Robust Bayesian Inference for Set-Identified Models [*]

Raffaella Giacomini[†] and Toru Kitagawa[‡]

This draft: 7 September 2020

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 and show that they 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 for the structural parameter given the reduced-form parameter. 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 provides 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, three anonymous referees 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.

In practice, we propose summarizing the information in the class of posteriors by reporting the 'posterior lower and upper probabilities' of an event and/or the 'set of posterior means (or quantiles)' in the class of posteriors and the associated 'robust credible region'. These outputs can be expressed in terms of the (single) posterior of the reduced-form parameter, so they can be obtained numerically if one can draw the reduced-form parameter randomly from its 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 would be 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), 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 that transforms reduced-form shocks into structural shocks[2]. Baumeister and Hamilton (2015) 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 our method in the context of SVARs. The conditions are mild or easy to verify, and cover a wide range of applications. In particular, the results on the types of 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 adds an optimization step to the algorithms used in the literature, such as those of Uhlig (2005) and Arias et al. (2018).

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 standard Bayesian output. Reporting the outputs from both approaches, together with the diagnostic tools, can help one separate the information contained in the data and in the 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 equality and sign restrictions typically used in the literature. The findings show that all 90% robust credible regions contain zero, casting doubts on the informativeness of such restrictions. In particular, sign restrictions alone have little identifying power, which means that standard Bayesian inference is largely driven by the choice of the unrevisable prior for the rotation matrix. The addition of zero restrictions tightens the estimated identified set, 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 has a long history in statistics. See Berger (1994) and references

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

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

therein. 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. See Canay and Shaikh (2017) for a survey of the literature and references therein. In the context of certain sign-restricted SVARs, Granziera et al. (2018) propose frequentist inference by inverting tests for a minimum-distance type criterion function, while Gafarov et al. (2018) apply the delta-method using directional derivatives. Our approach is complementary, as it accommodates a broader class of identifying restrictions, while relying on different conditions to attain asymptotic frequentist validity.

There is also a growing literature on Bayesian inference for set-identified models. Some propose posterior inference based on a single prior irrespective of the posterior sensitivity introduced by set identification (Baumeister and Hamilton (2015); Gustafson (2015)). Our paper does not intend to provide a normative argument as to whether one should adopt a single prior or multiple priors under set identification: our main goal is to offer new tools for inference and to show that they 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 dynamic discrete choice models. Kline and Tamer (2016) focus on moment inequality models and construct confidence regions that share some relation with our robust credible regions. Their proposal does not have a formal Bayesian interpretation, and their credible sets do not minimize volume, so the coverage can be conservative. If one were to extend our framework to moment inequality models, one could view our approach as providing a formal posterior interpretation for the confidence regions in Kline and Tamer (2016), while ensuring that they have minimum volume. In addition, we show how to construct the set of posterior means as a consistent estimator of the true identified set. Wan (2013) and Chen et al. (2018) propose using Bayesian Markov Chain Monte Carlo methods to overcome some computational challenges of the frequentist approach to inference about the identified set.

Finally, a key insight of this paper is to recognize that from the Bayesian perspective, under mild regularity conditions, an identification region can be viewed as a random closed set, and Bayesian inference on it can be carried out using elements of random set theory. This theory has proven very helpful for partial identification analysis since its introduction to econometrics (Beresteanu and Molinari (2008); Beresteanu et al. (2012)), and a novel contribution of our paper is to bring these tools into Bayesian inference.

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. Sections 4 and 5 are self-contained, so a reader interested in SVARs can focus on these sections. Section 6 contains the empirical application and Section 7 concludes. The proofs are in Appendix A. An online supplemental appendix (Appendix B in Giacomini and Kitagawa (2020)) contains additional results and discussion about the validity of the assumptions in SVARs.

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. We restrict attention to parametric models, so $\Theta \subset \mathcal{R}^d$, $d < \infty$. 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)). 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 $y \in \mathbf{Y}$ (see, e.g., Barankin (1960)). 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 (Koopmans and Reiersol (1950)), $\phi = g(\theta)$ is the reduced-form parameter that indexes the distribution of the 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 $p(y|\theta) = \hat{p}(y|g(\theta))$ for every $y \in \mathbf{Y}$ and $\theta \in \Theta$.[3]

Let the parameter of interest $\eta \in \mathcal{H}$ be a 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.*

[3] In Bayesian statistics, $\phi = g(\theta)$ is called the (minimal) sufficient parameters that satisfy conditional independence $Y \perp \theta|\phi$ (Barankin (1960).

(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 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) of θ and π_ϕ be the corresponding prior of 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

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

where $\pi_{\theta|\phi}$ denotes the conditional distribution of θ given ϕ,[4] 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*.

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 induces a conditional prior $\pi_{\theta|\phi}$ that is unique up to π_ϕ-almost sure equivalence. 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,

[4] To avoid the Borel paradox, we follow the definition of conditional distribution $\pi_{\theta|\phi}$ via the conditional expectation, i.e., $\pi_{\theta|\phi}(A) \equiv E(1_A(\theta)|\phi)$, where $E(\cdot|\phi)$ is the $\mathcal{B}$-measurable, π_ϕ-integrable function such that for any $B \in \mathcal{B}$, $E(\xi(\theta) \cdot 1_B(\theta)) = \int_B E(\xi(\theta)|\phi) d\pi_\phi$. The regular conditional distribution $\pi_{\theta|\phi}$ is unique up to a π_ϕ-null set of ϕ.

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 specified 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.

In this paper we shall not discuss how to select π_ϕ, and treat it as given. As the influence of this prior on posterior inference disappears asymptotically, any sensitivity issues in this respect potentially only concern small samples. Another reason for not introducing multiple priors for ϕ is to avoid possible issues of non-convergence of the class of posteriors, as discussed in the literature on global sensitivity analysis, e.g. Ruggeri and Sivaganesan (2000).

2.3 Posterior Lower and Upper Probabilities

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

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

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.

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

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 allows an improper prior with support larger than $g(\Theta)$ and imposes the assumption 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 measurability requirement. The closedness of $IS_\theta(\phi)$ and $IS_\eta(\phi)$ is implied, e.g., by continuity of $g(\cdot)$ and $h(\cdot)$.

The next theorem expresses the posterior lower and upper probabilities for the parameter of interest in terms of the posterior of ϕ. This provides the basis for the numerical approximation of these 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 $D \in \mathcal{D}$,*

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

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 D in terms of the posterior probability of ϕ. The intuition for this result can be best understood when $\eta = \theta$. The decomposition in equation (2.2) suggests that, to minimize the posterior probability on $\{\theta \in D\}$, we choose, if possible, a conditional prior $\pi_{\theta|\phi}$ that assigns all the probability outside D so as to attain $\pi_{\theta|\phi}(D) = 0$. Such choice of prior is, however, not possible for ϕ such that $IS_\theta(\phi) \subset D$, since the requirement $\pi_{\theta|\phi}(IS_\theta(\phi)) = 1$ binds and any choice of prior satisfies $\pi_{\theta|\phi}(D) = 1$. Symmetrically, the posterior probability on $\{\theta \in D\}$ is maximized by choosing, if possible (i.e., if $IS_\theta(\phi) \cap D \neq \emptyset$), a conditional prior that puts all the probability inside D. These constructions of the extreme conditional priors immediately lead to the expressions of the lower and upper probabilities in Theorem 1.[5]

[5]Viewing the identified-set correspondence $IS_\theta(\phi)$ as a random closed set defined on the probability space $(\Phi, \mathcal{B}, \pi_{\phi|X})$, let $\xi : \Phi \to \Theta$ be a measurable selection of $IS_\theta(\phi)$, i.e., $\xi(\phi)$ is $\mathcal{B}$-measurable and satisfies $\pi_{\phi|X}(\{\xi(\phi) \in IS_\theta(\phi)\}) = 1$. Artstein's inequality (Theorem 2.1 in Artstein (1983)) shows that the set of probability distributions for θ formed by measurable selections of $IS_\theta(\phi)$ is given by

$$\Pi^S_{\theta|Y} \equiv \left\{\pi_{\theta|Y} : \pi_{\theta|Y}(A) \leq \pi_{\phi|Y}(IS_\theta(\phi) \cap A \neq \emptyset), \ \ \forall A \in \mathcal{A}, \text{closed}\right\}.$$

Since a measurable selection of $IS_\theta(\phi)$ corresponds to degenerate conditional priors $\{\pi_{\theta|\phi} = 1_{\xi(\phi)} : \phi \in \Phi\}$ and it is included in the set of conditional priors used in our analysis, the set of posteriors for θ, $\Pi_{\theta|Y}$, defined in (2.4) with

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. Corollary A.1 in Appendix A shows that 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 the investigation of its analytical properties and the practical implementation of the method.

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 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 the 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

$\eta = \theta$ satisfies $\Pi_{\theta|Y} \supseteq \Pi^S_{\theta|Y}$. As implied by Artstein's inequality, however, the upper probability of $\Pi_{\theta|Y}$ obtained in Theorem 1 with $\eta = \theta$ agrees with the upper probability of $\Pi^S_{\theta|Y}$ on any closed measurable $A \in \mathcal{A}$.

[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)) $E^A_{\phi|Y}(X) \equiv \left\{E_{\phi|Y}(\xi) : \xi \in S^1(X)\right\}$.

$co(IS_\eta(\phi))$ 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 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 $IS_\eta(\phi)$. This lower (upper) bound is attained by the conditional priors $\{\pi_{\theta|\phi} : \phi \in \Phi\}$ that allocate probability one to θ attaining the lower (upper) bound of $IS_\eta(\phi)$. 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 sentence yields the usual interpretation of a 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 to resolve this multiplicity by choosing the C_α with the smallest volume:

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

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 as a *smallest robust credible region* with credibility α. The credible regions for the identified set proposed in Moon and Schorfheide (2011), Norets and Tang (2014), and Kline

[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.

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).[9]

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*

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

For each $\eta_c \in \mathcal{H}$, let $r_\alpha(\eta_c)$ be the α-th quantile of the distribution of $d\left(\eta_c, IS_\eta(\phi)\right)$ 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 in (2.7), with $\mathcal{C}$ restricted to the class of closed connected intervals, is a closed interval centered at $\eta^*_c = \arg\min_{\eta_c \in \mathcal{H}} r_\alpha(\eta_c)$ with radius $r^*_\alpha = r_\alpha(\eta^*_c)$.

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, $\tilde{\pi}_{\phi|Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$.[10] 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)).

[9] 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.

[10] 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.

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 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. [11]

The amount of information in the posterior provided by the choice of the unrevisable prior $\pi_{\theta|\phi}$ in a standard Bayesian analysis can be similarly measured by comparing the width of C_α satisfying (2.6) to the width of the standard Bayesian credible region obtained from the single prior:

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

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 quantities are useful from the frequentist perspective, we now analyze 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 our 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 $p(Y^T|\phi_0)$.

[11] The measure is meaningful only when the identified sets in both models are bounded. In this case the measure lies in the unit interval because $IS_\eta(\phi, M_s) \subseteq IS_\eta(\phi, M_l)$ for all ϕ.

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 correspondence $IS_\eta : \Phi \rightrightarrows H$ is continuous at $\phi = \phi_0$(see, e.g., Sundaram (1996) for the definition of continuity for correspondences).*

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

(iii) There exists $\delta > 0$ such that $IS_\eta(\phi)$ is $L_{1+\delta}$-integrable with respect to $\pi_{\phi|Y^T}$;

$$E_{\phi|Y^T}\left(\sup_{\eta \in IS_\eta(\phi)} \|\eta\|^{1+\delta}\right) < \infty, \quad p\left(Y^T|\phi_0\right)\text{-a.s.,} \tag{3.1}$$

for all large enough T.

Assumption 2(i) requires that the identified set of η is a continuous correspondence at 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. Since $\ell(\phi)$ and $u(\phi)$ can be viewed as the values of the optimizations $\ell(\phi) = \min_{\theta \in IS_\theta(\phi)} h(\theta)$ and $u(\phi) = \max_{\theta \in IS_\theta(\phi)} h(\theta)$, the theorem of maximum (e.g., Theorem 9.14 in Sundaram (1996)) shows that sufficient conditions for continuity of $\ell(\phi)$ and $u(\phi)$ are continuity of $h(\cdot)$ and continuity of the correspondence for $IS_\theta(\phi)$. In a common special case where $IS_\theta(\phi)$ is a polyhedron and $[\ell(\phi), u(\phi)]$ are the values of linear programming, continuity of the polyhedral correspondence is implied by the condition of dimension-stability of the polyhedron (e.g., Proposition 6 in Wets (1985)). For SVAR models, Appendix B.2 in Giacomini and Kitagawa (2020) show that the continuity property is mild and easily verifiable.

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 $(1+\delta)$-th moment. In the scalar case, Assumption 2(iii) holds with $\delta = 1$ if $\ell(\phi)$ and $u(\phi)$ have finite posterior variances.

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

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

(ii) Suppose Assumption 2 holds and the prior for ϕ, π_ϕ, is non-atomic, then 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 $IS_\eta(\phi_0)$

[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 $p\left(Y^\infty|\phi_0\right)$. For a finite-dimensional ϕ, posterior consistency is implied by higher-level conditions for the likelihood of ϕ. See Section 7.4 of Schervish (1995) for details.

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 $\mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ be the set of continuous functions from the k-dimensional unit sphere $\mathcal{S}^{k-1}$ to $\mathcal{R}$, and let $\hat{\phi}$ denote the maximum likelihood estimator of ϕ. For a sequence $a_T \to \infty$ as $T \to \infty$, define stochastic processes in $\mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$ indexed by $q \in \mathcal{S}^{k-1}$,*

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

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

(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 $\mathcal{C}\left(\mathcal{S}^{k-1}, \mathcal{R}\right)$, $\hat{c}_T(\cdot) \equiv a_T\left[s(C_\alpha, \cdot) - s\left(IS_\eta(\hat{\phi}), \cdot\right)\right]$, 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). 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 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 sample Y^T vanishes as $T \to \infty$. Beresteanu and Molinari (2008) and Kaido and Santos (2014) provide practical examples where the limiting process Z is a zero-mean tight Gaussian process in $\mathcal{C}(\mathcal{S}^{k-1}, \mathcal{R})$.

Assumptions 4(i) and (ii) are delicate assumptions and whether they hold depends on the geometry of the identified set. The working paper version of this paper discusses an example (Example C.2 in Appendix C) showing that, if the support function $s\left(IS_\eta(\phi), q\right)$ is not differentiable in ϕ at some q, this can lead to violation of Assumptions 4(i) and (ii). See also Kitagawa et al. (2020) for properties of the asymptotic posterior for a non-differentiable function of parameters satisfying the Bernstein-von Mises property such as ϕ.

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

[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.

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 shows pointwise asymptotic coverage rather than asymptotic uniform coverage over a class of the sampling processes ϕ_0. As stressed in the frequentist literature (e.g., Stoye (2009); Romano and Shaikh (2010); Andrews and Soares (2010), it is desirable for frequentist methods to attain asymptotically valid coverage in the uniform sense. Examining this property for our procedure is, however, challenging because it would require the Bernstein-von Mises condition for the support function processes (Assumption 4(i) and (ii)) to hold uniformly over a class of the sampling processes ϕ_0. To our knowledge, little is known to what extent the Bernstein-von Mises property holds in the uniform sense even for the standard case of identifiable parameters. We thus leave this investigation for future research.

Third, the confidence region considered by 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, the credible region proposed by these authors 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 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 is rather high-level, and could be difficult to check when η is a vector. For a scalar η, 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 parameter of interest η be a scalar. Denote the convex hull of the identified set by $co\left(IS_\eta(\phi)\right) = \left[\ell(\phi), u(\phi)\right]$.*

(i) The maximum likelihood estimator $\hat{\phi}$ is strongly consistent for ϕ_0, and the posterior of ϕ and the sampling distribution of $\hat{\phi}$ 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.

It is important to note that Assumption 5 (ii) is restrictive in some aspects. First, it does not allow $\ell(\phi)$ or $u(\phi)$ to be flat at ϕ_0. This is because the Bernstein-von Mises property for ϕ (Assumption 5 (i)) does not carry over to $\ell(\phi)$ or $u(\phi)$ through the second-order delta method if their first-order derivatives are zero at ϕ_0, and it leads to violation of Assumptions 4(i) and (ii). Second, non-differentiability of $\ell(\phi)$ and $u(\phi)$ arises if the projection bounds for η involve the max or min operations and the minimizers or maximizers are not unique at $\phi = \phi_0$. Bounds involving the max or min appear in the intersection bound analysis of Manski (1990) and the partial identification analysis via linear programming of Balke and Pearl (1997). As shown in Kitagawa et al. (2020), the Bernstein-von Mises property breaks down for non-differentiable functions of ϕ, and it leads to violation of Assumption 4(i) and (ii). Appendix B.3 in Giacomini and Kitagawa (2020) discusses the differentiability assumptions and their plausibility in the context of SVAR models.

Proposition 2. *Suppose Assumptions 3 and 5 hold. Then Assumptions 4(i) - (iii) hold 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,*

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

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 our method in the context of impulse-response analysis in set-identified SVARs. This section is self-contained. Consider an SVAR(p):

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

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 $y_1, \ldots, y_p$ are given. We assume that one always imposes the sign normalization restrictions that the diagonal elements of

A_0 are nonnegative. 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 $E(u_t u_t') \equiv \Sigma = A_0^{-1}\left(A_0^{-1}\right)'$. The reduced-form parameter is $\phi = (vec(B)', vech(\Sigma)')' \in \Phi \subset \mathcal{R}^{n+n^2p} \times \mathcal{R}^{n(n+1)/2}$, where $B = [b, B_1, \ldots, B_p]$. We restrict the domain Φ to the set of ϕ's such that the variance-covariance matrix of the reduced-form errors is nonsingular and the reduced-form VAR(p) model can be inverted into a VMA(∞) model.

Let $Q \in \mathcal{O}(n)$ be an $n \times n$ orthonormal 'rotation' matrix and $\mathcal{O}(n)$ be the set of $n \times n$ orthonormal matrices. As in Uhlig (2005) and Rubio-Ramírez et al. (2010), consider the transformation

$$B = A_0^{-1}[a, A_1, \ldots, A_p], \quad \Sigma = A_0^{-1}\left(A_0^{-1}\right)', \quad Q = \Sigma_{tr}^{-1}A_0^{-1},$$

where Σ_{tr} is the lower-triangular Cholesky factor of Σ with nonnegative diagonal elements. This transformation is one-to-one, as it is invertible with nonsingular Σ, so that $A_0 = Q'\Sigma_{tr}^{-1}$ and $[a, A_1, \ldots, A_p] = Q'\Sigma_{tr}^{-1}B$. Since θ in Section 2 can be any one-to-one transformation of the structural parameters, in this section we follow the convention in the literature and set $\theta = \left(\phi', vec(Q)'\right)'$.

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

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

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

The VMA(∞) representation of the model is:

$$y_t = c + \sum_{j=0}^{\infty} C_j u_{t-j} = c + \sum_{j=0}^{\infty} C_j \Sigma_{tr} Q \epsilon_{t-j},$$

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, $h = 0, 1, 2, \ldots$

$$IR^h = C_h \Sigma_{tr} Q, \tag{4.4}$$

and the long-run cumulative impulse-response matrix by

$$CIR^{\infty} = \sum_{h=0}^{\infty} IR^h = \left(\sum_{h=0}^{\infty} C_h\right) \Sigma_{tr} Q. \tag{4.5}$$

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

$$\eta = IR_{ij}^h \equiv e_i' C_h \Sigma_{tr} Q e_j \equiv c_{ih}'(\phi)\, q_j = \eta(\phi, Q), \tag{4.6}$$

where e_i is the i-th column vector of 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 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

$$IS_\eta(\phi) = \{\eta(\phi, Q) : Q \in \mathcal{Q}\}. \tag{4.7}$$

In the following we characterize the subspace $\mathcal{Q}$ under common types of 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 cumulative long-run responses CIR^∞ in (4.5).

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

$$\begin{array}{rcccl} ((i,j)\text{-th element of } A_0) & = & 0 & \Longleftrightarrow & \left(\Sigma_{tr}^{-1} e_j\right)' q_i = 0, \\ ((i,j)\text{-th element of } A_l) & = & 0 & \Longleftrightarrow & \left(\Sigma_{tr}^{-1} B_l e_j\right)' q_i = 0, \\ \left((i,j)\text{-th element of } A_0^{-1}\right) & = & 0 & \Longleftrightarrow & \left(e_i' \Sigma_{tr}\right) q_j = 0, \\ ((i,j)\text{-th element of } CIR^\infty) & = & 0 & \Longleftrightarrow & \left[e_i' \sum_{h=0}^{\infty} C_h(B)\, \Sigma_{tr}\right] q_j = 0. \end{array} \tag{4.8}$$

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

$$F(\phi, Q) \equiv \begin{pmatrix} F_1(\phi)\, q_1 \\ F_2(\phi)\, q_2 \\ \vdots \\ F_n(\phi)\, q_n \end{pmatrix} = \mathbf{0}, \tag{4.9}$$

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 (4.8), 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 (4.9)) 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* $f_{j-1} > f_j$.[14]

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

$$f_i \leq n - i \quad \text{for all } i = 1, \ldots, n, \tag{4.10}$$

with strict inequality for at least one $i \in \{1, \ldots, n\}$.[15]

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},

$$\begin{pmatrix} u_{\pi,t} \\ u_{y,t} \\ u_{m,t} \\ u_{i,t} \end{pmatrix} = \begin{pmatrix} a^{11} & a^{12} & 0 & 0 \\ a^{21} & a^{22} & 0 & 0 \\ a^{31} & a^{32} & a^{33} & a^{34} \\ a^{41} & a^{42} & a^{43} & a^{44} \end{pmatrix} \begin{pmatrix} \epsilon_{\pi,t} \\ \epsilon_{y,t} \\ \epsilon_{m,t} \\ \epsilon_{i,t} \end{pmatrix}. \tag{4.11}$$

Let the objects of interest be the impulse responses to $\epsilon_{i,t}$ *(a monetary policy shock). Let* $\left[q^\pi, q^y, q^m, q^i\right]$ *be a* 4×4 *orthonormal matrix. By (4.8), 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 (4.10). 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 ordering in Definition 3. If there are only sign restrictions, we order first the variable whose structural shock is of interest. Suppose there are sign restrictions on the responses to the j-th structural shock. Sign restrictions are linear constraints on the columns of Q:

[14] The assumption 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 convexity in Appendix B in Giacomini and Kitagawa (2020) is not affected by the ordering of the other variables as long as the f_i's are in decreasing order.

[15] 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 (4.10), 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.

$S_{hj}(\phi) q_j \geq \mathbf{0}$,[16] where $S_{hj}(\phi) \equiv D_{hj} C_h(B) \Sigma_{tr}$, with D_{hj} an $s_{hj} \times n$ matrix that selects the sign-restricted responses from the 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

$$S_j(\phi) q_j \geq \mathbf{0}, \tag{4.12}$$

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

Let $\mathcal{I}_S \subset \{1, 2, \ldots, n\}$ be such that $j \in \mathcal{I}_S$ if some of the impulse responses to the j-th structural shock are sign-constrained. We denote the set of all sign restrictions, $S_j(\phi) q_j \geq \mathbf{0}$ for $j \in \mathcal{I}_S$, as

$$S(\phi, Q) \geq \mathbf{0}. \tag{4.13}$$

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

$$IS_\eta(\phi|F, S) = \{\eta(\phi, Q) : Q \in \mathcal{Q}(\phi|F, S)\}, \tag{4.14}$$

where $\mathcal{Q}(\phi|F, S)$ is the set of Q's that jointly satisfy the sign restrictions (4.13), the zero restrictions (4.9) and the sign normalizations (4.3),

$$\mathcal{Q}(\phi|F, S) = \left\{Q \in \mathcal{O}(n) : S(\phi, Q) \geq \mathbf{0},\ F(\phi, Q) = \mathbf{0},\ diag\left(Q' \Sigma_{tr}^{-1}\right) \geq \mathbf{0}\right\}. \tag{4.15}$$

Proposition B.1 in Giacomini and Kitagawa (2020) shows that, unlike the cases with only zero restrictions, with sign restrictions the identified set of η can be empty.

4.4 Multiple Priors in SVARs

Let $\tilde{\pi}_\phi$ be a prior for the reduced-form parameter. We ensure that the prior for ϕ is consistent with Assumption 1(i) by trimming the support of $\tilde{\pi}_\phi$ as

$$\pi_\phi \equiv \frac{\tilde{\pi}_\phi 1\{\mathcal{Q}(\phi|F, S) \neq \emptyset\}}{\tilde{\pi}_\phi(\{\mathcal{Q}(\phi|F, S) \neq \emptyset\})}, \tag{4.16}$$

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

[16] For $y = (y_1, \ldots, y_m)'$, $y \geq \mathbf{0}$ means $y_i \geq 0$ for all i and $y > \mathbf{0}$ means $y_i \geq 0$ for all i and $y_i > 0$ for some i.

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 specifying only π_ϕ cannot yield a single posterior and one would also need to specify a prior $\pi_{Q|\phi}$. 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).[17]

The robust Bayesian procedure in this paper does not require specifying a prior $\pi_{Q|\phi}$, but considers the class of all priors $\pi_{Q|\phi}$ supported on $\mathcal{Q}(\phi|F, S)$,

$$\Pi_{Q|\phi} = \left\{\pi_{Q|\phi} : \pi_{Q|\phi}\left(\mathcal{Q}\left(\phi|F, S\right)\right) = 1,\ \pi_\phi\text{-almost surely}\right\}. \tag{4.17}$$

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

$$\Pi_{\theta|Y} = \left\{\pi_{\theta|Y} = \pi_{Q|\phi}\pi_{\phi|Y} : \pi_{Q|\phi} \in \Pi_{Q|\phi}\right\} \tag{4.18}$$

and a class of posteriors for the impulse response η,

$$\Pi_{\eta|Y} \equiv \left\{\pi_{\eta|Y}(\cdot) = \int \pi_{Q|Y}\left(\eta\left(\phi, Q\right) \in \cdot\right) d\pi_{\phi|Y} : \pi_{Q|Y} \in \Pi_{Q|Y}\right\}. \tag{4.19}$$

4.5 Set of Posterior Means and Robust Credible Region

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

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

where $\ell(\phi) = \inf\left\{\eta\left(\phi, Q\right) : Q \in \mathcal{Q}\left(\phi|F, S\right)\right\}$ and $u\left(\phi\right) = \sup\left\{\eta\left(\phi, Q\right) : Q \in \mathcal{Q}\left(\phi|F, S\right)\right\}$. Section 5 discusses how to compute $\ell(\phi)$ 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 (4.19).

[17] 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 $\pi_{Q|\phi}$ can be equivalently stated as the difficulty of specifying a prior for the structural parameters that is compatible with π_ϕ.

To validate the frequentist interpretation of the set of posterior means, Appendix B in Giacomini and Kitagawa (2020) provides conditions for 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, by Proposition 2, differentiability of the upper and lower bounds of the impulse response identified set in ϕ with non-zero derivatives guarantees that the robust credible region is an asymptotically valid confidence set for the true impulse response identified set.

5 Numerical Implementation

We present an algorithm to numerically approximate the set of posterior means, the robust credible region and the diagnostic tool in Section 2.6.1 for the case of SVARs. The algorithm assumes that the variables are ordered as in Definition 3 and any zero restriction satisfies (4.12). Matlab code implementing the procedure can be obtained from the authors' personal websites or upon request.

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.[18] *Estimate a Bayesian reduced-form VAR to obtain the posterior $\tilde{\pi}_{\phi|Y}$.*

(Step 2). Draw ϕ from $\tilde{\pi}_{\phi|Y}$. 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

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

where $\|\cdot\|$ is the Euclidian metric in $\mathcal{R}^n$. If $(\sigma^i)' \tilde{q}_i$ is zero for some i, set $sign\left((\sigma^i)' \tilde{q}_i\right)$ 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

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

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 obtained in (Step 2), compute the lower and upper bounds of $IS_\eta(\phi|S, F)$ by solving the following constrained nonlinear optimization problem:

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

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 $\min_\eta \hat{z}_\alpha(\eta)$.[19]

(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, e.g, 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. [20]

Second, Step 3 is a non-convex optimization problem and the convergence of gradient-based optimization methods in these problems 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). While they apply the result at $\hat{\phi}$ in a frequentist setting, 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.[21] The advantage of the analytical approach

[19] 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.

[20] The Matlab code we provide also offers the option of using a QR decomposition. In our experience, the two ways of drawing Q are comparable both in terms of the resulting distribution of Q and computational cost.

[21] This involves considering all possible combinations of $(n - 1)$ active restrictions and checking whether any one

is that we can assess the emptiness even when the identified set is narrow, and it is computationally faster. The advantage of the numerical approach is that it is applicable even when the restrictions involve multiple columns of Q (i.e., the restrictions are on multiple structural shocks).

Fourth, if there are concerns about the convergence properties of the numerical optimization step due to 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 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 estimator of the identified set is still consistent 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 optimization in Step 3.

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 considered by Granziera et al. (2018), which in turn is based on Aruoba and Schorfheide (2011). The observables are 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).[22] The data are quarterly from 1965:1 to 2006:1. The model is:

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

of the vectors solving the active restrictions satisfies all the non-active sign restrictions. In practice, we compute the unit-length vectors in the null space of the matrix containing the $(n-1)$ 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 $(n-1)$ active restrictions, we conclude that the identified set is nonempty. See also Giacomini et al. (2020).

[22]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 of Granziera et al. (2018) and Footnote 5 of Aruoba and Schorfheide (2011).

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 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 I coincides with their model.

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
$\tilde{\pi}_{\phi\vert Y}\left(\{\phi : IS_\eta(\phi) \neq \emptyset\}\right)$	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. The approach using Step 6 in Algorithm 1 (with a maximum of 3,000 draws of Q at each draw of ϕ) and the analytical approach described in the remarks after Algorithm 1 produce the same estimates

The prior for the reduced-form parameters, $\tilde{\pi}_\phi$, is the improper Jeffreys' prior, with density proportional to $|\Sigma|^{-\frac{4+1}{2}}$. This implies a normal-inverse-Wishart posterior from which it is easy to draw. The results are based on Algorithm 1, considering five starting values as discussed in the remarks in Section 5.[23] We draw ϕ's until we obtain 1,000 realizations of the nonempty identified set and check for convexity of the set at every draw of ϕ using Proposition B.1 in Appendix B in Giacomini and Kitagawa (2020). 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 differences across models are mainly due to different identifying restrictions.

We also compare our approach to standard Bayesian inference based on a uniform prior for Q. We obtain draws from the single posterior for the impulse-response by iterating Steps (2.1)–(2.3) of Algorithm 1, and retaining only the draws of Q that satisfy the sign restrictions.[24]

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 approaches. 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 approach; for the standard Bayesian approach, they report the posterior mean (dotted line) and the 90% highest posterior density region (dashed line).[25]

We can draw several conclusions. First, choosing a uniform prior for the rotation matrix affects posterior inference: in Model I this prior choice is more informative than the identifying restrictions; in Model III standard Bayesian analysis suggests that the output response is negative for some horizons, whereas the robust Bayesian lower probability of this event is very low, implying that this conclusion is largely driven by the choice of unrevisable prior. See Wolf (2020) for a similar finding.

Second, all 90% robust credible regions contain zero, casting doubts about the informativeness of these common under-identifying restrictions. In particular, sign restrictions alone (Model I) have little identifying power and result in set estimates that are too wide to draw informative inference about the sign of the impulse response. Adding a single zero restriction (Models II to IV) makes the identified set estimates tighter, although the identifying power varies across horizons: The restriction on the contemporaneous response (restriction (ii)) is more informative at short

[23]The results are visually indistinguishable when using the analytical approach discussed in the remarks in Section 5. Five initial values appear sufficient to achieve convergence of the numerical algorithm to the true optimum.

[24]In Models 0 and I, this is equivalent to Uhlig (2005), as it obtains draws from the uniform prior over the space of rotation 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 prior as Arias et al. (2018) gives visually indistinguishable results.

[25]These figures do not capture the dependence of the responses across different horizons.

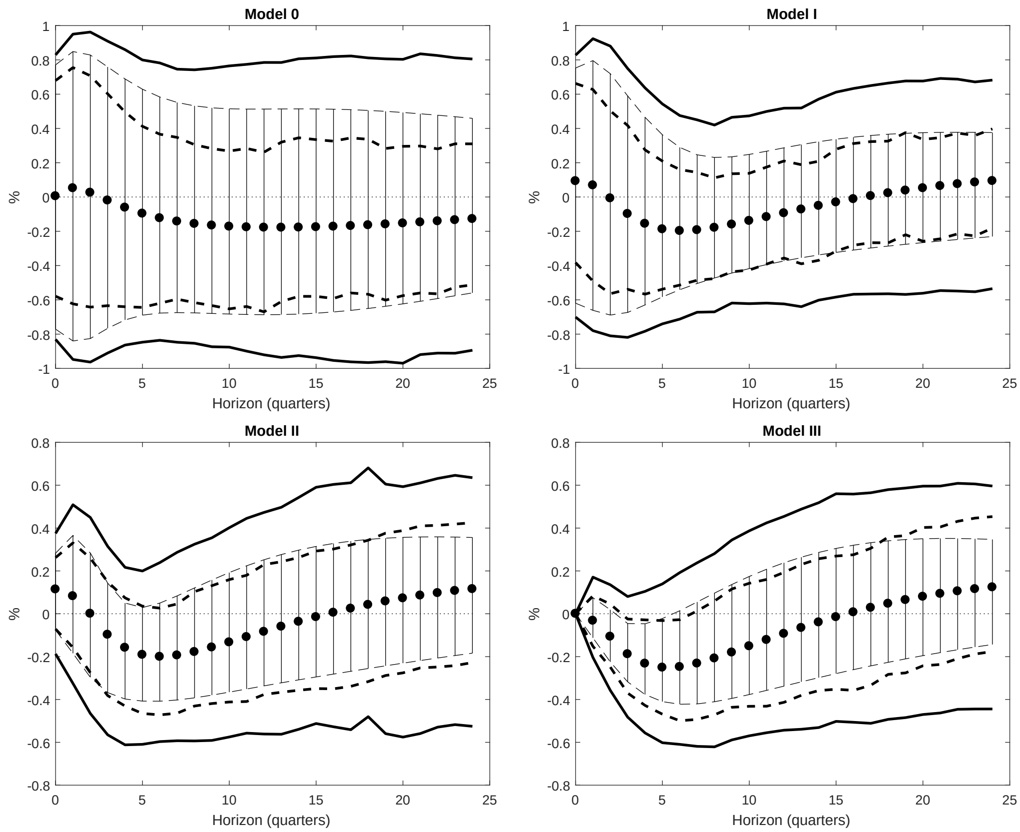
horizons and the long-run restriction (restriction (iii)) is more informative at long horizons. The zero restriction on A_0 (restriction (i)) is informative at both short- and long horizons.

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%

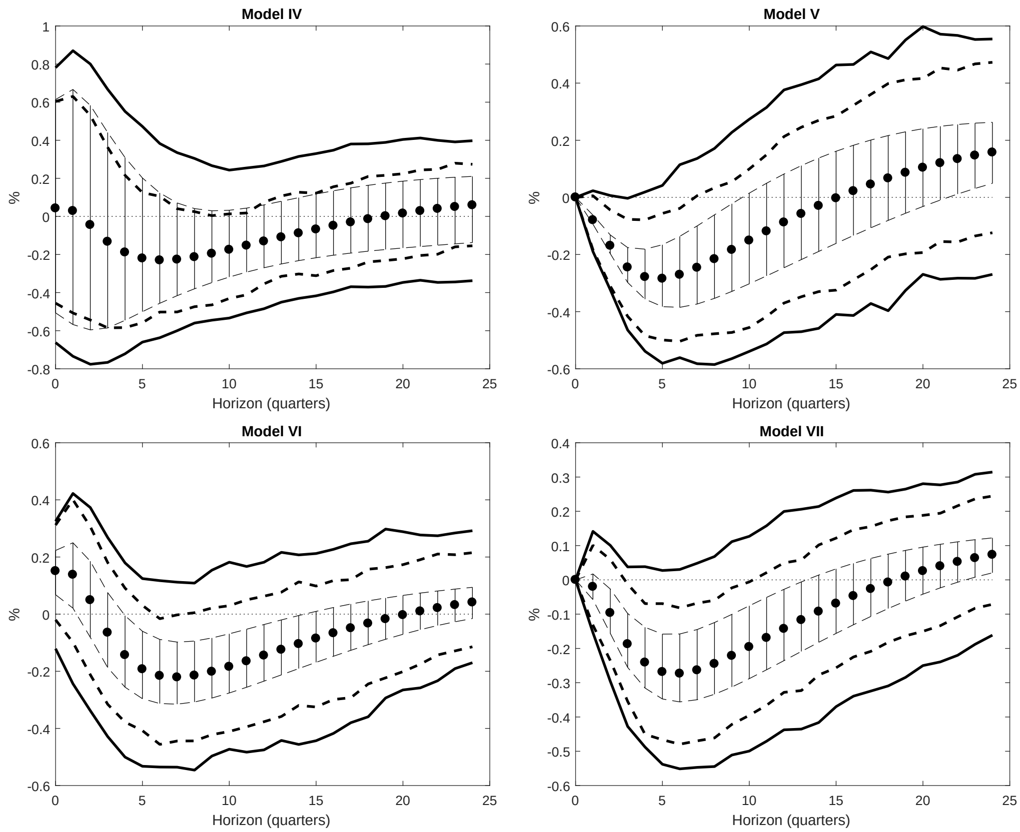
Third, imposing additional zero restrictions (Models V to VII) makes the identifying restrictions more informative than the choice of the prior and reduces the gap between the conclusions of standard- and robust Bayesian analysis. The robust Bayesian analysis also becomes informative for the sign of the output response.

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

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

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



Notes: See Figure 1.

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 and the associated smallest robust credible region that respectively satisfy the properties of consistency and correct frequentist 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 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 which 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 applications to macroeconomic policy analysis using SVARs the set of posterior means and the robust credible region may be too wide to draw informative policy recommendations should not be considered a disadvantage of the method. Wide sets may encourage the researcher to look for additional credible identifying 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.

Appendix

A Lemmas and Proofs

Lemmas A.1 - A.3 are used to prove Theorem 1.

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* $D \in \mathcal{H}$*,*

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

Proof of Lemma A.1. Closedness of $IS_\theta(\phi)$ and $IS_\eta(\phi)$ is implied directly by Assumption 1(ii) and 1(iii). To prove 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 a Polish space, $\{\phi : IS_\theta(\phi) \cap A \neq \emptyset\} \in \mathcal{B}$ 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.

Measurability of $\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\}$ follows since $\{\phi : IS_\eta(\phi) \cap D \neq \emptyset\} = \left\{\phi : IS_\theta(\phi) \cap h^{-1}(D) \neq \emptyset\right\}$ and $h^{-1}(D) \in \mathcal{A}$ (Assumption 1(iii)). The first result 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 of Lemma A.2. 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 $\int_B 1_{\Phi_1^A}(\phi) d\pi_\phi \leq \int_B \pi_{\theta|\phi}(A) d\pi_\phi$ 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 conditional probability. By the construction of Φ_1^A, $IS_\theta(B \cap \Phi_1^A) \subset A$ holds, so $\pi_\theta(A \cap IS_\theta(B \cap \Phi_1^A)) = \pi_\theta(IS_\theta(B \cap \Phi_1^A)) = \pi_\phi(B \cap \Phi_1^A) = \int_B 1_{\Phi_1^A}(\phi) d\pi_\phi$. Thus, the inequality is proven. $\blacksquare$

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 of Lemma A.3. The claim holds trivially when $A = \emptyset$ or $A = \Theta$. We hence prove the claim for A different from $\emptyset$ or Θ. Fix $A \in \mathcal{A}$ and let Φ_1^A be as in the proof of Lemma A.2 and

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

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 $g(\Theta) \subset \Phi$. Consider a Θ-valued measurable selection $\xi^A(\cdot)$ defined on Φ_2^A if nonempty, such that $\xi^A(\phi) \in (IS_\theta(\phi) \cap A^c)$ holds for π_ϕ-almost every $\phi \in \Phi_2^A$. 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 $d(\theta, A) = \inf_{\theta' \in A} \left\|\theta - \theta'\right\|$ 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 $\xi^A(\phi)$). If Φ_2^A is empty, we do not need to construct $\xi^A(\cdot)$ and the construction of the conditional probability distribution $\pi_{\theta|\phi*}^A$ shown below remains valid.

Pick an arbitrary $\pi_{\theta|\phi} \in \Pi_{\theta|\phi}$ and construct another $\pi_{\theta|\phi*}^A$ as, for $\tilde{A} \in \mathcal{A}$,

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

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

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

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

Proof of Theorem 1. We first show the special case of $\eta = \theta$. In the expression of the posterior of θ given in equation (2.2), $\pi_{\theta|Y}(A)$ is minimized over the prior class by plugging in the attainable pointwise lower bound of $\pi_{\theta|\phi}(A)$. 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 upper probability follows by its conjugacy, $\pi^*_{\theta|Y}(A) = 1 - \pi_{\theta|Y*}(A^c)$.

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

$$\pi_{\eta|Y*}(D) = \pi_{\theta|Y*}(h^{-1}(D)) = \pi_{\phi|Y}(\left\{\phi : IS_\theta(\phi) \subset h^{-1}(D)\right\}) \quad = \pi_{\phi|Y}(\{\phi : IS_\eta(\phi) \subset D\}).$$

The expression of the posterior upper probability follows again by the conjugacy property. ■

Corollary A.1 *Under Assumption 1, let $\pi_{\eta|Y*}(D)$ and $\pi^*_{\eta|Y}(D)$, $D \in \mathcal{D}$, be the posterior lower and upper probabilities obtained in Theorem 1. The set of posterior probabilities $\{\pi_{\eta|Y}(D) : \pi_{\eta|Y} \in \Pi_{\eta|Y}\}$ is a connected interval given by $[\pi_{\eta|Y*}(D), \pi^*_{\eta|Y}(D)]$.*

Proof of Corollary A.1. Consider first the case of $\eta = \theta$. Fix $A \in \mathcal{A}$ different from $\emptyset$ and Θ. Using the notation and the argument in the proof of Lemma A.3, it can be shown that the 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 $\pi^{A*}_{\theta|\phi}$ with $\pi^A_{\theta|\phi*}$ constructed in the proof of Lemma A.3, $\pi^\lambda_{\theta|\phi} \equiv \lambda\pi^A_{\theta|\phi*} + (1-\lambda)\pi^{A*}_{\theta|\phi}$, $\lambda \in [0,1]$. 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 $\left[\pi_{\theta|Y*}(A), \pi^*_{\theta|Y}(A)\right]$. The conclusion follows by setting $A = h^{-1}(D)$. ■

Proof of Theorem 2. At each ϕ in the support of $\pi_{\phi|Y}$, 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, $\Pi_{\theta|\phi}$ contains 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. Let $C_r(\eta_c)$ denote the closed interval centered at η_c with radius r. The event $\{IS_\eta(\phi) \subset C_r(\eta_c)\}$ happens if and only if $\{d(\eta_c, IS_\eta(\phi)) \leq r\}$. So, $r_\alpha(\eta_c) \equiv \inf\left\{r : \pi_{\phi|Y}(\{\phi : d(\eta_c, IS_\eta(\phi)) \leq r\}) \geq \alpha\right\}$ 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

$$\pi_{\phi|Y^T}\{\phi : d_H(IS(\phi), IS(\phi_0)) > \epsilon\} = \pi_{\phi|Y^T}(\{\phi : d_H(IS(\phi), IS(\phi_0)) > \epsilon\} \cap G^c) \leq \pi_{\phi|Y^T}(G^c),$$

where the equality 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 the 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

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

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

$\limsup_{T\to\infty} E_{\phi|Y^T}\left[\sup_{\eta\in IS_\eta(\phi)}\|\eta\|^{1+\delta}\right] < \infty$, 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.1) converges to zero $p(Y^\infty|\phi_0)$-a.s. Since ϵ is arbitrary, claim (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 $s(IS_\eta(\phi), q) \le s(C_\alpha, q)$ for all $q \in \mathcal{S}^{k-1}$. Therefore, we have

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

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 $P_{Y^T|\phi_0}(IS_\eta(\phi_0) \subset C_\alpha) = P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \le \hat{c}_T(\cdot)\right)$. Let P_X be the probability law of the limiting stochastic process $X(\cdot)$ of Assumption 4(i) and (ii).

Our aim is to prove the following convergence claims: under Assumption 4,

$$\text{(A)} \quad \left|\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \le \hat{c}_T(\cdot)\right) - P_X(X(\cdot) \le \hat{c}_T(\cdot))\right| \to 0, \text{ as } T \to \infty, p_{Y^\infty|\phi_0}\text{-a.s., and}$$

$$\text{(B)} \quad \left|P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \le \hat{c}_T(\cdot)\right) - P_X(X(\cdot) \le \hat{c}_T(\cdot))\right| \to 0 \text{ in } p_{Y^\infty|\phi_0}\text{-probability as } T \to \infty.$$

Since $\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \le \hat{c}(\cdot)\right) \ge \alpha$, convergence (A) implies $\liminf_{T\to\infty} P_X(X(\cdot) \le \hat{c}_T(\cdot)) \ge \alpha$, $p_{Y^\infty|\phi_0}$-a.s. Then, convergence (B) in turn implies our desired conclusion. If $\hat{c}_T$ is chosen to satisfy $\pi_{\phi|Y^T}\left(X_{\phi|Y^T}(\cdot) \le \hat{c}(\cdot)\right) = \alpha$, (A) and (B) imply $\lim_{T\to\infty} P_{Y^T|\phi_0}\left(X_{Y^T|\phi_0}(\cdot) \le \hat{c}_T(\cdot)\right) = \alpha$.

To show (A), 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 constrain $\mathcal{F}$ to equicontinuous functions because the support functions for bounded sets are Lipshitz continuous.

For (A), it suffices to show

$$\sup_{c\in\mathcal{F}} |P_{X_T}(X_T(\cdot) \le c(\cdot)) - P_X(X(\cdot) \le c(\cdot))| \to 0, \text{ as } T \to \infty \tag{A.1}$$

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

$$\left|P_{X_{T'}}(X_{T'}(\cdot) \le c_{T'}(\cdot)) - P_X(X(\cdot) \le c_{T'}(\cdot))\right| > \epsilon \tag{A.2}$$

holds for all T'. By Assumption 4(iv) and the Arzelà-Ascoli theorem, $\mathcal{F}$ 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), $P_X(X(\cdot) \le c_{T''}(\cdot)) \to P_X(X(\cdot) \le c^*(\cdot))$ as $T'' \to \infty$. By the assumption that $X_T \rightsquigarrow X$ and the continuous mapping theorem, $X_{T''} - c_{T''} \rightsquigarrow X - c^*$. Hence, Assumption 4(iii) and the Portmanteau theorem[26] imply that $P_{X_{T''}}(X_{T''}(\cdot) - c_{T'}(\cdot) \le 0) \to P_X(X(\cdot) - c^*(\cdot) \le 0)$ as $T'' \to \infty$. We have shown $\left|P_{X_{T''}}(X_{T''}(\cdot) \le c_{T''}(\cdot)) - P_X(X(\cdot) \le c_{T''}(\cdot))\right| \to 0$ along T'', which contradicts (A.2), so the convergence (A.1) holds.

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

Next, we show (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), 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 $P_Z\left(Z(\cdot) \leq c(\cdot)\right) = P_X\left(X(\cdot) \leq c(\cdot)\right)$. Furthermore, by 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 the two claims proves (B). ■

The proof of Proposition 2 uses the next lemma.

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

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

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 $\{\mathbf{c}_T : T = 1, 2, \ldots\}$ sequence of measurable selections of Lev_T.

Proof of Lemma A.4. To prove (a), suppose that the conclusion is false. That is, there exist a subsequence T', ϵ, $\delta > 0$, and $\mathbf{c} = (c_\ell, c_u) \in Lev$ such that $P_{Y^{T'}|\phi_0}\left(d(\mathbf{c}, Lev_{T'}) > \epsilon\right) > \delta$ for all T'. The event $d(\mathbf{c}, Lev_{T'}) > \epsilon$ 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

$$P_{Y^{T'}|\phi_0}\left(J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) < \alpha\right) > \delta \tag{A.3}$$

along T'. Under Assumption 4(i), however, $J_{T'}\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) - J\left(c_\ell + \frac{\epsilon}{2}, c_u + \frac{\epsilon}{2}\right) \to 0$, $p_{Y^\infty|\phi_0}$-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.3).

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

$$P_{Y^{T'}|\phi_0}\left(J\left(c_{\ell,T'} + \frac{\epsilon}{2}, c_{u,T'} + \frac{\epsilon}{2}\right) < \alpha\right) > \delta \tag{A.4}$$

holds along T'. Note, however, that

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

in $p_{Y^{T'}|\phi_0}$-probability, where the strict inequality follows from that $J(\cdot)$ is strictly monotonic and $J_T(\mathbf{c}_{T'}) \geq \alpha$, and the convergence in probability in the last line follows from the continuity of $J(\cdot)$ and $\sup_{\mathbf{c} \in \mathcal{R}^2} |J(\mathbf{c}) - J_{T'}(\mathbf{c})| \to 0$ for any sequence of distributions $J_{T'}$ converging weakly to a distribution with continuous CDF (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.4). ■

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 $IS_\eta(\phi)$,

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

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 $X_{\phi|Y^T}$, the 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 satisfies 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 $\mathbf{c} \equiv (c_\ell, c_u)'$,

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

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 $\hat{\mathbf{c}}_T \in \arg\min_{\mathbf{c}}\{c_\ell + c_u\}$ subject to $J_T(\mathbf{c}) \geq \alpha$. Having shown Assumption 4(i), let $J(\mathbf{c}) \equiv P_X\left(\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} X \leq \mathbf{c}\right)$. Note that the weak convergence $X_{\phi|Y^T} \rightsquigarrow X$ and continuity of $J(\cdot)$ imply $J_T(\mathbf{c}) \to J(\mathbf{c})$ 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: $\mathbf{c}^* \in \arg\min_{\mathbf{c}}\{c_\ell + c_u\}$ subject to $J(\mathbf{c}) \geq \alpha$. Since $\{\mathbf{c} : J(\mathbf{c}) \geq \alpha\}$ 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.

We next prove $\hat{\mathbf{c}}_T \rightarrow \mathbf{c}^*$ in $p_{Y^T|\phi_0}$-probability as $T \to \infty$. 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, i.e., there exist ϵ, $\delta > 0$, and subsequence T' such that

$$P_{Y^{T'}|\phi_0}(\|\hat{\mathbf{c}}_{T'} - \mathbf{c}^*\| > \epsilon) > \delta \tag{A.5}$$

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.5) 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 $P_{Y^{T'}|\phi_0}(\|\tilde{\mathbf{c}}_{T'} - \mathbf{c}^*\| > \epsilon) > \delta$ 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'. 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

$$P_{Y^{T'}|\phi_0}\left(\hat{f}_{T'} - f^* > \xi\right) > \delta, \tag{A.6}$$

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.6), $P_{Y^{T'}|\phi_0}\left(\hat{f}_{T'} - \left(\check{c}_{\ell,T'} + \check{c}_{u,T'}\right) > \xi\right) > \delta$ 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 of Assumption 4(iv). ■

References

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.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46, 313–324.

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.

Balke, A. and J. Pearl (1997): "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

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

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

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 (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965 – 2018.

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

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.

Giacomini, R. and T. Kitagawa (2020): "Supplement to 'Robust Bayesian Inference for Set-Identfied Models'," *unpublished manuscript.*

Giacomini, R., T. Kitagawa, and A. Volpicella (2020): "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.

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

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.

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, J. Payne, and A. Verez (2020): "Posterior Distributions of Nondifferentiable Functions," *Journal of Econometrics*, 217, 161–175.

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.

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

Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 3.

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

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.

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.

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.

Ruggeri, F. and S. Sivaganesan (2000): "On a Global Sensitivity Measure for Bayesian Inference," *Sankhyā: The Indian Journal of Statistics, Series A*, 62, 110–127.

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.

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

Wets, R. J.-B. (1985): "On the Continuity of the Value of a Linear Program and of Related Polyhedral-Valued Multifunctions," *Mathematical Programming Study*, 24, 14–29.

Wolf, C. (2020): "SVAR (Mis-)Identification and the Real Effects of Monetary Policy Shocks," *American Economic Journal: Macroeconomics, forthcoming.*