Federal Reserve Bank of Chicago

Robust Bayesian Analysis for Econometrics

Raffaella Giacomini, Toru Kitagawa, and Matthew Read

August 23, 2021

WP 2021-11

https://doi.org/10.21033/wp-2021-11

Robust Bayesian Analysis for Econometrics*

Raffaella Giacomini†, Toru Kitagawa‡, and Matthew Read§

This draft: 23 Aug 2021

Abstract

We review the literature on robust Bayesian analysis as a tool for global sensitivity analysis and for statistical decision-making under ambiguity. We discuss the methods proposed in the literature, including the different ways of constructing the set of priors that are the key input of the robust Bayesian analysis. We consider both a general set-up for Bayesian statistical decisions and inference and the special case of set-identified structural models. We provide new results that can be used to derive and compute the set of posterior moments for sensitivity analysis and to compute the optimal statistical decision under multiple priors. The paper ends with a self-contained discussion of three different approaches to robust Bayesian inference for set-identified structural vector autoregressions, including details about numerical implementation and an empirical illustration.

Keywords: ambiguity, Bayesian robustness, statistical decision theory, identifying restrictions, multiple priors, structural vector autoregression

*We 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). The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago, the Federal Reserve System or the Reserve Bank of Australia

†University College London, Department of Economics and Federal Reserve Bank of Chicago. Email: r.giacomini@ucl.ac.uk

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

§University College London, Department of Economics and Reserve Bank of Australia. Email: readm@rba.gov.au

1 Introduction

Bayesian analysis has many attractive features, such as finite-sample decision-theoretic optimality and computational tractability. The crucial assumption to enjoy these benefits is that the researcher can specify the inputs of the analysis: the likelihood, a prior distribution for the parameters, and a loss function if the analysis involves statistical decisions. In practice, however, researchers commonly face uncertainty about the choice of these inputs. Robust Bayesian analysis addresses this uncertainty by quantifying the sensitivity of the results of Bayesian inference to changes in these inputs. In this paper we present a selective review of the literature on robust Bayesian analysis.[1]

How sensitivity should be measured in a robust Bayesian analysis depends on how the input is perturbed. A local approach quantifies marginal changes in posterior quantities with respect to local perturbation in the input; for example, see Gustafson (2000), Müller (2012), and the references therein. In contrast, a global approach introduces a set of inputs and summarizes posterior sensitivity by reporting the corresponding set of posterior quantities.

The main focus of this paper is on sensitivity to the prior input and on global, rather than local, robust Bayesian analysis. There are several reasons for pursuing a global approach. First, the set of prior distributions to be specified as an input of the robust analysis can be viewed as a representation of ambiguous beliefs (Knightian uncertainty), which has been well studied in economic decision theory and experimental economics since the pioneering works of Ellsberg (1961) and Schmeidler (1989). Some statisticians have also argued that a set of priors is easier to elicit than a single prior (Good (1965)). Second, sets of posterior means or probabilities are easier to interpret than local sensitivity parameters such as the derivative of the posterior mean with respect to the prior mean. Third, although it is often argued that the set of posteriors is more difficult to compute than local sensitivity parameters, that is not the case for the structural vector autoregression (SVAR) models considered in detail in this paper.

We first consider a general environment for inference and statistical decision-making under multiple priors and discuss different ways of constructing the set of priors. We then specialize the discussion to set-identified structural models, where the posterior sensitivity is due to a component of the prior (the conditional prior for the structural parameter given the reduced-form parameter) that is never updated by the data. We illustrate the "full-ambiguity" approach of Giacomini and Kitagawa (2021) (henceforth, GK) to constructing a set of priors in general set-identified structural models. In addition, we provide new theoretical results that can be used to derive and compute the set of posterior moments for sensitivity analysis and that are also useful for computing the optimal statistical decision in the presence of multiple priors. These results are new to the literature and generalize some results in GK.

The paper ends with a detailed and self-contained discussion of robust Bayesian inference for

[1] See Berger (1994) and Ríos Insua and Ruggeri (2000) for previous surveys on the topic.

set-identified SVARs. Set-identification arises in SVARs when there are sign restrictions on impulse responses or under-identifying zero restrictions, or in SVARs identified using external instruments ("proxy SVARs") when there are multiple instruments for multiple shocks (see Giacomini, Kitagawa and Read (in press)). We review three robust Bayesian approaches for set-identified SVARs. The common feature of the approaches is that they replace the unrevisable component of the prior with multiple priors. We cast the approaches within a common framework, discuss their numerical implementation and illustrate their use in an empirical example. Our goal is to elucidate how and when these different approaches may be useful in eliminating or quantifying the influence of the unrevisable component of the prior on posterior inference. Ultimately, we argue that the robust Bayesian outputs generated by these methods should be reported alongside the standard Bayesian outputs that are typically reported in studies that use set-identified SVARs.

The first approach to robust Bayesian inference for set-identified SVARs is GK, which replaces the unrevisable component of the prior with the set of all priors that are consistent with the imposed identifying restrictions. This generates a set of posteriors, which can be summarised by a set of posterior means and a robust credible region, which is the shortest interval that is assigned at least a given posterior probability under all posteriors within the set. One can also report the lower or upper posterior probability of some event (e.g, that the output response to a monetary policy shock is negative at some horizon), which is the smallest or largest probability of the event over all posteriors in the set. GK show that, under certain conditions, the set of posterior means is a consistent estimator of the identified set and the robust confidence region attains valid frequentist coverage of the true identified set asymptotically. In contrast, under standard Bayesian inference, the posterior mean asymptotically lies at a point within the identified set that is determined entirely by the prior, and standard credible intervals lie strictly within the identified set asymptotically (Moon and Schorfheide 2012). The approach of GK therefore reconciles the asymptotic disagreement between frequentist and Bayesian inference in set-identified models.

The second approach is the "model-averaging" approach of Giacomini, Kitagawa and Volpicella (in press; henceforth, GKV). The approach extends Bayesian model averaging to a mix of single-prior and multiple-prior models. Given prior probabilities chosen by the user, the multiple-posterior models are averaged (posterior-by-posterior) with the single-posterior models, where the weights on each model are the posterior model probabilities. This averaging generates a set of posteriors, which can be summarised as in GK. For instance, when we have one point-identified model (which yields a single prior) and one set-identified model, the post-averaging set of posterior means shrinks the bounds of the set of posterior means in the set-identified model towards the posterior mean in the point-identified model. GKV explain the conditions under which prior model probabilities are revised, in which case the data may be informative about which identifying restrictions are more plausible.

The third approach is the "KL-neighborhood" approach proposed by Giacomini, Kitagawa

and Uhlig (2019; henceforth, GKU). GKU consider a set of priors in a Kullback-Leibler (KL) neghborhood of a 'benchmark' prior. The motivation for this proposal is that one may not want to entertain priors that are, in some sense, far from the benchmark prior, because the benchmark prior may be partially credible. Similarly to GK, this generates a set of posteriors, which can be summarised by a set of posterior means and/or quantiles. GKU also derive a point estimator solving a Bayesian statistical decision problem that allows for ambiguity over the set of priors within the KL-neighborhood around the benchmark prior (the 'posterior Gamma minimax problem').

The robustness issue reviewed in this article focuses exclusively on misspecification of and sensitivity to the prior distribution in the Bayesian setting. There is a vast literature on robust statistics from the frequentist perspective; for example, see Huber and Ronchetti (2009), Rieder (1994), and references therein for classical approaches to robust statistical methods. The frequentist approach to robustness typically concerns misspecification of the likelihood (contamination of the data-generating process), identifying assumptions, moment conditions, or the distribution of unobservables. The main focuses of this literature are to quantify sensitivity of estimation and inference to such mispecification and develop estimators that are robust against it. For recent advances in econometrics, see Kitamura et al. (2013), Andrews et al. (2017, 2020), Bonhomme and Weidner (2018), Christensen and Connault (2019) and Armstrong and Kolesár (2021).

The remainder of the paper is structured as follows. Section 2 presents a general overview of robust Bayesian analysis. Section 3 specializes the discussion to set-identified structural models. Section 4 presents new theoretical results regarding the set of posterior moments. Section 5 discusses three approaches to robust Bayesian analysis in set-identified SVARs. Section 6 contains details about the numerical implementation of the three approaches, emphasizing the choices that practitioners face during implementation. Section 7 applies the three approaches to the empirical application considered by Arias, Caldara and Rubio-Ramírez (2019). Sections 5–7 are self-contained. Section 8 concludes.

2 Robust Bayesian analysis

2.1 Bayesian statistical decisions and inference

We start from the classical framework of statistical decision theory as in Wald (1950). Let $\mathbf{Y} \in \mathcal{Y} \subset \mathbb{R}^T$ be a sample whose probability distribution is assumed to belong to a parametric family of distributions $P_{\mathbf{Y}|\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{d_\theta}$. We denote the Borel σ-algebra of $\boldsymbol{\theta}$ by $\mathcal{B}(\Theta)$. Let $\delta(\cdot) : \mathcal{Y} \to \mathcal{A}$, $\delta \in \mathcal{D}$, be a nonrandomized statistical decision rule that maps a sample $Y \in \mathcal{Y}$ to an action $a \in \mathcal{A}$, where $\mathcal{D}$ is a set of (possibly constrained) decision rules and $\mathcal{A}$ is the set of actions that the decision maker (DM) can take. Let $L(\boldsymbol{\theta}, a) : \Theta \times \mathcal{A} \to \mathbb{R}$ be the *loss* that the DM incurs when the true parameter value is $\boldsymbol{\theta}$ and the action taken is a. The *risk* of the decision rule δ, denoted by $R(\boldsymbol{\theta}, \delta)$,

measures the average loss under repeated sampling of the sample $\mathbf{Y} \sim P_{\mathbf{Y}|\boldsymbol{\theta}}$,

$$R(\boldsymbol{\theta}, \delta) \equiv E_{\mathbf{Y}|\boldsymbol{\theta}}(L(\boldsymbol{\theta}, \delta(\mathbf{Y}))) = \int_{\mathcal{Y}} L(\boldsymbol{\theta}, \delta(\mathbf{y})) dP_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}). \tag{2.1}$$

Since a decision rule that dominates the others uniformly over $\boldsymbol{\theta}$ is usually not available, the ranking of decision rules depends on how the DM handles uncertainty about the unknown parameter $\boldsymbol{\theta}$.

If the DM could express the uncertainty about $\boldsymbol{\theta}$ in the form of a probability distribution $\pi_{\boldsymbol{\theta}}$ of the measurable space $(\Theta, \mathcal{B}(\Theta))$, they would choose δ that performs best in terms of *Bayes risk*

$$r(\pi_{\boldsymbol{\theta}}, \delta) \equiv \int_{\Theta} R(\boldsymbol{\theta}, \delta) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}). \tag{2.2}$$

This corresponds to the Bayesian decision principle, which is easy to derive and implement, and is favored by statisticians and decision theorists on the basis of the likelihood principle, conditionality viewpoint, and guaranteed admissibility (see, for example, the discussions in Chapters 1 and 4 of Berger (1985)). Under weak regularity conditions (e.g., Brown and Purves (1973)), the Bayes decision δ_{bayes} minimizing (2.2) can be obtained by minimizing the *posterior expected loss* $\rho(\pi_{\boldsymbol{\theta}}, a)$ at each realization of the sample supported by the marginal likelihood, $m(\cdot) = \int_{\Theta} P_{\mathbf{Y}|\boldsymbol{\theta}}(\cdot) d\pi_{\boldsymbol{\theta}}$. That is, $\delta_{Bayes}(\mathbf{y})$ minimizes in a

$$\rho(\pi_{\boldsymbol{\theta}}, a) \equiv \int_{\Theta} L(\boldsymbol{\theta}, a) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}), \tag{2.3}$$

where $\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\cdot|\mathbf{y})$ is the posterior (distribution) of $\boldsymbol{\theta}$ given the realization of the sample $\mathbf{Y} = \mathbf{y}$. If the goal of the analysis is to summarize uncertainty about the unknown parameter $\boldsymbol{\theta}$ upon observing the data, it suffices to report the posterior $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$ or its summary statistics, which is feasible in many contexts thanks to advances in Monte Carlo sampling methods.

The Bayesian approach to statistical decision-making and inference is conceptually straightforward and easy to implement numerically as long as the DM specifies a triplet of loss function, likelihood and prior for $\boldsymbol{\theta}$. Specifying these inputs, however, can be a challenge in practice. The DM may not be sure about how to choose the loss function and/or the class of parametric distributions for the likelihood. Arguably, the dominant concern in Bayesian practice is how to organize the DM's belief for $\boldsymbol{\theta}$ (or the lack thereof) in terms of a prior. To cope with these concerns, robust Bayesian analysis allows for multiplicity in each of these inputs and assesses the set of posterior expected losses or Bayes risks spanned by the set of inputs. If the DM is interested in an optimal decision subject to the set of posterior losses or Bayes risks, the robust Bayesian literature has considered minimizing the upper bound of the posterior expected losses or Bayes risks. See Dey and Micheas (2000), Shyamalkumar (2000), and references therein for robust Bayesian analysis with multiple losses and likelihoods. The focus of this paper is on robust Bayesian analysis with multiple priors.

2.2 Robust Bayesian analysis with multiple priors

Let $\Pi_{\boldsymbol{\theta}}$ be a set of priors for $\boldsymbol{\theta}$. In the subjective robust Bayesian sense, $\Pi_{\boldsymbol{\theta}}$ represents ambiguity such that the DM considers any prior in $\Pi_{\boldsymbol{\theta}}$ plausible and cannot judge which one is more credible than the others.

To summarize the posterior uncertainty for $\boldsymbol{\theta}$, we update the set of priors $\Pi_{\boldsymbol{\theta}}$ based on the likelihood $P_{\mathbf{Y}|\boldsymbol{\theta}}$. One approach is prior-by-prior updating, which is often referred to as the full Bayesian updating rule.[2] It applies Bayes' rule to each prior in Π_θ to obtain the set of posteriors $\Pi_{\boldsymbol{\theta}|Y}$,

$$\Pi_{\boldsymbol{\theta}|\mathbf{Y}} \equiv \left\{ \pi_{\boldsymbol{\theta}|\mathbf{Y}}(\cdot) = \frac{\int_{\{\boldsymbol{\theta}\in\cdot\}} P_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int_\Theta P_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})} : \pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}} \right\}. \tag{2.4}$$

Given the set of posteriors, the analysis proceeds by reporting various posterior quantities. For instance, the *lower and upper posterior probabilities* for the hypothesis $\{\boldsymbol{\theta} \in A\}$ are the lower and upper bounds of $\pi_{\boldsymbol{\theta}|\mathbf{Y}}(A)$ on $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$; for an arbitrary measurable subset $A \in \mathcal{B}(\Theta)$, we have

$$\begin{aligned} &\text{lower posterior probability for } \boldsymbol{\theta}\text{:} \quad \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A) \equiv \inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(A), \\ &\text{upper posterior probability for } \boldsymbol{\theta}\text{:} \quad \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(A) \equiv \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(A) = 1 - \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A). \end{aligned} \tag{2.5}$$

For example, $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A)$ can be interpreted as saying that "the posterior credibility for the hypothesis $\{\boldsymbol{\theta} \in A\}$ is at least equal to $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A)$, no matter which prior in $\Pi_{\boldsymbol{\theta}}$ one assumes."

The corresponding probabilities for a parameter transformation $\eta = h(\boldsymbol{\theta}) \in \mathcal{H}$ are obtained as

$$\begin{aligned} &\text{lower posterior probability for } \eta = h(\boldsymbol{\theta})\text{:} \quad \pi_{\eta|\mathbf{Y}*}(D) \equiv \inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta}) \in D), \\ &\text{upper posterior probability for } \eta = h(\boldsymbol{\theta})\text{:} \quad \pi^*_{\eta|\mathbf{Y}}(D) \equiv \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta}) \in D) = 1 - \pi_{\eta|\mathbf{Y}*}(D), \end{aligned}$$

for any Borel set $D \subset \mathcal{H}$. If $\eta = h(\boldsymbol{\theta})$ is a scalar parameter of interest, quantities often reported in Bayesian global sensitivity analysis are the bounds for the posterior mean of η,

$$\left[\inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \int_\Theta h(\boldsymbol{\theta}) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}), \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \int_\Theta h(\boldsymbol{\theta}) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}) \right]. \tag{2.6}$$

In addition, the robust Bayesian counterpart of the highest posterior density region for η can be defined by a set $C_\alpha \subset \mathcal{H}$ such that the posterior lower probability is greater than or equal to α,

$$\pi_{\eta|\mathbf{Y}*}(C_\alpha) \geq \alpha. \tag{2.7}$$

[2] The literature has considered different ways to update the set of priors. For example, the maximum likelihood updating rule (Gilboa and Schmeidler (1993)) uses the observed sample to select a prior by maximizing the marginal likelihood and then applies Bayes' rule. This way of updating is known as Type-II maximum likelihood (Good (1965)) or, equivalently, as the empirical Bayes method (e.g., Robbins (1956), Berger and Berliner (1986)).

Such 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 set*". GK call C_α a *robust credible region with credibility* α.

The prior set $\Pi_{\boldsymbol{\theta}}$ can also generate the sets of Bayes risks and posterior expected losses; for any $\tilde{\pi}_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}$ and decision function $\delta \in \mathcal{D}$, it holds

$$\inf_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta) \leq r(\tilde{\pi}_{\boldsymbol{\theta}}, \delta) \leq \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta), \tag{2.8}$$

and for any action $a \in \mathcal{A}$ and sample realization $\mathbf{y} \in \mathcal{Y}$, it holds

$$\inf_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a) \leq \rho(\tilde{\pi}_{\boldsymbol{\theta}}, a) \leq \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a). \tag{2.9}$$

The set of posterior expected losses coincides with the set (2.6) with $h(\cdot)$ set to $L(\boldsymbol{\theta}, a)$.

Interpreting the set of priors as the DM's ambiguous belief, an unconditional optimality criterion attractive to the ambiguity-averse DM is the unconditional *Gamma minimax* criterion. It defines an optimal decision δ^* by minimizing the worst-case Bayes risk,

$$\delta^* \equiv \arg \inf_{\delta \in \mathcal{D}} \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta). \tag{2.10}$$

A similar but distinct optimality criterion is the *conditional Gamma minimax* criterion, which defines an optimal action by minimizing the worst-case posterior expected loss conditional on $\mathbf{y}$,

$$a^*_{\mathbf{y}} \equiv \arg \inf_{a \in \mathcal{A}} \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a). \tag{2.11}$$

The unconditional Gamma minimax decision δ^* and the conditional Gamma minimax action $a^*_{\mathbf{y}}$ do not generally agree. An advantage of the former is its guaranteed admissibility, since, under the regularity conditions leading to the minimax theorem, the Gamma minimax decision is Bayes optimal under a least-favorable prior. On the other hand, obtaining δ^* is challenging analytically and numerically, unless we limit the analysis to simple models with a particular set of priors. See Chamberlain (2000) for an application to portfolio choice and Vidakovic (2000) for a review. The conditional Gamma minimax action is easier to analyze and implement as the minimax problem only involves action a. However, a potential downside is that this criterion is not guaranteed to give an admissible decision. DasGupta and Studden (1989) consider conditional Gamma minimax estimators for the normal means model. See also Betrò and Ruggeri (1992) for other examples.

2.3 Examples of sets of priors

We review different constructions of the prior set $\Pi_{\boldsymbol{\theta}}$ that have been considered in the literature.

Example 1 *In one the earliest applications of Bayesian global sensitivity analysis, Chamberlain and Leamer (1976) and Leamer (1978, 1982) consider the regression model*

$$\underset{T\times 1}{\mathbf{y}} = \underset{T\times k}{\mathbf{X}} \underset{k\times 1}{\boldsymbol{\beta}} + \underset{T\times 1}{\boldsymbol{\epsilon}}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_T(0, \sigma^2 \mathbf{I}_T),$$

and specify a set of conjugate priors where the variance of the conjugate Gaussian prior for $\boldsymbol{\beta}$ varies over a certain set. These works derive a closed-form representation for the set of posterior means of $\boldsymbol{\beta}$ and study its analytical properties.

Example 2 (ϵ-contaminated set) *A well-studied set of priors is the ϵ-contaminated set of priors (e.g., Huber (1973), Berger (1984, 1985), Berger and Berliner (1986), Sivaganesan and Berger (1989)). Its canonical representation is*

$$\Pi_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}} = (1-\epsilon)\pi^0_{\boldsymbol{\theta}} + \epsilon q_{\boldsymbol{\theta}} : q_{\boldsymbol{\theta}} \in \mathcal{Q}_{\boldsymbol{\theta}}\}, \tag{2.12}$$

where the inputs to be specified by the user are $\pi^0_{\boldsymbol{\theta}}$, the base prior representing a benchmark prior with limited confidence, $\epsilon \in [0,1]$, the amount of contamination that gauges the uncertainty on the base prior, and $\mathcal{Q}_{\boldsymbol{\theta}}$, the set of contamination distributions which span the plausible priors. Huber (1973) considers a set of arbitrary distributions for $\mathcal{Q}_{\boldsymbol{\theta}}$ and derives a closed-form expression for the lower and upper posterior probabilities. Berger and Berliner (1986) consider the empirical Bayes posterior (Type-II maximum likelihood) with various sets of contamination distributions including unimodal and symmetric ones, and Sivaganesan and Berger (1989) derive the set of posteriors with contaminations that preserve unimodality of $\pi_{\boldsymbol{\theta}}$.

Example 3 (Priors with fixed marginal) *In the presence of multiple parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{d_\theta})$, $d_\theta \geq 2$, it is often feasible to elicit the marginal distribution of the prior for each component, while eliciting their dependence is difficult. Lavine et al. (1991) consider an ϵ-contaminated set of priors with $\mathcal{Q}_{\boldsymbol{\theta}}$ consisting of the set of priors sharing fixed marginals for each component in $\boldsymbol{\theta}$, while their dependence is unconstrained. They propose linearization techniques to solve the optimization in (2.6). Moreno and Cano (1995) fix the prior marginals of only a subset of the parameters in $\boldsymbol{\theta}$.*

Example 4 (Priors known up to coarsened domain) *Kudō (1967) and Manski (1981) study the set of posterior distributions and Gamma minimax statistical decisions when the analyst can elicit a prior distribution only up to a class of coarsened subsets of $\boldsymbol{\theta}$, i.e., a σ-algebra smaller than the Borel σ-algebra of $\boldsymbol{\theta}$, $\mathcal{B}(\Phi)$. Formally, consider a transformation $\boldsymbol{\phi} = g(\boldsymbol{\theta})$, where g is a many-to-one function $g : \Theta \to \Phi$ that coarsens the parameter space of $\boldsymbol{\theta}$, and $\boldsymbol{\phi}$ indexes the sets in the partition of $\boldsymbol{\theta}$. Given a unique prior $\pi_{\boldsymbol{\phi}}$ on $(\Phi, \mathcal{B}(\Phi))$, the set of priors for $\boldsymbol{\theta}$ consists of the distributions of $\boldsymbol{\theta}$ that imply that the distribution of $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ is the given $\pi_{\boldsymbol{\phi}}$,*

$$\Pi_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}}(g^{-1}(B)) = \pi_{\boldsymbol{\phi}}(B),\ \forall B \in \mathcal{B}(\Phi)\}. \tag{2.13}$$

This set of priors can be interpreted as a special case of the general construction in Wasserman (1990). Wasserman (1990) considers a set of priors whose lower and upper probabilities are given by the containment and capacity functional of a random set $\Gamma : \Phi \rightrightarrows \Theta$, i.e., given $\pi_{\boldsymbol{\phi}}$, a probability measure on $(\Phi, \mathcal{B}(\Phi))$,

$$\Pi_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}*}(A) \leq \pi_{\boldsymbol{\theta}}(A) \leq \pi^*_{\boldsymbol{\theta}}(A),\ \forall A \in \mathcal{B}(\Theta)\}, \tag{2.14}$$

where $\pi_{\boldsymbol{\theta}*}(A) = \pi_{\boldsymbol{\phi}}(\Gamma(\boldsymbol{\phi}) \subset A)$ *and* $\pi^*_{\boldsymbol{\theta}}(A) = \pi_{\boldsymbol{\phi}}(\Gamma(\boldsymbol{\phi}) \cap A \neq \emptyset)$*. In view of random set theory, this set of priors can be interpreted as the set of selectionable distributions from the random set* $\Gamma(\boldsymbol{\phi})$*,* $\boldsymbol{\phi} \sim \pi_{\boldsymbol{\phi}}$*. See Artstein (1983), Molchanov (2005), and Molchanov and Molinari (2018). This set of priors coincides with (2.13) in the special case where* $\Gamma(\boldsymbol{\phi})$ *is set to* $g^{-1}(\boldsymbol{\phi})$*. Wasserman (1990) derives analytically the lower and upper posterior probabilities when* $\Pi_{\boldsymbol{\theta}}$ *is given in the form (2.14).*

Example 5 (Priors in an information neighborhood) *In the minimax approaches to robust estimation and robust control, the set of distributions that one wishes to be robust against is formed by an information neighborhood around a benchmark distribution* $\pi^0_{\boldsymbol{\theta}}$*. One can consider a variety of statistical divergence criteria to define the neighborhood, including the KL divergence and the Hellinger distance; see, for example, Peterson et al. (2000), Hansen and Sargent (2001) and Kitamura et al. (2013). This approach offers a flexible and analytically tractable way to define the set of priors for robust Bayesian analysis. Along this line, Ho (2020) introduces the set of priors through the KL neighborhood centered at a benchmark prior* $\pi^0_{\boldsymbol{\theta}}$*,*

$$\Pi_{\boldsymbol{\theta}} = \left\{ \pi_{\boldsymbol{\theta}} : \int_{\Theta} \log \left(\frac{d\pi_{\boldsymbol{\theta}}}{d\pi^0_{\boldsymbol{\theta}}} \right) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \leq \lambda \right\}, \tag{2.15}$$

where $\lambda > 0$ *is the radius of the KL neighborhood specified by the user. Watson and Holmes (2016) analyze posterior Gamma minimax actions with a KL neighborhood around the benchmark posterior. As discussed in Ho (2020), a convenient feature of this approach is that it is easy to additionally impose moment constraints on the priors and/or posteriors.*

3 Robust Bayesian Analysis for Set-identified Models

This section discusses how the general framework of robust Bayesian analysis introduced in the previous section can be extended to a class of set-identified structural models.

3.1 Set-identified structural models

Non-identification of a structural parameter $\boldsymbol{\theta}$ arises when multiple values of $\boldsymbol{\theta}$ are observationally equivalent; that is, there exist $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ such that $p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}')$ for every $\mathbf{y} \in \mathcal{Y}$ (Rothenberg (1971)). Observational equivalence can be represented by a many-to-one function $g : (\Theta, \mathcal{A}) \to (\Phi, \mathcal{B})$, such that $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}')$ if and only if $p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}')$ for all $\mathbf{y} \in \mathcal{Y}$ (e.g., Barankin (1960)). This relationship partitions the parameter space Θ into equivalent classes, in each of which the likelihood of $\boldsymbol{\theta}$ is "flat" irrespective of observations, and $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ maps each of the equivalent classes to a point in a parameter space Φ. Following the terminology of structural models in econometrics (Koopmans and Reiersol (1950)), $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ is the reduced-form parameter indexing the distribution of the data. The likelihood depends on $\boldsymbol{\theta}$ only through $\boldsymbol{\phi} = g(\boldsymbol{\theta})$; that is, there exists a $\mathcal{B}(\Phi)$-measurable function $\hat{p}(\mathbf{y}|\cdot)$ such that $p(\mathbf{y}|\boldsymbol{\theta}) = \hat{p}(\mathbf{y}|g(\boldsymbol{\theta}))$ for every $\mathbf{y} \in \mathcal{Y}$ and $\boldsymbol{\theta} \in \Theta$.

The identified set of $\boldsymbol{\theta}$ is the inverse image of $g(\cdot)$: $IS_\theta(\boldsymbol{\phi}) = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) = \boldsymbol{\phi}\}$, where $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ and $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}')$ for $\boldsymbol{\phi} \neq \boldsymbol{\phi}'$ are disjoint and $\{IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi\}$ is a partition of Θ. For the parameter of interest $\eta = h(\boldsymbol{\theta})$ with $h : \Theta \to \mathcal{H}$, $\mathcal{H} \subset \mathbb{R}^{d_\eta}$, $d_\eta < \infty$, we define the identified set as the projection of $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ onto $\mathcal{H}$ through $h(\cdot)$, $IS_\eta(\boldsymbol{\phi}) \equiv \{h(\boldsymbol{\theta}) : \boldsymbol{\theta} \in IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})\}$. The parameter $\eta = h(\boldsymbol{\theta})$ is point- or set-identified at $\boldsymbol{\phi}$ if $IS_\eta(\boldsymbol{\phi})$ is a singleton or not a singleton, respectively. By the definition of observational equivalence, $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ and $IS_\eta(\boldsymbol{\phi})$ are the sharp identification regions at every distribution of data indexed by $\boldsymbol{\phi}$.

In SVARs with sign restrictions, the model can be observationally restrictive in the sense of Koopmans and Reiersol (1950). This means the model is falsifiable and $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ can be empty for some $\boldsymbol{\phi} \in \Phi$ on which the reduced-form likelihood is well defined.

3.2 Influence of prior choice under set-identification

Let $\pi_{\boldsymbol{\theta}}$ be a prior for $\boldsymbol{\theta}$ and $\pi_{\boldsymbol{\phi}}$ be the corresponding prior for $\boldsymbol{\phi}$ induced by $\pi_{\boldsymbol{\theta}}$ and $g(\cdot)$:

$$\pi_{\boldsymbol{\phi}}(B) = \pi_{\boldsymbol{\theta}}(IS_{\boldsymbol{\theta}}(B)) \quad \text{for all } B \in \mathcal{B}(\Phi). \tag{3.1}$$

From the definition of the reduced-form parameter, the likelihood for $\boldsymbol{\theta}$ is flat on $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ for any $\mathbf{y}$, which implies conditional independence $\boldsymbol{\theta} \perp \mathbf{Y}|\boldsymbol{\phi}$. Hence, as obtained by Poirier (1998), Moon and Schorfheide (2012), and Baumeister and Hamilton (2015), the posterior of $\boldsymbol{\theta}$, $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$, can be expressed as

$$\pi_{\boldsymbol{\theta}|\mathbf{Y}}(A) = \int_\Phi \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(A) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \quad A \in \mathcal{B}(\Theta), \tag{3.2}$$

where $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ is the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$ whose support agrees with or is contained in $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$, and $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$ is the posterior of $\boldsymbol{\phi}$. This expression shows that the prior of the reduced-form parameter, $\pi_{\boldsymbol{\phi}}$, can be updated by the data, whereas the conditional prior of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$ is never updated because the likelihood is flat on $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}) \subset \Theta$ for any realization of the sample. In this sense, one can interpret $\pi_{\boldsymbol{\phi}}$ as the *revisable prior knowledge* and the conditional priors, $\left\{\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\cdot|\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi\right\}$, as the *unrevisable prior knowledge*.

Marginalizing $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$ for $\eta = h(\boldsymbol{\theta})$ gives

$$\pi_{\eta|\mathbf{Y}}(D) = \int_\Phi \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta}) \in D) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} = \int_\Phi \pi_{\eta|\boldsymbol{\phi}}(D) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} \tag{3.3}$$

for $D \in \mathcal{B}(\mathcal{H})$, where $\pi_{\eta|\boldsymbol{\phi}}$ is the conditional prior for η given $\boldsymbol{\phi}$, which is by construction supported on $IS_\eta(\boldsymbol{\phi})$. If η is set-identified, its posterior thus has a non-degenerate unrevisable component $\pi_{\eta|\boldsymbol{\phi}}$.

The previous discussion clarifies the following features of Bayesian inference under set-identification:

1. As discussed in Poirier (1998), the lack of identification for η does not mean that the prior for η is not updated by the data. The prior for η can be updated, but this happens only

through the update of the prior for the reduced-form parameter $\boldsymbol{\phi}$. Comparing the prior and posterior for η therefore does not indicate whether or not the parameter is point-identified.

2. Since the posteriors for $\boldsymbol{\theta}$ and η involve nonrevisable priors, they are sensitive to the choice of prior even asymptotically. In particular, the posterior for η is sensitive to perturbations of the prior that change the shape of $\pi_{\eta|\boldsymbol{\phi}}$. This suggests that posterior sensitivity, rather than the comparison of the shapes of prior and posterior, is informative about the strength of identification. This feature is similar to the local sensitivity analysis in Müller (2012).

3. The reduced-form parameter $\boldsymbol{\phi}$ is identified by construction. If the likelihood for $\boldsymbol{\phi}$ converges to its true value $\boldsymbol{\phi}_0$ in large samples, the posterior for η converges to the conditional prior $\pi_{\eta|\boldsymbol{\phi}}$ given $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. Since the support of $\pi_{\eta|\boldsymbol{\phi}}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ is equal to or contained in $IS_\eta(\boldsymbol{\phi}_0)$, the asymptotic posterior does not lead to an estimate for η that lies outside of its identified set. However, the shape of the asymptotic posterior on $IS_\eta(\boldsymbol{\phi}_0)$ is determined entirely by the prior.

3.3 Full ambiguity for the unrevisable prior

The discussion in the previous subsection motivates a key feature of the robust Bayesian approaches for set-identified models that we discuss in this paper: they assume multiple priors for the unrevisable component of the prior (the prior for the structural parameter $\boldsymbol{\theta}$ given the reduced-form parameter $\boldsymbol{\phi}$), but maintain a single prior for the revisable component (the prior for $\boldsymbol{\phi}$).[3]

In this section we review the full-ambiguity approach of GK for general models, but one can consider a variety of approaches to refine the set of priors in GK to reflect partial prior knowledge about the unrevisable component of the prior. Examples include the ϵ-contaminated set of priors (Example 2) and the KL-neighborhood set of priors (Example 5). For SVARs, GKV investigate the former and GKU investigate a variation of the latter, as we review in Section 5 below.

GK construct a set of priors for $\boldsymbol{\theta}$ constrained by a single proper prior $\pi_{\boldsymbol{\phi}}$ for $\boldsymbol{\phi} = g(\boldsymbol{\theta})$, supported on $g(\Theta)$,

$$\Pi^{FA}_{\boldsymbol{\theta}} \equiv \{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}}(IS_{\boldsymbol{\theta}}(B)) = \pi_{\boldsymbol{\phi}}(B),\ \forall B \in \mathcal{B}(\Phi)\}, \tag{3.4}$$

where $IS_{\boldsymbol{\theta}}(B) = \cup_{\boldsymbol{\phi}\in B} IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$. Noting that $IS_{\boldsymbol{\theta}}(\cdot) = g^{-1}(\cdot)$, this takes the form as the set of priors in (2.13). An equivalent but perhaps more intuitive way to introduce ambiguity for the unrevisable prior is in terms of the set of conditional priors:

$$\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}} \equiv \left\{\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})) = 1, \pi_{\boldsymbol{\phi}} - \textit{almost surely}\right\}. \tag{3.5}$$

[3]GK discuss how this is motivated by the asymptotically negligible effect of the prior choice for $\boldsymbol{\phi}$ and by a desire to avoid possible issues of non-convergence of the set of posteriors. In addition, for $\boldsymbol{\phi}$ one can apply existing methods for constructing a non-informative prior such as Jeffreys' prior or for selecting a data-driven prior.

$\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ consists of arbitrary conditional priors as long as they assign probability one to the identified set of $\boldsymbol{\theta}$, and is linked to $\Pi^{FA}_{\boldsymbol{\theta}}$ in (3.4) by $\Pi^{FA}_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}}(\cdot) = \int \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\cdot) d\pi_{\boldsymbol{\phi}}(\boldsymbol{\phi}) : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}\}$.

Applying Bayes' rule to each prior in $\Pi^{FA}_{\boldsymbol{\theta}}$ gives a set of posteriors for $\boldsymbol{\theta}$. Marginalizing each posterior and invoking (3.3) and (3.5) generates a set of posteriors for the parameter of interest η,

$$\Pi^{FA}_{\eta|\mathbf{Y}} \equiv \left\{ \pi_{\eta|\mathbf{Y}}(\cdot) = \int_{\Phi} \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta}) \in \cdot) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \right\}. \tag{3.6}$$

Under mild regularity conditions (Assumption 1 in GK), GK derive the lower and upper posterior probabilities of $\Pi^{FA}_{\eta|\mathbf{Y}}$ as

$$\pi_{\eta|\mathbf{Y}*}(D) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(\{\boldsymbol{\phi} : IS_{\eta}(\boldsymbol{\phi}) \subset D\}), \quad \pi^{*}_{\eta|\mathbf{Y}}(D) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(\{\boldsymbol{\phi} : IS_{\eta}(\boldsymbol{\phi}) \cap D \neq \emptyset\}), \tag{3.7}$$

for $D \in \mathcal{B}(\mathcal{H})$, and show that the set of posterior probabilities $\{\pi_{\eta|\mathbf{Y}}(D) : \pi_{\eta|\mathbf{Y}} \in \Pi^{FA}_{\eta|\mathbf{Y}}\}$ coincides with the connected intervals $[\pi_{\eta|\mathbf{Y}*}(D), \pi^{*}_{\eta|\mathbf{Y}}(D)]$, which implies that any posterior probability in this set can be attained by some posterior in $\Pi^{FA}_{\eta|\mathbf{Y}}$. The expression for $\pi_{\eta|\mathbf{Y}*}(D)$ shows that the lower probability on D is the probability that the (random) identified set $IS_{\eta}(\boldsymbol{\phi})$ is contained in D in terms of the posterior probability of $\boldsymbol{\phi}$. The upper probability is the probability that the identified set hits D. These closed-form expressions of the lower and upper probabilities suggest how to compute them in practice. For instance, to approximate $\pi_{\eta|\mathbf{Y}*}(D)$, one obtains Monte Carlo draws of $\boldsymbol{\phi}$ from its posterior and computes the proportion of the draws that satisfy $IS_{\eta}(\boldsymbol{\phi}) \subset D$.

GK show that the set of posterior means of η coincides with the Aumann expectation of the convex hull of the identified set, $co(IS_{\eta}(\boldsymbol{\phi}))$, with respect to $\pi_{\boldsymbol{\phi}|Y}$. In particular, if η is a scalar and denoting the convexified identified set for η by $[\ell(\boldsymbol{\phi}), u(\boldsymbol{\phi})] = co(IS_{\eta}(\boldsymbol{\phi}))$, the set of posterior means for η is the interval connecting the posterior means of $\ell(\boldsymbol{\phi})$ and $u(\boldsymbol{\phi})$,

$$\left\{ E_{\eta|\mathbf{Y}}(\eta) : \pi_{\eta|\mathbf{Y}} \in \Pi^{FA}_{\eta|\mathbf{Y}} \right\} = \left[E_{\boldsymbol{\phi}|\mathbf{Y}}(\ell(\boldsymbol{\phi})), E_{\boldsymbol{\phi}|\mathbf{Y}}(u(\boldsymbol{\phi})) \right]. \tag{3.8}$$

The set of posterior τ-th quantiles of η can be computed by first applying (3.7) with $D = (-\infty, t]$, $-\infty < t < \infty$ to obtain the set of the posterior cumulative distribution functions of η for each t and then inverting the upper and lower bounds of this set at $\tau \in (0, 1)$.

Given the representation of the lower posterior probability (3.7), a robust credible region satisfying (2.7) with credibility $\alpha \in (0, 1)$ can be expressed as

$$\pi_{\eta|\mathbf{Y}*}(C_{\alpha}) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(IS_{\eta}(\boldsymbol{\phi}) \subset C_{\alpha})) \geq \alpha. \tag{3.9}$$

GK propose to report the smallest robust credible region (i.e., C_{α} with the smallest volume):

$$C^{*}_{\alpha} \in \arg\min_{C \in \mathcal{C}} Leb(C), \text{ s.t. } \pi_{\phi|\mathbf{Y}}(IS_{\eta}(\boldsymbol{\phi}) \subset C)) \geq \alpha, \tag{3.10}$$

where $Leb(C)$ is the volume of C in terms of the Lebesgue measure and $\mathcal{C}$ is a family of subsets in $\mathcal{H}$. The credible regions for the identified set proposed in Moon and Schorfheide (2011), Norets and Tang (2014) and Kline and Tamer (2016) satisfy (3.9), so they can be interpreted as robust credible regions, but they are not optimized in terms of volume.

4 Analytical results for set of posterior moments

In this section we present new and general theoretical results that link the set of posterior moments in (2.6) and the lower and upper posterior probabilities in (2.5). These results are useful because: 1) they provide a general approach to deriving and computing the set of posterior moments for sensitivity analysis; and 2) they help solve the posterior Gamma minimax problem.

The lower and upper posterior probabilities viewed as functions of $A \in \mathcal{B}(\Theta)$ are nonnegative and monotone set functions ($0 \leq \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A_1) \leq \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A_2)$ for $A_1 \subset A_2$), while they are non-additive; a measure μ defined on $\mathcal{B}(\Theta)$ is non-additive if $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) \neq \mu(A_1) + \mu(A_2)$ for some $A_1 \neq A_2$, $A_1, A_2 \in \mathcal{B}(\Theta)$. A non-additive measure μ is called *submodular* or *2-alternating* if

$$\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) \leq \mu(A_1) + \mu(A_2), \quad \forall A_1, A_2 \in \mathcal{B}(\Theta). \tag{4.1}$$

If the inequality in (4.1) is reversed, μ is called *supermodular* or *2-monotone*. The *core* of a non-additive measure μ is defined by

$$core(\mu) \equiv \{\pi \text{ probability measure } : \pi(A) \geq \mu(A) \text{ holds for all } A \in \mathcal{B}(\Theta)\}. \tag{4.2}$$

The next condition concerns the supermodular (submodular) property of the lower (upper) posterior probability and the richness of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ in the sense that $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ agrees with the core of its lower probability. The latter property is called representability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ by a lower probability (Huber (1973)).

Condition 4.1 *The set of posteriors $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ satisfies the following two conditions:*

(i) The lower probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$, $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}$, is supermodular, or equivalently, the upper probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$, $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$, is submodular.

(ii) $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ is representable by its lower probability $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}$, i.e., $\Pi_{\boldsymbol{\theta}|\mathbf{Y}} = core(\pi_{\boldsymbol{\theta}|\mathbf{Y}*})$ holds.

Under Condition 4.1, we obtain the following result that expresses the bounds of the posterior mean of $h(\boldsymbol{\theta})$ on $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ in terms of the Choquet expectation of $h(\boldsymbol{\theta})$ with respect to the upper probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$. This result follows directly from Proposition 10.3 of Denneberg (1994), so we omit a proof.

Theorem 4.1 *Let $h : \Theta \to \mathbb{R}$ be a measurable real-valued function. If Condition 4.1 holds, then*

$$\begin{aligned} \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} E_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta})) &= \int h(\boldsymbol{\theta}) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}} \\ \inf_{\pi_{\boldsymbol{\theta}|Y} \in \Pi_{\boldsymbol{\theta}|Y}} E_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta})) &= -\int (-h(\boldsymbol{\theta})) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}, \end{aligned} \tag{4.3}$$

where the integral with respect to non-additive measure $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$ is defined as the Choquet integral, i.e., for measurable real-valued function $f : \Theta \to \mathbb{R}$,

$$\int f(\boldsymbol{\theta}) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}} \equiv \int_0^\infty \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\}) dt + \int_{-\infty}^0 [\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\}) - 1] dt \tag{4.4}$$

$$= \int_0^\infty \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\}) dt - \int_{-\infty}^0 \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\}) dt. \tag{4.5}$$

In many cases, it is not straightforward to check whether Condition 4.1 holds. One important case in which this condition is guaranteed to hold is for a set of posteriors whose lower probability can be represented as the containment probability of some random set and the set of posteriors includes any measurable selections of the random set. The set of posteriors for set-identified models considered by GK satisfies this condition. This claim follows from the fact that the posterior lower probability of $\Pi^{FA}_{\boldsymbol{\theta}|\mathbf{Y}}$ in (3.7) is the containment probability of the random set $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$, $\boldsymbol{\phi} \sim \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, and from Artstein's inequality (Artstein (1983)) for selectionable distributions from the random set. Hence, Theorem 4.1 always applies to robust Bayesian analysis with the GK set of priors.

The equivalence relationship (4.3) in Theorem 4.1 is valid for general posterior sets as far as they satisfy Condition 4.1, and it is not limited to the GK set of priors for set-identified models. Setting $\eta = h(\boldsymbol{\theta})$ in this theorem gives the upper and lower bounds of the posterior mean of η in terms of the Choquet integral. If the Choquet integral (or the upper posterior probability $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$) is analytically or numerically tractable in a given context, this theorem offers a general approach to deriving and computing the set of posterior moments.

Theorem 4.1 is also useful for solving the posterior Gamma minimax problem. The theorem implies that the worst-case posterior expected loss coincides with the Choquet expectation with respect to $\pi^*_{\eta|\mathbf{Y}}$. Furthermore, for the GK set of priors, the Choquet expectation of the loss given action $a \in \mathcal{A}$ can be expressed as (e.g., Theorem 5.1 in Molchanov (2005))

$$\begin{aligned}
\int L(\eta, a) d\pi^*_{\eta|\mathbf{Y}} &= \int_0^\infty \pi_{\boldsymbol{\phi}|\mathbf{Y}} \left(\{\boldsymbol{\phi} : \{\eta : L(\eta, a) \geq t\} \cap IS_\eta(\boldsymbol{\phi}) \neq \emptyset\}\right) dt \\
&= \int_0^\infty \pi_{\boldsymbol{\phi}|\mathbf{Y}} \left(\left\{ \boldsymbol{\phi} : \sup_{\eta \in IS_\eta(\boldsymbol{\phi})} \{L(\eta, a)\} \geq t \right\} \right) dt \\
&= \int_\Phi \sup_{\eta \in IS_\eta(\boldsymbol{\phi})} L(\eta, a) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}).
\end{aligned}$$

We hence obtain the following theorem for the representation of the posterior Gamma minimax criterion:

Theorem 4.2 *Let $L(\eta, a)$ be a nonnegative loss function (e.g., the quadratic loss $L(\eta, a) = (\eta - a)^2$) and $\Pi_{\boldsymbol{\theta}}$ be the set of priors constructed in (3.4). With the set of posteriors $\Pi^{FA}_{\eta|\mathbf{Y}}$ obtained in (3.6), the upper posterior expected loss at action a satisfies*

$$\sup_{\pi_{\boldsymbol{\theta}} \in \Pi^{FA}_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a) = \int L(\eta, a) d\pi^*_{\eta|\mathbf{Y}} = \int_\Phi \sup_{\eta \in IS_\eta(\boldsymbol{\phi})} L(\eta, a) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \tag{4.6}$$

provided that the Choquet integral is finite, $\int L(\eta, a) d\pi^*_{\eta|\mathbf{Y}}(\eta) < \infty$.

The posterior Gamma minimax criterion shown in (4.6) combines the ambiguity about η (given $\boldsymbol{\phi}$, what we know about η is only that it lies within the identified set $IS_\eta(\boldsymbol{\phi})$) with the posterior uncertainty about the identified set $IS_\eta(\boldsymbol{\phi})$ (in finite samples, the identified set of η is known with some uncertainty as summarized by the posterior of $\boldsymbol{\phi}$). Since Theorem 4.2 imposes no assumption on the loss function other than its nonnegativity, this result is also applicable to a planner's policy decision problem under a set-identified social welfare criterion. See Manski (2000) for statistical decision theory applied to treatment choice under a set-identified welfare criterion.

Theorem 4.2 also suggests a simple numerical algorithm for computing the posterior Gamma minimax action using a Monte Carlo sample of $\boldsymbol{\phi}$ from its posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$. Let $\{\boldsymbol{\phi}_s\}_{s=1}^S$ be S random draws of $\boldsymbol{\phi}$ from the posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$. Then, the posterior Gamma minimax action $\hat{a}^* \in \arg\min_a \left\{\sup_{\pi_{\boldsymbol{\theta}} \in \Pi^{FA}_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a)\right\}$ can be approximated by

$$\hat{a}^* \in \arg\min_a \frac{1}{S} \sum_{s=1}^{S} \sup_{\eta \in IS_\eta(\boldsymbol{\phi}_s)} L(\eta, a).$$

The posterior Gamma minimax action does not generally coincide with an unconditional optimal Gamma minimax decision. This is also the case with the prior set (3.4), implying that $\hat{a}^*$ fails to be a Bayesian action with respect to any single prior in the set.

5 Robust Bayesian inference in SVARs

This section discusses the approaches to conducting robust Bayesian analysis in SVARs in GK, GKV and GLU. We first describe the SVAR framework and outline some commonly used identifying restrictions. We assume the parameter of interest is an individual impulse response, but the approaches easily extend to other parameters, such as forecast error variance decompositions.

5.1 Setup

Consider an SVAR(p) for the n-dimensional vector $\mathbf{y}_t$:

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \quad \text{for } t = 1, \ldots, T, \tag{5.1}$$

with $\boldsymbol{\varepsilon}_t$ a vector white noise process, normally distributed with mean zero and variance the identity matrix $\mathbf{I}_n$. The initial conditions $\mathbf{y}_1, \ldots, \mathbf{y}_p$ are treated as given.

The reduced-form VAR(p) model is

$$\mathbf{y}_t = \mathbf{b} + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{y}_{t-j} + \mathbf{u}_t, \tag{5.2}$$

where $\mathbf{b} = \mathbf{A}_0^{-1}\mathbf{a}$, $\mathbf{B}_j = \mathbf{A}_0^{-1}\mathbf{A}_j$, $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t$, and $\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right) \equiv \boldsymbol{\Sigma} = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$. The reduced-form parameter is $\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})')' \in \boldsymbol{\Phi}$, where $\mathbf{B} = [\mathbf{b}, \mathbf{B}_1, \ldots, \mathbf{B}_p]$. Let $\mathbf{Y}$ denote the sample.

In SVAR applications, we can set the structural parameter vector $\boldsymbol{\theta}$ considered in the general framework of the previous section as $\boldsymbol{\theta} = (\boldsymbol{\phi}', \text{vec}(\mathbf{Q})')'$, where $\mathbf{Q}$ is an $n \times n$ orthonormal matrix in the set $\mathcal{O}(n)$ of orthonormal matrices (see, e.g., Uhlig (2005) and Rubio-Ramírez et al. (2010)). As far as $\mathbf{A}_0$ is invertible, $\boldsymbol{\theta}$ transforms the SVAR structural parameters $[\mathbf{A}_0, \mathbf{a}, \mathbf{A}_1, \ldots, \mathbf{A}_p]$ as $\mathbf{B} = \mathbf{A}_0^{-1}[\mathbf{a}, \mathbf{A}_1, \ldots, \mathbf{A}_p]$, $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$ and $\mathbf{Q} = \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$, where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of Σ with nonnegative diagonal elements. This transformation is one-to-one and can be inverted as $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ and $[\mathbf{a}, \mathbf{A}_1, \ldots, \mathbf{A}_p] = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{B}$.

We assume that the reduced-form VAR(p) model can be inverted into a VMA(∞) model:

$$\mathbf{y}_t = \mathbf{c} + \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{u}_{t-j} = \mathbf{c} + \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_{t-j},$$

where $\mathbf{C}_j$ is the j-th coefficient matrix of $\left(\mathbf{I}_n - \sum_{j=1}^{p}\mathbf{B}_j L^j\right)^{-1}$.

The h-th horizon impulse response is the $n \times n$ matrix $\mathbf{IR}^h$, $h = 0, 1, 2, \ldots$

$$\mathbf{IR}^h = \mathbf{C}_h \boldsymbol{\Sigma}_{tr}\mathbf{Q}, \tag{5.3}$$

and the long-run cumulative impulse-response matrix is

$$\mathbf{CIR}^{\infty} = \sum_{h=0}^{\infty}\mathbf{IR}^h = \left(\sum_{h=0}^{\infty}\mathbf{C}_h\right)\boldsymbol{\Sigma}_{tr}\mathbf{Q} = \left(\mathbf{I}_n - \sum_{j=1}^{p}\mathbf{B}_j\right)^{-1}\boldsymbol{\Sigma}_{tr}\mathbf{Q}. \tag{5.4}$$

The scalar parameter of interest η is the impulse-response, i.e., the (i,j)-element of $\mathbf{IR}^h$:

$$\eta = IR_{ij}^h \equiv \mathbf{e}_i'\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_j \equiv \mathbf{c}_{ih}'(\boldsymbol{\phi})\,\mathbf{q}_j = h(\boldsymbol{\phi}, \mathbf{Q}), \tag{5.5}$$

where $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_n$ and $\mathbf{c}_{ih}'(\boldsymbol{\phi})$ is the i-th row of $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$.

5.2 Set-identifying restrictions in SVARs

An SVAR without identifying restrictions is set-identified because there are multiple values of $\mathbf{A}_0$ that are consistent with $\boldsymbol{\phi}$: $\left\{\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1} : \mathbf{Q} \in \mathcal{O}(n)\right\}$. Imposing zero and/or sign restrictions can be viewed as constraining the set of orthonormal matrices to lie in a subspace $\mathcal{Q}(\boldsymbol{\phi})$ of $\mathcal{O}(n)$, which in turn yields the following identified set for the impulse-response η:

$$IS_{\eta}(\boldsymbol{\phi}) = \left\{\eta(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi})\right\}. \tag{5.6}$$

We now characterize the subspace $\mathcal{Q}(\boldsymbol{\phi})$ under typical zero restrictions in SVARs, under restrictions induced by external instruments in proxy SVARs and under sign restrictions. In addition to any such restrictions, we assume one always imposes the sign normalization restrictions

$\text{diag}(\mathbf{A}_0) = \text{diag}\left(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1}$, which are inequalities involving the columns $\boldsymbol{\sigma}^i$ of $\boldsymbol{\Sigma}_{tr}^{-1}$ and $\mathbf{q}_i$ of $\mathbf{Q}$:

$$\left(\boldsymbol{\sigma}^i\right)' \mathbf{q}_i \geq 0 \quad \text{for all } i = 1, \ldots, n. \tag{5.7}$$

5.2.1 Zero restrictions

Commonly used zero restrictions in SVARs can be written as linear constraints on the columns of $\mathbf{Q}$. For example:

$$\begin{aligned}
((i,j)\text{-th element of } \mathbf{A}_0) &= 0 \iff \left(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_j\right)' \mathbf{q}_i = 0, \\
\left((i,j)\text{-th element of } \mathbf{A}_0^{-1}\right) &= 0 \iff \left(\mathbf{e}_i'\boldsymbol{\Sigma}_{tr}\right) \mathbf{q}_j = 0, \\
((i,j)\text{-th element of } \mathbf{CIR}^{\infty}) &= 0 \iff \left[\mathbf{e}_i' \left(\mathbf{I}_n - \sum_{j=1}^{p} \mathbf{B}_j\right)^{-1} \boldsymbol{\Sigma}_{tr}\right] \mathbf{q}_j = 0.
\end{aligned} \tag{5.8}$$

We represent a collection of zero restrictions as:

$$F(\boldsymbol{\phi}, \mathbf{Q}) \equiv \begin{pmatrix} F_1(\boldsymbol{\phi})\,\mathbf{q}_1 \\ F_2(\boldsymbol{\phi})\,\mathbf{q}_2 \\ \vdots \\ F_n(\boldsymbol{\phi})\,\mathbf{q}_n \end{pmatrix} = \mathbf{0}_{\sum_{i=1}^n f_i \times 1}, \tag{5.9}$$

with $F_i(\boldsymbol{\phi})$ an $f_i \times n$ matrix. We assume that the imposed zero restrictions satisfy $f_i \leq n - i$, $i = 1, \ldots, n$, so that we rule out the over-identifying zero restrictions and locally- but not globally-identified SVARs (Bacchiocchi and Kitagawa (2020)). We assume that the variables in the model are ordered such that the number of zero restrictions f_i imposed on the i-th column of $\mathbf{Q}$ satisfies $f_1 \geq f_2 \geq \cdots \geq f_n \geq 0$. In case of ties, if the impulse response of interest is that to the j-th structural shock, one should order the j-th variable first. If there are only sign restrictions, one should order first the variable whose structural shock is of interest.

The subspace $\mathcal{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then given by

$$\mathcal{Q}(\boldsymbol{\phi}) = \left\{\mathbf{Q} \in \mathcal{O}(n) : F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^n f_i \times}, \text{diag}\left(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1}\right\}. \tag{5.10}$$

5.2.2 Exogeneity restrictions in Proxy SVARs

Proxy SVARs rely on the assumption that there are instruments ('proxies') external to the SVAR that are correlated with particular structural shocks ('relevant') and uncorrelated with other shocks ('exogenous').[4] Set-identification arises in these models when there are multiple proxies for multiple shocks. Giacomini, Kitagawa and Read (in press) propose a robust Bayesian approach to inference

[4] See, for example, Mertens and Ravn (2013) and Stock and Watson (2018).

in this context that starts by writing the restrictions arising from exogeneity of the proxies as linear constraints on the columns of $\mathbf{Q}$, as follows.

Let $\boldsymbol{\varepsilon}_{(i:j),t} = (\boldsymbol{\varepsilon}_{i,t}, \boldsymbol{\varepsilon}_{i+1,t}, ..., \boldsymbol{\varepsilon}_{j-1,t}, \boldsymbol{\varepsilon}_{j,t})'$ for $i < j$. Assume that $\mathbf{m}_t$ is a $k \times 1$ vector of proxies (with $k < n$) that are correlated with the last k structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(n-k+1:n),t}) = \boldsymbol{\Psi}$ with $\boldsymbol{\Psi}$ a full-rank matrix, and uncorrelated with the first $n-k$ structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:n-k),t}) = \mathbf{0}_{k\times(n-k)}$. We assume that $\mathbf{m}_t$ follows an SVAR(p_m) with $\boldsymbol{\varepsilon}_t$ included as exogenous variables:

$$\boldsymbol{\Gamma}_0 \mathbf{m}_t = \boldsymbol{\gamma} + \boldsymbol{\Lambda} \boldsymbol{\varepsilon}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\nu}_t, \quad t = 1, ..., T, \tag{5.11}$$

with $\boldsymbol{\Gamma}_0$ invertible and $(\boldsymbol{\varepsilon}'_t, \boldsymbol{\nu}'_t)' | \mathcal{F}_{t-1} \sim N(0, \mathbf{I}_{n+k})$, where $\mathcal{F}_{t-1}$ is the information set at time $t-1$. Consider the 'first-stage regression'

$$\mathbf{m}_t = \mathbf{g} + \mathbf{D}\mathbf{y}_t + \mathbf{G}\mathbf{x}_t + \sum_{l=1}^{p_m} \mathbf{H}_l \mathbf{m}_{t-l} + \mathbf{v}_t, \tag{5.12}$$

where $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ and $\mathbb{E}(\mathbf{v}_t \mathbf{v}'_t) = \boldsymbol{\Upsilon}$. Since $\boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} = \mathbf{D}\mathbf{A}_0^{-1} = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, the instrument validity conditions imply that

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_t) = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q} = \left[\mathbf{0}_{k\times(n-k)}, \boldsymbol{\Psi}\right]. \tag{5.13}$$

The (i,j)th element of this matrix is $\mathbf{e}'_{i,k}\mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n} = \mathbf{d}'_i \mathbf{q}_j$, where $\mathbf{d}'_i \equiv \mathbf{e}'_{i,k}\mathbf{D}\boldsymbol{\Sigma}_{tr}$ is the ith row of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$. The exogeneity conditions in a proxy SVAR therefore generate linear zero restrictions on the first $n-k$ columns of $\mathbf{Q}$ given $\mathbf{D}$ and $\boldsymbol{\Sigma}_{tr}$. Similarly to the previous subsection, we can write these restrictions in the general form $F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{k(n-k)\times 1}$, where the reduced-form parameter is now

$$\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})', \text{vec}(\mathbf{g})', \text{vec}(\mathbf{D})', \text{vec}(\mathbf{G})', \text{vec}(\mathbf{H}_1)', \ldots, \text{vec}(\mathbf{H}_{p_m})', \text{vech}(\boldsymbol{\Upsilon})')'.$$

The subspace $\mathcal{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then defined as in (5.10).

5.2.3 Sign restrictions

Sign restrictions on impulse-responses can also be written as linear constraints on the columns of $\mathbf{Q}$: $S_{hj}(\boldsymbol{\phi})\,\mathbf{q}_j \geq \mathbf{0}_{s_{jh}\times 1}$,[5] where $S_{hj}(\boldsymbol{\phi}) \equiv \mathbf{D}_{hj}\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$, with $\mathbf{D}_{hj}$ a matrix that selects the sign-restricted responses from $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{q}_j$, with nonzero element 1 or -1 depending on whether the responses are positive or negative. By stacking $S_{hj}(\boldsymbol{\phi})$ over multiple horizons we obtain the set of sign restrictions on the responses to the j-th shock, $S_j(\boldsymbol{\phi})\,\mathbf{q}_j \geq \mathbf{0}_{s_j\times 1}$. We represent a collection of sign restrictions, $\{S_j(\boldsymbol{\phi})\,\mathbf{q}_j \geq \mathbf{0}_{s_j\times 1} \text{ for } j \in \mathcal{I}_S\}$ as

$$S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}, \tag{5.14}$$

[5] For $\mathbf{y} = (y_1, \ldots, y_m)'$, $\mathbf{y} \geq 0$ means $y_i \geq 0$ for all $i = 1, 2, \ldots, m$.

where $\mathcal{I}_S \subset \{1, 2, \ldots, n\}$ is such that $j \in \mathcal{I}_S$ if some responses to the j-th shock are restricted.

The subspace $\mathcal{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then given by

$$\mathcal{Q}(\boldsymbol{\phi}) = \left\{ \mathbf{Q} \in \mathcal{O}(n) : S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}, \operatorname{diag}\left(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1} \right\}. \tag{5.15}$$

Sign restrictions on other parameters, such as elements of $\mathbf{A}_0$, $\mathbf{CIR}^{\infty}$, or (in proxy SVARs) $\boldsymbol{\Psi}$, can be imposed similarly.

5.3 Multiple Priors in SVARs

In a standard Bayesian approach to SVAR estimation, one typically specifies a prior for $\boldsymbol{\theta} = (\boldsymbol{\phi}, \mathbf{Q})$, $\pi_{\boldsymbol{\theta}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}}$, by specifying single priors $\pi_{\boldsymbol{\phi}}$ and $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$. As discussed in Section 3, the former is updated by the data, while the latter is not updated (see also Baumeister and Hamilton (2015)).

Maintaining the choice of a single prior for $\boldsymbol{\phi}$, we discuss three approaches that rely on different specifications for the set of priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$: the full ambiguity approach of GK, which applies the set of priors of Section 3.3; the model-averaging approach of GKV, which can be viewed as robust Bayesian analysis with an ϵ-contaminated set (Example 2); and the robust control approach of GKU, which introduces the KL-neighborhood set of conditional priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$.

5.3.1 Full ambiguity (GK)

In terms of the current notation for SVAR applications, the set of conditional priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$ that represents full ambiguity for the unrevisable component of the prior can be represented as

$$\Pi_{\boldsymbol{\theta}}^{FA} = \{\pi_{\boldsymbol{\theta}} = \int \pi_{\mathbf{Q}|\boldsymbol{\phi}} d\pi_{\boldsymbol{\phi}}(\boldsymbol{\phi}) : \pi_{\mathbf{Q}|\boldsymbol{\phi}} \in \Pi_{\mathbf{Q}|\boldsymbol{\phi}}\}, \tag{5.16}$$

where $\Pi_{\mathbf{Q}|\boldsymbol{\phi}}$ allows all conditional priors supported on the subspace $\mathcal{Q}(\boldsymbol{\phi})$ of orthonormal matrices,

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}}^{FA} = \left\{ \pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}(\mathcal{Q}(\boldsymbol{\phi})) = 1,\ \pi_{\boldsymbol{\phi}}\text{-almost surely} \right\}. \tag{5.17}$$

As shown in (3.6), the resulting set of posteriors for the impulse response of interest $\eta = h(\mathbf{Q}, \boldsymbol{\phi})$ is

$$\Pi_{\eta|\mathbf{Y}}^{FA} = \left\{ \pi_{\eta|\mathbf{Y}}(\cdot) = \int \pi_{\mathbf{Q}|\phi}\left(h(\mathbf{Q}, \boldsymbol{\phi}) \in \cdot\right) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} : \pi_{\mathbf{Q}|\phi} \in \Pi_{\mathbf{Q}|\phi}^{FA} \right\}. \tag{5.18}$$

The general formulae for the set of posterior means (3.8) and the robust credible regions (3.9) shown in Section 3.3 can apply as they are. See Section 6.1 for algorithms to compute these quantities.

5.3.2 Model averaging (GKV)

GKV consider a set of priors for the impulse-response that averages single-prior and multiple-prior models. Focusing on the case of two models, the approach could be viewed as a refinement to GK when the researcher has access to a single prior for the impulse-response; for example, that implied by a prior on the SVAR's structural parameters (e.g., Baumeister and Hamilton (2015)) or a prior based on a Bayesian DSGE model. Another example is when a set of restrictions give point-identification but some of the restrictions are controversial. The single prior in this case corresponds to the point-identified model that imposes all restrictions, while the multiple-prior model corresponds to a set-identified model that relaxes the controversial restrictions.

Let M^s be the single-prior model and M^m the multiple-prior (set-identified) model, with corresponding prior probabilities $\pi_{M^s} \in [0,1]$ and $1-\pi_{M^s}$. The single-prior model admits a unique prior for $\boldsymbol{\theta}$, $\pi_{\boldsymbol{\theta}|M^s}$, while the input of the multiple-prior model is the GK set of priors for $\boldsymbol{\theta}$, $\Pi_{\boldsymbol{\theta}|M^m}$, given a unique prior for $\boldsymbol{\phi}$, $\pi_{\boldsymbol{\phi}|M^m}$. GKV obtain a set of posteriors for the impulse-response that combine the single posterior in model M^s and the set of posteriors in model M^m according to the posterior model probabilities.

This practice of averaging the single-prior (or point-identified) model and the multiple-prior (set-identified) model can be viewed as a robust Bayesian analysis with the following set of priors:

$$\Pi_{\boldsymbol{\theta}}^{Avg} \equiv \left\{ \pi_{\boldsymbol{\theta}} = \pi_{\boldsymbol{\theta}|M^p}\pi_{M^p} + \pi_{\boldsymbol{\theta}|M^s}\pi_{M^s} : \pi_{\boldsymbol{\theta}|M^s} \in \Pi_{\boldsymbol{\theta}|M^s}^{FA} \right\}. \tag{5.19}$$

This set of priors takes the form of an ϵ-contaminated set of priors as in (2.12), where the benchmark prior is from the single-prior (point-identified) model $\pi_{\boldsymbol{\theta}}^0 = \pi_{\boldsymbol{\theta}|M^s}$, the amount of contamination is the prior model probability assigned to the set-identified model $\epsilon = \pi_{M^m}$ and $\mathcal{Q}_{\boldsymbol{\theta}}$ corresponds to the full-ambiguity set of priors for the set-identified model $\Pi_{\boldsymbol{\theta}|M^s}^{FA}$. That is, if the single-prior (or point-identified) model is a possibly misspecified benchmark, averaging it with the set-identified model can be interpreted as performing Bayesian sensitivity analysis with respect to a contamination of the benchmark model by an amount π_{M^m} in every possible direction, while maintaining the set-identifying restrictions in M^m.

GKV show that the posterior model probabilities differ from the prior probabilities if the models are 'distinguishable' for some values of $\boldsymbol{\phi}$ and/or the two models consider different priors for $\boldsymbol{\phi}$. Models are distinguishable if they imply different reduced-form parameter spaces. Models that admit the same reduced-form representation (i.e., a VAR with the same variables and lag length) but differ in the identifying restrictions they impose are distinguishable if the restrictions rule out different values of $\boldsymbol{\phi}$ (for example, by yielding an empty identified set for some values of $\boldsymbol{\phi}$).

The posterior model probabilities are obtained as

$$\pi_{M^s|\mathbf{Y}} = \frac{p(\mathbf{Y}|M^s)\cdot\pi_{M^s}}{p(\mathbf{Y}|M^s)\cdot\pi_{M^s}+p(\mathbf{Y}|M^m)\cdot(1-\pi_{M^s})},$$
$$\pi_{M^m|\mathbf{Y}} = \frac{p(\mathbf{Y}|M^m)\cdot(1-\pi_{M^s})}{p(\mathbf{Y}|M^s)\cdot\pi_{M^s}+p(\mathbf{Y}|M^m)\cdot(1-\pi_{M^s})}, \tag{5.20}$$

where $p(\mathbf{Y}|M) \equiv \int p(\mathbf{Y}|\boldsymbol{\phi}, M)d\pi_{\boldsymbol{\phi}|M}(\boldsymbol{\phi})$ is the marginal likelihood of model M with $p(\mathbf{Y}|\boldsymbol{\phi}, M)$ the likelihood of the reduced-form parameter.

The set of posteriors can again be summarized by reporting a set of posterior means and a robust credible region. The set of posterior means for η is the weighted average of the posterior mean in model M^s and the set of posterior means in model M^m:

$$\begin{aligned}&\left[\inf_{\pi_{\eta|\mathbf{Y}}\in\Pi_{\eta|\mathbf{Y}}} E_{\eta|\mathbf{Y}}(\eta), \sup_{\pi_{\eta|\mathbf{Y}}\in\Pi_{\eta|\mathbf{Y}}} E_{\eta|\mathbf{Y}}(\eta)\right]\\ =&\pi_{M^s|\mathbf{Y}}E_{\eta|M^s,\mathbf{Y}}(\eta) + \pi_{M^m|\mathbf{Y}}\left[E_{\boldsymbol{\phi}|M^m,\mathbf{Y}}(\ell(\boldsymbol{\phi})), E_{\boldsymbol{\phi}|M^m,\mathbf{Y}}(u(\boldsymbol{\phi}))\right],\end{aligned} \tag{5.21}$$

where $(\ell(\boldsymbol{\phi}), u(\boldsymbol{\phi}))$ are as defined in (3.8) and $E_{\boldsymbol{\phi}|M^m,\mathbf{Y}}(\cdot)$ is the posterior mean with respect to the $\boldsymbol{\phi}$-prior in model M^m. Section 6.2 discusses how to compute the set of posterior moments and the robust credible regions.

A potentially useful analysis that can be carried out in this context is a reverse-engineering exercise that computes the prior weight w one would assign to the restrictions in M^s to obtain a given conclusion. For example, to find the smallest weight such that the set of posterior means is contained in the positive real half-line, one solves for w in the equation

$$\begin{aligned}\frac{p(\mathbf{Y}|M^s)\cdot w}{p(\mathbf{Y}|M^s)\cdot w+p(\mathbf{Y}|M^m)\cdot(1-w)}&E_{\eta|M^s,\mathbf{Y}}(\eta)\\ &+\frac{p(\mathbf{Y}|M^m)\cdot(1-w)}{p(\mathbf{Y}|M^s)\cdot w+p(\mathbf{Y}|M^m)\cdot(1-w)}E_{\boldsymbol{\phi}|M^m,\mathbf{Y}}(\ell(\boldsymbol{\phi}))=0.\end{aligned} \tag{5.22}$$

5.3.3 KL-neighborhood (GKU)

GKU consider a refinement of the set of priors in GK. The starting point is the availability of a benchmark conditional prior for $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$, $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. By considering the set of priors in a KL neighborhood of the benchmark prior with a given radius $\lambda > 0$,

$$\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda) = \left\{\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} : \int_\Theta \ln\left(\frac{d\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}}{d\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}}\right)d\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\boldsymbol{\theta}) \le \lambda,\ \pi_{\boldsymbol{\phi}}\text{–almost surely}\right\}, \tag{5.23}$$

one can obtain a set of posteriors for the impulse-response η as

$$\Pi^{KL}_{\eta|\mathbf{Y}}(\lambda) = \left\{\pi_{\eta|\mathbf{Y}}(\cdot) = \int \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta})\in\cdot)d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}) : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)\right\}. \tag{5.24}$$

$\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ refines $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ in the following aspects. First, by choosing a partially credible benchmark prior $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, one can anchor the set of priors to the plausible one, disregarding from $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ those that are far from the benchmark prior. Second, any prior $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ is absolutely continuous with respect to the benchmark prior. Hence the support of $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ is contained in that of $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, and they share a common dominating measure, implying that one can constrain the support of $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ by the choice of $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. Third, the choice of $\lambda > 0$ conveniently controls the size of the prior set. Specifically, in terms of the set of posterior means spanned, varying λ from 0 to ∞ lets $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ vary from a single-prior Bayes approach under the benchmark prior to the multiple-prior Bayes approach under $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. GKU suggest eliciting λ by assessing the set of prior means of η or other parameters that $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ spans and matching it with the researcher's partial prior knowledge.

The set of posteriors obtained in (5.24) can be used for sensitivity analysis by reporting, for example, the set of posterior means of a function of interest $f(\eta)$ (e.g., $f(\eta) = \eta$ or $f(\eta) = 1\{\eta \in D\}$). GKU show that this set of posterior means is given by

$$\left[\int_\Phi \left(\int_{-\infty}^{\infty} f(\eta) d\pi^{\ell}_{\eta|\boldsymbol{\phi}}(\eta)\right) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \int_\Phi \left(\int_{-\infty}^{\infty} f(\eta) d\pi^{u}_{\eta|\boldsymbol{\phi}}(\eta)\right) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi})\right], \tag{5.25}$$

where $\pi^{\ell}_{\eta|\boldsymbol{\phi}}$ and $\pi^{u}_{\eta|\boldsymbol{\phi}}$ are obtained by exponential tilting of the benchmark priors,

$$\begin{aligned}
d\pi^{\ell}_{\eta|\boldsymbol{\phi}} &\equiv \frac{\exp\{-f(\eta)/\kappa^{\ell}_{\lambda}(\boldsymbol{\phi})\}}{\int \exp\{-f(\eta)/\kappa^{\ell}_{\lambda}(\phi)\} d\pi^0_{\eta|\boldsymbol{\phi}}} \cdot d\pi^0_{\eta|\boldsymbol{\phi}}, \\
d\pi^{u}_{\eta|\boldsymbol{\phi}} &\equiv \frac{\exp\{f(\eta)/\kappa^{u}_{\lambda}(\boldsymbol{\phi})\}}{\int \exp\{f(\eta)/\kappa^{u}_{\lambda}(\boldsymbol{\phi})\} d\pi^0_{\eta|\boldsymbol{\phi}}} \cdot d\pi^0_{\eta|\boldsymbol{\phi}}, \\
\kappa^{\ell}_{\lambda}(\boldsymbol{\phi}) &\equiv \arg\min_{\kappa \geq 0} \left\{\kappa \ln \int \exp\left\{\frac{-f(\eta)}{\kappa}\right\} d\pi^0_{\eta|\boldsymbol{\phi}}(\eta) + \kappa\lambda\right\}, \\
\kappa^{u}_{\lambda}(\boldsymbol{\phi}) &\equiv \arg\min_{\kappa \geq 0} \left\{\kappa \ln \int \exp\left\{\frac{f(\eta)}{\kappa}\right\} d\pi^0_{\eta|\boldsymbol{\phi}}(\eta) + \kappa\lambda\right\},
\end{aligned} \tag{5.26}$$

where $\pi^0_{\eta|\boldsymbol{\phi}}$ is the benchmark conditional prior for η given $\boldsymbol{\phi}$ obtained by marginalizing $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ to η. See Section 6.3 for how to compute these bounds.

The posterior mean upper bound obtained in (5.25) is also useful for solving the posterior Gamma minimax problem. For instance, let $\delta(\mathbf{Y})$ be an estimator for η and $L(\delta(\mathbf{Y}), \eta)$ be an estimation loss function. The posterior Gamma minimax estimator $\delta_\lambda(\mathbf{Y})$ with prior set $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ can be obtained by

$$\delta_\lambda(\mathbf{Y}) \in \arg\min_a \int_\Phi \left[\int_{IS_\eta(\boldsymbol{\phi})} L(a, \eta) d\pi^{u}_{\eta|\boldsymbol{\phi}}(\eta)\right] d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \tag{5.27}$$

where

$$d\pi^{u}_{\eta|\boldsymbol{\phi}} = \frac{\exp\left\{L(a, \eta)/\kappa_\lambda(a, \boldsymbol{\phi})\right\}}{\int_{IS_\eta(\boldsymbol{\phi})} \exp\left\{L(a, \eta)/\kappa_\lambda(a, \boldsymbol{\phi})\right\} d\pi^0_{\eta|\boldsymbol{\phi}}} \cdot d\pi^0_{\eta|\boldsymbol{\phi}}$$

and $\kappa_\lambda(a, \boldsymbol{\phi}) > 0$ is the unique solution to the following convex minimization:

$$\min_{\kappa \geq 0} \left\{ \kappa \ln \int_{IS_\eta(\boldsymbol{\phi})} \exp\left\{ \frac{L(a,\eta)}{\kappa} \right\} d\pi^0_{\eta|\boldsymbol{\phi}} + \kappa\lambda \right\}.$$

We emphasize that $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ constructed above is distinct from the KL-neighborhood set for the unconditional prior discussed in Example 5. The main difference is that the set of priors in (2.15) allows multiple priors not only for the conditional prior of $\boldsymbol{\theta}|\boldsymbol{\phi}$ but also for the marginal prior of $\boldsymbol{\phi}$. Having multiple priors for $\boldsymbol{\phi}$ enables one to assess posterior sensitivity with respect to the prior for the identifiable reduced-form parameter, but masks the shape of the posterior distributions if the set contains a prior that fits the data poorly. This is because obtaining a large set of posteriors could be due to some priors for $\boldsymbol{\phi}$ that are severely in conflict with the data, rather than indicating a lack of information in the observed likelihood. With $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$, in contrast, all the posteriors in the set share the same value of the marginal likelihood, so we can assess posterior sensitivity while keeping the denominator of Bayes' rule constant.

5.4 Frequentist properties of the robust Bayesian approach

The methods described in the previous sections are valid from a Bayesian perspective as tools for robust Bayesian sensitivity analysis, but it can also be important to understand their frequentist properties. This section briefly summarizes the asymptotic frequentist properties of the multiple-prior estimation and inference procedures covered in this paper, and overviews other approaches. See the individual papers for precise regularity conditions, formal statements of the frequentist results, and proofs.

In the setting considered in Section 3, GK examine whether the robust Bayesian approach restores the asymptotic equivalence between frequentist and Bayesian inference – an equivalence that breaks down under set-identification if one adopts a single-prior Bayesian approach (Moon and Schorfheide (2012)). Under the assumptions that the Bernstein-von Mises property holds for estimation of the reduced-form parameter and that the identified-set mapping is convex, continuous and differentiable, the set of posterior means is consistent and the robust credible region has valid frequentist coverage for the true identified set asymptotically. GK also provide primitive conditions in SVARs under which the high-level conditions are satisfied; these typically require checking the pattern of zero and sign restrictions imposed. See also Liao and Simoni (2013) and Kline and Tamer (2016) for Bernstein-von Mises results of the Bayesian confidence sets for the identified set. Chen et al. (2018) develop a Monte Carlo-based confidence interval that projects a highest density posterior credible region for the parameter vector $\boldsymbol{\theta}$, and show its asymptotic frequentist validity without requiring differentiability of the identified-set mapping.

Giacomini et al. (in pressa) provide conditions for asymptotic frequentist validity of the robust Bayesian approach in the case of proxy SVARs, and discuss how the case of proxies that are only

weakly correlated with the structural shocks ('weak proxies') affects the asymptotic frequentist properties of the GK robust Bayesian credible sets.

The asymptotic frequentist validity of robust Bayesian inference shown in GK reveals that for set-identified models, ambiguity about the parameters represented by $\Pi_{\boldsymbol{\theta}}^{FA}$ can match the absence or removal of a prior in frequentist inference. This means that robust Bayesian inference under a set of priors Π_{θ} that is a strict subset of $\Pi_{\boldsymbol{\theta}}^{FA}$ generally leads to more informative inference than frequentist inference when the models are set-identified. Accordingly, the robust credible regions obtained under the GKV set of priors $\Pi_{\boldsymbol{\theta}}^{Avg}$ and the GKU set of priors $\Pi_{\boldsymbol{\theta}}^{KL}(\lambda)$, $\lambda < \infty$, yield posterior inference that is too optimistic in terms of frequentist coverage.

6 Numerical Implementation

This section explains how to numerically implement the three approaches described in the previous section. We emphasise the key choices that practitioners face when implementing the algorithms.

6.1 Full ambiguity (GK)

We present a general algorithm to numerically approximate the set of posterior means and the robust credible region. The algorithm assumes that $f_i \leq n - i$ for all $i = 1, \ldots, n$.

Algorithm 1. *Let $F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}$ and $S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}$ be the set of identifying restrictions, and let $\eta = \mathbf{c}_{ih}'(\boldsymbol{\phi})\,\mathbf{q}_{j^*}$ be the impulse response of interest.*

- **Step 1:** *Specify a prior for the reduced-form parameter, $\tilde{\pi}_{\boldsymbol{\phi}}$.*
- **Step 2:** *Draw $\boldsymbol{\phi}$ from its posterior, $\tilde{\pi}_{\boldsymbol{\phi}|\mathbf{Y}}$, and check whether the set of orthonormal matrices satisfying the identifying restrictions, $\mathcal{Q}(\boldsymbol{\phi})$, is empty. If so, repeat Step 2. Otherwise, proceed to Step 3.*
- **Step 3:** *Given $\boldsymbol{\phi}$ obtained in Step 2, compute the lower bound, $\ell(\boldsymbol{\phi})$, and upper bound, $u(\boldsymbol{\phi})$, of the identified set for η, $IS_{\eta}(\boldsymbol{\phi})$. $\ell(\boldsymbol{\phi})$ is defined by the following minimisation problem:*

$$\begin{aligned} & \ell(\boldsymbol{\phi}) = \arg\min_{\mathbf{Q}} \mathbf{c}_{ih}'(\boldsymbol{\phi})\,\mathbf{q}_j^*, \\ s.t. \quad & \mathbf{Q}'\mathbf{Q} = \mathbf{I}_n, \;\; F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}, \;\; \operatorname{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1}, \;\; S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}, \end{aligned}$$

  and $u(\boldsymbol{\phi}) = \arg\max_{\mathbf{Q}} \mathbf{c}_{ih}'(\boldsymbol{\phi})\,\mathbf{q}_j^*$ under the same set of constraints.
- **Step 4:** *Repeat Steps 2–3 M times to obtain $[l(\boldsymbol{\phi}_m), u(\boldsymbol{\phi}_m)]$, $m = 1, \ldots, M$. Approximate the set of posterior means by the sample averages of $l(\boldsymbol{\phi}_m)$ and $u(\boldsymbol{\phi}_m)$.*

- **Step 5:** *To obtain an approximation of the smallest robust credible region with credibility* $\alpha \in (0,1)$*, define* $d(\eta,\phi) = \max\{|\eta - \ell(\boldsymbol{\phi})|, |\eta - u(\boldsymbol{\phi})|\}$*, and let* $\hat{z}_\alpha(\eta)$ *be the sample* α*-th quantile of* $(d(\eta,\boldsymbol{\phi}_m) : m = 1,\dots,M)$*. An approximated smallest robust credible region for* η *is an interval centered at* $\arg\min_\eta \hat{z}_\alpha(\eta)$ *with radius* $\min_\eta \hat{z}_\alpha(\eta)$*.*[6]

The method used to draw $\boldsymbol{\phi}$ from its posterior in Step 2 depends on the prior specified in Step 1 (which may be improper). A commonly used prior for $\boldsymbol{\phi}$ is the normal-inverse-Wishart (e.g., Arias et al. (2018)); this prior induces a normal-inverse-Wishart posterior, from which it is easy to obtain independent draws (e.g., Del Negro and Schorfheide (2011)). It is also possible to apply this algorithm when the prior is specified for the structural parameters rather than the reduced-form parameters, provided that the prior for the structural parameters embeds exact zero restrictions and/or dogmatic sign restrictions (e.g., Baumeister and Hamilton (2015)). In this case, draws of the structural parameters (e.g., obtained via Markov Chain Monte Carlo methods) can be transformed into draws of the reduced-form parameters. When the identified set is empty at some values of $\boldsymbol{\phi}$ that receive positive prior probability under $\tilde{\pi}_{\boldsymbol{\phi}}$, the prior for $\boldsymbol{\phi}$ is implicitly trimmed by the algorithm to support only values of $\boldsymbol{\phi}$ that yield a nonempty identified set.

Step 2 requires checking whether the identified set for $\mathbf{Q}$ given $\boldsymbol{\phi}$ is empty. In the case where there are zero restrictions only subject to $f_i \leq n - i$, $i = 1,\dots,n$, the identified set is never empty. When there are sign restrictions (possibly alongside zero restrictions), there are different algorithms to check whether the identified set is empty, and their applicability depends on the types of restriction. The following algorithm in GK can be applied to any pattern of zero and sign restrictions.

Algorithm 2.

- **Step 1:** *Draw* $\mathbf{z}_1 \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ *and let* $\tilde{\mathbf{q}}_1 = \left[\mathbf{I}_n - \mathbf{F}_1'(\mathbf{F}_1\mathbf{F}_1')^{-1}\mathbf{F}_1\right]\mathbf{z}_1$*, then, for* $i = 2,\dots,n$*, draw* $\mathbf{z}_i \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ *and compute* $\tilde{\mathbf{q}}_i = \left[\mathbf{I}_n - \tilde{\mathbf{F}}_i'(\tilde{\mathbf{F}}_i\tilde{\mathbf{F}}_i')^{-1}\tilde{\mathbf{F}}_i\right]\mathbf{z}_i$*, where* $\mathbf{F}_i \equiv F_i(\phi)$ *and* $\tilde{\mathbf{F}}_i' = [\mathbf{F}_i', \tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_{i-1}]$*.*

- **Step 2:** *Compute*

$$\mathbf{Q}_0 = \left[\operatorname{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_1)\frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|}, \dots, \operatorname{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n,n})'\tilde{\mathbf{q}}_n)\frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|}\right].$$

- **Step 3:** *Check whether* $\mathbf{Q}_0$ *satisfies* $S(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$*. If so, conclude* $\mathcal{Q}(\boldsymbol{\phi})$ *is nonempty. Otherwise, repeat Steps 1–2 (up to a maximum of* L *times) until* $\mathbf{Q}_0$ *is obtained satisfying* $S(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$*. If no draws of* $\mathbf{Q}_0$ *satisfy* $S(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$*, approximate* $\mathcal{Q}(\boldsymbol{\phi})$ *as being empty.*

[6] The objective function in this minimisation is nondifferentiable in η, so we recommend obtaining this interval via grid search.

Step 1 of this algorithm generates vectors $(\tilde{\mathbf{q}}_1, ..., \tilde{\mathbf{q}}_n)$ that are orthogonal and satisfy the zero restrictions. Step 2 normalises these vectors to have unit length and imposes the sign normalisation that the diagonal elements of $\mathbf{A}_0$ are nonnegative. The resulting $\mathbf{Q}_0$ is an orthonormal matrix that satisfies the zero restrictions and the sign normalisations. Step 3 checks whether the drawn $\mathbf{Q}_0$ satisfies the sign restrictions. The advantage of this algorithm is its generality. A drawback is that, for a finite value of L, the algorithm may misclassify the identified set as being empty. Increasing the value of L reduces the chance of this happening, but at the cost of increased computing time when the identified set is actually empty at some values of $\boldsymbol{\phi}$. In practice, practitioners using this algorithm should check whether the chosen L is large enough by seeing whether the proportion of draws with empty identified set is sensitive to an increase in L.

When there are zero and sign restrictions on a single column of $\mathbf{Q}$, emptiness of the identified set can be determined without recourse to random sampling by using the following algorithm in GKV:

Algorithm 3. *Assume any zero and sign restrictions apply to* $\mathbf{q}_1$ *only and let the value of* $\boldsymbol{\phi}$ *be given. Let* $S(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$ *represent the sign restrictions (including the sign normalisation). Further, assume that the* $(n-1)\times n$ *matrix* $Z(\boldsymbol{\phi}) = [F(\boldsymbol{\phi})', \tilde{S}(\boldsymbol{\phi})']'$ *has rank* $n-1$ *for any* $(n - f_1 - 1)\times n$ *matrix* $\tilde{S}(\boldsymbol{\phi})$ *constructed from a selection of* $n - f_1 - 1$ *rows of* $S(\boldsymbol{\phi})$.

- **Step 1:** *Choose* $n - f_1 - 1$ *rows from* $S(\boldsymbol{\phi})$ *and collect these in* $\tilde{S}(\boldsymbol{\phi})$. *Construct* $Z(\boldsymbol{\phi}) = [F(\boldsymbol{\phi})', S(\boldsymbol{\phi})']'$.
- **Step 2:** *Compute an orthonormal basis for the null space of* $Z(\boldsymbol{\phi})$, $N(Z(\boldsymbol{\phi}))$, *which is an* $n \times 1$ *vector.*
- **Step 3:** *Check if either* $N(Z(\boldsymbol{\phi}))$ *or* $-N(Z(\boldsymbol{\phi}))$ *satisfies the remaining* $s - (n - f_1 - 1)$ *sign restrictions not contained in* $\tilde{S}(\boldsymbol{\phi})$. *If so, conclude that* $\mathcal{Q}(\boldsymbol{\phi})$ *is nonempty. Otherwise, return to Step 1 until all* $\binom{s}{n-f-1}$ *combinations are exhausted, in which case conclude that* $\mathcal{Q}(\boldsymbol{\phi})$ *is empty.*

This algorithm relies on the fact that any nonempty identified set for $\mathbf{q}_1$ must contain a vertex on the unit sphere where at least $n-1$ constraints are binding. The algorithm determines whether the identified set is nonempty by considering all possible combinations of $n - f_1 - 1$ binding sign restrictions and checking whether the implied vertex satisfies the remaining $s - (n - f_1 - 1)$ sign restrictions. Under the assumptions stated in the algorithm, the rank-nullity theorem implies that the null space of $Z(\boldsymbol{\phi})$ is one-dimensional, but if some $\mathbf{q}$ satisfies $Z(\boldsymbol{\phi})\mathbf{q} = \mathbf{0}_{(n-1)\times 1}$, then so too does $-\mathbf{q}$, so it is necessary to check whether $N(Z(\boldsymbol{\phi}))$ or $-N(Z(\boldsymbol{\phi}))$ satisfies the sign restrictions excluded from $\tilde{S}(\boldsymbol{\phi})$. The advantage of this algorithm over Algorithm 2 is that it will never misclassify the identified set as being nonempty. However, since the algorithm requires

checking $\binom{s}{n-f-1}$ combinations of restrictions, the algorithm may become slow or infeasible when there is a large number of sign restrictions.[7]

A third approach to checking whether the identified set is empty is the 'Chebyshev criterion' proposed in Amir-Ahmadi and Drautzburg (2021). This algorithm is applicable when there are sign restrictions that constrain a single column of $\mathbf{Q}$ and there are no zero restrictions. Read (2021) extends this algorithm to allow for zero restrictions.

Step 3 of Algorithm 1 requires computing the bounds of the identified set for η at each draw of $\boldsymbol{\phi}$. If one is interested in more than one scalar object at a time (e.g., impulse responses for multiple variables), this step is run repeatedly at each draw of $\boldsymbol{\phi}$. As for Step 2, there are multiple approaches for computing the bounds of the identified set. GK suggest two different approaches that are applicable under arbitrary configurations of zero and sign restrictions. The first is to use a numerical optimizer initialised at the value of $\mathbf{Q}_0$ obtained using Algorithm 2. This is a nonconvex optimization problem, so convergence to the true optimum is not guaranteed. The second approach is to repeat Algorithm 2 many times at each draw of $\boldsymbol{\phi}$ to obtain a large number of draws of $\mathbf{Q}$ from $\mathcal{Q}(\boldsymbol{\phi})$ and then to compute the minimum and maximum of η over the draws. This provides an approximated identified set that is smaller than the actual identified set, but that converges to the actual identified set as the number of draws goes to infinity. A third approach is available when the zero and sign restrictions constrain a single column of $\mathbf{Q}$ only, in which case the bounds of the identified set can be computed using the active-set algorithm in Gafarov et al. (2018). This approach may be prohibitively slow when there are many sign restrictions.

6.2 Model averaging (GKV)

This section presents a general algorithm to numerically approximate the set of posterior means when there is uncertainty over the set of identifying restrictions, as in GKV.

Algorithm 5.

- **Step 1:** *Draw a model $M \in \mathcal{M}$ from a multinomial distribution with parameters $(\pi_{M|\mathbf{Y}} : M \in \mathcal{M})$.*
- **Step 2:** *If the drawn M belongs to $\mathcal{M}_p$, draw $\eta \sim \pi_{\eta|M,\mathbf{Y}}$ and set $IS_\eta^{mix} = \{\eta\}$. If the drawn M belongs to $\mathcal{M}_s$, draw $\boldsymbol{\phi}_M \sim \pi_{\boldsymbol{\phi}|M,\mathbf{Y}}$ and set $IS_\eta^{mix} = IS_\eta(\boldsymbol{\phi}_M|M)$.*
- **Step 3:** *Repeat Steps 1 and 2 G times to obtain G draws of IS_η^{mix}.*

[7] This algorithm can be straightforwardly extended to the case where the first i^* columns of $\mathbf{Q}$ are exactly identified by zero restrictions and there are zero and/or sign restrictions on the (i^*+1) column, or when the first i^* columns of $\mathbf{Q}$ are determined up to an i^*-dimensional linear subspace of $\mathbb{R}^n$.

- **Step 4:** *Let l_g^{mix} and u_g^{mix} be the upper and lower bounds of IS_η^{mix}, respectively, where $l_g^{mix} = u_g^{mix}$ if the gth draw of M belongs to $\mathcal{M}_p$. Approximate the bounds of the set of posterior means by the sample averages of l_g^{mix} and u_g^{mix}.*

Step 1 of this algorithm requires computing the posterior model probabilities, $\pi_{M|\mathbf{Y}}$, for each model or, equivalently, the marginal likelihood for each model. Algorithms to compute the marginal likelihood include those in Chib and Jeliazkov (2001), Geweke (1999) and Sims et al. (2008). When the models admit an identical reduced form, it is unnecessary to compute the marginal likelihoods, because the posterior model probabilities depend only on posterior-prior plausibility ratios, which are the posterior probability that the identified set is nonempty divided by the prior probability that the identified set is nonempty in each model. These ratios can be computed using numerical approximations of the prior and posterior probabilities that the identified set is nonempty. Depending on the pattern of zero and sign restrictions considered, these probabilities can be computed by drawing $\boldsymbol{\phi}$ from its prior or posterior and using the algorithms described in the previous subsection to check whether the identified set is nonempty. Since computing these ratios requires drawing $\boldsymbol{\phi}$ from its prior, it is necessary for this prior to be proper.[8]

Step 2 requires drawing from the posterior of the object of interest η when the model is point-identified or the prior is for the structural parameters, which are standard problems. For example, when the prior is for the reduced-form parameters, one simply draws from the posterior for $\boldsymbol{\phi}$ and transforms the draw. When the sampled model is set-identified, Step 2 requires drawing $\boldsymbol{\phi}$ from its posterior and computing the identified set for the object of interest. Note that $\pi_{\boldsymbol{\phi}|M,\mathbf{Y}}$ supports only values of $\boldsymbol{\phi}$ with nonempty identified set. In practice, the practitioner may specify a prior that assigns positive probability to regions of the reduced-form parameter space with empty identified set and simply continue to draw $\boldsymbol{\phi}$ from its posterior at a given draw of M until the identified set is nonempty. The bounds of the identified set can be computed using any of the approaches described in the previous subsection (depending on the pattern of zero and sign restrictions). As in GK, the draws of l_g^{mix} and u_g^{mix} can be used to construct a robust credible interval.

6.3 KL-neighborhood (GKU)

This section discusses how to compute the set of posterior means and the posterior Gamma minimax decision (i.e., the point-estimator under ambiguity). The algorithms below assume that posterior draws of $\boldsymbol{\phi}$ are given and that η can be drawn from the benchmark conditional prior. GKU also present modifications of these algorithms when direct draws of η are unavailable but its probability

[8] It is possible to eliminate a source of Monte Carlo sampling variability arising from Step 1 by avoiding sampling M from a multinomial distribution. For example, if the posterior model probabilities are known to be 0.5 when averaging over two models, one simply needs to draw $G/2$ times from the relevant posterior for each model.

density can be evaluated up to a proportional constant. The first algorithm below describes how to compute the set of posterior means for some object of interest $f(\eta)$.

Algorithm 6. *Let G posterior draws of $\boldsymbol{\phi}$, $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_G\}$, and benchmark conditional prior $\pi^*_{\eta|\boldsymbol{\phi}}$ be given.*

- **Step 1:** *For each $g = 1, \ldots, G$, obtain K independent draws of η, $\eta_{gk} \sim \pi^*_{\eta|\boldsymbol{\phi}}$, $k = 1, \ldots, K$. Approximate the Lagrange multipliers, $\kappa^l_\lambda(\boldsymbol{\phi}_g)$ and $\kappa^u_\lambda(\boldsymbol{\phi}_g)$, by solving the following optimization problems:*

$$\hat{\kappa}^l_\lambda(\boldsymbol{\phi}_g) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(-\frac{f(\eta_{gk})}{\kappa} \right) \right) + \kappa\lambda \right\} \tag{6.1}$$

$$\hat{\kappa}^u_\lambda(\boldsymbol{\phi}_g) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(\frac{f(\eta_{gk})}{\kappa} \right) \right) + \kappa\lambda \right\}. \tag{6.2}$$

- **Step 2:** *Approximate the set of posterior means of $f(\eta)$ by*

$$\left[\frac{1}{G} \sum_{g=1}^{G} \left(\frac{\sum_{k=1}^{K} f(\eta_{gk}) \exp\left(-\frac{f(\eta_{gk})}{\hat{\kappa}^l_\lambda(\phi_g)} \right)}{\sum_{k=1}^{K} \exp\left(-\frac{f(\eta_{gk})}{\hat{\kappa}^l_\lambda(\phi_g)} \right)} \right), \frac{1}{G} \sum_{g=1}^{G} \left(\frac{\sum_{k=1}^{K} f(\eta_{gk}) \exp\left(\frac{f(\eta_{gk})}{\hat{\kappa}^u_\lambda(\phi_g)} \right)}{\sum_{k=1}^{K} \exp\left(\frac{f(\eta_{gk})}{\hat{\kappa}^u_\lambda(\phi_g)} \right)} \right) \right]. \tag{6.3}$$

The optimization problems in Step 1 are convex and can be solved reliably using a gradient-based numerical optimization routine, such as the interior-point algorithm in Matlab's 'fmincon' optimizer. Care should be taken when the Lagrange multipliers are close to zero, since terms in the expression for the set of posterior means may be very large and result in numerical overflow.

The following algorithm can be used to approximate the worst-case risk of decision (i.e., estimator) δ, which can then be used to compute the posterior Gamma minimax estimator.

Algorithm 7. *Let G posterior draws of $\boldsymbol{\phi}$, $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_G\}$, and benchmark conditional prior $\pi^*_{\eta|\boldsymbol{\phi}}$ be given. Let $h(\delta, \eta)$ be the loss function (e.g., quadratic or check).*

- **Step 1:** *For each $g = 1, \ldots, G$, obtain K independent draws of η, $\eta_{gk} \sim \pi^*_{\eta|\boldsymbol{\phi}}$, $k = 1, \ldots, K$. Approximate the Lagrange multiplier, $\kappa_\lambda(\boldsymbol{\phi}_g)$, by solving the following optimization problem:*

$$\hat{\kappa}_\lambda(\delta, \boldsymbol{\phi}_g) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(\frac{h(\delta, \eta_{gk}))}{\kappa} \right) \right) + \kappa\lambda \right\}. \tag{6.4}$$

- **Step 2:** *For each $g = 1, \ldots, G$, compute*

$$\hat{r}_\lambda(\delta, \boldsymbol{\phi}_g) = \frac{\sum_{k=1}^{K} h(\delta, \eta_{gk}) \exp\left(\frac{h(\delta, \eta_{gk})}{\hat{\kappa}_\lambda(\delta, \boldsymbol{\phi}_g)} \right)}{\sum_{k=1}^{K} \exp\left(\frac{h(\delta, \eta_{gk})}{\hat{\kappa}_\lambda(\delta, \boldsymbol{\phi}_g)} \right)}. \tag{6.5}$$

The posterior Gamma minimax estimator is then obtained by minimising $(1/G)\sum_{g=1}^{G}\hat{r}_\lambda(\delta,\boldsymbol{\phi}_g)$ with respect to δ. If the loss is differentiable in δ, this minimisation can be carried out using a gradient-based numerical optimization routine. Otherwise, the minimisation can be done via grid search, which is not computationally costly because δ is a scalar. In either case, the same draws of $\boldsymbol{\phi}$ and η (at each draw of $\boldsymbol{\phi}$) should be used for the optimization.

7 Empirical Illustration

This section illustrates how to apply the methods described above using an empirical example. The empirical application considered is from Arias et al. (2019; henceforth, ACR), who estimate the effects of monetary policy shocks in the United States using a mixture of zero and sign restrictions on the systematic response of the federal funds rate to macroeconomic variables.

Reduced-form VAR. The model's endogenous variables are real GDP (GDP_t), the GDP deflator ($GDPDEF_t$), a commodity price index (COM_t), total reserves (TR_t), non-borrowed reserves (NBR_t) (all in natural logarithms) and the federal funds rate (FFR_t). The data are monthly and run from January 1965 to June 2007. The VAR includes 12 lags and no deterministic terms.

In order to apply the approach in GKV, we need to compute the posterior-prior plausibility ratio. This requires drawing from the prior for $\boldsymbol{\phi}$, so this prior needs to be proper. ACR use an improper normal-inverse-Wishart prior, which is inappropriate for our purposes. Instead, we use a diffuse (but proper) normal-inverse-Wishart prior under which the prior means of the VAR coefficients imply that each variable in $\mathbf{y}_t$ follows a univariate random walk a priori. The posterior for the reduced-form parameters is then also a normal-inverse-Wishart distribution, from which it is straightforward to obtain independent draws (e.g., using the sampler described in Del Negro and Schorfheide (2011)).[9]

Identifying restrictions. Let $\mathbf{y}_t = (FFR_t, GDP_t, GDPDEF_t, COM_t, TR_t, NBR_t)'$. The monetary policy shock is ε_{1t} and the first equation of the SVAR can be interpreted as the monetary policy reaction function. ACR19 set-identify impulse responses to the monetary policy shock using a mixture of sign and zero restrictions on the systematic response of monetary policy to macroeconomic variables, which are restrictions on the first row of $\mathbf{A}_0$. The zero restrictions that they impose are that FFR_t does not react contemporaneously to TR_t and NBR_t (i.e., these two variables do not appear in the central bank's reaction function), which implies that $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{5,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{5,6})'\mathbf{q}_1 = 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{6,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{6,6})'\mathbf{q}_1 = 0$. The sign restrictions that they impose are that the contemporaneous reactions of FFR_t to GDP_t and $GDPDEF_t$ are non-negative, which – given the sign normalisation that $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{1,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,6})'\mathbf{q}_1 \geq 0$ – implies that

[9]The results when conducting standard Bayesian inference (as in ACR) and robust Bayesian inference (as in GK) reported below are very similar to those obtained using the same (improper) prior used in ACR.

$\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{2,6} = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{2,6})'\mathbf{q}_1 \leq 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{3,6} = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{3,6})'\mathbf{q}_1 \leq 0$. Additionally, the impact impulse response of the federal funds rate to a monetary policy shock is restricted to be nonnegative, which implies that $\mathbf{e}_{1,6}'\mathbf{A}_0^{-1}\mathbf{e}_{1,6} = \mathbf{e}_{1,6}'\mathbf{\Sigma}_{tr}\mathbf{q}_1 \geq 0$.

Standard Bayesian inference. The approach to Bayesian inference used in ACR assumes a uniform prior for $\mathbf{Q}$. Here, we assume that the conditional prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ is uniform over the space of orthonormal matrices satisfying the identifying restrictions.[10] We obtain 10,000 independent draws from the normal-inverse-Wishart posterior for $\boldsymbol{\phi}$ such that the identified set is nonempty; we check whether the identified set is nonempty using Algorithm 3. At each draw of $\boldsymbol{\phi}$, we obtain a draw of $\mathbf{Q}$ from the uniform distribution over $\mathcal{Q}(\boldsymbol{\phi}|S)$ using Algorithm 2. The resulting joint draw of $(\boldsymbol{\phi}, \mathbf{Q})$ is then used to compute the impulse responses.

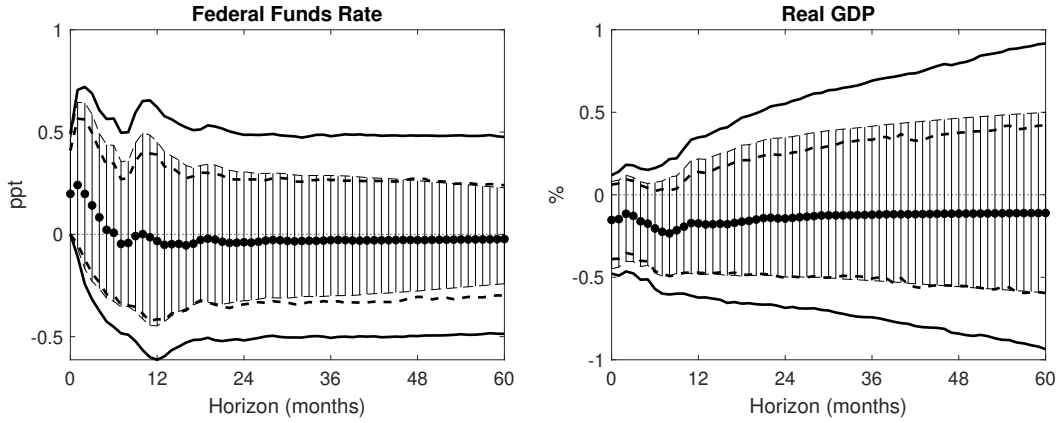
Figure 1 plots the impulse responses of the federal funds rate and real GDP to a positive standard-deviation monetary policy shock (the responses of the remaining variables are omitted for brevity). Based on the posterior mean (black circles), the federal funds rate increases by about 20 basis points in response to the shock before declining to be around its pre-shock value after six months. Output is around 0.2 per cent lower in the year after the shock and the posterior probability of a negative output response is reasonably high at short horizons (e.g., around 85 per cent on impact and one year after the shock). Overall, the results essentially replicate those in ACR and suggest that a positive monetary policy shock results in an economic contraction.

Full ambiguity (GK). As discussed above, the results obtained under the single prior may be sensitive to the choice of prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$, which is not updated by the data. To address this concern, the robust Bayesian approach of GK replaces the unrevisable prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ with the set of all (conditional) priors that are consistent with the identifying restrictions in the sense that the prior places probability one on the identified set given $\boldsymbol{\phi}$. This generates a set of posteriors, which can be summarised by a set of posterior means (an estimator of the identified set) and a robust credible region (the shortest interval that is assigned at least a given posterior probability under all posteriors within the set). To obtain these quantities, it is necessary to compute the lower and upper bound of the identified set for the object of interest (i.e., the impulse response at each horizon) at each draw of $\boldsymbol{\phi}$. As discussed above, there are several ways to this. Here, we apply the active-set algorithm described in Gafarov et al. (2018) at each draw of $\boldsymbol{\phi}$ from its posterior.

The set of posterior means (the vertical lines) includes zero at all horizons. This means that there exist (unrevisable) priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$ that are consistent with the identifying restrictions

[10] This differs slightly from the prior used in ACR, who assume a uniform-normal-inverse-Wishart prior for $\mathbf{Q}$ and $\boldsymbol{\phi}$. In terms of implementation, our prior requires a single draw of $\mathbf{Q}$ to be obtained at each draw of $\boldsymbol{\phi}$ (with nonempty identified set). In contrast, the prior in ACR is imposed by making a joint draw of $\boldsymbol{\phi}$ from the normal-inverse-Wishart posterior and $\mathbf{Q}$ from the uniform distribution over $\mathcal{O}(n)$, and rejecting joint draws that do not satisfy the sign restrictions. See Uhlig (2017) for a discussion of this point. This difference in priors does not substantively affect the results.

Figure 1: Impulse Responses to a Monetary Policy Shock – Standard and Robust Bayesian Inference



Notes: Circles and dashed lines are, respectively, posterior means and 95 per cent (pointwise) highest posterior density intervals under the uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$. Vertical bars are sets of posterior means and solid lines are 95 per cent (pointwise) robust credible regions. Impulse responses are to a standard-deviation shock.

and yield positive posterior mean output responses. The posterior lower probability – the lowest probability over all posteriors generated by the set of priors – of a negative output response is close to zero for all horizons considered.[11] This suggests that the result that output falls with high posterior probability that is obtained under standard Bayesian inference is sensitive to the choice of unrevisable prior. GK propose to quantify the influence of the choice of single prior on posterior inference by comparing the width of the highest posterior density intervals (the dashed lines) against that of the robust Bayesian credible intervals (the solid lines). On average across the horizons considered, the width of the 95 per cent highest posterior density intervals for the output response is 40 per cent that of the robust credible intervals, which suggests that the unrevisable prior contributes a substantial amount of the information contained in the standard Bayesian posterior.

Model averaging (GKV). To illustrate the application of the (robust) Bayesian model-averaging procedure in GKV, we consider a second set of identifying restrictions in addition to the set considered above. Specifically, we use the classic recursiveness assumption considered in, for example, Christiano, Eichenbaum and Evans (1999). Under this set of restrictions, GDP_t, $GDPDEF_t$ and COM_t do not respond contemporaneously to a monetary policy shock (i.e., a shock to FFR_t), while FFR_t does not respond contemporaneously to shocks in TR_t and NBR_t. After re-ordering the variables so that $y_t = (GDP_t, GDPDEF_t, COM_t, FFR_t, TR_t, NBR_t)'$, the restrictions imply

[11] The posterior lower probability that the output response is negative at a given horizon is approximated by the share of draws from the posterior of $\boldsymbol{\phi}$ where the upper bound of the identified set for the output response is negative (i.e., $u(\boldsymbol{\phi}) < 0$).

that $\mathbf{A}_0^{-1}$ is block lower-triangular. These restrictions are sufficient to point-identify the impulse responses to a monetary policy shock; in particular, the fourth column of $\mathbf{A}_0^{-1}$ is identified by the fourth column of $\mathbf{\Sigma}_{tr}$.

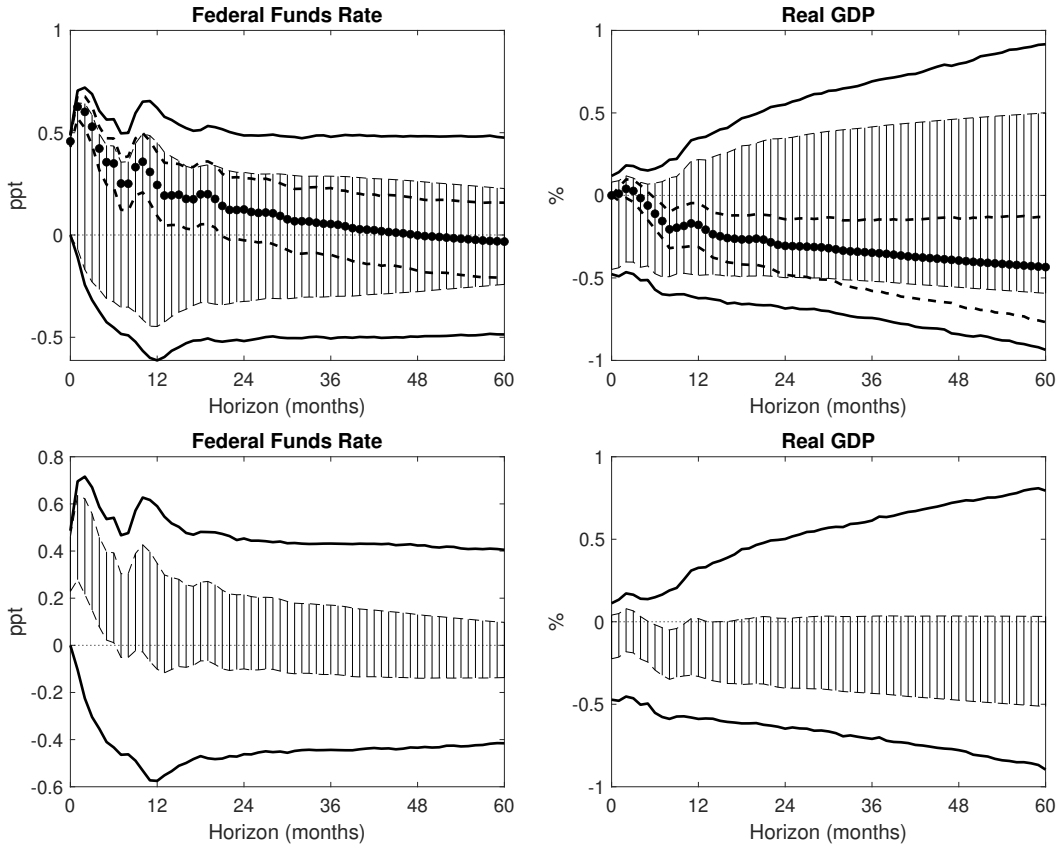
We assume there is uncertainty over the set of identifying restrictions. For the sake of illustration, we place equal weights on the two sets of restrictions. GKV discuss when the prior model probabilities are updated by the data. In particular, the prior model probabilities are not updated when the models are 'indistinguishable' (i.e., they admit an identical reduced-form defined on a common parameter space) and share a common prior for $\boldsymbol{\phi}$, where the prior for $\boldsymbol{\phi}$ is the notional reduced-form prior truncated to the region with nonempty identified set. Under the set-identifying restrictions considered, the identified set is never empty at any value of $\boldsymbol{\phi}$ supported by the reduced-form prior. The identified set under the point-identifying restrictions (which is a singleton) is also never empty. The two models share a common reduced-form prior and the posterior model probabilities are equal to the prior model probabilities. The set of posteriors that allows for uncertainty over the identifying restrictions is then given by the simple average of the set of posteriors under the set-identifying restrictions and the single posterior under the point-identifying restrictions.

The top panels of Figure 2 plot the posterior means and 95 per cent highest posterior density intervals under the point-identifying restrictions alongside the robust Bayesian output under the set-identifying restrictions (i.e., the robust Bayesian output plotted in Figure 1). Under the point-identifying restrictions, output falls with high posterior probability. The bottom panels plot the model-averaged set of posterior means and robust credible intervals. The set of posterior means under the set-identifying restrictions has been shrunk towards the posterior mean under the point-identifying restrictions; since the posterior model probabilities are equal, the lower (upper) bound of the set of posterior means is the average of the lower (upper) bound of the set of posterior means in the set-identified model and the posterior mean in the point-identified model. As a result, the set of posterior means for the output response now excludes zero at horizons of less than one year. Nevertheless, the robust credible intervals contain zero at all horizons and the posterior lower probability that the output response is negative at the one-year horizon is only 50 per cent. Equally weighting models identified using the set-identifying restrictions in ACR and classic recursive restrictions thus provides little evidence that output falls following a monetary policy shock once one allows for ambiguity over the unrevisable prior in the set-identified model.

This approach can also be used to back out the prior model probabilities that would lead to particular posterior inferences. For example, one would need to place a prior probability of at least 0.55 on the point-identified model for the model-averaged set of posterior means to unambiguously imply a negative output response at the one-year horizon.

KL-neighborhood (GKU). To illustrate the approach in GKU, we treat the normal-inverse-Wishart prior for $\boldsymbol{\phi}$ and the conditionally uniform prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ as a benchmark prior and conduct a posterior sensitivity exercise. The conditionally uniform prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ implies a

Figure 2: Impulse Responses to a Monetary Policy Shock – Uncertain Identification

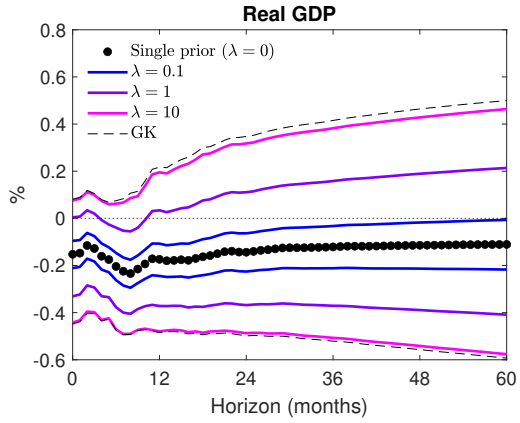


Notes: Top panels plot robust Bayesian output under set-identifying restrictions and standard Bayesian output under point-identifying restrictions. Bottom panels plot robust Bayesian output from GKV, assuming equal prior probabilities on the two sets of restrictions. Circles and dashed lines are, respectively, posterior means and 95 per cent (pointwise) highest posterior density intervals under point-identifying restrictions. Vertical bars are sets of posterior means and solid lines are 95 per cent (pointwise) robust credible regions.

benchmark (conditional) prior for the impulse response of interest, η. We consider perturbations of the prior for η within a neighborhood of the benchmark prior. The size of the neighborhood is determined by the KL distance, λ. We consider different values of λ and document how posterior inference changes depending on the size of the neighborhood.

At each draw of $\boldsymbol{\phi}$ from its posterior, $\{\boldsymbol{\phi}_m\}_{m=1}^M$, we obtain N draws of $\mathbf{Q}$ from the conditionally uniform distribution over the identified set. We transform these draws into impulse-response space, so we have a set of draws of η from the benchmark prior at each draw of $\boldsymbol{\phi}$, $\{\eta_{mi}\}_{i=1}^N$. We then approximate the Lagrange multipliers, $\kappa_\lambda^l(\boldsymbol{\phi})$ and $\kappa_\lambda^u(\boldsymbol{\phi})$, using an interior-point algorithm implemented within Matlab's 'fmincon' optimizer. After computing the Lagrange multipliers, we compute the set of posterior means using the sample analogue of (5.25) (i.e., with integration

Figure 3: Impulse Responses to a Monetary Policy Shock – Posterior Sensitivity Analysis



Notes: Circles are posterior means under the benchmark conditional prior, coloured solid lines are bounds of sets of posterior means for different values of the Kullback-Leibler distance λ and dashed lines are bounds of sets of posterior means using the approach from GK (i.e., from the right panel of Figure 1).

replaced by averaging over the draws of η and ϕ). For the purposes of illustration, we conduct the sensitivity exercise under different values of $\lambda \in \{0.1, 1, 10\}$.

Figure 3 presents the set of posterior means of the output response to a monetary policy shock at each horizon of interest and for each value of λ considered. For comparison, the figure also plots the posterior mean under the benchmark prior (which is equivalent to the case where $\lambda = 0$) and the set of posterior means obtained using the approach from GK described above (which is equivalent to the case where $\lambda \to \infty$). When $\lambda = 0.1$, the set of priors is constrained to lie within a relatively small KL distance of the benchmark prior, and the resulting set of posterior means lies entirely below zero at all horizons considered; in other words, for priors within a relatively small neighborhood of the benchmark prior, the set of posterior means supports the conclusion that output falls following a positive monetary policy shock. As λ increases, we allow for priors that are further from the benchmark prior and the set of posterior means expands. At $\lambda = 1$, the set of posterior means excludes zero at only some horizons shorter than one year, and at $\lambda = 10$ the set of posterior means includes zero at all horizons. For large values of λ, the set of priors grows to include all priors that are consistent with the identifying restrictions. Consequently, as λ increases, the set of posterior means converges towards that obtained using the approach from GK, which allows for full ambiguity over the set of priors that are consistent with the identifying restrictions.

8 Conclusion

We overviewed how Robust Bayesian analysis provides useful tools for Bayesian econometricians and decision-makers who want to assess the sensitivity of inferences and statistical decisions to the choice of a prior. The main idea is to construct a set of priors, which in turn delivers: 1) a set of posterior quantities that can be used for inference; and/or 2) an optimal statistical decision for ambiguity-averse decision-makers. We discussed how the sensitivity concerns are particularly salient in set-identified structural models due to the fact that a component of the prior is not revised by the data even in large samples. We reviewed different ways to construct the set of priors and discussed in detail how to implement the methods in macroeconometric applications using set-identified SVARs.

References

Amir-Ahmadi, P. and T. Drautzburg (2021): "Identification and Inference with Ranking Restrictions," *Quantitative Economics*, 12, 1–39.

Andrews, I., M. Gentzkow, and J. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132, 1553–1592.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2020): "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 88, 2231–2258.

Arias, J., D. Caldara, and J. Rubio-Ramírez (2019): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," *Journal of Monetary Economics*, 101, 1–13.

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.

Armstrong, T. and M. Kolesár (2021): "Sensitivity Analysis using Approximate Moment Condition Models," *Quantitative Economics*, 12, 77–108.

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

Bacchiocchi, E. and T. Kitagawa (2020): "Locally- But Not Globally-Identified SVARs," cemmap Working Paper CWP40/20.

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.

Berger, J. (1984): "The Robust Bayesian Viewpoint," in *Robustness of Bayesian Analysis*, ed. by J. Kadane, Amsterdam, Netherlands: North-Holland.

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

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

Berger, J. O. (1994): "An Overview of Robust Bayesian Analysis," *TEST*, 3, 5–58.

Betrò, B. and F. Ruggeri (1992): "Conditional Γ-minimax Actions Under Convex Losses," *Communications in Statistics, Part A - Theory and Methods*, 21, 1051–1066.

Bonhomme, S. and M. Weidner (2018): "Minimizing Sensitivity to Model Misspecification," Tech. Rep. CWP59/18.

Brown, L. and R. Purves (1973): "Measurable Selections of Extrema," *Annals of Statistics*, 1, 902–912.

Chamberlain, G. (2000): "Econometric Applications of Maxmin Expected Utilities," *Journal of Applied Econometrics*, 15, 625–644.

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.

Chib, S. and I. Jeliazkov (2001): "Likelihoods from the Metropolis Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

Christensen, T. and B. Connault (2019): "Counterfactual Sensitivity and Robustness," *unpublished manuscript*.

Christiano, L., M. Eichenbaum, and C. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?" in *Handbook of Macroeconomics*, ed. by J. Taylor and M. Woodford, Elsevier, vol. 1, Part A, 65–148.

DasGupta, A. and W. Studden (1989): "Frequentist Behavior of Robust Bayes Estimates of Normal Means," *Statistics and Decisions*, 7, 333–361.

Del Negro, M. and F. Schorfheide (2011): "Bayesian Macroeconometrics," in *Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. V. Dijk, Oxford University Press, 293—389.

Denneberg, D. (1994): *Non-additive Measures and Integral*, Kluwer Academic Publishers.

Dey, D. K. and A. C. Micheas (2000): "Ranges of posterior expected losses and ϵ-robust actions," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 145–159.

Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *The Quarterly Journal of Economics*, 75, 643–669.

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.

Geweke, J. (1999): "Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–126.

Giacomini, R. and T. Kitagawa (2021): "Robust Bayesian Inference for Set-identified Models," *Econometrica*, 89, 1519–1556.

Giacomini, R., T. Kitagawa, and M. Read (in pressa): "Robust Bayesian Inference in Proxy SVARs," *Journal of Econometrics*.

Giacomini, R., T. Kitagawa, and H. Uhlig (2019): "Estimation Under Ambiguity," cemmap Working Paper CWP24/19.

Giacomini, R., T. Kitagawa, and A. Volpicella (in pressb): "Uncertain Identification," *Quantitative Economics*.

Gilboa, I. and D. Schmeidler (1993): "Updating ambiguous beliefs," *Journal of Economic Theory*, 59, 33–49.

Good, I. (1965): *The Estimation of Probabilities*, MIT Press.

Gustafson, P. (2000): "Local Robustness in Bayesian Analysis," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 71–88.

Hansen, L. P. and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review, AEA Papers and Proceedings*, 91, 60–66.

Ho, P. (2020): "Global Robust Bayesian Analysis in Large Models," Federal Reserve Bank of Richmond WP 20-07.

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

Huber, P. J. and E. M. Ronchetti (2009): *Robust Statistics, 2nd edition*, John Wiley and Sons.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81, 1185–1201.

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.

Kudō, H. (1967): "On partial prior information and the property of parametric sufficiency," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. L. Cam and J. Neyman, vol. 1, 251–265.

Lavine, M., L. Wasserman, and R. Wolpert (1991): "Bayesian inference with specified marginals," *Journal of the American Statistical Association*, 86, 964–971.

Leamer, E. (1978): *Specification Searches*, New York: Wiley.

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

Liao, Y. and A. Simoni (2013): "Semi-parametric Bayesian Partially Identified Models based on Support Function," *unpublished manuscript*.

Manski, C. (1981): "Learning and decision making when subjective probabilities have subjective domains," *Annals of Statistics*, 9, 59–65.

Manski, C. F. (2000): "Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice," *Journal of Econometrics*, 95, 415–442.

Mertens, K. and M. Ravn (2013): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 103, 1212–47.

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

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

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

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

Moreno, E. and J. Cano (1995): "Classes of bidimensional priors specified on a collection of sets," *Journal of Statistical Inference and Planning*, 46, 325–334.

Müller, U. K. (2012): "Measuring prior sensitivity and prior informativeness in large Bayesian models," *Journal of Monetary Economics*, 59, 581–597.

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

Peterson, I. R., M. R. James, and P. Dupuis (2000): "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints," *ISSS Transactions on Automatic Control*, 45, 398–412.

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

Read, M. (2021): "Algorithms for Inference in SVARs Identified with Sign and Zero Restrictions," Unpublished manuscript.

Rieder, H. (1994): *Robust Asymptotic Statistics*, Springer.

Ríos Insua, D. and F. Ruggeri, eds. (2000): *Robust Bayesian Analysis*, Springer.

Robbins, H. (1956): "An empirical Bayes approach to Statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 157–163.

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.

Schmeidler, D. (1989): "Subjective probability and expected utility without additivity," *Econometrica*, 57, 571–587.

Shyamalkumar, N. (2000): "Likelihood robustness," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 127–143.

Sims, C., D. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146, 255–274.

Sivaganesan, S. and J. Berger (1989): "Ranges of posterior measures for priors with unimodal contaminations," *Annals of Statistics*, 17, 868–889.

Stock, J. and W. Watson (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128, 917–948.

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.

——— (2017): "Shocks, Sign Restrictions, and Identification," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. H. A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, vol. 2, 95–127.

Vidakovic, B. (2000): "Γ-minimax: A paradigm for conservative robust Bayesians," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 241–259.

Wald, A. (1950): *Statistical Decision Functions*, New York: Wiley.

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

Watson, J. and C. Holmes (2016): "Approximate models and robust decisions," *Statistical Science*, 31, 465–489.