmatrix} q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi) \\ (\sigma^{j^*})' \end{pmatrix} q > \mathbf{0}, \tag{4.23}$$

then the identified set of η, $IS_\eta(\phi|F,S)$, is nonempty and convex for every i and h.

Proposition 3 shows that when a set of zero restrictions satisfies $f_i \leq n-i$ for all $i = 1, 2, \ldots, n$, the identified set for the impulse response is never empty, so the zero restrictions cannot be refuted by data. The plausibility of the identifying restrictions defined in Section 2.6.1 is always one in this case. When there are also sign restrictions, we can have an empty identified set and a non-trivial value for the plausibility of the identifying restrictions.

Lemma 1 of Granziera et al. (2018) shows convexity of the impulse-response identified set for the special case where zero and sign restrictions are imposed only on responses to the j^*-th shock, i.e., $j^* = 1$, $f_i = 0$ for all $i = 2, \ldots, n$, and $\mathcal{I}_S = \{1\}$ in our notation. Proposition 3 extends their result to the case where zero restrictions are placed on the columns of Q other than q_{j^*}. The inequality conditions (iv) and (v) of Proposition 3 imply that the set of feasible q's does not collapse to a one-dimensional subspace in $\mathcal{R}^n$. If the set of feasible q's becomes degenerate, non-convexity arises since the intersection of a one-dimensional subspace in $\mathcal{R}^n$ with the unit sphere consists of two disconnected points.[20]

4.6.2 Continuity and Differentiability

One of the key assumptions for Theorem 3 is the continuity of $IS_\eta(\phi)$ at $\phi = \phi_0$ (Assumption 2(i)).[21] The next proposition shows that in SVARs this continuity property is ensured by mild regularity conditions on the coefficient matrices of the zero and sign restrictions.

[20]If the set of ϕ's that leads to such degeneracy has measure zero in Φ, then, as a corollary of Proposition 3, we can claim that the impulse response identified set is convex for almost all ϕ conditional on it being nonempty.

[21]Proposition 3 shows boundedness of $IS_\eta(\phi|F,S)$ for all ϕ so that Assumption 2 (iii) also holds.

Proposition 4 (Continuity). *Let* $\eta = c'_{ih}(\phi) q_{j^*}$, $i \in \{1,\ldots,n\}$, $h \in \{0,1,2,\ldots\}$, *be the impulse response of interest. Suppose that the variables are ordered according to Definition 3 and sign restrictions are placed only on the responses to the* j^**-th structural shock, i.e.,* $\mathcal{I}_S = \{j^*\}$.

(i) Suppose that the zero restrictions satisfy one of Conditions (i) and (ii) of Proposition 3. If there exists an open neighborhood of ϕ_0, $G \subset \Phi$, *such that* $rank(F_{j^*}(\phi)) = f_{j^*}$ *for all* $\phi \in G$, *and if there exists a unit-length vector* $q \in \mathcal{R}^n$ *such that*

$$F_{j^*}(\phi_0) q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi_0) \\ (\sigma^{j^*}(\phi_0))' \end{pmatrix} q >> \mathbf{0},$$

then the identified set correspondence $IS_\eta(\phi|F,S)$ *is continuous at* $\phi = \phi_0$ *for every* i *and* h.[22] *(ii) Suppose that the zero restrictions satisfy condition (iii) of Proposition 3, and let* $[q_1(\phi),\ldots,q_{i^*}(\phi)]$ *be the first* i^**-th column vectors of* Q *that are exactly identified. If there exists an open neighborhood of* ϕ_0, $G \subset \Phi$, *such that* $\begin{pmatrix} F_{j^*}(\phi) \\ q'_1(\phi) \\ \vdots \\ q'_{i^*}(\phi) \end{pmatrix}$ *is a full row-rank matrix for all* $\phi \in G$, *and if there exists a unit-length vector* $q \in \mathcal{R}^n$ *such that*

$$\begin{pmatrix} F_{j^*}(\phi_0) \\ q'_1(\phi_0) \\ \vdots \\ q'_{i^*}(\phi_0) \end{pmatrix} q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi_0) \\ (\sigma^{j^*}(\phi_0))' \end{pmatrix} q >> \mathbf{0},$$

then the identified-set correspondence $IS_\eta(\phi|F,S)$ *is continuous at* $\phi = \phi_0$ *for every* i *and* h.

In the development of a delta method for the endpoints of the impulse-response identified set, Theorem 2 in Gafarov et al. (2018) shows their directional differentiability. We restrict our analysis to settings of Proposition 3 where the identified set is guaranteed to be convex. Adopting and extending Theorem 2 of Gafarov et al. (2018), we obtain the following sufficient condition for differentiability of $\ell(\phi)$ and $u(\phi)$

Proposition 5 (Differentiability). *Let* $\eta = c'_{ih}(\phi) q_{j^*}$, $i \in \{1,\ldots,n\}$, $h \in \{0,1,2,\ldots\}$, *be the impulse response of interest. Suppose that the variables are ordered according to Definition 3 and sign restrictions are placed only on the responses to the* j^**-th structural shock, i.e.,* $\mathcal{I}_S = \{j^*\}$.

(i) Suppose that the zero restrictions satisfy one of Conditions (i) or (ii) of Proposition 3 and the column vectors of $[F_{j^*}(\phi_0)', S_{j^*}(\phi_0)', \sigma^{j^*}(\phi_0)]$ *are linearly independent. If the set of solutions of the following optimization problem,*

$$\min_{q \in \mathcal{S}^{n-1}} \left\{ c'_{ih}(\phi_0) q \right\} \text{ (resp. } \max_{q \in \mathcal{S}^{n-1}} \left\{ c'_{ih}(\phi_0) q \right\}) \text{ s.t. } F_{j^*}(\phi_0) q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi_0) \\ (\sigma^{j^*}(\phi_0))' \end{pmatrix} q \geq \mathbf{0},$$

[22]For a vector $y = (y_1,\ldots,y_m)'$, $y >> \mathbf{0}$ means $y_i > 0$ for all $i = 1,\ldots,m$.

$$(4.24)$$

is singleton, the optimized value $\ell(\phi_0)$ (resp. $u(\phi_0)$) is nonzero, and the number of binding sign restrictions at the optimum is less than or equal to $n - f_{j^*} - 1$, then $\ell(\phi)$ (resp. $u(\phi)$) is differentiable at $\phi = \phi_0$.

(ii) Suppose that the zero restrictions satisfy Conditions (iii) of Proposition 3. Let $[q_1(\phi_0), \ldots, q_{i^*}(\phi_0)]$ be the first i^*-th column vectors of Q that are exactly identified at $\phi = \phi_0$. Assume that the column vectors of $[F_{j^*}(\phi_0)', S_{j^*}(\phi_0)', \sigma^{j^*}(\phi_0), q_1(\phi_0), \ldots, q_{i^*}(\phi_0)]$ are linearly independent. If the set of solutions of the following optimization problem,

$$\min_{q \in \mathcal{S}^{n-1}} \left\{ c_{ih}'(\phi_0)\, q \right\} \quad (resp. \max_{q \in \mathcal{S}^{n-1}} \left\{ c_{ih}'(\phi_0)\, q \right\}) \quad s.t. \begin{pmatrix} F_{j^*}(\phi_0) \\ q_1(\phi_0)' \\ \vdots \\ q_{i^*}(\phi_0)' \end{pmatrix} q = 0 \text{ and } \begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} q \geq \mathbf{0}, \tag{4.25}$$

is singleton, the optimized value $\ell(\phi_0)$ (resp. $u(\phi_0)$) is nonzero, and the number of binding sign restrictions at the optimum is less than or equal to $n - f_{j^*} - i^* - 1$, then $\ell(\phi)$ (resp. $u(\phi)$) is differentiable at $\phi = \phi_0$.

Theorem 2 in Gafarov et al. (2018) concerns Case (i) of Proposition 3 with sign restrictions placed on $\mathcal{I}_S = \{1\}$ and no zero restrictions on the other shocks, $f_2 = \cdots = f_n = 0$. Proposition 5 extends Theorem 2 in Gafarov et al. (2018) to the setting where we impose the zero restrictions on the column vectors of Q other than j^* subject to the conditions for convexity of the identified set characterized in Proposition 3.[23]

5 Numerical Implementation

We present an algorithm to numerically approximate the set of posterior means, the robust credible region and the diagnostic tool discussed in Section 2.6.1 for the case of impulse-response analysis in SVARs. The algorithm assumes that the variables are ordered according to Definition 3 and any imposed zero restriction satisfies (4.12).

Matlab code implementing the procedure can be obtained from the authors' personal websites or upon request. The code checks the conditions in Definition 3, condition (4.12) and also the

[23]The statement of Theorem 2 of Gafarov et al. (2018) does not explicitly constrain the maximal number of binding inequality restrictions at the optimum (cf. Proposition 5 in this paper), while their proof implicitly does so. The condition for the maximal number of binding inequality restrictions implies that if the optimum is attained at one of the vertices on the constrained surface in $\mathcal{S}^{n-1}$, this vertex has to be a basic feasible solution (i.e., exactly $n-1$ equality and sign restrictions have to be active).

convexity of the identified set using Proposition 3. The code further gives the user various options, such as reporting the standard Bayesian output in addition to the robust Bayesian output or computing the bounds analytically using the method of Gafarov et al. (2018), when it is applicable.

Algorithm 1 *Let $F(\phi, Q) = \mathbf{0}$ and $S(\phi, Q) \geq \mathbf{0}$ be the set of identifying restrictions (one or both of which may be empty), and let $\eta = c'_{ih}(\phi) q_{j^*}$ be the impulse response of interest.*

(Step 1). Specify $\tilde{\pi}_\phi$, the prior for the reduced-form parameter ϕ.[24] *Estimate a Bayesian reduced-form VAR to obtain the posterior $\tilde{\pi}_{\phi|Y}$.*

(Step 2). Draw ϕ from $\tilde{\pi}_{\phi|Y}$.[25] *Given the draw of ϕ, check whether $\mathcal{Q}(\phi|F, S)$ is empty by following the subroutine (Step 2.1) – (Step 2.3) below.*

(Step 2.1). Let $z_1 \sim \mathcal{N}(\mathbf{0}, I_n)$ be a draw of an n-variate standard normal random variable. Let $\tilde{q}_1 = \mathcal{M}_1 z_1$ be the $n \times 1$ residual vector in the linear projection of z_1 onto an $n \times f_1$ regressor matrix $F_1(\phi)'$. For $i = 2, 3, \ldots, n$, run the following procedure sequentially: draw $z_i \sim \mathcal{N}(0, I_n)$ and compute $\tilde{q}_i = \mathcal{M}_i z_i$, where $\mathcal{M}_i z_i$ is the residual vector in the linear projection of z_i onto the $n \times (f_i + i - 1)$ matrix $\left[F_i(\phi)', \tilde{q}_1, \ldots, \tilde{q}_{i-1}\right]$. The vectors $\tilde{q}_1, ..., \tilde{q}_n$ are orthogonal and satisfy the equality restrictions.

(Step 2.2). Given $\tilde{q}_1, \ldots, \tilde{q}_n$ obtained in the previous step, define

$$Q = \left[sign\left(\left(\sigma^1\right)' \tilde{q}_1 \right) \frac{\tilde{q}_1}{\|\tilde{q}_1\|}, \ldots, sign\left(\left(\sigma^n\right)' \tilde{q}_n \right) \frac{\tilde{q}_n}{\|\tilde{q}_n\|} \right],$$

where $\|\cdot\|$ is the Euclidian metric in $\mathcal{R}^n$. If $\left(\sigma^i\right)' \tilde{q}_i$ is zero for some i, set $sign\left(\left(\sigma^i\right)' \tilde{q}_i\right)$ equal to 1 or -1 with equal probability. This step imposes the sign normalization that the diagonal elements of A_0 are nonnegative.

(Step 2.3). Check whether Q obtained in (Step 2.2) satisfies the sign restrictions $S(\phi, Q) \geq \mathbf{0}$. If so, retain this Q and proceed to (Step 3). Otherwise, repeat (Step 2.1) and (Step 2.2) a maximum of L times (e.g. $L = 3000$) or until Q is obtained satisfying $S(\phi, Q) \geq \mathbf{0}$. If none of the L draws of Q satisfies $S(\phi, Q) \geq \mathbf{0}$, approximate $\mathcal{Q}(\phi|F, S)$ as being empty and return to Step 2 to obtain a new draw of ϕ.

(Step 3). Given ϕ and Q obtained in (Step 2), compute the lower and upper bounds of $IS_\eta(\phi|S, F)$

[24] $\tilde{\pi}_\phi$ need not be proper, nor satisfy the condition $\tilde{\pi}_\phi(\{\phi : \mathcal{Q}(\phi|F, S) \neq \emptyset\}) = 1$ (that is, the prior may assign positive probability to regions of the reduced-form parameter space that yield an empty set of Q's satisfying the zero and sign restrictions).

[25] Available methods for drawing ϕ depend on the form of the posterior. For example, given a normal-inverse-Wishart conjugate prior, ϕ can be drawn from the normal-inverse-Wishart posterior (see, for example, Arias et al. (2018)). If the posterior is non-standard, Markov Chain Monte Carlo methods could be used. For the purpose of our algorithm, all that matters is that ϕ can be drawn from its posterior.

by solving the following constrained nonlinear optimization problem:

$$\begin{aligned} & \ell(\phi) = \arg\min_{Q} c_{ih}'(\phi)\, q_{j^*}, \\ s.t. \quad & Q'Q = I_n, \;\; F(\phi, Q) = \mathbf{0}, \\ & diag(Q'\Sigma_{tr}^{-1}) \geq 0, \text{ and } S(\phi, Q) \geq \mathbf{0}, \end{aligned}$$

and $u(\phi) = \arg\max_{Q} c_{ih}'(\phi)\, q_{j^*}$ *under the same set of constraints.*

(Step 4). Repeat (Step 2) – (Step 3) M times to obtain $[\ell(\phi_m), u(\phi_m)]$, $m = 1, \ldots, M$. *Approximate the set of posterior means by the sample averages of* $(\ell(\phi_m) : m = 1, \ldots, M)$ *and* $(u(\phi_m) : m = 1, \ldots, M)$.

(Step 5). To obtain an approximation of the smallest robust credible region with credibility $\alpha \in (0, 1)$, *define* $d(\eta, \phi) = \max\{|\eta - \ell(\phi)|, |\eta - u(\phi)|\}$, *and let* $\hat{z}_\alpha(\eta)$ *be the sample* α*-th quantile of* $(d(\eta, \phi_m) : m = 1, \ldots, M)$. *An approximated smallest robust credible region for* η *is an interval centered at* $\arg\min_\eta \hat{z}_\alpha(\eta)$ *with radius* $\min_\eta \hat{z}_\alpha(\eta)$.[26]

(Step 6). The proportion of drawn ϕ'*s that pass Step 2.3 is an approximation of the posterior probability of having a nonempty identified set,* $\tilde{\pi}_{\phi|Y}(\{\phi : \mathcal{Q}(\phi|F, S) \neq \emptyset\})$, *corresponding to the diagnostic tool discussed in Section 2.6.1.*

Remarks: First, the step of the algorithm drawing orthonormal Q's subject to zero- and sign restrictions (Step 2) is common to our approach and the existing standard Bayesian approach of, for example, Arias et al. (2018). In particular, Step 2.1 is similar to Steps 2 and 3 in Algorithm 2 of Arias et al. (2018), but uses a linear projection instead of their QR decomposition and imposes different sign normalizations. The Matlab code we provide also offers the option of using a QR decomposition.[27]

Second, the optimization step (Step 3) is a non-convex optimization problem and the convergence of gradient-based optimization methods like the one we use in the Matlab code is not guaranteed. To mitigate this problem, at each draw of ϕ one can draw multiple values of Q from $\mathcal{Q}(\phi|F, S)$ to use as starting values in the optimization step, and then take the optimum over the solutions obtained from the different starting values.

Third, if the zero and sign restrictions restrict only a single column of Q, Steps 2.1–2.3 and 3 can be substituted by an analytical computation of the bounds of the identified set at each draw of ϕ, using the result of Gafarov *et al.* (2018).[28] Their paper applies the result at $\hat{\phi}$ in a frequentist

[26] In practice we obtain this interval by grid search using a fine grid over η. The objective function in this problem is non-differentiable, so gradient-based optimization methods are inappropriate.

[27] In our experience, the two ways of drawing Q are comparable both in terms of the resulting distribution of Q and computational cost.

[28] Given a set of active restrictions, Gafarov *et al.* (2018) provide an analytical expression for the value functions of the optimization problems associated with finding the bounds of the identified set. Their algorithm involves considering each possible combination of active restrictions, computing the associated value function and the (potential)

setting, whereas here we apply it at each draw from the posterior of ϕ. Step 6 can also be replaced by analytically checking whether the identified set is empty at each draw of ϕ. This is done by considering all possible combinations of $(n-1)$-number of active restrictions and checking whether any one of the vectors solving the active restrictions satisfies all the non-active sign restrictions.[29] The advantage of the analytical approach is that we can precisely assess emptiness of the identified set even when the identified set is very narrow, and it is computationally much faster. The advantage of the numerical optimization approach is that it is applicable even when restrictions are placed on multiple columns of Q, which is the case whenever the restrictions involve more than one structural shock.[30]

Fourth, if there are concerns about the convergence properties of the numerical optimization step due to, say, a large number of variables and/or constraints, but there are restrictions on multiple columns of Q (so the analytical approach cannot be applied), one could alternatively use the following algorithm.

Algorithm 2. *In Algorithm 1 replace (Step 3) with the following:*

(Step 3'). Iterate (Step 2.1) – (Step 2.3) K times and let $\left(Q_l : l = 1, \ldots, \tilde{K}\right)$ *be the draws that satisfy the sign restrictions. (If none of the draws satisfy the sign restrictions, draw a new* ϕ *and iterate (Step 2.1) – (Step 2.3) again). Let* $q_{j^*,k}$*,* $k = 1, \ldots, \tilde{K}$*, be the* j^**-th column vector of* Q_k*. Approximate* $[\ell(\phi), u(\phi)]$ *by* $[\min_k c'_{ih}(\phi)\, q_{j^*,k}, \max_k c'_{ih}(\phi)\, q_{j^*,k}]$.

A downside of this alternative is that the approximated identified set is smaller than $IS_\eta(\phi|F,S)$ at every draw of ϕ. Nonetheless, the alternative bounds still provide a consistent estimator of the identified set as the number of draws of Q goes to infinity. Comparing the bounds obtained using Algorithms 1 and 2 may also provide a useful check on the convergence properties of the numerical optimization in Step 3.

Fifth, the measure of the plausibility of the identifying restrictions computed in Step 6 is not meaningful if there are only zero restrictions, since in this case the identified set is never empty (see Proposition 3) and the data cannot detect violation of the restrictions irrespective of the choice of $\tilde{\pi}_\phi$. With sign restrictions the identified set can be empty for some ϕ, so if the chosen $\tilde{\pi}_\phi$ supports the entire Φ, the data can update the belief about the plausibility of the restrictions.

---

solution of the optimization problem and checking whether the solution is feasible (i.e. whether it satisfies the non-active sign restrictions). The bounds of the identified set are then obtained by computing the smallest and largest values corresponding to feasible solutions across the different combinations of active restrictions.

[29]In practice, we compute the unit-length vectors in the null space of the matrix containing the $(n-1)$-number of active restrictions using the 'null' function in Matlab. Since the null space has dimension one, there are only two unit-length vectors, which differ only in their signs. We check whether either one of the vectors satisfy the non-active restrictions. If we can pass this check for at least one combination of $(n-1)$-number of active restrictions, we conclude that the identified set is nonempty. Our Matlab code has an option to implement this procedure. See also Giacomini et al. (2018).

[30]See footnote 1 in Gafarov, et al. (2018) for references imposing restrictions on multiple structural shocks.

6 Empirical Application

We illustrate how our method can be used to: (1) perform robust Bayesian inference in SVARs without specifying a prior for the rotation matrix Q; (2) obtain a consistent estimator of the impulse-response identified set; and 3) if a prior for Q is available, disentangle the information introduced by this choice of prior from that solely contained in the identifying restrictions.

The model is the four-variable SVAR used in Granziera et al. (2018), which in turn is based on Aruoba and Schorfheide (2011). The vector of observables is the federal funds rate (i_t), real GDP per capita as a deviation from a linear trend (y_t), inflation as measured by the GDP deflator (π_t), and real money balances (m_t).[31] The data are quarterly from 1965:1 to 2006:1. The model is:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} i_t \\ y_t \\ \pi_t \\ m_t \end{pmatrix} = a + \sum_{j=1}^{2} A_j \begin{pmatrix} i_{t-j} \\ y_{t-j} \\ \pi_{t-j} \\ m_{t-j} \end{pmatrix} + \begin{pmatrix} \epsilon_{i,t} \\ \epsilon_{y,t} \\ \epsilon_{\pi,t} \\ \epsilon_{m,t} \end{pmatrix},$$

and the impulse response of interest is the output response to a monetary policy shock, $\frac{\partial y_{t+h}}{\partial \epsilon_{i,t}}$, so $j^* = 1$. The sign normalization restrictions (non-negative diagonal elements of the matrix on the left-hand side) and the assumption that the covariance matrix of the structural shocks is the identity matrix imply that the output response is with respect to a unit standard deviation positive (contractionary) monetary policy shock.

We consider different combinations of the following zero and sign restrictions:

(i) $a_{12} = 0$: the monetary authority does not respond contemporaneously to output.

(ii) $IR^0(y, i) = 0$: the instantaneous impulse response of output to a monetary policy shock is zero.

(iii) $IR^\infty(y, i) = 0$: the long-run impulse response of output to a monetary policy shock is zero.

(iv) Sign restrictions: following a contractionary monetary policy shock, the responses of inflation and real money balances are nonpositive on impact and after one quarter ($\frac{\partial \pi_{t+h}}{\partial \epsilon_{i,t}} \leq 0$ and $\frac{\partial m_{t+h}}{\partial \epsilon_{i,t}} \leq 0$ for $h = 0, 1$), and the response of the interest rate is nonnegative on impact and after one quarter ($\frac{\partial i_{t+h}}{\partial \epsilon_{i,t}} \geq 0$ for $h = 0, 1$).

We start from a model that does not impose any identifying restrictions (Model 0). We then impose different combinations of the restrictions, summarized in Table 1, which all give rise to set identification. Restrictions (i)–(iii) are zero restrictions that constrain the first column of Q, so $f_1 = 1$ if only one restriction out of (i)–(iii) is imposed (Models II to IV), and $f_1 = 2$ if two

[31] The data are from Frank Schorfheide's website: https://web.sas.upenn.edu/schorf/. For details on the construction of the series, see Appendix D from Granziera et al. (2018) and Footnote 5 of Aruoba and Schorfheide (2011).

restrictions are imposed (Models V to VII). No zero restrictions are placed on the remaining columns of Q, so for all models $f_2 = f_3 = f_4 = 0$, and the order of the variables satisfies Definition 3.

All models impose the sign restrictions in (iv), which are those considered in Granziera, Moon and Schorfheide (2017). This implies that Model 1 coincides with their model.

The bottom row of Table 1 reports the posterior plausibility of the imposed restrictions (i.e. the posterior probability that the identified set is nonempty), computed both numerically and analytically.[32] This measure is close to one for all models.

The prior of the reduced-form parameters, $\tilde{\pi}_{\phi}$, is the improper Jeffreys' prior, with density function proportional to $|\Sigma|^{-\frac{4+1}{2}}$. This implies that the posterior for ϕ is normal-inverse-Wishart, from which it is easy to draw. The posterior for ϕ is nearly identical to the likelihood. In implementing Algorithm 1, we draw ϕ's until we obtain 1,000 realizations of the nonempty identified set. We check for convexity of the identified set at every draw of ϕ using Proposition 3. The reported results are based on Algorithm 1, considering five starting values as discussed in the remarks in Section 5.[33] Since the prior for ϕ is the same in all models and the posterior probabilities of a nonempty identified set are all close to one, the posterior bounds differ across models mainly due to the different identifying restrictions.

We compare our approach to standard Bayesian inference based on choosing a uniform prior for Q to assess how this choice of unrevisable prior affects posterior inference. We obtain draws from the single posterior for the impulse responses by iterating Step (2.1)–(2.3) of Algorithm 1, and retaining the draws of Q that satisfy the sign restrictions.[34]

Table 2 provides the posterior inference results for the output responses at $h = 1$ (3 months), $h = 10$ (2 years and 6 months), and $h = 20$ (5 years) in each model, for both the robust Bayesian and the standard Bayesian approach. The table also shows the posterior lower probability that the impulse response is negative, $\pi_{\eta|Y*}(\eta < 0)$, as well as the diagnostic tools from Section 2.4.

Figures 1 and 2 report the set of posterior means for the impulse responses (vertical bars) and the smallest robust credible region with credibility 90% (continuous line), for the robust Bayesian

[32] The numerical computation considers a maximum of 3,000 draws of Q at each draw of ϕ. The fact that the posterior plausibility is the same using both numerical and analytical approaches suggests that this number of draws is sufficient to accurately verify whether the identified set is nonempty.

[33] The results are visually indistinguishable when obtaining the bounds using the analytical approach discussed in the remarks in Section 5. Moreover, five initial values appear sufficient to achieve convergence of the numerical algorithm to the true optimum computed analytically in more than 99% of the draws of ϕ. As a robustness check, we implemented Algorithm 2 with $\bar{K} = 50,000$ for Model IV, and found that the widths of the set of posterior means differ from the widths of those reported in Figure 2 by 0.4% on average.

[34] In Models 0 and I, this is equivalent to Uhlig (2005), as it obtains draws from the uniform distribution (or Haar measure) over the space of orthonormal matrices satisfying the sign normalizations and sign restrictions (if any). In models with both zero and sign restrictions, this is comparable to Arias et al. (2018), aside from the small differences in the algorithms discussed in Section 5 and the fact that they use a normal-inverse-Wishart prior for the reduced-form parameter. Using the same normal-inverse-Wishart prior as Arias et al. (2018) gives visually indistinguishable results in our application.

approach; for the standard Bayesian approach, they report the posterior mean (dotted line) and the 90% highest posterior density region (dashed line).[35]

We can draw several conclusions. First, choosing a uniform (unrevisable) prior for the rotation matrix can have large effects on posterior inference: in Model I this prior choice is more informative than the identifying restrictions (cf. the measures of informativeness); in Model III this choice would lead to the conclusion that the output response is negative for some horizons, whereas the robust Bayesian lower probability of this event is very low (cf. Figure 1 and the lower probability in Table 2), implying that the conclusion that the output response is negative for some horizons is largely driven by the unrevisable prior.

Second, sign restrictions alone (Model I) have little identifying power and result in identified set estimates that are too wide to draw any informative inference about the sign of the impulse response. Adding a single zero restriction (Models II to IV) makes the identified set estimates substantially tighter, although the identifying power of the zero restrictions varies across horizons (cf. the measure of informativeness in Table 2 and Figure 1). Unsurprisingly, the restriction on the contemporaneous response (restriction (ii)) is more informative at short horizons and the long-run restriction (restriction (iii)) is more informative at long horizons. The zero restriction on A_0 (restriction (i)) is informative at both short- and long horizons.

Third, imposing additional zero restrictions (Models V to VII) makes the identifying restrictions much more informative than the choice of the prior (cf. the measures of informativeness in Table 2) and reduces the gap between the conclusions of standard- and robust Bayesian analysis. The robust Bayesian analysis further becomes informative for the sign of the output response (cf. the lower probabilities in Table 2 and Figure 2). Since in these models the identifying restrictions carry a lot of information and result in narrow identified sets, we can understand how a given inferential conclusion depends on individual (or small sets of) identifying restrictions. We find that the sign of the output response crucially depends on which pair of the three zero restrictions (i)-(iii) one imposes: in Models V and VII, the response is negative at short-to-medium horizons; in Model V, the response is positive at short horizons.

Fourth, by comparing the results for Model 1 in Figure 1 to Figure 5 in Granziera et al. (2018), one can see that the robust Bayesian output is very similar to the estimates of the identified sets and the frequentist confidence intervals for the same model that are reported in that paper. This is compatible with the consistency property shown in Theorem 3.

Finally, note that in Models V to VII the estimator of the identified set lies inside the standard Bayesian credible region. This may seem to contradict the result of Moon and Schorfheide (2012) that standard Bayesian credible regions (asymptotically) lie inside the true identified set. A possible reason is that whether the asymptotic result approximates finite-sample behavior de-

[35]These figures summarize the marginal distribution of the impulse response at each horizon, and do not capture the dependence of the responses across different horizons.

pends on the width of the identified set and how accurately it is estimated. When the width of $IS_\eta(\phi|F,S) = [\ell(\phi), u(\phi)]$ is small relative to the posterior variances of $(\ell(\phi), u(\phi))$, the standard Bayesian credible region can be as wide as the credible region for $\ell(\phi)$ or $u(\phi)$, because the posterior of the impulse response is similar to the posterior of $\ell(\phi)$ or $u(\phi)$. On the other hand, the set of posterior means can be tight even for large variances of $(\ell(\phi), u(\phi))$ as they are affected only by the means of the posteriors of $\ell(\phi)$ and $u(\phi)$. This implies that the standard Bayesian credible region can be wider than the estimator of the identified set. The relationship between the standard Bayesian credible region and the smallest robust credible region, on the other hand, stays stable across the models (see the prior informativeness in Table 2), with the former 20% to 40% shorter than the latter in all models.

7 Conclusion

We develop a robust Bayesian inference procedure for set-identified models, providing Bayesian inference that is asymptotically equivalent to frequentist inference about the identified set. The main idea is to remove the need to specify a prior that is not revised by the data, but allow for ambiguous beliefs (multiple priors) for the unrevisable component of the prior. We show how to compute an estimator of the identified set (the set of posterior means) and the smallest robust credible region that respectively satisfy the properties of consistency and correct coverage asymptotically.

We conclude by summarizing the recommended uses and advantages of our method. First, by reporting the robust Bayesian output, one can learn what inferential conclusions can be supported solely by the imposed identifying restrictions and the posterior for the reduced-form parameter. Even if a user has a credible prior for parameters for which the data are not informative, the robust Bayesian output will help communicate with other users who may have different priors. Second, by comparing the output across different sets of identifying restrictions, one can learn and report which identifying restrictions are crucial in drawing a given inferential conclusion. Third, the procedure can be a useful tool for separating the information contained in the data from any prior input that is not revised by the data.

The fact that in practical applications the set of posterior means and the robust credible region for a given set of identifying restrictions may be too wide to draw informative policy recommendations should not be considered a disadvantage of the method. Wide bounds may encourage the researcher to look for additional credible restrictions and/or to refine the set of priors, by inspecting how the data are collected, by considering empirical evidence from other studies, and by turning to economic theory. If additional restrictions are not available, our analysis informs the researcher about the amount of ambiguity that the policy decision will be subject to. As Manski (2013) argues, knowing what we do not know is an important premise for a policy decision without incredible certitude.

8 Tables and Figures

Table 1: Model Definition and Plausibility of Identifying Restrictions

Restrictions \Model	0	I	II	III	IV	V	VI	VII
(i) $a_{12} = 0$	-	-	x	-	-	x	x	-
(ii) $IR^0(y, i) = 0$	-	-	-	x	-	x	-	x
(iii) $IR^\infty(y, i) = 0$	-	-	-	-	x	-	x	x
(iv) Sign restrictions	-	x	x	x	x	x	x	x
$\tilde{\pi}_{\phi\|Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$ (numerical)	1.0000	1.0000	1.0000	1.0000	0.9950	0.9042	0.9421	0.9728
$\tilde{\pi}_{\phi\|Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$ (analytical)	1.0000	1.0000	1.0000	1.0000	0.9950	0.9042	0.9421	0.9728

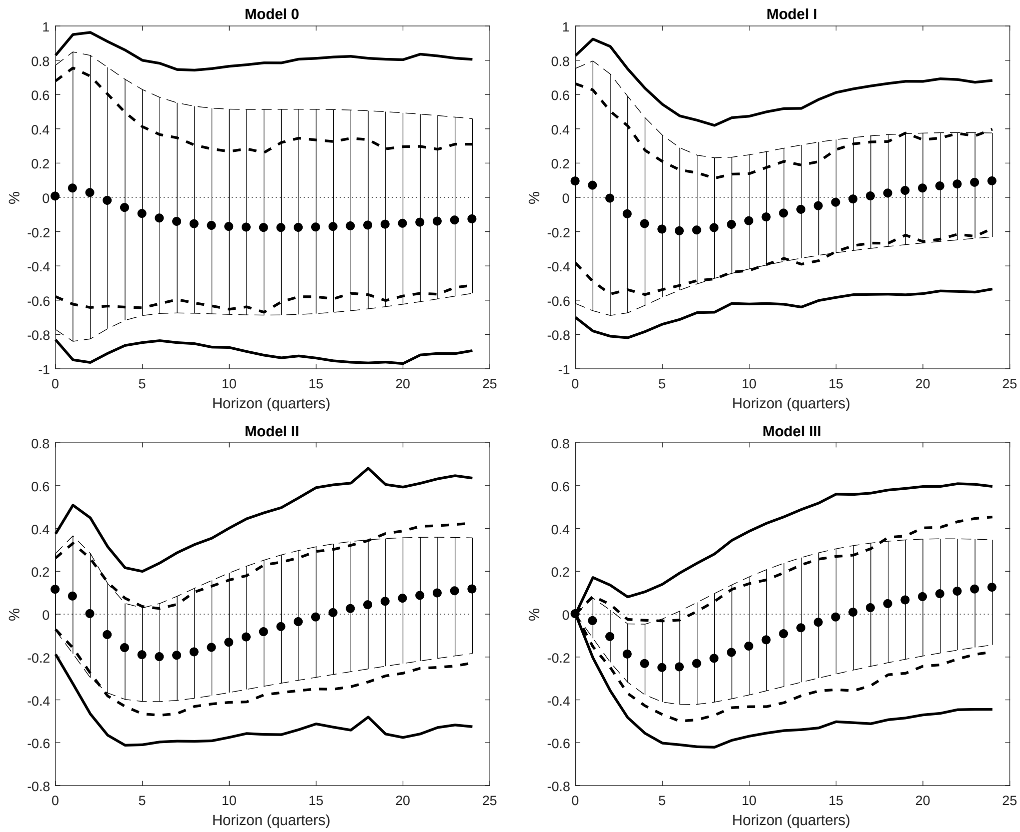
Notes: 'x' indicates the restriction is imposed; $\tilde{\pi}_{\phi|Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$ is the measure of the plausibility of the identifying restrictions described in Section 2.6.1, where 'numerical' results are obtained using Step 6 in Algorithm 1 (with a maximum of 3,000 draws of Q at each draw of ϕ) and 'analytical' results are obtained using the analytical approach described in the remarks after Algorithm 1.

Table 2: Output responses at $h = 1, 10$, and 20: Standard Bayes (SB) vs. Robust Bayes (RB)

	Model 0			Model I		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
SB: posterior mean	0.05	-0.17	-0.15	0.07	-0.14	0.05
SB: 90% credible region	[-0.62,0.76]	[-0.65,0.27]	[-0.58,0.30]	[-0.49,0.63]	[-0.43,0.14]	[-0.26,0.34]
RB: set of posterior means	[-0.84,0.85]	[-0.68,0.51]	[-0.62,0.49]	[-0.66,0.80]	[-0.42,0.25]	[-0.27,0.38]
RB: 90% robust credible region	[-0.95,0.95]	[-0.88,0.76]	[-0.97,0.80]	[-0.78,0.92]	[-0.62,0.47]	[-0.56,0.68]
Lower probability: $\pi_{\eta\|Y*}(\eta<0)^{\dagger}$	0.000	0.000	0.000	0.000	0.013	0.001
Informativeness of restrictions*	–	–	–	0.14	0.44	0.43
Informativeness of prior**	0.27	0.44	0.51	0.34	0.48	0.52
	Model II			**Model III**		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
SB: posterior mean	0.08	-0.13	0.07	-0.03	-0.15	0.08
SB: 90% credible region	[-0.16,0.33]	[-0.41,0.16]	[-0.28,0.39]	[-0.15,0.08]	[-0.43,0.14]	[-0.24,0.40]
RB: set of posterior means	[-0.18,0.37]	[-0.37,0.19]	[-0.23,0.36]	[-0.11,0.07]	[-0.38,0.17]	[-0.20,0.35]
RB: 90% robust credible region	[-0.33,0.51]	[-0.58,0.40]	[-0.58,0.59]	[-0.20,0.17]	[-0.57,0.39]	[-0.47,0.60]
Lower probability: $\pi_{\eta\|Y*}(\eta<0)^{\dagger}$	0.000	0.055	0.003	0.130	0.065	0.002
Informativeness of restrictions*	0.67	0.53	0.47	0.89	0.54	0.51
Informativeness of prior**	0.41	0.42	0.43	0.37	0.4	0.39
	Model IV			**Model V**		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
SB: posterior mean	0.03	-0.18	0.02	-0.08	-0.15	0.1
SB: 90% credible region	[-0.51,0.63]	[-0.43,0.01]	[-0.22,0.22]	[-0.18,0.01]	[-0.46,0.10]	[-0.19,0.42]
RB: set of posterior means	[-0.57,0.67]	[-0.32,0.03]	[-0.17,0.19]	[-0.09,-0.06]	[-0.30,0.01]	[-0.03,0.24]
RB: 90% robust credible region	[-0.73,0.87]	[-0.53,0.24]	[-0.35,0.40]	[-0.19,0.02]	[-0.54,0.27]	[-0.27,0.60]
Lower probability: $\pi_{\eta\|Y*}(\eta<0)^{\dagger}$	0.016	0.448	0.041	0.866	0.442	0.076
Informativeness of restrictions*	0.27	0.71	0.69	0.98	0.74	0.76
Informativeness of prior**	0.29	0.43	0.41	0.11	0.32	0.3
	Model VI			**Model VII**		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
SB: posterior mean	0.14	-0.18	0	-0.02	-0.2	0.03
SB: 90% credible region	[-0.10,0.40]	[-0.41,0.03]	[-0.20,0.17]	[-0.13,0.10]	[-0.40,-0.01]	[-0.15,0.19]
RB: set of posterior means	[0.02,0.25]	[-0.28,-0.07]	[-0.07,0.07]	[-0.06,0.02]	[-0.29,-0.08]	[-0.04,0.10]
RB: 90% robust credible region	[-0.24,0.42]	[-0.47,0.18]	[-0.27,0.29]	[-0.15,0.14]	[-0.50,0.13]	[-0.25,0.28]
Lower probability: $\pi_{\eta\|Y*}(\eta<0)^{\dagger}$	0.024	0.734	0.287	0.405	0.749	0.145
Informativeness of restrictions*	0.86	0.83	0.88	0.96	0.82	0.88
Informativeness of prior**	0.25	0.33	0.32	0.21	0.38	0.36

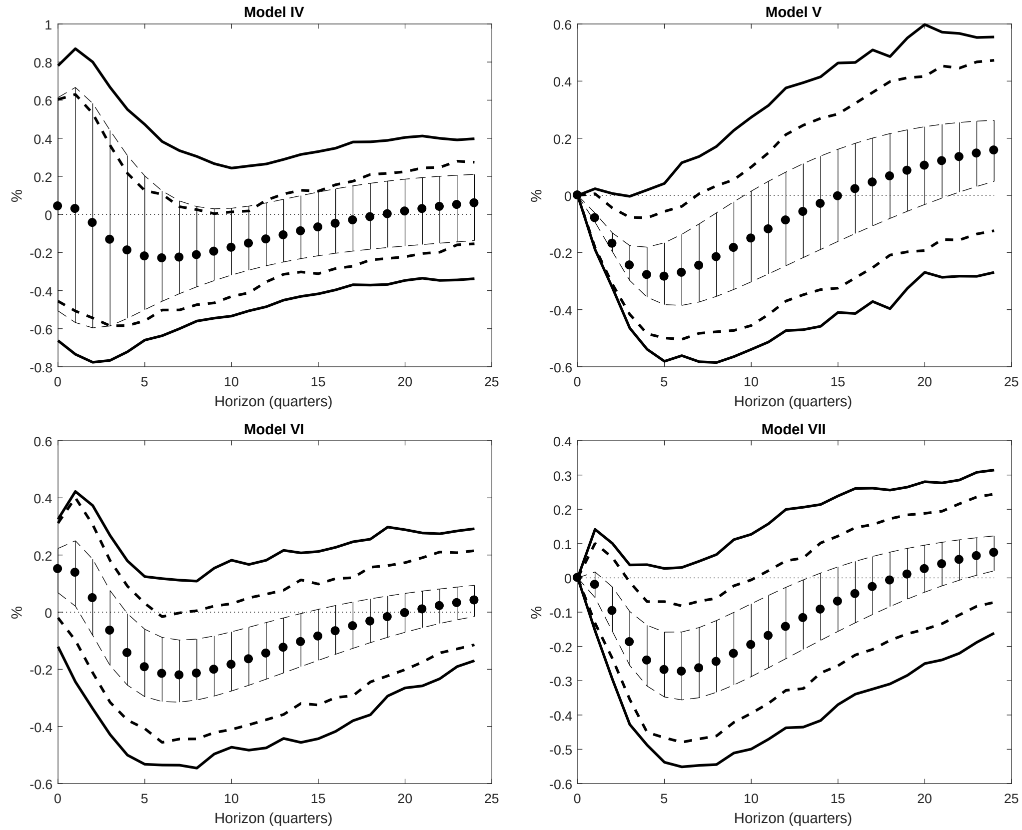
Notes: Robust credible regions reported are smallest ones defined in (2.7). * see eq. (2.8) for definition. The model informativeness is measured relative to Model 0. ** see eq. (2.9) for definition. † the posterior lower probability (see Theorem 1) is computed as the proportion of draws where the upper bound of the identified set estimator is less than zero (conditional on the set being nonempty).

Figure 1: Plots of Output Impulse Responses to a Monetary Policy Shock (Models 0–III)



Notes: See Table 1 for the definition of models. In each figure, the points are the standard Bayesian posterior means, the vertical bars are the set of posterior means, the dashed curves are the upper and lower bounds of the standard Bayesian highest posterior density regions with credibility 90%, and the solid curves are the upper and lower bounds of the robust credible regions with credibility 90%

Figure 2: Plots of Output Impulse Responses to a Monetary Policy Shock (Models IV–VII)



Notes: See Figure 1.

Appendix

A Lemmas and Proofs

Lemma A.1 *Assume* $(\Theta, \mathcal{A})$ *and* $(\Phi, \mathcal{B})$ *are measurable spaces in which* Θ *and* Φ *are complete separable metric spaces. Under Assumption 1,* $IS_\theta(\phi)$ *and* $IS_\eta(\phi)$ *are random closed sets induced by a probability measure on* $(\Phi, \mathcal{B})$*, i.e.,* $IS_\theta(\phi)$ *and* $IS_\eta(\phi)$ *are closed and, for* $A \in \mathcal{A}$ *and* $D \in \mathcal{H}$*,*

$$\begin{aligned} \{\phi : IS_\theta(\phi) \cap A \neq \emptyset\} &\in \mathcal{B} \quad \text{for} \quad A \in \mathcal{A}, \\ \{\phi : IS_\eta(\phi) \cap D \neq \emptyset\} &\in \mathcal{B} \quad \text{for} \quad D \in \mathcal{H}. \end{aligned}$$

Proof. Closedness of $IS_\theta(\phi)$ and $IS_\eta(\phi)$ is implied directly by Assumption 1 (ii) and 1 (iii). To prove the measurability of $\{\phi : IS_\theta(\phi) \cap A \neq \emptyset\}$, Theorem 2.6 in Chapter 1 of Molchanov (2005) is invoked, which states that, given Θ as Polish, $\{\phi : IS_\theta(\phi) \cap A \neq \emptyset\} \in \mathcal{B}$ holds if and only if $\{\phi : \theta \in IS_\theta(\phi)\} \in \mathcal{B}$ is true for every $\theta \in \Theta$. Since $IS_\theta(\phi)$ is an inverse image of the many-to-one mapping, $g : \Theta \to \Phi$, $\{\phi : \theta \in IS_\theta(\phi)\}$ is a singleton for each $\theta \in \Theta$. Any singleton set of ϕ belongs to $\mathcal{B}$, since Φ is a metric space. Hence, $\{\phi : \theta \in IS_\theta(\phi)\} \in \mathcal{B}$ holds.

To verify the measurability of $\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\}$, note that

$$\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\} = \left\{\phi : IS_\theta(\phi) \cap h^{-1}(D) \neq \emptyset\right\}.$$

Since $h^{-1}(D) \in \mathcal{A}$, by the measurability of h (Assumption 1 (iii)), the first statement of this lemma implies $\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\} \in \mathcal{B}$. ■

Lemma A.2 *Under Assumption 1, let $A \in \mathcal{A}$ be an arbitrary fixed subset of Θ. For every $\pi_{\theta|\phi} \in \Pi_{\theta|\phi}$, $1_{\{IS_\theta(\phi) \subset A\}}(\phi) \leq \pi_{\theta|\phi}(A|\phi)$ holds π_ϕ-almost surely.*

Proof. For the given subset A, define $\Phi_1^A = \{\phi : IS_\theta(\phi) \subset A, IS_\theta(\phi) \neq \emptyset\} = \{\phi : IS_\theta(\phi) \cap A^c \neq \emptyset\}^c$. Note that, by Lemma A.1, Φ_1^A belongs to $\mathcal{B}$. To prove the claim, it suffices to show that

$$\int_B 1_{\Phi_1^A}(\phi) d\pi_\phi \leq \int_B \pi_{\theta|\phi}(A) d\pi_\phi$$

for every $\pi_{\theta|\phi} \in \Pi_{\theta|\phi}$ and $B \in \mathcal{B}$. Consider

$$\int_B \pi_{\theta|\phi}(A) d\pi_\phi \geq \int_{B \cap \Phi_1^A} \pi_{\theta|\phi}(A) d\pi_\phi = \pi_\theta(A \cap IS_\theta(B \cap \Phi_1^A)),$$

where the equality follows by the definition of the conditional probability. By the construction of Φ_1^A, $IS_\theta(B \cap \Phi_1^A) \subset A$ holds, so

$$\begin{aligned}
\pi_\theta(A \cap IS_\theta(B \cap \Phi_1^A)) &= \pi_\theta(IS_\theta(B \cap \Phi_1^A)) \\
&= \pi_\phi(B \cap \Phi_1^A) \\
&= \int_B 1_{\Phi_1^A}(\phi) d\pi_\phi.
\end{aligned}$$

Thus, the inequality is proven. ■

Lemma A.3 *Under Assumption 1, for each $A \in \mathcal{A}$, there exists $\pi_{\theta|\phi*}^A \in \Pi_{\theta|\phi}$ that achieves the lower bound of $\pi_{\theta|\phi}(A)$ obtained in Lemma A.2, π_ϕ-almost surely.*

Proof. Fix $A \in \mathcal{A}$ and consider subsets of Φ based on the relationship between $IS_\theta(\phi)$ and A,

$$\begin{aligned}
\Phi_0^A &= \{\phi : IS_\theta(\phi) \cap A = \emptyset, IS_\theta(\phi) \neq \emptyset\}, \\
\Phi_1^A &= \{\phi : IS_\theta(\phi) \subset A, IS_\theta(\phi) \neq \emptyset\}, \\
\Phi_2^A &= \{\phi : IS_\theta(\phi) \cap A \neq \emptyset \text{ and } IS_\theta(\phi) \cap A^c \neq \emptyset\},
\end{aligned}$$

where each of Φ_0^A, Φ_1^A, and Φ_2^A belongs to $\mathcal{B}$ by Lemma A.1. Note that Φ_0^A, Φ_1^A, and Φ_2^A are mutually disjoint and constitute a partition of $g(\Theta) \subset \Phi$.

Now, consider a Θ-valued measurable selection $\xi^{A}(\cdot)$ defined on Φ_2^A such that $\xi^{A}(\phi) \in [IS_\theta(\phi) \cap A^c]$ holds for π_ϕ-almost every $\phi \in \Phi_2^A$. Note that such measurable selection $\xi^A(\phi)$ can be constructed, for instance, by $\xi^A(\phi) = \arg\max_{\theta \in IS_\theta(\phi) \cap A^\epsilon} d(\theta, A)$, where $d(\theta, A) = \inf_{\theta' \in A} \|\theta - \theta'\|$ and $A^\epsilon = \{\theta : d(\theta, A) \leq \epsilon\}$ is the closed ϵ-enlargement of A (see Theorem 2.27 in Chapter 1 of Molchanov (2005) for $\mathcal{B}$-measurability of such $\xi^A(\phi)$). Let us pick an arbitrary conditional probability distribution from the class $\pi_{\theta|\phi} \in \Pi_{\theta|\phi}$ and construct another conditional probability distribution $\pi^A_{\theta|\phi*}$ by, for $\tilde{A} \in \mathcal{A}$,

$$\pi^A_{\theta|\phi*}\left(\tilde{A}\right) = \begin{cases} \pi_{\theta|\phi}\left(\tilde{A}\right) & \text{for } \phi \in \Phi_0^A \cup \Phi_1^A, \\ 1_{\left\{\xi^A(\phi) \in \tilde{A}\right\}}(\phi) & \text{for } \phi \in \Phi_2^A. \end{cases}$$

It can be checked that $\pi^A_{\theta|\phi*}(\cdot)$ is a probability measure on $(\Theta, \mathcal{A})$: $\pi_{\theta*}$ satisfies $\pi_{\theta|\phi*}(\emptyset) = 0$, $\pi_{\theta|\phi*}(\Theta) = 1$, and is countably additive. Furthermore, $\pi^A_{\theta|\phi*}$ belongs to $\Pi_{\theta|\phi}$ because $\pi^A_{\theta|\phi*}(IS_\theta(\phi)) = 1$ holds, π_ϕ a.s., by the construction of $\xi^A(\phi)$. With the thus-constructed $\pi^A_{\theta|\phi*}$ and an arbitrary subset $B \in \mathcal{B}$, consider

$$\begin{aligned}
\int_B \pi^A_{\theta|\phi*}(A) d\pi_\phi &= \int_B \pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) d\pi_\phi \\
&= \int_{B \cap \Phi_0^A} \pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) d\pi_\phi \\
&\quad + \int_{B \cap \Phi_1^A} \pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) d\pi_\phi + \int_{B \cap \Phi_2^A} \pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) d\pi_\phi \\
&= 0 + \int_{B \cap \Phi_1^A} \pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) d\pi_\phi + 0 \\
&= \int_B 1_{\Phi_1^A}(\phi) d\pi_\phi,
\end{aligned}$$

where the first equality follows by $\pi^A_{\theta|\phi*} \in \Pi_{\theta|\phi}$, the third equality follows since $A \cap IS_\theta(\phi) = \emptyset$ for $\phi \in \Phi_0^A$ and $\pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) = 1_{\left\{\xi^A(\phi) \in A\right\}}(\phi) = 0$ for $\phi \in \Phi_2^A$, and the fourth equality follows since $\pi^A_{\theta|\phi*}(A \cap IS_\theta(\phi)) = 1$ for $\phi \in \Phi_1^A$. Since $B \in \mathcal{B}$ is arbitrary, this implies that $\pi^A_{\theta|\phi*}(A) = 1_{\Phi_1^A}(\phi)$, π_ϕ-almost surely, implying that $\pi^A_{\theta|\phi*}$ achieves the lower bound shown in Lemma A.2. ■

Proof of Theorem 1. We first show the special case of $\eta = \theta$. The posterior of θ is given by (see equation (2.2))

$$\pi_{\theta|Y}(A) = \int_\Phi \pi_{\theta|\phi}(A) d\pi_{\phi|Y}(\phi).$$

By the monotonicity of the integral, $\pi_{\theta|Y}(A)$ is minimized over the prior class by plugging in the attainable pointwise lower bound of $\pi_{\theta|\phi}(A)$ into the integrand. By Lemmas A.2 and A.3, the attainable pointwise lower bound of $\pi_{\theta|\phi}(A)$ is given by $1_{\{IS_\theta(\phi) \subset A\}}(\phi)$. Hence,

$$\pi_{\theta|Y*}(A) = \int_\Phi 1_{\{IS_\theta(\phi) \subset A\}}(\phi) d\pi_{\phi|Y}(\phi) = \pi_{\phi|Y}(\{\phi : IS_\theta(\phi) \subset A\}).$$

The posterior upper probability follows by its conjugacy property with the lower probability, $\pi^*_{\theta|Y}(A) = 1 - \pi_{\theta|Y*}(A^c)$.

By repeating the argument in the proof of Lemma A.3, it can be shown that this upper probability can be attained by setting the conditional prior as, for $\tilde{A} \in \mathcal{A}$,

$$\pi^{A*}_{\theta|\phi}\left(\tilde{A}\right) = \begin{cases} \pi_{\theta|\phi}\left(\tilde{A}\right) & \text{for } \phi \in \Phi^A_0 \cup \Phi^A_1, \\ 1_{\left\{\xi^{A^c}(\phi) \in \tilde{A}\right\}}(\phi) & \text{for } \phi \in \Phi^A_2, \end{cases}$$

where $\xi^{A^c}(\cdot)$ is a Θ-valued measurable selection defined for $\phi \in \Phi^A_2$ such that $\xi^{A^c}(\phi) \in [IS_\theta(\phi) \cap A]$ holds for π_ϕ-almost every $\phi \in \Phi^A_2$. Consider mixing these extreme conditional priors, $\pi^\lambda_{\theta|\phi} \equiv \lambda\pi^A_{\theta|\phi*} + (1-\lambda)\pi^{A*}_{\theta|\phi}$, $\lambda \in [0,1]$. Note that $\pi^\lambda_{\theta|\phi}$ belongs to $\Pi_{\theta|\phi}$ for any $\lambda \in [0,1]$ since $\pi^\lambda_{\theta|\phi}(IS_\theta(\phi)) = 1$. The posterior probability for $\{\theta \in A\}$ with conditional prior $\pi^\lambda_{\theta|\phi}$ is the λ-convex combination of the posterior lower and upper probabilities, $\lambda\pi_{\theta|Y*}(A) + (1-\lambda)\pi^*_{\theta|Y}(A)$. Since $\lambda \in [0,1]$ is arbitrary, the set of the posterior probabilities for $\{\theta \in A\}$ is the connected interval $\left[\pi_{\theta|Y*}(A), \pi^*_{\theta|Y}(A)\right]$.

For the general case $\eta = h(\theta)$, the expression of the posterior lower probability follows from

$$\begin{aligned} \pi_{\eta|Y*}(D) &= \pi_{\theta|Y*}(h^{-1}(D)) \\ &= \pi_{\phi|Y}(\{\phi : IS_\theta(\phi) \subset h^{-1}(D)\}) \\ &= \pi_{\phi|Y}(\{\phi : IS_\eta(\phi) \subset D\}). \end{aligned} \tag{A.1}$$

The expression of the posterior upper probability follows again by its conjugacy with the lower probability. The convexity of the set of $\pi_{\eta|Y}(D)$ follows by setting $A = h^{-1}(D)$ in the connected set for $\pi_{\theta|Y}(A) \in \left[\pi_{\theta|Y*}(A), \pi^*_{\theta|Y}(A)\right]$ established above. ■

Proof of Theorem 2. At each ϕ in the support of $\pi_{\phi|Y}$, note that the set $\left\{E_{\eta|\phi}(\eta) : \pi_{\eta|\phi}(IS_\eta(\phi)) = 1\right\}$ agrees with $co(IS_\eta(\phi))$. Hence, $\left(E_{\eta|\phi}(\eta) : \phi \in g(\Theta)\right)$ pinned down by selecting $\pi_{\theta|\phi}$ from $\Pi_{\theta|\phi}$ can be viewed as a selection from $co(IS_\eta)$. Since the prior class $\Pi_{\theta|\phi}$ does not constrain choices of $\pi_{\theta|\phi}$ over different ϕ's, priors in $\Pi_{\theta|\phi}$ can exhaust any selection of $co(IS_\eta)$. Having assumed that $co(IS_\eta(\phi))$ is a $\pi_{\phi|Y}$-integrable random closed set, the set $\left\{E_{\eta|Y}(\eta) = E_{\phi|Y}\left[E_{\eta|\phi}(\eta)\right] : \pi_{\theta|\phi} \in \Pi_{\theta|\phi}\right\}$ agrees with $E^A_{\phi|Y}[co(IS_\eta)]$ by the definition of the Aumann integral. Its convexity follows by the assumption that $IS_\eta(\phi)$ is closed and integrable and Theorem 1.26 of Molchanov (2005). ■

Proof of Proposition 1. The event $\{IS_\eta(\phi) \subset C_r(\eta_c)\}$ happens if and only if $\left\{\bar{d}(\eta_c, IS_\eta(\phi)) \leq r\right\}$. So, $r_\alpha(\eta_c) \equiv \inf\left\{r : \pi_{\phi|Y}\left(\left\{\phi : \bar{d}(\eta_c, IS_\eta(\phi)) \leq r\right\}\right) \geq \alpha\right\}$ is the radius of the smallest interval centered at η_c that contains random sets $IS_\eta(\phi)$ with a posterior probability of at least α. Therefore, finding a minimizer of $r_\alpha(\eta_c)$ in η_c is equivalent to searching for the center of the smallest interval that contains $IS_\eta(\phi)$ with posterior probability α. The attained minimum of $r_\alpha(\eta_c)$ is its radius. ■

Proof of Theorem 3. (i) Let $\epsilon > 0$ be arbitrary. Since Assumption 2 (i) implies that $IS_\eta(\cdot)$ is compact-valued in an open neighborhood of ϕ_0, continuity of the identified set correspondence at ϕ_0 is equivalent to continuity of $IS_\eta(\cdot)$ at ϕ_0 in terms of the Hausdorff metric (see, e.g., Proposition 5 in Chapter E of Ok (2007)). This implies that there exists an open neighborhood G of ϕ_0 such that $d_H(IS(\phi), IS(\phi_0)) < \epsilon$ holds for all $\phi \in G$. Consider

$$\begin{aligned}
\pi_{\phi|Y^T}\{\phi : d_H(IS(\phi), IS(\phi_0)) > \epsilon\} &= \pi_{\phi|Y^T}(\{\phi : d_H(IS(\phi), IS(\phi_0)) > \epsilon\} \cap G) \\
&\quad + \pi_{\phi|Y^T}(\{\phi : d_H(IS(\phi), IS(\phi_0)) > \epsilon\} \cap G^c) \\
&\le \pi_{\phi|Y^T}(G^c),
\end{aligned}$$

where the last line follows because $\{\phi : d_H(IS_\eta(\phi), IS_\eta(\phi_0)) > \epsilon\} \cap G = \emptyset$ by the construction of G. The posterior consistency of ϕ yields $\lim_{T\to\infty} \pi_{\phi|Y^T}(G^c) = 0$, $p(Y^\infty|\phi_0)$-a.s.

(ii) Let $s(co(IS_\eta), u) = \sup_{\eta \in co(IS_\eta(\phi))} \eta' u$, $u \in \mathcal{S}^{k-1}$, be the support function of closed and convex set $co(IS_\eta)$, where $\mathcal{S}^{k-1}$ is the unit sphere in $\mathcal{R}^k$. Let $\epsilon > 0$ be arbitrary and let G be an open neighborhood of ϕ_0 such that $d_H(IS_\eta(\phi), IS_\eta(\phi_0)) < \epsilon$ holds for all $\phi \in G$. Under Assumption 2 (iii), $E^A_{\phi|Y^T}[co(IS_\eta(\phi))]$ is bounded, so using the support function, the Hausdorff distance between $E^A_{\phi|Y^T}[co(IS_\eta(\phi))]$ and $co(IS_\eta(\phi_0))$ can be bounded above by

$$\begin{aligned}
& d_H\left(E^A_{\phi|Y^T}[co(IS_\eta(\phi))], co(IS_\eta(\phi_0))\right) \\
&= \sup_{u\in\mathcal{S}^{k-1}} \left| s\left(E^A_{\phi|Y^T}[co(IS_\eta(\phi))], u\right) - s(co(IS_\eta(\phi_0)), u)\right| \\
&= \sup_{u\in\mathcal{S}^{k-1}} \left| E_{\phi|Y^T}[s(co(IS_\eta(\phi)), u) - s(co(IS_\eta(\phi_0)), u)]\right| \\
&\le \sup_{u\in\mathcal{S}^{k-1}} \left| E_{\phi|Y^T}[\{s(co(IS_\eta(\phi)), u) - s(co(IS_\eta(\phi_0)), u)\} \cdot 1_G(\phi)]\right| \\
&\quad + \sup_{u\in\mathcal{S}^{k-1}} \left| E_{\phi|Y^T}[\{s(co(IS_\eta(\phi)), u) - s(co(IS_\eta(\phi_0)), u)\} \cdot 1_{G^c}(\phi)]\right| \\
&\le E_{\phi|Y^T}[d_H(co(IS_\eta(\phi)), co(IS_\eta(\phi_0))) \cdot 1_G(\phi)] \\
&\quad + \sup_{u\in\mathcal{S}^{k-1}} \sqrt{E_{\phi|Y^T}[\{s(co(IS_\eta(\phi)), u) - s(co(IS_\eta(\phi_0)), u)\}^2]\pi_{\phi|Y^T}(G^c)} \\
&\le \epsilon + \sqrt{2E_{\phi|Y^T}\left[\sup_{\eta\in IS_\eta(\phi)} \|\eta\|^2\right] + 2\sup_{\eta\in IS_\eta(\phi_0)} \|\eta\|^2} \cdot \sqrt{\pi_{\phi|Y^T}(G^c)},
\end{aligned}$$

where the first line uses the identity $d_H(D, D') = \sup_{u\in\mathcal{S}^{k-1}} |s(D,u) - s(D',u)|$ that holds for any convex and compact sets D, $D' \subset \mathcal{R}^k$, the second line uses the identity $s\left(E^A_{\phi|Y^T}[co(IS_\eta(\phi))], u\right) = E_{\phi|Y^T}[s(co(IS_\eta(\phi)), u)]$ (see, e.g., Theorem 1.26 in Chap. 2 of Molchanov (2005)), the fourth line applies the Cauchy-Schwartz inequality to the term involving $1_{G^c}(\phi)$, and the final line follows since $d_H(co(IS_\eta(\phi)), co(IS_\eta(\phi_0))) < \epsilon$ on G and $\sup_{u\in\mathcal{S}^{k-1}} s(co(IS_\eta(\phi)), u)^2 = \sup_{\eta\in IS_\eta(\phi)} \|\eta\|^2$. By Assumptions 2 (i), (iii), and posterior consistency of $\pi_{\phi|Y^T}$, we have $\sup_{\eta\in IS_\eta(\phi_0)} \|\eta\|^2 < \infty$,

$\limsup_{T\to\infty} E_{\phi|Y^T}\left[\sup_{\eta\in IS_\eta(\phi)} \|\eta\|^2\right] < \infty$, and $\lim_{T\to\infty} \pi_{\phi|Y^T}(G^c) = 0$, $p(Y^\infty|\phi_0)$-a.s. Hence, the second term in (A.2) converges to zero $p(Y^\infty|\phi_0)$-a.s. Since ϵ is arbitrary, the claim of (ii) follows. ■

Proof of Theorem 4. Since C_α is convex by Assumption 3 (ii), $IS_\eta(\phi) \subset C_\alpha$ holds if and only if $s(IS_\eta(\phi), q) \leq s(C_\alpha, q)$ for all $q \in \mathcal{S}^{k-1}$. Therefore, we have

$$\begin{aligned}\pi_{\phi|Y^T}(IS_\eta(\phi) \subset C_\alpha) &= \pi_{\phi|Y^T}(s(IS_\eta(\phi), \cdot) \leq s(C_\alpha, \cdot)) \\ &= \pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \leq \hat{c}_T(\cdot)\right),\end{aligned}$$

and robust credible region C_α with α posterior probability of covering $IS_\eta(\phi)$ satisfies

$$\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \leq \hat{c}_T(\cdot)\right) \geq \alpha,$$

for all Y^T and $T = 1, 2, \ldots$. Similarly, the frequentist coverage probability of C_α for $IS_\eta(\phi_0)$ can be expressed as

$$P_{Y^T|\phi_0}(IS_\eta(\phi_0) \subset C_\alpha) = P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \leq \hat{c}_T(\cdot)\right).$$

Let P_X be the probability law of limiting stochastic process $X(\cdot)$ introduced in Assumption 4 (i) and (ii). In what follows, our aim is to prove the following convergence claims: under Assumption 4,

$$\begin{aligned}&\text{(A)} \quad \left|\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \leq \hat{c}_T(\cdot)\right) - P_X(X(\cdot) \leq \hat{c}_T(\cdot))\right| \to 0, \text{ as } T \to \infty,\ p_{Y^\infty|\phi_0}\text{-a.s., and} \\ &\text{(B)} \quad \left|P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \leq \hat{c}_T(\cdot)\right) - P_X(X(\cdot) \leq \hat{c}_T(\cdot))\right| \to 0 \text{ in } p_{Y^\infty|\phi_0}\text{-probability as } T \to \infty.\end{aligned}$$

Since $\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \leq \hat{c}(\cdot)\right) \geq \alpha$, convergence (A) implies $\liminf_{T\to\infty} P_X(X(\cdot) \leq \hat{c}_T(\cdot)) \geq \alpha$, $p_{Y^\infty|\phi_0}$-a.s. Then, convergence (B) in turn implies our desired conclusion,

$$\liminf_{T\to\infty} P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \leq \hat{c}_T(\cdot)\right) \geq \alpha,$$

as otherwise it contradicts $\liminf_{T\to\infty} P_X(X(\cdot) \leq \hat{c}_T(\cdot)) \geq \alpha$, $p_{Y^\infty|\phi_0}$-a.s. In the case that $\hat{c}_T$ is chosen to satisfy $\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \leq \hat{c}(\cdot)\right) = \alpha$, the convergences (A) and (B) imply

$$\lim_{T\to\infty} P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \leq \hat{c}_T(\cdot)\right) = \alpha.$$

To show (A), we note that any weakly converging sequence of stochastic processes in $\mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ is tight (see, e.g., Lemma 16.2 and Theorem 16.3 in Kallenberg (2001)). Hence, Assumption 4 (i) implies that for almost every sampling sequence of Y^T, there exists a class of bounded functions

$\mathcal{F} \subset \mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ such that $\mathcal{F}$ contains $\{\hat{c}_T(\cdot)\}$ for all large T. Furthermore, we can set $\mathcal{F}$ to be constrained to equicontinuous functions because the support functions for bounded sets are Lipshitz continuous.

To prove convergence (A), it suffices to show

$$\sup_{c \in \mathcal{F}} \left| P_{X_T}\left(X_T(\cdot) \leq c(\cdot)\right) - P_X\left(X(\cdot) \leq c(\cdot)\right)\right| \to 0, \text{ as } T \to \infty \tag{A.2}$$

for any weakly converging stochastic processes, $X_T \rightsquigarrow X$. Suppose this claim does not hold. Then, there exists a subsequence T' of T, a sequence of functions $\{c_{T'}(\cdot) \in \mathcal{F}\}$, and $\epsilon > 0$ such that

$$\left| P_{X_{T'}}\left(X_{T'}(\cdot) \leq c_{T'}(\cdot)\right) - P_X\left(X(\cdot) \leq c_{T'}(\cdot)\right)\right| > \epsilon \tag{A.3}$$

holds for all T'. By Assumption 4 (iv) and the Arzelà-Ascoli theorem,[36], $\mathcal{F}$ is relatively compact. Hence, there exists a subsequence T'' of T' such that $c_{T''}$ converges to $c^* \in \mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ (in the supremum metric) as $T'' \to \infty$. By Assumption 4 (iii), $P_X\left(X(\cdot) \leq c_{T''}(\cdot)\right) \to P_X\left(X(\cdot) \leq c^*(\cdot)\right)$ as $T'' \to \infty$. On the other hand, by the assumption of $X_T \rightsquigarrow X$ and the continuous mapping theorem, $X_{T''} - c_{T''} \rightsquigarrow X - c^*$. Hence, Assumption 4 (iii) and the Portmanteau theorem[37] imply that $P_{X_{T''}}\left(X_{T''}(\cdot) - c_{T'}(\cdot) \leq 0\right) \to P_X\left(X(\cdot) - c^*(\cdot) \leq 0\right)$ as $T'' \to \infty$. Combining these, we have shown $\left| P_{X_{T''}}\left(X_{T''}(\cdot) \leq c_{T''}(\cdot)\right) - P_X\left(X(\cdot) \leq c_{T''}(\cdot)\right)\right| \to 0$ along T''. This contradicts (A.3), so the convergence (A.2) holds.

Next, we show convergence (B). By Assumption 4 (iv), $X_{Y^T|\phi_0} - \hat{c}_T \rightsquigarrow Z - c$. Since Z is distributed identically to X by Assumption 4 (ii) and X is continuously distributed in the sense of Assumption 4 (iii), an application of the Portmanteau theorem gives convergence of $P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \leq \hat{c}_{T'}(\cdot)\right)$ to $P_Z\left(Z(\cdot) \leq c(\cdot)\right) = P_X\left(X(\cdot) \leq c(\cdot)\right)$. On the other hand, with Assumption 4 (iii) and (iv), the continuous mapping theorem implies $P_X\left(X(\cdot) \leq \hat{c}(\cdot)\right) \to_{p_{Y^T|\phi_0}} P_X\left(X(\cdot) \leq c(\cdot)\right)$. Combining these two convergence claims, convergence (B) is obtained. ■

Proof of Proposition 2. We first show that Assumption 5 implies Assumption 4 (i) - (iii). Set $a_T = \sqrt{T}$. When $k = 1$, the domain of the support function of $IS_\eta(\phi)$ consists of two points $\mathcal{S}^0 = \{-1, 1\}$, and the stochastic processes considered in Assumption 4 (i) and (ii) are reduced to bivariate random variables corresponding to the lower and upper bounds of $IS_\eta(\phi)$,

$$\begin{aligned} X_{\phi|Y^T} &= \sqrt{T}\begin{pmatrix} \ell(\phi) - \ell\left(\hat{\phi}\right) \\ u(\phi) - u\left(\hat{\phi}\right) \end{pmatrix}, \\ X_{Y^T|\phi_0} &= \sqrt{T}\begin{pmatrix} \ell(\phi_0) - \ell\left(\hat{\phi}\right) \\ u(\phi_0) - u\left(\hat{\phi}\right) \end{pmatrix}. \end{aligned}$$

[36] See, e.g., pp. 264 of Ok (2007).

[37] See, e.g., Theorem 4.25 of Kallenberg (2001).

By the delta method, the asymptotic distribution of $X_{Y^T|\phi_0}$ is

$$X_{Y^T|\phi_0} \rightsquigarrow \mathcal{N}\left(G'_{\phi_0}\Sigma_\phi G_{\phi_0}\right),$$

where $G_\phi \equiv \left(\frac{\partial \ell}{\partial \phi}(\phi), \frac{\partial u}{\partial \phi}(\phi)\right)$.

For $X_{\phi|Y^T}$, first order mean value expansion at $\hat{\phi}$ leads to

$$X_{\phi|Y^T} = G'_{\tilde{\phi}} \cdot \sqrt{T}\left(\phi - \hat{\phi}\right),$$

where $\tilde{\phi} = \lambda_\phi \phi + (1-\lambda_\phi)\hat{\phi}$, for some $\lambda_\phi \in [0,1]$. Since $\hat{\phi}$ is assumed to be strongly consistent to ϕ_0 and Assumption 5 (i) implies that $\tilde{\phi}$ converges in $\pi_{\phi|Y^T}$-probability to $\hat{\phi}$, $p_{Y^\infty|\phi_0}$-a.s., $G_{\tilde{\phi}}$ converges in $\pi_{\phi|Y^T}$-probability to G_{ϕ_0}, $p_{Y^\infty|\phi_0}$-a.s. Combining with $\left(\phi - \hat{\phi}\right)|Y^T \rightsquigarrow \mathcal{N}(0, \Sigma_\phi)$, $p_{Y^\infty|\phi_0}$-a.s., we conclude $X_{\phi|Y^T} \rightsquigarrow \mathcal{N}\left(G'_{\phi_0}\Sigma_\phi G_{\phi_0}\right)$, $p_{Y^\infty|\phi_0}$-a.s. Hence, Assumption 4 (i) and (ii) follow. Assumption 4 (iii) clearly holds by the properties of the bivariate normal distribution.

Next, we show that C^*_α meets Assumption 4 (iv). We represent connected intervals by $C = \left[\ell\left(\hat{\phi}\right) - c_\ell/\sqrt{T}, u\left(\hat{\phi}\right) + c_u/\sqrt{T}\right]$, $(c_\ell, c_u) \in \mathcal{R}^2$. Denote the posterior lower probability of C as a function of $\mathbf{c} \equiv (c_\ell, c_u)'$,

$$\begin{aligned} J_T(\mathbf{c}) &\equiv \pi_{\eta|Y^T}(C) \\ &= \pi_{\phi|Y^T}\left(\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} X_{\phi|Y^T} \le \mathbf{c}\right). \end{aligned}$$

Denoting the shortest robust credible region as $C^*_\alpha = \left[\ell\left(\hat{\phi}\right) - \hat{c}_{\ell,T}/\sqrt{T}, u\left(\hat{\phi}\right) + \hat{c}_{u,T}/\sqrt{T}\right]$, $\hat{\mathbf{c}}_T \equiv (\hat{c}_{\ell,T}, \hat{c}_{u,T})'$ is obtained by

$$\begin{aligned} &\hat{\mathbf{c}}_T \in \arg\min_{\mathbf{c}} \{c_\ell + c_u\} \\ &\text{s.t.} \quad J_T(\mathbf{c}) \ge \alpha. \end{aligned}$$

Having shown Assumption 4 (i), $X_{\phi|Y^T} \rightsquigarrow X$ as $T \to \infty$, $p_{Y^\infty|\phi_0}$-a.s., holds, let

$$J(\mathbf{c}) \equiv P_X\left(\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} X \le \mathbf{c}\right).$$

Note that the weak convergence $X_{\phi|Y^T} \rightsquigarrow X$ and continuity of $J(\cdot)$ implies $J_T(\mathbf{c}) \to J(\mathbf{c})$ as $T \to \infty$, $p_{Y^\infty|\phi_0}$-a.s., for any $\mathbf{c} \in \mathcal{R}^2$. Let $\mathbf{c}^* = (c^*_\ell, c^*_u)'$ be a solution of the following minimization problem

$$\begin{aligned} &\mathbf{c}^* \in \arg\min_{\mathbf{c}} \{c_\ell + c_u\} \\ &\text{s.t.} \quad J(\mathbf{c}) \ge \alpha. \end{aligned}$$

Since $\{\mathbf{c} : J(\mathbf{c}) \ge \alpha\}$ is the upper level set of the bivariate normal CDF, which is strictly convex and bounded from below, and the objective function is linear in $\mathbf{c}$, $\mathbf{c}^*$ is unique.

In what follows, we prove $\hat{\mathbf{c}}_T \to \mathbf{c}^*$ in $p_{Y^T|\phi_0}$-probability as $T \to \infty$. Our proof uses the following lemma, whose proof is given after completing the proof of the current proposition.

Lemma A.4 *Let Lev_T and Lev be the α-level sets of $J_T(\cdot)$ and $J(\cdot)$, respectively,*

$$\begin{aligned} Lev_T &= \left\{\mathbf{c} \in \mathcal{R}^2 : J_T(\mathbf{c}) \geq \alpha\right\}, \\ Lev_\alpha &= \left\{\mathbf{c} \in \mathcal{R}^2 : J(\mathbf{c}) \geq \alpha\right\}. \end{aligned}$$

Define a distance from point $\mathbf{c} \in \mathcal{R}^2$ to set $F \subset \mathcal{R}^2$ by $d(\mathbf{c}, F) \equiv \inf_{\mathbf{c}' \in F} \|\mathbf{c} - \mathbf{c}'\|$, where $\|\cdot\|$ is the Eucledean distance. Under Assumption 2, (a) $d(\mathbf{c}, Lev_T) \to 0$ in $p_{Y^T|\phi}$-probability for every $\mathbf{c} \in Lev$, and (b) $d(\mathbf{c}_T, Lev) \to 0$ in $p_{Y^T|\phi}$-probability for every $\{\mathbf{c}_T : T = 1, 2, \dots\}$ sequence of measurable selections of Lev_T.

Let $K_T = \arg\min\{c_l + c_u : J_T(\mathbf{c}) \geq \alpha\}$, and suppose that $\hat{\mathbf{c}}_T \to \mathbf{c}^*$ in $p_{Y^T|\phi_0}$-probability is false. That is, suppose that there exist ϵ, $\delta > 0$, and subsequence T' such that

$$P_{Y^{T'}|\phi_0}\left(\|\hat{\mathbf{c}}_{T'} - \mathbf{c}^*\| > \epsilon\right) > \delta \tag{A.4}$$

holds for all T'. Since $\hat{\mathbf{c}}_{T'}$ is a selection from Lev_T, Lemma A.4 (b) ensures that there exists a sequence of selections in Lev, $\tilde{\mathbf{c}}_T = (\tilde{c}_{\ell,T}, \tilde{c}_{u,T})'$, such that $\|\hat{\mathbf{c}}_{T''} - \tilde{\mathbf{c}}_{T'}\| \to 0$ in $p_{Y^{T'}|\phi_0}$-probability along T'. Consequently, (A.4) implies that an analogous statement holds also for $\tilde{\mathbf{c}}_{T'}$ for all large T'. Let $\hat{f}_T = \hat{c}_{\ell,T} + \hat{c}_{u,T}$, $\tilde{f}_T = \tilde{c}_{\ell,T} + \tilde{c}_{u,T}$, and $f^* = c^*_\ell + c^*_u$. By continuity of the value function, the claim $P_{Y^{T'}|\phi_0}\left(\|\tilde{\mathbf{c}}_{T'} - \mathbf{c}^*\| > \epsilon\right) > \delta$ for all large T' and $\tilde{\mathbf{c}}_T \in Lev$ imply existence of $\xi > 0$ such that $P_{Y^{T'}|\phi_0}\left(\tilde{f}_{T'} - f^* > \xi\right) > \delta$ for all large T'. Also, since $\|\hat{\mathbf{c}}_{T''} - \tilde{\mathbf{c}}_{T'}\| \to 0$ in $p_{Y^{T'}|\phi_0}$-probability implies $\left|\hat{f}_{T'} - \tilde{f}_{T'}\right| \to 0$ in $p_{Y^{T'}|\phi_0}$-probability, it also holds

$$P_{Y^{T'}|\phi_0}\left(\hat{f}_{T'} - f^* > \xi\right) > \delta, \tag{A.5}$$

for all large T'.

In order to derive a contradiction, apply Lemma A.4 (a) to construct a sequence $\check{\mathbf{c}}_{T'} = \left(\check{c}_{\ell,T'}, \check{c}_{u,T'}\right) \in Lev_T$ such that $\|\check{\mathbf{c}}_{T'} - \mathbf{c}^*\| \to 0$ in $p_{Y^{T'}|\phi_0}$-probability. Then, we have $f^* - \left(\check{c}_{\ell,T'} + \check{c}_{u,T'}\right) \to 0$ in $p_{Y^{T'}|\phi_0}$-probability and, combined with (A.5),

$$P_{Y^{T'}|\phi_0}\left(\hat{f}_{T'} - \left(\check{c}_{\ell,T'} + \check{c}_{u,T'}\right) > \xi\right) > \delta,$$

for all large T'. This means that the value of the objective function evaluated at feasible point $\check{\mathbf{c}}_{T'} \in Lev_{T'}$ is strictly smaller than the value evaluated at $\hat{\mathbf{c}}_{T'}$ with a positive probability for all large T'. This contradicts that $\hat{\mathbf{c}}_T$ is a minimizer in Lev_T for all T. This completes the proof for Assumption 4 (iv). ■

Proof of Lemma A.4. To prove (a), suppose that the conclusion is false. That is, there exist subsequence T', ϵ, $\delta > 0$, and $\mathbf{c} = (c_\ell, c_u) \in Lev$ such that $P_{Y^{T'}|\phi_0}\left(d\left(\mathbf{c}, Lev_{T'}\right) > \epsilon\right) > \delta$ for all T'. Event $d\left(\mathbf{c}, Lev_{T'}\right) > \epsilon$ implies $J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) < \alpha$ since $\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) \notin Lev_{T'}$. Therefore, it holds that

$$P_{Y^{T'}|\phi_0}\left(J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) < \alpha\right) > \delta \tag{A.6}$$

along T'. Under Assumption 4 (i), however,

$$J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) - J\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) \to 0,\ p_{Y^\infty|\phi_0}\text{-a.s.},$$

This convergence combined with strict monotonicity of $J(\cdot)$ implies

$$J\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) > J(\mathbf{c}) \geq \alpha$$

Hence, $P_{Y^{T'}|\phi_0}\left(J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) > \alpha\right) \to 1$ as $T' \to \infty$, but this contradicts (A.6).

To prove (b), suppose again that the conclusion is false. This implies there exist subsequence T', ϵ, $\delta > 0$, and a sequence of (random) measurable selections $\mathbf{c}_T = (c_{\ell,T}, c_{u,T})'$ from Lev_T such that $P_{Y^{T'}|\phi_0}\left(d\left(\mathbf{c}_{T'}, Lev\right) > \epsilon\right) > \delta$ for all T'. Since $d\left(\mathbf{c}_{T'}, Lev\right) > \epsilon$ implies $J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) < \alpha$, it holds

$$P_{Y^{T'}|\phi_0}\left(J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) < \alpha\right) > \delta \tag{A.7}$$

along T'. To find contradiction, note that

$$\begin{aligned} J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) &= \left[J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) - J\left(\mathbf{c}_{T'}\right)\right] \\ &\quad + \left[J\left(\mathbf{c}_{T'}\right) - J_T\left(\mathbf{c}_{T'}\right)\right] + J_T\left(\mathbf{c}_{T'}\right) \\ &> \left[J\left(\mathbf{c}_{T'}\right) - J_{T'}\left(\mathbf{c}_{T'}\right)\right] + \alpha \to \alpha \end{aligned}$$

in $p_{Y^{T'}|\phi_0}$-probability, where the strict inequality follows from that $J(\cdot)$ is strictly monotonic and $J_T\left(\mathbf{c}_{T'}\right) \geq \alpha$, and the convergence in probability in the last line follows from the continuity of $J(\cdot)$ and $\sup_{\mathbf{c}\in\mathcal{R}^2}\left|J(\mathbf{c}) - J_{T'}(\mathbf{c})\right| \to 0$ for any sequence of distributions $J_{T'}$ converging weakly to a distribution with continuous CDF (see, e.g., Lemma 2.11 in van der Vaart (1998)). This in turn implies $P_{Y^{T'}|\phi_0}\left(J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) \geq \alpha\right) \to 1$ as $T' \to \infty$, which contradicts (A.7). ■

Notation: The proofs given below use the following notation. For given $\phi \in \Phi$ and $i = 1, \ldots, n$, let $\tilde{f}_i(\phi) \equiv rank\left(F_i(\phi)\right)$. Since the rank of $F_i(\phi)$ is determined by its row rank, $\tilde{f}_i(\phi) \leq f_i(\phi)$ holds. Let $\mathcal{F}_i^\perp(\phi)$ be the linear subspace of $\mathcal{R}^n$ that is orthogonal to the row vectors of $F_i(\phi)$. If no zero restrictions are placed on q_i, we interpret $\mathcal{F}_i^\perp(\phi)$ to be $\mathcal{R}^n$. Note that the dimension of $\mathcal{F}_i^\perp(\phi)$ is equal to $n - \tilde{f}_i(\phi)$. We let $\mathcal{H}_i(\phi)$ be the half-space in $\mathcal{R}^n$ defined by the sign normalization restriction $\left\{z \in \mathcal{R}^n : \left(\sigma^i\right)' z \geq 0\right\}$, where σ^i is the i-th column vector of Σ_{tr}^{-1}. Given linearly independent vectors, $A = [a_1, \ldots, a_j] \in \mathcal{R}^{n \times j}$, denote the linear subspace in $\mathcal{R}^n$ that is orthogonal to the column vectors of A by $\mathcal{P}(A)$. Note that the dimension of $\mathcal{P}(A)$ is $n - j$.

Proof of Proposition 3. Consider first the case with only zero restrictions (Case (I)). Fix $\phi \in \Phi$. Let $Q_{1:i} = [q_1, \ldots, q_i]$, $i = 2, \ldots, (n-1)$, be an $n \times i$ matrix of orthonormal vectors in $\mathcal{R}^n$. The set of feasible Q's satisfying the zero restrictions and the sign normalizations, $\mathcal{Q}(\phi|F)$, can be written in the following recursive manner,

$$
\begin{aligned}
Q &= [q_1, \ldots, q_n] \in \mathcal{Q}(\phi|F) \\
& \quad \text{if and only if } Q = [q_1, \ldots, q_n] \text{ satisfies} \\
q_1 &\in D_1(\phi) \equiv \mathcal{F}_1^{\perp}(\phi) \cap \mathcal{H}_1(\phi) \cap \mathcal{S}^{n-1}, \\
q_2 &\in D_2(\phi, q_1) \equiv \mathcal{F}_2^{\perp}(\phi) \cap \mathcal{H}_2(\phi) \cap \mathcal{P}(q_1) \cap \mathcal{S}^{n-1}, \\
q_3 &\in D_3(\phi, Q_{1:2}) \equiv \mathcal{F}_3^{\perp}(\phi) \cap \mathcal{H}_3(\phi) \cap \mathcal{P}(Q_{1:2}) \cap \mathcal{S}^{n-1}, \\
& \quad \vdots \\
q_j &\in D_j\left(\phi, Q_{1:(j-1)}\right) \equiv \mathcal{F}_j^{\perp}(\phi) \cap \mathcal{H}_j(\phi) \cap \mathcal{P}(Q_{1:(j-1)}) \cap \mathcal{S}^{n-1}, \\
& \quad \vdots \\
q_n &\in D_n\left(\phi, Q_{1:(n-1)}\right) \equiv \mathcal{F}_n^{\perp}(\phi) \cap \mathcal{H}_n(\phi) \cap \mathcal{P}(Q_{1:(n-1)}) \cap \mathcal{S}^{n-1}.
\end{aligned}
\tag{A.8}
$$

where $D_i\left(\phi, Q_{1:(i-1)}\right) \subset \mathcal{R}^n$ denotes the set of feasible q_i's given $Q_{1:(i-1)} = [q_1, \ldots, q_{i-1}]$, the set of $(i-1)$ orthonormal vectors in $\mathcal{R}^n$ preceding i. Nonemptiness of the identified set for $\eta = c_{ih}(\phi)\, q_j$ follows if the feasible domain of the orthogonal vector $D_i\left(\phi, Q_{1:(i-1)}\right)$ is nonempty at every $i = 1, \ldots, n$.

Note that by the assumption $f_1 \leq n-1$, $\mathcal{F}_1^{\perp}(\phi) \cap \mathcal{H}_1(\phi)$ is the half-space of the linear subspace of $\mathcal{R}^n$ with dimension $n - \tilde{f}_1(\phi) \geq n - f_1 \geq 1$. Hence, $D_1(\phi)$ is nonempty for every $\phi \in \Phi$. For $i = 2, \ldots, n$, $\mathcal{F}_i^{\perp}(\phi) \cap \mathcal{H}_i(\phi) \cap \mathcal{P}(Q_{1:(i-1)})$ is the half-space of the linear subspace of $\mathcal{R}^n$ with dimension at least

$$
\begin{aligned}
n - \tilde{f}_i(\phi) - \dim(\mathcal{P}(Q_{1:(n-1)})) &\geq n - f_i - (i-1) \\
&\geq 1,
\end{aligned}
$$

where the last inequality follows from the assumption $f_i \leq n - i$. Hence, $D_i\left(\phi, Q_{1:(i-1)}\right)$ is nonempty for every $\phi \in \Phi$. We thus conclude that $\mathcal{Q}(\phi|F)$ is nonempty, and this implies nonemptiness of the impulse response identified sets for every $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, n\}$, and $h = 0, 1, 2, \ldots$. The boundedness of the identified sets follows since $|c_{ih}(\phi)\, q_j| \leq \|c_{ih}(\phi)\| < \infty$ for any $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, n\}$, and $h = 0, 1, 2, \ldots$, where the boundedness of $\|c_{ih}(\phi)\|$ is ensured by the restriction on ϕ that the reduced-form VAR is invertible to VMA(∞).

Next we show convexity of the identified set of the impulse response to the j^*-th shock under each one of conditions (i) - (iii). Suppose $j^* = 1$ and $f_1 < n-1$ (Condition (i)). Since $\tilde{f}_1(\phi) < n-1$

for all $\phi \in \Phi$, $D_1(\phi)$ is a path-connected set because it is an intersection of the half-space with dimension at least 2 and the unit sphere. Since the impulse response is a continuous function of q_1, the identified set of $\eta = c_{ih}(\phi) q_1$ is an interval, as the set of a continuous function with a path-connected domain is always an interval (see, e.g., Propositions 12.11 and 12.23 in Sutherland (2009)).

Suppose $j^* \geq 2$ and assume condition (ii) holds. Denote the set of feasible q_{j^*}'s by $\mathcal{E}_{j^*}(\phi) \equiv \left\{q_{j^*} \in \mathcal{S}^{n-1} : Q \in \mathcal{Q}(\phi|F)\right\}$. The next lemma provides a specific expression for $\mathcal{E}_{j^*}(\phi)$. We defer its proof to a later part of this appendix.

Lemma A.5 *Suppose $j^* \geq 2$ and assume condition (ii) of Proposition 3 holds. Then $\mathcal{E}_{j^*}(\phi) = \mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \mathcal{S}^{n-1}$.*

This lemma shows that $\mathcal{E}_{j^*}(\phi)$ is an intersection of a half-space of a linear subspace with dimension $n - f_{j^*} \geq j^* \geq 2$ with the unit sphere. Hence, $\mathcal{E}_{j^*}(\phi)$ is a path-connected set on $\mathcal{S}^{n-1}$ and convexity of $IS_\eta(\phi|F)$ follows.

Next, suppose condition (iii) holds. Let $Q_{1:i^*}(\phi) \equiv [q_1(\phi), \ldots, q_{i^*}(\phi)]$ be the first i^* columns of feasible $Q \in \mathcal{Q}(\phi|F)$ that are common for all $Q \in \mathcal{Q}(\phi|F)$ by the assumption of exact identification of the first i^* columns. In this case, the set of feasible q_{j^*}'s can be expressed as in the next lemma (see a later part of this appendix for its proof).

Lemma A.6 *Suppose $j^* \geq 2$ and assume condition (iii) of Proposition 3 holds. Then, whenever $Q_{1:i^*}(\phi) = (q_1(\phi), \ldots, q_{i^*}(\phi))$ is uniquely determined as a function of ϕ (this is the case for almost every $\phi \in \Phi$ by the assumption of exact identification), $\mathcal{E}_{j^*}(\phi) = \mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \mathcal{P}(Q_{1:i^*}(\phi)) \cap \mathcal{S}^{n-1}$.*

This lemma shows that $\mathcal{E}_{j^*}(\phi)$ is an intersection of a half-space of a linear subspace with dimension $n - f_{j^*} - i^* \geq j^* + 1 - i^* \geq 2$ with the unit sphere. Hence, $\mathcal{E}_{j^*}(\phi)$ is a path-connected set on $\mathcal{S}^{n-1}$ and convexity of $IS_\eta(\phi|F)$ follows.

For the cases under condition (i) or (ii), since $\phi \in \Phi$ is arbitrary, the convexity of the impulse-response identified set holds for every $\phi \in \Phi$. As for the case of condition (iii), the exact identification of $[q_1(\phi), \ldots, q_{i^*}(\phi)]$ assumes its unique determination up to almost every $\phi \in \Phi$, so convexity of the identified set holds for almost every $\phi \in \Phi$.

Next, consider the case with both zero and sign restrictions (Case (II)). Suppose $j^* = 1$ and $f_1 < n - 1$ (condition (i)). Following (A.8), the set of feasible q_1's can be denoted by $D_1(\phi) \cap \{x \in \mathcal{R}^n : S_1(\phi)x \geq 0\}$. Let $\tilde{q}_1 \in D_1(\phi)$ be a unit-length vector that satisfies $\begin{pmatrix} S_1(\phi) \\ (\sigma^1)' \end{pmatrix} \tilde{q}_1 > \mathbf{0}$. Such $\tilde{q}_1$ is guaranteed to exist by the assumption stated in the current proposition. Let $q_1 \in D_1(\phi) \cap \{x \in \mathcal{R}^n : S_1(\phi)x \geq 0\}$ be arbitrary. Note that $q_1 \neq -\tilde{q}_1$ must hold, since otherwise some of the sign restrictions are violated. Consider

$$q_1(\lambda) = \frac{\lambda q_1 + (1-\lambda)\tilde{q}_1}{\|\lambda q_1 + (1-\lambda)\tilde{q}_1\|}, \quad \lambda \in [0,1],$$

which is a connected path in $D_1(\phi) \cap \{x \in \mathcal{R}^n : S_1(\phi)x \geq \mathbf{0}\}$ since the denominator is nonzero for all $\lambda \in [0,1]$ by the fact that $q_1 \neq -\tilde{q}_1$. Since q_1 is arbitrary, we can connect any points in $D_1(\phi) \cap \{x \in \mathcal{R}^n : S_1(\phi)x \geq \mathbf{0}\}$ by connected paths via $\tilde{q}_1$. Hence, $D_1(\phi) \cap \{x \in \mathcal{R}^n : S_1(\phi)x \geq \mathbf{0}\}$ is path-connected, and convexity of the impulse-response identified set follows.

Suppose $j^* \geq 2$ and assume that the imposed zero restrictions satisfy condition (ii). Let $\mathcal{E}_{j^*}(\phi) \equiv \{q_{j^*} \in \mathcal{S}^{n-1} : Q \in \mathcal{Q}(\phi|F,S)\}$, and let $\tilde{q}_{j^*} \in \mathcal{E}_{j^*}(\phi)$ be chosen so as to satisfy $\begin{pmatrix} S_{j^*}(\phi) \\ [\sigma^{j^*}(\phi)]' \end{pmatrix} \tilde{q}_{j^*} > \mathbf{0}$. Such $\tilde{q}_{j^*}$ exists by the assumption stated in the current proposition. For any $q_{j^*} \in \mathcal{E}_{j^*}(\phi)$, $q_{j^*} \neq -\tilde{q}_{j^*}$ must be true, since otherwise q_{j^*} would violate some of the imposed sign restrictions. Consider constructing a path between q_{j^*} and $\tilde{q}_{j^*}$ as follows. For $\lambda \in [0,1]$, let

$$q_{j^*}(\lambda) = \frac{\lambda \tilde{q}_{j^*} + (1-\lambda) q_{j^*}}{\|\lambda \tilde{q}_{j^*} + (1-\lambda) q_{j^*}\|}, \tag{A.9}$$

which is a continuous path on the unit sphere since the denominator is nonzero for all $\lambda \in [0,1]$ by the construction of $\tilde{q}_{j^*}$. Along this path, $F_{j^*}(\phi) q_{j^*}(\lambda) = \mathbf{0}$ and the sign restrictions hold. Hence, for every $\lambda \in [0,1]$, if there exists $Q(\lambda) \equiv [q_1(\lambda), \ldots, q_{j^*}(\lambda), \ldots, q_n(\lambda)] \in \mathcal{Q}(\phi|F,S)$, then the path-connectedness of $\mathcal{E}_{j^*}(\phi)$ follows. A recursive construction similar to Algorithm 3 in Appendix B can be used to construct such $Q(\lambda) \in \mathcal{Q}(\phi|F,S)$. For $i = 1, \ldots, (j^*-1)$, we recursively obtain $q_i(\lambda)$ that solves

$$\begin{pmatrix} F_i(\phi) \\ q_1'(\lambda) \\ \vdots \\ q_{i-1}'(\lambda) \\ q_{j^*}'(\lambda) \end{pmatrix} q_i(\lambda) = \mathbf{0}, \tag{A.10}$$

and satisfies $[\sigma^i(\phi)]' q_i(\lambda) \geq 0$. Such a $q_i(\lambda)$ always exists since the rank of the matrix multiplied to $q_i(\lambda)$ is at most $f_i + i$, which is less than n under condition (ii). For $i = (j^*+1), \ldots, n$, a direct application of Algorithm 3 yields a feasible $q_i(\lambda)$. Thus, existence of $Q(\lambda) \in \mathcal{Q}(\phi|F,S)$, $\lambda \in [0,1]$, is established. We therefore conclude that $\mathcal{E}_{j^*}(\phi)$ is path-connected under condition (ii), and the convexity of impulse-response identified sets holds for every variable and every horizon. This completes the proof for case (iv) of the current proposition.

Last, we consider case (v). Suppose that the imposed zero restrictions satisfy condition (iii) of the current proposition. Let $[q_1(\phi), \ldots, q_{i^*}(\phi)]$ be the first i^*-th columns of feasible $Q's$, that are common for all $Q \in \mathcal{Q}(\phi|F,S)$, ϕ-a.s., by exact identification of the first i^*-columns. Let

$\tilde{q}_{j^*} \in \mathcal{E}_{j^*}(\phi)$ be chosen so as to satisfy $\begin{pmatrix} S_{j^*}(\phi) \\ (\sigma^{j^*})' \end{pmatrix} \tilde{q}_{j^*} > \mathbf{0}$, and $q_{j^*} \in \mathcal{E}_{j^*}(\phi)$ be arbitrary. Consider $q_{j^*}(\lambda)$ in (A.9) and construct $Q(\lambda) \in \mathcal{Q}(\phi|F,S)$ as follows. The first i^*-th column of $Q(\lambda)$ must be $[q_1(\phi), \ldots, q_{i^*}(\phi)]$, ϕ-a.s., by the assumption of exact identification. For $i = (i^*+1), \ldots, (j^*-1)$, we can recursively obtain $q_i(\lambda)$ that solves

$$\begin{pmatrix} F_i(\phi) \\ q_1'(\phi) \\ \vdots \\ q_{i^*}'(\phi) \\ q_{i^*+1}'(\lambda) \\ \vdots \\ q_{i-1}'(\lambda) \\ q_{j^*}'(\lambda) \end{pmatrix} q_i(\lambda) = \mathbf{0} \tag{A.11}$$

and satisfies $[\sigma^i(\phi)]' q_i(\lambda) \geq 0$. There always exist such $q_i(\lambda)$ because $f_i < n-i$ for all $i = (i^*+1), \ldots, (j^*-1)$. The rest of the column vectors $q_i(\lambda)$, $i = j^*+1, \ldots, n$, of $Q(\lambda)$ are obtained successively by applying Algorithm 3 in Appendix B. Having shown a feasible construction of $Q(\lambda) \in \mathcal{Q}(\phi|F,S)$ for $\lambda \in [0,1]$, we conclude that $\mathcal{E}_{j^*}(\phi)$ is path-connected, and convexity of the impulse-response identified sets follows for every variable and every horizon. ■

In what follows, we provide proofs for the lemmas used in the proof of Proposition 3.

Proof of Lemma A.5. Given zero restrictions $F(\phi, Q) = \mathbf{0}$ and the set of feasible orthogonal matrices $\mathcal{Q}(\phi|F)$, define the projection of $\mathcal{Q}(\phi|F)$ with respect to the first i column vectors,

$$\mathcal{Q}_{1:i}(\phi|F) \equiv \{[q_1, \ldots, q_i] : Q \in \mathcal{Q}(\phi|F)\}.$$

Following the recursive representation of (A.8), $\mathcal{E}_{j^*}(\phi) \equiv \{q_{j^*} \in \mathcal{S}^{n-1} : Q \in \mathcal{Q}(\phi|F)\}$ can be written as

$$\begin{aligned} \mathcal{E}_{j^*}(\phi) &= \bigcup_{Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)} \left[\mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \mathcal{P}(Q_{1:(j^*-1)}) \cap \mathcal{S}^{n-1}\right] \\ &= \mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \left[\bigcup_{Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)} \mathcal{P}(Q_{1:(j^*-1)})\right] \cap \mathcal{S}^{n-1}. \end{aligned}$$

Hence, the conclusion follows if we can show $\bigcup_{Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)} \mathcal{P}(Q_{1:(j^*-1)}) = \mathcal{S}^{n-1}$. To show this claim, let $q \in \mathcal{S}^{n-1}$ be arbitrary, and we construct $Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)$ such that

$q \in \mathcal{P}(Q_{1:(j^*-1)})$ holds. Specifically, construct q_i, $i = 1, \ldots, (j^*-1)$, successively, by solving

$$\begin{pmatrix} F_i(\phi) \\ q_1' \\ \vdots \\ q_{i-1}' \\ q' \end{pmatrix} q_i = \mathbf{0},$$

and choose the sign of q_i to satisfy its sign normalization. Under condition (ii) of Proposition 3, $q_i \in \mathcal{S}^{n-1}$ solving these equalities exists since the rank of the coefficient matrix is at most $f_i + i < n$. The obtained $Q_{1:(j^*-1)} = [q_1, \ldots, q_{j^*-1}]$ belongs to $\mathcal{Q}_{1:(j^*-1)}(\phi|F)$ by construction, and it is orthogonal to q. Hence, $q \in \mathcal{P}(Q_{1:(j^*-1)})$. Since q is arbitrary, we obtain $\bigcup_{Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)} \mathcal{P}(Q_{1:(j^*-1)}) = \mathcal{S}^{n-1}$. ■

Proof of Lemma A.6. Let $Q_{1:i^*}(\phi) \equiv [q_1(\phi), \ldots, q_{i^*}(\phi)]$ be the first i^*-th columns of feasible $Q \in \mathcal{Q}(\phi|F)$, that are common for all $Q \in \mathcal{Q}(\phi|F)$, ϕ-a.s., by exact identification of the first i^*-columns. As in the proof of Lemma A.1, $\mathcal{E}_{j^*}(\phi)$ can be written as

$$\begin{aligned} \mathcal{E}_{j^*}(\phi) &= \mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \left[\bigcup_{Q_{1:(j^*-1)} \in \mathcal{Q}_{1:(j^*-1)}(\phi|F)} \mathcal{P}(Q_{1:(j^*-1)}) \right] \cap \mathcal{S}^{n-1} \\ &= \mathcal{F}_{j^*}^{\perp}(\phi) \cap \mathcal{H}_{j^*}(\phi) \cap \mathcal{P}(Q_{1:i^*}(\phi)) \cap \bigcup_{Q_{(i^*+1):(j^*-1)} \in \mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F)} \mathcal{P}(Q_{(i^*+1):(j^*-1)}) \cap \mathcal{S}^{n-1}, \end{aligned}$$

where $\mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F) = \left\{ Q_{(i^*+1):(j^*-1)} = [q_{i^*+1}, \ldots, q_{j^*-1}] : Q \in \mathcal{Q}(\phi|F) \right\}$ is the projection of $\mathcal{Q}(\phi|F)$ with respect to the (i^*+1)-th to (j^*-1)-th columns of Q. We now show that, under condition (iii) of Proposition 3, $\bigcup_{Q_{(i^*+1):(j^*-1)} \in \mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F)} \mathcal{P}(Q_{(i^*+1):(j^*-1)}) = \mathcal{S}^{n-1}$ holds. Let $q \in \mathcal{S}^{n-1}$ be arbitrary, and we consider constructing $Q_{(i^*+1):(j^*-1)} \in \mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F)$ such that $q \in \mathcal{P}(Q_{(i^*+1):(j^*-1)})$ holds. For $i = (i^*+1), \ldots, (j^*-1)$, we recursively obtain q_i by solving

$$\begin{pmatrix} F_i(\phi) \\ q_1'(\phi) \\ \vdots \\ q_{i^*}'(\phi) \\ q_{i^*+1}' \\ \vdots \\ q_{i-1}' \\ q' \end{pmatrix} q_i = \mathbf{0},$$

and choose the sign of q_i to be consistent with the sign normalization. Under condition (iii) of Proposition 3, $q_i \in \mathcal{S}^{n-1}$ solving these equalities exists since the rank of the coefficient matrix is at most $f_i + i < n$ for all $i = (i^* + 1), \ldots, (j^* - 1)$. The obtained $Q_{(i^*+1):(j^*-1)} = [q_{i^*+1}, \ldots, q_{j^*-1}]$ belongs to $\mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F)$ by construction, and it is orthogonal to q. Hence, $q \in \mathcal{P}(Q_{(i^*+1):(j^*-1)})$. Since q is arbitrary we have that

$$\bigcup_{Q_{(i^*+1):(j^*-1)} \in \mathcal{Q}_{(i^*+1):(j^*-1)}(\phi|F)} \mathcal{P}(Q_{(i^*+1):(j^*-1)}) = \mathcal{S}^{n-1}.$$

■

Proof of Proposition 4. (i) Following the notation introduced in the proof of Proposition 3, the upper and lower bounds of the impulse-response identified set are written as

$$\begin{aligned} u(\phi)/\ell(\phi) &= \max/\min_{q_{j^*}} c_{ih}'(\phi) q_{j^*}, \\ \text{s.t.,} \quad q_{j^*} &\in \mathcal{E}_{j^*}(\phi) \text{ and } S_{j^*}(\phi) q_{j^*} \geq \mathbf{0}. \end{aligned} \tag{A.12}$$

When $j^* = 1$ (Case (i) of Proposition 3), $\mathcal{E}_1(\phi)$ is given by $D_1(\phi)$ defined in (A.8). On the other hand, when $j^* \geq 2$ and Case (ii) of Proposition 3 applies, Lemma A.5 provides a concrete expression for $\mathcal{E}_{j^*}(\phi)$. Accordingly, in either case, the constrained set of q_{j^*} in (A.12) can be expressed as

$$\tilde{\mathcal{E}}_{j^*}(\phi) \equiv \left\{ q \in \mathcal{S}^{n-1} : F_{j^*}(\phi) q = \mathbf{0}, \begin{pmatrix} S_{j^*}(\phi) \\ (\sigma^{j^*}(\phi))' \end{pmatrix} q \geq \mathbf{0} \right\}.$$

The objective function of (A.12) is continuous in q_{j^*}, so, by the Theorem of Maximum (see, e.g., Theorem 9.14 of Sundaram (1996)), the continuity of $u(\phi)$ and $\ell(\phi)$ is obtained if $\tilde{\mathcal{E}}_{j^*}(\phi)$ is shown to be a continuous correspondence at $\phi = \phi_0$.

To show continuity of $\tilde{\mathcal{E}}_{j^*}(\phi)$, note first that $\tilde{\mathcal{E}}_{j^*}(\phi)$ is a closed and bounded correspondence, so upper-semicontinuity and lower-semicontinuity of $\tilde{\mathcal{E}}_{j^*}(\phi)$ can be defined in terms of sequences (see, e.g., Propositions 21 of Border (2013)):

- $\tilde{\mathcal{E}}_{j^*}(\phi)$ is upper-semicontinuous (usc) at $\phi = \phi_0$ if and only if, for any sequence $\phi^v \to \phi_0$, $v = 1, 2, \ldots$, and any $q_{j^*}^v \in \tilde{\mathcal{E}}_{j^*}(\phi^v)$, there is a subsequence of $q_{j^*}^v$ with limit in $\tilde{\mathcal{E}}_{j^*}(\phi_0)$.
- $\tilde{\mathcal{E}}_{j^*}(\phi)$ is lower-semicontinuous (lsc) at $\phi = \phi_0$ if and only if, $\phi^v \to \phi_0$, $v = 1, 2, \ldots$, and $q_{j^*}^0 \in \tilde{\mathcal{E}}_{j^*}(\phi_0)$ imply that there is a sequence $q_{j^*}^v \in \tilde{\mathcal{E}}_{j^*}(\phi^v)$ with $q_{j^*}^v \to q_{j^*}^0$.

In the proofs below, we use the same index v to denote a subsequence for brevity of notation.

Usc: Since $q_{j^*}^v$ is a sequence on the unit-sphere, it has a convergent subsequence $q_{j^*}^v \to q_{j^*}$. Since $q_{j^*}^v \in \tilde{\mathcal{E}}_{j^*}(\phi^v)$, $F_{j^*}(\phi^v) q_{j^*}^v = \mathbf{0}$ and $\begin{pmatrix} S_{j^*}(\phi^v) \\ (\sigma^{j^*}(\phi^v))' \end{pmatrix} q_{j^*}^v \geq \mathbf{0}$ hold for all v. Since $F_{j^*}(\cdot)$ and $\begin{pmatrix} S_{j^*}(\cdot) \\ (\sigma^{j^*}(\cdot))' \end{pmatrix}$ are continuous in ϕ, these equality and sign restrictions hold at the limit as well. Hence, $q_{j^*} \in \tilde{\mathcal{E}}_{j^*}(\phi_0)$.

Lsc: Our proof of lsc proceeds similarly to the proof of Lemma 3 in the 2013 working paper version of Granziera et al. (2018). Let $\phi^v \to \phi_0$ be arbitrary. Let $q^0_{j^*} \in \tilde{\mathcal{E}}_{j^*}(\phi_0)$, and define $\mathbf{P}^0 = F_{j^*}(\phi_0)' \left[F_{j^*}(\phi_0)F_{j^*}(\phi_0)'\right]^{-1} F_{j^*}(\phi_0)$ be the projection matrix onto the space spanned by the row vectors of $F_{j^*}(\phi_0)$. By the assumption of the current proposition, $F_{j^*}(\phi)$ has full row-rank in the open neighborhood of ϕ_0, so $\mathbf{P}^0$ and $\mathbf{P}^v = F_{j^*}(\phi^v)' \left[F_{j^*}(\phi^v)F_{j^*}(\phi^v)'\right]^{-1} F_{j^*}(\phi^v)$ are well-defined for all large v. Let $\boldsymbol{\xi}^* \in \mathcal{R}^n$ be a vector satisfying $\begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^0\right] \boldsymbol{\xi}^* >> \mathbf{0}$, which exists by the assumption. Let

$$\zeta = \min \left\{ \begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^0\right] \boldsymbol{\xi}^* \right\} > 0,$$

and define

$$\begin{aligned}
\boldsymbol{\xi} &= \frac{2}{\eta}\boldsymbol{\xi}^*, \\
\epsilon^v &= \left\| \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^v\right] - \begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^0\right] \right\|, \\
q^v_{j^*} &= \frac{\left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right]}{\left\| \left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right] \right\|}.
\end{aligned}$$

Since $\mathbf{P}^v$ converges to $\mathbf{P}^0$, $\epsilon^v \to 0$. Furthermore, $\left[I_n - \mathbf{P}^0\right] q^0_{j^*} = q^0_{j^*}$ implies that $q^v_{j^*}$ converges to $q^0_{j^*}$ as $v \to \infty$. Note that $q^v_{j^*}$ is orthogonal to $F_{j^*}(\phi^v)$ by construction. Furthermore, note that

$$\begin{aligned}
& \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} q^v_{j^*} \\
=\ & \frac{1}{\left\| \left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right] \right\|} \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[\left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right]\right] \\
\geq\ & \frac{1}{\left\| \left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right] \right\|} \begin{pmatrix} \left(\begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^v\right] - \begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^0\right] \right) q^0_{j^*} \\ + \epsilon^v \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^v\right] \boldsymbol{\xi} \end{pmatrix} \\
\geq\ & \frac{1}{\left\| \left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right] \right\|} \left(-\epsilon^v \left\| q^0_{j^*} \right\| \mathbf{1} + \epsilon^v \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^v\right] \boldsymbol{\xi} \right) \\
=\ & \frac{\epsilon^v}{\left\| \left[I_n - \mathbf{P}^v\right]\left[q^0_{j^*} + \epsilon^v \boldsymbol{\xi}\right] \right\|} \left(\frac{2}{\eta} \begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^v\right] \boldsymbol{\xi}^* - \mathbf{1} \right),
\end{aligned}$$

where the third line follows by $\begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} \left[I_n - \mathbf{P}^0\right] q^0_{j^*} = \begin{pmatrix} S_{j^*}(\phi_0) \\ \left(\sigma^{j^*}(\phi_0)\right)' \end{pmatrix} q^0_{j^*} \geq \mathbf{0}$. By the construction of $\boldsymbol{\xi}^*$ and ζ, $\frac{2}{\zeta}\begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix}[I_n - \mathbf{P}^v]\,\boldsymbol{\xi}^* > \mathbf{1}$ holds for all large v. This implies that $\begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} q_{j^*}^v \geq 0$ holds for all large v, implying that $q_{j^*}^v \in \tilde{\mathcal{E}}_{j^*}(\phi^v)$ for all large v. Hence, $\tilde{\mathcal{E}}_{j^*}(\phi)$ is lsc at $\phi = \phi_0$.

(ii) *Usc*: Under Case (iii) of Proposition 3, Lemma A.6 implies that the constraint set of q_{j^*} in (A.12) can be expressed as

$$\tilde{\mathcal{E}}_{j^*}(\phi) \equiv \left\{ q \in \mathcal{S}^{n-1} : \begin{pmatrix} F_{j^*}(\phi) \\ q_1'(\phi) \\ \vdots \\ q_{i^*}'(\phi) \end{pmatrix} q = \mathbf{0}, \ \begin{pmatrix} S_{j^*}(\phi) \\ \left(\sigma^{j^*}(\phi)\right)' \end{pmatrix} q \geq \mathbf{0} \right\}.$$

Let $q_{j^*}^v$, $v = 1, 2, \ldots$, be a sequence on the unit sphere, such that $q_{j^*}^v \in \tilde{\mathcal{E}}_{j^*}(\phi^v)$ holds for all v. This has a convergent subsequence $q_{j^*}^v \to q_{j^*}$. Since $F_i(\phi)$ are continuous in ϕ for all $i = 1, \ldots, i^*$, $q_i(\phi)$, $i = 1, \ldots, i^*$, are continuous in ϕ as well, implying that the equality restrictions and the sign restrictions, $\begin{pmatrix} F_{j^*}(\phi^v) \\ q_1'(\phi^v) \\ \vdots \\ q_{i^*}'(\phi^v) \end{pmatrix} q_{j^*}^v = \mathbf{0}$ and $\begin{pmatrix} S_{j^*}(\phi^v) \\ \left(\sigma^{j^*}(\phi^v)\right)' \end{pmatrix} q_{j^*}^v \geq \mathbf{0}$ must hold at the limit $v \to \infty$. Hence, $q_{j^*} \in \tilde{\mathcal{E}}_{j^*}(\phi_0)$.

Lsc: Define $\mathbf{P}^0$ and $\mathbf{P}^v$ as the projection matrices onto the row vectors of $\begin{pmatrix} F_{j^*}(\phi_0) \\ q_1'(\phi_0) \\ \vdots \\ q_{i^*}'(\phi_0) \end{pmatrix}$ and $\begin{pmatrix} F_{j^*}(\phi^v) \\ q_1'(\phi^v) \\ \vdots \\ q_{i^*}'(\phi^v) \end{pmatrix}$, respectively. The imposed assumptions imply that $\mathbf{P}^v$ and $\mathbf{P}^0$ are well-defined for all large v, and $\mathbf{P}^v \to \mathbf{P}^0$. With the current definition of $\mathbf{P}^v$ and $\mathbf{P}^0$, lower-semicontinuity of $\tilde{\mathcal{E}}_{j^*}(\phi)$ can be shown by repeating the same argument as in the proof of part (i) of the current proposition. We omit details for brevity. ■

Proof of Proposition 5. We show that in each of cases (i) - (iii) of Proposition 3 with the sign restrictions imposed on the j^*-th shock only, the optimization problem to be solved for the endpoints of the identified set $(\ell(\phi), u(\phi))$ is reduced to the optimization problem that Gafarov et al. (2018) analyze. The differentiability of the endpoints in ϕ then follows by directly applying Theorem 2 of Gafarov et al. (2018). Our proof focuses on the lower bound $\ell(\phi_0)$ only, as the conclusion for the upper bound can be proved similarly.

To show claim (i) of this proposition, assume $j^* = 1$ and $f_1 < n-1$ (i.e., case (i) of Proposition 3 with $\mathcal{I}_S = \{1\}$). The choice set of q_1 is given by $D_1(\phi_0) \cap \left\{q \in \mathcal{S}^{n-1} : S_1(\phi_0)q \geq 0\right\}$, where $D_1(\phi)$ is as defined in (A.8), and the optimization problem to obtain $\ell(\phi)$ can be written as (4.24) with $j^* = 1$. One-to-one differentiable reparametrization of $q \in \mathcal{S}^{n-1}$ by $x = \Sigma_{tr} q$ leads to the optimization problem in equation (2.5) of Gafarov et al. (2018). Hence, under the assumptions stated in claim (i) of the current proposition, their Theorem 2 proves differentiability of $\ell(\phi_0)$.

Assume that the imposed zero restrictions satisfy case (ii) of Proposition 3 with $\mathcal{I}_S = \{j^*\}$. By applying Lemma A.5, the choice set of q_{j^*} is given by $\mathcal{F}_{j^*}^{\perp}(\phi_0) \cap \mathcal{H}_{j^*}(\phi) \cap \left\{q \in \mathcal{S}^{n-1} : S_{j^*}(\phi_0)q \geq 0\right\}$, and the optimization problem to obtain $\ell(\phi_0)$ can be written as (4.25). One-to-one differentiable reparametrization of $q \in \mathcal{S}^{n-1}$ by $x = \Sigma_{tr} q$ leads to the optimization problem in equation (2.5) of Gafarov et al. (2018), so the conclusion follows by their Theorem 2.

Last, assume that the imposed zero restrictions satisfy case (iii) of Proposition 3 with $\mathcal{I}_S = \{j^*\}$. By applying Lemma A.6, the choice set of q_{j^*} is given by $\mathcal{F}_{j^*}^{\perp}(\phi_0) \cap \mathcal{H}_{j^*}(\phi_0) \cap \mathcal{P}(Q_{1:i^*}(\phi_0)) \cap \left\{q \in \mathcal{S}^{n-1} : S_{j^*}(\phi_0)q \geq 0\right\}$ with $Q_{1:i^*}(\phi_0) = [q_1(\phi_0), \ldots, q_{i^*}(\phi_0)]$ pinned down uniquely by the assumption of exact identification. Accordingly, the optimization problem to obtain $\ell(\phi_0)$ can be written as (4.25). One-to-one differentiable reparametrization of $q \in \mathcal{S}^{n-1}$ by $x = \Sigma_{tr} q$ leads to the optimization problem in equation (2.5) of Gafarov et al. (2018) with the expanded set of equality restrictions consisting of $F_{j^*}(\phi_0)(\Sigma_{tr})^{-1}x = 0$ and $Q_{1:i^*}(\phi_0)'(\Sigma_{tr})^{-1}x = 0$. Hence, under the assumptions stated in claim (ii) of the current proposition, their Theorem 2 implies differentiability of $\ell(\phi_0)$. ■

B Further Results on Convexity

The main results on convexity of the impulse-response identified set are discussed in Proposition 3 in the body of the paper. In this appendix we provide additional discussion, examples and results.

To gain some intuition behind the convexity results of Proposition 3, consider the case of equality restrictions that restrict a single column q_j by linear constraints of the form (4.11). Convexity of the identified set for η then follows if the subspace of q_j's constrained by the zero restrictions has dimension greater than one. The reason is that the set of feasible q_j's becomes a subset on the unit sphere in $\mathcal{R}^n$ where any two elements q_j and $q_{j'}$ are path-connected, which in turn implies a convex identified set for the impulse response because the impulse response is a continuous function of q_j. When the subspace has dimension one, non-convexity can occur if, for example, the identified set consists of two disconnected points, which means that the impulse response is locally, but not globally, identified. This argument implies that for almost every $\phi \in \Phi$, we can guarantee convexity of the identified set by finding a condition on the number of zero restrictions that yields a subspace of feasible q_j's with dimension greater than one.

As discussed in footnote 19, the following algorithm can be used to verify condition (iii) of Proposition 3.

Algorithm 3 *(Successive construction of orthonormal vectors, Algorithm 1 in Rubio-Ramírez et al. (2010)). Consider a collection of zero restrictions of the form given by (4.11), where the order of the variables is consistent with* $f_1 \geq f_2 \geq \cdots \geq f_n \geq 0$*. Assume* $f_i = n - i$ *for all* $i = 1, \ldots, i^*$*, and* $rank(F_i(\phi)) = f_i$ *for all* $i = 1, \ldots, i^*$*,* ϕ*-a.s. Let* q_1 *be a unit-length vector satisfying* $F_1(\phi)q_1 = 0$*, which is unique up to sign since* $rank(F_1(\phi)) = n - 1$ *by assumption. Given* q_1*, find orthonormal vectors* $q_2, \ldots, q_{i^*}$*, by solving*

$$\begin{pmatrix} F_i(\phi) \\ q_1' \\ \vdots \\ q_{i-1}' \end{pmatrix} q_i = 0,$$

successively for $i = 2, 3, \ldots, i^*$*. If*

$$rank \begin{pmatrix} F_i(\phi) \\ q_1' \\ \vdots \\ q_{i-1}' \end{pmatrix} = n - 1 \text{ for } i = 2, \ldots, i^*, \tag{B.1}$$

and q_i*,* $i = 1, \ldots, i^*$*, obtained by this algorithm satisfies* $(\sigma^i)' q_i \neq 0$ *for almost all* $\phi \in \Phi$*, i.e., the sign normalization restrictions determine a unique sign for the* $q_i's$*, then* $[q_1, \ldots, q_{i^*}]$ *is exactly identified.*[38]

The following examples illustrate how one can verify the conditions for convexity of the impulse-response identified set using Proposition 3.

Example 1 *Consider the partial causal ordering case considered in Example 1 in Section 4. If the object of interest is an impulse response to the monetary policy shock* $\epsilon_{i,t}$*, we order the variables as* $(i_t, m_t, \pi_t, y_t)'$ *and have* $(f_1, f_2, f_3, f_4) = (2, 2, 0, 0)$ *with* $j^* = 1$*. Since* $f_1 = 2$*, condition (i) of Proposition 3 guarantees that the impulse-response identified set is* ϕ*-a.s. convex. If the object of interest is an impulse response to a demand shock* $\epsilon_{y,t}$*, we order the variables as* (i_t, m_t, y_t, π_t)*, and* $j^* = 3$*. None of the conditions of Proposition 3 apply in this case, so convexity of the impulse-response identified set is not guaranteed.*

[38] A special situation where the rank conditions (B.1) are guaranteed at almost every ϕ is when σ^i is linearly independent of the row vectors in $F_i(\phi)$ for all $i = 1, \ldots, n$, and the row vectors of $F_i(\phi)$ are spanned by the row vectors of $F_{i-1}(\phi)$ for all $i = 2, \ldots, i^*$. This condition holds in the recursive identification scheme, which imposes a triangularity restriction on A_0^{-1}. See Example 2 in Appendix B.

Example 2 *Consider adding to Example 1 in Section 4 a long-run money neutrality restriction, which sets the long-run impulse response of output to a monetary policy shock* ($\epsilon_{i,t}$) *to zero. This adds a zero restriction on the* $(2,4)$*-th element of the long-run cumulative impulse response matrix* CIR^{∞} *and implies one more restriction on* q^i. *We can order the variables as* $(i_t, m_t, \pi_t, y_t)'$ *and we have* $(f_1, f_2, f_3, f_4) = (3, 2, 0, 0)$. *It can be shown that in this case the first two columns* $[q_1, q_2]$ *are exactly identified,*[39] *which implies that the impulse responses to* $\epsilon_{i,t}$ *and* $\epsilon_{m,t}$ *are point-identified. The impulse responses to* $\epsilon_{y,t}$ *are instead set-identified and their identified sets are convex, as condition (iii) of Proposition 3 applies to* $(i_t, m_t, y_t, \pi_t)'$ *with* $j^* = 3$.

The next corollary presents a formal result to establish whether the addition of identifying restrictions tightens the identified set.

Corollary 1 *Let a set of zero restrictions, an ordering of variables* $(1, \ldots, j^*, \ldots, n)$*, and the corresponding number of zero restrictions* $(f_1, \ldots, f_n)$ *satisfy* $f_i \leq n - i$ *for all* i, $f_1 \geq \cdots \geq f_n \geq 0$*, and* $f_{j^*-1} > f_{j^*}$*, as in Definition 3. Consider imposing additional zero restrictions. Let* $\pi(\cdot) : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *be a permutation that reorders the variables to be consistent with Definition 3 after adding the new restrictions, and let* $\left(\tilde{f}_{\pi(1)}, \ldots, \tilde{f}_{\pi(n)}\right)$ *be the new number of restrictions. If* $\tilde{f}_{\pi(i)} \leq n - \pi(i)$ *for all* $i = 1, \ldots, n$, $(\pi(1), \ldots, \pi(j^*)) = (1, \ldots, j^*)$*, and* $(f_1, \ldots, f_{j^*}) = \left(\tilde{f}_1, \ldots, \tilde{f}_{j^*}\right)$*, i.e., adding the zero restrictions does not change the order of the variables and the number of restrictions for the first* j^* *variables, then the additional restrictions do not tighten the identified sets for the impulse response to the* j^**-th shock for every* $\phi \in \Phi$.

Proof. The successive construction of the feasible column vectors q_i, $i = 1, \ldots, n$, shows that the additional zero restrictions that do not change the order of variables and the zero restrictions for those preceding j^* do not constrain the set of feasible q_{j^*}'s. ■

Example 3 *Consider adding to Example 1 in Section 4 the restriction* $a^{12} = 0$. *Then, an ordering of the variables when the objects of interest are the impulse responses to* $\epsilon_{i,t}$ *is* $(i_t, m_t, y_t, \pi_t)'$ *with* $j^* = 1$ *and* $(f_1, f_2, f_3, f_4) = (2, 2, 1, 0)$. *Compared to Example 1 in Section 4, imposing* $a^{12} = 0$ *does not change* j^*. *Corollary 1 then implies that the restriction does not bring any additional identifying information for the impulse responses.*

The next corollary shows invariance of the identified sets when relaxing the zero restrictions, which partially overlaps with the implications of Corollary 1.

Corollary 2 *Let a set of zero restrictions, an ordering of variables* $(1, \ldots, j^*, \ldots, n)$*, and the corresponding number of zero restrictions* $(f_1, \ldots, f_n)$ *satisfy* $f_i \leq n - i$ *for all* i, $f_1 \geq \cdots \geq f_n \geq 0$,

[39] In this case $F_2(\phi)$ is a submatrix of $F_1(\phi)$, which implies that the vector space spanned by the rows of $F_1(\phi)$ contains the vector space spanned by the rows of $F_2(\phi)$ for every $\phi \in \Phi$. Hence, the rank condition for exact identification (B.1) holds,

and $f_{j^*-1} > f_{j^*}$, as in Definition 3. Under any of the conditions (i) - (iii) of Proposition 3, the identified set for the impulse responses to the j^*-th structural shock does not change when relaxing any or all of the zero restrictions on $q_{j^*+1}, \ldots, q_{n-1}$. Furthermore, if condition (ii) of Proposition 3 is satisfied, the identified set for the impulse responses to the j^*-th structural shock does not change when relaxing any or all of the zero restrictions on $q_1, \ldots, q_{j^*-1}$. When condition (iii) of Proposition 3 is satisfied, the identified set for the impulse responses to the j^*-th shock does not change when relaxing any or all of the zero restrictions on $q_{i^*+1}, \ldots, q_{j^*-1}$.

Proof. Dropping the zero restrictions imposed for those following the j^*-th variable does not change the order of variables nor the construction of the set of feasible q_{j^*}'s. Under condition (ii) of Proposition 3, Lemma A.1 in Appendix A shows that the set of feasible q_{j^*}'s does not depend on any of $F_i(\phi)$, $i = 1, \ldots, (j^* - 1)$. Hence, removing or altering them (as long as condition (ii) of Proposition 3 holds) does not affect the set of feasible q_{j^*}'s. Under condition (iii) of Proposition 3, Lemma A.6 shows that the set of feasible q_{j^*}'s does not depend on any $F_i(\phi)$, $i = (i^* + 1), \ldots, (j^* - 1)$. Hence, relaxing the zero restrictions constraining $[q_{i^*+1}, \ldots, q_{j^*-1}]$ does not affect the set of feasible q_{j^*}'s. ■

Example 4 *Consider relaxing one of the zero restrictions in (4.13),*

$$\begin{pmatrix} u_{\pi,t} \\ u_{y,t} \\ u_{m,t} \\ u_{i,t} \end{pmatrix} = \begin{pmatrix} a^{11} & a^{12} & 0 & 0 \\ a^{21} & a^{22} & 0 & a^{24} \\ a^{31} & a^{32} & a^{33} & a^{34} \\ a^{41} & a^{42} & a^{43} & a^{44} \end{pmatrix} \begin{pmatrix} \epsilon_{\pi,t} \\ \epsilon_{y,t} \\ \epsilon_{m,t} \\ \epsilon_{i,t} \end{pmatrix},$$

where the $(2,4)$-th element of A_0^{-1} is now unconstrained, i.e., the aggregate demand equation is allowed to respond contemporaneously to the monetary policy shock. If the interest is on the impulse responses to the monetary policy shock $\epsilon_{i,t}$, the variables can be ordered as $(m_t, i_t, \pi_t, y_t)'$ with $j^* = 2$. Condition (ii) of Proposition 3 is satisfied and the impulse-response identified sets are convex. In fact, Lemma A.1 in Appendix A implies that in situations where condition (ii) of Proposition 3 applies, the zero restrictions imposed on the preceding shocks to the j^*-th structural shocks do not tighten the identified sets for the j^*-th shock impulse responses compared to the case with no zero restrictions. In the current context, this means that dropping the two zero restrictions on q_m does not change the identified sets for the impulse responses to $\epsilon_{i,t}$.

If sign restrictions are imposed on impulse responses to some structural shock other than the j^*-th shock, the identified set for an impulse response can become non-convex, as we show in the next example.[40]

[40]See also the example in Section 4.4 of Rubio-Ramírez et al. (2010), where $n = 3$ and the zero restrictions satisfy

Example 5 *Consider an SVAR(0) model,*

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = A_0^{-1} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}.$$

Let $\Sigma_{tr} = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$*, where* $\sigma_{11} \geq 0$ *and* $\sigma_{22} \geq 0$*. Positive semidefiniteness of* $\Sigma = \Sigma_{tr}\Sigma_{tr}'$ *requires* $\sigma_{22} \geq 1$*, while* σ_{21} *is left unconstrained. Denoting an orthonormal matrix by* $Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}$*, we can express the contemporaneous impulse-response matrix as*

$$IR^0 = \begin{pmatrix} \sigma_{11}q_{11} & \sigma_{11}q_{12} \\ \sigma_{21}q_{11} + \sigma_{22}q_{21} & \sigma_{21}q_{12} + \sigma_{22}q_{22} \end{pmatrix}.$$

Consider restricting the sign of the $(1,2)$*-th element of* IR^0 *to being positive,* $\sigma_{11}q_{12} \geq 0$*. Since* $\Sigma_{tr}^{-1} = (\sigma_{11}\sigma_{22})^{-1} \begin{pmatrix} \sigma_{22} & 0 \\ -\sigma_{21} & \sigma_{11} \end{pmatrix}$*, the sign normalization restrictions give* $\sigma_{22}q_{11} - \sigma_{21}q_{21} \geq 0$ *and* $\sigma_{11}q_{22} \geq 0$*. We now show that the identified set for the* $(1,1)$*-th element of* IR^0 *is non-convex for a set of* Σ *with a positive measure. Note first that the second column vector of* Q *is constrained to* $\{q_{12} \geq 0, q_{22} \geq 0\}$*, so that the set of* $(q_{11}, q_{21})'$ *orthogonal to* $(q_{12}, q_{22})'$ *is constrained to*

$$\{q_{11} \geq 0, q_{21} \leq 0\} \cup \{q_{11} \leq 0, q_{21} \geq 0\}.$$

When $\sigma_{21} < 0$*, intersecting this union set with the half-space defined by the first sign normalization restriction* $\{\sigma_{22}q_{11} - \sigma_{21}q_{21} \geq 0\}$ *yields two disconnected arcs,*

$$\left\{ \begin{pmatrix} q_{11} \\ q_{21} \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} : \theta \in \left(\left[\frac{1}{2}\pi, \frac{1}{2}\pi + \psi\right] \cup \left[\frac{3}{2}\pi + \psi, 2\pi\right] \right) \right\},$$

where $\psi = \arccos\left(\frac{\sigma_{22}}{\sqrt{\sigma_{22}^2 + \sigma_{21}^2}}\right) \in \left[0, \frac{1}{2}\pi\right]$*. Accordingly, the identified set for* $r = \sigma_{11}q_{11}$ *is given by the union of two disconnected intervals*

$$\left[\sigma_{11}\cos\left(\frac{1}{2}\pi + \psi\right), 0\right] \cup \left[\sigma_{11}\cos\left(\frac{3}{2}\pi + \psi\right), \sigma_{11}\right].$$

Since $\{\sigma_{21} < 0\}$ *has a positive measure in the space of* Σ*, the identified set is non-convex with a positive measure.*

$f_1 = f_2 = f_3 = 1$. Their paper shows that the identified set for an impulse response consists of two distinct points. If we interpret the zero restrictions on the second and third variables as pairs of linear inequality restrictions for q_2 and q_3 with opposite signs, convexity of the impulse-response identified set fails. In this example, the assumption that sign restrictions are only placed on q_j fails.

References

ANDREWS, D. AND P. GUGGENBERGER (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

ANDREWS, D. AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

ARIAS, J., J. RUBIO-RAMÍREZ, AND D. WAGGONER (2018): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685–720.

ARUOBA, S. AND F. SCHORFHEIDE (2011): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, 3, 60–90.

AUMANN, R. (1965): "Integrals of Set-Valued Functions," *Journal of Mathematical Analysis and Applications*, 12, 1–12.

BARANKIN, E. (1960): "Sufficient Parameters: Solution of the Minimal Dimensionality Problem," *Annals of the Institute of Statistical Mathematics*, 12, 91–118.

BAUMEISTER, C. AND J. HAMILTON (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2012): "Partial Identification Using Random Set Theory," *Journal of Econometrics*, 166, 17–32.

BERESTEANU, A. AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

BERGER, J. (1985): *Statistical Decision Theory and Bayesian Analysis*, New York, NY: Springer-Verlag, 2nd ed.

——— (1994): "The Robust Bayesian Viewpoint," in *Robustness of Bayesian Analysis*, ed. by J. Kadane, Amsterdam, Netherlands: North-Holland.

BERGER, J. AND L. BERLINER (1986): "Robust Bayes and Empirical Bayes Analysis with ϵ-contaminated Priors," *The Annals of Statistics*, 14, 461–486.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

BORDER, K. (2013): "Introduction to Correspondences," *unpublished manuscript.*

Canay, I. and A. Shaikh (2017): "Practical and Theoretical Advances for Inference in Partially Identified Models," in *Advances in Economics and Econometrics: Volume 2: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, Econometric Society Monographs, 271–306.

Canova, F. and G. D. Nicolo (2002): "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1121–1159.

Chamberlain, G. and E. Leamer (1976): "Matrix Weighted Averages and Posterior Bounds,," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 73–84.

Chen, X., T. Christensen, and E. Tamer (forthcoming): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*.

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, Elsevier, vol. 1, Part A, 65–148.

Dawid, A. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41, 1–31.

DeRobertis, L. and J. Hartigan (1981): "Bayesian Inference Using Intervals of Measures," *The Annals of Statistics*, 9, 235–244.

Drèze, J. (1974): "Bayesian Theory of Identification in Simultaneous Equations Models," in *Studies in Bayesian Econometrics and Statistics, In Honor of Leonard J. Savage*, ed. by S. Fienberg and A. Zellner, Amsterdam, Netherlands: North-Holland.

Epstein, L. and K. Seo (2014): "Bayesian Inference and Non-Bayesian Prediction and Choice: Foundations and an Application to Entry Games with Multiple Equilibria," *unpublished manuscript*.

Faust, J. (1998): "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Conference Series on Public Policy*, 48, 207–244.

Florens, J. and M. Mouchart (1977): "Reduction of Bayesian Experiments," CORE Discussion Paper 7737, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

Florens, J., M. Mouchart, and J. Rolin (1990): *Elements of Bayesian Statistics*, New York, NY: Marcel Dekker.

Gafarov, B., M. Meier, and J. Montiel-Olea (2018): "Delta-method Inference for a Class of Set-identified SVARs," *Journal of Econometrics*, 203, 316–327.

Galichon, A. and M. Henry (2009): "A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 152, 186–196.

Giacomini, R., T. Kitagawa, and A. Volpicella (2018): "Uncertain Identification," *Cemmap working paper*.

Granziera, E., H. Moon, and F. Schorfheide (2018): "Inference for VARs Identified with Sign Restrictions," *Quantitative Economics*.

Gustafson, P. (2015): *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*, Chapman and Hall/CRC.

Horowitz, J. and C. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84.

Huber, P. (1973): "The Use of Choquet Capacities in Statistics," *Bulletin of the International Statistical Institute*, 45, 181–191.

Hurwicz, L. (1950): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models, Cowles Commision Monograph 10*, New York, NY: John Wiley and Sons.

Imbens, G. and C. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Inoue, A. and L. Kilian (2013): "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 177, 1–13.

Kadane, J. (1974): "The Role of Identification in Bayesian Theory," in *Studies in Bayesian Econometrics and Statistics, In Honor of Leonard J. Savage*, ed. by S. Fienberg and A. Zellner, Amsterdam, Netherlands: North-Holland.

Kaido, H. (2016): "A Dual Approach to Inference for Partially Identified Econometric Models," *Journal of Econometrics*, 192, 269–290.

Kaido, H. and A. Santos (2014): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, 82, 387–413.

Kallenberg, O. (2001): *Foundations of Modern Probability*, New York, NY: Springer-Verlag, 2nd ed.

Kitagawa, T., J. Montiel-Olea, and J. Payne (2017): "Posterior Distributions of Nondifferentiable Functions," *Cemmap working paper*.

Kline, B. and E. Tamer (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7, 329–366.

Koopmans, T. and R. Reiersol (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21, 165–181.

Leamer, E. (1982): "Sets of Posterior Means with Bounded Variance Priors," *Econometrica*, 50, 725–736.

Liao, Y. and W. Jiang (2010): "Bayesian Analysis in Moment Inequality Models," *Annals of Statistics*, 38, 275–316.

Manski, C. (2007): *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press.

——— (2013): *Public Policy in an Uncertain World*, Cambridge, MA: Harvard University Press.

Molchanov, I. (2005): *Theory of Random Sets*, London, England: Springer-Verlag.

Molchanov, I. and F. Molinari (2018): *Random Sets in Econometrics*, Econometric Society Monographs, Cambridge University Press.

Moon, H. and F. Schorfheide (2011): "Bayesian and Frequentist Inference in Partially Identified Models," *NBER working paper*.

——— (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

Mountford, A. and H. Uhlig (2009): "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics*, 24, 960–992.

Norets, A. and X. Tang (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, 81, 1229–1262.

Ok, E. (2007): *Real Analysis with Economic Applications*, Princeton, NJ: Princeton University Press.

Picci, G. (1977): "Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem," *SIAM Journal on Applied Mathematics*, 33, 383–398.

Poirier, D. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14, 483–509.

Rockafellar, R. (1970): *Convex Analysis*, Princeton, NJ: Princeton University Press.

Romano, J. and A. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

Rothenberg, T. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Rubio-Ramírez, J., D. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Schervish, M. (1995): *Theory of Statistics*, New York, NY: Springer-Verlag.

Sims, C. and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67, 1113–1155.

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

Sundaram, R. K. (1996): *A First Course in Optimization Theory*, Cambridge, MA: Cambridge University Press.

Sutherland, W. A. (2009): *Introduction to Metric and Topological Spaces*, Oxford University Press, second ed.

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.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge, MA: Cambridge University Press.

Wan, Y. (2013): "An Integration-based Approach to Moment Inequality Models," *unpublished manuscript.*

Wasserman, L. (1989): "A Robust Bayesian Interpretation of Likelihood Regions," *The Annals of Statistics*, 17, 1387–1393.

——— (1990): "Prior Envelopes Based on Belief Functions," *The Annals of Statistics*, 18, 454–464.

Wasserman, L. and J. Kadane (1990): "Bayes' Theorem for Choquet Capacities," *The Annals of Statistics*, 18, 1328–1339.