Quantitative Economics 13 (2022), 95–123 1759-7331/20220095

Uncertain identification

Raffaella Giacomini
Department of Economics, University College London and Federal Reserve Bank of Chicago

Toru Kitagawa
Department of Economics, University College London and Department of Economics, Kyoto University

Alessio Volpicella
School of Economics, University of Surrey

Uncertainty about the choice of identifying assumptions is common in causal studies, but is often ignored in empirical practice. This paper considers uncertainty over models that impose different identifying assumptions, which can lead to a mix of point- and set-identified models. We propose performing inference in the presence of such uncertainty by generalizing Bayesian model averaging. The method considers multiple posteriors for the set-identified models and combines them with a single posterior for models that are either point-identified or that impose nondogmatic assumptions. The output is a set of posteriors (*post-averaging ambiguous belief*), which can be summarized by reporting the set of posterior means and the associated credible region. We clarify when the prior model probabilities are updated and characterize the asymptotic behavior of the posterior model probabilities. The method provides a formal framework for conducting sensitivity analysis of empirical findings to the choice of identifying assumptions. For example, we find that in a standard monetary model one would need to attach a prior probability greater than 0.28 to the validity of the assumption that prices do not react contemporaneously to a monetary policy shock, in order to obtain a negative response of output to the shock.

Keywords. Partial identification, sensitivity analysis, model averaging, Bayesian robustness, ambiguity.

JEL classification. C11, C32, C52.

Raffaella Giacomini: r.giacomini@ucl.ac.uk
Toru Kitagawa: t.kitagawa@ucl.ac.uk
Alessio Volpicella: a.volpicella@surrey.ac.uk

We would like to thank Frank Kleibergen, José Montiel-Olea, Ulrich Müller, John Pepper, James Stock, three anonymous referees, and participants at numerous seminars and conferences for valuable comments and beneficial discussions. Financial support from ERC grants (numbers 536284 for R. Giacomini and 715940 for T. Kitagawa) and the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001) is gratefully acknowledged. This paper is based on and develops the first chapter of the Ph.D. thesis, Volpicella (2020). The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

 https://doi.org/10.3982/QE1671

1. Introduction

The choice of identifying assumptions is the crucial step that allows researchers to draw causal inferences using observational data. This is often a controversial choice, and there can be uncertainty about which assumptions to impose from a menu of plausible ones, but this uncertainty and its effects on inference are typically ignored in empirical work. This paper proposes a formal framework for sensitivity analysis via Bayesian model averaging in the presence of uncertain identification, which we characterize as uncertainty over a class of models that impose different sets of identifying assumptions. The class of models can include ones where parameters are set-identified, which occurs when the assumptions are underidentifying or take the form of inequality restrictions. For these models, we advocate adopting the multiple-prior approach of Giacomini and Kitagawa (2021). In our context, the approach has the additional advantage of isolating the component of each model that depends on the identifying restrictions, making it possible, for example, to compare models that only differ in the restrictions they impose.

The paper makes both a methodological and a theoretical contribution. The methodological contribution is to extend Bayesian model averaging/selection to allow for models characterized by multiple priors (associated here with set identification). The theoretical contribution is to clarify how the different components of the models affect inference in terms of model averaging/selection in finite samples and asymptotically.

There are several examples in economics where empirical researchers face uncertainty about identifying assumptions that lead to point- or set-identification of a common causal parameter of interest. The first is macroeconomic policy analysis based on structural vector autoregressions (SVARs), where assumptions include causal ordering restrictions (Sims (1980)) and long-run neutrality restrictions (Blanchard and Quah (1993)). Subsets of these assumptions deliver set-identified impulse-responses, as do sign restrictions (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)). The second example is microeconometric causal effect studies with assumptions such as selection on observables (Ashenfelter (1978) and Rosenbaum and Rubin (1983)), selection on observables and unobservables (Altonji, Elder, and Taber (2005)), exclusion and monotonicity restrictions in instrumental variables methods (Imbens and Angrist (1994), yielding set-identification of the average treatment effect), and monotone instrument assumptions (Manski and Pepper (2000), also yielding set-identification). The third example is missing data with assumptions such as missing at random, Bayesian imputation (Rubin (1987)), and unknown missing mechanism (Manski (1989), yielding set-identification). Finally, estimation of structural models with multiple equilibria relies on assumptions about the equilibrium selection rule, with different assumptions (or lack thereof) delivering point- or set-identification (e.g., Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2011), and Ciliberto and Tamer (2009)).

The common practice in empirical work is to report results based on what is deemed the most credible set of identifying assumptions, or sometimes, based on a small number of alternative assumptions, viewed as an informal sensitivity analysis. Our method provides a formal framework for investigating the sensitivity of empirical findings to

specific identifying assumptions and/or for aggregating results based on different identifying assumptions, which can be more practical than reporting separate results when there are many restrictions.[1]

The idea of model averaging has a long history in econometrics and statistics since the pioneering works of Bates and Granger (1969) and Leamer (1978). The literature has considered Bayesian approaches (see, e.g., Hoeting, Madigan, Raftery, and Volinsky (1999)), frequentist approaches (Hansen (2007, 2014), Hjort and Claeskens (2003), Liu and Okui (2013)), and hybrid approaches (Hjort and Claeskens (2003), Kitagawa and Muris (2016), and Magnus, Powell, and Prüfer (2010)), but none of them allows for set-identification/multiple priors in any candidate model.

This paper takes a Bayesian perspective. The standard approach to Bayesian model averaging delivers a single posterior that is a mixture of the posteriors of the models, with weights equal to the posterior model probabilities.[2] This approach could in principle be extended to our context if one could obtain a single posterior for every model, including set-identified ones. Assuming a single prior under set identification is however problematic from a robustness viewpoint as the choice of a single prior can lead to spuriously informative posterior inference for the object of interest (Baumeister and Hamilton (2015)). The severity of the problem is magnified by the fact that the effect of the prior choice persists asymptotically, unlike in the case of point-identified models (Moon and Schorfheide (2012), Poirier (1998), among others).

The key innovation of our approach to Bayesian model averaging is that we do not assume availability of a single posterior for the set-identified models. Rather, we allow for multiple priors (*an ambiguous belief*) within the set-identified models (as in Giacomini and Kitagawa (2021)), and then combine the corresponding multiple posteriors with single posteriors for models that are either point-identified or that impose nondogmatic identifying assumptions in the form of a Bayesian prior for the structural parameters (as in Baumeister and Hamilton (2015)). The output of the procedure is a set of posteriors (*post-averaging ambiguous belief*), that are mixtures of the single posteriors and any element of the set of multiple posteriors, with weights equal to the posterior model probabilities. To summarize and visualize the post-averaging ambiguous belief, one can report the set of posterior quantities (e.g., the mean or median) and the associated credible region (an interval to which any posterior in the class assigns a certain credibility level), which are easy to compute in practice.

The method proposed in this paper provides a formal framework for conducting sensitivity analysis of causal inferences to the choice of identifying assumptions. First, one can perform reverse-engineering exercises that compute the minimal prior probability one would need to attach to a set of identifying assumptions in order for the

[1]For example, the SVAR literature often considers models with a large number of sign restrictions (e.g., Amir-Ahmadi and Uhlig (2015), Korobilis (2020), Furlanetto, Ravazzolo, and Sarferaz (2019), and Matthes and Schwartzman (2019)).

[2]When a constrained model is a lower dimensional submodel of a large model, performing inference conditional on the constrained model may suffer from the Borel paradox; see, for example, Drèze and Richard (1983). Bayesian model averaging offers a practical way to avoid the Borel paradox in such context.

averaging to obtain a certain conclusion (e.g., that the set of posterior means for the impact response of output to a monetary policy shock is contained in the negative real half-line). This exercise has a similar motivation as the breakdown frontier analysis in Horowitz and Manski (1995) and Masten and Poirier (2020). Second, when a set-identified model nests a point-identified model, our method can be used to assess the posterior sensitivity in the point-identified model with respect to perturbations of the prior in the direction of relaxing some of the point-identifying assumptions. This exercise can be seen as an example of the ϵ-contamination sensitivity analysis developed in Huber (1973) and Berger and Berliner (1986). Our approach to sensitivity analysis therefore differs from and complements the approaches proposed by Giacomini, Kitagawa, and Uhlig (2019) and Ho (2019), which specify the class of priors as a Kullback–Leibler neighborhood of a benchmark prior.

Our method can also be viewed as bridging the gap between point- and set-identification. When focusing solely on a point-identified model, a researcher who is not fully confident about the choice of identifying assumptions may doubt the robustness of the conclusions. On the other hand, discarding some of the point-identifying assumptions and reporting estimates of the identified set may appear "excessively agnostic," and often results in uninformative conclusions. Our averaging procedure reconciles these two extreme representations of the posterior beliefs by exploiting the prior weights that one can assign to alternative sets of identifying assumptions.

This paper contributes to the growing literature on Bayesian inference for partially identified models (Giacomini and Kitagawa (2021), Kline and Tamer (2016), Moon and Schorfheide (2012)). We follow the multiple-prior approach to model the lack of knowledge within the identified set as in Giacomini and Kitagawa (2021). When a set-identified model is the only model considered, the set of posteriors generated by the approach provides posterior inference for the identified set. When there is uncertainty about the identifying assumptions, however, the usual definition of identified set is not available without conditioning on the model. The multiple prior viewpoint has an advantage in this case since the set of posteriors has a well-defined subjective interpretation even in the presence of model uncertainty.

The paper makes two main analytical contributions to the literature on Bayesian model selection and averaging. First, we clarify under which conditions the prior model probabilities can be updated by data. We show that the updating occurs if some models are "distinguishable" for some distribution of data and/or the priors for the reduced-form parameters differ across models. Second, we investigate the asymptotic properties of the posterior model probabilities and of the averaging method. We show that, when only one model is consistent with the true distribution of the data, our method asymptotically assigns probability one to it. When multiple models are observationally equivalent and "not falsified" at the true data generating process, the posterior model probabilities asymptotically assign nontrivial weights to them. We clarify what part of the prior input determines the asymptotic posterior model probabilities in such case. The consistency property of Bayesian model selection has been well studied in the statistics literature (e.g., Claeskens and Hjort (2008) and references therein), but there is no discussion about the asymptotic behavior of posterior model probabilities when the models differ

in terms of the identifying assumptions but can be observationally equivalent in terms of their reduced form representations. These new results therefore could be of separate interest.

The empirical application in this paper considers SVAR analysis with uncertainty over the classes of identifying assumptions typically used in empirical work. The choice of identifying assumptions has often been a source of controversy in this literature, and researchers have differing opinions about their credibility. To our knowledge, little work has been done on multimodel inference in the SVAR literature, and the methods proposed in this paper could therefore prove helpful in reconciling the controversies about the identifying assumptions that are widespread in this literature. As an example, the empirical application documents the high sensitivity of the conclusion in standard monetary SVARs that output decreases after a contractionary monetary policy shock to the choice of identifying assumptions.

The remainder of the paper is organized as follows. Section 2 illustrates the motivation and the implementation of the method in the context of a simple model. Section 3 presents the formal analysis in a general framework and provides a computational algorithm to implement the procedure. Section 4 applies our method to impulse response analysis in monetary SVARs. The Appendix in the Online Supplementary Material (Giacomini, Kitagawa, and Volpicella (2021)) contains proofs and details about computation.

2. Illustrative example

We present the key ideas and the implementation of the method in a price-quantity static model, subject to common types of identifying assumptions. The model is

$$A\begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, t = 1, \ldots, T, \tag{2.1}$$

where (q_t, p_t) are price and quantity of a certain good/service in a given market and $(\epsilon_t^d, \epsilon_t^s)$ is an i.i.d. normally distributed vector of demand and supply shocks with variance-covariance the identity matrix. A is the structural parameter and the contemporaneous impulse responses are elements of A^{-1}. For example, in the labor market (q_t, p_t) can be replaced by employment and wages, respectively.

The reduced-form model is indexed by Σ, the variance-covariance matrix of (q_t, p_t), which satisfies $\Sigma = A^{-1}(A^{-1})'$. Denote its lower triangular Cholesky decomposition with nonnegative diagonal elements by $\Sigma_{\text{tr}} = \left(\begin{smallmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{smallmatrix}\right)$ with $\sigma_{11} \geq 0$ and $\sigma_{22} \geq 0$, and define the reduced form parameter as $\phi = (\sigma_{11}, \sigma_{21}, \sigma_{22}) \in \Phi = \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.[3] Let the mapping from the structural parameter to the reduced-form parameter be denoted by $\phi = g(A)$.

Suppose the object of interest is the response of the first variable to a unit positive shock in the first variable, $\alpha \equiv (1, 1)$-element of A^{-1}. Without identifying assumptions, the structural parameter is set-identified since knowledge of the reduced-form parameter ϕ cannot uniquely pin down the structural parameter ($\phi = g(A)$ is a many-to-one

[3]The positive semidefiniteness of Σ does not constrain the value of ϕ other than $\sigma_{11} \geq 0$ and $\sigma_{22} \geq 0$.

mapping). Imposing assumptions can lead to a set or a point for α, depending on the type and number of assumptions.

A Bayesian model is the combination of a likelihood and a prior input. The prior input can be either a single prior or multiple priors. In point-identified models, the prior input is a single prior for the structural parameter A, which implies the prior for the reduced-form parameter ϕ. In set-identified models, one could either specify a single prior for A (e.g., as a way of imposing nondogmatic identifying assumptions) or consider multiple priors as in Giacomini and Kitagawa (2021). In the latter case, a model is the combination of a likelihood, a single prior for the reduced-form parameter ϕ (which is revised) and multiple priors for $A|\phi$ (which are not revised).[4]

The division that we introduce in the paper is between single-prior models (which could be point- or set-identified) and multiple-prior models (which are always set-identified). We now illustrate how this interplays with identifying assumptions in two examples.

2.1 *Dogmatic identifying assumptions*

First, consider dogmatic identifying assumptions, which are equality or inequality restrictions on (functions of) the structural parameter that hold with probability one.

Scenario 1: Candidate models

- *Model M^p (point-identified)*: The demand is inelastic to price, $a_{12} = 0$.
- *Model M^s (set-identified)*: The price elasticity of demand is nonpositive, $a_{12} \geq 0$, and the price elasticity of supply is nonnegative, $a_{21} \leq 0$.

Model M^p restricts A to be lower-triangular, as in the classical causal ordering assumptions of Sims (1980) and Bernanke (1986). Combined with the sign normalization restrictions requiring the diagonal elements of A to be nonnegative, the assumption implies that the impulse responses can be identified by $A^{-1} = \Sigma_{\text{tr}}$. The parameter of interest is $\alpha = \alpha_{M^p}(\phi) \equiv \sigma_{11}$.

Model M^s imposes sign restrictions that only set-identify α. The Appendix in the Online Supplementary Material shows that the identified set for α is

$$\text{IS}_\alpha(\phi) \equiv \begin{cases} \left[\sigma_{11}\cos\left(\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right), \sigma_{11}\right], & \text{for } \sigma_{21} > 0, \\ \left[0, \sigma_{11}\cos\left(\arctan\left(-\frac{\sigma_{21}}{\sigma_{22}}\right)\right)\right], & \text{for } \sigma_{21} \leq 0. \end{cases} \tag{2.2}$$

Note that the identified set is nonempty for any ϕ. Hence, models M^p and M^s are observationally equivalent at any $\phi \in \Phi$ and neither of them is falsifiable, that is, for any

[4]See Giacomini and Kitagawa (2021) for a discussion about and motivation for assuming a single prior for ϕ. An additional advantage of this assumption in the context of model selection is that it allows one to isolate the component of the model that depends on the identifying restrictions. This enables one, for example, to compare models that only differ in the restrictions they impose.

$\phi \in \Phi$ in both models there exists a structural parameter A that satisfies the identifying assumptions.[5]

We start by specifying a prior for ϕ in each model. Given the observational equivalence of the two models, it might be reasonable to specify the same prior:

$$\pi_{\phi|M^p} = \pi_{\phi|M^s} = \tilde{\pi}_\phi, \tag{2.3}$$

where $\tilde{\pi}_\phi$ is a *proper* prior, such as the one induced by a Wishart prior on Σ. The same prior for ϕ in observationally equivalent models leads to the same posterior:

$$\pi_{\phi|M^p,Y} = \pi_{\phi|M^s,Y} = \tilde{\pi}_{\phi|Y}. \tag{2.4}$$

In model M^p, the posterior for ϕ implies a unique posterior for α, $\pi_{\alpha|M^p,Y}$, via the mapping $\alpha = \alpha_{M^p}(\phi)$. In model M^s, on the other hand, the posterior for ϕ does not yield a unique posterior for α, since the mapping in (2.2) is generally set-valued. Following Giacomini and Kitagawa (2021), we formulate the lack of prior knowledge by considering multiple priors (ambiguous belief). Formally, given the single prior $\pi_{\phi|M^s}$, we form the class of priors for A by admitting arbitrary conditional priors for A given ϕ, as long as they are consistent with the identifying assumptions:

$$\Pi_{A|M^s} \equiv \left\{ \pi_{A|M^s} = \int_\Phi \pi_{A|M^s,\phi}\, d\pi_{\phi|M^s} : \pi_{A|M^s,\phi}\big(\mathcal{A}_{\text{sign}} \cap g^{-1}(\phi)\big) = 1,\ \pi_{\phi|M^s}\text{-a.s.} \right\},$$

where $\mathcal{A}_{\text{sign}} = \{A : a_{12} \geq 0,\ a_{21} \leq 0,\ \text{diag}(A) \geq 0\}$ is the set of structural parameters that satisfy the sign restrictions and the sign normalizations and $g^{-1}(\phi)$ is the set of observationally equivalent structural parameters given the reduced-form parameter ϕ.

Since the likelihood depends on the structural parameter only through the reduced-form parameter, applying Bayes' rule to each prior in the class only updates the prior for ϕ, and thus leads to the following class of posteriors for A:

$$\Pi_{A|M^s,Y} \equiv \left\{ \pi_{A|M^s,Y} = \int_\Phi \pi_{A|M^s,\phi}\, d\pi_{\phi|M^s,Y} : \pi_{A|M^s,\phi}\big(\mathcal{A}_{\text{sign}} \cap g^{-1}(\phi)\big) = 1,\ \pi_{\phi|M^s}\text{-a.s.} \right\}. \tag{2.5}$$

Marginalizing the posteriors in $\Pi_{A|M^s,Y}$ to α leads to the class of α-posteriors:

$$\Pi_{\alpha|M^s,Y} \equiv \left\{ \pi_{\alpha|M^s,Y} = \int_{\tilde{\Phi}} \pi_{\alpha|M^s,\phi}\, d\pi_{\phi|M^s,Y} : \pi_{\alpha|M^s,\phi}(\text{IS}_\alpha(\phi)) = 1,\ \pi_{\phi|M^s}\text{-a.s.} \right\}. \tag{2.6}$$

We view this class as a representation of the posterior uncertainty about α in the set-identified model. The class contains any α-posterior that assigns probability one to the identified set, and it represents the lack of belief therein in terms of Knightian uncertainty (ambiguity). This is a key departure from the standard approach to Bayesian

[5]When $\sigma_{21} > 0$, the point-identified α in model M^p is the upper bound of the identified set in model M^s, whereas when $\sigma_{21} < 0$, the identified set in model M^s does not contain the point-identified α. This is because in model M^p we have $a_{12} = -\frac{\sigma_{21}}{\sigma_{11}\sigma_{22}}$, which is positive if $\sigma_{21} < 0$, meaning that the point-identifying assumptions $a_{12} = 0$ and $\sigma_{21} < 0$ are not compatible with the restriction $a_{21} \leq 0$.

model averaging, which requires a single posterior for all models, including those where the parameter is set-identified.

Suppose that the researcher's prior uncertainty over the two models can be represented by prior probabilities $\pi_{M^p} \in [0, 1]$ for model M^p and $(1 - \pi_{M^p})$ for model M^s. Our proposal is to combine the single posterior for α in model M^p and the set of posteriors for α in model M^s according to the posterior model probabilities $\pi_{M^p|Y}$ and $\pi_{M^s|Y}$ (the posterior model probability for model M^s depends only on the single prior for the reduced-form parameter, so it is unique in spite of the multiple priors for the structural parameter). The combination delivers a class of posteriors $\Pi_{\alpha|Y}$, the *post-averaging ambiguous belief*:

$$\Pi_{\alpha|Y} = \{\pi_{\alpha|M^p,Y}\pi_{M^p|Y} + \pi_{\alpha|M^s,Y}\pi_{M^s|Y} : \pi_{\alpha|M^s,Y} \in \Pi_{\alpha|M^s,Y}\}. \tag{2.7}$$

As we show in Section 4.1, our proposal can be interpreted as applying Bayes' rule to each prior in a class that has the form of an ϵ-contaminated class of priors (Berger and Berliner (1986)).

A key result of the paper is to establish conditions under which the prior model probabilities are updated by the data, which we show occurs when the models are "distinguishable" for some reduced-form parameter values and/or they specify different priors for ϕ (see Lemma 3.1 below). In the current scenario, the two models are indistinguishable, so the prior model probabilities are not updated if they use a common ϕ-prior.

In practice, we recommend reporting as the output of the procedure the post-averaging set of posterior means or quantiles of $\Pi_{\alpha|Y}$ and its associated *robust credible region* with credibility $\gamma \in (0, 1)$, defined as the shortest interval that receives posterior probability at least γ for every posterior in $\Pi_{\alpha|Y}$. Proposition 3.1 shows that the set of posterior means is the weighted average of the posterior mean in model M^p and the set of posterior means in model M^s:

$$\begin{aligned}&\Big[\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha), \sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha)\Big] \\ &\quad = \pi_{M^p|Y}E_{\alpha|M^p,Y}(\alpha) + \pi_{M^s|Y}\big[E_{\phi|M^s,Y}\big(l(\phi)\big), E_{\phi|M^s,Y}\big(u(\phi)\big)\big],\end{aligned} \tag{2.8}$$

where $(l(\phi), u(\phi))$ are the lower and upper bounds of the nonempty identified set for α shown in (2.2), $a + b[c, d]$ stands for $[a + bc, a + bd]$, and $E_{\phi|M^s,Y}(\cdot)$ denotes the posterior mean with respect to $\pi_{\phi|M^s,Y} = \tilde{\pi}_{\phi|Y}$. Since the set of posterior means can be viewed as an estimator for the identified set in model M^s, our procedure effectively shrinks the estimate of the identified set in the set-identified model toward the point estimate in the point-identified model, with the amount of shrinkage determined by the posterior model probabilities.

The robust credible region for α with credibility γ can be computed as follows. We first draw $z_1, \ldots, z_G$ randomly from a Bernoulli distribution with mean $\pi_{M^p|Y}$ and then generate $g = 1, \ldots, G$ random draws of the "mixture identified set" for α according to

$$\mathrm{IS}_\alpha^{\mathrm{mix}}(\phi_g) = \begin{cases} \{\alpha(\phi_g)\}, & \phi_g \sim \pi_{\phi|M^p,Y} = \tilde{\pi}_{\phi|Y} \quad \text{if } z_g = 1, \\ \big[l(\phi_g), u(\phi_g)\big], & \phi_g \sim \pi_{\phi|M^s,Y} = \tilde{\pi}_{\phi|Y} \quad \text{if } z_g = 0. \end{cases} \tag{2.9}$$

Intuitively, with probability $\pi_{M^p|Y}$, a draw of the mixture identified set is a singleton corresponding to the point-identified value of α, and with probability $\pi_{M^s|Y}$ it is a nonempty identified set for α. The robust credible region with credibility level γ is approximated by an interval that contains the γ-fraction of the drawn $\mathrm{IS}_\alpha^{\mathrm{mix}}(\phi)$'s. The minimization problem in Step 5 of Algorithm 4.1 in Giacomini and Kitagawa (2021) is solved to obtain the shortest-width robust credible region.

Our method lends itself to reverse-engineering exercises that help shed light on the role of identifying assumptions in drawing inferences. For instance, we can compute the prior weight w one would assign to the restriction in M^p such that the set of posterior means is contained in the positive real half-line. In the current example, the prior probabilities are not updated, since the two models are observationally equivalent. We would hence obtain the weights w by solving $wE_{\alpha|M^p,Y}(\alpha) + (1 - w)[E_{\phi|M^s,Y}(l(\phi)), E_{\phi|M^s,Y}(u(\phi))] \geq 0$ as a function of w.

2.2 *Nondogmatic identifying assumptions*

Our method allows for identifying assumptions that are expressed as a nondogmatic prior for the structural parameter.

Scenario 2: Candidate models

- *Model M^B (single prior)*: A prior for the structural parameter A.
- *Model M^s (multiple priors)*: Same as the set-identified model in Scenario 1.

Model M^B assumes availability of a prior for the whole structural parameter. This prior can reflect Bayesian probabilistic uncertainty about identifying assumptions expressed as equalities (see, e.g., Baumeister and Hamilton (2015), who propose a prior for a dynamic version of the current model based on a metaanalysis of the literature). Another key example of a model that implies a single prior for the structural parameter is a Bayesian DSGE model.

Model M^B always yields a single posterior for α. However, the influence of prior choice does not vanish asymptotically due to the lack of identification. In principle, if the researcher were confident about the prior specification in model M^B, she could perform standard Bayesian inference and obtain a credible posterior, despite the identification issues. In practice, this is rather rare. For instance, the prior considered by Baumeister and Hamilton (2015) is based on the elicitation of first and second moments and the remaining characteristics of the distribution are chosen for analytical or computational convenience. Further, eliciting dependence among structural parameters is challenging, and an independent prior could lead to unintended or counterintuitive effects on posterior inference.[6] These robustness concerns can be addressed by averaging model M^B with the set-identified model M^s, which accommodates the lack of prior knowledge about the structural parameter (beyond the inequality restrictions).

One important consideration in this scenario is that the single prior for A in model M^B implies a single prior for ϕ. Here, we thus allow the prior for ϕ in model M^s to differ

[6]"Knowing no dependence" among the parameters differs from "not knowing their dependence."

from that in model M^B. This, in turn, affects the posterior model probabilities, which are given by

$$\begin{aligned}\pi_{M^B|Y} &= \frac{p(Y|M^B)\cdot\pi_{M^B}}{p(Y|M^B)\cdot\pi_{M^B}+p(Y|M^s)\cdot(1-\pi_{M^B})},\\ \pi_{M^s|Y} &= \frac{p(Y|M^s)\cdot(1-\pi_{M^B})}{p(Y|M^B)\cdot\pi_{M^B}+p(Y|M^s)\cdot(1-\pi_{M^B})},\end{aligned} \tag{2.10}$$

where π_{M^B} is the prior weight assigned to model M^B, $p(Y|M)\equiv\int_\Phi p(Y|\phi,M)\,d\pi_{\phi|M}(\phi)$, $M=M^B,M^s$, are the marginal likelihoods of model M with $p(Y|\phi,M)$ the likelihood of the reduced form parameters. In this scenario, the different priors for ϕ imply $p(Y|M^B)\neq p(Y|M^s)$ and, therefore, the prior model probabilities can be updated by the data.[7]

Given these posterior model probabilities, the construction of the post-averaging ambiguous belief proceeds as in (2.7). The set of posterior means for α can be obtained similar to (2.8), where M^B replaces M^p. The robust credible region can be constructed as in Scenario 1, by drawing i.i.d. draws $z_1,\dots,z_G\sim\text{Bernoulli}(\pi_{M^B|Y})$ and letting

$$\text{IS}^{\text{mix}}_{\alpha,g}=\begin{cases}\{\alpha\}, & \alpha\sim\pi_{\alpha|M^B,Y} \quad \text{if } z_g=1,\\ \big[l(\phi_g),u(\phi_g)\big], & \phi_g\sim\pi_{\phi|M^s,Y} \quad \text{if } z_g=0.\end{cases} \tag{2.11}$$

The reverse-engineering described at the end of Section 2.1 can also be applied in this scenario, with the difference that the weights w and $1-w$ are now substituted by the updated posterior model probabilities, $\pi_{M^B|Y}=\frac{p(Y|M^B)\cdot w}{p(Y|M^B)\cdot w+p(Y|M^s)\cdot(1-w)}$ and $\pi_{M^s|Y}=\frac{p(Y|M^s)\cdot(1-w)}{p(Y|M^B)\cdot w+p(Y|M^s)\cdot(1-w)}$.

3. Formal analysis

3.1 *Notation and definitions*

Consider $J+K\geq 2$ models, $J,K\geq 0$, where J models are *single-prior models* collected in the class $\mathcal{M}^p$ and K models are *multiple-prior models* collected in the class $\mathcal{M}^s$.

Let $\mathcal{M}\equiv\mathcal{M}^p\cup\mathcal{M}^s$. The structural parameters in model $M\in\mathcal{M}$ is $\theta_M\in\Theta_M$, where Θ_M embeds the identifying assumptions imposed in model M. We assume that the scalar parameter of interest $\alpha=\alpha_M(\theta_M)\in\mathbb{R}$ is well-defined as a function of θ_M and it carries a common (causal) interpretation in all models. The reduced-form parameter is $\phi_M=g_M(\theta_M)\in\mathbb{R}^{d_M}$, where $g_M(\cdot)$ maps a set of observationally equivalent structural parameters subject to the identifying assumptions in model M to a point in the

[7]Section 3 shows that, if we add identifying restrictions so that M^s becomes falsifiable, the posterior model probabilities become $\pi_{M^B|Y}=\frac{p(Y|M^B)\cdot\pi_{M^B}}{p(Y|M^B)\cdot\pi_{M^B}+p(Y|M^s)\cdot O_{M^s}\cdot(1-\pi_{M^B})}$ and $\pi_{M^s|Y}=\frac{p(Y|M^s)\cdot O_{M^s}\cdot(1-\pi_{M^B})}{p(Y|M^B)\cdot\pi_{M^B}+p(Y|M^s)\cdot O_{M^s}\cdot(1-\pi_{M^B})}$, where O_{M^s} is the posterior-prior plausibility ratio.

reduced-form parameter space $\Phi_M = g_M(\Theta_M)$.[8] Our most general set-up allows the parameter space of structural and reduced-form parameters to differ across models. We express the likelihood in model $M \in \mathcal{M}$ in terms of the reduced-form parameter by $p(Y|\phi_M, M)$.[9] For $M \in \mathcal{M}^s$, we define the identified set of α by $\mathrm{IS}_\alpha(\phi_M|M) = \{\alpha_M(\theta_M) : \theta_M \in \Theta_M \cap g_M^{-1}(\phi_M)\}$, which is a set-valued mapping from Φ_M to $\mathbb{R}$.

We next introduce the concept of identical reduced-forms.

DEFINITION 3.1. A class of models $\mathcal{M}$ **admits an identical reduced-form** if:

(a) Φ_M can be embedded into a common d-dimensional Euclidean space $\mathbb{R}^d$ for all $M \in \mathcal{M}$ (hence ϕ_M can be denoted by $\phi \in \mathbb{R}^d$).

(b) For every $M \in \mathcal{M}$, $p(Y|\phi_M = \phi, M)$ defines a probability distribution of Y on the extended domain $\phi \in \Phi \equiv \bigcup_{M\in\mathcal{M}} \Phi_M$, and $p(Y|\phi_M = \phi, M) = p(Y|\phi)$ holds for all $\phi \in \Phi$, where $p(Y|\phi)$ is the likelihood common among $M \in \mathcal{M}$.

Definition 3.1 formalizes the situation where models imposing different identifying assumptions lead to the same family of distributions for the observables (different identifying assumptions, nonetheless, can lead to different Φ_M). For instance, if $\mathcal{M}$ consists of SVAR models with the same set of variables but different identifying assumptions, Definition 3.1 is satisfied when the reduced-form VARs implied by the models feature the same variables and lag length.

We next introduce the concepts of observational equivalence and distinguishability.

DEFINITION 3.2. (i) The models in $\mathcal{M}$ are **observationally equivalent at** ϕ if $\mathcal{M}$ admits an identical reduced-form and $\phi \in \bigcap_{M\in\mathcal{M}} \Phi_M$.

(ii) $M, M' \in \mathcal{M}$ that admit an identical reduced-form are **distinguishable** if $\Phi_M \neq \Phi_{M'}$.

(iii) The models in $\mathcal{M}$ are **indistinguishable** if $\mathcal{M}$ admits an identical reduced-form and $\Phi_M = \Phi$ for all $M \in \mathcal{M}$.

Note that our definition of observational equivalence is local to ϕ, and it does not constrain the relationship among the reduced-form parameter spaces for different models (except that they must have a nonempty intersection). On the other hand, indistinguishability can be interpreted as observational equivalence of the models in a global sense—if the models are indistinguishable, one could not find support for one model rather than the others based on the data, regardless of any available knowledge about the distribution of observables.

[8] Φ_M incorporates any testable implications of the imposed identifying assumptions. For a set-identified model, Φ_{M^s} is equivalent to the set of ϕ_M's that yield a nonempty identified set, $\Phi_{M^s} = \{\phi_{M^s} \in \mathbb{R}^{d_{M^s}} : \mathrm{IS}_\alpha(\phi_{M^s}|M^s) \neq \emptyset\}$.

[9] The likelihood $\tilde{p}(Y|\theta_M, M)$ depends on θ_M only through the reduced-form parameters $g_M(\theta_M)$ for any realization of Y, that is, there exists $p(Y|\cdot, M)$ such that $\tilde{p}(Y|\theta_M, M) = p(Y|g_M(\theta_M), M)$ holds for every Y and $\phi_M = g_M(\theta_M)$ is identifiable.

3.2 *Prior and posterior model probabilities*

This section shows when and how the data update the prior model probabilities.

Let $(\pi_M : M \in \mathcal{M})$, $\sum_{M\in\mathcal{M}} \pi_M = 1$, be prior probabilities assigned over $\mathcal{M}$. By Bayes' rule, the posterior probability for each model is

$$\pi_{M|Y} = \frac{p(Y|M)\pi_M}{\sum_{M'\in\mathcal{M}} p\big(Y|M'\big)\pi_{M'}}. \tag{3.1}$$

Since the marginal likelihood depends only on the ϕ_M-prior, which we assume to be a single prior for all $M \in \mathcal{M}$, the posterior model probabilities are unique for all models.[10]

The next lemma obtaines posterior model probabilities when the models admit an identical reduced form.

LEMMA 3.1. (i) *Suppose that $M^s \in \mathcal{M}^s$ admit an identical reduced form with $\phi \in \Phi = \bigcup_{M^s\in\mathcal{M}^s} \Phi_{M^s} \subset \mathbb{R}^d$. Let $\tilde{\pi}_\phi$ be a proper prior on Φ and assume that $\tilde{\pi}_\phi(\Phi_{M^s}) = \tilde{\pi}_\phi(\mathrm{IS}_\alpha(\phi|M^s) \neq \emptyset) > 0$ for all $M^s \in \mathcal{M}^s$. Let $\tilde{\pi}_{\phi|Y}$ be the posterior obtained by updating $\tilde{\pi}_\phi$ with the common likelihood $p(Y|\phi)$. Suppose that the ϕ-prior is obtained by trimming the support of $\tilde{\pi}_\phi$ to Φ_{M^s}:*

$$\pi_{\phi|M^s}(B) = \frac{\tilde{\pi}_\phi(B \cap \Phi_{M^s})}{\tilde{\pi}_\phi(\Phi_{M^s})}, \quad B \in \mathcal{B}(\Phi), \tag{3.2}$$

where $\mathcal{B}(\Phi)$ is the Borel σ-algebra of Φ. Then the posterior model probabilities are given by

$$\begin{cases} \pi_{M^p|Y} = \dfrac{p\big(Y|M^p\big)\pi_{M^p}}{\sum_{M^{p\prime}\in\mathcal{M}^p} p\big(Y|M^{p\prime}\big)\pi_{M^{p\prime}} + \tilde{p}(Y)\sum_{M^{s\prime}\in\mathcal{M}^s} O_{M^{s\prime}}\pi_{M^{s\prime}}}, & \text{for } M^p \in \mathcal{M}^p, \\ \pi_{M^s|Y} = \dfrac{\tilde{p}(Y)O_{M^s}\pi_{M^s}}{\sum_{M^{p\prime}\in\mathcal{M}^p} p\big(Y|M^{p\prime}\big)\pi_{M^{p\prime}} + \tilde{p}(Y)\sum_{M^{s\prime}\in\mathcal{M}^s} O_{M^{s\prime}}\pi_{M^{s\prime}}}, & \text{for } M^s \in \mathcal{M}^s, \end{cases} \tag{3.3}$$

where O_{M^s} is the posterior-prior plausibility ratio of the set-identifying assumptions of model $M^s \in \mathcal{M}^s$ and $\tilde{p}(Y)$ is the marginal likelihood with respect to $\tilde{\pi}_\phi$,

$$\begin{aligned} O_{M^s} &\equiv \frac{\tilde{\pi}_{\phi|Y}(\Phi_{M^s})}{\tilde{\pi}_\phi(\Phi_{M^s})} = \frac{\tilde{\pi}_{\phi|Y}\big(\mathrm{IS}_\alpha\big(\phi|M^s\big) \neq \emptyset\big)}{\tilde{\pi}_\phi\big(\mathrm{IS}_\alpha\big(\phi|M^s\big) \neq \emptyset\big)}, \\ \tilde{p}(Y) &= \int_\Phi p(Y|\phi)\, d\tilde{\pi}_\phi(\phi). \end{aligned} \tag{3.4}$$

(ii) *Suppose that, in addition to $\mathcal{M}^s$, all the models in $\mathcal{M}^p$ admit an identical reduced form. Let $\tilde{\pi}_\phi$ be as in* (i) *of the current lemma and assume $\tilde{\pi}_\phi(\Phi_M) > 0$ for all*

[10]Note that our method introduces ambiguous beliefs for the nonidentifiable parameters, while it assumes availability of prior model probabilities even when the models are indistinguishable. Hence, we are not treating nonidentifiability of the parameters and of the models in a symmetric way.

$M \in \mathcal{M}$. If the ϕ-prior satisfies (3.2) in every $M \in \mathcal{M}$, then the posterior model probabilities further simplify to

$$\pi_{M|Y} = \frac{O_M \pi_M}{\sum_{M' \in \mathcal{M}} O_{M'} \pi_{M'}} \quad \text{for } M \in \mathcal{M}, \tag{3.5}$$

where $O_M = \frac{\tilde{\pi}_{\phi|Y}(\Phi_M)}{\tilde{\pi}_{\phi}(\Phi_M)}$.

(iii) *If all models are indistinguishable and the ϕ-prior is common, then the model probabilities are never updated, $\pi_{M|Y} = \pi_M$ for all $M \in \mathcal{M}$ and for any realization of Y.*

Lemma 3.1 clarifies the sources of updating of the prior model probabilities. In claim (i), the specification of the ϕ-prior as in (3.2) simplifies the marginal likelihood of $M^s \in \mathcal{M}^s$ to $\tilde{p}(Y)O_{M^s}$. If all the models admit an identical reduced-form (claim (ii)), the posterior model probabilities only depend on $\{O_M : M \in \mathcal{M}\}$. Claim (iii) shows the intuitive result that model probabilities are not updated if all the models are indistinguishable and share a unique ϕ-prior.

3.3 *Post-averaging ambiguous belief and the set of posteriors*

Estimation of the single-prior models proceeds in the standard Bayesian way. We therefore take the posterior $\pi_{\alpha|M^p,Y}$ as given.

We perform posterior inference for model $M^s \in \mathcal{M}^s$ in the robust Bayesian way: we specify a single proper prior $\pi_{\phi_{M^s}|M^s}$ that is supported on Φ_{M^s}, and form the set of priors for θ_{M^s} as

$$\Pi_{\theta_{M^s}|M^s} \equiv \left\{\pi_{\theta_{M^s}|M^s} : \pi_{\theta_{M^s}|M^s}\left(\Theta_{M^s} \cap g_{M^s}^{-1}(B)\right) = \pi_{\phi_{M^s}|M^s}(B), \forall B \in \mathcal{B}(\Phi_{M^s})\right\}, \tag{3.6}$$

where $\mathcal{B}(\Phi_{M^s})$ is the Borel σ-algebra of Φ_{M^s}.[11] Applying Bayes' rule to each θ_M-prior in $\Pi_{\theta_{M^s}|M^s}$ with the likelihood, $\tilde{p}(Y|\theta_{M^s}, M^s)$,[12] and marginalizing the resulting posterior of θ_M via $\alpha = \alpha_M(\theta_M)$, we obtain the following set of posteriors for α:[13]

$$\Pi_{\alpha|M^s,Y} \equiv \left\{\pi_{\alpha|M^s,Y} = \int_{\Phi_M} \pi_{\alpha|M^s,\phi_{M^s}} \, d\pi_{\phi_{M^s}|M^s,Y} : \pi_{\alpha|M^s,\phi_{M^s}}\left(\mathrm{IS}_\alpha\left(\phi_{M^s}|M^s\right)\right) = 1, \right.$$
$$\left. \pi_{\phi_{M^s}|M^s}\text{-a.s.}\right\}. \tag{3.7}$$

[11] By noting that the constraints in (3.6) are rewritten as $\int_B \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi)) \, d\pi_{\phi_{M^s}|M^s}(\phi_{M^s}) = \pi_{\phi_{M^s}|M^s}(B)$ for all $B \in \mathcal{B}(\Phi_{M^s})$, the prior class (3.6) can be equivalently represented as

$$\Pi_{\theta_{M^s}|M^s} = \left\{\int_{\Phi_{M^s}} \pi_{\theta_{M^s}|\phi_{M^s},M^s} \, d\pi_{\Phi_{M^s}|M^s} : \pi_{\theta_{M^s}|\phi_{M^s},M^s}\left(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s})\right) = 1, \, \pi_{\phi_{M^s}|M^s,Y}\text{-a.s.}\right\}.$$

This alternative expression is exploited in the illustrative example of Section 2.

[12] The likelihood of θ_M is linked to the likelihood of ϕ_M via $\tilde{p}(Y|\theta_{M^s}, M^s) = p(Y|g(\theta_{M^s}), M^s)$ by the definition of reduced-form parameters.

[13] Lemma A.1 in the Appendix of the Online Supplementary Material shows a formal derivation of $\Pi_{\alpha|M^s,Y}$.

Given the posterior model probabilities, a posterior for α with the models averaged out is

$$\pi_{\alpha|Y} = \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y} \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\alpha|M^s,Y} \pi_{M^s|Y},$$

where the α-posterior for $M^p \in \mathcal{M}^p$ is unique, while there are multiple α-posteriors for $M^s \in \mathcal{M}^s$ as shown in (3.7). The set of averaged posteriors can be represented as

$$\Pi_{\alpha|Y} = \left\{ \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y} \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\alpha|M^s,Y} \pi_{M^s|Y} : \pi_{\alpha|M^s,Y} \in \Pi_{\alpha|M^s,Y} \ \forall M^s \in \mathcal{M}^s \right\}. \tag{3.8}$$

The next proposition provides a formal robust Bayes' justification for our averaging formula (3.8) when the structural parameters are common across all models.[14]

PROPOSITION 3.1. *Suppose that structural parameters are common in all models, $\theta_M = \theta \in \mathbb{R}^{d_\theta}$ for all $M \in \mathcal{M}$, and define $\Theta = \bigcup_{M \in \mathcal{M}} \Theta_M \subset \mathbb{R}^{d_\theta}$. Consider prior model probabilities $(\pi_M : M \in \mathcal{M})$, a prior $\pi_{\theta|M^p}$ for θ in $M^p \in \mathcal{M}^p$, and a prior for the reduced-form parameters in $M^s \in \mathcal{M}^s$. Define a set of priors for $(\theta, M) \in \Theta \times \mathcal{M}$:*

$$\Pi_{\theta,M} \equiv \left\{ \pi_{\theta,M} = \pi_{\theta|M} \pi_M : \pi_{\theta|M^s} \in \Pi_{\theta|M^s} \text{ for every } M^s \in \mathcal{M}^s \right\}, \tag{3.9}$$

where $\Pi_{\theta|M^s}$ is defined in (3.6). Then Bayes' rule applied to each prior in $\Pi_{\theta,M}$ with likelihood $\tilde{p}(Y|\theta, M)$ and marginalization to α yields (3.8) as the class of posteriors for α.

The next proposition derives the set of posterior means and the posterior probabilities when the posterior for α varies within $\Pi_{\alpha|Y}$.

PROPOSITION 3.2. *Let $[l(\phi_{M^s}|M^s), u(\phi_{M^s}|M^s)]$ be the convex hull of the identified set $\mathrm{IS}_\alpha(\phi_{M^s}|M^s)$ in model $M^s \in \mathcal{M}^s$.*

(i) *The set of posterior means of $\Pi_{\alpha|Y}$ is the convex interval with lower and upper bounds*:

$$\begin{aligned}
\inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = &\sum_{M^p \in \mathcal{M}^p} E_{\alpha|M^p,Y}(\alpha) \pi_{M^p|Y} \\
&+ \sum_{M^s \in \mathcal{M}^s} E_{\phi_{M^s}|Y,M^s}\left[l\left(\phi_{M^s}|M^s\right)\right] \pi_{M^s|Y}, \\
\sup_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = &\sum_{M^p \in \mathcal{M}^p} E_{\alpha|M^p,Y}(\alpha) \pi_{M^p|Y} \\
&+ \sum_{M^s \in \mathcal{M}^s} E_{\phi_{M^s}|Y,M^s}\left[u\left(\phi_{M^s}|M^s\right)\right] \pi_{M^s|Y},
\end{aligned}$$

where $E_{\phi_{M^s}|Y,M^s}(\cdot)$ is the expectation with respect to the posterior of ϕ_{M^s}.

[14]The reason we assume a common structural parameter space is to ensure that we can construct a prior distribution on the product space of the structural parameter space and the model space.

(ii) *For any measurable subset H in $\mathbb{R}$, the lower and upper bounds of the posterior probabilities on $\{\alpha \in H\}$ in the class $\Pi_{\alpha|Y}$ (the lower and upper posterior probabilities of $\Pi_{\alpha|Y}$) are*

$$\begin{aligned}\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) &= \sum_{M^p\in\mathcal{M}^p} \pi_{\alpha|M^p,Y}(H)\pi_{M^p|Y} \\ &\quad + \sum_{M^s\in\mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}\big(\mathrm{IS}_\alpha\big(\phi_{M^s}|M^s\big)\subset H\big)\cdot\pi_{M^s|Y},\\ \sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) &= \sum_{M^p\in\mathcal{M}^p} \pi_{\alpha|M^p,Y}(H)\pi_{M^p|Y} \\ &\quad + \sum_{M^s\in\mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}\big(\mathrm{IS}_\alpha\big(\phi_{M^s}|M^s\big)\cap H\neq\emptyset\big)\cdot\pi_{M^s|Y}.\end{aligned}$$

If $\mathrm{IS}_\alpha(\phi_{M^s}|M^s)$ is a connected interval at every reduced-form parameter value, then we can view $[E_{\phi_{M^s}|Y,M^s}[l(\phi_{M^s}|M^s)], E_{\phi_{M^s}|Y,M^s}[u(\phi_{M^s}|M^s)]]$ as an estimator of the identified set in model M^s. We can thus interpret the set of post-averaging posterior means as the weighted Minkowski sum of the Bayesian point estimators in the point-identified models and the identified set estimators in the set-identified models. The second claim of the proposition provides an analytical expression for the lower probability of $\Pi_{\alpha|Y}$ as a mixture of the containment functionals of the random sets, which in turn can be viewed as the containment functional of the *mixture random sets* $\Pr(\mathrm{IS}_\alpha^{\mathrm{mix}}\subset A)$, where $\mathrm{IS}_\alpha^{\mathrm{mix}}$ is generated according to

$$\begin{aligned}&M\sim\mathrm{Multinomial}\big(\{\pi_{M|Y}\}_{M\in\mathcal{M}}\big),\\ &\mathrm{IS}_\alpha^{\mathrm{mix}} = \begin{cases}\{\alpha\}, & \alpha|(M^p,Y)\sim\pi_{\alpha|M^p,Y}, \quad \text{for } M^p\in\mathcal{M}^p,\\ \mathrm{IS}_\alpha\big(\phi_{M^s}|M^s\big), & \phi_{M^s}|(M^s,Y)\sim\pi_{\phi_{M^s}|M^s,Y}, \quad \text{for } M^s\in\mathcal{M}^s.\end{cases}\end{aligned} \tag{3.10}$$

This way of interpreting the lower probability of $\Pi_{\alpha|Y}$ simplifies its computation and justifies the algorithm presented in (2.9).

3.4 *Computation*

To report the set of posteriors based on the analytical expressions in Proposition 3.2, we need to compute (i) the posterior model probabilities (equivalently, the marginal likelihood in each $M\in\mathcal{M}$), (ii) the posterior for α for each single-prior model, and (iii) the identified set $\mathrm{IS}_\alpha(\phi_{M^s}|M^s)$ and the posterior for ϕ_{M^s} for each multiple-prior model. Estimation of the single-prior models in (ii) is standard, and we assume some suitable posterior sampling algorithm is applicable to obtain Monte Carlo draws of $\alpha\sim\pi_{\alpha|M^p,Y}$. For (i), efficient and reliable algorithms to compute the marginal likelihood are available in the literature; for example, see Chib and Jeliazkov (2001), Geweke (1999), and Sims, Waggoner, and Zha (2008). When all the models admit an identical reduced-form, computing the marginal likelihoods is not necessary since the posterior model probabilities depend only on the posterior-prior plausibility ratios O_M.

In each multiple-prior model, the posterior-prior plausibility ratio O_{M^s} can be computed by plugging in numerical approximations for the prior and posterior probabilities of the non-emptiness of the identified set into (3.4). The denominator of O_{M^s} is computed by drawing many ϕ's from the prior $\tilde{\pi}_\phi$ and computing the fraction of draws that yield nonempty identified sets. The numerator of O_{M^s} is computed similarly except that the ϕ's are drawn from the posterior $\tilde{\pi}_{\phi|Y}$. Whether checking the nonemptiness of $\mathrm{IS}_\alpha(\phi|M^s)$ is simple or not depends on the application. In the application in Section 4 to SVARs with sign restrictions, we consider two ways to check the nonemptiness of $\mathrm{IS}_\alpha(\phi|M^s)$. The first (Algorithm A.1 in the Appendix of the Online Supplementary Material) builds on Algorithm 1 of Giacomini and Kitagawa (2021) and assesses nonemptiness based on the Monte Carlo draws of the impulse responses. The second approach (Algorithm A.2 in the Appendix of the Online Supplementary Material), which is novel in the literature and can be of independent interest, exploits the analytical features of the identifying restrictions in sign restricted SVARs. See the Appendix for the details of these algorithms.

Monte Carlo draws of the lower and upper bounds of the identified set in model $M \in \mathcal{M}^s$ can be obtained by first drawing ϕ's from the posterior $\tilde{\pi}_{\phi|Y}$, then retaining the draws of ϕ that yield a nonempty $\mathrm{IS}_\alpha(\phi|M^s)$, and computing the corresponding $l(\phi|M^s)$ and $u(\phi|M^s)$. Their sample averages approximate $E_{\phi|M^s,Y}(l(\phi|M^s))$ and $E_{\phi|M^s,Y}(u(\phi|M^s))$. Implementation of this procedure requires computability of the lower and upper bounds of the identified set for each ϕ. In the SVAR application of Section 4, we compute $l(\phi|M^s)$ and $u(\phi|M^s)$ by numerical optimization.

Utilizing the mixture random set representation shown in (3.10), we can use the following algorithm to approximate the lower posterior probability.

Algorithm 3.1.

Step 1: Draw a model $M \in \mathcal{M}$ from a multinomial distribution with parameters $(\pi_{M|Y} : M \in \mathcal{M})$.

Step 2: If the drawn M belongs to $\mathcal{M}^p$, then draw $\alpha \sim \pi_{\alpha|M,Y}$ and set $\mathrm{IS}_\alpha^{\mathrm{mix}} = \{\alpha\}$ (a singleton). If the drawn M belongs to $\mathcal{M}^s$, draw $\phi_M \sim \pi_{\phi|M,Y}$ and set $\mathrm{IS}_\alpha^{\mathrm{mix}} = \mathrm{IS}_\alpha(\phi_M|M)$.[15]

Step 3: Repeat Steps 1 and 2 many (G) times and obtain G draws of $\mathrm{IS}_\alpha^{\mathrm{mix}}$: $\mathrm{IS}_{\alpha,1}^{\mathrm{mix}}, \ldots, \mathrm{IS}_{\alpha,G}^{\mathrm{mix}}$.

Step 4: Let $[l_g^{\mathrm{mix}}, u_g^{\mathrm{mix}}]$ be the lower and upper bounds of $\mathrm{IS}_{\alpha,g}^{\mathrm{mix}}$, $g = 1, \ldots, G$, where $l_g^{\mathrm{mix}} = u_g^{\mathrm{mix}}$ if $\mathrm{IS}_{\alpha,g}^{\mathrm{mix}}$ is a singleton (i.e., gth draw of M belongs to $\mathcal{M}^p$). Approximate the mean bounds of the post-average posterior class by

$$\inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = \frac{1}{G}\sum_{g=1}^{G} l_g^{\mathrm{mix}}, \qquad \sup_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = \frac{1}{G}\sum_{g=1}^{G} u_g^{\mathrm{mix}}. \tag{3.11}$$

[15] Note that since $\pi_{\phi|M,Y}$ is supported only on the set of ϕ's yielding a nonempty identified set, $\mathrm{IS}_\alpha(\phi|M)$ computed subsequently is nonempty.

Approximate the lower probability of the post-averaging posterior class at $H \subset \mathbb{R}$ by

$$\inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) \approx \frac{1}{G} \sum_{g=1}^{G} 1\{\mathrm{IS}_{\alpha,g}^{\mathrm{mix}} \subset H\}. \tag{3.12}$$

The draws of $\mathrm{IS}_{\alpha}^{\mathrm{mix}}$ obtained in Steps 1–3 in Algorithm 3.1 are also useful for constructing robust credible regions, which are intervals that attain a certain level of credibility uniformly over the posterior class. Applying Proposition 1 of Giacomini and Kitagawa (2021) to the Monte Carlo draws of $\mathrm{IS}_{\alpha}^{\mathrm{mix}}$, we can easily approximate the shortest robust credible region for α.

3.5 *Asymptotic properties*

This section analyzes the asymptotic properties of our method. The method is finite-sample exact (up to Monte Carlo approximation errors), but the asymptotic analysis can be valuable to understand what aspects of the prior input, if any, remain influential in large samples. In this section, we make the sample size explicit by denoting a size n sample by Y^n.

We assume that at least one model is correctly specified, so that the data-generating process is given by $p(Y^n|\phi_{\mathrm{true}})$, where $\phi_{\mathrm{true}} \in \Phi$ is the true reduced-form parameter value. We denote the unconstrained maximum likelihood estimator for ϕ by $\hat{\phi} \equiv \arg\max_{\phi \in \Phi} p(Y^n|\phi)$ and the true probability law of the sampling sequence $\{Y^n : n = 1, 2, \dots\}$ by $P_{Y^\infty|\phi_{\mathrm{true}}}$.

We impose the following regularity assumptions.

Assumption 3.2. (i) *$\mathcal{M}$ admits an identical reduced-form* (*Definition* 3.1) *and every $M \in \mathcal{M}$ satisfies either one of the following conditions*:

(A) *Φ_M contains ϕ_{true} in its interior.*

(B) *Φ_M^c contains ϕ_{true} in its interior.*

$\mathcal{M}_A$, denoting the set of models satisfying condition (A), *is nonempty.*

(ii) *Let $l_n(\phi) \equiv n^{-1} \log p(Y^n|\phi)$. There exist an open neighborhood B of ϕ_{true} and $n_0 \geq 1$, such that for any $\{Y^n : n = n_0, n_0 + 1, \dots\}$, $l_n(\cdot)$ is third-time differentiable with the third-order derivatives bounded uniformly on B.*

(iii) *Let $H_n(\hat{\phi}) \equiv -\frac{\partial^2 l_n(\hat{\phi})}{\partial \phi' \partial \phi}$. $H_n(\hat{\phi})$ is a positive definite matrix and*

$$\liminf_{n \to \infty} \det\big(H_n(\hat{\phi})\big) > 0,$$

with $P_{Y^\infty|\phi_{\mathrm{true}}}$-probability one.

(iv) *For any open neighborhood B of ϕ_{true},*

$$\limsup_{n\to\infty} \sup_{\phi\in\Phi\setminus B} \{l_n(\phi) - l_n(\phi_{\text{true}})\} < 0$$

holds with $P_{Y^\infty|\phi_{\text{true}}}$-probability one.

(v) *For every $M \in \mathcal{M}$, $\pi_{\phi|M}$ has probability density $f_{\phi|M}(\phi) \equiv \frac{d\pi_{\phi|M}}{d\phi}(\phi)$ with respect to the Lebesgue measure on Φ_M and $f_{\phi|M}(\phi)$ is continuously differentiable with a uniformly bounded derivative. For every $M \in \mathcal{M}_A$, $f_{\phi|M}(\phi_{\text{true}}) > 0$.*

Assumption 3.2(i) implies that none of the models has ϕ_{true} on the boundary of its reduced-form parameter space. Assumptions 3.2(iii) and (iv) impose regularity conditions that imply almost sure consistency of $\hat{\phi}$. Assumptions 3.2(ii) and (v) allow an application of the Laplace method to approximate the large sample marginal likelihood. Assumptions similar to Assumptions 3.2(ii)–(v) appear in Kass, Tierney, and Kadane (1990) in their validation of the higher-order expansion of the marginal likelihood.

The next proposition derives the limits of the posterior model probabilities.

PROPOSITION 3.3. (i) *Suppose Assumption 3.2 holds. Then*

$$\pi_{M|Y^\infty} \equiv \lim_{n\to\infty} \pi_{M|Y^n} = \begin{cases} \dfrac{f_{\phi|M}(\phi_{\text{true}})\cdot\pi_M}{\sum_{M'\in\mathcal{M}_A} f_{\phi|M'}(\phi_{\text{true}})\cdot\pi_{M'}}, & \text{for } M \in \mathcal{M}_A, \\ 0, & \text{for } M \notin \mathcal{M}_A. \end{cases} \tag{3.13}$$

with $P_{Y^\infty|\phi_{\text{true}}}$-probability one.

(ii) *Suppose that Assumption 3.2 holds and a prior for ϕ given M is constructed according to (3.2) with a proper prior $\tilde{\pi}_\phi$. If $\tilde{\pi}_\phi(\Phi_M) > 0$ for all $M \in \mathcal{M}$,*

$$\pi_{M|Y^\infty} = \begin{cases} \dfrac{\tilde{\pi}_\phi(\Phi_M)^{-1}\cdot\pi_M}{\sum_{M'\in\mathcal{M}_A} \tilde{\pi}_\phi(\Phi_{M'})^{-1}\cdot\pi_{M'}}, & \text{for } M \in \mathcal{M}_A, \\ 0, & \text{for } M \notin \mathcal{M}_A. \end{cases} \tag{3.14}$$

with $P_{Y^\infty|\phi_{\text{true}}}$-probability one.

(iii) *Under the assumptions of Lemma 3.1(iii), $\pi_{M|Y^\infty} = \pi_M$ holds for every $M \in \mathcal{M}$ for any sampling sequence $\{Y^n : n = 1, 2, \dots\}$.*

The proposition clarifies the large sample behavior of the posterior model probabilities when the models admit an identical reduced-form. First, it shows that our procedure asymptotically screens out models whose identifying assumptions are misspecified, $M \notin \mathcal{M}_A$, as their posterior probabilities converge to zero. If there is only one model consistent with the data-generating process, asymptotically it has probability one. Second, if $\mathcal{M}_A$ contains multiple models, their asymptotic probabilities depend

on the prior model probabilities and on the ϕ-priors evaluated at ϕ_{true}. This implies that the post-averaging posterior is asymptotically sensitive to the choices of ϕ-priors and prior model probabilities when multiple models are observationally equivalent at ϕ_{true}. Third, when the ϕ-priors are common, the asymptotic model probabilities are proportional to the reciprocal of the prior probability that the data is consistent with the identifying assumptions. Hence, the asymptotic posterior model probabilities are higher for more observationally restrictive models, that is, if $\Phi_{M_1} \subset \Phi_{M_2}$ for $M_1, M_2 \in \mathcal{M}_A$, we have $\pi_{M_1|Y^\infty} \geq \pi_{M_2|Y^\infty}$. This result is in line with the principle of parsimony (Ockham's razor)— we should prefer a more parsimonious model among those that explain the data equally well.[16]

3.6 *Discussion*

We discuss how our method relates to the literature on ϵ-contaminated class of priors and to a hierarchical Bayesian way to bridge the gap between structural and reduced-form models.

Our method can be directly linked to performing robust Bayes' analysis using an *ϵ-contaminated class of priors* (Huber (1973), Berger and Berliner (1986)). Consider the case of one single-posterior model and one multiple-posterior model, $\mathcal{M} = \{M^p, M^s\}$ that share the same parameterization of the structural model and where the likelihood for the common structural parameters θ does not depend on the model.

Given (π_{M^p}, π_{M^s}), $\pi_{\theta|M^p}$, and $\Pi_{\theta|M^s}$ as in (3.6), consider the set of priors for θ constructed by marginalizing $\Pi_{\theta,M}$ of Proposition 3.1 to θ,

$$\Pi_\theta \equiv \{\pi_\theta = \pi_{\theta|M^p}\pi_{M^p} + \pi_{\theta|M^s}\pi_{M^s} : \pi_{\theta|M^s} \in \Pi_{\theta|M^s}\}. \tag{3.15}$$

A general formulation of an ϵ-contaminated class of priors is given by

$$\Pi_\theta^\epsilon \equiv \left\{\pi_\theta = (1-\epsilon)\pi_\theta^0 + \epsilon q_\theta : q_\theta \in \mathcal{Q}\right\}, \tag{3.16}$$

where $0 \leq \epsilon \leq 1$ is a prespecified constant, π_θ^0 is a *benchmark* prior for θ, and $\mathcal{Q}$ is a set of priors of θ. Following Berger and Berliner (1986), ϵ is interpreted as the amount of contamination, q_θ captures how π_θ^0 differs from the most credible prior and $\mathcal{Q}$ is the set of possible departures. The prior input of our procedure in (3.15) has the same form as the ϵ-contaminated class of priors (3.16)—Π_θ is an ϵ-contaminated class of priors where the benchmark prior is from the single-prior (point-identified) model $\pi_\theta^0 = \pi_{\theta|M^p}$, the amount of contamination is the prior model probability assigned to the set-identified model $\epsilon = \pi_{M^s}$ and $\mathcal{Q}$ corresponds to the multiple priors for the set-identified model $\Pi_{\theta|M^s}$. This clarifies a robust Bayes interpretation of our method: If the point-identified model is a possibly misspecified benchmark, averaging it with the set-identified model with weight π_{M^s} can be interpreted as performing sensitivity analysis by contaminating

[16]For instance, in a SVAR, a model point-identified by equality restrictions is not observationally restrictive, while a model set-identified by sign restrictions can be observationally restrictive. If the ϕ-priors satisfy (3.2) and the models are observationally equivalent at ϕ_{true}, then, relative to the prior model weights, the sign-restricted model receives a larger weight than the point-identified model in large samples.

the prior of the point-identified model by an amount π_{M^s} in every possible direction subject to the set-identifying assumptions.

Our method could be viewed as a way to bridge the gap between structural and reduced-form models, for example, as an alternative to the hierarchical Bayesian approach of, for example, (Del Negro and Schorfheide (2004)), in which the structural parameters in a DSGE model act as hyperparameters of a prior for SVAR parameters. The two approaches differ in several ways. First, the hierarchical Bayesian approach always leads to a single posterior for the parameter, even if it is not identified in the SVAR model. If the parameter is not identified, this means that the priors have some part that is unrevisable by the data, leading to posterior sensitivity. In contrast, our procedure would classify the DSGE model as a single-prior model and the set-identified SVAR as a multiple-prior model, thus removing sensitivity to the choice of prior. Second, in the hierarchical Bayesian approach the prior confidence assigned to the structural model is the tightness of the prior predicted by the DSGE model, while in our procedure it is the model probability. It is however important to distinguish the notions of confidence in the two approaches, since the former is in terms of Bayesian probabilistic uncertainty while the latter is in terms of Knightian uncertainty.

4. Empirical application

We illustrate our method in the context of a conventional monetary SVAR for the federal funds rate i_t, real output growth Δgdp_t and inflation π_t, as in Aruoba and Schorfheide (2011). The model has three lags (as selected by the HQ information criterion). Following Definition 3 in Giacomini and Kitagawa (2021), we order the variables so that we can verify the conditions guaranteeing convexity of the identified set using their Proposition B.1:

$$A_0 y_t = c + \sum_{j=1}^{3} A_j y_{t-j} + \epsilon_t, \quad \text{for } t = 1, \ldots, T, \tag{4.1}$$

where $y_t = (i_t, \Delta gdp_t, \pi_t)'$ and

$$A_0 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{4.2}$$

Assume $\epsilon_t = (\epsilon_t^{\text{mp}}, \epsilon_t^d, \epsilon_t^s)'$ are i.i.d. normally distributed with mean zero and variance-covariance the identity matrix I_3. The first equation in (4.1) is interpreted as a monetary policy function, while the second and third represent aggregate demand (AD) and aggregate supply (AS), respectively. Thus, ϵ_t^{mp}, ϵ_t^d, and ϵ_t^s are monetary policy-, aggregate demand-, and aggregate supply shocks, respectively. The data are quarterly observations from 1965:1 to 2005:1 from the FRED2 database.

The reduced-form VAR is

$$y_t = b + \sum_{j=1}^{3} B_j y_{t-j} + u_t, \tag{4.3}$$

where $b = A_0^{-1}c$, $B_j = A_0^{-1}A_j$, $u_t = A_0^{-1}\epsilon_t$, $\text{var}(u_t) = E(u_t u_t') = \Sigma = A_0^{-1}(A_0^{-1})'$. The reduced form parameter is $\phi = (b, B_1, \ldots, B_4, \Sigma)$.

The prior belongs to the normal inverse-Wishart family:

$$\Sigma \sim \mathcal{IW}(\Psi, d), \qquad \beta|\Sigma \sim \mathcal{N}(\bar{b}, \Sigma \otimes \Omega),$$

where $\beta \equiv \text{vec}([b, B_1, \ldots, B_4]')$. $\Psi = I_3$ is the location matrix of Σ, $d = 4$ is a scalar degrees of freedom hyperparameter and $\Omega = 100 I_{10}$ is the variance-covariance matrix of β. The prior mean $\bar{b}$ is consistent with a random walk representation for the observables. In what follows, we perform Algorithm 3.1 with 1000 draws of ϕ's from the normal inverse-Wishart posterior. Following Christiano, Eichenbaum, and Evans (1999), we always impose the sign normalization restrictions so that the diagonal elements of A_0 are nonnegative.

4.1 *Averaging indistinguishable models*

Suppose we are interested in the cumulative output growth response[17] to a unit (contractionary) monetary policy shock ϵ_t^{mp} at horizon h, $IR^h_{\Delta gdp,\text{mp}}$, and consider the following two sets of identifying assumptions.

- *Model 1 (M1, point-identified)*
  Assume that output growth and inflation do not react on impact to the monetary policy shock, so that the (2,1) and (3,1) elements of the matrix of contemporaneous impulse responses $IR^0 = A_0^{-1}$ are zero. This identification scheme point identifies $IR^h_{\Delta gdp,\text{mp}}$.
- *Model 2 (M2, set-identified through zero restrictions)*
  The identification scheme in Model 1 is controversial.[18] Thus, in Model 2 we leave inflation unrestricted and the zero restriction is only imposed on the (2,1) element of A_0^{-1}. By Proposition B.1 in Giacomini and Kitagawa (2021), Model 2 delivers a convex identified set for $IR^h_{\Delta gdp,\text{mp}}$.

Panels (a), (b), and (e) of Figure 1 focus on the output response at horizon $h = 3$ implied by Model 1, Model 2, and their average for uniform prior model probabilities. In panel (a), the vertical solid lines for Model 1 are the 90% credible region for the point-identified output response based on a single posterior; in panel (b), the vertical dashed lines for Model 2 are the posterior mean bounds (consistent estimator of the identified set) for the output response and the solid line represents credible regions piled up from the 95% (bottom) to 5% (top) with increasing credibility by 5%. Panel (e) reports the model averaging results. The vertical dashed lines for the averaged model can be viewed as shrinking the identified set estimator from Model 2 toward the point estimator from Model 1. Figure 2 reports the results for multiple horizons.

[17] From now on, any impulse response is cumulative.

[18] See Kilian (2013) for a discussion.

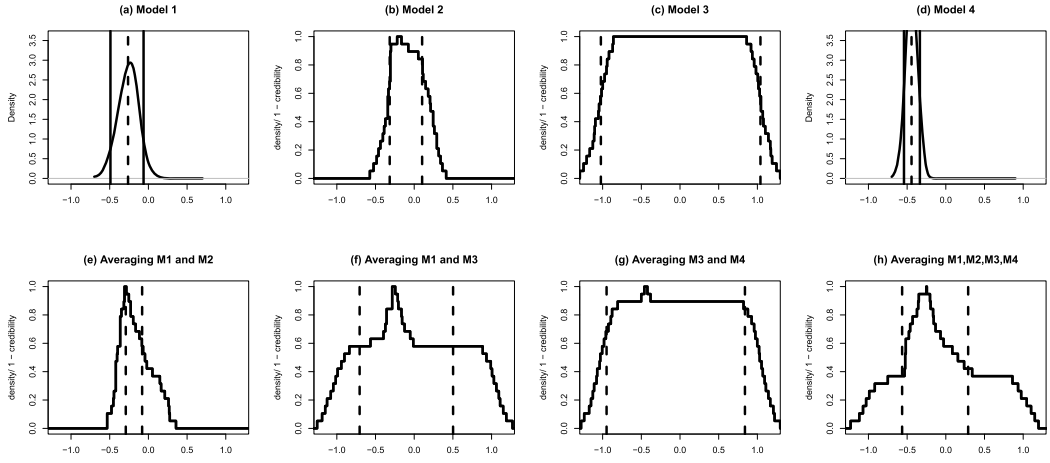

FIGURE 1. Density and robust credible region of output impulse responses. *Note*: Figure 1 reports output impulse response at horizon $h = 3$. For set-identified models (panel (b), (c) (e), (f), (g), (h)), step lines represent the Robust Credible Region (RCR) at different credibility levels. The vertical dashed lines represent the posterior mean bounds. For point-identified models (panel (a) and (d)), the vertical solid lines display the standard credible region. In such a case, we report its posterior density.

Note that, as is common for point-identified small-scale SVARs, Model 1 shows a negative response of output in the short run, whereas the set-identified Model 2 is consistent with both positive and negative effects. This is confirmed by the last row in Ta-

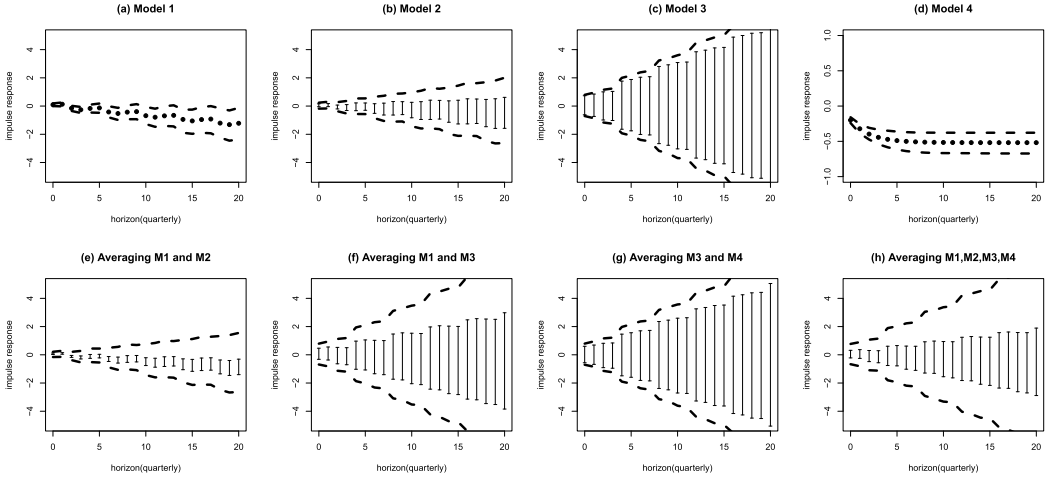

FIGURE 2. Plots of output impulse responses. *Note*: For set-identified models (panel (b), (c) (e), (f), (g), (h)), the vertical bars show the posterior mean bounds and the dashed curves connect the upper/lower bounds of posterior robust credible regions with credibility 90%. For point-identified models (panel (a) and (d)), the points plot the (unique) posterior mean and the dashed curve represent the highest posterior density regions with credibility 90%.

ble 2, reporting the lower and upper probability that the post-averaging interval of posterior means of the output response lies in the negative real half-line. Averaging the models still does not rule out a positive output response, as the 90% robust credibility region always contains positive values. Note that, since the models are indistinguishable, the prior model probabilities are not updated by the data.

4.2 *Averaging distinguishable models*

We now consider a case where the prior model probabilities are updated, by adding two popular models: a sign-restricted SVAR and a structural DSGE model:

- *Model 3 (M3, set-identified through sign restrictions)*
  We consider the following sign restrictions: the inflation response to a contractionary monetary policy shock is nonpositive and the interest rate response is nonnegative at $h = 0, 1$. As in Uhlig (2005), the output response is unrestricted. By Lemma 5.2 in Giacomini and Kitagawa (2021), the identified set in Model 3 is convex.

Consider averaging Model 1 and Model 3 with equal prior probabilities. In contrast to the previous example, the prior probabilities can now be updated using equation (3.5) because the models are distinguishable due to the observationally restrictive sign restrictions. The Appendix in the Online Supplementary Material provides two algorithms for approximating the posterior-prior plausibility ratio for the sign-restricted SVARs. We report results based on Algorithm A.1 (Algorithm A.2 produces almost identical results).

Panel (f) of Figures 1 and 2 reports the results of averaging the two models: as in the case of Model 2, Model 3 does not rule out a positive output response (this is also the conclusion of Uhlig (2005), however based on a single-prior approach). Table 1 shows that the posterior model probabilities favor Model 3 (with posterior probability 0.55), and the average of the two models does not exclude a positive output response.

- *Model 4 (M4, DSGE)*
  We consider the Bayesian DSGE model in An and Schorfheide (2007), which is a simplified version of Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005). In order to estimate the model, we rely on the prior specification in An and Schorfheide (2007), Table 2 and use output, inflation, and interest rate as observables. We use the Laplace approximation to compute the marginal likelihood.

Panel (g) of Figures 1 and 2 shows the results of averaging Models 3 and 4; note the different scale for Model 4. These models do not admit an identical reduced form, so the (equal) prior probabilities are updated according to equation (3.3). We see that Model 4 implies a negative output response; however, its posterior model probability is only 0.13, and the averaged model is consistent with both a positive and negative output response.

Finally, Panel (h) of Figures 1 and 2 reports the results of averaging all models (with equal prior weights). The posterior model probabilities (Table 1, last column) show evidence supporting the sign-restricted SVAR, while the support for the DSGE model is again weak. As in all previous cases, the averaged model does not rule out a positive output response.

TABLE 1. Output responses: prior and posterior weights.

	Averaging M1, M2	Averaging M1, M3	Averaging M3, M4	Averaging M1, M2, M3, M4
Prior w_1	0.50	0.50	/	0.25
Prior w_2	0.50	/	/	0.25
Prior w_3	/	0.50	0.50	0.25
Prior w_4	/	/	0.50	0.25
O_1	1	1	/	1
O_2	1	/	/	1
O_3	/	1.21	1.21	1.21
O_4	/	/	1	1
$\ln \tilde{p}(Y)$	−779.61	−779.61	−779.61	−779.61
$\ln p(Y\|M^1)$	−779.61	−779.61	−779.61	−779.61
$\ln p(Y\|M^4)$	/	/	−781.29	−781.29
Posterior w_1^*	0.50	0.45	/	0.29
Posterior w_2^*	0.50	/	/	0.29
Posterior w_3^*	/	0.55	0.87	0.36
Posterior w_4^*	/	/	0.13	0.06

Note: Prior w_i, O_i, and posterior w_i^* denote prior model probability, posterior-prior credibility ratio, and posterior model probability for candidate Model i, respectively; $\ln \tilde{p}(Y)$, $\ln p(Y|M^1)$, and $\ln p(Y|M^4)$ represent log marginal likelihood for the common reduced form, for Model 1 and for Model 4, respectively.

4.3 *Reverse-engineering prior model probabilities*

We now conduct the reverse engineering exercise discussed in Section 2, which computes the prior weight one would need to assign to a set of restrictions in order for the posterior mean bounds for the output response to be contained in the negative real halfline.

First, consider Model 1 and Model 2. Letting w be the prior probability of Model 1, the post-averaging interval of posterior means is

$$\Big[\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha), \sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha)\Big]$$
$$= \pi_{M^1|Y} E_{\alpha|M^1,Y}(\alpha) + \pi_{M^2|Y}\big[E_{\phi|Y,M^2}\big(l\big(\phi_{M^2}|M^2\big)\big), E_{\phi|Y,M^2}\big(u\big(\phi_{M^2}|M^2\big)\big)\big]$$

and the posterior model probabilities are equal to the prior probabilities (since the models are indistinguishable), that is, $\pi_{M^1|Y} = w$ and $\pi_{M^2|Y} = 1 - w$.

We compute the prior model probability w such that the post-averaging interval of posterior means is contained in the negative real half-line for $h = 3$. We find that one would need $w > 0.28$ to support the conclusion.

We next consider Model 1 and Model 3 (set-identification through sign restrictions). The only difference is that now the posterior model probabilities are updated and are equal to

$$\pi_{M^1|Y} = \frac{O_1 \cdot w}{O_1 \cdot w + O_3 \cdot (1-w)} \quad \text{and} \quad \pi_{M^3|Y} = \frac{O_3 \cdot (1-w)}{O_1 \cdot w + O_3 \cdot (1-w)}.$$

We find that one would need to attach very high prior probability ($w > 0.83$) to the point-identifying restrictions in Model 1 to obtain a negative output response.

TABLE 2. Output responses: estimation and inference.

	M1			M2		
	$h=2$	$h=10$	$h=20$	$h=2$	$h=10$	$h=20$
Post. Mean	−0.16	−0.68	−1.22	/	/	/
90% CR	[−0.35, 0.02]	[−1.28, −0.06]	[−2.37, −0.13]	/	/	/
Post. Mean Bounds	/	/	/	[−0.20, 0.07]	[−0.83, 0.27]	[−1.58, 0.62]
90% robust CR	/	/	/	[−0.37, 0.28]	[−1.43, 0.97]	[−2.62, 2.00]
Set of $\Pi_{IR^h\|Y}\{IR^h<0\}$	0.92	0.97	0.97	[0.25, 0.99]	[0.28, 0.99]	[0.24, 0.99]

	M3			M4		
	$h=2$	$h=10$	$h=20$	$h=2$	$h=10$	$h=20$
Post. Mean	/	/	/	−0.40	−0.52	−0.52
90% CR	/	/	/	[−0.47, −0.31]	[−0.67, −0.38]	[−0.67, −0.38]
Post. Mean Bounds	[−0.99, 1.01]	[−3.03, 3.08]	[−5.73, 5.93]	/	/	/
90% robust CR	[−1.16, 1.16]	[−3.69, 3.60]	[−6.93, 6.94]	/	/	/
Set of $\Pi_{IR^h\|Y}\{IR^h<0\}$	[0, 1]	[0, 1]	[0, 1]	1	1	1

	Averaging M1, M2			Averaging M1, M3		
	$h=2$	$h=10$	$h=20$	$h=2$	$h=10$	$h=20$
Post. Mean	/	/	/	/	/	/
90% CR	/	/	/	/	/	/
Post. Mean Bounds	[−0.18, −0.04]	[−0.76, −0.21]	[−1.41, −0.31]	[−0.64, 0.52]	[−2.05, 1.53]	[−3.85, 2.98]
90% robust CR	[−0.38, 0.21]	[−1.44, 0.76]	[−2.63, 1.55]	[−1.13, 1.12]	[−3.52, 3.48]	[−6.60, 6.72]
Set of $\Pi_{IR^h\|Y}\{IR^h<0\}$	[0.59, 0.95]	[0.62, 0.99]	[0.61, 0.98]	[0.38, 0.96]	[0.40, 0.98]	[0.40, 0.98]

	Averaging M3, M4			Averaging M1, M2, M3, M4		
	$h=2$	$h=10$	$h=20$	$h=2$	$h=10$	$h=20$
Post. Mean	/	/	/	/	/	/
90% CR	/	/	/	/	/	/
Post. Mean Bounds	[−0.91, 0.82]	[−2.71, 2.60]	[−5.05, 5.05]	[−0.48, 0.30]	[−1.55, 0.94]	[−2.89, 1.89]
90% robust CR	[−1.14, 1.16]	[−3.61, 3.57]	[−6.88, 6.80]	[−1.09, 1.07]	[−3.32, 3.37]	[−6.27, 6.51]
Set of $\Pi_{IR^h\|Y}\{IR^h<0\}$	[0.13, 1]	[0.13, 1]	[0.13, 1]	[0.40, 0.96]	[0.43, 0.99]	[0.42, 0.99]

Similar reverse engineering exercises could usefully shed light on the role of identifying assumptions in generating so-called price and liquidity puzzles in monetary SVARs.[19]

5. CONCLUSION

We proposed a method to average point-identified models and set-identified models from the multiple prior (ambiguous belief) viewpoint. The method combines single priors in point-identified models with multiple priors in set-identified models, and delivers

[19]The price puzzle occurs when contractionary monetary policy shocks produce a positive response of the price level (Sims (1992)). The liquidity puzzle refers to positive shocks in monetary aggregates leading to an initial rising rather than falling of interest rates (Reichenstein (1987)).

a set of posteriors. The post-averaging set of posteriors can be summarized by the set of posterior means and robust credible regions, which are easy to compute numerically. Our averaging method can effectively reduce the amount of ambiguity (the size of the posterior class) relative to the analysis based on a set-identified model only, and hence offers a simple and flexible way to introduce additional identifying information into the set-identified model. In the opposite direction, when the set-identified model nests the point-identified model, our method offers a simple and flexible way to conduct sensitivity analysis for the point-identified model.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools." *Journal of Political Economy*, 113, 151–184. [96]

Amir-Ahmadi, P. and H. Uhlig (2015), "Sign restrictions in Bayesian FAVARs with an application to monetary policy shocks." National Bureau of Economic Research. [97]

An, S. and F. Schorfheide (2007), "Bayesian analysis of DSGE models." *Econometric reviews*, 26, 113–172. [117]

Aruoba, B. and F. Schorfheide (2011), "Sticky prices versus monetary frictions: An estimation of policy trade-offs." *American Economic Journal: Macroeconomics*, 3, 60–90. [114]

Ashenfelter, O. (1978), "Estimating the effect of training programs on earnings." *Review of Economics and Statistics*, 60, 47–57. [96]

Bajari, P., H. Hong, and S. P. Ryan (2010), "Identification and estimation of a discrete game of complete information." *Econometrica*, 78, 1529–1568. [96]

Bates, J. and C. Granger (1969), "The combination of forecasts." *Operational Research Quarterly*, 20, 451–468. [97]

Baumeister, C. and J. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83, 1963–1999. [97, 103]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79, 1785–1821. [96]

Berger, J. and L. Berliner (1986), "Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors." *The Annals of Statistics*, 14, 461–486. [98, 102, 113]

Bernanke, B. (1986), "Alternative explorations of the money-income correlation." *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99. [100]

Blanchard, O. and D. Quah (1993), "The dynamic effects of aggregate demand and supply disturbances." *American Economic Review*, 83, 655–673. [96]

Canova, F. and G. D. Nicolo (2002), "Monetary disturbances matter for business fluctuations in the G-7." *Journal of Monetary Economics*, 49, 1121–1159. [96]

Chib, S. and I. Jeliazkov (2001), "Marginal likelihoods from the Metropolis Hastings output." *Journal of the American Statistical Association*, 96, 270–281. [109]

Christiano, L., M. Eichenbaum, and C. Evans (2005), "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113, 1–45. [117]

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999), "Monetary policy shock: What have we learned and to what end?" In *Handbook of Macroeconomics*, Vol. 1, Part A (J. B. Taylor and M. Woodford, eds.), 65–148, Elsevier. [115]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828. [96]

Claeskens, G. and N. L. Hjort (2008), *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, UK. [98]

Del Negro, M. and F. Schorfheide (2004), "Priors from general equilibrium models for vars." *International Economic Review*, 45, 643–673. [114]

Drèze, J. and J.-F. Richard (1983), "Bayesian analysis of simultaneous equation models." In *Handbook of Econometrics*, Vol. 1, Chapter 9 (Z. Griliches and M. D. Intriligator, eds.), 517–598, North-Holland. [97]

Faust, J. (1998), "The robustness of identified VAR conclusions about money." *Carnegie-Rochester Conference Series on Public Policy*, 48, 207–244. [96]

Furlanetto, F., F. Ravazzolo, and S. Sarferaz (2019), "Identification of financial factors in economic fluctuations." *The Economic Journal*, 129, 311–337. [97]

Geweke, J. (1999), "Using simulation methods for Bayesian econometric models: Inference, development, and communication." *Econometric Reviews*, 18, 1–126. [109]

Giacomini, R., T. Kitagawa, and H. Uhlig (2019), "Estimation under ambiguity." Cemmap Working Paper, University College London. [98]

Giacomini, R. and T. Kitagawa (2021), "Robust Bayesian inference for set-identified models." *Econometrica*, 89, 1519–1556. [96, 97, 98, 100, 101, 103, 110, 111, 114, 115, 117]

Giacomini, R., T. Kitagawa, and A. Volpicella (2022), "Supplement to 'Uncertain identification'." *Quantitative Economics Supplemental Material*, 13, https://doi.org/10.3982/QE1671. [99]

Hansen, B. E. (2007), "Least squares model averaging." *Econometrica*, 75, 1175–1189. [97]

Hansen, B. E. (2014), "Model averaging, asymptotic risk, and regressor groups." *Quantitative Economics*, 5, 495–530. [97]

Hjort, N. L. and G. Claeskens (2003), "Frequentist model average estimators." *Journal of the American Statistical Association*, 98, 879–899. [97]

Ho, P. (2019), "Global robust Bayesian analysis in large models." Working Paper. [98]

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), "Bayesian model averaging: A tutorial." *Statistical Science*, 14, 382–417. [97]

Horowitz, J. L. and C. F. Manski (1995), "Identification and robustness with contaminated and corrupted data." *Econometrica*, 63, 281–302. [98]

Huber, P. (1973), "The use of Choquet capacities in statistics." *Bulletin of the International Statistical Institute*, 45, 181–191. [98, 113]

Imbens, G. and J. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62, 467–475. [96]

Kass, R. E., L. Tierney, and J. B. Kadane (1990), "The validity of posterior expansions based on Laplace method." In *Bayesian and Likelihood Methods in Statistics and Econometrics* (S. Geisser, J. Hodges, S. Press, and A. Zellner, eds.), 473–488, Elsevier, North-Holland. [112]

Kilian, L. (2013), "Structural vector autoregressions." In *Handbook of Research Methods and Applications in Empirical Macroeconomics* (M. Hashimzade and M. A.Thornton, eds.), 515–554, Edward Elgar, Cheltenham, UK. [115]

Kitagawa, T. and C. Muris (2016), "Model averaging in semiparametric estimation of treatment effects." *Journal of Econometrics*, 193, 271–289. [97]

Kline, B. and E. Tamer (2016), "Bayesian inference in a class of partially identified models." *Quantitative Economics*, 7, 329–366. [98]

Korobilis, D. (2020), "Sign restrictions in high-dimensional vector autoregressions." Available at SSRN. [97]

Leamer, E. E. (1978), *Specification Searches*. Wiley, New York. [97]

Liu, Q. and R. Okui (2013), "Heteroskedasticity-robust C_p model averaging." *Econometrics Journal*, 16, 463–472. [97]

Magnus, J. R., O. Powell, and P. Prüfer (2010), "A comparison of two model averaging techniques with an application to growth empirics." *Journal of Econometrics*, 154, 139–153. [97]

Manski, C. F. (1989), "Anatomy of the selection problems." *Journal of Human Resources*, 24, 343–360. [96]

Manski, C. F. and J. V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68, 997–1010. [96]

Masten, M. A. and A. Poirier (2020), "Inference on breakdown frontiers." *Quantitative Economics*, 11, 41–111. [98]

Matthes, C. and F. Schwartzman (2019), "What do sectoral dynamics tell us about the origins of business cycles?" FRB Working Paper, 19. [97]

Moon, H. and F. Schorfheide (2012), "Bayesian and frequentist inference in partially identified models." *Econometrica*, 80, 755–782. [97, 98]

Poirier, D. (1998), "Revising beliefs in nonidentified models." *Econometric Theory*, 14, 483–509. [97]

Reichenstein, W. (1987), "The impact of money on short-term interest rates." *Economic Inquiry*, 25, 67–82. [119]

Rosenbaum, P. and D. B. Rubin (1983), "The central role of the propensity score in observational studies." *Biometrika*, 70, 41–55. [96]

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Hoboken, New Jersey. [96]

Sims, C. (1980), "Macroeconomics and reality." *Econometrica*, 48, 1–48. [96, 100]

Sims, C. (1992), "Interpreting the macroeconomic time series facts: The effects of monetary policy." *European Economic Review*, 36, 975–1000. [119]

Sims, C., D. Waggoner, and T. Zha (2008), "Methods for inference in large multiple-equation Markov-switching models." *Journal of Econometrics*, 146, 255–274. [109]

Smets, F. and R. Wouters (2003), "An estimated dynamic stochastic general equilibrium model of the euro area." *Journal of the European economic association*, 1, 1123–1175. [117]

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. [96, 117]

Volpicella, A. (2020), *Essays in Structural Vector Autoregressions (SVARs)*. Ph.D. thesis, Queen Mary University of London. [95]

Co-editor Andres Santos handled this manuscript.

Manuscript received 28 June, 2020; final version accepted 25 June, 2021; available online 3 August, 2021.