Uncertain Identification*

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

This draft: January 2017

Abstract

Uncertainty about the choice of identifying assumptions is common in causal studies, but has been often ignored in empirical practice. This paper considers uncertainty over a class of models that impose different sets of identifying assumptions, which, in general, leads to a mix of point- and set-identified models. We propose a method for performing inference in the presence of this type of uncertainty by generalizing Bayesian model averaging. Our proposal is to consider ambiguous belief (multiple posteriors) for the set-identified models, and to combine them with a single posterior in a model that is either point-identified or that imposes non-dogmatic identifying assumptions in the form of a Bayesian prior. 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 mixture weights the posterior probabilities of the models. We propose to summarize the post-averaging ambiguous belief by reporting the range of posterior means and the associated credible regions, and offer a simple algorithm to compute these quantities. We establish conditions under which the data are informative about model probabilities, which occurs when the models are distinguishable for some distribution of data and/or specify different priors for reduced-form parameters, and examine the asymptotic behavior of the posterior model probabilities. The method is general and allows for dogmatic and non-dogmatic identifying assumptions, multiple point-identified models, multiple set-identified models, and nested or non-nested models.

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

*We would like to thank Frank Kleibergen, José Montiel-Olea, John Pepper, James Stock, and seminar participants at University of Bonn, University of North Carolina, and University of Virginia for valuable comments. We also thank participants at IAAE 2016 (Milan), RES 2016 (Brighton), AMES 2016 (Kyoto), MMF 2016 (Bath), ERFIN 2016 (Warsaw) and the workshop of "New Identification Approaches to Macroeconometrics" (Oxford) for beneficial discussions. Financial support from 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. Email: t.kitagawa@ucl.ac.uk

[§]Queen Mary University of London, Department of Economics. Email: a.volpicella@qmul.ac.uk

1 Introduction

The choice of identifying assumptions is the crucial step that allows empirical researchers to draw causal inferences using observational data. This is a controversial choice in empirical practice, and the researcher often faces uncertainty about which identifying assumptions to impose from a menu of plausible ones. This uncertainty and its effects on inference have been typically ignored in empirical work until now. We propose a formal method for drawing inferences about causal effects in the presence of uncertainty about identifying assumptions, which we characterize as uncertainty over a class of models that are characterized by different sets of assumptions. The method can be viewed as a generalization of Bayesian model averaging to include set-identified models, which arise in the common case when the assumptions are under-identifying or when they are expressed as inequality restrictions.

There are several important examples of this set-up within economics. The first is macroeconomic policy analysis using structural vector autoregressions (SVARs), where identifying assumptions include causal ordering restrictions (Bernanke (1986) and Sims (1980)), long-run neutrality restrictions (Blanchard and Quah (1993)), and Bayesian prior mean restrictions implied by a structural model (Del Negro and Schorfheide (2004)). Subsets of these assumptions will deliver set-identified impulse-responses, as will the widespread use of sign restrictions (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)). The second example is microeconometric causal effect studies with identifying assumptions including selection on observables (Ashenfelter (1978) and Rosenbaum and Rubin (1983)), selection on observables and unobservables (Altonji, Elder, and Taber (2005)), exclusion and monotonicity restrictions in instrumental variables methods (Imbens and Angrist (1994), yielding set-identification of the average treatment effect), and monotone instrument assumptions (Manski and Pepper (2000), also yielding set-identification). The third example is missing data with identifying 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-identified models (Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2011), and Ciliberto and Tamer (2009), to list a few).

How should one draw inference for the object of interest in the presence of uncertainty over point- and set-identified models? The common practice in empirical work is to report the estimation result from what is deemed the most credible set of identifying assumptions, or, in some cases, the results from a number of alternative sets of assumptions, viewed as an informal way to conduct sensitivity analysis. This paper proposes a method that formalizes the sensitivity analysis and provides a way to aggregate estimation results relying on point-identifying assumptions and those relying only on set-identifying assumptions by generalizing

Bayesian model averaging.

Standard Bayesian model averaging delivers a single posterior distribution that is a mixture of the posterior distributions of the candidate models with weights equal to the posterior model probabilities. This approach could in principle be extended to our context if one could obtain a single posterior distribution for every candidate model, including those that are set-identified. Obtaining a single posterior in the set-identified model is however a challenging task, as has been recently pointed out in the context of SVARs with sign restrictions (Baumeister and Hamilton (2015) and Giacomini and Kitagawa (2015)): when the available prior knowledge is exhausted by an under-identifying set of assumptions, choosing a single prior for the non-identified parameters (even an apparently uninformative one) may result in spuriously informative posterior inference for the parameters of interest, an effect that persists even in large samples. This problem does not occur in point-identified models, where the effect of prior choice vanishes asymptotically. Choosing a single prior in set-identified models that can precisely represent the lack of knowledge for the non-identified part of the models is generally infeasible.[1]

The approach in this paper, in contrast, does not assume availability of a single posterior for the set-identified model. The key innovation is to introduce multiple priors (*an ambiguous belief*) into Bayesian model averaging by representing the posterior information in the set-identified model by a set of posteriors. We then combine the single posterior probabilistic belief obtained from the point-identified models with the ambiguous belief obtained from the set-identified models. The output of the procedure is a set of posteriors (*post-averaging ambiguous belief*), that consists of mixtures of the posterior distribution in the point-identified model and any posterior belonging to the set of posteriors of the set-identified model, with weights equal to the posterior model probabilities. To summarize and visualize the post-averaging ambiguous belief, we recommend to report the range of posterior quantities such as the mean or median, and associated credible region (an interval on which any posterior in the class allocates a certain credibility level). We show that these quantities have analytically simple expressions and they are easy to compute in practice.

Our motivation for introducing multiple priors in set-identified models builds on the robust Bayes interpretation of multiple priors as a way to express the "lack of knowledge". As argued by Giacomini and Kitagawa (2015) and Kitagawa (2012), allowing for ambiguity over the unrevisable component of the prior knowledge (i.e., the prior for the non-identified parameters) and conducting posterior inference based on the resulting class of posteriors offers a rationale

[1]The Bayesian analysis for set-identified SVARs commonly specifies the rotation invariant prior for non-identified parameters (the uniform prior on a Stiefel manifold) as a representation of "noninformative" prior (Uhlig (2005) and Arias, Rubio-Ramirez, and Waggoner (2013)). It is important to acknowledge that such noninformative prior assigns unique weights over the non-identified parameters and it fails to accommodate the researcher's inability to specify a belief for them.

for focusing on the posterior distribution of the identified set, as proposed in Kline and Tamer (2016), Moon and Schorfheide (2011), and Liao and Simoni (2013). Conditional on a set-identified model, the state of knowledge we assume is that one cannot assess what parameter values are more credible than the others within the identified set, and such ambiguity within the identified set is represented by the resulting set of posteriors for the parameter of interest. Conditional on a point-identified model, on the other hand, the state of knowledge we assume is that the researcher can specify a prior distribution for the reduced-form parameters or for the structural parameters subject to the point-identifying assumptions, so that she can summarize the posterior information for the parameter of interest by the single posterior distribution. Given prior model weights specified by the user, our averaging procedure aggregates these distinct forms of posterior information for the common parameter of interest in a way coherent to the robust Bayes prior-by-prior updating rule.

The method proposed in this paper can also be viewed as bridging the gap between point- and set-identification. When focusing solely on the estimates from a point-identified model, the researcher who is not fully confident about her choice of identifying assumptions may doubt the robustness of the obtained conclusions. At the same time, discarding some of the point-identifying assumptions entirely and reporting the estimate of the identified set under the weaker restrictions may appear "excessively agnostic" and can result in overwhelmingly uninformative conclusions. Our averaging procedure reconciles these two extreme representations of the posterior beliefs by exploiting the prior weights that the researcher can assign to alternative sets of identifying assumptions. We show that the range of posterior means for the post-averaging ambiguous belief is given by the weighted average (in the sense of Minkowski sum) of the posterior mean of the point-identified model and the range of posterior means (interval) in the set-identified model. When the underlying identified set is a connected interval, the range of posterior means can be viewed as an estimate of the identified set (Giacomini and Kitagawa (2015)). Hence, our averaging procedure can effectively shrink the identified set estimate toward the point estimate in the point-identified model. The degree of shrinkage is governed by the posterior model probabilities. A key result in the paper is to clarify under which conditions the prior model probabilities can be updated by data. We show that the updating occurs if some candidate models are distinguishable for some distribution of data and/or the priors for the reduced-form parameters are different across the models. We also perform the asymptotic analysis for the posterior model probabilities. We show that in the situation where only one candidate model is consistent with the true distribution of data, the posterior model probability asymptotically assigns probability one to the correctly specified model. In situations where multiple candidate models are observationally equivalent and non-falsified at the true data generating process, we show that the posterior model probabilities asymptotically assign nontrivial weights among them. We clarify what part of the prior input determines the asymptotic posterior model probabilities in such case. The large sample consistency property of the Bayesian model selection has been well-studied in the statistics literature (see, e.g., Claeskens and Hjort (2008) and references therein). However, little is known about the asymptotic behaviors of the posterior model probabilities when the candidate models, which differ in terms of the identifying restrictions, can be observationally equivalent in terms of the reduced form model. Our asymptotic analysis of the posterior model probabilities therefore adds new contributions to the literature of the Bayesian model selection and they could be of independent interest.

The method proposed in this paper contributes to the empirical practice both in macroeconomics and microeconomics by offering simple and flexible ways to conduct sensitivity analysis of causal inference to the choice of identifying assumptions. First, when the set-identified model nests the point-identified model, it can be used to assess posterior sensitivity in the point-identified model with respect to arbitrary perturbations of the prior input in the direction of relaxing some of the point-identifying assumptions, with the maximal magnitude of the perturbation is specified by the prior model probability assigned to the set-identified model. From this viewpoint, we can formally interpret our averaging method as an ϵ-contamination sensitivity analysis developed in Berger and Berliner (1986) with a particular construction of the prior class. Second, if the point-identified model can be considered as a reasonable benchmark model, 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 and disciplined manner.

The remainder of the paper is organized as follows. Section 2 illustrates the main results of this paper and the implementation of our averaging method in the context of a simple two-variable SVAR. Section 3 presents formal analysis in a general framework and provides a computational algorithm to implement the procedure. Section 4 discusses the relationships between our method and existing (robust) Bayesian methods, and elicitation of model probabilities. In Section 5, we apply our method to impulse response analysis in SVARs. The Appendix contains proofs omitted from the main text and one microeconometric application.

1.1 Related Literature

The idea of model averaging has a long history in econometrics and statistics since the pioneering works by Bates and Granger (1969) and Leamer (1978). There are mainly two approaches in the literature. One is Bayesian model averaging (see e.g., Hoeting, Madigan, Raftery, and Volinsky (1999) and Claeskens and Hjort (2008) for a review of the literature). Another is frequentist model averaging, with important developments made by Hansen (2007) and Hjort and Claeskens (2003). See also Hansen (2014), Liu (2015), Liu and Okui (2013), Hansen and Racine (2012), and Zhang and Liang (2011) for recent advances in frequentist model averaging, and Hjort and Claeskens (2003), Kitagawa and Muris (2016), and Magnus, Powell, and Prüfer (2010) for proposals that compromise between Bayesian and frequentist averaging. None of these works considers averaging point- and set-identified models. We tackle this problem from the angle of Bayesian model averaging since it is not obvious how to approach it from the frequentist standpoint. Unlike the standard Bayesian model averaging, however, our averaging introduces ambiguity and does not require the full prior specification in the set-identified model. To the best of our knowledge, this is the first paper that formally considers aggregating of probabilistic beliefs and ambiguous beliefs in the context of model averaging.

This paper also contributes to the growing literature on Bayesian inference for partially identified models (Giacomini and Kitagawa (2015), Kitagawa (2012), Kline and Tamer (2016), Moon and Schorfheide (2012), Norets and Tang (2014), Liao and Simoni (2013)). This paper follows the robust Bayes approach with multiple priors to model the lack of belief within the identified set (Giacomini and Kitagawa (2015) and Kitagawa (2012)). When the set-identified model is the only model considered, the range of posteriors generated by the robust Bayes formulation 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 uncertainty on the identifying assumptions is present, however, the usual definition of the identified set is not available without conditioning on the model. The multiple prior viewpoint enjoys an advantage in this case since the range of posteriors still has a well-defined subjective interpretation even in the presence of model uncertainty.

The empirical application in this paper concerns SVAR analysis with classical causal ordering restrictions (Bernanke (1986) and Sims (1980)), dynamic stochastic general equilibrium (DSGE) model restrictions (Del Negro and Schorfheide (2004)), and sign restrictions (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)). In many SVAR applications, what set of identifying assumptions to impose becomes a source of controversy given that the researchers have heterogeneous opinions about the credibility of the identifying assumptions. One popular empirical practice is to impose weaker assumptions in the form of sign restrictions on the impulse responses or on the structural parameters. Although the resulting model generally only set-identifies the impulse responses of interest, the common practice is to condier Bayesian estimation by adding a "noninformative" prior on the non-identified part of the model. Empirical studies using this Bayesian approach include 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. As alternative methods, Moon, Schorfheide, and Granziera (2013) and Gafarov, Meier, and Montiel-Olea (2016a,b) develop frequentist inference for the identified set and Giacomini and

Kitagawa (2015) proposes 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 prove helpful in reconciling the controversies about the identifying assumptions that are widespread in this literature.

2 Illustrative Example

We begin with a simple example that illustrates the analytical framework, motivations, and the implementation of our proposal.[2]

Consider a model of static labor supply and demand:

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

where $(\Delta n_t, \Delta w_t)$, $t = 1, \ldots, T$, are, respectively, the growth rates of employment and wages and $(\epsilon_t^d, \epsilon_t^s)$ is a vector of demand and supply shocks that are independent of the past realizations of $(\epsilon_t^d, \epsilon_t^s)$. We assume that the structural shocks are normally distributed with covariance equal to the identity matrix, so that the impulse responses are with respect to a unit standard deviation shock. A is the matrix of structural coefficients and the contemporaneous impulse response matrix is A^{-1}.

The reduced-form model is indexed by Σ the variance-covariance matrix of $(\Delta n_t, \Delta w_t)$, which is determined by A via $\Sigma = A^{-1}(A^{-1})'$. We denote its lower triangular Cholesky decomposition 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 parameters by $\phi = (\sigma_{11}, \sigma_{21}, \sigma_{22}) \in \Phi = \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.[3] We denote the map from the structural parameters to the reduced-form parameters by $\phi = g(A)$.

Let the response of the first variable to a unit positive shock in the first variable be the object of interest, $\alpha \equiv$ (1,1)-element of A^{-1}. In the absence of identifying assumptions, the structural parameters are not identified, i.e., knowledge of the reduced-form parameters ϕ cannot uniquely pin down the structural parameters, since $\phi = g(A)$ is a many-to-one mapping.

We motivate our proposal and illustrate its implementation in two distinct scenarios.

2.1 Dogmatic Identifying Assumptions

Suppose that the researcher is uncertain about the following two sets of identifying assumptions, which are dogmatic, in the sense that they are imposed as exact equalities or inequality

[2] The scope of application of our averaging method includes not only macroeconometric applications but also microeconometric applications. In Appendix A.2, we provide an application of our method to a treatment effect model with noncompliance.

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

restrictions on (functions of) the structural parameters.

Scenario 1: Candidate Models

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

Model M^p restricts A to be lower-triangular and it leads to the classical causal ordering assumption of Sims (1980) and Bernanke (1986). Combined with the sign normalization restrictions such that the diagonal elements of A are nonnegative, the contemporaneous impulse responses can be identified by $A^{-1} = \Sigma_{tr}$. We can accordingly express the point-identified α as $\alpha = \alpha_{M^p}(\phi) \equiv \sigma_{11}$.

In contrast, model M^s uses only sign restrictions and it can only set-identify α. Appendix A shows that the identified set for α, which is viewed as a set-valued map from Φ to $\mathbb{R}$, is a connected interval given by:

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

Note that this identified set is non-empty for any ϕ. Hence, model M^p and model M^s are observationally equivalent at any $\phi \in \Phi$ and neither of them is falsifiable, i.e., for any $\phi \in \Phi$ and in both model M^p and M^s, there exist structural parameters A that satisfy the imposed identifying assumptions. When $\sigma_{21} > 0$, the point-identified α in model M^p coincides with the upper-bound of the identified set in model M^s. On the other hand, when $\sigma_{21} < 0$, the identified set in model M^s does not contain the point-identified α.[4]

Suppose that the researcher's prior uncertainty over the two models can be represented by a weight assigned to model M^p, $w \in [0, 1]$, and its complement $(1 - w)$ assigned to model M^s.[5]

Consider specifying a prior distribution for the *reduced-form* parameters in each model. Note that this prior is updated by the data, and conditional on the model, such a choice does not affect the conclusion about the parameter of interest asymptotically.[6] Bearing in mind the observational equivalence of the two models at every $\phi \in \Phi$, it could be reasonable to specify the same prior for ϕ between the two models:

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

[4]This is because in model M^p, the parameter a_{21} is given by $-\frac{\sigma_{21}}{\sigma_{11}\sigma_{22}}$, which is positive if $\sigma_{21} < 0$. That is, the point-identifying assumptions $a_{12} = 0$ and $\sigma_{21} < 0$ are not compatible with the upward sloping supply restriction $a_{21} \leq 0$.

[5]We discuss interpretation and elicitation of the prior model probabilities in Section 4.2 below.

[6]As we show in Section 3.5 below, the choice of a prior for reduced-form parameters in each model can influence the posterior model probabilities even asymptotically.

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

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

where Y denotes the sample.

In model M^p, the posterior for ϕ implies a unique posterior for α via the mapping $\alpha = \alpha_{M^p}(\phi)$. We denote the posterior of α in this model by $\pi_{\alpha|M^p,Y}$.

On the other hand, in model M^s, having one prior for ϕ does not yield a unique posterior for α, since the mapping given in (2.2) is generally set-valued. Following Giacomini and Kitagawa (2015) and Kitagawa (2012), we formulate the lack of prior knowledge other than the sign restrictions by multiple priors (an ambiguous belief). Formally, with $\pi_{\phi|M^s}$ given, 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 imposed identifying restrictions (including the sign normalization restrictions):

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

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, $g^{-1}(\phi)$ is the set of observationally equivalent A's given reduced-form parameters ϕ.

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

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

We then form the class of posteriors for α by marginalizing the posteriors in $\Pi_{A|M^s,Y}$ to α. The resulting class of α-posteriors can be represented as

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

We use $\Pi_{\alpha|M^s,Y}$ as the representation of the posterior uncertainty for α in the set-identified model. $\Pi_{\alpha|M^s,Y}$ consists of any α-posteriors that assign probability one on its identified set, and it represents the lack of belief therein in terms of Knightian uncertainty (ambiguity). This is an important departure from the standard approach to Bayesian model averaging, which requires a single posterior distribution in every candidate model including the one where the parameter of interest is non-identified.

We combine the single posterior of α in model M^p and the set of posteriors of α in model M^s according to their posterior model probabilities denoted by $\pi_{M^p|Y}$ and $\pi_{M^s|Y}$. It is important to note that the posterior model probability on model M^s depends only on the prior for the reduced-form parameters. Hence, with $\pi_{\phi|M^s}$ fixed in the construction of $\Pi_{\alpha|M^s,Y}$, $\pi_{M^s|Y}$ can be uniquely determined in spite of the multiple prior input for the structural parameters. An important part of analysis is to establish conditions under which the prior model probabilities are updated by data — the update occurs if the models are distinguishable for some reduced-form parameter values and/or they have different priors for ϕ (see Lemma 3.1 below). In the current scenario, the two models are observationally equivalent at any $\phi \in \Phi$, so the data never update the prior model probabilities, $(\pi_{M^p|Y}, \pi_{M^s|Y}) = (w, 1-w)$.

The main proposal in this paper aggregates the posterior beliefs for α by forming the following set posteriors:

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

We refer to $\Pi_{\alpha|Y}$ as the *post-averaging ambiguous belief*; the class of mixture distributions in which the mixture weights are the posterior model probabilities $(\pi_{M^p|Y}, \pi_{M^s|Y})$, the component distribution corresponding to model M^p is $\pi_{\alpha|M^p,Y}$, and the component distribution corresponding to model M^s varies over the posterior class $\Pi_{\alpha|M^s,Y}$ given in (2.6).[7] To summarize the post-averaging set of posteriors, we suggest to report the range of posterior means or quantiles of $\Pi_{\alpha|Y}$ and the *robust credible region* with credibility $\gamma \in (0,1)$, which is defined by the shortest interval that receives posterior probability at least γ for every posterior in $\Pi_{\alpha|Y}$. Proposition 3.1 below shows that the range of posterior means for α spanned by $\Pi_{\alpha|Y}$ is

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

where $(l(\phi), u(\phi))$ are the lower and upper bounds of the nonempty identified set for α shown in (2.2), $a + b[c,d]$ stands for $[a+bc, a+bd]$, and $E_{\phi|M^s,Y}(\cdot)$ denotes the posterior mean with respect to $\pi_{\phi|M^s,Y} = \tilde{\pi}_{\phi|Y}$. This expression for the range of posterior means is intuitive and simple to interpret; the range of averaged posterior means is the weighted average (Minkowski sum) of the posterior mean in model M^p and the range of posterior means (interval) in model

[7] As we show in Section 4.1 below, averaging formula (2.7) can be derived by applying the Bayes rule prior-by-prior to a certain class of priors that has the form of an ϵ-contaminated class priors.

M^s. Noting that the range of posterior means can be viewed as an estimator for the identified set in model M^s, our model averaging procedure can be viewed as a method to shrink the estimate of the identified set in the set-identified model toward the point estimate in the point-identified model. The amount of shrinkage is determined by the posterior model probabilities, which in the current case are not updated by the data (cf. Scenario 2 below and Example 2 in Appendix A, where the prior model probabilities can be updated).

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

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

That is, 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 $\pi_{M^s|Y}$, a draw of the mixture identified set is a nonempty identified set for α. The robust credible region with credibility level γ is approximated by an interval that contains the γ-fraction of the drawn $IS_\alpha^{mix}(\phi)$'s. The minimization problem presented in Step 5 of Algorithm 4.1 in Giacomini and Kitagawa (2015) 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 distribution of the structural parameters.

Scenario 2: Candidate Models

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

An important feature of Model M^B is the availability of a prior distribution for the whole structural parameters. The prior in model M^B can reflect Bayesian probabilistic uncertainty about the equality identifying assumptions. See, for instance, Baumeister and Hamilton (2015) for a careful construction of a prior distribution for a dynamic version of the current model based on a meta-analysis of the literature.[8]

Model M^B always yields a single posterior distribution for α. However, the influence of prior choice does not vanish even asymptotically due to the lack of identification. If the researcher could be *fully confident* with the prior specification in model M^B, she could perform

[8]In the SVAR application in Section 5, we include a DSGE-VAR model (Del Negro and Schorfheide (2004)) as one of the single-posterior models.

standard Bayesian inference and obtain a *credible posterior* despite the identification issue. This situation is, in practice, rather rare. For instance, the prior specification for the elasticity parameters considered in Baumeister and Hamilton (2015) is based on the elicitation of their first and second moments and the remaining distributional characteristics of the prior distribution are chosen according to analytical or computational convenience. Further, elicitation of the dependence among structural parameters is a challenging task, and a naively-specified independent prior could lead to unintended or counter-intuitive influences to the posterior inference.[9] These robustness concerns motivate us to mix the Bayesian model M^B with the set-identified model M^s that can accommodate the lack of prior knowledge for the structural parameters further than the dogmatic set-identifying restrictions.

The single prior for A specified in model M^B implies a single prior for the reduced form parameters $\pi_{\phi|M^B}$. In Scenario 2, we therefore relax the restrictions (2.3), and allow the prior for ϕ in model M^s to be different from the prior for ϕ in model M^B. This, in turn, affects the posterior model probabilities. Let $w \in [0,1]$ be a prior weight assigned to model M^B, and $p(Y|M^B) \equiv \int_\Phi p(Y|\phi) d\pi_{\phi|M^B}(\phi)$ and $\tilde{p}(Y) \equiv \int_\Phi p(Y|\phi) d\tilde{\pi}_\phi(\phi)$ be the marginal likelihoods of model M^B and model M^s, respectively, where $p(Y|\phi)$ is the likelihood of the reduced form parameters (that is common across the models as reflected in the notation). The posterior model probabilities are

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

The different priors for ϕ imply $p(Y|M^B) \neq \tilde{p}(Y)$, and thereby the prior model probabilities can be updated by the data.

With these posterior model probabilities, the construction of the post-averaging ambiguous belief proceeds as in (2.7). The range of posterior means for α can be obtained similarly to (2.8), where M^B replaces M^p. The robust credible regions can be constructed in the same manner as in Scenario 1: with iid draws of binary $z_1, \ldots, z_G \sim Bernoulli(\pi_{M^B|Y})$, we generate

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

3 Formal Analysis

In this section, we formalize the idea in a more general setting and verify the analytical claims made in the previous section.

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

3.1 Notation and Framework

Consider $J+K \geq 2$ candidate models, $J, K \geq 0$, that can differ in various aspects including the identifying assumptions imposed and the parametrizations of the structural models. We call a member of J models as a *single-posterior model*, whose prior input always (i.e., independent of the realization of data) leads to a single posterior distribution for the parameter of interest. For instance, a model that imposes dogmatic point-identifying assumptions with a single prior distribution for the reduced-form parameters (such as model M^p in Scenario 1) belongs to this class, as does a model that assumes a single prior distribution for the whole structural parameters (such as model M^B in Scenario 2). We denote the collection of single-posterior models by $\mathcal{M}^p$.

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

Let $\mathcal{M} \equiv \mathcal{M}^p \cup \mathcal{M}^s$. We denote a vector of structural parameters in model $M \in \mathcal{M}$ by $\theta_M \in \Theta_M$, where we define the domain Θ_M as the set of structural parameters that satisfy the identifying assumptions imposed in model M. We assume that the 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 any of the candidate models. The reduced-form parameter ϕ_M is a function of the structural parameters, $\phi_M = g_M(\theta_M) \in \mathbb{R}^{d_M}$, where $g_M(\cdot)$ maps a set of observationally equivalent structural parameters subject to the identifying assumptions of model M to a point in the reduced-form parameter space.[10] The domain of the reduced-form parameters is defined by $\Phi_M = g_M(\Theta_M)$. As reflected in the notations, our most general set-up allows the parameter spaces of the structural parameters as well as those of the reduced-form parameters to differ across the models.[11] We express the likelihood in model $M \in \mathcal{M}$ in terms of the reduced-form parameters by $p(Y|\phi_M, M)$. For a multiple-posterior 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 $\mathbb{R}$.

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

[11]For instance, in the context of the simultaneous equation model considered in Section 2, the reduced-form parameter space differs depending on how many lagged endogenous variables and/or other exogenous variables are included in the equation system.

Note that, by its construction, the domain of the reduced form parameters 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 nonempty identified set, $\Phi_{M^s} = \{\phi_{M^s} \in \mathbb{R}^{d_{M^s}} : IS_\alpha(\phi_{M^s}|M^s) \neq \emptyset\}$.[12]

We introduce the following concepts concerning the relationships of the reduced-form parameter spaces among candidate models.

Definition 3.1 *Let $\mathcal{M}$ be a collection of candidate models.*

*(i) $\mathcal{M}$ **admits an identical reduced-form** if the following three condition holds:*

(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) The reduced-form likelihood $p(Y|\phi_M = \phi, M)$ defines a probability distribution of Y for all $\phi \in \Phi \equiv \cup_{M\in\mathcal{M}}\Phi_M$.

(c) Y is conditionally independent of M given ϕ, i.e., $p(Y|\phi_M = \phi, M) = p(Y|\phi)$ holds for all $\phi \in \Phi$ and $M \in \mathcal{M}$, where $p(Y|\phi)$ is the likelihood function common among $M \in \mathcal{M}$.

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

*(iii) Two distinct models $M, M' \in \mathcal{M}$ are **distinguishable** if $\Phi_M \neq \Phi_{M'}$.*

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

The concept of an identical reduced-form formalizes the situation where the models differentiated in terms of identifying assumptions lead to an identical parametric family of distributions of observables (Condition (a) in Definition 3.1 (i)). Even when the models admit an identical reduced-form, the imposed identifying assumptions can be observationally restrictive, and thereby $\{\Phi_M\}$ the supports of the reduced-form parameters can differ across the models. Important features in Definition 3.1 (i) are Conditions (b) and (c). They require that the distribution of Y in model M (indexed by ϕ) can be well-defined over the extended domain $\Phi = \cup_{M\in\mathcal{M}}\Phi_M$. For instance, if $\mathcal{M}$ consists of SVAR(p) models with different choices of identifying assumptions

[12] For instance, in SVAR(p) with observationally restrictive sign restrictions, Φ_M is the set of reduced-form parameters in VAR(p) yielding the nonempty impulse response identified set, which can be a proper subset of the reduced-form parameter space of VAR(p).

(including observationally restrictive ones such as SVARs with sign restrictions), the conditions of Definition 3.1(i) are satisfied with the identical reduced-form being VAR(p).[13]

Models that are observationally equivalent at ϕ (Definition 3.1 (ii)) fit equally well to the given distribution of data (corresponding to ϕ). Note that our definition of the observational equivalence is local to the given ϕ, and it does not constrain relationships among Φ_M's except that they have the non-empty intersection. In contrast, the concept of (in)distinguishability in Definition 3.1 (iii) and (iv) compares the domains of the reduced-form parameters across the models. In particular, Definition 3.1 (iv) is interpreted as observational equivalence in the global sense — if the models in $\mathcal{M}$ are indistinguishable, no matter what knowledge we have for the distribution of data, we never be able to judge what model fits better than the others (e.g., the candidate models of SVAR(0) considered in Section 2).

3.2 Posterior Model Probabilities

This section derives posterior model probabilities to be used for our averaging procedure. Lemma 3.1 below shows analytical results in the special case where the candidate models admit an identical reduced-form.

By the definition of reduced-form parameters, the value of the likelihood depends on θ_M only through ϕ_M. This means that the marginal likelihood can be computed uniquely for every model including any multiple-posterior model since it specifies a single prior for ϕ_{M^s}.

Let $(\pi_M : M \in \mathcal{M})$, $\sum_{M \in \mathcal{M}} \pi_M = 1$, be prior probability weights assigned over $\mathcal{M}$. By the Bayes' rule, the posterior model probabilities $(\pi_{M|Y} : M \in \mathcal{M})$ are

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

In the special cases addressed precisely in the next lemma, we can obtain simple formulae for the posterior model probabilities.

Lemma 3.1 *Suppose that the multiple-posterior models $M^s \in \mathcal{M}^s$ admit an identical reduced-form, so that we denote their reduced-form parameters by $\phi_{M^s} = \phi \in \mathbb{R}^d$. Let $\Phi = \cup_{M^s \in \mathcal{M}^s} \Phi_{M^s} \subset \mathbb{R}^d$. Suppose also that for a proper prior $\tilde{\pi}_\phi$ on Φ, $\tilde{\pi}_\phi(\Phi_{M^s}) = \tilde{\pi}_\phi(IS_\alpha(\phi|M^s) \neq \emptyset) > 0$ 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 $p(Y|\phi)$ that is common among $M^s \in \mathcal{M}^s$.*

(i) If the ϕ-prior in each multiple-posterior model satisfies for every measurable subset $B \subset \Phi$,

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

[13] The treatment effect models of Appendix A.2 shows another example that admits an identical reduced-form.

i.e., the prior of ϕ in model $M^s \in \mathcal{M}^s$ is constructed by trimming the support of $\tilde{\pi}_\phi$ to $\Phi_{M^s} = \{IS_\alpha(\phi|M^s) \neq \emptyset\}$, 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 prior $\tilde{\pi}_\phi$,

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

(ii) Suppose that, in addition to $\mathcal{M}^s$, all the single-posterior models $\mathcal{M}^p$ admit the identical reduced-form. If a prior for ϕ in every $M \in \mathcal{M}$ satisfies (3.2) with common $\tilde{\pi}_\phi$ and $\tilde{\pi}_\phi(\Phi_M) > 0$, then the posterior model probabilities can be 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 is the posterior-prior plausibility ratio $O_M = \frac{\tilde{\pi}_{\phi|Y}(\Phi_M)}{\tilde{\pi}_\phi(\Phi_M)}$.

(iii) If all the candidate models are indistinguishable and a prior for ϕ is common, then the model probabilities are not updated, $\pi_{M|Y} = \pi_M$ for all $M \in \mathcal{M}$.

Lemma 3.1 clarifies the sources of updating for the model probabilities when the candidate models yield the identical reduced-form model but differ in terms of the identifying assumptions. In the first claim, the specification of ϕ-prior (3.2) simplifies the marginal likelihood of the set-identified model $M^s \in \mathcal{M}^s$ to $\tilde{p}(Y)O_{M^s}$. Since computation for $\tilde{p}(Y)$ and O_{M^s} requires only one set of Monte Carlo draws of ϕ from each $\tilde{\pi}_{\phi|Y}$ and $\tilde{\pi}_\phi$, we save computation by avoiding to run separate Monte Carlo simulations across the set-identified models. In the setting of Lemma 3.1 (ii), we do not even need to compute the marginal likelihoods. The posterior model probabilities are updated only through the posterior-prior plausibility ratios of the candidate models, implying that draws of the reduced form parameters from prior $\tilde{\pi}_\phi$ and posterior $\tilde{\pi}_{\phi|Y}$, and assessments of the validity of the identifying assumptions at each drawn ϕ are all we need to approximate them. In particular, the claim in (iii) says if all the candidate models are indistinguishable, the prior model probabilities can never be updated.

In the example presented in Section 2, Scenario 1 represents a case where Lemma 3.1 (iii) applies: no update occurs for the model probabilities. Scenario 2 represents the case of Lemma 3.1 (i) with $O_{M^s} = 1$, since the identified set is never empty. In the example of the treatment effect model presented 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) offers the formula of their posterior model probabilities.

3.3 Post-Averaging Ambiguous Belief and the Range of Posteriors

Estimation of the single-posterior models proceeds in the standard Bayesian way. We therefore take $\pi_{\alpha|M^p,Y}$, the posterior for α in each single-posterior 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 $\pi_{\phi_{M^s}|M^s}$ for the reduced form parameters with support 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}.[14] In words, $\Pi_{\theta_{M^s}|M^s}$ collects prior distributions of θ_{M^s} that meet the imposed identifying restrictions 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 prior-by-prior updating to $\Pi_{\theta_{M^s}|M^s}$ with the likelihood $\tilde{p}(Y|\theta_{M^s}, M^s)$ and marginalizing the resulting posteriors in terms of α, we obtain the set of posteriors for α with the following form:[15]

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

Given the posterior model probabilities, an averaged posterior for α is a mixture,

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

where $(\pi_{\alpha|M^p,Y} : M^p \in \mathcal{M}^p)$ are given uniquely, while for $M^s \in \mathcal{M}^s$, the posterior information of α is summarized by the set of posteriors given in (3.7). Since there are no cross-model restrictions that constrain selections of posteriors across the different multiple posterior models, the set of averaged posteriors spanned by $\{\Pi_{\alpha|M^s,Y} : M^s \in \mathcal{M}^s\}$ is obtained as

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

[14]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.

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

Our presentation of the averaging step is heuristic and does not explicitly justify it in relation to any robust Bayesian updating rule. If the structural parameters are common across the candidate models (i.e., the subscript in θ_M can be dropped), we can formally derive the formula of post-averaging ambiguous belief by applying the prior-by-prior updating rule to a certain class of priors, as shown in the next proposition.

Proposition 3.1 *Suppose that the structural parameters in any of the candidate models can be embedded into a common parameter space, i,e,, $\theta_M = \theta \in \mathbb{R}^{d_\theta}$ for all $M \in \mathcal{M}$. Accordingly, define $\Theta = \cup_{M \in \mathcal{M}} \Theta_M \subset \mathbb{R}^{d_\theta}$. Let prior models probabilities $(\pi_M : M \in \mathcal{M})$, $\pi_{\theta|M^p}$ a unique prior for θ in $M^p \in \mathcal{M}^p$, and a unique prior for reduced-form paramters $\pi_{\phi_{M^s}|M^s}$ in $M^s \in \mathcal{M}^s$ be given. Define the following set of priors for $(\theta, M) \in \Theta \times \mathcal{M}$:*

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

The prior-by-prior Bayesian updating rule applied to $\Pi_{\theta,M}$ with likelihood $\tilde{p}(Y|\theta, M)$ yields (3.8) as the class of posteriors marginalized to α.

The next proposition shows the range 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 range of posterior means of $\Pi_{\alpha|Y}$ is the convex interval with the lower and upper bounds given by

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

where $E_{\phi_{M^s}|Y,M^s}(\cdot)$ is the expectation with respect to the posterior distribution of ϕ_{M^s} in model $M^s \in \mathcal{M}^s$.
(ii) For any measurable subset A in $\mathbb{R}$, the lower bound of the posterior probabilities on $\{\alpha \in A\}$ in the class $\Pi_{\alpha|Y}$ (the lower posterior probability of $\Pi_{\alpha|Y}$) is

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

(iii) The lower and upper bounds of the cumulative distribution functions (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 range 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 a point estimator for the identified set in model M^s. We can therefore interpret the range of post-averaging posterior means as the weighted Minkowski sum of the Bayesian point estimators (posterior mean) 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 of 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 Monte Carlo algorithm presented in (2.9).

3.4 Computations

To report the range 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 distributions of α for each single-posterior model, and (III) the identified set $IS_\alpha(\phi_{M^s}|M^s)$ and its probability law $\pi_{\phi_{M^s}|M^s,Y}$ for each multiple-posterior model. Since the estimation of the single-posterior model in (II) is the standard Bayesian inference, we assume some suitable posterior sampling algorithm is applicable to have 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, Waggoner, and Zha (2008). When Lemma 3.1 (i) or (ii) applies, such as in the empirical application below, the computation of the marginal likelihoods for multiple-posterior models can be reduced to the computation of the posterior-prior plausibility ratios O_M. Since O_M's and the quantities in (III) are less common, this section briefly discusses how to compute them under the setting of Lemma 3.1(i) or (ii).

In each multiple-posterior model, if we can assess non-emptiness of the identified set at each $\phi \in \Phi$, the posterior-prior plausibility ratio O_{M^s} can be computed simply by plugging in numerical approximations for the prior and posterior probabilities of non-emptiness of the identified set into (3.4). The denominator of O_{M^s} is computed by drawing many ϕ's from the prior $\tilde{\pi}_\phi$ and computing the fraction of draws that yield nonempty identified sets. The numerator of O_{M^s} is computed similarly except that ϕ's are drawn from the posterior $\tilde{\pi}_{\phi|Y}$.[16]

Monte Carlo draws of the lower and upper bounds of the identified set in model $M \in \mathcal{M}^s$ can be obtained in the following two steps. In the first step, we draw ϕ's from the posterior $\tilde{\pi}_{\phi|Y}$. In the second step, we retain the draws of ϕ that yield nonempty $IS_\alpha(\phi|M^s)$, and compute corresponding $l(\phi|M^s)$ and $u(\phi|M^s)$. By taking their sample averages, we can approximate $E_{\phi|M^s,Y}(l(\phi|M^s))$ and $E_{\phi|M^s,Y}(u(\phi|M^s))$. Implementation of this computational procedure relies on computability of the lower and upper bounds of the identified set for each ϕ. Whether it is a simple task or not depends on applications. 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, Hong, and Tamer (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 probability:

Algorithm 3.1

Step 1: Draw a model $M \in \mathcal{M}$ from the 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 set). If the drawn M belongs to $\mathcal{M}^s$, draw $\phi_M \sim \pi_{\phi|M,Y}$ and set $IS_\alpha^{mix} = IS_\alpha(\phi_M|M)$.[17]

[16] For instance, in the SVAR application considered in Section 5, we can assess non-emptiness of the identified set by drawing many non-identified parameters (rotation matrices) from the uniform distribution (Haar measure on the space of orthonormal matrices) using the sampling algorithm of Uhlig (2005), and then verifying if any of the draws satisfy the imposed sign restrictions. See also Algorithm 5.1 in Giacomini and Kitagawa (2015).

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

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

Step 4: Let $[l_g^{mix}, u_g^{mix}]$ be the lower and upper bound 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 $\mathcal{M}^p$). We approximate the mean bounds of the post-average posterior class by

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

We approximate the lower probability of the post-averaging posterior class at $A \subset \mathbb{R}$ by

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

The Monte Carlo 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 posterior class assigns probability at least γ;

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

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 5.1 of Kitagawa (2012) and it can be readily applied to the current context.

3.5 Asymptotic Behaviour

This section analyzes the asymptotic properties of our averaging proposal. The proposed procedure is indeed finite-sample exact (up to Monte Carlo approximation errors) and does not rely on any asymptotic approximation. The asymptotic analysis is nevertheless valuable, as it informs what aspect of prior input can be influential to the posterior output independent of the sample size. In this section, we make the sample size explicit in our notation by denoting a size n sample by Y^n.

To keep tight links to the analytical results of Lemma 3.1 and the empirical application below, we assume throughout this section that $\mathcal{M}$ admits an identical reduced-form in the sense of Definition 3.1 (i). We assume that at least one model is correctly specified so that the true 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

$\hat{\phi} \equiv \arg\max_{\phi \in \Phi} p(Y^n|\phi)$. In what follows, we denote the true probability law of the sampling sequence $\{Y^n : n = 1, 2, \ldots\}$ by $P_{Y^\infty|\phi_{true}}$.

Our asymptotic results shown below uses 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 nonempty.

(ii) Let $l_n(\phi) \equiv n^{-1} \log p(Y^n|\phi)$. There exists B an open neighborhood of ϕ_{true} and $n_0 \geq 1$ such that for any $\{Y^n : n = n_0, n_0 + 1, \ldots\}$, $l_n(\cdot)$ 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 B open neighborhood 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 model $M \in \mathcal{M}$, $\pi_{\phi|M}$ has the 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 the 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 3.1 (ii). The requirement for ϕ_{true} being in the interior of Φ_M implies that Φ_M, $M \in \mathcal{M}_A$, has a nonempty interior in $\mathbb{R}^d$. In addition, for a set-identified model, condition (A) implies that $M^s \in \mathcal{M}_A$ has nonempty $IS_\alpha(\phi|M^s)$ on an open neighborhood of ϕ_{true}, and condition (B) implies that $M^s \in \mathcal{M}^s \setminus \mathcal{M}_A$ has empty $IS_\alpha(\phi|M^s)$ on an open neighborhood of ϕ_{true}. Assumptions 3.2 (iii) and (iv) impose regularity conditions for the likelihood functions that imply strong (almost sure) consistency of $\hat{\phi}$. Assumptions 3.2 (ii) and (v) that restrict smoothness of the log-likelihood and ϕ-prior allow an application of the Laplace method to approximate the large sample marginal likelihood. A set of assumptions similar to Assumptions 3.2 (ii) - (v) appears in Kass, Tierney, and Kadane (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, summarizes the large sample limits of the posterior model probabilities.

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

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

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 $\{Y^n : n = 1, 2, \dots\}$.

This proposition clarifies the large sample behaviors of the averaging weights in the situations where the candidate models admit an identical reduced-form. First, the posterior model probabilities can correctly screen out all the misspecified models $M \notin \mathcal{M}_A$, as whose posterior weights converge to zero asymptotically irrespective of an initial choice of prior model probabilities. Accordingly, if there is only one model that is consistent with the true data generating process, our averaging procedure asymptotically assigns weight one to it. Second, if $\mathcal{M}_A$ contains multiple models, the asymptotic weights allocated among $M \in \mathcal{M}_A$ are determined by the prior model probabilities $\{\pi_M : M \in \mathcal{M}_A\}$ and the densities of ϕ-priors evaluated at ϕ_{true}. This implies that the sensitivity of post-averaging posterior to $\pi_{\phi|M}$ and π_M does not vanish asymptotically when multiple candidate models are observationally equivalent at ϕ_{true}. Third, when 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 $M_1, M_2 \in \mathcal{M}_A$, it holds $\pi_{M_1|Y^\infty} \geq \pi_{M_2|Y^\infty}$. 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" consistent with the claim of (ii) corresponds to the size of the reduced-form parameter spaces, and has nothing to do with the strength of identifying assumptions often measured by the width of the α's identified set.[18]

A combination of the asymptotic posterior model probabilities obtained in Proposition 3.3 and the asymptotic behaviors of $\pi_{\alpha|M,Y^n}$ of single-posterior models and $\Pi_{\alpha|M,Y^n}$ of multiple-posterior models yields the asymptotic convergence properties on the range of the post-averaging ambiguous belief. To be specific, in addition to Assumption 3.2, we assume that (i) the posterior for ϕ is consistent to 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 $M^s \in \mathcal{M}^s \cap \mathcal{M}_A$, the identified set correspondence $IS_\alpha(\phi|M^s)$, 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 range of post-averaging posterior means considered in Proposition 3.2 (i) has the following large sample limits:

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

4 Discussion

4.1 Relationship with Robust Bayes Analysis with ϵ-contaminated Class of Priors

The method proposed in this paper has a close link to the robust Bayes analysis with an *ϵ-contaminated class of priors* (Huber (1973), Berger and Berliner (1986)). To clarify this, consider a simple case where the candidate models are one single posterior model and one multiple posterior model, $\mathcal{M} = \{M^p, M^s\}$. Furtheremore, we assume that these two models share a parametrization of the structural model, i.e., the likelihood for common structural parameters θ does not depend on the models, i.e., M^p and M^s differ only in the identifying assumptions, implying $\Theta_{M^p} \neq \Theta_{M^s}$.

[18]For instance, in the SVAR context, a model just-identified by a set of equality restrictions is not observationally restrictive, while a model set-identified by sign restrictions is observationally restrictive if the number of sign restrictions is larger than the number of variables in the SVAR system. If specifications of ϕ-prior satisfy (3.2) and these 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.

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 and marginalizing to α.

A general formulation of 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 distribution 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 ranges of posterior probabilities for various specifications of $\mathcal{Q}$ when a prior varies over Π_θ^ϵ.

Despite 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 exactly the same form as the ϵ-contaminated class of priors (4.2) — Π_θ is an ϵ-contaminated class of priors with benchmark prior coming from the single-prior (point-identified) model $\pi_\theta^0 = \pi_{\theta|M^p}$, the amount of contamination is the prior model probability assigned to the set-identified model $\epsilon = \pi_{M^s}$, and the set of priors $\mathcal{Q}$ corresponds to the multiple prior input of the set-identified model $\Pi_{\theta|M^s}$. This coincidence clarifies the robust Bayes interpretation behind our averaging scheme.[19] If the single-posterior (point-identified) model plays the role of a sensible benchmark 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.

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 the partial identification analysis, our analysis offers a new way to specify $\mathcal{Q}$ without losing analytical and numerical tractability.

[19] 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 the constant marginal likelihood over $\pi_{\theta|M^s} \in \Pi_{\theta|M^s}$.

4.2 Eliciting Prior Model Probabilities

The key prior input of our procedure is the prior model probability. The robust Bayesian viewpoint along the ϵ-contaminated class of priors clarifies interpretation of the prior model probability and facilitates its elicitation in practice.

Suppose again the set of candidate models consists of one point-identified model M^p and one set-identified model M^s. In addition, we assume M^p is nested in M^s in the sense that the identifying assumptions in M^p include those imposed in M^s. In this case, the prior model probability assigned to M^p should 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 put 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 the 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.

4.3 Relationship with Hierarchical Bayesian Approach

Point-identifying assumptions or a prior for structural parameters sometime come from a structural econometric model utilizing 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 commonly used DSGE model as a single-posterior model and a sign restricted SVAR model as a set-identified model.

In such context, averaging them offers a way to combine the structural modelling approach and a more "reduced-form" approach.[20] The macroeconometrics literature has proposed uses

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

of hierarchical Bayesian methods to bridge the gap between the structural modelling approach and the "reduced-form" approach (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 context, differs from the hierarchical Bayesian approach in the following aspects. First, the hierarchical Bayesian approach always leads to a single posterior for the causal parameters (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 prior inputs 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-posterior model. Limited credibility in the prior for the Bayesian DSGE model can be incorporated into the posterior inference by mixing 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 the Bayesian probabilistic uncertainty while the latter is in the scale of Knightian uncertainty.

5 Empirical Application

We illustrate our method in the context of a monetary SVAR. We consider a 4-variable SVAR with two lags. Let $y_t = [i_t, \Delta y_t, \pi_t, m_t]'$ denote the endogenous variables, where i_t is the federal funds rate, Δy_t is real output growth, π_t is inflation and m_t is a measure of real money. Following Notation 3.1 in Giacomini and Kitagawa (2015), we order the variables so that we can easily verify the conditions guaranteeing convexity of the identified set using their Lemmas 5.1 and 5.2.

$$A_0 \begin{pmatrix} i_t \\ \Delta y_t \\ \pi_t \\ m_t \end{pmatrix} = c + \sum_{j=1}^{2} A_j \begin{pmatrix} i_{t-j} \\ \Delta y_{t-j} \\ \pi_{t-j} \\ m_{t-j} \end{pmatrix} + \begin{pmatrix} \epsilon_t^i \\ \epsilon_t^{\Delta y} \\ \epsilon_t^{\pi} \\ \epsilon_t^m \end{pmatrix} \tag{5.1}$$

where

$$A_0 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}. \tag{5.2}$$

The reduced-form VAR is as follows

$$y_t = b + \sum_{j=1}^{2} B_j y_{t-j} + u_t \tag{5.3}$$

for $t = 1, \ldots, T$, where $b = A_0^{-1}c, B_j = A_0^{-1}A_j, u_t = A_0^{-1}\epsilon_t, var(u_t) = E(u_t u_t') = \Sigma = A_0^{-1}(A_0^{-1})'$. Let $\phi = (b, B_1, B_2, \Sigma)$ collect the reduced-form parameters.

We interpret the first equation as a monetary policy function: the Federal Reserve reacts to price, GDP and money, as well as lags of all variables. Any additional change would be an exogenous monetary policy shock. The second and third equations represent aggregate demand (AD) and aggregate supply (AS), respectively. The last equation is a money demand equation derived from the well-known relation $MV = PY$, where Y is the real income and V represents velocity. In this perspective, ϵ_t^m stands for a velocity shock, provided that real GDP is the real income. The data are quarterly observations from 1965:1 to 2005:1 and are from the FRED2 database of the Federal Reserve Bank of St. Louis. The data set is from Aruoba and Schorfheide (2011), and it is the same as that used by Moon, Schorfheide, and Granziera (2013) and Giacomini and Kitagawa (2015).

The prior for the reduced-form parameters belongs to the Normal-Wishart family and is common among the models unless otherwise stated:

$$\Sigma^{-1} \sim \mathcal{W}(\Psi^{-1}, d), \;\; \beta|\Sigma \sim \mathcal{N}(\bar{b}, \Sigma \otimes \Omega),$$

where $\beta \equiv vec([b, B_1, B_2]')$. Let Ψ be the location matrix of Σ and d a scalar corresponding to the degrees of freedom. Let $\bar{b}$ and Ω denote the prior mean and variance-covariance matrix of β, respectively.

Following Christiano, Eichenbaum, and Evans (1999), we impose the sign normalization restrictions so that the diagonal elements of A_0 are all nonnegative.

5.1 Averaging Non-falsifiable Models

Suppose we are interested in the output response to a unit positive shock in the federal funds rate ϵ_t^i, denoted by $IR^h_{\Delta yi}$, and consider the following two sets of identifying assumptions.

- *Model 1 (M1, point-identified)*

  Consider the standard recursive causal ordering restrictions (Bernanke (1986) and Sims (1980)). In particular, we adopt the Cholesky decomposition to identify the interest rate shock and follow the literature in ordering it fourth after output, inflation, and the monetary aggregate:

$$\begin{pmatrix} a_{11}^{M1} & 0 & 0 & 0 \\ a_{21}^{M1} & a_{22}^{M1} & 0 & 0 \\ a_{31}^{M1} & a_{32}^{M1} & a_{33}^{M1} & 0 \\ a_{41}^{M1} & a_{42}^{M1} & a_{43}^{M1} & a_{44}^{M1} \end{pmatrix} \begin{pmatrix} \Delta y_t \\ \pi_t \\ m_t \\ i_t \end{pmatrix} = c + \sum_{j=1}^{2} A_j \begin{pmatrix} \Delta y_{t-j} \\ \pi_{t-j} \\ m_{t-j} \\ i_{t-j} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{\Delta y} \\ \epsilon_t^{\pi} \\ \epsilon_t^{m} \\ \epsilon_t^{i} \end{pmatrix} \tag{5.4}$$

  The economic interpretation is that AD, AS and money demand do not react to the contemporaneous interest rate shock. Since we are interested in identifying a single shock, the identification scheme restricts A_0 in equations (5.1) and (5.2) so that $a_{21} = a_{31} = a_{41} = 0$.

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

  The recursive identification scheme in Model 1 is controversial[21]. For example, assumption $a_{31} = 0$, assuming that prices do not react contemporaneously to the interest rate shock, can be difficult to justify if the researcher relies on the stock price index rather than the GDP deflator. Thus, in Model 2 we leave AS unrestricted, i.e., AS can react to the interest rate within the quarter. By Lemma 5.1 in Giacomini and Kitagawa (2015), Model 2 now delivers a convex identified set for $IR^h_{\Delta yi}$ for every value of the reduced form parameters.

The prior weights assigned to the two models are w_1 and $w_2 = 1 - w_1$, respectively. In other words, with prior weight w_1 the researcher believes that AS does not respond to the interest rate within a quarter (Model 1); with prior weight $(1 - w_1)$, the researcher believes that AS reacts to the interest rate within a quarter (Model 2).

Figure 1 reports the credible region for the output response based on Model 1 and Model 2 at horizon $h = 3$. In the top panel, the vertical solid lines for Model 1 are the credible region for the point-identified model based on a single posterior for the impulse response, and the vertical dashed lines for Model 2 are the posterior mean bounds (consistent estimator of the identified set) for the output response. The bottom left panel reports the combination results for uniform prior weights, $w_1 = w_2 = 0.5$. The vertical dashed lines for the averaged model can be interpreted as shrinking the identified set estimator from Model 2 towards the point estimator from Model 1. Figure 2 reports the output impulse response credible sets for multiple horizons for the same set of models as in Figure 1.

[21] See Kilian (2013) for details over the limitations of point-identifying assumptions in such a model.

First of all, since the models are observationally equivalent and non-falsifiable, the prior weights are not updated and hence directly affect the averaged model. Second, as common for standard recursive causal ordering restrictions in small-scale SVAR, Model 1 shows output puzzle[22], while set-identified Model 2 supports neutrality of money. Under $w_1 = 0.5$, the prior weight attached to Model 1 is high enough to favour output puzzle in the averaged impact response, i.e., the range of averaged posterior means at $h = 0$ lies entirely above zero and so does most of the correspondent robust credible region. It can be shown that output puzzle vanishes as prior model probability for Model 1 is equal or less than 0.24, i.e., zero is now contained in the range of averaged posterior means and the robust credible region is (almost) equally split above and below zero.

We also perform averaging exercise for $w_1 = 0.8$. As expected, output puzzle is still present, the range of averaged posterior means almost collapses to the (single) posterior mean of Model 1, and the robust credible region gets tighter.

5.2 Averaging with Falsifiable Models

In addition to the previous two sets of assumptions, we further consider two sets of identifying assumptions that are widely used in empirical applications and that respectively lead to a set-identified falsifiable model and a point identified model: sign restrictions and restrictions from a 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 for two quarters; the interest rate response is non-negative for two quarters; the response of money is nonpositive for two quarters. By Lemma 5.2 in Giacomini and Kitagawa (2015), the identified set in Model 3 is convex.

Let the prior weights for Model 1, Model 2, and Model 3 be denoted by (w_1, w_2, w_3). In contrast to the previous examples, we are now able to update the prior weights by using equation (3.6), starting from uniform prior weights, $w_1 = w_2 = w_3 = 0.33$. Figure 3 and Figure 4 report the results of averaging the three models: while the interpretation of Model 1 and Model 2 is the same as before, Model 3 favours neutrality of money as common in literature (Uhlig 2005, but he relies on a single prior). Table 1 shows that the posterior model probabilities strongly support Model 3, implying that the weighted model is dominated by sign-restricted framework. Note that the posterior weights for Model 1 and Model 2 are the same as they are observationally equivalent.

[22] Output puzzle refers to contractionary shocks in monetary policy leading to an initial rising rather than falling of output.

- *Model 4 (M4, DSGE-VAR)*

  We consider the DSGE-VAR approach suggested by Del Negro and Schorfheide (2004). We use a standard New Keynesian DSGE model as a prior for the vector autoregression. Thus, we now allow the reduced-form and its prior to differ across models. The DSGE model consists of three main endogenous processes: inflation, real output, and a nominal interest rate, represented in the form of a Phillips curve, a forward-looking IS equation, and a monetary policy function, respectively[23]. In order to obtain the VAR prior from the DSGE model, we simulate time-series data from the DSGE model and fit a VAR to these data. Specifically, we replace the sample moments of the simulated data by population moments computed from the DSGE model solution. The DSGE model relies on unknown structural parameters, so we use a hierarchical prior by setting up a distribution on the DSGE model parameters. A tightness parameter controls the weight on the DSGE model prior relative to the weight of the actual sample. Finally, Markov Chain Monte Carlo methods generate draws from the joint posterior distribution of the VAR and DSGE model parameters and the Laplace approximation is used to compute the marginal likelihood.

  In order to identify the interest rate shock, we construct an orthonormal matrix from the VAR approximation of the DSGE model to map the reduced-form innovations into structural shocks. This procedure induces a DSGE model-based prior distribution for the VAR impulse responses. In other words, we first QR-factorize the state space representation of the DSGE model and then replace the unidentified rotation matrix of the SVAR with it. Thus, this framework can be viewed as a way to obtain a non-degenerate prior for the rotation matrix using the DSGE model and a prior for its parameters. Note that, since the DSGE-VAR is a single-posterior model, what matters for updating the prior weight for the model is its marginal likelihood.

  It is noteworthy that the extent to which the SVAR impulse responses resemble the DSGE model's responses will depend on the tightness of the prior. The larger the latter, the closer the responses will be. We choose the tightness of the prior endogenously based on maximization of the marginal data density, as in equation (34) in Del Negro and Schorfheide (2004). However, in our data the marginal data density is almost flat over the tightness parameter, meaning that updating of posterior probabilities is not affected by choice of tightness parameter.

We now propose to weight Model 1, Model 2, and Model 4. Let the prior weights be $w_1 = w_2 = w_4 = 0.33$. Given that the prior for reduced-form changes across point-identified models, i.e., Model 1 and Model 4, we can update prior model probabilities by using equation (3.4). Figure 5 and Figure 6 display the results of the averaging exercise: while Model 4 strongly

[23] As reference see Gali (2008); Lawless and Whelan (2011); Woodford (2003).

supports textbook theory (negative impact of contractionary monetary shock on output), its data-driven posterior probability converges to 1 (see Table 1). Thus, the weighted model collapses to Model 4 and, in contrast to the previous examples, is now point-identified. We observe a similar outcome as we weight Model 1, Model 2, Model 3, and Model 4 (Figure 7 and Figure 8), where $w_1 = w_2 = w_3 = w_4 = 0.25$: the averaged model collapses to Model 4 as its posterior probability goes to 1 again (Table 1). Interestingly, the last two averaging exercises seem to support evidence of negative impact of (contractionary) monetary policy shock on output[24].

5.3 Methodological Note

In each multiple-posterior model (Model 2 and Model 3 above), the prior posterior odds ratio O_M can be computed simply by plugging in numerical approximates of the prior and posterior probabilities for non-emptiness of the identified set into (3.5).

Specifically, the denominator of O_M can be computed by drawing many ϕ's from the reduced-form prior distribution and getting the fraction of draws that yield a nonempty identified set. In doing so, for each draw of ϕ, we need to draw many non-identified parameters (rotation matrices Q) and check if any of the draws satisfies the model's assumptions. For details, see Algorithm 5.1 in Giacomini and Kitagawa (2015). The numerator of O_M can be computed similarly except that ϕ's are drawn from the posterior distribution.

Note that this procedure presents two drawbacks. First, it is computationally costly and time-consuming when the number of variables/sign restrictions is large because it relies on drawing many non-identified parameters[25]. Second, it is just a rough assessment of the identified set non-emptiness. As a result, for each draw ϕ[26], we get around the subroutine of drawing many Q's and rely on alternative theorems in Border (2013), where construction of the identified set employs the criterion function approach in Chernozhukov, Hong, and Tamer (2007). In this perspective, the emptiness of the identified set can be reduced to existence of a feasible point in the linear system shaped by equality and inequality constraints, i.e., zero and sign restrictions, respectively. To put it another way, we check for the emptiness of the polyhedron created by such a system, meaning that the whole routine is reduced to a linear optimization problem. We find that this procedure is sometimes faster than drawing many Q's and becomes more efficient as we increase the number of zero and sign restrictions. In turn, this leads to a more precise estimate of the proportion of times that the identified set is empty.

[24]This statement is not supposed to be general and needs to be carefully considered. It is strictly conditional on dataset and class of candidate models.

[25]For instance, Giacomini and Kitagawa (2015) propose to draw 3000 Q's for each draw of reduced-form parameters.

[26]In our algorithm, we draw 1000 ϕ's.

5.4 Reverse-Engineering Prior Model Weights

We assumed a uniform distribution of prior weights for the majority of the empirical application. However, our procedure can readily accommodate non-uniform weights. For pioneering contributions on how to determine prior weights in Bayesian averaging, see, 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 one should be split between that model and its duplicates. Among others, Chipman (1996) attaches smaller prior weights to models that are unlikely, Hoeting, Madigan, Raftery, and Volinsky (1999) rely on variable selection in regression models to determine prior weights and Clyde and George (2004) propose a Bernoulli specification.

We now propose an exercise of reverse engineering. Suppose that we aim at selecting prior weights such that the response of output to a contractionary interest rate shock in the averaged model is initially non-positive, i.e., output puzzle. The economic rationale behind this exercise is to shed light on the role of identification schemes in our understanding of contractionary monetary policy shocks.

Reverse engineering constrains the following convex interval to be above zero for $h = 0, \ldots, \bar{H}$:

$$\left[\inf_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha), \sup_{\pi_{\alpha|Y}\in\Pi_{\alpha|Y}} E_{\alpha|Y}(\alpha)\right] \subseteq [0,\infty)\,,$$

where

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

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

We follow the usual notation and α is now the IRF of output to the interest rate shock. This is a system of inequalities, where the unknowns are prior weights π_M.

For simplicity of exposition, let us consider only Model 3 (set-identification through sign restrictions) and Model 1 (point-identification through causal ordering restrictions) as described in section 5.1 and 5.2. We thus have a system of inequalities, where the prior weight for Model 1 w is the unknown:

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

$$= \pi_{M^1|Y} E_{\alpha|M^1,Y}(\alpha) + \pi_{M^3|Y}\left[E_{\phi|Y,M^3}(l(\phi_{M^3}|M^3)), E_{\phi|Y,M^3}(u(\phi_{M^3}|M^3))\right] \subseteq [0,\infty)\,,$$

and

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

We find that the averaged IRF between Model 1 and Model 3 is above zero for $h = 0, 1, 2$ as $w \geq 0.95$. This result is consistent with the monetary policy literature, where the evidence for the output puzzle is very strong for small-scale SVAR with causal ordering restrictions. Similar reverse engineering exercises can be useful to shed light on the role of identification assumptions in generating so-called price and liquidity puzzles[27]. For simplicity, we considered two frameworks. However, the number of models can be increased straightforwardly at the cost of complicating the system of inequalities.

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 prior(s) in point-identified model(s) with multiple priors in set-identified model(s), and delivers a set of posteriors. The post-averaging set of posteriors can be summarized by the range of posterior means and robust credible regions, which are easy to compute based on the MCMC draws of the identified sets in each candidate model. Our averaging method can effectively reduce the amount of ambiguity (the size of the posterior class) relative to the analysis with 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 averaging 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;

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

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

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 range as ρ varies over the range 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 $\arctan(-\sigma_{21}/\sigma_{22}) \leq \theta \leq \arctan(\sigma_{22}/\sigma_{21})$. 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 $M^p \in \mathcal{M}^p$ agrees with $\tilde{p}(Y) O_{M^p}$. Hence, (3.5) follows.

(iii) The claim follows immediately by 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

$$\Pi_{\theta_{M^s}|M^s,Y} = \left\{\int_{\Phi_{M^s}} \pi_{\theta_{M^s}|\phi_{M^s},M^s} d\pi_{\Phi_{M^s}|Y} : \pi_{\theta_{M^s}|\phi_{M^s},M^s}(\Theta_{M^s} \cap g_{M^s}^{-1}(\phi_{M^s})) = 1, \pi_{\phi_{M^s}|M^s,Y}\text{-a.s.}\right\}. \tag{A.5}$$

Note that we have

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

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 $\pi_{\theta,M} \in \Pi_{\theta,M}$ be a prior. The corresponding posterior for θ with M integrated out can be computed as follows: for any measurable subset $A \subset \Theta$,

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

where the second line uses

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

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 Proposition 4.1 (ii) of Giacomini and Kitagawa (2015) shows $\inf_{\pi_{\alpha|M,Y}\in\Pi_{\alpha|M,Y}}\left\{E_{\alpha|M,Y}(\alpha)\right\} = E_{\phi|M,Y}(l(\phi|M))$. The claim of the mean lower bound therefore follows. The mean upper bound can be shown similarly.
(ii) Note that

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

Proposition 3.1 of Kitagawa (2012) shows

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

(iii) By setting A 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 range of quantiles then follows by inverting these cdf bounds. ■

Next, we show two lemmas 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 $\phi = (\phi_1, \ldots, \phi_d)$ 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

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

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

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

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 $(2\pi)^{d/2}p(Y^n|\hat{\phi})n^{-d/2}\det(H_n(\hat{\phi}))^{1/2}$:

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

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 $y \in \{1, 0\}$:

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

Ruling out the marginal distribution of Z, the structural parameters index a joint distribution of (Y_1, Y_0, T):

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

where Θ is the 16-dimensional probability simplex.

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

- *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.

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

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 (2009) 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).

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

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

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.

[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).

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 range 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.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113(1), 151–184.

Arias, J. E., J. F. Rubio-Ramirez, and D. Waggoner (2013): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," *unpublished manuscript.*

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(5), 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(6), 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*, pp. 17–36.

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, vol. 1, Part A, pp. 65–148. Elsevier.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 17911828.

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

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*, pp. 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(2), 512–549.

Del Negro, M., and F. Schorfheide (2004): "Priors from general equilibrium models for vars," *International Economic Review*, 45(2), 643–673.

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(1), 89–121.

Gafarov, B., M. Meier, and J.-L. Montiel-Olea (2016a): "Delta-Method Inference for a Class of Set-identified SVARs," *unpublished manuscript.*

——— (2016b): "Projection Inferene for Set-identified SVARs," *unpublished manuscript.*

Gali, J. (2008): *Monetary Policy, Inflation, and the Business Cycle*. Princeton University Press.

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 (2015): "Robust Inference about Partially-identified SVARs," *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(4), 1175–1189.

——— (2014): "Model Averaging, Asymptotic Risk, and Regressor Groups," *Quantitative Economics*, 5(3), 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(4), 382–417.

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, pp. 473–488. Elsevier, North-Holland.

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

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. (2009): "Identification Region of the Potential Outcome Distributions under Instrument Independence," *unpublished manuscript.*

Kitagawa, T. (2012): "Estimation and Inference for Set-identified Parameters Using Posterior Lower Probabilities," *unpublished manuscript.*

Kitagawa, T., and C. Muris (2016): "Model Averaging in Semiparametric Estimation of Treatment Effects," *Journal of Econometrics*, 193(1), 271–289.

Kline, B., and E. Tamer (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7(2), 329–366.

Lawless, M., and K. T. Whelan (2011): "Understanding the dynamics of labor shares and inflation," *Journal of Macroeconomics*, 33(2), 121–136.

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

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(3), 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(4), 997–1010.

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(5), 597–621.

Mountford, A., and H. Uhlig (2009): "What Are the Effects of Fiscal Policy Shocks?," *Journal of applied econometrics*, 24(6), 960–992.

Norets, A., and X. Tang (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, 81(3), 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(3), 727–750.

Rafiq, S., and S. Mallick (2008): "The Effect of Monetary Policy on Output in EMU3: A Sign Restriction Approach," *Journal of Macroeconomics*, 30(4), 1756–1791.

Reichenstein, W. (1987): "The Impact of money on short-term interest rates," *Economic Inquiry*, 25(1), 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*. John Wiley & Sons, Hoboken, New Jersey.

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), 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(5), 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.

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(3), 977–990.

Woodford, M. (2003): "Interest and prices: Foundations of a theory of monetary policy," .

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

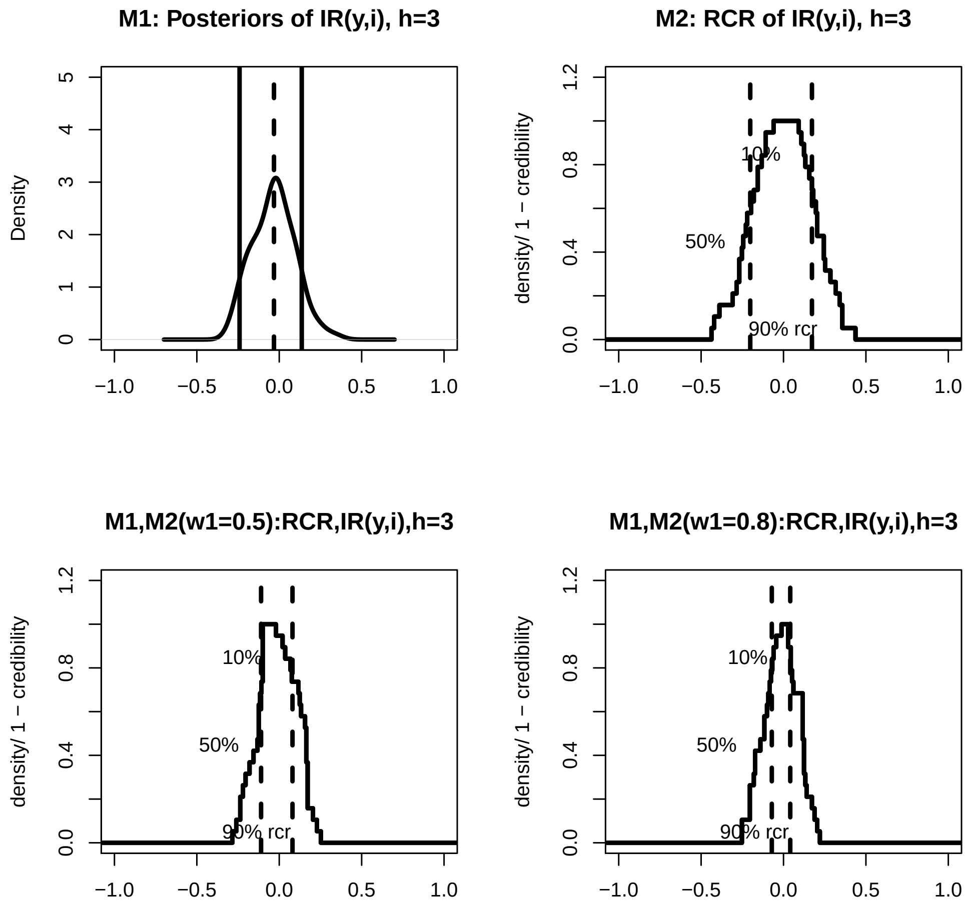

Figure 1: Density and Robust Credible Region of Output Impulse Responses

Note: Output Impulse Response at horizon $h = 3$. For set-identified models, step lines represent the Robustified Credible Region (RCR) at different credibility levels (90%, 50%, 10% levels are explicitly indicated) as described in the last paragraph of Section 2.1 by modifying (Step 5) of Algorithm 4.1 in Giacomini and Kitagawa (2015). The vertical dashed lines represent the posterior mean bounds. For point-identified models (Model 1 and Model 4 in Figure 3), the vertical solid lines display the standard credible region. In such a case, we reported its posterior density, since the posterior mean bounds collapse to a singleton.

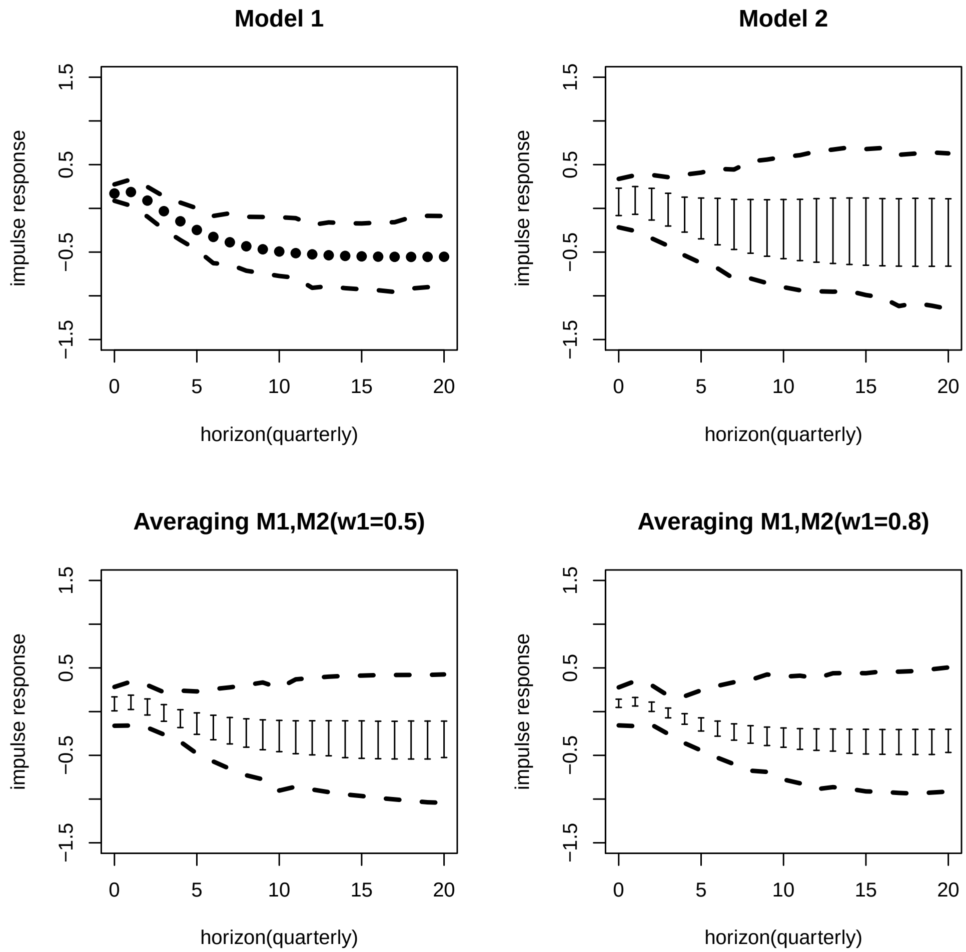

Figure 2: Plots of Output Impulse Responses

Note: for point-identified models, the points plot the (unique) posterior mean and the dashed curve represent the highest posterior density regions with credibility 90%. For set-identified models (Model 2, the averaged models and Model 3 in Figure 4), 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%.

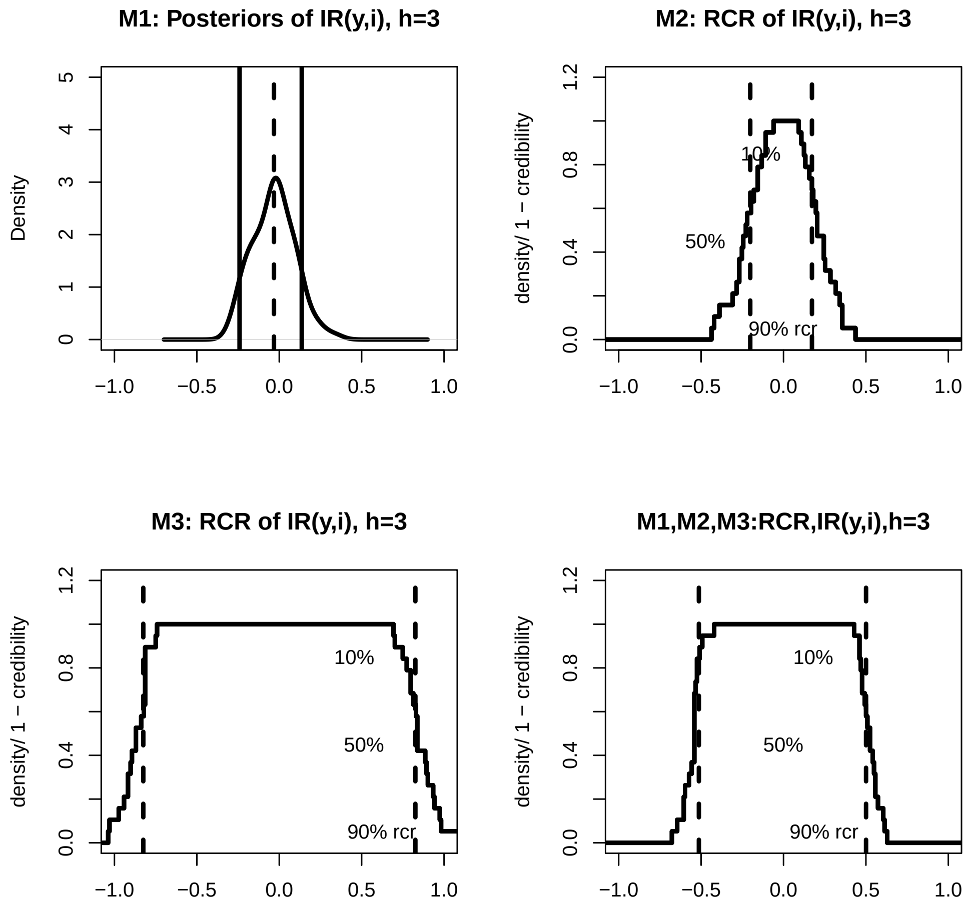

Figure 3: Density and Robust Credible Region of Output Impulse Responses

See the caption of Figure 1 for remarks.

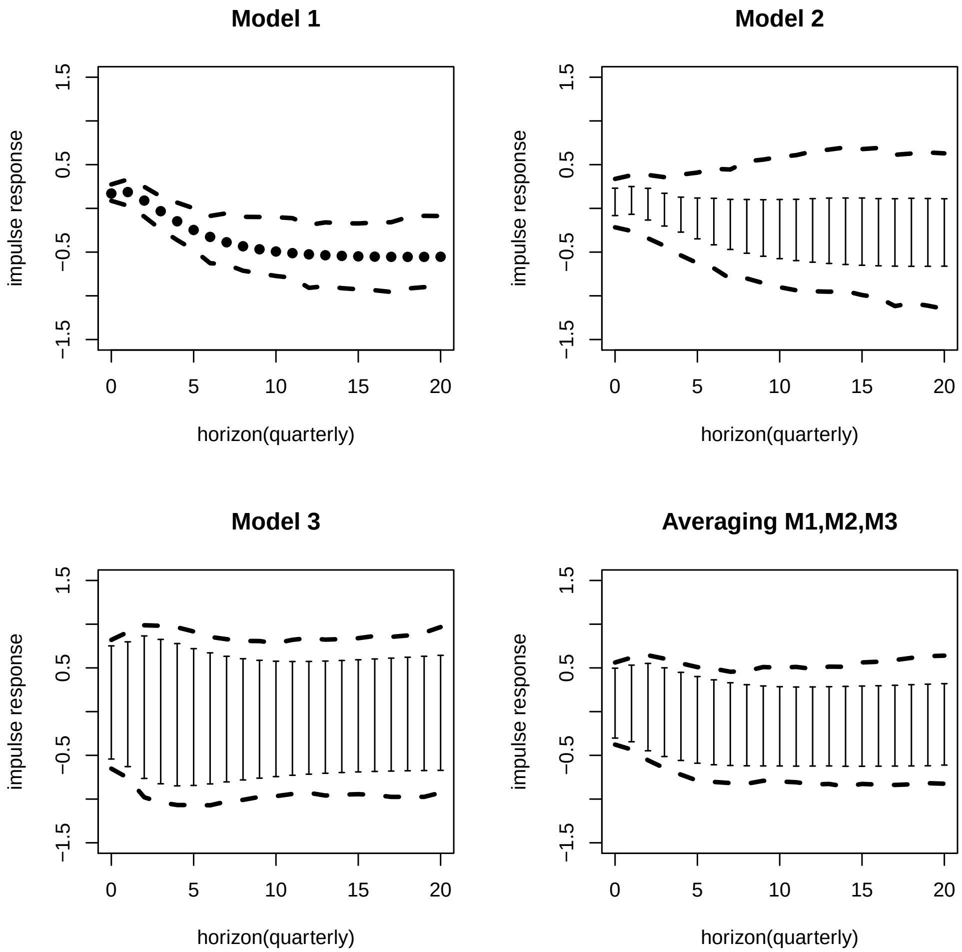

Figure 4: Plots of Output Impulse Responses

See the caption of Figure 2 for remarks.

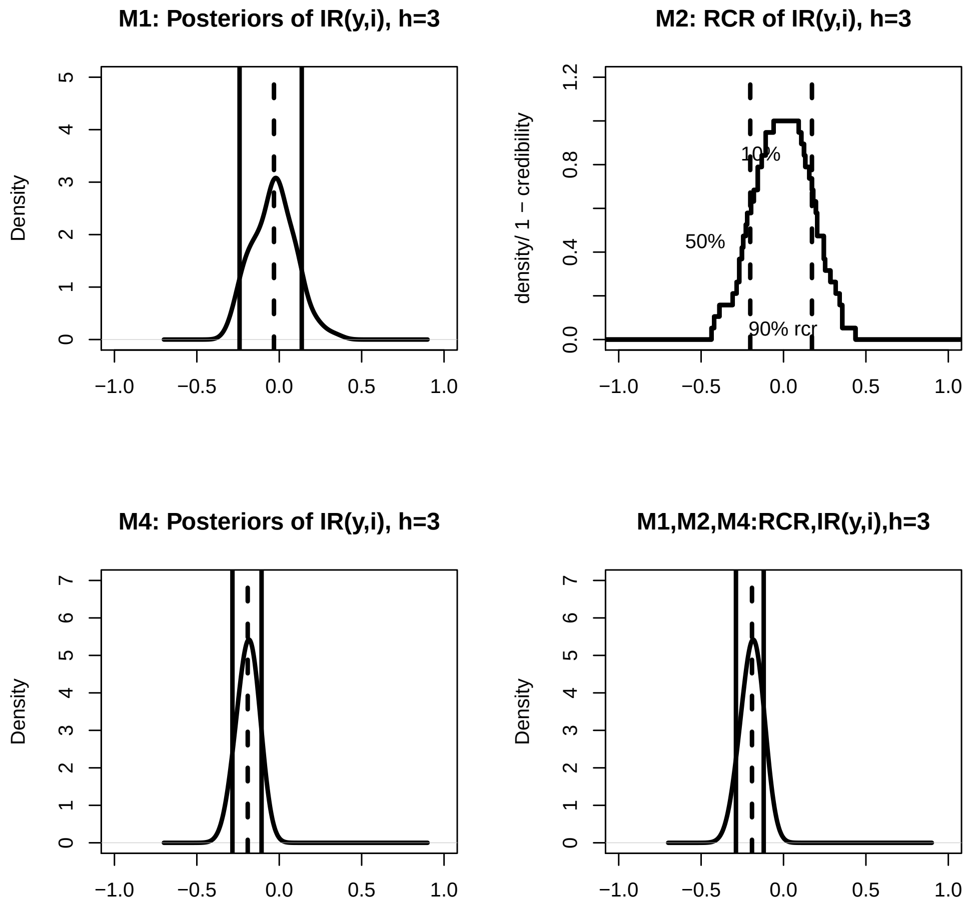

Figure 5: Density and Robust Credible Region of Output Impulse Responses

See the caption of Figure 1 for remarks.

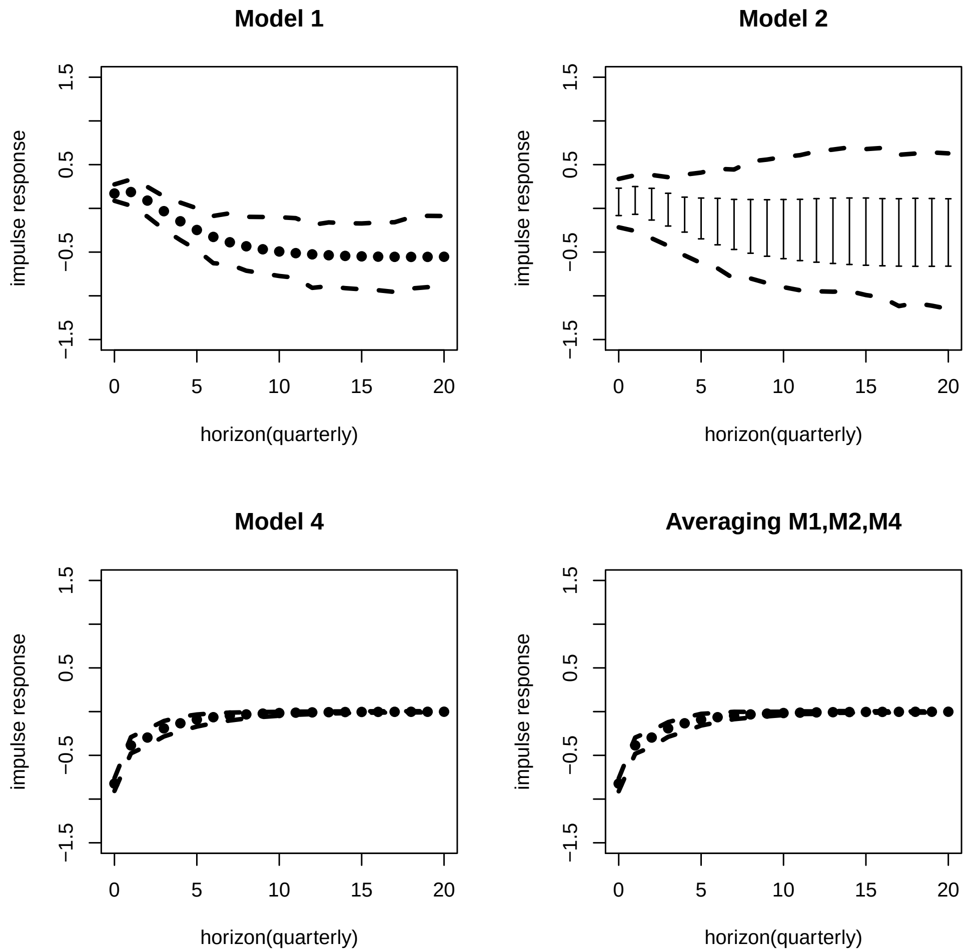

Figure 6: Plots of Output Impulse Responses

See the caption of Figure 2 for remarks.

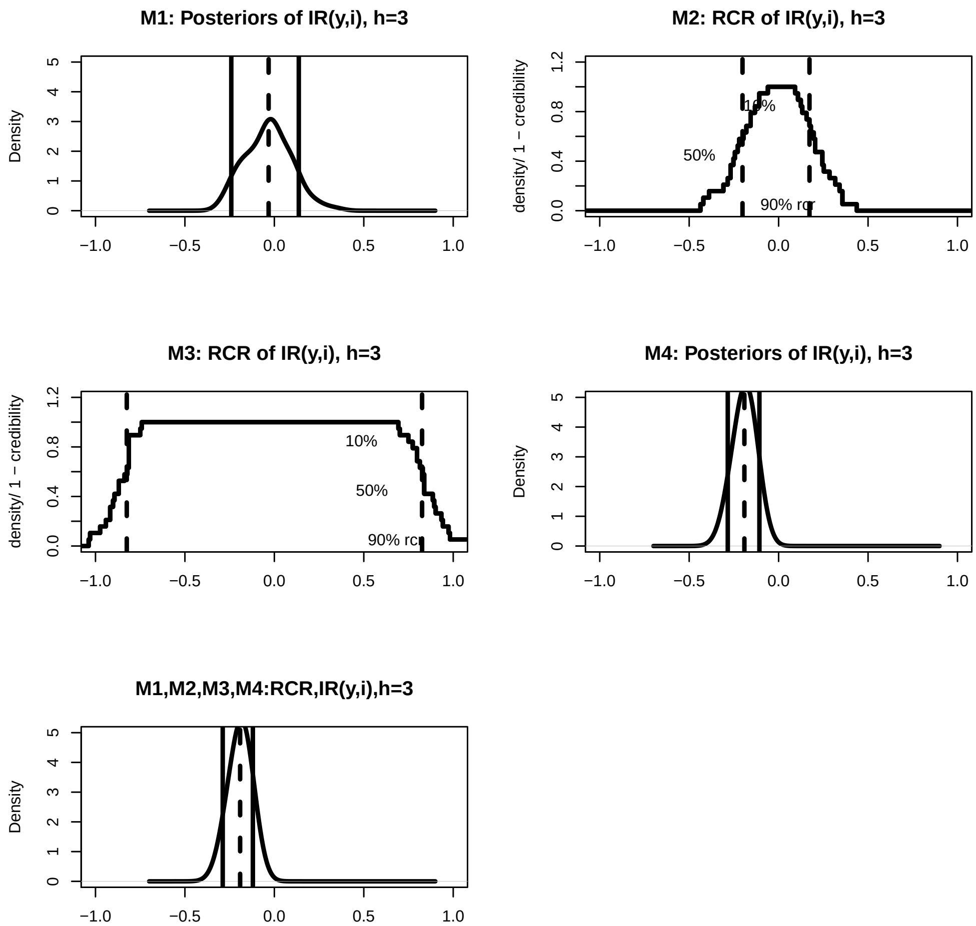

Figure 7: Density and Robust Credible Region of Output Impulse Responses

See the caption of Figure 1 for remarks.

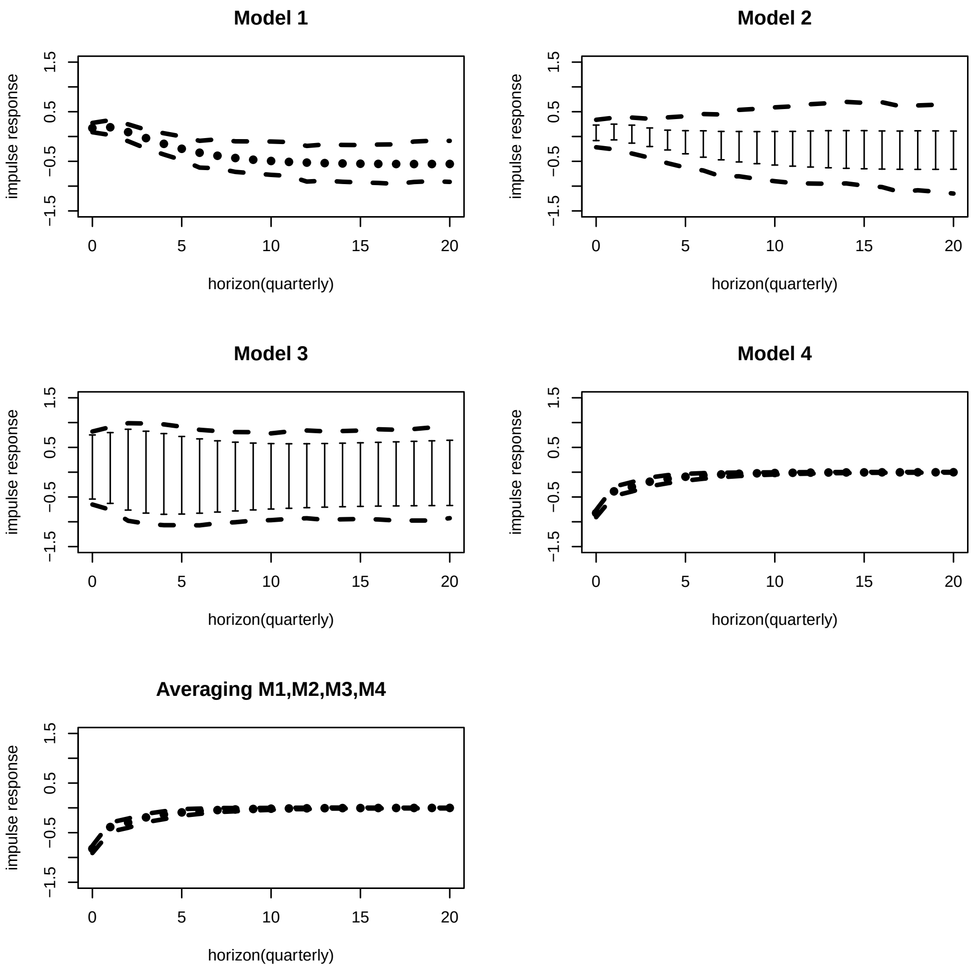

Figure 8: Plots of Output Impulse Responses

See the caption of Figure 2 for remarks.

	Averaging M1, M2	Averaging M1,M2	Averaging M1,M2,M3	Averaging M1,M2,M4	Averaging M1,M2,M3,M4
Prior w_1	0.50	0.80	0.33	0.33	0.25
Prior w_2	0.50	0.20	0.33	0.33	0.25
Prior w_3	/	/	0.33	/	0.25
Prior w_4	/	/	/	0.33	0.25
O_1	1	1	1	1	1
O_2	1	1	1	1	1
O_3	/	/	2.69	/	2.69
O_4	/	/	/	1	1
$ln\ \tilde{p}(Y)$	-815.23	-815.23	-815.23	-815.23	-815.23
$ln\ p(Y\|M^1)$	-815.23	-815.23	-815.23	-815.23	-815.23
$ln\ p(Y\|M^4)$	/	/	/	-468.82	-468.82
Posterior w_1^*	0.50	0.80	0.21	0	0
Posterior w_2^*	0.50	0.20	0.21	0	0
Posterior w_3^*	/	/	0.58	/	0
Posterior w_4^*	/	/	/	1	1

Table 1: Output Responses: Prior and Posterior Weights

	M1			M2		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	.187	$-.493$	$-.553$	/	/	/
90% CR	$[.030, .331]$	$[-.773, -.101]$	$[-.915, -.087]$	/	/	/
Post. Mean Bounds	/	/	/	$[-.068, .250]$	$[-.575, .102]$	$[-.661, .110]$
90% robustified CR	/	/	/	$[-.259, .379]$	$[-.901, .588]$	$[-1.150, .629]$
	M3			M4		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	$-.378$	$-.027$	$-.003$
90% CR	/	/	/	$[-.509, -.261]$	$[-.076, -.003]$	$[-.915, -.087]$
Post. Mean Bounds	$[-.629, .798]$	$[-.743, .577]$	$[-.672, .644]$	/	/	/
90% robustified CR	$[-.757, .910]$	$[-.967, .782]$	$[-.927, .967]$	/	/	/
	Averaging M1,M2($w1=0.5$)			Averaging M1,M2($w1=0.8$)		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	/	/	/
90% CR	/	/	/	/	/	/
Post. Mean Bounds	$[.036, .181]$	$[-.443, -.125]$	$[-.508, -.135]$	$[.077, .153]$	$[-.391, -.216]$	$[-.447, -.233]$
90% robustified CR	$[-.172, .341]$	$[-.809, .336]$	$[-1.004, .467]$	$[-.162, .363]$	$[-.758, .445]$	$[-.877, .516]$
	Averaging M1,M2,M3			Averaging M1,M2,M4		
	$h=1$	$h=10$	$h=20$	$h=1$	$h=10$	$h=20$
Post. Mean	/	/	/	$-.378$	$-.027$	$-.003$
90% CR	/	/	/	$[-.509, -.261]$	$[-.076, -.003]$	$[-.915, -.087]$
Post. Mean Bounds	$[-.384, .568]$	$[-.639, .326]$	$[-.620, .365]$	/	/	/
90% robustified CR	$[-.477, .646]$	$[-.817, .520]$	$[-.819, .699]$	/	/	/
	Averaging M1,M2,M3,M4					
	$h=1$	$h=10$	$h=20$			
Post. Mean	$-.378$	$-.027$	$-.003$			
90% CR	$[-.509, -.261]$	$[-.076, -.003]$	$[-.915, -.087]$			
Post. Mean Bounds	/	/	/			
90% robustified CR	/	/	/			

Table 2: Output Responses: Estimation and Inference