Averaging Point- and Set-identified Models*

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

This draft: September 2015

Abstract

Researchers in causal studies are often unsure about which identifying assumptions to impose among a set of plausible ones. Such uncertainty about the identifying assumptions often leads to the model uncertainty over point- and set-identified models; a combination of plausible identifying assumptions point-identifies the object of interest, while some other combination can only set-identify it. This paper proposes a novel and simple method to cope with this type of model uncertainty by generalizing Bayesian model averaging. For the point-identified model, we treat the standard Bayesian posterior as the representation of the posterior belief, while as an innovative feature of our averaging method, for the set-identified model, we use the set of posteriors (ambiguous belief) as a representation of the posterior belief. Our averaging procedure combines the probabilistic belief from the point-identified model with the ambiguous belief from the set-identified model, and outputs the set of posteriors (post-averaging ambiguous belief) that consists of a mixture of the posterior distribution in the point-identified model and any one of those belonging to the class of posteriors in the set-identified model, with the mixture weights being the posterior probabilities over the models. We propose to summarize this post-averaging ambiguous belief by the range of posterior means and the robustified credible regions, and offer an simple-to-implement algorithm to compute these quantities. We argue that if the point-identified model comes from a reasonable benchmark model, our averaging procedure can be used as a simple and flexible tool to inject additional identifying information into the set-identified model. The method covers the cases with multiple point-identified models and multiple set-identified models, and they can be non-nested. We apply the method to an impulse response analysis of structural vector autoregressions, where a point-identified model relies on a particular causal ordering among endogenous variables and a set-identified model makes use of sign restrictions.

*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

[§]University College London, Department of Economics. Email: alessio.volpicella.14@ucl.ac.uk

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

1 Introduction

Estimation of causal impact of a policy crucially relies on the researcher's choice of identifying assumptions. In observational studies, however, the researcher often faces uncertainty about which identifying assumptions to impose among a set of plausible ones. Such uncertainty about the identifying assumptions commonly leads to the model uncertainty over point- and set-identified models, since a combination of plausible identifying assumptions may point-identify the object of interest, while some other combination can only set-identify it. For instance, when an empirical macroeconomist conducts impulse response analysis using structural vector autoregressions (SVAR), there are wide variety of ways to specify identifying assumptions. These include the classical causal ordering restrictions (Bernanke (1986) and Sims (1980)), restrictions on the long-run responses (Blanchard and Quah (1993)), and sign restrictions on the structural parameters and/or the impulse responses (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)). A particular causal ordering assumption, which may be only partially credible to the researcher, can deliver point-identification of the impulse response, while a combination of sign restrictions, which may be also only partially credible to her, typically leads to the set-identified impulse responses. In the context of program evaluation study in microeconometrics, the unconfoundedness assumption (selection on observables) point-identifies the population average causal effect, while replacing the often controversial unconfoundedness assumption with a combination of the monotone treatment response and monotone selection response assumptions proposed in Manski and Pepper (2000) delivers only set-identification of the average causal effects. In the current empirical practice, what is commonly done is either to report the estimation result exclusively from a most credible model or to report the estimation results of a few candidate models parallelly without making an effort to aggregate them.

How should we draw statistical inference for the object of interest in the presence of uncertainty over point- and set-identified models? This paper proposes a solution to this practically important question by proposing a generalization of the Bayesian model averaging. In the standard Bayesian model averaging, the posterior inference is drawn by mixing the posterior distributions of the candidate models with respect to the posterior model probabilities. This principle could be certainly implemented in the current context if one could come up with a single posterior distribution in every candidate model including those that lack identification. Availability of the single posterior in a non-identified model, however, is arguable, since it demands the analyst to become extremely careful in specifying the prior for the non-identified models due to the fact that the posterior becomes highly sensitive to the prior irrespective of the sample size. In particular, choosing the prior can become a quite challenging task if one's available prior knowledge is well exhausted by her choice of identifying assumptions. Our averaging proposal, in contrast, does not assume availability of the single posterior for the set-identified models. The key innovation of our averaging method is to introduce the set of priors or, equivalently, ambiguous belief to Bayesian model averaging. Specifically, we introduce the set of posteriors as a representation of the posterior belief in the set-identified model. We then combine the single posterior probabilistic belief obtained from the point-identified model with the ambiguous belief obtained from the set-identified model, and output the set of posteriors (*post-averaging ambiguous belief*) that consists of a mixture of the posterior distribution in the point-identified model and any one of those belonging to the set of posteriors in the set-identified model, with the mixture weights being the posterior model probabilities. As a way to summarize and visualize the post-averaging ambiguous belief, we recommend to report the range of posterior means and the lower bound of the posterior probabilities (*lower probability*). We show that these quantities have analytically simple expressions and they are easy to compute in practice using Markov chain Monte Carlo draws of the reduced form parameters (parameters that index the distribution of data).

Our motivation of introducing ambiguous belief (rather than a single prior) in the candidate set-identified model is consistent with the way that the partial identification literature interprets the identified set and the way that the robust Bayes literature introduces the "lack of knowledge". As argued in detail by Giacomini and Kitagawa (2015) and Kitagawa (2012), introducing the ambiguous belief to unrevisable part of prior knowledge (i.e., prior for non-identified parameters) and operating a posterior inference upon the resulting class of posteriors offer us a profound subjective probability-based justification for focusing on the posterior distribution of the identified set, as suggested by (Kline and Tamer (2013), Moon and Schorfheide (2011), and Liao and Simoni (2013)). This means, given a cadidate set-identified model, if the researcher cannot add any additional prior information about what parameter values are more credible than the others within the identified set, then representing the posterior knowledge by the set of posteriors provides an honest and precise description of her/his posterior knowledge on the object of interest.

The framework in this paper assumes that the every candidate model has the same reduced-form model, in the sense that the distribution of data in each model is indexed by the same reduced form parameters $\phi \in \Phi$. For example, in the context of SVARs, this assumption implies that the number of endogenous variables and lags are same for all the candidate models so that they yield the identical reduced form VARs. We assume that an object of interest r is scalar and that its interpretation is common among the models. In a point identified model M^p, the knowledge of ϕ pins down r, and hence $r = r(\phi|M^p)$ is a function of ϕ only. In set-identified model M^s, r can be represented as $r(\theta, \phi)$ a function of ϕ and auxiliary parameters $\theta \in \Theta$, where θ is a non-identified parameter(s) that is necessary to pin down a value of r (e.g. in SVAR, the orthogonal matrix that rotates the impulse response matrix). The identified set for r in model M^s, $IS_r(\phi|M^s)$, is the set-valued map of ϕ defined by the range of $r(\theta, \phi)$ spanned by the set of θ's that satisfy the imposed identifying restrictions. To implement our averaging procedure, prior inputs that the user must specify are a prior distribution for the reduced form parameters ϕ and prior weights over the candidate models (prior model probabilities). We constrain the prior for ϕ to be common across the candidate models (if some models are observationally restrictive, the support of ϕ prior in these models is trimmed to the set of compatible ϕ's without changing the kernel of the prior). Indeed, the prior for ϕ can be updated by data to give its posterior. In contrast, whether the prior model probabilities can be updated by data or not depends on refutability property of the candidate models. Under the assumption that the priors for the reduced form parameters are mutually consistent among the models, we derive a simple formula for the posterior model probabilities and show that the prior model probabilities are updated by data if some candidate set-identified models can generate an empty identified set with a positive probability in terms of the posterior of ϕ.

The range of posterior means for the post-averaging ambiguous belief has the following expression. Let $[l(\phi|M^s), u(\phi|M^s)]$ be the convex hull of the identified set $IS_r(\phi|M^s)$, and denote the posterior model probabilities for the point- and set-identified models by $\pi_{M^p|Y}$ and $\pi_{M^s|Y}$, respectively, where Y denotes the sample and $\pi_{M^p|Y} + \pi_{M^s|Y} = 1$. We show that the range of posterior means of the post-averaging ambiguous belief is given by

$$\pi_{M^p|Y} E_{\phi|Y}(r(\phi|M^p)) + \pi_{M^s|Y} \left[E_{\phi|Y,\tilde{\Phi}}(l(\phi|M^s)), E_{\phi|Y,\tilde{\Phi}}(u(\phi|M^s)) \right], \tag{1.1}$$

where $a + b[c, d]$ means $[a + bc, a + bd]$, $E_{\phi|Y}(\cdot)$ is the posterior expectation in ϕ, and $E_{\phi|Y,\tilde{\Phi}}$ is the posterior expectation in ϕ conditional on that $\phi \in \tilde{\Phi} \equiv \{\phi : IS_r(\phi|M^s) \neq \emptyset\}$. Note that $E_{\phi|Y}(r(\phi|M^p))$ were the posterior mean that one would report if she would only considered the point-identified model, and $\left[E_{\phi|Y,\tilde{\Phi}}(l(\phi|M^s)), E_{\phi|Y,\tilde{\Phi}}(u(\phi|M^s)) \right]$ were the range of posterior means that she would report if she would only considered the set-identified model (Giacomini and Kitagawa (2015)). That is, the mean bounds (1.1) can be seen as a weighted average of the point and the interval estimates in each of the two models. That is, what our averaging effectively does is to shrink the identified set estimate in model M^s toward the point estimate in model M^p. The degree of this shrinkage is governed by the posterior model probabilities. This shrinkage interpretation of the posterior mean bounds suggest that if the point-identified model can be viewed as a useful benchmark model, our averaging procedure offers a simple and flexible way to bring in the additional identifying information to the set-identified model.

The remainder of the paper is organized as follows. Section 2 overviews the main results of this paper and implementation of our averaging method through a simple example of two-variable SVAR. Section 3 presents a formal analysis of our averaging method in a general framework and provides computational algorithm to implement the procedure. In Section 4, using the data from Aruoba and Schorfheide (2011), we apply our method to impulse response analysis to a monetary policy shock based on a 4-variable SVAR.

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). In the current literature of model averaging, there are mainly two approaches. One is Bayesian model averaging (see e.g., Hoeting, Madigan, Raftery, and Volinsky (1999) and Hjort and Claeskens (2008) for a review of the literature). Another is frequentist model averaging, whose important developments can be found in Hansen (2007) and Hjort and Claeskens (2003). See also Hansen (2014), Liu (2015), Liu and Okui (forthcoming), Hansen and Racine (2012), and Zhang and Liang (2011) for recent advancements of the frequentist model averaging, and Hjort and Claeskens (2003), Kitagawa and Muris (2015), and Magnus, Powell, and Prüfer (2010) for proposals to compromise the Bayesian and frequentist averaging. None of these works considers averaging the point- and set-identified models. We tackle this problem from the angle of Bayesian model averaging, since it is not at all obvious how to define the frequentist risk criterion in the current problem. Unlike the standard Bayesian model avearaging, however, our averaging introduces amibugous belief and does not require the full prior specification in the set-identified models. To the best of our knowledge, this is the first paper that formally considers averaging of the probabilistic belief and the ambiguous belief.

Another strand of literature that this paper contributes is the growing literature on Bayesian inference for partially identified models (Giacomini and Kitagawa (2015), Kitagawa (2012), Kline and Tamer (2013), Moon and Schorfheide (2012), Norets and Tang (2014), Liao and Simoni (2013)). This paper follows the robust Bayes approach with multiple priors proposed by Giacomini and Kitagawa (2015) and Kitagawa (2012) to model the lack of belief within the identified set. 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 (2013), Liao and Simoni (2013), and Moon and Schorfheide (2011). When model uncertainty is present, however, the identified set cannot be well defined without conditioning on the model. The multiple prior approach enjoys an advantage in this case since the range of posteriors is still a well-defined object and its subjective interpretation remains the same as in the case with the single set-identified model.

The empirical application in this paper concerns SVAR analysis with sign restrictions (Canova and Nicolo (2002), Faust (1998), and Uhlig (2005)). In many SVAR applications, what set of identifying restrictions to impose becomes a source of controversies due to the fact that the researchers have different opinions about credibility of the identifying restrictions. The current empirical practice pursues to impose a weakest set of assumptions, which many of them in the field can agree upon, in the form of sign restrictions on the impulse responses or structural parameters. The resulting model generally only set-identifies the impulse responses of interest, while the common practice is to run the Bayesian estimation by putting 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) develops frequentist inference for the identified set and Giacomini and Kitagawa (2015) proposes a robust Bayesian approach. Although averaging is a natural way to pacify the controversies about the identifying restrictions, little work has been done on multi-model inference in the SVAR literature.

2 An Illustrating Example and Overview of Results

We illustrate the analytical framework introduced above and our proposal of averaging using a simple model of two-variable SVAR. Consider

$$A_0 \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad \Leftrightarrow \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = A_0^{-1} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

where (y_{1t}, y_{2t}), $t = 1, \ldots, T$, is a vector of endogenous variables, $(\epsilon_{1t}, \epsilon_{2t})$ is a vector of structural shocks that is independent of the past realizations of (y_{1t}, y_{2t}) and follow iid normal with the covariance matrix being the identity matrix, and A_0 is the matrix of structural coefficients. Since there are no intercept nor lags in this model, the reduced-form parameters consists of only Σ the variance-covariance matrix of (y_{1t}, y_{2t}). 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 $\sigma_{22} \geq 0$. We parametrize the reduced form parameters by $\phi = (\sigma_{11}, \sigma_{12}, \sigma_{22})$.[1] The contemporaneous impulse response

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

matrix is $A_0^{-1} \equiv \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$, and we let the contemporaneous response of the first variable to the unit positive shock in the first variable be the object of interest, $r = r_{11}$. Following Uhlig (2005), we reparametrize the structural coefficients in A_0 via the rotation matrix $Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$ with spherical coordinate $\theta \in [0, 2\pi]$, so that we write r as a function of ϕ and the non-identified parameter θ indexing the rotation matrix. Specifically,

$$A_0^{-1} = \Sigma_{tr} Q = \begin{pmatrix} \sigma_{11} q_{11} & \sigma_{11} q_{12} \\ \sigma_{21} q_{11} + \sigma_{22} q_{21} & \sigma_{21} q_{12} + \sigma_{22} q_{22} \end{pmatrix}$$

leads to $r = r(\theta, \phi) \equiv \sigma_{11} \cos\theta$. We impose the sign normalizations for the structural parameters by constraining the diagonal elements of A_0 to being nonnegative, which deliver

$$\sigma_{22} q_{11} - \sigma_{21} q_{21} \geq 0 \tag{2.1}$$

and

$$\sigma_{11} q_{22} \geq 0. \tag{2.2}$$

Suppose that the researcher's prior belief about the identifying restriction is given as follows:

Prior belief about identifying restrictions

- *Model M^p (point-identified)*: With prior weight $w \in (0, 1)$, she/he believes y_{1t} causally precedes y_{2t}, so that the causal ordering assumption such that $(1, 2)$-entry of A_0^{-1} matrix $\sigma_{11} q_{12}$ is equal to zero.
- *Model M^s (set-identified)*: With prior weight $(1 - w)$, she/he is lack of the belief about the causal ordering between (y_{1t}, y_{2t}).

The prior weight w specified by the researcher represents her/his degree of credibility about the causal ordering assumption. The alternative model M^s corresponds to the "no assumption" model where the causal ordering assumption is dropped, and the impulse response is only set-identified. Since there are only two candidate models, model M^s receives prior weight $1 - w$.

Under model M^p, the zero restriction and the sign normalization restrictions pin down Q uniquely to $Q = I$, and point-identified r is given by $r(\phi | M^p) \equiv \sigma_{11}$. Under model M^s, the identified set for r is characterized solely by the sign normalization restrictions (2.1) and (2.2), and it is given by a connected interval,

$$IS_r(\phi | M^s) = [l(\phi | M^s), u(\phi | M^s)] \equiv \left[-\sqrt{\frac{\sigma_{21}^2}{\sigma_{21}^2 + \sigma_{22}^2}} \sigma_{11}, \sigma_{11} \right].$$

Note that this identified set is non-empty for any value of ϕ. In this particular case, the point-identified r in model M^p is the upper-bound of the identified set in model M^s, i.e., $r(\phi|M^p) = u(\phi|M^s)$ for all ϕ.

We now describe implementation of our averaging procedure. Given the candidate models, we next specify $\pi_{\phi|M}$ a prior distribution (probability density) for the reduced form parameters ϕ conditional on each model $M \in \{M^p, M^s\}$. Specifically, we introduce the following model-conditioned priors for ϕ:

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

$$\pi_{\phi|M^s}(\phi) = \frac{\tilde{\pi}_\phi(\phi) 1\{IS_r(\phi|M^s) \neq \emptyset\}}{\tilde{\pi}_\phi(IS_r(\phi|M^s) \neq \emptyset\}} \propto \tilde{\pi}_\phi(\phi) 1\{IS_r(\phi|M^s) \neq \emptyset\}, \tag{2.4}$$

where $\tilde{\pi}_\phi(\phi)$ is a *proper* prior distribution for ϕ, which in the current context, can be the one induced from a Wishart prior of Σ, and, with abuse of notation, we denote the prior probability of having an empty identified set in terms of $\tilde{\pi}_\phi$ by $\tilde{\pi}_\phi(IS_r(\phi|M^s) \neq \emptyset\} \equiv \int_\Phi 1\{IS_r(\phi|M^s) \neq \emptyset\}\tilde{\pi}_\phi(\phi)d\phi$. The prior for ϕ in model M^s is the same as the one of model M^p except that its support is restricted to those ϕ's that yield nonempty identified sets. In the current specific example, an empty identified set never occurs so that $\pi_{\phi|M^p} = \pi_{\phi|M^s}$ holds.

In model M^p, the Bayes theorem applied to the prior for ϕ given in (2.3) yields the posterior for r, since under point-identification, $r = r(\phi|M^p)$ is a function of ϕ. We denote the posterior of r in model M^p by $\pi_{r|M^p,Y}(r)$.

In contrast, in model M^s, unless one further specifies a conditional prior of θ given ϕ, only the prior for ϕ cannot lead to a unique posterior for r since $r = r(\theta, \phi)$ depends also on non-identified parameter θ. By adopting the way that Giacomini and Kitagawa (2015) and Kitagawa (2012) specify the lack of prior knowledge about θ (this is the situation meant in model M^s), our averaging procedure introduces arbitrary conditional priors of θ given ϕ. This results in the following class of posteriors for r,

$$\Pi_{r|M^s,Y} \equiv \left\{ \pi_{r|M^s,Y}(r) = \int_\Phi \pi_{r|M^s,\phi}(r)\pi_{\phi|M^s,Y}(\phi)d\phi : supp(\pi_{r|M^s,\phi}) \subset IS_r(\phi|M^s), \forall \phi \in \tilde{\Phi} \right\}, \tag{2.5}$$

where $supp(\pi_{r|M^s,\phi})$ stands for the support of the conditional distribution of r given ϕ (induced by a conditional prior of θ given ϕ), and $\tilde{\Phi} = \{\phi \in \Phi : IS_r(\phi|M^s) \neq \emptyset\}$, which contains the support of $\pi_{\phi|M^s,Y}(\phi)$ by the construction of prior (2.4). We use the class of posteriors shown in (2.5) as a representation of the posterior belief for r in the set-identified model. This is an important departure from the standard Bayesian model averaging, as the latter does not allow for any form of ambiguity, and forces one to have a single posterior distribution for every candidate model.

Other components that one needs to know to implement out averaging scheme are the posterior model probabilities denoted by $\pi_{M_{p|Y}}$ and $\pi_{M_{s|Y}}$. To derive them, let us define the *posterior-prior odds ratio for plausibility of the identifying assumptions* for set-identified model M^s:

$$O_{M^s} \equiv \frac{\tilde{\pi}_{\phi|Y}(IS_r(\phi|M^s) \neq \emptyset)}{\tilde{\pi}_{\phi}(IS_r(\phi|M^s) \neq \emptyset)},$$

where the numerator in O_{M^s} is the *posterior* probability of having an empty identified set with the posterior for ϕ obtained from the Bayesian updating of $\tilde{\pi}_\phi$. An easy-to-verify but remarkable result shown in Lemma 3.1 below says that when a prior for ϕ in each model is specified as in (2.3) and (2.4), the posterior model probabilities are obtained as

$$\pi_{M^p|Y} = \frac{w}{w + O_{M^s} \cdot (1-w)} \text{ and } \pi_{M^s|Y} = \frac{O_{M^s} \cdot (1-w)}{w + O_{M^s} \cdot (1-w)}. \tag{2.6}$$

Note that in the current example, model M^s is not observationally restrictive in the sense that $IS_r(\phi|M_{s)}$ is non-empty for all ϕ. Hence, the posterior-prior odds ratio O_{M^s} is one and the formula of the posterior model probabilities (2.6) yields $(\pi_{M^p|Y}, \pi_{M^s|Y}) = (w, 1-w)$, i.e., no update occurs for the prior model probabilities.

Combining the posterior belief for r in each model with the posterior model probabilities obtained in (2.6), we form the following set of averaged posteriors, which we refer to as *post-averaging set of posteriors*:

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

$\Pi_{r|Y}$ is a set of mixture distributions with the mixture weight set at the posterior model probabilities $(\pi_{M^p|Y}, \pi_{M^s|Y})$, which equal to $(w, 1-w)$ in the current example, and the component distribution corresponding to model M^p fixed at $\pi_{r|M^p,Y}$, whereas a component distribution corresponding to model M^s ranges over the posterior class $\Pi_{r|M^s,Y}$ shown in (2.5).

To summarize the post-averaging set of posteriors, we suggest to report the range of posterior means of $\Pi_{r|Y}$ and the robustified credible regions with credibility $\alpha \in (0,1)$ defined by the shortest interval that receives posterior probability at least α for every posterior in $\Pi_{r|Y}$ (Giacomini and Kitagawa (2015)). The mixture representation given in (2.7) leads to the range of posterior means as

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

where $a + b[c, d]$ stands for $[a + bc, a + bd]$ and $E_{\phi|M^s,\tilde{\Phi}}(\cdot)$ denotes the posterior mean with respect to $\pi_{\phi|M^s,Y} = \frac{\tilde{\pi}_{\phi|Y} \cdot 1\{IS_r(\phi|M^s) \neq \emptyset\}}{\tilde{\pi}_{\phi|Y}(IS_r(\phi|M^s) \neq \emptyset)}$. This expression of the range posterior means is intuitive and simple to interpret: the range of averaged posterior means is the weighted average (Minkowski sum) of the point corresponding to the posterior mean in model M^p and the interval corresponding to the range of posterior means in model M^s. Hence, in words, our model averaging procedure can be viewed as a method to shrink the identified set toward the point estimate in the point-identified model with exploiting the belief assigned to the point-identified model. The degree of this shrinkage is governed by the posterior model probabilities, which can be updated if the set-identified model is observationally restrictive, and otherwise, they coincide with the prior model probabilities.

Following Giacomini and Kitagawa (2015), a robustified credible region for r with credibility α constructed upon the set of posteriors (2.7) is an interval C_α such that $\pi_{r|Y}(r \in C_\alpha) \geq \alpha$ holds for all $\pi_{r|Y} \in \Pi_{r|Y}$. A robustified credible region with the shortest width can be computed by modifying (Step 5) of Algorithm 4.1 in Giacomini and Kitagawa (2015). We first draw $z_1, \ldots, z_L$ randomly from the Bernoulli distribution with mean $\pi_{M^p|Y}$. We then generate $l = 1, \ldots, L$ random draws of the "mixture identified set" for r by

$$IS_r^{avg}(\phi_l) = \begin{cases} \{r(\phi_l|M^p)\}, \ \ \phi_l \sim \tilde{\pi}_{\phi|Y}, & \text{if } z_l = 1, \\ [l(\phi_l|M^s), u(\phi_l|M^s)], \ \ , \phi_l \sim \tilde{\pi}_{\phi|Y,\tilde{\Phi}} & \text{if } z_l = 0. \end{cases} \tag{2.8}$$

That is, with probability $\pi_{M^p|Y}$, a draw of the "mixture identified set" for r is a singleton set corresponding to the point-identified model, and with the other probability $\pi_{M^s|Y}$, a draw of the "mixture identified set" is a nonempty identified set in the set-identified model. Based on these mixture draws, the minimization problem presented in Step 5 of Algorithm 4.1 in Giacomini and Kitagawa (2015) yields the shortest-width robustified credible region.

3 Robust Bayes Averaging: General Case

This section provides formal analytical claims that justify the procedures presented in the illustrating example in a general framework.

3.1 Analytical results

Suppose that one faces model uncertainty among $K + J$ models, $K, J \geq 1$, that differ each other in terms of the identifying assumptions. The K distinct models correspond to point-identified and observationally non-restrictive models (i.e., over-identified models are ruled out). We denote them by $\mathcal{M}^p \equiv \{M_k^p : k = 1, \ldots, K\}$. The J distinct models correspond to the set-identified models, which we denote by $\mathcal{M}^s \equiv \{M_j^s : j = 1, \ldots, J\}$. Let $\mathcal{M} \equiv \mathcal{M}^p \cup \mathcal{M}^s$ denote the whole collection of candidate models to be averaged. We assume that all the candidate models have the identical reduced-form in the sense that the dimension and the definition of the reduced form parameters ϕ are common across all the models. This implies that among the point identified models $M_k^p \in \mathcal{M}^p$, the likelihood function $p(Y|\phi, M_j^p)$ does not depend on $M_j^p \in \mathcal{M}^p$. The object of interest $r \in \mathbb{R}$ is a scalar quantity, and it carries the same interpretation in every model. Let $r(\phi|M_k^p)$, $k = 1, \ldots, K$, be a real-valued function that returns the value of r in point-identified model M_k^p when a value of reduced form parameters is set at ϕ. For the set-identified models, we denote by $IS_r(\phi|M_j^s) \subset \mathbb{R}$, $j = 1, \ldots, J$, the identified set for r in model M_j^s, which is viewed as a set-valued map from ϕ to the real line. Some of the set-identified models can be observationally restrictive in the sense that $IS_r(\phi|M_j^s)$ is empty for some value of ϕ. When model $M_j^s \in \mathcal{M}^s$ is observationally restrictive in this sense, the value of the likelihood $p(Y|\phi, M_j^s)$ at any ϕ yielding a nonempty identified set equals to that of the likelihood in a point-identified model, i.e., $p(Y|\phi, M_j^s) = p(Y|\phi, M_k^p) \equiv p(Y|\phi)$ for all ϕ such that $IS_r(\phi|M_j^s) \neq \emptyset$.

The next set of assumptions specifies the regularity conditions for our averaging procedure.

Assumption 3.1 *Let $\tilde{\pi}_\phi$ be a measure defined on the space of reduced-form parameters Φ. With abuse of notation, $\tilde{\pi}_\phi(\phi)$ denotes the density function (the Radon-Nykodim derivative) of $\tilde{\pi}_\phi$ with respect to the Lebesgue measure on Φ.*

(i) $\tilde{\pi}_\phi$ is proper.

(ii) The point identified models in $\mathcal{M}^p$ are not observationally restrictive relative to the reduced-form parameter space Φ, in the sense that $r(\phi|M^p)$ is well defined for all $\phi \in \Phi$, $\tilde{\pi}_\phi$-a.s., and $\tilde{\pi}_\phi(IS_r(\phi|M_j^s) \neq \emptyset) > 0$ for all set-identified models $M_j^s \in \mathcal{M}^s$,

(iii) Prior model weights $(\pi_M : M \in \mathcal{M})$ are nonnegative and sum up to one (probability weights).

(iv) A prior for ϕ in each model has the following probability density:

$$\pi_{\phi|M}(\phi) = \begin{cases} \tilde{\pi}_\phi(\phi) & \text{for } M \in \mathcal{M}^p, \\ \frac{\tilde{\pi}_\phi(\phi) 1\{IS_r(\phi|M_j^s) \neq \emptyset\}}{\tilde{\pi}_\phi(IS_r(\phi|M_j^s) \neq \emptyset)} & \text{for } M \in \mathcal{M}^s, \end{cases} \tag{3.1}$$

Assumption 3.1 (i) assumes that the base measure for ϕ that generates the priors for ϕ in each model according to (3.1) is a probability measure. The specification for $\tilde{\pi}_\phi$ is an important input in our analysis, while its influence to posterior inference is less significant when the reduced form parameters are accurately estimated by the likelihood. Assumption 3.1 (ii) concerns the properties of the candidate models. It assumes that all the point-identified models are just-identified and do not constrain the distribution of data. On the other hand, it assumes that the set-identified models can be observationally restrictive and can have empty identified set for some but not all ϕ. Assumption 3.1 (iv) assumes that the prior for ϕ in each model is common across the models except for its support. The prior specification (3.1) says, for a refutable set-identified model, the support of the prior for ϕ is constrained to those ϕ's that yield non-empty identified set.

When the priors for ϕ satisfy Assumption 3.1, the posterior model probabilities have the following simple expressions.

Lemma 3.1 *Suppose that prior model probability* $(\pi_M : M \in \mathcal{M})$ *and a prior of* ϕ *in each model satisfy Assumption 3.1. Define the posterior-prior odds of model plausibility of set-identified model* $M \in \mathcal{M}^s$ *by*

$$O_M \equiv \frac{\tilde{\pi}_{\phi|Y}(IS_r(\phi|M) \neq \emptyset)}{\tilde{\pi}_{\phi}(IS_r(\phi|M) \neq \emptyset)}. \tag{3.2}$$

where $\tilde{\pi}_{\phi|Y}$ *be the posterior of* ϕ *obtained from prior* $\tilde{\pi}_{\phi}$*. Then, the posterior model probability is given by*

$$\pi_{M|Y} = \begin{cases} \frac{\pi_M}{\sum_{M \in \mathcal{M}^p} \pi_M + \sum_{M \in \mathcal{M}^s} O_M \pi_M}, & \text{for } M \in \mathcal{M}^p, \\ \frac{O_M \pi_M}{\sum_{M \in \mathcal{M}^p} \pi_M + \sum_{M \in \mathcal{M}^s} O_M \pi_M}, & \text{for } M \in \mathcal{M}^s. \end{cases} \tag{3.3}$$

Proof. By the Bayes theorem, the posterior model probability of M satisifes

$$\pi_{M|Y} = \frac{p(Y|M)\pi_M}{\sum_{M \in \mathcal{M}} p(Y|M)\pi_M}, \tag{3.4}$$

where $p(Y|M)$ is the marginal likelihood, $\int_{\phi} p(Y|\phi, M) d\pi_{\phi|M}(\phi)$. Since the reduced-form model is common over $\mathcal{M}$, the likelihood $p(Y|\phi, M)$ does not depend on M. Under Assumption 3.1 (iii), the prior for ϕ for point-identified model $M \in \mathcal{M}$ does not depend on M and is given by $\pi_{\phi|M} = \tilde{\pi}_{\phi}$. Hence, $p(Y|M) = \int_{\phi} p(Y|\phi) d\tilde{\pi}_{\phi}(\phi) \equiv \tilde{p}(Y)$ for all $M \in \mathcal{M}^p$.

For set-identified model $M \in \mathcal{M}^s$, the marginal likelihood can be written as

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

where the second line follows since $p(Y|\phi, M)\tilde{\pi}_\phi(\phi) = \tilde{p}(Y)\tilde{\pi}_{\phi|Y}(\phi)$ holds for any ϕ such that $IS_r(\phi|M) \neq \emptyset$. Plugging the marginal likelihoods of these point- and set-identified models into (3.4) leads to the conclusion. ■

This lemma shows that the model probabilities can be updated by data only when in some candidate set-identified model, the identified set is empty with a positive probability in terms of $\tilde{\pi}_\phi$. Hence, if none of the set-identified models is refutable, the prior and posterior model probabilities coincide.

Given the posterior model probabilities, an averaged posterior for r can be expressed by the following mixture form,

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

where $\pi_{r|M,Y}$ is a posterior of r in model M. For the point identified model, the posterior of ϕ induces unique $\pi_{r|M,Y}$. For the set-identified models, however, having the posterior of ϕ does not suffice to pin down a unique posterior for $r = r(\theta, \phi|M^s)$. Instead, if we admit any posterior distributions of r that are compatible with $\pi_{\phi|Y}$, we end up with the set of posteriors (see Giacomini and Kitagawa (2015) and Kitagawa (2012) for details and more precise claims on this). Let us denote the set of posteriors of r in model $M \in \mathcal{M}^s$ by $\Pi_{r|M,Y}$. If there is no restriction that links posteriors in $\Pi_{r|M,Y}$ across different models, the set of averaged posteriors spanned by $\{\Pi_{r|M,Y} : M \in \mathcal{M}^s\}$ can be represented by

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

The next proposition shows the range of posterior means and the lower probability of $\Pi_{r|Y}$.

Proposition 3.1 *Suppose Assumption 3.1 holds. For $M \in \mathcal{M}^s$, let $[l(\phi|M), u(\phi|M)]$ be the convex hull of $IS_r(\phi|M)$. Then, the range of posterior means of $\Pi_{r|Y}$ is the convex interval with the lower and upper bounds given by*

$$\inf_{\pi_{r|Y} \in \Pi_{r|Y}} E_{r|Y}(r) = \sum_{M \in \mathcal{M}^p} E_{r|M,Y}(r)\pi_{M|Y} + \sum_{M \in \mathcal{M}^s} E_{\phi|Y,\tilde{\Phi}_M} l(\phi|M)\pi_{M|Y},$$
$$\sup_{\pi_{r|Y} \in \Pi_{r|Y}} E_{r|Y}(r) = \sum_{M \in \mathcal{M}^p} E_{r|M,Y}(r)\pi_{M|Y} + \sum_{M \in \mathcal{M}^s} E_{\phi|Y,\tilde{\Phi}_M} u(\phi|M)\pi_{M|Y},$$

where $\tilde{\Phi}_M = \{\phi : IS_r(\phi|M) \neq \emptyset\}$ and $E_{\phi|Y,\tilde{\Phi}_M}(\cdot)$ is the expectation with respect to the posterior distribution of ϕ trimmed on $\tilde{\Phi}_M$. Furthermore, for any measurable subsets A in $\mathbb{R}$, the lower *bound for the posterior probability of $\{r \in A\}$ in the class $\Pi_{r|Y}$ is*

$$\inf_{\pi_{r|Y} \in \Pi_{r|Y}} \pi_{r|Y}(r \in A) = \sum_{M \in \mathcal{M}^p} \pi_{r|M,Y}(r \in A) \cdot \pi_{M|Y} + \sum_{M \in \mathcal{M}^s} \tilde{\pi}_{\phi|Y,\tilde{\Phi}_M}(IS_r(\phi|M) \subset A) \cdot \pi_{M|Y}.$$

Proof. Since there is no constraint across the posteriors in $\{\Pi_{r|M,Y} : M \in \mathcal{M}^s\}$, it holds

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

By applying Proposition 4.1 (ii) of Giacomini and Kitagawa (2015), it holds $\inf_{\pi_{r|M,Y} \in \Pi_{r|M,Y}} \left\{E_{r|M,Y}(r)\right\} = E_{\phi|M,Y,\tilde{\Phi}_M}(l(\phi|M))$. By Assumption 3.1 (iv), the posterior distribution of ϕ depends on $M \in \mathcal{M}^s$ only through the support $\tilde{\Phi}_M$, so that $E_{\phi|M,Y,\tilde{\Phi}_M}(l(\phi|M)) = E_{\phi|Y,\tilde{\Phi}_M}(l(\phi|M))$ holds. The claim of the mean lower bound therefore follows. The mean upper bound can be shown similarly.

For the lower probability, note that

$$\inf_{\pi_{r|Y} \in \Pi_{r|Y}} \pi_{r|Y}(r \in A) = \sum_{M \in \mathcal{M}^p} \pi_{r|M,Y}(r \in A) \cdot \pi_{M|Y} + \sum_{M \in \mathcal{M}^s} \inf_{\pi_{r|M,Y} \in \Pi_{r|M,Y}} \left\{\pi_{r|M,Y}(r \in A)\right\} \cdot \pi_{M|Y}. \tag{3.5}$$

Proposition 4.1 (i) of Giacomini and Kitagawa (2015) shows $\inf_{\pi_{r|M,Y} \in \Pi_{r|M,Y}} \left\{\pi_{r|M,Y}(r \in A)\right\} = \tilde{\pi}_{\phi|M,Y,\tilde{\Phi}_M}(IS_r(\phi|M) \subset A)$. Again, by Assumption 3.1 (iv), $\tilde{\pi}_{\phi|M,Y,\tilde{\Phi}_M} = \tilde{\pi}_{\phi|Y,\tilde{\Phi}_M}$ holds and the claim follows. ■

3.2 Computing the Range of Post-averaging Posteriors

In order to report the mean bounds of the post-averaging posterior class shown in Proposition 3.1, we need to be able to calculate (I) the prior posterior odds ratio given in (3.2), (II) the posterior mean of r in each point-identified model, and (III) the posterior means of the lower and upper bounds of the identified set in each set-identified model. Since the estimation of a point-identified model is the standard Bayesian inference, any of posterior sampling algorithms can be applied to compute the posterior mean of r in the point-identified model. The other two quantities are less standard in the standard Bayesian inference, and hence this section briefly discusses how to compute them.

If we can assess non-emptiness of the identified set for each given ϕ, 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 (3.2). Specifically, the denominator of O_M is computed by drawing many ϕ's from the prior $\tilde{\pi}_\phi$ and getting the fraction of draws that yield nonempty identified sets, and the numerator of O_M is computed similarly except that ϕ's are drawn from the posterior $\tilde{\pi}_{\phi|Y}$. In the SVAR example of Section 4, we assess non-emptiness of the identified set by drawing many non-identified parameters (rotation matrix) from the uniform distribution (Haar measure on the space of orthonormal matrices) using the sampling algorithm of Uhlig (2005), and checking if any of the draws satisfy the imposed sign restrictions. See also Algorithm 4.1 in Giacomini and Kitagawa (2015).

The computation of the posterior means of the lower and upper bounds of the identified set in model $M \in \mathcal{M}^s$ can be done in the following two steps. In the first step, we draw ϕ's from the posterior $\tilde{\pi}_{\phi|Y}$ and retain those that yield nonempty $IS_r(\phi|M)$. In the second step, for each of the retained draws of ϕ, we compute $l(\phi|M^s)$ and $u(\phi|M^s)$ and take their sample averages. We then use these sample averages as approximations for $E_{\phi|Y,\tilde{\Phi}_M}(l(\phi|M))$ and $E_{\phi|Y,\tilde{\Phi}_M}(u(\phi|M))$. Implementation of this computational algorithm 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 4, we implement the optimization approach considered in Algorithm 4.1 of Giacomini and Kitagawa (2015). When the construction of the identified set employs the criterion function approach of Chernozhukov, Hong, and Tamer (2007), computation of the lower and upper bounds of the identified set can be facilitated by the slice sampling algorithm as proposed by Kline and Tamer (2013).

The analytical expression of the lower probability of the post-averaging ambiguous belief shown in Proposition 3.1 suggests that it can be represented by the *containment functional* of a properly defined mixture of random closed sets. To show this point, define a mixture identified sets generated from a probability measure on the product space of $\mathcal{M} \times \Phi$:

$$IS_r^{avg}(M, \phi) \equiv \begin{cases} \{r(\phi|M)\} & \text{for } M \in \mathcal{M}^p \text{ and } \phi \in \Phi, \\ IS_r(\phi|M) & \text{for } M \in \mathcal{M}^s \text{ and } \phi \in \tilde{\Phi}_M, \\ \emptyset & \text{for } M \in \mathcal{M}^s \text{ and } \phi \notin \tilde{\Phi}_M. \end{cases} \tag{3.6}$$

By setting the probability measure on $\mathcal{M} \times \Phi$ at the posterior joint probability for (M, ϕ), $\pi_{M,\phi|Y} \equiv \pi_{\phi|Y,M} \cdot \pi_{M|Y} = \tilde{\pi}_{\phi|Y,\tilde{\Phi}_M} \cdot \pi_{M|Y}$ (i.e., the second equality follows from Assumption 3.1 (iv)), we can view $IS_r^{avg}(M, \phi)$ as $\pi_{M,\phi|Y}$ - a.s. nonempty random closed sets. Then, we can easily see that the lower probability of Proposition 3.1 can be represented by

$$\inf_{\pi_{r|Y} \in \Pi_{r|Y}} \pi_{r|Y}(r \in A) = \pi_{M,\phi|Y}(IS_r^{avg}(M, \phi) \subset A). \tag{3.7}$$

This representation is particularly useful from the computation perspective. It says the lower probability of the post-averaging ambiguous belief can be approximated by drawing $IS_r^{avg}(M, \phi)$ according to $(M, \phi) \sim \pi_{M,\phi|Y}$ and computing the frequency of event $\{IS_r^{avg}(M, \phi) \subset A\}$ being true. Note that drawing (M, ϕ) from $\pi_{M,\phi|Y}$ is simple as well by noticing its mixture structure.

Specifically, we first draw M from the multinomial distribution with the probabilities set at the posterior model probabilities, and in the second step, conditional on M, we draw ϕ from $\tilde{\pi}_{\phi|Y,\tilde{\Phi}_M}$.

Thus-obtained Monte Carlo draws of $IS_r^{avg}(M,\phi)$ are also useful for constructing the robustified credible region considered in Kitagawa (2012) and Giacomini and Kitagawa (2015). The robustified credible region with credibility $\alpha \in (0,1)$ is defined by the shortest interval on which every posterior in the posterior class assigns the probability at least α;

$$C_\alpha \equiv \arg\min_{C \in \mathcal{C}} \text{length}(C), \quad \text{s.t.} \inf_{\pi_{r|Y} \in \Pi_{r|Y}} \pi_{r|Y}(r \in C) \geq \alpha, \tag{3.8}$$

where $\mathcal{C}$ is the class of connected intervals in $\mathbb{R}$. Since the constraint in (3.8) can be written equivalently as $\pi_{M,\phi|Y}(IS_r^{avg}(M,\phi) \subset C) \geq \alpha$, the computation of C_α can be reduced to finding a shortest interval that contains the α-proportion of the Monte Carlo draws of $IS_r^{avg}(M,\phi)$. 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.

4 Empirical Application

We illustrate our method with a SVAR analysis on monetary policy. We consider 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 the real output growth, π_t is the 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 check the convexity of the identified set using their Lemmas 5.1 and 5.2.

$$\begin{pmatrix} i_t \\ \Delta y_t \\ \pi_t \\ m_t \end{pmatrix} = A_0^{-1} c + A_0^{-1} \sum_{j=1}^{2} A_j \begin{pmatrix} i_{t-j} \\ \Delta y_{t-j} \\ \pi_{t-j} \\ m_{t-j} \end{pmatrix} + A_0^{-1} \begin{pmatrix} \epsilon_t^i \\ \epsilon_t^{\Delta y} \\ \epsilon_t^\pi \\ \epsilon_t^m \end{pmatrix} \tag{4.1}$$

where

$$A_0^{-1} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}. \tag{4.2}$$

We interpret the first equation as a monetary policy function: the Federal Reserve reacts to price, GDP and money shock, as well as lags of all variables. Any additional change would be an exogenous monetary policy shock. Such shocks could arise from changes in the composition of the Federal Open Market Committee or reflect idiosyncratic events such as 9/11 or the housing crisis. The second and third equation are aggregate demand (AD) and aggregate supply (AS) curves, respectively. Finally, 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 in 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. Following the conventional wisdom, we also consider standard Bayesian inference, i.e., a single prior over Q rather than the approach in Giacomini and Kitagawa (2015). In doing so, we use the same agnostic prior as in Uhlig (2005). If models are point-identified, i.e., Model 1 below, using the single prior is equivalent to the method in Giacomini and Kitagawa (2015) as Q is uniquely identified.

4.1 Observationally Non-restrictive Models

Suppose that we are interested in the output impulse response function (IRF) to the unit positive shock in the federal funds rate ϵ_t^i, i.e., $IR^h_{\Delta yi}$, and consider the following two models, i.e., prior beliefs about identifying restrictions with prior weight w_1 and $w_2 = 1 - w_1$, respectively:

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

  According to Algorithm 1 in Rubio-Ramirez, Waggoner, and Zha (2010), we need three zero restrictions on q_i to get point-identification of responses to ϵ_t^i shock. We assume that $a_{21} = a_{31} = a_{41} = 0$, meaning that that AD, AS and money demand do not react to interest rate within a quarter, respectively. These point-identifying assumptions are hard to justify with full confidence[2]. For example, constraining a_{31} can be challenging to justify if the researcher relies on commodity price index rather than GDP deflator.

- *Model 2 (M2, set-identified)*

  In order to get set-identification, we leave a_{31} unrestricted, i.e., AS can now react to interest rate within a quarter. By Lemma 5.1 in Giacomini and Kitagawa (2015), Model 2 delivers a convex identified set for $IR^h_{\Delta yi}$ for every value of the reduced form parameters.

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

In other words, with prior weight w_1 the researcher believes that AS does not respond to interest rate within a quarter, i.e., $a_{31} = 0$ under Model 1; with prior weight $(1 - w_1)$, the researcher believes that AS reacts to interest rate within a quarter, i.e., $a_{31} \neq 0$ under Model 2.

The panel at the bottom on the right of Figure 2 averages the two models when researcher is *ex-ante* indifferent over the contemporaneous reaction of AS to interest rate, i.e., $w_1 = w_2 = 0.5$. As shown in Figure 2 and Table 2, the posterior mean bounds of weighted model is the linear combination of Model 1 and Model 2. Both posterior mean bounds and robustified credible region get smaller with respect to Model 2. As expected, Table 1 shows that the posterior-prior odds ratio is equal to one for both models, i.e., $O_1 = O_2 = 1$, meaning that the prior weights coincide with posterior weights, i.e., $w_1 = w_1^* = 0.5$ and $w_2 = w_2^* = 0.5$. As additional example, the panel at the bottom on the left depicts the case where the researcher is quite confident that AS does not react, i.e., $w_1 = 0.8$.

4.2 Observationally Restrictive Models

In addition to the models above, let us introduce one more model that is observationally restrictive:

- *Model 3 (M3, set-identified)*

  Consider the zero restrictions in Model 2 and add the following sign restrictions used in Moon, Schorfheide, and Granziera (2013): the inflation response to a contractionary monetary policy shock is nonpositive for one quarter; the interest rate response is nonnegative for one quarter; the response of the real money balances is nonpositive for one quarter. By Lemma 5.2 in Giacomini and Kitagawa (2015), the identified set in Model 3 is also convex.

Let the prior weights over the models be denoted by (w_1, w_2, w_3). In contrast to the previous example, we are able to update the prior weights as the Model 3 is observationally restrictive, i.e., Model 3 can have an empty identified set for some value of ϕ. We assume that the researcher is *ex-ante* indifferent among models, i.e. $w_1 = w_2 = w_3 = 0.33$. Table 1 shows that $O_3 = 0.87$, meaning that the plausibility in favour of Model 3 is not very strong. As a result, note that the posterior weights for Model 1 and Model 2 get bigger, while the posterior w_3^* is smaller than the prior w_3.

5 Conclusion

We proposed a method to average point-identified models and set-identified models from the multiple prior Bayesian viewpoint. The method combines the probabilistic belief(s) obtained from the point-identified model(s) with the ambiguous belief(s) obtained from the set-identified model(s), and outputs the set of posteriors. The post-averaging set of posteriors can be summarized by the range of posterior means and robustified credible regions, which are easy to compute based on the MCMC draws of the identified sets in each candidate model. Our method offers a simple and flexible way to inject an additional identifying information to the set-identified model, since averaging can effectively reduce the amount of ambiguity (the size of the posterior class) compared with the case with the set-identified model only. 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 a sensitivity analysis for the point-identified model. For simplicity of exposition, this paper assumed that the reduced form models are common over the candidate models and the point-identified models are just-identified and non-observationally restrictive. The restrictions can be relaxed straightforwardly at a cost of complicating notations and computations.

References

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

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

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.

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.

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.

Giacomini, R., and T. Kitagawa (2015): "Robust Inference about Partially-identified SVARs," *cemmap working paper, University College London.*

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.

——— (2008): *Model Selection and Model Averaging.* Cambridge University Press, Cambridge, UK.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–417.

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

Kitagawa, T., and C. Muris (2015): "Model Averaging in Semiparametric Estimation of Treatment Effects," *unpublished manuscript.*

KLINE, B., AND E. TAMER (2013): "Default Bayesian Inference in a Class of Partially Identified Models," *unpublished manuscript*.

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 (forthcoming): "Heteroskedasticity-robust C_p Model Averaging," *Econometrics Journal*.

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., 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.

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.

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

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.

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.

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

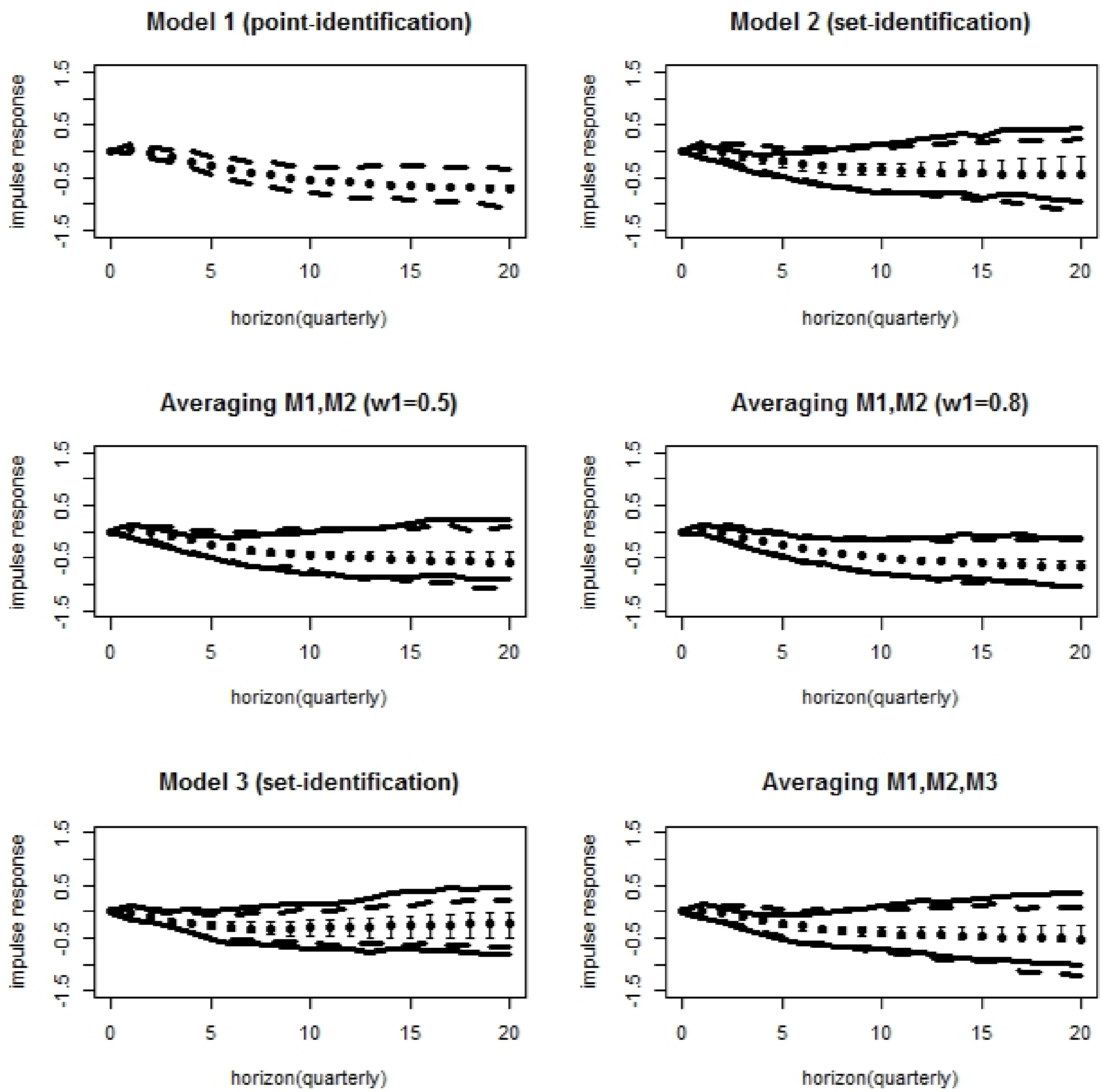

Figure 1: Plots of Output Impulse Responses

Note: the points plot the posterior means with the single prior for Q, the vertical bars show the posterior mean bounds with the multiple priors for Q, the dashed curves connect the upper/lower bounds of the highest posterior density regions with credibility. Being Model 1 point-identified, using single prior is equivalent to the method in Giacomini and Kitagawa (2015) as Q is uniquely identified.

	Averaging M1, M2	Averaging M1,M2	Averaging M1,M2,M3
Prior w_1	0.50	0.80	0.33
Prior w_2	0.50	0.20	0.33
Prior w_3	/	/	0.33
O_1	1	1	1
O_2	1	1	1
O_3	/	/	0.87
Posterior w_1^*	0.50	0.80	0.35
Posterior w_2^*	0.50	0.20	0.35
Posterior w_3^*	/	/	0.30

Table 1: Output Responses: Prior and Posterior Weights

	M1			M2		
	$h = 1$	$h = 10$	$h = 20$	$h = 1$	$h = 10$	$h = 20$
Post. Mean	.05	$-.53$	$-.72$	.04	$-.35$	$-.46$
90% CR	$[-.04, .13]$	$[-.80, -.30]$	$[-1.06, -.33]$	$[-.10, .15]$	$[-.75, .07]$	$[-1.11, .24]$
Post. Mean Bounds	/	/	/	$[-.02, .02]$	$[-.46, -.24]$	$[-.49, -.10]$
90% robustified CR	/	/	/	$[-.15, .11]$	$[-.77, 0.15]$	$[-.94, .43]$
	M3			Averaging M1,M2($w = 0.5$)		
	$h = 1$	$h = 10$	$h = 20$	$h = 1$	$h = 10$	$h = 20$
Post. Mean	$-.03$	$-.31$	$-.23$	.04	$-.44$	$-.59$
90% CR	$[-.13, .07]$	$[-.57, .03]$	$[-.66, .22]$	$[-.09, .13]$	$[-.71, .08]$	$[-1.07, .08]$
Post. Mean Bounds	$[-.04, .01]$	$[-.46, -.16,]$	$[-.51, -.01]$	$[.01, .03]$	$[-.49, -.38]$	$[-.58, -.38]$
90% robustified CR	$[-.14, .10]$	$[-.71, .14]$	$[-.82, .46]$	$[-.10, .13]$	$[-.79, -.00]$	$[-.91, .24]$
	Averaging M1,M2($w = 0.8$)			Averaging M1,M2,M3		
	$h = 1$	$h = 10$	$h = 20$	$h = 1$	$h = 10$	$h = 20$
Post. Mean	.05	$-.50$	$-.67$	.03	$-.40$	$-.52$
90% CR	$[-.05, .13]$	$[-.79, -.13]$	$[-1.08, -.16]$	$[-.10, .16]$	$[-.77, .04]$	$[-1.21, .08]$
Post. Mean Bounds	$[.03, .04]$	$[-.50, -.46]$	$[-.63, -.55]$	$[-0.00, .02]$	$[-.48, -.31]$	$[-.56, -.27]$
90% robustified CR	$[-.05, .14]$	$[-.78, -.14]$	$[-1.04, -.12]$	$[-.13, .14]$	$[-.71, .11]$	[**-1.00,.36**]

Table 2: Output Responses: Estimation and Inference