cemmap

centre for microdata methods and practice

Uncertain identification

Raffaella Giacomini
Toru Kitagawa
Alessio Volpicella

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper
CWP33/20

Economic
and Social
Research Council

Uncertain Identification*

Raffaella Giacomini[†], Toru Kitagawa[‡], and Alessio Volpicella[§]

This draft: June 2020

Abstract

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, in general, leads 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 non-dogmatic assumptions. The output is a set of posteriors (*post-averaging ambiguous belief*) that are mixtures of the single posterior and any element of the class of multiple posteriors, with weights equal to the posterior model probabilities. We suggest reporting the set of posterior means and the associated credible region in practice, and provide a simple algorithm to compute them. We establish that the prior model probabilities are updated when the models are "distinguishable" and/or they specify different priors for reduced-form parameters, 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. In a standard monetary model, for example, we show that, in order to support a negative response of output to a contractionary monetary policy shock, one would need to attach a prior probability greater than 0.05 to the validity of the assumption that prices do not react contemporaneously to the shock.

Keywords: Partial Identification, Sensitivity Analysis, Model Averaging, Bayesian Robustness, Ambiguity.

*We would like to thank Frank Kleibergen, Ulrich Müller, José Montiel-Olea, 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.

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

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

[§]University of Surrey, School of Economics. Email: a.volpicella@qmul.ac.uk

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 Bayesian model averaging/selection 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 under-identifying or take the form of inequality restrictions. For these models, we advocate adopting the multiple-prior approach of Giacomini and Kitagawa (2020). 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 identifying restrictions.

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 (Bernanke (1986) and Sims (1980)), long-run neutrality restrictions (Blanchard and Quah (1993)), and Bayesian prior restrictions implied by a structural model (Del Negro and Schorfheide (2004)). Subsets of these assumptions deliver set-identified impulse-responses, as does the use of 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 et al. (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 et al. (2010), Beresteanu et al. (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 proposed method provides a formal framework for investigating the sensitivity of empirical findings to specific identifying assumptions and/or for aggregating results based on different sets of identifying assumptions, which can be more practical than reporting inference for all models separately when there are many alternative sets of 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 et al. (1999) and Claeskens and Hjort (2008)), frequentist approaches (Hansen (2007, 2014), Hjort and Claeskens (2003), Hansen and Racine (2012), Liu (2015), Liu and Okui (2013), and Zhang and Liang (2011)), and hybrid approaches (Hjort and Claeskens (2003), Kitagawa and Muris (2016), and Magnus et al. (2010)), but none of them allows for set-identification/multiple priors in any candidate model.

We tackle this problem from the angle of Bayesian model averaging. Standard Bayesian model averaging delivers a single posterior that is a mixture of the posteriors of the candidate 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, even an apparently uninformative one, 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 (2020)), and then combine the corresponding multiple posteriors with single posteriors for models that are either point-identified or that impose non-dogmatic 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*),

[1] There are several examples in the empirical literature of sign-restricted models considering a large number of restrictions. For example, Korobilis (2020) considers a model with 63 sign restrictions; sign-restricted SVARs that consider many restrictions include, among others, Matthes and Schwartzman (2019) in the context of business cycles analysis, Ahmadi and Uhlig (2015), who investigate monetary policy, Furlanetto et al. (2019), who focus on financial shocks, and Antolin-Diaz et al. (2020), who study monetary policy and financial shocks.

[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, e.g., Drèze and Richard (1983). Bayesian model averaging offers a practical way to avoid the Borel paradox in such context.

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, we recommend reporting 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). We show that these quantities have analytically simple expressions and are easy to compute in practice.

This paper contributes to the growing literature on Bayesian inference for partially identified models (Giacomini and Kitagawa (2020), Kline and Tamer (2016), Moon and Schorfheide (2012), Norets and Tang (2014), Liao and Simoni (2013)). We follow the multiple-prior approach to model the lack of knowledge within the identified set as in Giacomini and Kitagawa (2020). When a set-identified model is the only model considered, the set of posteriors generated by the approach leads to the posterior inference for the identified set proposed in Kline and Tamer (2016), Liao and Simoni (2013), and Moon and Schorfheide (2011). 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 method proposed in this paper provides a formal framework for conducting sensitivity analysis of causal inferences to the choice of identifying assumptions. For example, when a set-identified model nests a point-identified model, the 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. In this case, we can formally interpret our averaging method as an example of the ϵ-contamination sensitivity analysis developed in Huber (1973) and Berger and Berliner (1986), with a particular construction of the prior class. In another example, if the point-identified model can be considered a reasonable benchmark, the method offers a simple and flexible way to add non-dogmatic identifying information to the set-identified model, which results in increasing informativeness of the conclusions in a transparent manner. Finally, the method can be used to 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 averaging to preserve a given empirical conclusion (e.g., the so-called price and liquidity puzzles in monetary SVARs, respectively discussed by (Sims, 1992) and (Reichenstein, 1987)). Obtaining this threshold in terms of the prior model probability shares the motivation with the breakdown frontier analysis proposed in Horowitz and Manski (1995) and Masten and Poirier (2020), where the breakdown frontier is the population quantity (as opposed to the belief) that measures the minimal violation of the benchmark point-identifying assumption to support the conclusion.

Our proposed 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. The output of the procedure is a weighted average of the posterior mean in the point-identified model and the set of posterior means in the set-identified model. When the identified set is a connected interval, the set of posterior means can be viewed as an estimate of the identified set (Giacomini and Kitagawa (2020)), and thus our averaging procedure effectively shrinks the identified set estimate toward the point estimate from the point-identified model, with the degree of shrinkage governed by the posterior model probabilities.

In addition to developing a novel approach to Bayesian model averaging, we make 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: causal ordering restrictions (Bernanke (1986) and Sims (1980)), sign restrictions (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)), and restrictions implied by a Dynamic Stochastic General Equilibrium (DSGE) model. The choice of identifying assumptions has often been a source of controversy in this literature, given that researchers have differing opinions about their credibility. One popular choice is the use of sign restrictions. Although the resulting model is set-identified and the approach therefore raises serious robustness concern as we discussed above, the common practice is to consider single-prior Bayesian inference in set-identified SVARs. The large body of the empirical literature adopting this approach includes Canova and Nicolo (2002), Faust (1998), Mountford (2005), Rafiq and Mallick (2008), Scholl and Uhlig (2008), Uhlig (2005), and Vargas-Silva (2008) for applications to monetary policy, Dedola and Neri (2007), Fujita (2011), and Peersman and Straub (2009) for applications to business cycle model, Mountford and Uhlig (2009) for applications to fiscal policy, Kilian and Murphy (2012) for applications to oil prices. Alternative approaches that do not suffer from the pitfalls of single-prior Bayesian inference are Moon et al. (2013) and Gafarov et al. (2018, 2016), who consider frequentist inference for the identified set and Giacomini and Kitagawa (2020), who propose a robust Bayesian approach. To our knowledge, little work has been done on multi-model 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 averaging 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 discusses the relationship between our method and existing Bayesian methods, and discusses elicitation of model probabilities. Section 5 applies our method to impulse response analysis in monetary SVARs. The Appendix contains proofs and a microeconometric application.

2 Illustrative Example

We present the key ideas and the implementation of the method in a static model of labor supply and demand, subject to common types of identifying assumptions.[3] The model is:

$$A \begin{pmatrix} \Delta n_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \text{ t=1,...,T}, \tag{2.1}$$

where $(\Delta n_t, \Delta w_t)$ are the growth rates of employment and wages 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}.

The reduced-form model is indexed by Σ, the variance-covariance matrix of $(\Delta n_t, \Delta w_t)$, which satisfies $\Sigma = A^{-1}(A^{-1})'$. Denote its lower triangular Cholesky decomposition with

[3]See Appendix A.2 for a microeconometric application to a treatment effect model with noncompliance.

nonnegative diagonal elements by $\Sigma_{tr} = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ with $\sigma_{11} \geq 0$ and $\sigma_{22} \geq 0$, and define the reduced form parameter as $\phi = (\sigma_{11}, \sigma_{21}, \sigma_{22}) \in \Phi = \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.[4] 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 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 θ, which in this case is equivalent to a prior for the reduced-form parameter ϕ. In set-identified models, one could either specify a single prior for θ (e.g., as a way of imposing non-dogmatic identifying assumptions) or consider multiple priors as in Giacomini and Kitagawa (2020). 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 $\theta|\phi$ (which are not revised). As in Giacomini and Kitagawa (2020) there are several reasons for adopting a single prior for ϕ: first, this prior is revised by the data so any sensitivity concerns in this respect are only present in finite samples; second, multiple priors for ϕ would lead to problematic non-convergence issues, as discussed by Ruggeri and Sivaganesan (2000); third, the single prior for ϕ facilitates computational implementation. An additional advantage of separating revisable and unrevisable prior knowledge 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.

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 of possible empirical interest.

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 labor demand is inelastic to wage, $a_{12} = 0$.

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

- *Model M^s (set-identified)*: The wage elasticity of demand is non-positive, $a_{12} \geq 0$, and the wage elasticity of supply is non-negative, $a_{21} \leq 0$.

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 $A^{-1} = \Sigma_{tr}$. The parameter of interest is $\alpha = \alpha_{M^p}(\phi) \equiv \sigma_{11}$.

Model M^s imposes sign restrictions that only set-identify α. Appendix A shows that the identified set for α is:

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

Note that the identified set is non-empty for any ϕ. Hence, models M^p and M^s are observationally equivalent at any $\phi \in \Phi$ and neither of them is falsifiable, i.e., for any $\phi \in \Phi$ 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 (2020), 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:

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

[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 $a_{21} \leq 0$.

where $\mathcal{A}_{sign} = \{A : a_{12} \geq 0, a_{21} \leq 0, diag(A) \geq 0\}$ 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:

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

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

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

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 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.[6]

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

[6]We discuss interpretation and elicitation of the prior model probabilities in Section 4.3.

$\gamma \in (0,1)$, 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:

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

where $(l(\phi), u(\phi))$ are the lower and upper bounds of the non-empty 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 $z_1,\dots,z_G$ randomly from a Bernoulli distribution with mean $\pi_{M^p|Y}$ and then generate $g=1,\dots,G$ random draws of the "mixture identified set" for α according to

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

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 non-empty 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 $IS_{\alpha}^{mix}(\phi)$'s. The minimization problem in Step 5 of Algorithm 4.1 in Giacomini and Kitagawa (2020) is solved to obtain the shortest-width robust credible region.

2.2 Non-dogmatic Identifying Assumptions

Our method allows for identifying assumptions that are expressed as a non-dogmatic 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 meta-analysis 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 counter-intuitive effects on posterior inference.[7] These robustness concerns can be addressed by averaging the Bayesian 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 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.

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 similarly to (2.8), where M^B replaces M^p. The robust credible region can be constructed as in Scenario 1, by drawing iid draws $z_1, \ldots, z_G \sim Bernoulli(\pi_{M^B|Y})$ and letting

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

3 Formal Analysis

This section formalizes the idea in a general setting and proves the analytical claims made in the previous section.

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

3.1 Notation and Definitions

Consider $J + K \geq 2$ candidate models, $J, K \geq 0$, that can differ in various aspects, including the identifying assumptions and the parameterization of the structural model. The class of J models consists of *single-prior models*, whose prior input always (i.e., independent of the realization of the data) leads to a single posterior for the parameter of interest. Examples are models that impose dogmatic point-identifying assumptions with a single prior for the reduced-form parameter (such as model M^p in Scenario 1), or models that assume a single prior for the structural parameter in spite of it being set-identified (such as model M^B in Scenario 2). We denote the class of single-prior models by $\mathcal{M}^p$.

The class of K models consists of *multiple-prior models*, defined by the following features: (1) under the identifying assumptions the parameter of interest is set-identified, i.e., knowledge of the distribution of observables (value of the reduced-form parameter) does not pin down a unique value for the parameter of interest, and (2) they specify a single prior for the reduced-form parameter. The posterior information in a multiple-prior model is characterized by the set of posteriors. We denote the class of multiple-prior models by $\mathcal{M}^s$.

Let $\mathcal{M} \equiv \mathcal{M}^p \cup \mathcal{M}^s$. The vector of structural parameters in model $M \in \mathcal{M}$ is $\theta_M \in \Theta_M$, where Θ_M is the set of structural parameters that satisfy 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 ϕ_M is a function of the structural parameter, $\phi_M = g_M(\theta_M) \in \mathbb{R}^{d_M}$, 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 reduced-form parameter space, defined as $\Phi_M = g_M(\Theta_M)$.[8] As reflected in the notation, our most general set-up allows the parameter space of both structural and reduced-form parameters to differ across models.[9] We express the likelihood in model $M \in \mathcal{M}$ in terms of the reduced-form parameter by $p(Y|\phi_M, M)$. For a multiple-prior model $M \in \mathcal{M}^s$, define the identified set of α by $IS_\alpha(\phi_M|M) = \{\alpha_M(\theta_M) : \theta_M \in \Theta_M \cap g_M^{-1}(\phi_M)\}$, which is a set-valued mapping from Φ_M to $\mathbb{R}$.

Note that, by construction, the parameter space of the reduced form parameter Φ_M incorporates the testable implications, if any, of the imposed identifying assumptions. For a set-identified model $M^s \in \mathcal{M}^s$, Φ_{M^s} is equivalent to the set of ϕ_M's that yield a non-empty identified set, $\Phi_{M^s} = \{\phi_{M^s} \in \mathbb{R}^{d_{M^s}} : IS_\alpha(\phi_{M^s}|M^s) \neq \emptyset\}$.[10]

[8]The likelihood $\tilde{p}(Y|\theta_M, M)$ in model M depends on θ_M only through the reduced-form parameters $g_M(\theta_M)$ for any realization of Y, i.e., 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. The statistics literature refers to the reduced-form parameter as the minimally sufficient parameter (see, e.g., Dawid (1979)).

[9]For instance, in the model considered in Section 2, the reduced-form parameter space can differ depending on how many lagged endogenous variables and/or exogenous variables are included in each model.

[10]For instance, in a SVAR with observationally restrictive sign restrictions, Φ_M is the set of reduced-form

The next definition introduces the concept of identical reduced-forms among the candidate models. Our analytical results about the posterior model probabilities shown below (Lemma 3.1 and Proposition 3.3) assume that some or all of the candidate models admit an identical reduced-form.

Definition 3.1 *Let $\mathcal{M}$ be a collection of models. $\mathcal{M}$* ***admits an identical reduced-form*** *if the following conditions hold:*

(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 $M \in \mathcal{M}$, the reduced-form likelihood $p(Y|\phi_M = \phi, M)$ defines a probability distribution of Y on the extended domain $\phi \in \Phi \equiv \cup_{M \in \mathcal{M}} \Phi_M$, 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 parametric family of distributions for the observables (Condition (a)). Different identifying assumptions, nonetheless, can constrain the class of distributions of observables in the sense that the domain of reduced-form parameters Φ_M can differ among the models. The key condition in Definition 3.1 is (b), requiring that the distribution of the data Y in model M (indexed by ϕ) is well-defined over the extended domain $\Phi = \cup_{M \in \mathcal{M}} \Phi_M$ and the likelihood of ϕ is common among the models $M \in \mathcal{M}$. For instance, if $\mathcal{M}$ consists of SVAR models with the same set of variables but subject to different identifying assumptions (including observationally restrictive ones such as sign restrictions), the conditions of Definition 3.1 are satisfied when when the reduced-form VARs implied by the models feature the same variables and lag length. See also the treatment effect models of Appendix A.2 as a microeconometrics example where all the candidate models admit an identical reduced-form. In what follows, whenever we assume that $\mathcal{M}$ admits an identical reduced-form, we denote the common reduced-form parameters by ϕ and the common reduced-form likelihood by $p(Y|\phi)$.

The next set of definitions introduces the concepts of observational equivalence and distinguishability of the candidate models.

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

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

parameters in the VAR yielding a non-empty impulse response identified set, which can be a proper subset of the reduced-form parameter space of the VAR.

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

Models that are observationally equivalent at ϕ (Definition 3.2 (i)) generate the same distribution of data (corresponding to ϕ), implying that knowledge of ϕ fails to uniquely identify what model generated the data. Note that our definition of observational equivalence is local to the given ϕ, and it does not constrain the relationship among the reduced-form parameter spaces for different models except that they must have a non-empty intersection. In contrast, the concept of (in)distinguishability in Definition 3.2 (ii) and (iii) concerns the relationship among the reduced-form parameter spaces across models. If two models admitting an identical reduced-form are distinguishable, then there exists some reduced-form parameter value that allows one to falsify one model in favor of the other. On the other hand, indistinguishability of Definition 3.1 (iii) 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.

3.2 Prior and Posterior Model Probabilities

This section shows when and how the data update the prior model probabilities when some or all of the candidate models admit an identical reduced form.

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 model probability for each model in the class is

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

By the definition of reduced-form parameters, the value of the likelihood depends on θ_M only through ϕ_M, for which we assume a single prior. This implies that the marginal likelihood depends only on the ϕ_M-prior, and thus it can be computed uniquely for all models since every $M \in \mathcal{M}$ assumes a single prior for ϕ_M (including including multiple-prior models).

In situations where the models admit an identical reduced-form, we can simplify the expression of the posterior model probabilities, as shown in the next lemma.

Lemma 3.1 *(i) Suppose that the multiple-prior models $M^s \in \mathcal{M}^s$ admit an identical reduced-form with reduced-form parameters $\phi \in \Phi = \cup_{M^s \in \mathcal{M}^s} \Phi_{M^s} \subset \mathbb{R}^d$. Let $\tilde{\pi}_\phi$ be a proper prior on Φ and assume that $\tilde{\pi}_\phi(\Phi_{M^s}) = \tilde{\pi}_\phi(IS_\alpha(\phi|M^s) \neq \emptyset) > 0$ holds for all $M^s \in \mathcal{M}^s$. Let $\tilde{\pi}_{\phi|Y}$ be the posterior of ϕ obtained by updating $\tilde{\pi}_\phi$ with the likelihood $p(Y|\phi)$, which is common among all $M^s \in \mathcal{M}^s$. Suppose that the ϕ-prior in each model is specified according to*

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

where $\mathcal{B}(\Phi)$ is the Borel σ-algebra of Φ, i.e., the ϕ-prior is constructed by trimming the support of $\tilde{\pi}_\phi$ to Φ_{M^s}. Then the posterior model probabilities are given by

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

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 $\tilde{\pi}_\phi$,

$$O_{M^s} \equiv \frac{\tilde{\pi}_{\phi|Y}(\Phi_{M^s})}{\tilde{\pi}_\phi(\Phi_{M^s})} = \frac{\tilde{\pi}_{\phi|Y}(IS_\alpha(\phi|M^s) \neq \emptyset)}{\tilde{\pi}_\phi(IS_\alpha(\phi|M^s) \neq \emptyset)}, \quad \tilde{p}(Y) = \int_\Phi p(Y|\phi) d\tilde{\pi}_\phi(\phi). \tag{3.4}$$

(ii) Suppose that, in addition to $\mathcal{M}^s$, all the single-prior models $\mathcal{M}^p$ admit an identical reduced-form. Let $\tilde{\pi}_\phi$ be as defined in (i) of the current lemma and assume $\tilde{\pi}_\phi(\Phi_M) > 0$ holds for all $M \in \mathcal{M}$. If the ϕ-prior satisfies (3.2) in every $M \in \mathcal{M}$, then the posterior model probabilities are further simplified 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 candidate models are indistinguishable and the ϕ-prior is common among them, 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 the first claim, the specification of the ϕ-prior (3.2) simplifies the marginal likelihood of the set-identified model $M^s \in \mathcal{M}^s$ to $\tilde{p}(Y)O_{M^s}$. The computation of $\tilde{p}(Y)$ and O_{M^s} requires one set of Monte Carlo draws of ϕ each from the prior $\tilde{\pi}_\phi$ and from the posterior $\tilde{\pi}_{\phi|Y}$, as well as an assessment of the validity of the identifying assumptions at the drawn ϕ's (the emptiness of the corresponding identified set). Hence, computation time can be saved by avoiding to run separate algorithms for each set-identified model. If all the candidate models admit an identical reduced-form (Lemma 3.1 (ii)), the posterior model probabilities only depend on $\{O_M : M \in \mathcal{M}\}$, so one does not even need to compute the marginal likelihoods. The claim in (iii) says that, if all the candidate models are indistinguishable and share a unique ϕ-prior, the prior model probabilities can never be updated. This result is intuitive: assuming the same prior knowledge for ϕ in the indistinguishable models (i.e. a common support of ϕ), all models have the same marginal likelihood, which therefore cancels out in (3.1).

Scenario 1 in Section 2 satisfies Lemma 3.1 (iii) and thus no update occurs for the model probabilities. Scenario 2 satisfies Lemma 3.1 (i) with $O_{M^s} = 1$, since the identified set in M^s is never empty. In the example of the treatment effect model in Appendix A.2, the point-identified and set-identified models are distinguishable since they have distinct testable implications. Hence, if the common kernel of the prior is maintained as in (3.2), Lemma 3.1 (ii) gives the formula of the posterior model probabilities.

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 $\pi_{\alpha|M^p,Y}$, the posterior for α in each single-prior model $M^p \in \mathcal{M}^p$, 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

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

where $\mathcal{B}(\Phi_{M^s})$ is the Borel σ-algebra of Φ_{M^s}.[11] In words, $\Pi_{\theta_{M^s}|M^s}$ collects priors for θ_{M^s} that satisfy the identifying assumptions with probability one (i.e., $\pi_{\theta_{M^s}|M^s}(\Theta_{M^s}) = 1$) and whose ϕ_{M^s}-marginals coincide with the specified ϕ_{M^s}-prior. 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]

$$\begin{aligned}&\Pi_{\alpha|M^s,Y}\\ &\equiv \left\{\pi_{\alpha|M^s,Y} = \int_{\Phi_M} \pi_{\alpha|M^s,\phi_{M^s}} d\pi_{\phi_{M^s}|M^s,Y} : \pi_{\alpha|M^s,\phi_{M^s}}(IS_\alpha(\phi_{M^s}|M^s)) = 1,\ \pi_{\phi_{M^s}|M^s}\text{-a.s.}\right\}.\end{aligned} \tag{3.7}$$

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

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

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). Since there is no restriction that constrains the choice of posterior across

[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

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

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 Appendix A shows a formal derivation of $\Pi_{\alpha|M^s,Y}$.

the set of posteriors, the set of averaged posteriors can be represented as

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

This is a representation of the post-averaging ambiguous belief that generalizes the two-model case shown in (2.7).

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] in which case (3.8) can be obtained by applying Bayes' rule to each prior in a certain well-defined class of priors.

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 $\Theta = \cup_{M \in \mathcal{M}} \Theta_M \subset \mathbb{R}^{d_\theta}$. 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, posterior quantiles, 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 $IS_\alpha(\phi_{M^s}|M^s)$ 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:

$$\inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = \sum_{M^p \in \mathcal{M}^p} E_{\alpha|M^p,Y}(\alpha) \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} E_{\phi_{M^s}|Y,M^s}[l(\phi_{M^s}|M^s)] \pi_{M^s|Y},$$

$$\sup_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = \sum_{M^p \in \mathcal{M}^p} E_{\alpha|M^p,Y}(\alpha) \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} E_{\phi_{M^s}|Y,M^s}[u(\phi_{M^s}|M^s)] \pi_{M^s|Y},$$

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

$$\inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) = \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y}(H) \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}(IS_\alpha(\phi_{M^s}|M^s) \subset H) \cdot \pi_{M^s|Y},$$

$$\sup_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) = \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y}(H) \pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}(IS_\alpha(\phi_{M^s}|M^s) \cap H \neq \emptyset) \cdot \pi_{M^s|Y}.$$

[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.

(iii) The lower and upper bounds of the cumulative distribution function (cdf) of $\pi_{\alpha|Y} \in \Pi_{\alpha|Y}$ *are*

$$\begin{aligned}
\underline{\pi}_{\alpha|Y}(a) &\equiv \inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}([-\infty, a]) \\
&= \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y}([-\infty, a])\pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}\left(\{u(\phi_{M^s}|M^s) \leq a\}\right)\pi_{M^s|Y}, \\
\bar{\pi}_{\alpha|Y}(a) &\equiv \sup_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}([-\infty, a]) \\
&= \sum_{M^p \in \mathcal{M}^p} \pi_{\alpha|M^p,Y}([-\infty, a])\pi_{M^p|Y} + \sum_{M^s \in \mathcal{M}^s} \pi_{\phi_{M^s}|Y,M^s}\left(\{l(\phi_{M^s}|M^s) \leq a\}\right)\pi_{M^s|Y},
\end{aligned}$$

and the set of posterior τ*-th quantiles,* $\tau \in (0,1)$*, is* $\left[\inf\{a : \bar{\pi}_{\alpha|Y}(a) \geq \tau\}, \inf\{a : \underline{\pi}_{\alpha|Y}(a) \geq \tau\}\right]$.

If a set-identified model delivers $IS_\alpha(\phi_{M^s}|M^s)$ as a connected interval at every reduced-form parameter value, then we can view $\left[E_{\phi_{M^s}|Y,M^s}[l(\phi_{M^s}|M^s)], E_{\phi_{M^s}|Y,M^s}[u(\phi_{M^s}|M^s)]\right]$ as an estimator of the identified set in model M^s. We can therefore interpret the set of post-averaging posterior means as the weighted Minkowski sum of the Bayesian point estimators (posterior means) 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 $\Pi_{\alpha|Y}$. This lower probability is 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* $\Pr(IS_\alpha^{mix} \subset A)$, where IS_α^{mix} is generated according to

$$M \sim \text{Multinomial}\left(\{\pi_{M|Y}\}_{M \in \mathcal{M}}\right), \tag{3.10}$$

$$IS_\alpha^{mix} = \begin{cases} \{\alpha\}, & \alpha|(M^p, Y) \sim \pi_{\alpha|M^p,Y} \text{ for } M^p \in \mathcal{M}^p, \\ IS_\alpha(\phi_{M^s}|M^s), & \phi_{M^s}|(M^s, Y) \sim \pi_{\phi_{M^s}|M^s,Y} \text{ for } M^s \in \mathcal{M}^s. \end{cases}$$

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

Note that our method introduces ambiguous beliefs for the non-identifiable parameters, while it assumes availability of prior model probabilities even when the models are indistinguishable. Hence, we are not treating non-identifiability of the parameters and of the models (choices of identifying assumptions) in a symmetric way. We do not have a normative argument for this asymmetric treatment, and our view is that whether one wants to introduce ambiguity for the parameters only or for both the parameters and the models should depend on the user's prior knowledge. In our observation of empirical practice, researchers typically motivate the credibility of point-identifying assumptions, whereas if the identifying assumptions only set-identify the parameter of interest, they tend to express ambiguity in the form of identified sets. We hence believe that, regardless of whether the models are distinguishable or not, assuming availability of probabilistic weights over different sets of identifying assumptions is not too demanding for potential users, and our asymmetric treatment is not too distant from the way that empirical research is currently performed.

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 $IS_\alpha(\phi_{M^s}|M^s)$ 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, e.g., see Chib and Jeliazkov (2001), Geweke (1999), and Sims et al. (2008). When Lemma 3.1 (ii) applies, i.e., when all the models admit an identical reduced-form, such as in one of the specifications in the empirical application of Section 5, 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 non-empty 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 non-emptiness of $IS_\alpha(\phi|M^s)$ is simple or not depends on the application. In the application in Section 5 to SVARs with sign restrictions, we consider two ways to check the non-emptiness of $IS_\alpha(\phi|M^s)$. The first (Algorithm A.1) builds on Algorithm 1 of Giacomini and Kitagawa (2020) and assesses non-emptiness based on the Monte Carlo draws of the impulse responses. The second approach (Algorithm A.2), 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 Appendix A.3 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 non-empty $IS_\alpha(\phi|M^s)$, 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 5, we compute $l(\phi|M^s)$ and $u(\phi|M^s)$ by numerical optimization. Alternatively, adopting the criterion function approach of Chernozhukov et al. (2007), the computation of the lower and upper bounds of the identified set can be facilitated by applying the slice sampling algorithm proposed by Kline and Tamer (2016).

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 $IS_\alpha^{mix} = \{\alpha\}$ (a singleton). If the drawn M belongs to $\mathcal{M}^s$, draw $\phi_M \sim \pi_{\phi|M,Y}$ and set $IS_\alpha^{mix} = IS_\alpha(\phi_M|M)$.[15]

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

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

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

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

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

The draws of IS_α^{mix} obtained in Steps 1-3 in Algorithm 3.1 are also useful for constructing the robust credible regions. The robust credible region with credibility $\gamma \in (0,1)$ is defined as the shortest interval to which every posterior in the class assigns probability at least γ;

$$C_\gamma \equiv \arg\min_{C \in \mathcal{C}} \text{length}(C), \;\; \text{s.t.} \inf_{\pi_{\alpha|Y} \in \Pi_{\alpha|Y}} \pi_{\alpha|Y}(C) \geq \gamma, \tag{3.13}$$

where $\mathcal{C}$ is the class of connected intervals in $\mathbb{R}$. Since the constraint in (3.13) can be interpreted equivalently as $\Pr(IS_\alpha^{mix} \subset C) \geq \gamma$, the computation of C_γ can be reduced to finding the shortest interval that contains the γ-proportion of the Monte Carlo draws of IS_α^{mix}. A simple computation algorithm for this optimization problem is shown in Proposition 1 of Giacomini and Kitagawa (2020) and it can be readily applied to the current context.

3.5 Asymptotic Properties

This section analyzes the asymptotic properties of our method. The procedure is finite-sample exact (up to Monte Carlo approximation errors) and does not rely on asymptotic approximations. The asymptotic analysis is nevertheless valuable, as it highlights what aspects of the

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

prior input, if any, remain influential even in large samples. In this section, we make the sample size explicit in our notation by denoting a size n sample by Y^n.

We assume that $\mathcal{M}$ admits an identical reduced-form (Definition 3.1) and that at least one model is correctly specified, so that the data-generating process is given by $p(Y^n|\phi_{true})$, where $\phi_{true} \in \Phi$ 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 $P_{Y^\infty|\phi_{true}}$.

For our asymptotic analysis, we impose the following regularity assumptions:

Assumption 3.2 *(i) $\mathcal{M}$ admits an identical reduced-form 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 non-empty.

(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 $\liminf_{n\to\infty} \det(H_n(\hat{\phi})) > 0$, with $P_{Y^\infty|\phi_{true}}$-probability one.

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

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

holds with $P_{Y^\infty|\phi_{true}}$-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 $M \in \mathcal{M}_A$, $f_{\phi|M}(\phi_{true}) > 0$.

Assumption 3.2 (i) implies that none of the models has ϕ_{true} on the boundary of its reduced-form parameter space. $\mathcal{M}_A$ defined in Assumption 3.2 (i) collects the models that are observationally equivalent at ϕ_{true} in the sense of Definition 3.2 (i). The requirement that ϕ_{true} be in the interior of Φ_M implies that Φ_M, $M \in \mathcal{M}_A$, has a non-empty interior in $\mathbb{R}^d$. For a set-identified model, condition (A) implies that $M^s \in \mathcal{M}_A$ has a non-empty identified set in an open neighborhood of ϕ_{true}, and condition (B) implies that $M^s \in \mathcal{M}^s \setminus \mathcal{M}_A$ has an empty identified set in an open neighborhood of ϕ_{true}. Assumptions 3.2 (iii) and (iv) impose regularity conditions that imply almost sure consistency of $\hat{\phi}$. Assumptions 3.2 (ii) and (v), imposing smoothness of the log-likelihood and ϕ-prior, 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 et al. (1990) in their validation of the higher-order expansion of the marginal likelihood.

The next proposition, which is a large sample analogue of Lemma 3.1, derives the limits of the posterior model probabilities.

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

$$\pi_{M|Y^\infty} \equiv \lim_{n\to\infty} \pi_{M|Y^n} = \begin{cases} \frac{f_{\phi|M}(\phi_{true})\cdot\pi_M}{\sum_{M'\in\mathcal{M}_A} f_{\phi|M'}(\phi_{true})\cdot\pi_{M'}}, & \text{for } M \in \mathcal{M}_A, \\ 0, & \text{for } M \notin \mathcal{M}_A. \end{cases} \tag{3.14}$$

with $P_{Y^\infty|\phi_{true}}$-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 $M \in \mathcal{M}$,

$$\pi_{M|Y^\infty} = \begin{cases} \frac{\tilde{\pi}_\phi(\Phi_M)^{-1}\cdot\pi_M}{\sum_{M'\in\mathcal{M}_A} \tilde{\pi}_\phi(\Phi_{M'})^{-1}\cdot\pi_{M'}}, & \text{for } M \in \mathcal{M}_A, \\ 0, & \text{for } M \notin \mathcal{M}_A. \end{cases} \tag{3.15}$$

with $P_{Y^\infty|\phi_{true}}$-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 irrespective of the prior probabilities. 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 are determined by the prior model probabilities and the densities of the ϕ-priors evaluated at ϕ_{true}. This implies that the sensitivity of the post-averaging posterior to the choices of ϕ-priors and prior model probabilities does not vanish asymptotically when multiple models are observationally equivalent at ϕ_{true}. Third, when the ϕ-priors share a common kernel, as assumed in Proposition 3.3 (ii), the asymptotic model probabilities are proportional to the reciprocal of the prior probability (in terms of $\tilde{\pi}_\phi$) that the distribution of data is consistent with the identifying assumptions. Hence, the asymptotic posterior model probabilities are higher for more observationally restrictive models, i.e., 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), which the standard Bayesian model selection/averaging is typically equipped with — we should prefer a more parsimonious model among those that explain the data equally well. Note that the notion of parsimony here refers to the size of the reduced-form parameter spaces, and has nothing to do with the strength of the identifying assumptions (often measured by the width of the identified set for α).[16]

Proposition 3.3, which assumes the reduced-form model is correctly specified and common among the candidate models, can be extended to the case where the common reduced-form model is misspecified. In such case, if we interpret ϕ_{true} as the unique pseudo-true parameter and maintain Assumption 3.2, we can show that the maximum likelihood estimator $\hat{\phi}$ and the posterior for ϕ are consistent estimators of the pseudo-true parameters in models such that Φ_M contains ϕ_{true}. The proof of Proposition 3.3 then carries over to show that the asymptotic model probabilities remain valid even with misspecified reduced-form models.

A combination of the asymptotic posterior model probabilities obtained in Proposition 3.3 and the asymptotic behavior of $\pi_{\alpha|M,Y^n}$ for single-prior models and of $\Pi_{\alpha|M,Y^n}$ for multiple-prior models yields the asymptotic convergence properties of the set of post-averaging posteriors. To be specific, in addition to Assumption 3.2, we assume that (i) the posterior for ϕ is consistent for ϕ_{true} with $P_{Y^\infty|\phi_{true}}$-probability one, (ii) for $M^p \in \mathcal{M}^p \cap \mathcal{M}_A$, $\alpha_{M^p}(\cdot)$ is continuous at ϕ_{true} and the posterior of $\alpha_{M^p}(\phi)$ is uniformly integrable with $P_{Y^\infty|\phi_{true}}$-probability one, and (iii) for $M^s \in \mathcal{M}^s \cap \mathcal{M}_A$, $IS_\alpha(\phi|M^s)$ is a compact and continuous correspondence at ϕ_{true} and the posteriors of $l(\phi|M^s)$ and $u(\phi|M^s)$ are uniformly integrable with $P_{Y^\infty|\phi_{true}}$-probability one. Then, the set of post-averaging posterior means considered in Proposition 3.2 (i) has the following limits:

$$
\begin{aligned}
&\lim_{n\to\infty} \left[\inf_{\pi_{\alpha|Y^n} \in \Pi_{\alpha|Y^n}} E_{\alpha|Y^n}(\alpha), \sup_{\pi_{\alpha|Y^n} \in \Pi_{\alpha|Y^n}} E_{\alpha|Y^n}(\alpha) \right] \\
&= \sum_{M^p \in \mathcal{M}^p \cap \mathcal{M}_A} \alpha_{M^p}(\phi_{true}) \pi_{M^p|Y^\infty} + \left[\sum_{M^s \in \mathcal{M}^s \cap \mathcal{M}_A} l(\phi_{true}|M^s) \pi_{M^s|Y^\infty}, \sum_{M^s \in \mathcal{M}^s \cap \mathcal{M}_A} u(\phi_{true}|M^s) \pi_{M^s|Y^\infty} \right].
\end{aligned}
$$

[16]For instance, in a SVAR, a model point-identified by a set of 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 two 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 sample.

4 Discussion

4.1 Relationship with ϵ-contaminated Class of Priors

The method proposed in this paper has a close link to performing robust Bayes analysis using an *ϵ-contaminated class of priors* (Huber (1973), Berger and Berliner (1986)). To clarify this, consider the simple case of one single posterior model and one multiple posterior model, $\mathcal{M} = \{M^p, M^s\}$. Further assume that the models share the same parameterization of the structural model and 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 $\Pi_{\theta|M^s}$ in the form of (3.6), consider the set of priors for θ constructed by marginalizing $\Pi_{\theta,M}$ of Proposition 3.1 to θ;

$$\Pi_\theta \equiv \left\{ \pi_\theta = \pi_{\theta|M^p}\pi_{M^p} + \pi_{\theta|M^s}\pi_{M^s} : \pi_{\theta|M^s} \in \Pi_{\theta|M^s} \right\}. \tag{4.1}$$

Similarly to Proposition 3.1, we obtain the post-averaging ambiguous belief $\Pi_{\alpha|Y}$ by updating Π_θ prior-by-prior with the common likelihood of θ and marginalizing to α.

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

$$\Pi_\theta^\epsilon \equiv \left\{ \pi_\theta = (1-\epsilon)\pi_\theta^0 + \epsilon q_\theta : q_\theta \in \mathcal{Q} \right\}, \tag{4.2}$$

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), a motivation for considering the ϵ-contaminated class of priors can be stated as follows. The researcher can express an initial believable prior for θ as π_θ^0, but the elicitation process is subject to error by some amount specified by ϵ. q_θ captures in what way π_θ^0 differs from the most credible prior and $\mathcal{Q}$ specifies the set of possible departures. Huber (1973) and Berger and Berliner (1986) show the sets of posterior probabilities for various specifications of $\mathcal{Q}$ when a prior varies over Π_θ^ϵ.

Despite the fact that the motivation for our averaging procedure differs from the original motivation of the ϵ-contaminated class of priors, the prior input of our averaging procedure specified in (4.1) has the same form as the ϵ-contaminated class of priors (4.2) — Π_θ is an ϵ-contaminated class of priors where the benchmark prior is 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 the set of priors $\mathcal{Q}$ corresponds to the multiple priors for the set-identified model $\Pi_{\theta|M^s}$. This clarifies a robust Bayes interpretation of our averaging method.[17] If the single-prior (point-identified) model plays the role of a sensible benchmark

[17] As an alternative to the prior-by-prior updating, Berger and Berliner (1986) also considers the Type-II Maximum Likelihood updating rule (empirical Bayes updating rule) of Good (1965). This alternative approach resolves ambiguity by selecting from the class a prior that maximizes the marginal likelihood. Note that the Type-II Maximum Likelihood procedure fails to select a unique prior from Π_θ, because $\pi_{\theta|M^s} \in \Pi_{\theta|M^s}$ sharing a common prior for ϕ has a constant marginal likelihood.

model subject to potential misspecification, averaging it with the set-identified model with weight π_{M^s} can be interpreted as performing sensitivity analysis by contaminating the prior of the point-identified model by an amount π_{M^s} in every possible direction subject to the set-identifying assumptions. Similarly to fixing ϵ in the robust Bayes analysis with an ϵ-contaminated class of priors, our model averaging procedure fixes π_{M^s} no matter whether M^p and M^s are distinguishable or not. To treat indistinguishable models and non-identifiable parameters in M^s in the same way, we could additionally introduce ambiguity for π_{M^s}. If M^p is nested in M^s, however, full ambiguity of $\pi_{M^s} \in [0, 1]$ would lead to the same set of posteriors that would be obtained from the set-identified model only, i.e., $\pi_{M^s} = 1$.

This perspective also contrasts our averaging procedure with standard Bayesian model averaging. If π_θ^0 in (4.2) is degenerate for some θ due to the point-identifying assumption while q_θ is not, π_θ can be viewed as a spike-and-slab prior. It is well known that Bayesian model averaging with a model involving a dogmatic constraint can be replicated by the Bayesian procedure with the spike-and-slab prior. Our model averaging procedure, in contrast, exploits the *class of spike-and-slab priors* spanned by $q_\theta \in \mathcal{Q}$ and draws robust conclusions that hold for all $q_\theta \in \mathcal{Q}$.

The robust Bayes literature on ϵ-contaminated priors has considered several specifications of $\mathcal{Q}$ that lead to analytically tractable classes of posteriors (Berger and Berliner (1986)). To our knowledge, however, the class of priors in the form of $\Pi_{\theta|M^s}$ has not been investigated. Motivated by partial identification analysis, our analysis offers a new way to specify $\mathcal{Q}$ without losing analytical and numerical tractability.

4.2 Relationship with Hierarchical Bayesian Approach

Point-identifying assumptions or a prior for structural parameters sometimes come from a structural econometric model based on economic theory. A set-identified model, in contrast, may represent a "semi-structural" heuristic description of the underlying causal mechanisms with a flexible functional form. For instance, in empirical macroeconomic policy analysis, we can view a DSGE model as a single-prior model and a sign restricted SVAR model as a set-identified model.

In such contexts, averaging models offers a way to combine the structural modelling approach and a more "reduced-form" approach.[18] The macroeconometrics literature has proposed using hierarchical Bayesian methods to bridge the gap between structural and "reduced-form" approaches (Del Negro and Schorfheide (2004)), in which the structural parameters in the DSGE model act as hyperparameters of a prior for SVAR parameters.

The robust Bayes averaging approach, albeit similar in motivation in such contexts, differs

[18] What we mean by "reduced-form" approach here differs from the technical terminology of the reduced-form model/parameters in our expositions.

from the hierarchical Bayesian approach in several ways. First, the hierarchical Bayesian approach always leads to a single posterior for the impulse responses, no matter whether they are identified or not in the SVAR model. If they are not, this means that the prior for the structural parameters in the DSGE model and the prior for the SVAR parameters (given the hyperparameters) have some part that is unrevisable by the data. Hence, if one cannot specify these priors with full confidence, posterior sensitivity may well become a concern. In contrast, our procedure classifies the DSGE model as a single posterior model and the set-identified SVAR as a multiple-prior model. Limited credibility in the prior for the Bayesian DSGE model can be incorporated into the posterior inference by averaging it with the set-identified SVAR model with carefully specified π_{M^s}. Second, in the hierarchical Bayesian approach, tightness of the prior around the mean predicted by the DSGE model plays the role of prior confidence assigned to the structural model. In our procedure, the model probability assigned to the structural model governs the degree of confidence. It is however important to distinguish the notions of confidence between the two approaches, since the former is in the scale of Bayesian probabilistic uncertainty while the latter is in the scale of ambiguity (Knightian uncertainty).

4.3 Eliciting Prior Model Probabilities

The key prior input of our procedure is the prior model probability. A natural starting point is to assume a uniform distribution of prior probabilities, however our procedure can readily accommodate non-uniform probabilities. Discussions on how to determine prior probabilities in Bayesian averaging are in, e.g., George (1999) in the discussion of Clyde (1999), where, in order to prevent from overvaluing similar models, he suggests a "dilution" technique, i.e., if some models are similar, the weight attached to the original model should be split between that model and its duplicates. Among others, Chipman (1996) attaches smaller prior probabilities to models that are unlikely, Hoeting et al. (1999) rely on variable selection in regression models to determine prior probabilities and Clyde and George (2004) propose a Bernoulli specification.

In our context, the robust Bayesian viewpoint based on the ϵ-contaminated class of priors can help clarify the interpretation of the prior model probabilities and facilitate their elicitation.

Suppose again that the set of candidate models consists of one point-identified model M^p and one set-identified model M^s. Assume in addition that M^p is nested in M^s, in the sense that the identifying assumptions in M^p include those in M^s. In this case, the prior model probability assigned to M^p can be interpreted as the *minimal* amount of credibility assigned to the identifying assumptions in model M^p, and the prior model probability assigned to the set-identified model can be interpreted as the *maximal* amount of contamination given to the point-identifying assumptions imposed in M^p but not in M^s. The reason that π_{M^p} is giving the credibility lower bound for model M^p is that, when model M^s nests model M^p, the set of priors specified in model M^s contains beliefs that assign full or partial credibility to the identifying assumptions in M^p. As a result, any prior probability between $[\pi_{M^p}, 1]$ can be attained for the credibility of the identifying assumptions in M^p.

The interpretation of the prior model probabilities differs when the identifying assumptions in models M^p and M^s are non-overlapping. In this case, the prior model probabilities are interpreted as the standard probabilistic belief assigned over mutually exclusive models.

When the identifying assumptions in models M^p and M^s are non-nested but overlapping (e.g., Scenario 1 in Section 2), interpreting the model probabilities may not appear as clear-cut as in the previous two cases. However, the lower credibility bound interpretation of π_{M^p} given in the nested case above remains valid. What differs from the nested case is that the maximal credibility that can be assigned to the identifying assumptions in M^p can be strictly less than one.

5 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) and Moon et al. (2013). We employ the Hannan-Quinn (HQ) information criterion to select the number of lags, that is three. Following Definition 3 in Giacomini and Kitagawa (2020), we order the variables so that we can easily verify the conditions guaranteeing convexity of the identified set using their Proposition B.1.

$$A_0 y_t = a + \sum_{j=1}^{3} A_j y_{t-j} + \epsilon_t, \quad \textit{for } t = 1, \ldots, T \tag{5.1}$$

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

$$A_0 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{5.2}$$

Assume $\epsilon_t = (\epsilon_t^i, \epsilon_t^{\Delta gdp}, \epsilon_t^{\pi})'$ are i.i.d. normally distributed with mean zero and variance-covariance the identity matrix I_3. The corresponding reduced-form VAR is:

$$y_t = b + \sum_{j=1}^{3} B_j y_{t-j} + u_t, \tag{5.3}$$

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

The first equation in (5.1) is interpreted as a monetary policy function: the Federal Reserve reacts to price and GDP, as well as lags of all variables. The second and third equations represent aggregate demand (AD) and aggregate supply (AS), respectively. The data are quarterly observations from 1965:1 to 2005:1 from the FRED2 database.

The prior for the reduced-form parameters is a Random Walk, relatively flat and belongs to the Normal Inverse-Wishart family:[19]

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

where $\beta \equiv vec([b, B_1, \ldots, B_4]')$. $\Psi = I_3$ 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. We set $\bar{b}$, which is the prior mean of β, such that each endogenous time series is a Random Walk a-priori. In what follows, we perform Algorithm 3.1 with the 1000 number of draws of ϕ's from the Normal Inverse-Wishart posterior.

Following Christiano et al. (1999), we impose the sign normalization restrictions so that the diagonal elements of A_0 are nonnegative. As a result, we can interpret a unit positive change in a structural shock as a one standard-deviation positive shock to the corresponding variable.

5.1 Averaging Indistinguishable Models

Suppose we are interested in the cumulative output growth response[20] to a unit positive shock in the federal funds rate ϵ_t^i at horizon h, $IR^h_{\Delta gdp,i}$, and consider the following two sets of identifying assumptions.

- *Model 1 (M1, point-identified)* Assume that AD and AS do not react on impact to the interest rate shock: $a_{21} = a_{31} = 0$ in (5.1) and (5.2). This identification scheme point-identifies $IR^h_{\Delta yi}$; note that such identifying restrictions are also implied by standard recursive causal ordering restrictions (Bernanke (1986) and Sims (1980)).[21]

- *Model 2 (M2, set-identified through zero restrictions)*

  The identification scheme in Model 1 is controversial. For example, assumption $a_{31} = 0$, implying that prices do not react contemporaneously to the interest rate shock, can be difficult to justify if the researcher relies on quarterly data or on the stock price index rather than the GDP deflator.[22] Thus, in Model 2 we leave AS unrestricted, i.e., AS can react to the interest rate within a quarter and the zero restrictions are now $a_{21} = 0$. By

[19] In order to reduce the computational burden, we use a conjugate prior as its posterior and marginal likelihood is analytically available.

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

[21] Causal ordering restrictions point-identify all the shocks in (5.1) also assuming $a_{23} = 0$. However, as long as we are interested in point-identifying monetary shock only, restrictions in Model 1 are equivalent to causal ordering restrictions: $a_{23} = 0$ does not affect the identified set of $IR^h_{\Delta gdp,i}$ (see Corollary B.1 in Giacomini and Kitagawa (2020)).

[22] See Kilian (2013) for details over the limitations of point-identifying assumptions.

Proposition B.1 in Giacomini and Kitagawa (2020), Model 2 delivers a convex identified set for $IR^h_{\Delta gdp,i}$ for every value of the reduced form parameters.

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 for the impulse response; 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 average results. The vertical dashed lines for the averaged model can be viewed as shrinking the identified set estimator from Model 2 towards the point estimator from Model 1. Figure 2 reports the output response estimation and inference for multiple horizons for the same models as in Figure 1.

Note that, as is common for standard recursive causal ordering restrictions in small-scale SVARs, the point-identified Model 1 shows a negative response of output in the short run, whereas the set-identified Model 2 supports both positive and negative effects. This is confirmed by the last row in Table 2, reporting the lower and upper probability that the post-averaging interval of posterior means of the output response is negative.[23] Averaging the models still does not support a negative output response, as the 90% robust credibility region always crosses the zero line. Note that in this case the models are indistinguishable and so the prior model probabilities are not updated by the data. If researcher attached a higher prior probability to Model 1, we would observe a bigger shrinkage of the post-averaging interval of posterior means and robust credibility region towards the point-identified case.[24]

5.2 Averaging Distinguishable Models

In the previous example, model prior probabilities are not updated. Here we consider more interesting cases, in which model prior probabilities get updated, by adding two models that are widely used in empirical applications: 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 (2020), the identified set in Model 3 is convex.

[23] This is computed as the probability that the identified set lies entirely in the negative real halfline and the probability that the identified set intersects with the negative real halfline. For point-identified models, the set of probabilities collapses to a singleton.

[24] The case for $w_1 = 0.8$ is available upon request.

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 candidate models are distinguishable due to the imposition of observationally restrictive sign restrictions. Appendix A.3 provides two algorithms for approximating the posterior-prior plausibility ratio for the sign-restricted SVARs. Here, we report the results based on Algorithm A.1, as the other algorithm (Algorithm A.2) produces almost identical results.[25]

Panel (f) of Figure 1 and 2 reports the results of averaging the two models: as in the case of Model 2, Model 3 does not support a negative 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 favour Model 3 (with posterior probability 0.55), and the average of the two models does not support a negative 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 et al. (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 Figure 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 does not support a negative response.

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

5.3 Reverse-Engineering Prior Model Probabilities

Our method lends itself to useful reverse-engineering exercises that help shed light on the role of identifying assumptions in drawing inferences. Specifically, we compute the prior probability one needs to attach to a set of assumptions in order for the averaging to preserve certain model's

[25] Only a few sign restrictions imposed, the set of q_1's satisfying the sign restrictions is not small for most of the draws of ϕ. Hence, for assessing non-emptiness of the identified set, the numerical approximation of Algorithm A.1 works as good as the analytical method of Algorithm A.2. In terms of computation time, when $K = 1000$ draws of ϕ is used, Algorithm A.2 takes 12.35 seconds to compute O_3, while Algorithm A.1 with $L = 90000$ draws of q_1 takes 1377.79 seconds.

conclusions. In our application, for example, we saw that the negative response of output to a contractionary monetary policy shock disappears once relaxing either standard causal ordering restrictions or the restrictions embedded in a DSGE model. We can thus compute the prior weight one would assign to a given set of restrictions such that the posterior mean bounds, or the robust credible region, is contained in the negative part of the real line.

First consider Model 1 (point-identification through causal ordering restrictions) and Model 2 (set-identification by relaxing one restriction from Model 1). Letting w be the prior probability of Model 1, the post-averaging interval of posterior means is

$$\left[\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha), \sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha)\right] =$$

$$= \pi_{M^1|Y} E_{\alpha|M^1,Y}(\alpha) + \pi_{M^2|Y}\left[E_{\phi|Y,M^2}(l(\phi_{M^2}|M^2)), E_{\phi|Y,M^2}(u(\phi_{M^2}|M^2))\right]$$

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

We want to compute the prior model probability w such that the post-averaging interval of posterior means is contained in the negative part of the real line at $h = 3$. This is equivalent to solving a system of inequalities where w is the unknown.

We find that one would need to attach a prior probability greater than 0.05 to the validity of the assumption that prices do not react contemporaneously to an interest rate shock in order to have the post-averaging interval of posterior means contained in the negative real halfline.

We next consider Model 1 and Model 3 (set-identification through sign restrictions). The reverse-engineering exercise proceeds as before, with the only difference that now the posterior model probabilities are updated and are equal to

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

We find that the post-averaging interval of posterior means is contained in the negative real halfline only if $w > 0.76$. As expected, one would need to attach very high prior probability to the causal ordering restrictions to obtain a negative output response.

Another possibility is to conduct reverse engineering on robust credible region rather than on post-averaging interval of posterior means. We can thus compute the prior weight one would assign to a given set of restrictions such that the 90% robust credibility region is contained in the negative real halfline. When averaging Model 1 and Model 3 one would need to attach a prior probability greater than 0.87 to the validity of the assumptions in Model 1 in order to have the 90% robust credibility region contained in the negative real halfline at $h = 3$.[26]

[26]The prior model probability drops to 0.44 if we weight Model 1 and Model 2.

However, this exercise is sensitive to the choice of the credibility level of the robust credible region we focus, as the robust credible region tightens up as soon as the credibility level falls below the model probability assigned to the single posterior model.

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.[27]

[27]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).

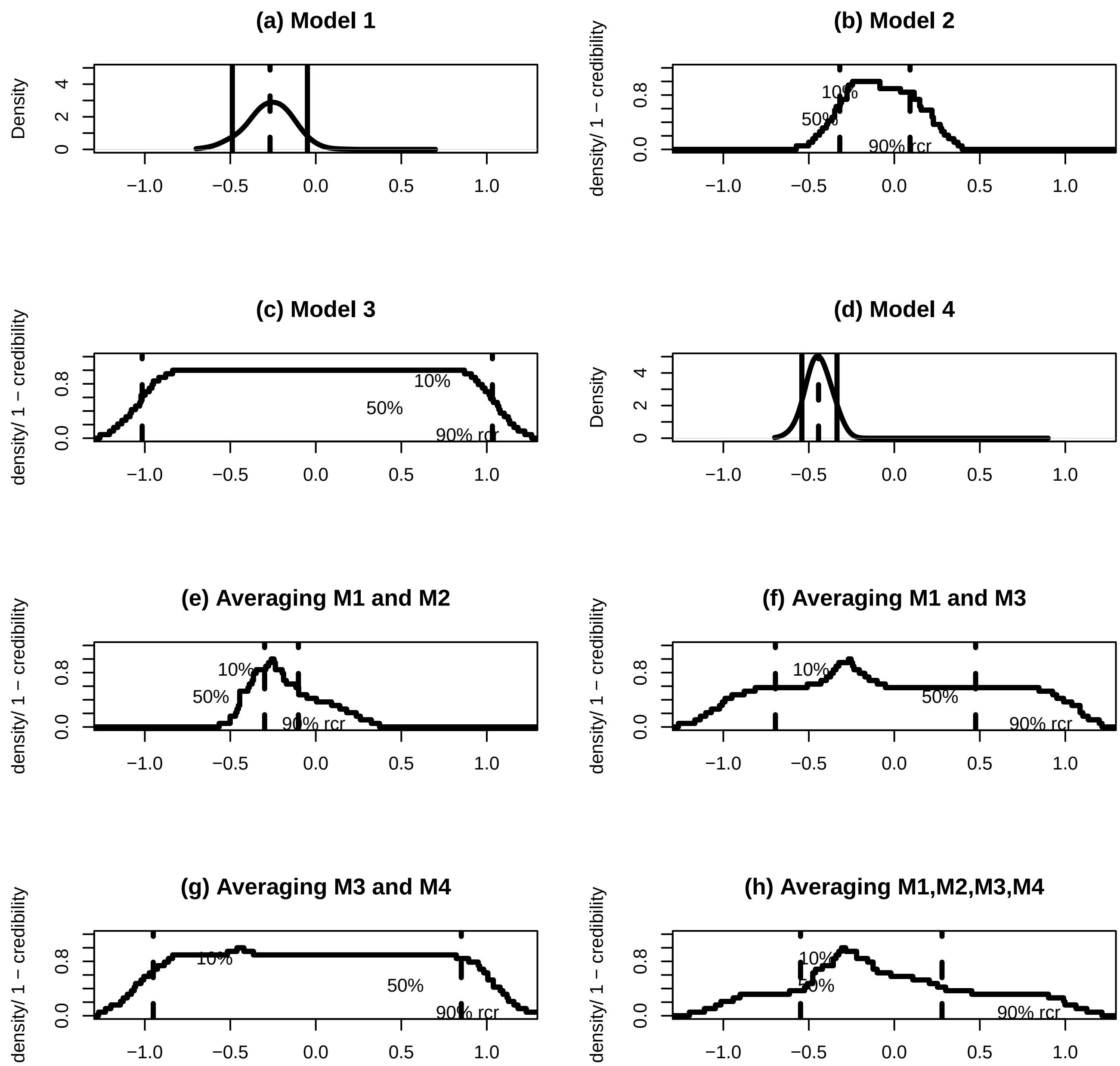

Figure 1: Density and Robust Credible Region of Output Impulse Responses

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 (90%, 50%, 10% levels are explicitly indicated). 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.

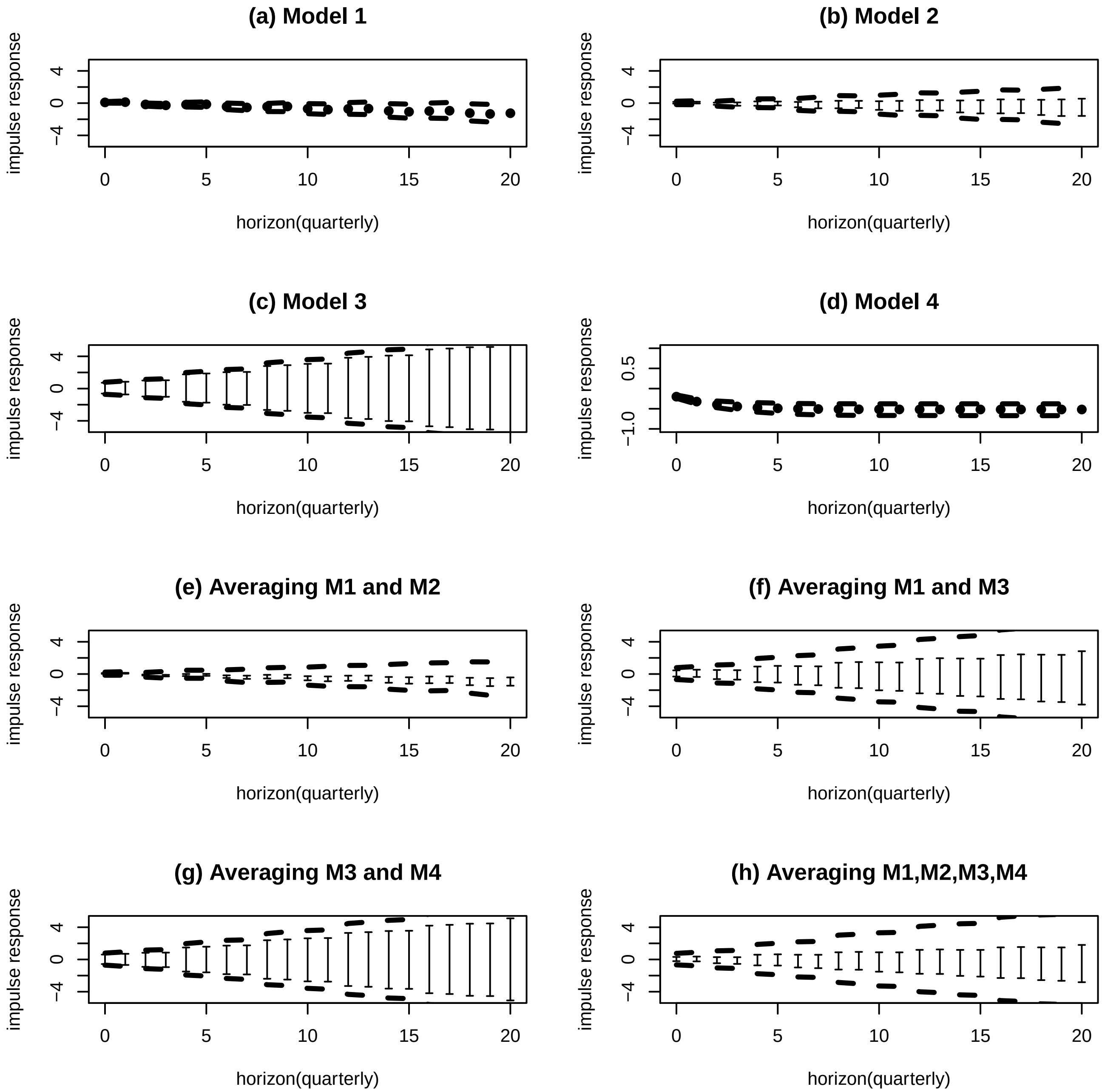

Figure 2: Plots of Output Impulse Responses

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%.

	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

Table 1: Output Responses: Prior and Posterior Weights

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.

	M1			M2		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	.01	-1.09	-1.97	/	/	/
90% CR	$[-.09, .09]$	$[-1.58, -.61]$	$[-2.88, -1.12]$	/	/	/
Post. Mean Bounds	/	/	/	$[-.07, .07]$	$[-1.12, .05]$	$[-2.03, .10]$
90% robust CR	/	/	/	$[-.15, .15]$	$[-1.58, .50]$	$[-2.87, .93]$
Set of $\Pi_{IR^h\|Y}\{IR^h < 0\}$	0.46	1	1	$[0.08, 0.89]$	$[0.43, 1]$	$[0.41, 1]$
	M3			M4		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	$-.32$	$-.52$	$-.52$
90% CR	/	/	/	$[-.39, -.26]$	$[-.67, -.38]$	$[-.67, -.38]$
Post. Mean Bounds	$[-.73, .85]$	$[-3.03, 3.08]$	$[-5.73, 5.91]$	/	/	/
90% robust CR	$[-.86, .96]$	$[-3.56, 3.65]$	$[-6.75, 6.95]$	/	/	/
Set of $\Pi_{IR^h\|Y}\{IR^h < 0\}$	$[0, 1]$	$[0, 1]$	$[0, 1]$	1	1	1
	Averaging M1,M2			Averaging M1,M3		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	/	/	/
90% CR	/	/	/	/	/	/
Post. Mean Bounds	$[-.03, .04]$	$[-1.11, -.50]$	$[-2.01, -.89]$	$[-.40, .47]$	$[-2.17, 1.21]$	$[-4.06, 2.39]$
90% robust CR	$[-.13, .14]$	$[-1.58, .39]$	$[-2.83, .74]$	$[-.83, .93]$	$[-3.50, 3.47]$	$[-6.66, 6.65]$
Set of $\Pi_{IR^h\|Y}\{IR^h < 0\}$	$[0.27, 0.68]$	$[0.70, 1]$	$[0.69, 1]$	$[0.19, 0.74]$	$[0.45, 1]$	$[0.45, 1]$
	Averaging M3,M4			Averaging M1,M2,M3,M4		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	/	/	/
90% CR	/	/	/	/	/	/
Post. Mean Bounds	$[-.69, .72]$	$[-2.77, 2.70]$	$[-5.19, 5.24]$	$[-.29, .30]$	$[-1.76, .73]$	$[-3.25, 1.46]$
90% robust CR	$[-.85, .96]$	$[-3.54, 3.57]$	$[-6.70, 6.86]$	$[-.79, .91]$	$[-3.38, 3.29]$	$[-6.36, 6.37]$
Set of $\Pi_{IR^h\|Y}\{IR^h < 0\}$	$[0.11, 1]$	$[0.11, 1]$	$[0.11, 1]$	$[0.22, 0.81]$	$[0.48, 1]$	$[0.48, 1]$

Table 2: Output Responses: Estimation and Inference

6 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 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 MCMC methods. 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 can also offer a simple and flexible way to conduct sensitivity analysis for the point-identified model.

A Appendix

A.1 Omitted Proofs

Derivation of identified set (2.2). Following Uhlig (2005), we reparameterize A via the Cholesky matrix Σ_{tr} and a rotation matrix $Q = \begin{pmatrix} cos\rho & -sin\rho \\ sin\rho & cos\rho \end{pmatrix}$ with spherical coordinate $\rho \in [0, 2\pi]$. We can then express α as a function of ϕ and the non-identified parameter ρ indexing a rotation matrix;

$$A^{-1} = \Sigma_{tr} Q = \begin{pmatrix} \sigma_{11}\cos\rho & -\sigma_{11}\sin\rho \\ \sigma_{21}\cos\rho + \sigma_{22}\sin\rho & -\sigma_{21}\sin\rho + \sigma_{22}\cos\rho \end{pmatrix}$$

and the parameter of interest is $\alpha = \alpha(\rho, \phi) \equiv \sigma_{11}\cos\rho$. We impose the sign normalization restrictions throughout by constraining the diagonal elements of A to being nonnegative,

$$\sigma_{22}\cos\rho - \sigma_{21}\sin\rho \geq 0 \text{ and } \sigma_{11}\cos\rho \geq 0. \tag{A.1}$$

The sign restrictions $a_{12} \geq 0$ and $a_{21} \leq 0$ are expressed as

$$\sigma_{11}\sin\rho \geq 0 \tag{A.2}$$

$$-\sigma_{22}\sin\rho - \sigma_{21}\cos\rho \leq 0. \tag{A.3}$$

Given ϕ, the identified set for $\alpha = \sigma_{11}\cos\rho$ is given by its set as ρ varies over the set characterized by the restrictions (A.1) - (A.3). Since the second constraint in (A.1) and (A.2) imply $\rho \in [0, \pi/2]$, we focus on how the other two restrictions (the first constraint in (A.1) and (A.3)) tighten up $\rho \in [0, \pi/2]$.

Assume $\sigma_{21} > 0$. Then, they imply $\arctan(-\sigma_{21}/\sigma_{22}) \leq \rho \leq \arctan(\sigma_{22}/\sigma_{21})$. Intersecting this interval with $\rho \in [0, \pi/2]$ leads to $[0, \arctan(\sigma_{22}/\sigma_{21})]$ as the identified set for ρ. Hence, the identified set for α in the $\sigma_{21} > 0$ case follows. A similar argument leads to the α identified set for the $\sigma_{21} \leq 0$ case. ■

Proof of Lemma 3.1. (i) By the construction of ϕ-prior (3.2), the marginal likelihood for $M \in \mathcal{M}^s$ can be rewritten as

$$\begin{aligned} p(Y|M) &= \int_\Phi p(Y|\phi, M) d\pi_{\phi|M}(\phi) \\ &= \int_\Phi p(Y|\phi) \cdot \frac{1\{IS_\alpha(\phi|M) \neq \emptyset\}}{\tilde{\pi}_\phi(IS_\alpha(\phi|M) \neq \emptyset)} d\tilde{\pi}_\phi(\phi) \\ &= \tilde{p}(Y) \int_\phi \frac{1\{IS_\alpha(\phi|M) \neq \emptyset\}}{\tilde{\pi}_\phi(IS_\alpha(\phi|M) \neq \emptyset)} d\tilde{\pi}_{\phi|Y}(\phi) \\ &= \tilde{p}(Y) \frac{\tilde{\pi}_{\phi|Y}(IS_r(\phi|M) \neq \emptyset)}{\tilde{\pi}_\phi(IS_r(\phi|M) \neq \emptyset)} = \tilde{p}(Y) O_M, \end{aligned}$$

where the second line uses the assumption that the set-identified models admit an identical reduced-form and the third line follows from the Bayes theorem for the reduced-form parameters, $p(Y|\phi)\tilde{\pi}_{\phi}(\phi) = \tilde{p}(Y)\tilde{\pi}_{\phi|Y}(\phi)$. Plugging this expression of the marginal likelihood into (3.1) leads to the claim.

(ii) Under the additionally imposed assumptions, the marginal likelihood of model $M^p \in \mathcal{M}^p$ is given by $\tilde{p}(Y)O_{M^p}$. Hence, combined with $p(Y|M^s) = \tilde{p}(Y)O_{M^s}$ shown in part (i), (3.5) follows.

(iii) The claim follows immediately by noting that the imposed assumptions imply $O_M = 1$ for all $M \in \mathcal{M}$ and setting O_M, $M \in \mathcal{M}$, to one in (3.5). ■

Derivation of $\Pi_{\alpha|M^s,Y}$ in equation (3.7). We derive $\Pi_{\alpha|M^s,Y}$ in the next lemma:

Lemma A.1 *For a set-identified model M^s with the structural parameters $\theta_{M^s} \in \Theta_{M^s}$ and reduced-form parameters $\phi_{M^s} = g_{M^s}(\theta_{M^s}) \in \Phi_{M^s} = g_{M^s}(\Theta_{M^s})$, let a prior for ϕ_{M^s}, $\pi_{\phi_{M^s}|M^s}$ be given. Define the class of priors of θ_{M^s} by*

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

Updating $\Pi_{\theta_{M^s}|M^s}$ prior-by-prior with the likelihood $\tilde{p}(Y|\theta_{M^s}, M^s)$ and marginalizing the resulting posteriors via $\alpha = \alpha_{M^s}(\theta_{M^s})$ leads to the following set of posteriors for α:

$$\begin{aligned}&\Pi_{\alpha|M^s,Y}\\ &\equiv \left\{\pi_{\alpha|M^s,Y} = \int_{\Phi_M} \pi_{\alpha|M^s,\phi_{M^s}} d\pi_{\phi_{M^s}|M^s,Y} : \pi_{\alpha|M^s,\phi_{M^s}}(IS_{\alpha}(\phi_{M^s}|M^s)) = 1,\ \pi_{\phi_{M^s}|M^s}\text{-}a.s.\right\}.\end{aligned} \tag{A.4}$$

■

Proof of Lemma A.1. The prior-by-prior updating rule updates $\Pi_{\theta_{M^s}|M^s}$ to

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

Since $\pi_{\theta_{M^s}|M^s,Y}(\Theta_{M^s} \cap g_{M^s}^{-1}(B))$ can be written as

$$\pi_{\theta_{M^s}|M^s,Y}(\Theta_{M^s} \cap g_{M^s}^{-1}(B)) = \int_B \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s}))d\pi_{\phi_{M^s}|M^s,Y}(\phi_{M^s}),$$

the ϕ_{M^s}-marginal constraints for $\pi_{\theta_{M^s}|M^s,Y}$ are equivalent to

$$\int_B \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s}))d\pi_{\phi_{M^s}|M^s,Y}(\phi_{M^s}) = \pi_{\phi_{M^s}|M^s,Y}(B).$$

This equality holds for all $B \in \mathcal{B}(\Phi_{M^s})$ if and only if $\pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s})) = 1$, $\pi_{\phi_{M^s}|M^s,Y}$-a.s. Accordingly, an equivalent representation of the class of posteriors for θ_{M^s} is

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

(A.5)

Note that we have

$$\pi_{\alpha|\phi_{M^s},M^s}(IS_\alpha(\phi_{M^s}|M^s)) = \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\alpha_{M^s}^{-1}(IS_\alpha(\phi_{M^s}|M^s))) \\ = \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s})),$$

where the second equality follows by the definition of the identified set of α. Hence, $\pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s})) = 1$, $\pi_{\phi_{M^s}|M^s,Y}$-a.s. holds if and only if $\pi_{\alpha|\phi_{M^s},M^s}(IS_\alpha(\phi_{M^s}|M^s)) = 1$, $\pi_{\phi_{M^s}|M^s,Y}$-a.s. The class of marginalized posteriors for α (A.4) therefore follows. ■

Proof of Proposition 3.1. Let $\pi_{\theta,M}$ be a prior of (θ, M) belonging to the proposed $\Pi_{\theta,M}$. The corresponding posterior for θ with M integrated out can be computed as follows: for any measurable subset $H \subset \Theta$,

$$\begin{aligned}
\pi_{\theta|Y}(H) &= \frac{\sum_{M\in\mathcal{M}} \int_H \tilde{p}(Y|\theta, M) d\pi_{\theta|M}(\theta)\pi_M}{\sum_{M\in\mathcal{M}} \left[\int_{\Theta_M} \tilde{p}(Y|\theta, M) d\pi_{\theta|M}(\theta)\right] \pi_M} \\
&= \frac{\begin{pmatrix} \sum_{M^p\in\mathcal{M}^p} \pi_{\theta|M^p,Y}(H)p(Y|M^p)\pi_{M^p} \\ + \sum_{M^s\in\mathcal{M}^s} \left[\int_{\Phi_{M^s}} \pi_{\theta|\phi_{M^s},M^s}(H)p(Y|\phi_{M^s}, M^s)d\pi_{\phi_{M^s}|M^s}(\phi_{M^s})\right] \pi_{M^s} \end{pmatrix}}{\sum_{M^p\in\mathcal{M}^p} p(Y|M^p)\pi_{M^p} + \sum_{M^s\in\mathcal{M}^s} \left[\int_{\Phi_{M^s}} p(Y|\phi_{M^s}, M^s)d\pi_{\phi_{M^s}|M^s}(\phi_{M^s})\right] \pi_{M^s}} \\
&= \sum_{M^p\in\mathcal{M}^p} \pi_{\theta|M^p}(H)\pi_{M^p|Y} + \sum_{M^s\in\mathcal{M}^s} \left[\int_{\Phi_{M^s}} \pi_{\theta|\phi_{M^s},M^s}(H)d\pi_{\phi_{M^s}|M^s,Y}(\phi_{M^s})\right] \pi_{M^s|Y}
\end{aligned}$$

where the second line uses

$$\begin{aligned}
\int_H \tilde{p}(Y|\theta, M)d\pi_{\theta|M}(\theta) &= \int_{\Phi_M} \left[\int_\Theta 1\{\theta \in H\}\tilde{p}(Y|\theta, M)d\pi_{\theta|\phi_M,M}(\theta)\right] d\pi_{\phi_M|M}(\phi_M) \\
&= \int_{\Phi_M} \left[\int_\Theta 1\{\theta \in H\}d\pi_{\theta|\phi_M,M}(\theta)\right] p(Y|\phi_M, M)d\pi_{\phi_M|M}(\phi_M) \\
&= \int_{\Phi_M} \pi_{\theta|\phi_M,M}(H)p(Y|\phi_M, M)d\pi_{\phi_M|M}(\phi_M).
\end{aligned}$$

The class of posteriors for θ can be therefore represented as

$$\Pi_{\theta|Y} \equiv \left\{ \sum_{M^p\in\mathcal{M}^p} \pi_{\theta|M^p,Y}\pi_{M^p|Y} + \sum_{M^s\in\mathcal{M}^s} \pi_{\theta|M^s,Y}\pi_{M^s|Y} : \pi_{\theta|M^s,Y} \in \Pi_{\theta|M^s,Y},\ \forall M^s \in \mathcal{M}^s \right\},$$

where $\Pi_{\theta|M^s,Y}$ is as defined in (A.5). As shown in the proof of Lemma A.1 above, marginalizing $\Pi_{\theta|M^s,Y}$ to α leads to $\Pi_{\alpha|M^s,Y}$ defined in (3.7). We therefore conclude that marginalizing $\Pi_{\theta|Y}$ to α results in $\Pi_{\alpha|Y}$ shown in (3.8). ■

Proof of Proposition 3.2. (i) Since there is no constraint across the posteriors belonging to different posterior classes, it holds

$$\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha) = \sum_{M^p\in\mathcal{M}^p} E_{\alpha|M^p,Y}(\alpha)\pi_{M^p|Y} + \sum_{M^s\in\mathcal{M}^s} \inf_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \left\{E_{\alpha|M^s,Y}(\alpha)\right\}\cdot\pi_{M^s|Y}.$$

By the construction of $\Pi_{\alpha|M^s,Y}$, an application of Theorem 2 of Giacomini and Kitagawa (2020) shows $\inf_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}}\left\{E_{\alpha|M^s,Y}(\alpha)\right\} = E_{\phi_{M^s}|M^s,Y}(l(\phi_{M^s}|M^s))$. The claim of the mean lower bound therefore follows. The mean upper bound can be shown similarly.
(ii) Note that

$$\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) = \sum_{M^p\in\mathcal{M}^p} \pi_{\alpha|M^p,Y}(H)\cdot\pi_{M^p|Y} + \sum_{M^s\in\mathcal{M}^s} \inf_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \left\{\pi_{\alpha|M^s,Y}(H)\right\}\cdot\pi_{M^s|Y},$$

$$\sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} \pi_{\alpha|Y}(H) = \sum_{M^p\in\mathcal{M}^p} \pi_{\alpha|M^p,Y}(H)\cdot\pi_{M^p|Y} + \sum_{M^s\in\mathcal{M}^s} \sup_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \left\{\pi_{\alpha|M^s,Y}(H)\right\}\cdot\pi_{M^s|Y}.$$

Theorem 1 of Giacomini and Kitagawa (2020) shows

$$\inf_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \left\{\pi_{\alpha|M^s,Y}(H)\right\} = \pi_{\phi_{M^s}|M^s,Y}(IS_\alpha(\phi_{M^s}|M^s)\subset H),$$

$$\sup_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \left\{\pi_{\alpha|M^s,Y}(H)\right\} = \pi_{\phi_{M^s}|M^s,Y}(IS_\alpha(\phi_{M^s}|M^s)\cap H\neq\emptyset),$$

so the conclusion follows.
(iii) By setting H to $[-\infty,a]$, the lower probability obtained in part (ii) yields the lower bound of the cdfs, since the event $IS_\alpha(\phi_{M^s}|M^s)\subset[-\infty,a]$ is equivalent to $u(\phi_{M^s}|M^s)\leq a$. The upper bound follows by noting

$$\begin{aligned}\sup_{\pi_{\alpha|M^s,Y}\in\Pi_{\alpha|M^s,Y}} \pi_{\alpha|M^s,Y}([\infty,a]) &= \pi_{\phi_{M^s}|M^s,Y}(IS_\alpha(\phi_{M^s}|M^s)\cap[\infty,a]\neq\emptyset)\\ &= \pi_{\phi_{M^s}|M^s,Y}(l(\phi_{M^s}|M^s)\leq a).\end{aligned}$$

The set of quantiles then follows by inverting these cdf bounds. ■

Next, we show two lemmas to be used to prove Proposition 3.3. We denote the set of candidate models satisfying condition (A) of Assumption 3.2 (i) by $\mathcal{M}_A$ and the set of those satisfying condition (B) by $\mathcal{M}_B$. Under Assumption 3.2 (i), $\mathcal{M}=\mathcal{M}_A\cup\mathcal{M}_B$ holds. Note that through these lemmas and the proof of Proposition 3.3, $\mathcal{M}$ is assumed to admit an identical reduced-form with reduced-form parameter dimension $d\geq 1$.

Lemma A.2 *Suppose Assumption 3.2 holds. For $M\in\mathcal{M}_A$,*

$$\frac{n^{d/2}\det(H_n(\hat{\phi}))^{1/2}p(Y^n|M)}{(2\pi)^{d/2}p(Y^n|\hat{\phi})} - f_{\phi|M}(\hat{\phi}) = O(n^{-1/2}),$$

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

Proof of Lemma A.2. Denote the reduced-form parameter vector by $\phi = (\phi_1, \dots, \phi_d)$ and the third-derivative of $l_n(\cdot)$ by $h_{ijk}(\cdot) \equiv \frac{\partial^3}{\partial\phi_i \partial\phi_j \partial\phi_k} l_n(\cdot)$, $1 \leq i, j, k \leq d$. By Assumptions 3.2 (i), (ii) and (iv), there exists B^* an open neighborhood of ϕ_{true} such that $B^* \subset \Phi_M$ holds for all $M \in \mathcal{M}_A$, and

$$\sup_{\phi \in B^*} \max_{1 \leq i,j,k \leq d} |h_{ijk}(\phi)| < \infty, \tag{A.6}$$

and

$$\lim\sup_{n\to\infty} \sup_{\phi \in \Phi \setminus B^*} \{l_n(\phi) - l_n(\phi_{true})\} < 0, \quad \text{with } P_{Y^\infty|\phi_{true}}\text{-probability one} \tag{A.7}$$

hold. Since Assumptions 3.2 (iii) and (iv) imply the strong convergence of $\hat{\phi}$, for all sufficiently large n, $\hat{\phi} \in B^*$ holds. Given $\hat{\phi} \in B^*$, consider the third-order mean value expansions of $nl_n(\phi)$:

$$\begin{aligned}
nl_n(\phi) &= nl_n(\hat{\phi}) - \frac{n}{2}(\phi - \hat{\phi})' H_n(\hat{\phi})(\phi - \hat{\phi}) + \frac{n}{6} \sum_{1 \leq i,j,k \leq d} h_{ijk}(\tilde{\phi})(\phi_i - \hat{\phi}_i)(\phi_j - \hat{\phi}_j)(\phi_k - \hat{\phi}_k) \\
&= nl_n(\hat{\phi}) - \frac{1}{2} u' H_n(\hat{\phi}) u + \frac{1}{\sqrt{n}} R_{1n}(u),
\end{aligned}$$

where $\tilde{\phi}$ is a convex combination of ϕ and $\hat{\phi}$, $u \equiv \sqrt{n}(\phi - \hat{\phi})$, and $R_{1n}(u) = \frac{1}{6}\sum_{1 \leq i,j,k \leq d} h_{ijk}(\tilde{\phi}) u_i u_j u_k$, where u_i is the i-th entry of vector u. By the boundedness of h_{ijk} on B^*, $R_{1n}(u)$ can be bounded by a third-order polynomial of u with bounded coefficients on $\sqrt{n}(B^* - \hat{\phi})$, where $\sqrt{n}(B^* - \hat{\phi})$ is the subset in $\mathbb{R}^d$ that translates B^* by $\hat{\phi}$ and scales up by $\sqrt{n}$. Plugging this expansion in $p(Y^n|\phi) = \exp(nl_n(\phi))$ and combining it with the first-order expansion of $f_{\phi|M}(\phi)$, we obtain on $\phi \in B^*$ (or equivalently on $u \in \sqrt{n}(B^* - \hat{\phi})$)

$$\begin{aligned}
p(Y^n|\phi) f_{\phi|M}(\phi) &= \exp\left\{nl_n(\hat{\phi}) - \frac{1}{2} u' H_n(\hat{\phi}) u\right\} \left\{1 + \frac{1}{\sqrt{n}} R_{1n}(u) + \frac{1}{2n} R_{1n}(u)^2 + \cdots\right\} \\
&\quad \times \left\{f_{\phi|M}(\hat{\phi}) + \frac{1}{\sqrt{n}} R_{2n}(u)\right\} \\
&= \exp\left\{nl_n(\hat{\phi}) - \frac{1}{2} u' H_n(\hat{\phi}) u\right\} \left\{f_{\phi|M}(\hat{\phi}) + \frac{1}{\sqrt{n}} R_{3n}(u)\right\},
\end{aligned} \tag{A.8}$$

where the first equality invokes the expansion of $\exp(x) = 1 + x + 2^{-1}x^2 + \cdots$, $R_{2n} = f'_{\phi|M}(\tilde{\phi})u$, and R_{3n} collects the residual terms that can be bounded uniformly on $\sqrt{n}(B^* - \hat{\phi})$ by a finite order polynomial of u with bounded coefficients.

Integration of $p(Y^n|\phi) f_{\phi|M}(\phi)$ over $\phi \in B^*$ is equivalent to integrating (A.8) in u over

$\sqrt{n}(B^* - \hat{\phi})$:

$$\begin{aligned}
&\int_{B^*} p(Y^n|\phi) f_{\phi|M}(\phi) d\phi \\
=& n^{-d/2} \exp\{n l_n(\hat{\phi})\} \left(\int_{\sqrt{n}(B^*-\phi_{true})} \left(f_{\phi|M}(\hat{\phi}) + R_{3n}(u) \right) \exp\left\{ -\frac{1}{2} u' H_n(\hat{\phi}) u \right\} du \right) \\
=& (2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2} \left(f_{\phi|M}(\hat{\phi}) E_{H_n}[1_{\sqrt{n}(B^*-\hat{\phi})}(u)] + n^{-1/2} E_{H_n}[R_{3n}(u) \cdot 1_{\sqrt{n}(B^*-\hat{\phi})}(u)] \right) \\
=& (2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2} \left(f_{\phi|M}(\hat{\phi}) + O(n^{-1/2}) \right), \qquad \text{(A.9)}
\end{aligned}$$

where $E_{H_n}(\cdot)$ is the expectation taken with respect to $u \sim \mathcal{N}(0, H_n(\hat{\phi})^{-1})$. Note that the third equality follows since the replacement of $\sqrt{n}(B^* - \hat{\phi})$ with $\mathbb{R}^d$ incurs an error of exponentially decreasing order and $E_{H_n}(R_{3n}(u))$ is finite, i.e., the multivariate normal distribution has finite moments at any order.

Consider now integrating $p(Y^n|\phi) f_{\phi|M}(\phi)$ over $\Phi_M \setminus B^*$.

$$\begin{aligned}
&\int_{\Phi_M \setminus B^*} p(Y^n|\phi) f_{\phi|M}(\phi) d\phi \\
=& (2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2} \\
&\times \left((2\pi)^{-d/2} n^{d/2} \det(H_n(\hat{\phi}))^{-1/2} \int_{\Phi_M \setminus B^*} \exp\{n(l_n(\phi) - l_n(\hat{\phi}))\} f_{\phi|M}(\phi) d\phi \right) \\
\leq& (2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2} \\
&\times \left((2\pi)^{-d/2} n^{d/2} \det(H_n(\hat{\phi}))^{-1/2} \bar{f}_{\phi|M} \sup_{\phi \in \Phi \setminus B^*} \{\exp\{n(l_n(\phi) - l_n(\phi_{true}))\}\} \right), \qquad \text{(A.10)}
\end{aligned}$$

where by Assumption 3.2 (v), $\bar{f}_{\phi|M} \equiv \sup_{\phi \in \Phi} f_{\phi|M}(\phi) < \infty$. Assumptions 3.2 (iii) and (iv) imply that the term in the parentheses of (A.10) converges to zero faster than $n^{-1/2}$-rate with $P_{Y^\infty|\phi_{true}}$-probability one. Summing up (A.9) and (A.10) gives the following asymptotic approximation of the marginal likelihood in model $M \in \mathcal{M}_A$.

$$\begin{aligned}
p(Y^n|M) &= \int_{B^*} p(Y^n|\phi) f_{\phi|M}(\phi) d\phi + \int_{\Phi_M \setminus B^*} p(Y^n|\phi) f_{\phi|M}(\phi) d\phi \\
&= (2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2} \left(f_{\phi|M}(\hat{\phi}) + O(n^{-1/2}) \right), \qquad \text{(A.11)}
\end{aligned}$$

with $P_{Y^\infty|\phi_{true}}$-probability one. Bringing the multiplicative terms in the right-hand side of (A.11) to the left-hand side completes the proof. ■

Lemma A.3 *Suppose Assumption 3.2 holds. For model $M \in \mathcal{M}_B$,*

$$\frac{n^{d/2} \det(H_n(\hat{\phi}))^{1/2} p(Y^n|M)}{(2\pi)^{d/2} p(Y^n|\hat{\phi})} = o(n^{-1/2}),$$

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

Proof of Lemma A.3. Let B^* be an open neighborhood of ϕ_{true} as defined in the proof of Lemma A.2.

Consider the marginal likelihood of model $M \in \mathcal{M}_B$ divided by $(2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2}$:

$$\begin{aligned}
\frac{n^{d/2}\det(H_n(\hat{\phi}))^{1/2} p(Y^n|M)}{(2\pi)^{d/2} p(Y^n|\hat{\phi})} &= \frac{n^{d/2}\det(H_n(\hat{\phi}))^{1/2}}{(2\pi)^{d/2}} \int_{\Phi_M} \exp\{n(l_n(\phi) - l_n(\hat{\phi}))\} f_{\phi|M}(\phi) d\phi \\
&\leq \frac{n^{d/2}\det(H_n(\hat{\phi}))^{1/2}}{(2\pi)^{d/2}} \bar{f}_{\phi|M} \sup_{\phi\in\Phi_M} \exp\{n(l_n(\phi) - l_n(\hat{\phi}))\} \\
&\leq \frac{n^{d/2}\det(H_n(\hat{\phi}))^{1/2}}{(2\pi)^{d/2}} \bar{f}_{\phi|M} \sup_{\phi\in\Phi\backslash B^*} \exp\{n(l_n(\phi) - l_n(\phi_{true}))\},
\end{aligned} \tag{A.12}$$

where $\bar{f}_{\phi|M} = \sup_\phi f_{\phi|M}(\phi) < \infty$, and the third line follows since $B^* \subset \Phi_M^c$ implies $\Phi_M \subset \Phi\backslash B^*$. Note that by Assumption 3.2 (iv), the upper bound shown in (A.12) converges to zero faster than the polynomial rate of $n^{-1/2}$ with $P_{Y^\infty|\phi_{true}}$-probability one. ■

Proof of Proposition 3.3. (i) Under Assumption 3.2 (i), the posterior model probability of model $M \in \mathcal{M}$ can be written as

$$\pi_{M|Y^n} = \frac{p(Y^n|M)\pi_M}{\sum_{M'\in\mathcal{M}_A} p(Y^n|M')\pi_{M'} + \sum_{M'\in\mathcal{M}_B} p(Y^n|M')\pi_{M'}}$$

By dividing both the numerator and denominator by $(2\pi)^{d/2} p(Y^n|\hat{\phi}) n^{-d/2} \det(H_n(\hat{\phi}))^{1/2}$ and applying Lemmas A.2 and A.3, we have

$$\pi_{M|Y^n} = \begin{cases} \frac{f_{\phi|M}(\hat{\phi})\pi_M}{\sum_{M'\in\mathcal{M}_A} f_{\phi|M'}(\hat{\phi})\pi_{M'}} + O(n^{-1/2}), & \text{for } M \in \mathcal{M}_A, \\ o(n^{-1/2}), & \text{for } M \in \mathcal{M}_B, \end{cases}$$

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

Since $f_{\phi|M}(\cdot)$ is assumed to be continuous and Assumptions 3.2 (iii) and (iv) imply almost sure convergence of $\hat{\phi}$ to ϕ_{true}, $\pi_{M|Y^\infty}$ of the current proposition follows.

(ii) With the given specifications of the ϕ-prior, $f_{\phi|M}(\phi_{true})$ is proportional to $\tilde{\pi}(\Phi_M)^{-1}$ up to the model-independent constant (the Lebesgue density of $\tilde{\pi}_\phi$ evaluated at $\phi = \phi_{true}$). Hence, (i) of the current proposition is reduced to the asymptotic model probabilities of (ii).

(iii) This trivially follows from Lemma 3.1 (iii). ■

A.2 Example 2: Treatment Effect Model with an Instrument

This appendix illustrate applicability of our averaging proposal to the treatment effect model with noncompliance and a binary instrumental variable $Z \in \{0, 1\}$ (Imbens and Angrist (1994)).

Assume that the treatment status and the outcome of interest are both binary. Let $(W_1, W_0) \in \{1,0\}^2$ be the potential treatment status in response to the instrument, and $W = ZW_1 + (1-Z)W_0$ be the observed treatment status. $(Y_1, Y_0) \in \{1,0\}^2$ is a pair of treated and control outcomes and $Y = WY_1 + (1-W)Y_0$ is the observed outcome. Following Imbens and Angrist (1994), consider partitioning the population into four subpopulations defined in terms of the potential treatment-selection responses:

$$T = \begin{cases} c & \text{if } W_1 = 1 \text{ and } W_0 = 0 & : \text{complier}, \\ at & \text{if } W_1 = W_0 = 1 & : \text{always-taker}, \\ nt & \text{if } W_1 = W_0 = 0 & : \text{never-taker}, \\ d & \text{if } W_1 = 0 \text{ and } W_0 = 1 & : \text{defier}, \end{cases}$$

where T is the indicator for the types of selection responses.

Assume that the instrument is randomized in the sense that $Z \perp (Y_1, Y_0, W_1, W_0)$.[28] Then, the distribution of observables and the distribution of potential outcomes satisfy the following equalities for $y \in \{1,0\}$:

$$\begin{aligned} \Pr(Y = y, W = 1|Z = 1) &= \Pr(Y_1 = y, T = c) + \Pr(Y_1 = y, T = at), \\ \Pr(Y = y, W = 1|Z = 0) &= \Pr(Y_1 = y, T = d) + \Pr(Y_1 = y, T = at), \\ \Pr(Y = y, W = 0|Z = 1) &= \Pr(Y_0 = y, T = d) + \Pr(Y_1 = y, T = nt), \\ \Pr(Y = y, W = 0|Z = 0) &= \Pr(Y_0 = y, T = c) + \Pr(Y_1 = y, T = nt). \end{aligned} \tag{A.13}$$

Since the marginal distribution of Z plays no role for identification of the potential outcome distributions, we let the vector of structural parameters θ consist of the parameters that index a joint distribution of (Y_1, Y_0, T):

$$\theta = \left(\Pr(Y_1 = y, Y_0 = y', T = t) : y = 1,0, \quad y' = 1,0, \quad t = c, nt, at, d\right) \in \Theta,$$

where Θ is the probability simplex in $\mathbb{R}^{16}$.

Let the average treatment effect (ATE) be the parameter of interest.

$$\begin{aligned} \alpha \;\equiv\; E(Y_1 - Y_0) &= \sum_{t=c,nt,at,d} [\Pr(Y_1 = 1, T = t) - \Pr(Y_0 = 1, T = t)] \\ &= \sum_{t=c,nt,at,d} \sum_{y=1,0} [\Pr(Y_1 = 1, Y_0 = y, T = t) - \Pr(Y_1 = y, Y_0 = 1, T = t)]. \end{aligned}$$

The reduced-form parameter vector consists of the eight probability masses:

$$\phi = (\Pr(Y = y, W = w|Z = z) : y = 1,0, \quad d = 1,0, \quad z = 1,0).$$

Consider the following two candidate models.

Candidate Models

[28] As reflected in the notation of the potential outcomes (Y_1, Y_0), we assume the exclusion restriction of the instrument.

- *Model M^p (point-identified)*: In addition to the randomized instrument assumption $Z \perp (Y_1, Y_0, W_1, W_0)$, the *instrument monotonicity* (no-defier) assumption of Imbens and Angrist (1994) holds and the causal effects are *homogeneous* in the sense that $E(Y_1 - Y_0|T = c) = E(Y_1 - Y_0|T = at) = E(Y_1 - Y_0|T = nt) = E(Y_1 - Y_0)$.

- *Model M^s (set-identified)*: The randomized instrument assumption holds. Heterogeneity of the treatment effects is unrestricted.

In model M^p, the complier's average treatment effect is identified by the Wald estimand (Imbens and Angrist (1994)), and combined with the homogeneity of the causal effects, we achieve the point-identification of ATE,

$$\alpha_{M^p}(\phi) = \frac{\Pr(Y = 1|Z = 1) - \Pr(Y = 1|Z = 0)}{\Pr(W = 1|Z = 1) - \Pr(W = 1|Z = 0)}.$$

In model M^s, what the Wald estimand identifies is the complier's average treatment effect, while ATE becomes set-identified. See Balke and Pearl (1997) for the construction of the ATE identified set, $IS_\alpha(\phi|M^s)$.

The two models considered admit the identical reduced-form (the distribution of $(Y, W)|Z$), whereas these two models are distinguishable, since they have different testable implications. The testable implication for model M^p is given by the testable implication for the joint restriction of randomized instrument and instrument monotonicity shown by Balke and Pearl (1997):[29]

$$\begin{aligned}
&\Pr(Y = 1, D = 1|Z = 1) \geq \Pr(Y = 1, D = 1|Z = 0),\\
&\Pr(Y = 0, D = 1|Z = 1) \geq \Pr(Y = 0, D = 1|Z = 0),\\
&\Pr(Y = 1, D = 0|Z = 1) \leq \Pr(Y = 1, D = 0|Z = 0),\\
&\Pr(Y = 0, D = 0|Z = 1) \geq \Pr(Y = 0, D = 0|Z = 0).
\end{aligned}$$

Accordingly, Φ_{M^p} is given by the set of ϕ's that satisfy these four inequalities.

Kitagawa (2020) shows that the instrument inequality of Pearl (1995) gives the sharp testable implication for the randomized instrument assumption, i.e., $IS_\alpha(\phi|M^s))$ is empty if and only if

$$\max_w \sum_y \max_z \{\Pr(Y = y, W = w)|Z = z\} \leq 1. \tag{A.14}$$

Hence, the reduced-form parameter space of model M^s, Φ_{M^s}, is obtained as the set of ϕ's that fulfills (A.14).

[29] Under the joint restriction of randomized instrument and instrument monotonicity, additionally imposing homogeneity of the treatment effects does not strengthen the testable implication of Balke and Pearl (1997).

Set prior model probabilities at $(\pi_{M^p}, \pi_{M^s}) = (w, 1-w)$. Construct a prior for ϕ in each model as

$$\pi_{\phi|M^p}(B) = \frac{\tilde{\pi}_\phi(B \cap \Phi_{M^p})}{\tilde{\pi}_\phi(\Phi_{M^p})},$$
$$\pi_{\phi|M^s}(B) = \frac{\tilde{\pi}_\phi(B \cap \Phi_{M^s})}{\tilde{\pi}_\phi(\Phi_{M^s})}.$$

for any measurable subset B in the probability simplex that ϕ lies, where $\tilde{\pi}_\phi$ is a prior for ϕ such as a Dirichlet distribution.

The two models M^p and M^s are distinguishable since Φ_{M^p} is a proper subset of Φ_{M^s}. With the current construction of the priors for ϕ, Lemma 3.1 (ii) gives their posterior model probabilities,

$$\pi_{M^p|Y} = \frac{O_{M^p} \cdot w}{O_{M^p} \cdot w + O_{M^s} \cdot (1-w)},$$
$$\pi_{M^s|Y} = \frac{O_{M^s} \cdot (1-w)}{O_{M^p} \cdot w + O_{M^s} \cdot (1-w)},$$

where O_{M^p} and O_{M^s} are the posterior-prior plausibility ratio as defined in Lemma 3.1.

With these posterior model probabilities, the robust Bayes averaging operates as presented in Scenario 1 of Example 1. The resulting set of posterior means shrinks the Balke and Pearl's ATE identified set toward the posterior mean of the Wald estimand that one would report in the point-identified model. Since the posterior model probabilities can differ from the prior ones, the degree of shrinkage can reflect how well the identifying assumptions fit the data. The current analysis offers one way to aggregate the Wald instrumental variable estimator and the ATE bounds with exploiting a partially credible assumption on homogeneity of the causal effects.

A.3 Computing Plausibility Ratios for Sign-restricted SVARs

This appendix provides details on how to compute the posterior-prior plausibility ratios O_M for the SVAR models set-identified by the under-identifying zero restrictions and the sign restriction. The crucial step in the computation is to check if the identified set $IS_\alpha(\phi)$ is empty or not at ϕ drawn from $\tilde{\pi}_{\phi|Y}$. The first proposal (Algorithm A.1), which is a special case of Algorithm 1 in Giacomini and Kitagawa (2020), makes use of random draws of the impulse responses and assesses whether any of these satisfies the imposed sign restrictions. The second proposal (Algorithm A.2) is novel in the literature. It directly checks a necessary and sufficient condition for non-emptiness of the identified set. The first algorithm is simple to implement, while it can give a wrong conclusion if the identified set is tiny. The second algorithm is guaranteed to give the right answer, while the computation can become cumbersome if the number of sign restrictions is large.

A.3.1 Notation

To show the computation procedures in a general setting, we generalize the representations of the SVAR (5.1) and the reduced-form VAR (5.3) to have n endogenous variables and $p \geq 0$ lags. Let $Q \in \mathcal{O}(n)$ be an $n \times n$ orthonormal matrix and $\mathcal{O}(n)$ be the set of $n \times n$ orthonormal matrices. Following Uhlig (2005) and Rubio-Ramirez et al. (2010), we transform the structural parameters $(A_0, a, A_1, \ldots, A_p)$ into the parameter vector consisting of the reduced-form parameters augmented by Q, $(\phi', vec(Q)')' \in \tilde{\Phi} \times vec(\mathcal{O}(n))$:

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

where Σ_{tr} denotes the lower-triangular Cholesky factor of Σ with nonnegative diagonal elements. We then set $\theta = (\phi', vec(Q)')'$ and its domain as $\Theta = \left\{(\phi', vec(Q)')' \in \Phi \times vec(\mathcal{O}(n)) : diag\left(Q'\Sigma_{tr}^{-1}\right) \geq 0\right\}$. Here, $diag\left(Q'\Sigma_{tr}^{-1}\right) \geq 0$ is the sign normalization restrictions:

$$
\left(\sigma^i\right)' q_i \geq 0 \quad \text{for all } i = 1, \ldots, n, \tag{A.15}
$$

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

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

$$
\begin{aligned}
y_t &= c + \sum_{j=0}^{\infty} C_j u_{t-j} \qquad \text{(A.16)} \\
&= c + \sum_{j=0}^{\infty} C_j \Sigma_{tr} Q \epsilon_{t-j},
\end{aligned}
$$

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

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

$$
IR^h = C_h \Sigma_{tr} Q. \tag{A.17}
$$

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

$$
\alpha = IR^h_{ij} \equiv e_i' C_h \Sigma_{tr} Q e_j \equiv c_{ih}'(\phi)\, q_j, \tag{A.18}
$$

where e_i is the i-th column vector of the identity matrix I_n and $c_{ih}'(\phi)$ is the i-th row vector of $C_h \Sigma_{tr}$.

Zero restrictions used in the literature are restrictions on some off-diagonal elements of A_0, on the lagged coefficients $\{A_l : l = 1, \ldots p\}$, on contemporaneous impulse responses $IR^0 = A_0^{-1}$, and on the cumulative long-run responses. All these restrictions can be viewed as linear constraints on the columns of Q. For example:

$$\begin{aligned} ((j,i)\text{-th element of } A_0) &= 0 \iff \left(\Sigma_{tr}^{-1} e_i\right)' q_j = 0, \\ ((j,i)\text{-th element of } A_l) &= 0 \iff \left(\Sigma_{tr}^{-1} B_l e_i\right)' q_j = 0, \\ \left((i,j)\text{-th element of } A_0^{-1}\right) &= 0 \iff \left(e_i' \Sigma_{tr}\right) q_j = 0, \\ \left((i,j)\text{-th element of } IR^h\right) &= 0 \iff \left[e_i' C_h \Sigma_{tr}\right] q_j = 0. \end{aligned} \tag{A.19}$$

We restrict our analysis to the case that the imposed zero restrictions constrain only one column vector of Q. Ordering the variables in such way that q_1 becomes the constrained column vector of Q, we can represent a collection of zero restrictions as

$$F(\phi)\, q_1 = \mathbf{0}, \tag{A.20}$$

where $F(\phi)$ is an $f \times n$ matrix. $F(\phi)$ stacks all the coefficient vectors that multiply q_1 into a matrix. Hence, f is the number of imposed zero restrictions. We consider under-identifying zero restrictions, so we assume $f \leq n - 2$.

We suppose there are sign restrictions on the responses to the first structural shock. Sign restrictions are linear constraints on the first column of Q: $S_h(\phi)\, q_1 \geq \mathbf{0}$, where $S_h(\phi) \equiv D_h C_h \Sigma_{tr}$ is an $s_h \times n$ matrix, and D_h is an $s_h \times n$ matrix that selects the sign-restricted responses from the $n \times 1$ impulse-response vector $C_h \Sigma_{tr} q_1$. The nonzero elements of D_h equal 1 or -1 depending on whether the corresponding impulse responses are positive or negative.

Stacking $S_h(\phi)$ over multiple horizons gives the set of sign restrictions

$$S(\phi)\, q_1 \geq \mathbf{0}, \tag{A.21}$$

where $S(\phi)$ is a $s \times n$ matrix $S(\phi) = \left[S_0(\phi)', \ldots, S_{\bar{h}}(\phi)\right]'$, where $s = \sum_{h=0}^{\bar{h}} s_h$ is the number of sign constraints and $0 \leq \bar{h} \leq \infty$ is the maximal horizon in the impulse-response analysis.[30]

A.3.2 Algorithms

For multiple posterior models, the plausibility ratio O_M can be computed by plugging into (3.5) numerical approximations of the prior and posterior probabilities for non-emptiness of the identified set. Specifically, the denominator of O_M can be computed by drawing many ϕ's from the prior and finding the fraction of draws that yield a non-empty identified set. The numerator of O_M can be computed similarly except that the ϕ's are drawn from the posterior.

[30] If there are no sign restrictions on the $\tilde{h}$-th horizon responses, $\tilde{h} \in \{0, \ldots, \bar{h}\}$, $s_{\tilde{h}} = 0$ and $S_{\tilde{h}}(\phi)$ is not present in $S(\phi)$.

Our first algorithm to approximate O_M draws many q_1's from a distribution supported only on the unit sphere, and check if any of the draws satisfies the model's assumptions given ϕ.

Algorithm A.1 *Suppose the identifying restrictions of model M consist of zero and sign restrictions as defined in (A.20) and (A.21), respectively. The following algorithm can be used to approximate $\tilde{p}_\phi(\Phi_M)$, where $\tilde{p}_\phi$ is a probability measure on $\tilde{\Phi}$, which can be $\tilde{\pi}_\phi$ or $\tilde{\pi}_{\phi|Y}$.*

1. *Draw ϕ from $\tilde{p}_\phi$.*

2. *Let $z \sim \mathcal{N}(\mathbf{0}, I_n)$ be a draw of an n-variate standard normal random variable. Let $\tilde{q}_1 = \mathcal{M}z$ be the $n \times 1$ residual vector in the linear projection of z onto the $n \times f$ regressor matrix $F(\phi)'$. Set $q_1 = sign\left(\left(\sigma^1\right)' \tilde{q}_1\right) \frac{\tilde{q}_1}{\|\tilde{q}_1\|}$. If $\left(\sigma^i\right)' \tilde{q}_i$ is zero for some i, set $sign\left(\left(\sigma^i\right)' \tilde{q}_i\right)$ equal to 1 or -1 with equal probability.*

3. *Check if q_1 satisfies the sign restrictions $S(\phi)q_1 \geq 0$. If it does, we conclude $IS_\alpha(\phi) \neq \emptyset$. Otherwise, repeat Step 2 a maximum of L times until q_1 satisfying $S(\phi)q_1 \geq \mathbf{0}$ is obtained. If none of the L draws of q_1 satisfies $S(\phi, Q) \geq \mathbf{0}$, approximate $IS_\alpha(\phi)$ as being empty and return to Step 1 to obtain a new draw of ϕ.*

4. *Repeat Steps 1 – 3 for K times. The proportion of drawn ϕ's that gives non-empty $IS_\alpha(\phi)$ in Step 3 approximates $\tilde{p}_\phi(\Phi_M)$.*

This procedure is simple to implement and easy to scale up to the situation where the number of sign restrictions is large. On the other hand, it only delivers an approximate assessment of the identified set non-emptiness. The approximation quality can become poor if the set of q's satisfying the sign restrictions is so thin that the finite number of q_1 draws may miss it. Also, as the dimension of variables gets larger, the dimension of q_1 increases and we require more draws of q_1.

To overcome this drawback of Algorithm A.1, the next algorithm exploits the linear structure of the identifying assumptions. A key observation is that any non-empty identified set for q_1 contains a vertex on the unit sphere on which at least $n-1$ number of equality and inequality constraints are binding. We can exhaust all the possible candidates for such vertex by selecting any combination of $n-1$ constraints and setting them binding. If we could find a vertex that satisfies the other $f+s-(n-1)$ constraints ruled out in the selection, we can claim this vertex is contained in the identified set for q_1, allowing us to conclude it is non-empty. If instead we cannot find any such vertex, we conclude the identified set is empty. This alternative approach assesses emptiness of the identified set without any approximation. We find that this procedure tends to be faster and become more efficient than Algorithm A.1 when the number of sign restrictions is small to moderate.

Algorithm A.2 *Suppose the identifying restrictions of model M consist of zero and sign restrictions as defined in (A.20) and (A.21), respectively. The following algorithm can be used to approximate $\tilde{p}_\phi(\Phi_M)$, where $\tilde{p}_\phi$ is a probability measure on $\tilde{\Phi}$, which can be $\tilde{\pi}_\phi$ or $\tilde{\pi}_{\phi|Y}$.*

1. *Draw ϕ from $\tilde{p}_\phi$.*

2. *Find q_1^* and $-q_1^*$ satisfying the system of "active constraints" (in the language of Gafarov et al. (2018)):*

$$\begin{cases} F(\phi)q = 0 \\ \tilde{S}(\phi)q = 0 \end{cases} \tag{A.22}$$

   where $\tilde{S}(\phi)$ is $\tilde{s} \times n$ matrix of active sign restrictions. It is set by picking $\tilde{s}$ rows from $S(\phi)$ matrix, where $f + \tilde{s} = n - 1$. Check if q_1^* or $-q_1^*$ satisfy the "inactive constraints," namely the rest of sign restrictions and the sign normalization restriction for q_1. If so, $IS_\alpha(\phi)$ is non-empty. Otherwise, keep constructing $\tilde{S}(\phi)$ with different combinations of $\tilde{s}$ active constraints and verify if the corresponding solution satisfy the inactive constraints. If none of the solutions satisfies the inactive restrictions, $IS_\alpha(\phi)$ is empty.

3. *Repeat Step 1 – 2 K times.*

4. *Approximate $\tilde{p}_\phi(\Phi_M)$ by the proportion of K draws of ϕ that delivers non-empty identified set in Step 2.*

In contrast to Algorithm A.1, this algorithm is advantageous in characterizing the non-emptiness of $IS_\alpha(\phi)$ with no error. As the number of sign restrictions increases, however, the number of combinations of the active constraints to be checked in Step 2 grows in the polynomial order of $n - 1 - f$, and it can be computational burdensome if a large number of sign restrictions is imposed. See Footnote 23 in the main text for the comparison of the computational time between the two algorithms in the empirical application.

References

Ahmadi, P. A. and H. Uhlig (2015): "Sign Restrictions in Bayesian FAVARs with an Application to Monetary Policy Shocks," *National Bureau of Economic Research.*

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.

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

Antolin-Diaz, J., I. Petrella, and J. F. Rubio-Ramirez (2020): "Structural Scenario Analysis with SVARs," *Journal of Monetary Economics.*

Aruoba, B. and F. Schorfheide (2011): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, 3, 60–90.

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 60, 47–57.

Bajari, P., H. Hong, and S. P. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78, 1529–1568.

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

Bates, J. and C. Granger (1969): "The Combination of Forecasts," *Operational Research Quarterly*, 20, 451–468.

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

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

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

Bernanke, B. (1986): "Alternative Explorations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99.

Blanchard, O. and D. Quah (1993): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 83, 655–673.

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

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

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

Chipman, H. (1996): "Bayesian variable selection with related predictors," *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 17–36.

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.

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*, ed. by J. B. Taylor and M. Woodford, Elsevier, vol. 1, Part A, 65–148.

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

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

Clyde, M. (1999): "Bayesian model averaging and model search strategies," *Bayesian statistics*, 6, 157.

Clyde, M. and E. I. George (2004): "Model uncertainty," *Statistical science*, 81–94.

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

Dedola, L. and S. Neri (2007): "What Does a Technology Shock Do? A VAR Analysis with Model-Based Sign Restrictions," *Journal of Monetary Economics*, 54, 512–549.

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

Drèze, J. and J.-F. Richard (1983): "Bayesian Analysis of Simultaneous Equation Models," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, North-Holland, vol. 1, Chapter 9, 517–598.

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

Fujita, S. (2011): "Dynamics of Worker Flows and Vacancies: Evidence from the Sign Restriction Approach," *Journal of Applied Econometrics*, 26, 89–121.

Furlanetto, F., F. Ravazzolo, and S. Sarferaz (2019): "Identification of Financial Factors in Economic Fluctuations," *The Economic Journal*, 129, 311–337.

Gafarov, B., M. Meier, and J.-L. Montiel-Olea (2016): "Projection Inferene for Set-identified SVARs," *unpublished manuscript.*

Gafarov, B., M. Meier, and J. L. M. Olea (2018): "Delta-Method inference for a class of set-identified SVARs," *Journal of Econometrics, forthcoming.*

George, E. (1999): "Discussion of "Model averaging and model search strategies" by M. Clyde," in *Bayesian Statistics 6–Proceedings of the Sixth Valencia International Meeting.*

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

Giacomini, R. and T. Kitagawa (2020): "Robust Bayesian Inference for Set-identified Models," *cemmap working paper, University College London.*

Good, I. J. (1965): *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, The MIT Press.

Hansen, B. E. (2007): "Least Squares Model Averaging," *Econometrica*, 75, 1175–1189.

——— (2014): "Model Averaging, Asymptotic Risk, and Regressor Groups," *Quantitative Economics*, 5, 495–530.

Hansen, B. E. and J. S. Racine (2012): "Jackknife Model Averaging," *Journal of Econometrics*, 167, 38–46.

Hjort, N. L. and G. Claeskens (2003): "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 98, 879–899.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–417.

Horowitz, J. L. and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

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

Imbens, G. and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

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*, ed. by S. Geisser, J. Hodges, S. Press, and A. Zellner, Elsevier, North-Holland, 473–488.

Kilian, L. (2013): "Structural Vector Autoregressions," in *Handbook of Research Methods and Applications in Empirical Macroeconomics*, ed. by M. Hashimzade, N. & Thornton, Edward Elgar, Cheltenham, UK, 515–554.

Kilian, L. and D. Murphy (2012): "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association*, 10, 1166–1188.

Kitagawa, T. (2020): "The Identification Region of the Potential Outcome Distributions under Instrument Independence," *cemmap working paper, University College London.*

Kitagawa, T. and C. Muris (2016): "Model Averaging in Semiparametric Estimation of Treatment Effects," *Journal of Econometrics*, 193, 271–289.

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

Korobilis, D. (2020): "Sign Restrictions in High-Dimensional Vector Autoregressions," *Available at SSRN.*

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

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

Liu, C.-A. (2015): "Distribution Theory of the Least Squares Model Averaging Estimator," *Journal of Econometrics*, 186, 142–159.

Liu, Q. and R. Okui (2013): "Heteroskedasticity-robust C_p Model Averaging," *Econometrics Journal*, 16, 463–472.

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.

Manski, C. F. (1989): "Anatomy of the Selection Problems," *Journal of Human Resources*, 24, 343–360.

Manski, C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

Masten, M. A. and A. Poirier (2020): "Inference on Breakdown Frontiers," *Quantitative Economics*, 11, 41–111.

Matthes, C. and F. Schwartzman (2019): "What Do Sectoral Dynamics Tell Us About the Origins of Business Cycles?" *FRB Working Paper*, 19.

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.

Moon, H., F. Schorfheide, and E. Granziera (2013): "Inference for VARs Identified with Sign Restrictions," *unpublished manuscript.*

Mountford, A. (2005): "Leaning into the Wind: A Structural VAR Investigation of UK Monetary Policy," *Oxford Bulletin of Economics and Statistics*, 67, 597–621.

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

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

Pearl, J. (1995): "On the Testability of Causal Models with Latent and Instrumental Variables," *Uncertainty in Artificial Intelligence*, 11, 435–443.

Peersman, G. and R. Straub (2009): "Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR," *International Economic Review*, 50, 727–750.

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

Rafiq, S. and S. Mallick (2008): "The Effect of Monetary Policy on Output in EMU3: A Sign Restriction Approach," *Journal of Macroeconomics*, 30, 1756–1791.

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

Rosenbaum, P. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies," *Biometrika*, 70, 41–55.

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

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

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

Scholl, A. and H. Uhlig (2008): "New Evidence on the Puzzles: Results from Agnostic Identification on Monetary Policy and Exchange Rates," *Journal of International Economics*, 76, 1–13.

Sims, C. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

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

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

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.

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.

Vargas-Silva, C. (2008): "Monetary Policy and the US Housing Market: A VAR Analysis Imposing Sign Restrictions," *Journal of Macroeconomics*, 30, 977–990.

Zhang, X. and H. Liang (2011): "Focused Information Criterion and Model Averaging for Generalized Additive Partial Linear Models," *Annals of Statistics*, 39, 174–200.