Intersection Bounds, Robust Bayes, and Updating Ambiguous Beliefs.

Toru Kitagawa*
CeMMAP and *Department of Economics, UCL*

Preliminary Draft
November, 2011

Abstract

This paper develops multiple-prior Bayesian inference for a set-identified parameter whose identified set is constructed by an intersection of two identified sets. We formulate an econometrician's practice of "adding an assumption" as "updating ambiguous beliefs." Among several ways to update ambiguous beliefs proposed in the literature, we consider the Dempster-Shafer updating rule (Dempster (1968) and Shafer (1976)) and the full Bayesian updating rule (Fagin and Halpern (1991) and Jaffray (1992)), and argue that the Dempster-Shafer updating rule rather than the full Bayesian updating rule better matches with an econometrician's common adoption of the analogy principle (Manski (1988)) in the context of intersection bound analysis.

Keywords: Intersection Bounds, Multiple Priors, Bayesian Robustness, Dempster-Shafer Theory, Updating Ambiguity, Imprecise Probability, Gamma-minimax, Random Set,

JEL Classification: C12, C15, C21.

*Email: t.kitagawa@ucl.ac.uk. Financial support from the ESRC through the ESRC Center for Microdata Methods and Practice (CEMMAP) (grant number RES-589-28-0001) is gratefully acknowledged.

1 Introduction

The intersection bound analysis proposed by Manski (1990, 2003) provides a way to aggregate identifying information for a common parameter of interest by taking the intersection of multiple identified sets. This way of constructing and defining the identified set innovates a new identification scheme in econometrics, and it has been applied to a wide range of empirical studies, e.g., Manski and Pepper (2000) and Blundell, Gosling, Ichimura, and Meghir (2007)). Recently, Chernozhukov, Lee, and Rosen (2009) develop estimation and inference for the intersected identified sets from the classical perspective and develop asymptotically valid confidence intervals for intersected identified sets.

In this paper, we analyze inference and decision for this class of partially identified models from the multiple-prior Bayes perspective. Kitagawa (2011) develops a framework of multiple-prior Bayes analysis that can explicitly take into account robustness/agnosticism pursued in the partial identification analysis. This paper extends the approach of Kitagawa (2011) to the intersection bound analysis with paying special attention to the following questions. Are there any multiple-prior Bayes (subjective probability) formulation that induces the operation of intersecting multiple identified sets? If so, what kind of subjective-probability-based reasoning do we have to invoke?

We approach to these questions by modelling an econometrcian's practice of "imposing an assumption" that leads him to intersect multiple identified sets" as "updating ambiguous beliefs". Among several ways to update ambiguous beliefs proposed in the literature, we consider the Dempster-Shafer updating rule (Dempster (1968) and Shafer (1976)) and the full Bayesian updating rule (Fagin and Halpern (1991) and Jaffray (1992)). Our main finding is that the Dempster-Shafer updating rule instead of the full Bayesian updating rule leads to an aggregation rule of intersecting multiple (random) identified sets, and it replicates well the econometrician's common adoption of analogy principle (Manski (1988)) in the context of intersection bound analysis. This result replicates the belief function analysis of Dempster (1967a) and Shafer (1973), who deduce an aggregation rule of ambiguous information as intersecting random sets. Also, an axiomatic analysis on updating ambiguous beliefs by Gilboa and Schmeidler (1993) provides lucid multiple-prior interpretation behind the Dempster-Shafer updating rule, which also readily applies to our econometric framework. Given these early results, we consider contributions of this paper are (i) to clarify a link between the aggregation rule in the Dempster-Shafer's belief function analysis and growing literatures on inference on partial identified model, and (ii) to provide decision theorists some outside-lab evidence that the econometrician's common way of updating ambiguous beliefs is in line with the Dempster-Shafer updating rule rather than the full Bayesian updating rule.

To keep a tight focus on our theoretical development, we develop our analysis along with the following simple model of missing data with an instrumental variable (Manski (1990, 2003)). Suppose a survey targets at inferring the population distribution of a binary variable $Y \in \{1, 0\}$ (e.g., employed or not). In data, not all the sampled subjects respond to the survey, and the response indicator of a sampled subject is denoted by $D \in \{1, 0\}$: $D = 1$ if Y is observed and $D = 0$ if Y is missing. Suppose that the survey is conducted by two modes, say, either by E-mail or by phone. Let us indicate the survey mode by a binary random variable $Z \in \{1, 2\}$: $Z = 1$ if the individual is surveyed by E-mail and $Z = 2$ if he/she is surveyed by phone. If the survey modes are randomized, then it is reasonable to assume that the survey mode indicator Z is independent of the underlying outcome Y. Associated with this exogeneity restriction, we consider using Z as an instrumental variable in the following manner.

Consider

$$\Pr(Y = 1|Z = 1) = \Pr(Y = 1, D = 1|Z = 1) + \Pr(Y = 1|D = 0, Z = 1)\Pr(D = 0|Z = 1),$$
$$\Pr(Y = 1|Z = 2) = \Pr(Y = 1, D = 1|Z = 2) + \Pr(Y = 1|D = 0, Z = 2)\Pr(D = 0|Z = 2).$$

In the right hand side of the first equation, the data let us consistently estimate $\Pr(Y = 1, D = 1|Z = 1)$ and $\Pr(D = 0|Z = 1)$, while the data are silent about the distribution of missing outcomes $\Pr(Y = 1|D = 0, Z = 1)$. Hence, without *any assumptions* on $\Pr(Y = 1|D = 0, Z = 1)$, what we could say about $\Pr(Y = 1|Z = 1)$ given complete knowledge on the distribution of data is

$$\Pr(Y = 1|Z = 1) \in [\Pr(Y = 1, D = 1|Z = 1), \Pr(Y = 1, D = 1|Z = 1) + \Pr(D = 0|Z = 1)]. \tag{1.1}$$

Similarly, without any assumptions on $\Pr(Y = 1|D = 0, Z = 2)$, it holds that

$$\Pr(Y = 1|Z = 2) \in [\Pr(Y = 1, D = 1|Z = 2), \Pr(Y = 1, D = 1|Z = 2) + \Pr(D = 0|Z = 2)]. \tag{1.2}$$

The instrument exogeneity restriction, $Y \perp Z$, then plays a role to combine these two bounds: $\Pr(Y = 1|Z = 1) = \Pr(Y = 1|Z = 2) = \Pr(Y = 1)$ implies that the parameter of interest $\Pr(Y = 1)$ must lie within the intersection of the two bounds (1.1) and (1.2),

$$\begin{aligned} & \max\left\{ \begin{array}{c} \Pr(Y = 1, D = 1|Z = 1) \\ \Pr(Y = 1, D = 1|Z = 2) \end{array} \right\} \\ \leq\ & \Pr(Y = 1) \\ \leq\ & \min\left\{ \begin{array}{c} \Pr(Y = 1, D = 1|Z = 1) + \Pr(D = 0|Z = 1) \\ \Pr(Y = 1, D = 1|Z = 0) + \Pr(D = 0|Z = 2) \end{array} \right\}. \end{aligned} \tag{1.3}$$

We use intersecting two identified sets as a procedure to aggregate two independent pieces of set-identifying information of the common parameter.

As far as identification is concerned, it is fine to say the complete knowledge on distribution of data is available, and therefore the operation of intersecting the two identified sets corresponds to an application of the Boolean logic. With the finite number of observations, however, it is not obvious how we can extrapolate the identification scheme of intersection bounds into the finite sample situation where we wish to use the language of probabilistic judgement for the parameter of interest. The main goal of this paper is to answer this question with focusing on a set of beliefs that represents ambiguity of missing data as well as ambiguous belief for the imposed restriction.

As emphasized in Manski (1990), another interesting feature of the intersection bounds is the refutability property. It means that, if the intersection bounds turn out to be empty, we can refute the imposed restriction which the operation of intersection relies on. The above intersection bounds (1.3) possess the refutability property, i.e., there exists a distribution of data that makes the intersected bounds empty. This paper also investigates how to incorporate ambiguity of the imposed restriction into posterior inference for the parameter of interest. In the above missing data example, the parameter of interest $\Pr(Y = 1)$ is well-defined no matter whether the exogeneity restriction is correctly specified or not. From the single-prior Bayesian point of view, it is natural to incorporate uncertainty on validity of exogeneity restriction into posterior inference by utilizing Bayesian model averaging. This paper explores how to extend the standard Bayesian model averaging to the multiple prior set-up. We derive and analyze the class of model-averaged posteriors , and discuss how to use it for the subsequent inference and decision for the parameter of interest.

The rest of the paper is organized as follows. In Section 2, we introduce our analytical framework. Section 3 provides the main result of the paper; the Dempster-Shafer updating rule and the full Bayesian updating rule are implemented and compared. In Section 4, we analyze point estimation and set inference for the set-identified parameter using the class of beliefs updated by the Dempster-Shafer rule. Model averaging with ambiguous beliefs on imposed assumption is discussed in Section 5, and Section 6 concludes. Proofs and lemma are provided in Appendix A.

2 Multiple-Prior Framework: Preparation

2.1 Setup and Notation

We lay out the framework of our analysis with focusing on the missing data example given in Introduction. We divide the population of study into two subpopulations that are indexed by a value of an assigned binary instrument, e.g., a subpopulation to be surveyed by E-mail and another to be surveyed by phone. Observations are randomly sampled from each of those. We use subscript $j = 1, 2$ to index each subpopulation, and we denote the likelihood function of each sample by $p(X_j|\theta_j)$, $\theta_j \in \Theta_j$, where $X_j = (Y_{ji}D_{ji}, D_{ji} : i = 1, \ldots, n_j)$ denotes observations generated from subpopulation j (the assigned instrument is $Z_i = j$) and θ_j is an unknown parameter vector for subpopulation j. The size of a sample generated from subpopulation j is denoted by n_j. A specification of θ_j must meet the following two requirements, (i) it pins down a distribution of data in subpopulation j, and (ii) θ_j pins down the value of a parameter to which a cross-population restriction is imposed. In the missing data example, θ_j can be specified as follows; for $j = 1, 2$,

$$\theta_j = \left(\theta_{yd|j} : \ y = 1, 0, \ d = 1, 0\right) \in \Theta_j \quad \text{where} \ \ \theta_{yd|j} = \Pr\left(Y = y, D = d|Z = j\right),$$

and Θ_j is four-dimensional probability simplex. A cross-population restriction will be imposed on the conditional mean of Y given Z, $\eta_j \equiv \Pr\left(Y = 1|Z = j\right)$, $j = 1, 2$, which is clearly determined by θ_j, $\eta_j = h_j(\theta_j) = \theta_{10|j} + \theta_{11|j} \in [0, 1]$.

Non-identification of θ_j is defined formally by observational equivalence: θ_j and θ'_j are observationally equivalent if $p(X_j|\theta_j) = p(X_j|\theta'_j)$ for every X_j (e.g., Rothenberg (1971) and Kadane (1974)). Observational equivalence implies that there exists a reduction of parameters $g_j : \Theta_j \to \Phi_j$ such that the likelihood satisfies $p(X_j|\theta_j) = \hat{p}\left(X_j|g_j(\theta_j)\right)$. Reduced-form parameters in subpopulation j, which is also called sufficient parameters in the statistics literature (e.g., Barankin (1960), Dawid (1979)), $\phi_j \equiv g_j(\theta_j) \in \Phi_j$ is defined by a function of θ_j that maps each observationally equivalent classes of θ_j to a point in another parameter space Φ_j. In the example of missing data with an instrumental variable, the observed data likelihood conditional on the instrument is written as a function of

$$\begin{aligned} \phi_j &= \left(\phi_{11|j}, \phi_{01|j}, \phi_{mis|j}\right) \\ &\equiv \left(\Pr\left(Y = 1, D = 1|Z = j\right), \Pr\left(Y = 0, D = 1|Z = j\right), \Pr\left(D = 0|Z = j\right)\right), \end{aligned}$$

for $j = 1, 2$, so,

$$\phi_j = g_j(\theta_j) = (\theta_{11|j}, \theta_{01|j}, \theta_{10|j} + \theta_{00|j}). \in \Phi_j$$

are reduced-form parameters in subpopulation j, where Φ_j is the three-dimensional probability simplex. Denote the inverse image of $g_j(\cdot)$ by $\Gamma_j : \Phi_j \rightrightarrows \Theta_j$. In the missing data example, it is written as

$$\Gamma_j(\phi_j) = \left\{ \theta_j \in \Theta_j : \theta_{11|j} = \phi_{11|j},\ \theta_{01|j} = \phi_{01|j},\ \theta_{10|j} + \theta_{00|j} = \phi_{mis|j} \right\}.$$

$\left\{ \Gamma_j(\phi_j) : \phi_j \in \Phi_j \right\}$ partition Θ_j into regions on each of which the likelihood for θ_j is flat irrespective of observations X_j. Note, by construction, $\Gamma_j(\phi_j) \neq \emptyset$.

Define the identified set for $\eta_j = h_j(\theta_j)$ by the range of $h_j(\theta_j)$ when the domain of θ_j is given by $\Gamma_j(\phi_j) \subset \Theta_j$.

$$H_j(\phi_j) = \left\{ h_j(\theta_j) \in \mathcal{H} : \theta_j \in \Gamma_j(\phi_j) \right\}$$

In the missing data example, we have $H_j(\phi_j) = \left[\phi_{11|j}, \phi_{11|j} + \phi_{mis|j} \right]$ and $\mathcal{H} = [0,1]$. This is identical to the Manski's bounds (Manski (1989)). In what follows, we use the following short-hand notations, $\theta \equiv (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \equiv \Theta$, $\phi \equiv (\phi_1, \phi_2) \in \Phi_1 \times \Phi_2 \equiv \Phi$, $\eta = h(\theta) \equiv (h_1(\theta_1), h_2(\theta_2)) = (\eta_1, \eta_2) \in [0,1]^2$, $X \equiv (X_1, X_2)$, $\mathbf{\Gamma}(\phi) \equiv [\Gamma_1(\phi_1) \times \Gamma_2(\phi_2)] \subset \Theta$, and $H(\phi) \equiv [H_1(\phi_1) \times H_2(\phi_2)] \subset [0,1]^2$.

In the missing data example, the parameter of ultimate interest is the marginal distribution of Y, $\Pr(Y = 1)$. We shall denote it by $\tau \in [0,1]$, which relates to η by

$$\tau = \lambda \eta_1 + (1 - \lambda) \eta_2,$$

where $\lambda = \Pr(Z = 1)$. In our development of inference for τ, we ignore estimation for λ and assume it is known.

2.2 Multiple Priors and Posterior Lower and Upper Probabilities

We assume that the two samples are independent in the sense that the likelihood of the entire data $X = (X_1, X_2)$ is written as

$$\begin{aligned} p(X|\theta) &= p(X_1|\theta_1)p(X_2|\theta_2). \\ &= \hat{p}(X_1|g_1(\theta_1))\hat{p}(X_2|g_2(\theta_2)) \\ &= \hat{p}(X_1|\phi_1)\hat{p}(X_2|\phi_2) \\ &\equiv \hat{p}(X|\phi). \end{aligned}$$

Consider the standard Bayesian inference based on μ_θ a single prior distribution on Θ. For the given μ_θ, the mapping between parameter θ and the reduced-form parameters $\phi = (g_1(\theta_1), g_2(\theta_2))$

yields μ_ϕ a unique prior distribution of the reduced-form parameters $\phi \in \Phi$. The relationship between μ_θ and μ_ϕ is written as, for every measurable subset $B \subset \Phi$,

$$\mu_\phi(B) = \mu_\theta\left(\Gamma\left(B\right)\right),$$

where $\Gamma\left(B\right) = \cup_{\phi \in B}\left[\Gamma(\phi)\right]$. In the presence of reduced-form parameters (sufficient parameters), the posterior distribution of θ denoted by $F_{\theta|X}(\cdot)$ is obtained as (see, e.g., Barankin (1960), Dawid (1979), Poirier (1998)).

$$F_{\theta|X}(A) = \int_\Phi \mu_{\theta|\phi}(A|\phi) dF_{\phi|X}(\phi), \qquad A \subset \Theta, \tag{2.1}$$

where $F_{\phi|X}(\cdot|x)$ is the posterior distribution of the reduced-form parameter ϕ and $\mu_{\theta|\phi}(\cdot|\phi)$ is the conditional prior of θ given ϕ implied by the initial specification of μ_θ. The expression (2.1) shows that the conditional prior for θ given ϕ is never be updated by data, and only the prior information for the reduced form parameters are updated because the value of the likelihood varies only depends on ϕ. Therefore, the shape of posterior for θ remains to be sensitive to the shape of $\mu_{\theta|\phi}(\cdot|\phi)$ implied by a specification of μ_θ no matter how many observations are available in data. This sensitivity of the posterior of θ to the conditional prior $\mu_{\theta|\phi}(\cdot|\phi)$ is also carried over to the posterior of $\eta = \left(h_1\left(\theta_1\right), h_2\left(\theta_2\right)\right)$ and τ if they are set-identified.

Ambiguity stemming from the lack of identification of θ_1 and θ_2 can be modelled in the robust Bayesian framework by introducing multiple priors. A class of priors for θ that is suitable to our context consists of a class of μ_θ that allows for *arbitrary* conditional prior $\mu_{\theta|\phi}(\cdot|\phi)$. Formally, we can formulate such class of priors as, given a single prior for the reduced-form parameter μ_ϕ,

$$\mathcal{M}(\mu_\phi) = \left\{\mu_\theta : \mu_\theta\left(\Gamma\left(B\right)\right) = \mu_\phi(B) \quad \text{for all measurable } B \subset \Phi\right\},$$

This class of prior is indexed by μ_ϕ, meaning that the analysis requires a single prior for the reduced-form parameters ϕ. In other words, we admit a single belief for the distribution of data. A rational for this is that fear for misspecification or the lack of prior knowledge is less severe since we know the prior for ϕ will be well updated by data. We can in principle adopt the existing selection rules for non-informative priors such as the Jeffreys' prior, the reference prior, and the empirical Bayes rule in order to specify a "reasonably objective" prior for the reduced-form parameters ϕ (see Kitagawa (2011) for further discussions).

We use the Bayes rule to update each prior $\mu_\theta \in \mathcal{M}(\mu_\phi)$, and obtain the class of posteriors of θ, which we denote by $\mathcal{F}_{\theta|X}$.[1] We marginalize each posterior of θ in $\mathcal{F}_{\theta|X}$ to form the class of posteriors of η. We denote thus-constructed class of posteriors of η by $\mathcal{F}_{\eta|X}$,

$$\mathcal{F}_{\eta|X} \equiv \left\{F_{\eta|X} : F_{\eta|X}\left(\cdot\right) = F_{\theta|X}\left(h\left(\theta\right) \in \cdot|X\right), \, \mu_\theta \in \mathcal{M}\left(\mu_\phi\right)\right\}, \tag{2.2}$$

[1] For the given specification of prior class, the prior-by-prior updating rule and the Dempster-Shafer updating

where $F_{\eta|X}$ is a posterior distribution of $\eta \in [0,1]^2$. Note that $\mathcal{F}_{\eta|X}$ depends on prior for ϕ through a specification of prior class $\mathcal{M}(\mu_\phi)$, although our notation does not make it explicit. We summarize the class of posteriors $\mathcal{F}_{\eta|X}$ by its lower envelope and upper envelope, the so-called *posterior lower probability and posterior upper probability*: for $D \subset [0,1]^2$,

$$\text{posterior lower probability:} \qquad F_{\eta|X*}(D) \equiv \inf_{F_{\eta|X} \in \mathcal{F}_{\eta|X}} \left\{ F_{\eta|X}(D) \right\},$$

$$\text{posterior upper probability:} \qquad F^*_{\eta|X}(D) \equiv \sup_{F_{\eta|X} \in \mathcal{F}_{\eta|X}} \left\{ F_{\eta|X}(D) \right\}.$$

Note that the posterior lower probability and the upper probability in general have the conjugation property, $F_{\eta|X*}(D) = 1 - F^*_{\eta|X}(D^c)$, which we will frequently refer to in our analysis. Since the prior class $\mathcal{M}(\mu_\phi)$ is designed to represent the collection of prior knowledge (assumptions) that will never be updated by data, we can interpret the value of posterior lower probability as "the posterior probability of $\eta \in D$ being at least $F_{\eta|X*}(D)$ irrespective of the unrevisable prior knowledge." The posterior upper probability is interpreted similarly by replacing "at least" in the previous statement with "at most."

The next theorem provides closed form expressions of $F_{\eta|X*}(\cdot)$ and $F^*_{\eta|X}(\cdot)$ and a list of their analytical properties.

Theorem 2.1 *(i) For measurable subset $D \subset [0,1]^2$,*

$$\begin{aligned} F_{\eta|X*}(D) &= F_{\phi|X}(\{\phi : H(\phi) \subset D\}), \\ F^*_{\eta|X}(D) &= F_{\phi|X}(\{\phi : H(\phi) \cap D \neq \emptyset\}). \end{aligned}$$

(ii) $F_{\eta|X*}(\cdot)$ is supermodular and $F^*_{\eta|X}(\cdot)$ is submodular, i.e., for measurable D_1, $D_2 \subset [0,1]^2$,

$$\begin{aligned} F_{\eta|X*}(D_1 \cup D_2) + F_{\eta|X*}(D_1 \cap D_2) &\geq F_{\eta|X*}(D_1) + F_{\eta|X*}(D_2), \\ F^*_{\eta|X}(D_1 \cup D_2) + F^*_{\eta|X}(D_1 \cap D_2) &\leq F^*_{\eta|X}(D_1) + F^*_{\eta|X}(D_2). \end{aligned}$$

Proof. For a proof of (i), see Theorem 3.1 in Kitagawa (2011). $F_{\eta|X*}(D)$ and $F^*_{\eta|X}(D)$ are containment and capacity functional of random closed sets induced by the posterior distribution of ϕ, so their supermodularity and submodularity are implied by the Choquet Theorem (see, e.g., Molchanov (2005)). ■

rule (the maximum likelihood updating rule) produces the same class of posteriors for θ. This is because

$$\int p(X|\theta)\, d\mu_\theta = \int \hat{p}(X|\phi)\, d\mu_\phi$$

holds and, therefore, the probability of observing the sample is identical for any $\mu_\theta \in \mathcal{M}(\mu_\phi)$. Hence, the prior class $\mathcal{M}(\mu_\phi)$ never shrinks even when we apply the Dempster-Shafer (maximum likelihood) updating rule. This phenomenon is in accordance with the condition for dynamic consistency (rectangurarity property of a prior class) discovered by Epstein and Schneider (2003).

Statement (i) of this theorem says the posterior lower and upper probability correspond to the containment functional and capacity functional of random closed sets (rectangles) in $[0,1]^2$. In the Dempster-Shafer theory, such functionals are called a *plausibility function* and a *belief function*, respectively. In the context partial identified model, thus-constructed lower and upper probability represent the posterior probability law of the identified sets of η induced by the posterior distribution for the identified parameters in the model. Our multiple-prior framework based on prior class $\mathcal{M}\left(\mu_\phi\right)$ highlights a seamless link among the random set theory, Dempster-Shafer theory, and set-identified model in econometrics.

3 Updating Ambiguous Posterior Beliefs

So far, there is no discussion about how to impose assumptions on η. Assumptions to be imposed for η can be in general represented as a subset $D_A \subset [0,1]^2$ referred to as an *assumption subset*. For instance, the instrument exogeneity restriction in the missing data example specifies $D_A = \{\eta : \eta_1 = \eta_2\}$, which is the 45-degree line in $[0,1]^2$. We interpret "imposing an assumption for η" as *updating the class of posteriors $\mathcal{F}_{\eta|X}$ with a conditioning set given by assumption subset D_A*. What we shall get after updating $\mathcal{F}_{\eta|X}$ is another class of posteriors for η. By marginalizing each posterior of η in the updated class for the ultimate parameter of interest $\tau = \lambda\eta_1 + (1-\lambda)\,\eta_2$, we obtain the updated class of posteriors for the ultimate parameter of interest τ, which we denote by $\mathcal{F}_{\tau|X,D_A}$. Our goal is to summarize thus-constructed $\mathcal{F}_{\tau|X,D_A}$ by its lower probability and use it for posterior inference of τ.

Literatures has proposed several ways to update a class of probability measures (see, e.g., Gilboa and Marinacci (2011) for a survey). As of this date, however, there does not appear general agreement on which update rule should be preferred to others. In this paper, we shall focus on two major updating rules, the *Dempster-Shafer updating rule* (Dempster (1967), Shafer (1973)), synonymously called the *maximum likelihood updating rule* (Gilboa and Schmeidler (1993)), and the *full Bayesian updating rule* (Fagin and Halpern (1991) and Jaffray (1992)).[2] We do not intend to provide normative argument on which updating rule should be applied in this context, whereas we compare them in order to argue which update rule agrees with the common adoption of the analogy principle (Manski (1988)) in the context of the intersection bound analysis.

Given assumption subset D_A, the class of posteriors for τ updated by the Demspter-Shafer updating rule has the following form: write $\tau = t(\eta) = \lambda\eta_1 + (1-\lambda)\,\eta_2 \in [0,1]$ and let $t^{-1}(\cdot)$ be

[2] Axiomatizations for these updating rules are done by Gilboa and Schmeidler (1993) for the Dempster-Shafer updating rule and by Pires (2002) for the full Bayesian updating rule.

its inverse image,

$$\mathcal{F}^{DS}_{\tau|X,D_A} \equiv \left\{ F_{\tau|X,D_A}(\cdot) = \frac{F_{\eta|X}\left(t^{-1}(\cdot) \cap D_A\right)}{F_{\eta|X}(D_A)} : F_{\eta|X} \in \mathcal{F}^{*}_{\eta|X} \right\},$$

where $\mathcal{F}^{*}_{\eta|X}$ is the class of posteriors of η defined by

$$\mathcal{F}^{*}_{\eta|X} \equiv \left\{ F_{\eta|X} : F_{\eta|X} \in \arg \max_{F_{\eta|X} \in \mathcal{F}_{\eta|X}} F_{\eta|X}(D_A) \right\}.$$

On the other hand, the updated class of posteriors for τ with the full Bayesian updating rule is written as

$$\mathcal{F}^{FB}_{\tau|X,D_A} \equiv \left\{ F_{\tau|X,D_A}(\cdot) = \frac{F_{\eta|X}\left(t^{-1}(\cdot) \cap D_A\right)}{F_{\eta|X}(D_A)} : F_{\eta|X} \in \mathcal{F}_{\eta|X} \right\},$$

where $\mathcal{F}_{\eta|X}$ is as defined in (2.2). A comparison of these definitions highlight the difference between the Dempster-Shafer updating rule and the full Bayesian updating rule. The Dempster-Shafer rule first reduces class of posteriors $\mathcal{F}_{\eta|X}$ by discarding $F_{\eta|X}$ that fails to put maximal belief on the assumption subset D_A, and, subsequently, applies the standard Bayes rule with conditioning set D_A to each of the remaining ones. By contrast, the full Bayesian updating rule retains all the posteriors in $\mathcal{F}_{\eta|X}$ irrespective of what value $F_{\eta|X} \in \mathcal{F}_{\eta|X}$ puts on D_A, and applies the standard Bayes rule to all members in $\mathcal{F}_{\eta|X}$.

With a metaphor of multiple experts as used in Gilboa and Marinacci (2011), the difference between the two updating rules in our context can be illustrated as follows. Consider a situation that we consult multiple experts about their opinion on τ. Assume they all agrees with the single prior for ϕ, but each of them has different belief for non-identified part of the model, i.e., $\mu_{\theta|\phi}$ differs among them. If we decide to "assume D_A" and to obtain updated opinions by applying the Dempster-Shafer rule, what we actually do is to collect the updated belief of only those experts who are most optimistic on D_A, and, on the other hand, we completely ignore the opinions of the rest of experts. Such stringent selection of multiple experts corresponds to the reduction of posterior class to $\mathcal{F}^{*}_{\eta|X}$. If we "assume D_A" and apply the full Bayesian updating rule, we do ask the opinions (conditional on D_A is true) of all the experts no matter how much they believe in D_A.

The next theorem is the main result of this paper that provides the lower probability of $\mathcal{F}^{DS}_{\tau|X,D_A}$ and $\mathcal{F}^{FB}_{\tau|X,D_A}$.

Theorem 3.1 *Let D_A be the assumption subset corresponding to the instrument exogeneity restriction, $D_A = \{\eta : \eta_1 = \eta_2\}$, and let $\mathcal{F}_{\eta|X}$ be as obtained in (2.2). Denote the intersected identified set by $H_\cap(\phi) \equiv [H_1(\phi_1) \cap H_2(\phi_2)] \subset [0,1]$.*

(i)

$$\begin{aligned} F_{\eta|X*}(D_A) &= F_{\phi|X}\left(\{\phi : H_1(\phi_1) \text{ and } H_2(\phi_2) \text{ are singletons and } H_1(\phi_1) = H_2(\phi_2)\}\right) \\ F^*_{\eta|X}(D_A) &= F_{\phi|X}\left(\{\phi : H_\cap(\phi) \neq \emptyset\}\right), \end{aligned}$$

(ii) If $F^*_{\eta|X}(D_A)$ is bounded away from zero, then the posterior lower probability for τ induced by the Dempster-Shafer updating rule is well defined and given by, for $T \subset [0,1]$,

$$\begin{aligned} F^{DS}_{\tau|X,D_A*}(T) &\equiv \inf\left\{F_{\tau|X,D_A}(T) : F_{\tau|X,D_A} \in \mathcal{F}^{DS}_{\tau|X,D_A}\right\} \\ &= F_{\phi|H_\cap(\phi)\neq\emptyset,X}\left(H_\cap(\phi) \subset T\right). \end{aligned} \tag{3.1}$$

where $F_{\phi|H_\cap(\phi)\neq\emptyset,X}(\cdot)$ is a posterior distribution of $\phi \in \Phi$ conditional on that $H_\cap(\phi) \neq \emptyset$.

(iii) If $F_{\eta|X*}(D_A)$ is bounded away from zero, then the posterior lower probability for τ induced by the full Bayesian updating rule is well-defined and given by, for $T \subset [0,1]$,

$$\begin{aligned} F^{FB}_{\tau|X,D_A*}(T) &\equiv \inf\left\{F_{\tau|X,D_A}(T) : F_{\tau|X,D_A} \in \mathcal{F}^{FB}_{\tau|X,D_A}\right\} \\ &= \frac{F_{\phi|X}\left(H_1(\phi_1) = H_2(\phi_2) = \{\tau\} \subset T\right)}{F_{\phi|X}\left(H_\cap(\phi) \cap T^c \neq \emptyset\right) + F_{\phi|X}\left(H_1(\phi_1) = H_2(\phi_2) = \{\tau\} \subset T\right)}. \end{aligned}$$

(iv) If $F^{DS}_{\tau|X,D_A*}(\cdot)$ and $F^{FB}_{\tau|X,D_A*}(\cdot)$ are well-defined, i.e., $F_{\eta|X*}(D_A)$ is bounded away from zero, then $F^{DS}_{\tau|X,D_A*}(\cdot) \geq F^{FB}_{\tau|X,D_A*}(\cdot)$ holds for every sample X.

Proof. See Appendix A. ■

Statement (i) shows that the posterior lower and upper probabilities of $\{\eta \in D_A\}$ (i.e., credibility for the exogeneity restriction) correspond to the posterior probabilities that the two identified sets become identical singletons and that they intersect, respectively. Intuition of this result is that, given the prior for ϕ, most optimistic belief for exogeneity restriction $\eta_1 = \eta_2$ is formed by the posterior belief that the empirical evidence is *compatible* with $\eta_1 = \eta_2$, i.e., $H_1(\phi_1) \cap H_2(\phi_2) \neq \emptyset$. On the other hand, the most conservative belief for $\eta_1 = \eta_2$ is formed by the posterior belief that the empirical evidence *implies* $\eta_1 = \eta_2$, i.e., $H_1(\phi_1)$ and $H_2(\phi_2)$ are identical singletons. These two extreme ways of forming belief determine the range of the posterior beliefs for $\eta_1 = \eta_2$.

The second result (ii) clarifies that updating $\mathcal{F}_{\eta|X}$ by the Dempster-Shafer rule yields the lower probability for τ that represents the posterior probability law of nonempty intersected identified sets. If we view the posterior distribution of ϕ as the source of belief in the language of Dempster-Shafer theory, this result coincides with the aggregation rule of the belief function proposed in Dempster (1967a) and Shafer (1973).

In statement (iii), the condition of $F_{\eta|X*}(D_A) > 0$ needed for the full Bayesian updating rule requires that the prior for ϕ must put a positive probability to the event $\{\phi : H_1(\phi_1)$ and $H_2(\phi_2)$ are singletons and

Since this event has Lebesgue measure zero in Φ, we are not able to apply the full Bayesian updating rule if μ_ϕ is absolutely continuous with respect to the Lebesgue measure. By contrast, implementability of the Dempster-Shafer updating rule requires a much weaker condition for μ_ϕ. It is known that $F^{FB}_{\tau|X,D_A*}(T)$ is supermodular (Proposition 2.5 in Denneberg (1994) whose proof refers to Sundberg and Wagner (1992) and Jaffray (1992)), while, being different from $F^{DS}_{\tau|X,D_A*}(\cdot)$, $F^{FB}_{\tau|X,D_A*}(\cdot)$ cannot be interpreted as a containment functional of some random sets.

Which of the above update rules would match with econometrician's practice of "imposing an assumption" in a set-identified model? It is common among econometricians to develop an estimation and inference procedure by examining the distribution of a "sample analogue" of the identified estimand. This is also the case in partially identified models with intersection bounds: investigation of the distribution of the sample analogue of the lower and upper bounds of the true identified set is considered as a stepping-stone to inference for set-identified parameters. As shown in the above theorem, drawing posterior inference for τ based on the Dempster-Shafer updated class is equivalent to drawing probabilistic judgement based on a posterior probability law of nonempty intersected identified sets $H_\cap(\phi)$. In contrast, such interpretation is not available if we employ the full Bayesian updating. We therefore think, from the multiple-prior Bayes perspective, what econometricians mean for "adding an assumption" is in line with updating (conditioning) ambiguous beliefs by the Dempster-Shafer updating rule.

4 Set Inference and Decision for τ Based on Ambiguous Beliefs $\mathcal{F}^{DS}_{\tau|X,D_A}$

4.1 Posterior Lower Credible Region

One difficulty of the lower-probability-based inference is that it is not possible to visualize it as we do for a posterior density in the standard Bayesian inference. To overcome this problem, Kitagawa (2011) proposes to report contour sets of the lower probability referred to as the lower credible region. The general definition of the lower credible region $C_{1-\alpha*}$ of a lower probability for $\tau \in [0,1]$, say $F_*(\cdot)$, with credibility level $(1-\alpha) \in [0,1]$ is given by

$$
\begin{aligned}
& C_{1-\alpha*} \equiv \arg\min_{C \in \mathcal{C}} Vol(C) \\
& \text{s.t.} \quad F_*(C) \geq 1-\alpha,
\end{aligned}
\tag{4.1}
$$

where $Vol(C)$ is the volume of subset C in terms of the Lebesgue measure, and $\mathcal{C}$ is a family of subsets in $[0,1]$ over which the volume minimizing credible region is searched. When $F_*(\cdot)$ is given by $F^{DS}_{\tau|X,D_A*}(\cdot)$, we can interpret $C_{1-\alpha*}$ as a smallest set on which posterior credibility for τ is at least $(1-\alpha)$ when posterior beliefs for τ vary over $\mathcal{F}^{DS}_{\tau|X,D_A}$. As Kitagawa (2011) shows, it is feasible to compute the thus-defined posterior credible regions at each credibility level $1-\alpha$, when the lower probability corresponds to a containment functional of convex random intervals, so, we can readily plot the contour sets of $F^{DS}_{\tau|X,D_A*}(\cdot)$.

To obtain the lower credible region of $F^{DS}_{\tau|X,D_A*}(\cdot)$, what we want to compute is a smallest subset $C \subset [0,1]$ that satisfies

$$F_{\phi|H_\cap(\phi)\neq\emptyset,X}\left(H_\cap\left(\phi\right) \subset C\right) \geq 1-\alpha. \tag{4.2}$$

A computational algorithm for computing it is summarized as follows. Let $\{\phi_s : s = 1, \ldots, S\}$ be random draws of ϕ from its posterior, and let $\{\phi_{s^*}\}_{s^*=1}^{S^*}$ be the subset of $\{\phi_s : s = 1, \ldots, S\}$ such that $H_\cap\left(\phi_{s^*}\right)$ is nonempty. Define a distance from $\tau \in \mathcal{H}$ to $H_\cap\left(\phi_{s^*}\right)$ by

$$d\left(\tau, H_\cap(\phi_{s^*})\right) \equiv \sup_{\tau' \in H_\cap(\phi_{s^*})} \left\{\left|\tau - \tau'\right|\right\}.$$

At each $\tau \in [0,1]$, we compute the empirical $(1-\alpha)$-th quantile of $\{d(\tau, H_\cap(\phi_{s^*}))\}_{s^*=1}^{S^*}$, and we find $\tau_{1-\alpha*} \in [0,1]$ that minimizes it. The volume minimizing posterior lower credible region $C_{1-\alpha*}$ is approximated by the interval centered at $\tau_{1-\alpha*}$ with radius equal to the minimized value of the empirical $(1-\alpha)$-th quantile.

In contrast to $F^{DS}_{\tau|X,D_A*}(\cdot)$, $F^{FB}_{\tau|X,D_A*}(\cdot)$ cannot be computed by a containment probability of some random sets, so the above algorithm is not applicable for computing the posterior lower credible regions based on $F^{FB}_{\tau|X,D_A*}(\cdot)$. We leave how to compute them for future research.

4.2 Gamma Minimax Decision with Ambiguous Beliefs

In this section, we derive point estimation for the parameter of interest $\tau = \lambda\eta_1 + (1-\lambda)\,\eta_2$ by solving a statistical decision problem. Specifically, we formulate the *conditional* decision problem with adopting the *posterior gamma-minimax criterion* (see, e.g., Berger (1985, p205), Betro and Ruggeri (1992), and Vidakovic (2000)). As Kitagawa (2011) demonstrates, an algorithm to approximate a solution of the gamma minimax decision problem is available when a lower probability of a class of posteriors is a containment functional of random sets. Along that approach, this section solves the gamma-minimax problem when a class of posteriors for τ is given by $\mathcal{F}^{DS}_{\tau|X,D_A}$.

Let $a \in [0,1]$ be an action, which is interpreted as reporting a particular point estimate for τ. Given action a is taken and τ_0 being the true state of nature, a loss function $L(\tau_0, a)$: $[0,1] \times [0,1] \to \mathcal{R}_+$ yields how much cost the decision maker owes by taking such action. The *posterior risk conditional on* exogeneity restriction $D_A = \{\eta : \eta_1 = \eta_2\}$ is defined by

$$\rho(a, F_{\tau|X,D_A}) \equiv \int_0^1 L(\tau,a)dF_{\tau|X,D_A}(\tau) \tag{4.3}$$

where the second argument $F_{\tau|X,D_A}$ represents the dependence of the criterion on posterior for τ. Our posterior analysis deals with multiple posterior distributions in $\mathcal{F}^{DS}_{\tau|X,D_A}$, so a class of posterior risks, $\left\{\rho(a, F_{\tau|X,D_A}) : F_{\tau|X,D_A} \in \mathcal{F}^{DS}_{\tau|X,D_A}\right\}$ will be considered.

Definition 4.1 *Define the posterior upper risk over* $F_{\tau|X,D_A} \in \mathcal{F}^{DS}_{\tau|X,D_A}$ *by*

$$\rho^*(a, \mathcal{F}^{DS}_{\tau|X,D_A}) \equiv \sup\left\{\rho(a, F_{\tau|X,D_A}) : F_{\tau|X,D_A} \in \mathcal{F}^{DS}_{\tau|X,D_A}\right\},$$

A posterior gamma-minimax action with the class of posteriors $\mathcal{F}^{DS}_{\tau|X,D_A}$ *is an action that minimizes* $\rho^*(a, \mathcal{F}^{DS}_{\tau|X,D_A})$.

The next proposition shows that these posterior gamma minimax actions can be obtained by minimizing certain objective functions that can be approximated if draws of ϕ from its posterior are available.

Proposition 4.1 *Assume loss function* $L(\tau, a)$ *is nonnegative. The gamma-minimax action with class of posteriors* $\mathcal{F}^{DS}_{\tau|X,D_A}$ *solves*

$$a^* = \arg\min_{a\in[0,1]} \int_{\Phi} \left[\sup_{\tau\in H_\cap(\phi)} L(\tau, a)\right] dF_{\phi|H_\cap(\phi)\neq\emptyset,X}(\phi), \tag{4.4}$$

Proof. A proof proceeds by writing the gamma minimax criteria in the form of Choquet expectations. See the proof of Proposition 4.1 in Kitagawa (2011) for further details. ■

The expression of (4.4) shows that the posterior gamma-minimax criterion conditional on exogeneity restriction D_A is written as the expectation of the worst-case loss $\sup_{\eta\in H_\cap(\phi)} L(\eta, a)$ when the bounds of τ are nonempty intersection bounds $H_\cap(\phi)$. The supremum part comes from the researcher's consideration of the worst-case scenario associated with ambiguity of τ: what the researcher knows about τ is only that it lies within the intersected identified set $H_\cap(\phi)$, but he does not have any probabilistic judgement on where the true τ is likely to be within $H_\cap(\phi)$. On the other hand, the expectation with respect to $F_{\phi|H_\cap(\phi)\neq\emptyset,X}(\cdot)$ represents posterior uncertainty for the nonempty identified set $H_\cap(\phi)$.

A closed form expression of a^* is not in general available, while this proposition suggests a simple numerical algorithm to approximate it. Let $\{\phi_s\}_{s=1}^{S}$ be S random draws of ϕ from its posterior $F_{\phi|X}(\phi)$. Among these S draws of ϕ, let $\{\phi_{s^*}\}_{s^*=1}^{S^*}$ be the subset of S draws that yield nonempty intersection bounds $H_\cap(\phi_s) \neq \emptyset$. Then the posterior upper risk appearing in (4.4) can be approximated by

$$\frac{1}{S^*}\sum_{s^*=1}^{S^*} \sup_{\tau\in H_\cap(\phi_{s^*})} L(\tau, a).$$

So, an approximation of a^* is obtained by numerically finding a minimizer of this function.

5 Extensions and Discussions

5.1 Model Averaging with Multiple Priors

In some intersection bound analysis, we can refute the imposed assumptions (given the complete knowledge of distribution of data) if the intersection bounds become empty. As we can see from the expression of (3.1), the lower probability (ambiguous beliefs) updated by the Dempster-Shafer rule focuses only on the nonempty identified sets ($H_{\cap}(\phi) \neq \emptyset$) by leaving the empty ones out of the conditioning event in the probability calculation. In the example of missing data, the parameter of interest $\tau = \Pr(Y=1)$ is well defined no matter whether $\eta_1 = \eta_2$ is valid or not. Hence, even when $\eta_1 = \eta_2$ is refuted ($H(\phi) = \emptyset$), we can still construct the identified set for τ without invoking the exogeneity restriction. In such situation, we may be interested in incorporating uncertainty/ambiguity about $\eta_1 = \eta_2$ into posterior inference on τ. One widely applied Bayesian treatment for incorporating model or assumption uncertainty is model averaging: take the weighted average of the posterior distribution conditional on the assumption being valid and the one conditional on the restriction not-being valid with the weight corresponding to the posterior probability for validity of the restriction. In what follows, we examine whether such model averaging idea can be formulated in the intersection bound analysis with refutability property.

Consider a posterior distribution of $\tau = t(\eta) = \lambda\eta_1 + (1-\lambda)\eta_2$,

$$\begin{aligned} F_{\tau|X}(T) &= F_{\eta|X}\left(t^{-1}(T)\right) \\ &= F_{\eta|X,D_A}\left(t^{-1}(T)\right) F_{\eta|X}(D_A) + F_{\eta|X,D_A^c}\left(t^{-1}(T)\right) F_{\eta|X}(D_A^c). \end{aligned}$$

The results of Theorem 3.1 concerns the lower probability of $F_{\eta|X,D_A}\left(t^{-1}(T)\right)$ appearing in the first term. The next theorem, in contrast, concerns the lower probability of $F_{\tau|X}(T)$ appearing in the left hand side. In order to state the result, define $H_\lambda(\phi)$ as the weighted average (Minkowski sum) of $H_1(\phi_1)$ and $H_1(\phi_2)$ with weight λ, i.e.,

$$H_\lambda(\phi) = \{\lambda\eta_1 + (1-\lambda)\eta_2 : \eta_1 \in H_1(\phi_1), \ \eta_2 \in H_2(\phi_2)\}. \tag{5.1}$$

In the missing data example, $H_\lambda(\phi)$ is nothing else than Manski (1989)'s bounds without an instrument,

$$\begin{aligned} H_\lambda(\phi) &= \left[\lambda\phi_{11|1} + (1-\lambda)\phi_{11|2}, \lambda\left[\phi_{11|1} + \phi_{mis|1}\right] + (1-\lambda)\left[\phi_{11|2} + \phi_{mis|2}\right]\right] \\ &= \left[\Pr(Y=1, D=1), \Pr(Y=1, D=1) + \Pr(D=0)\right]. \end{aligned}$$

Theorem 5.1 *(i) Let $H_\cap(\phi) = H_1(\phi_1) \cap H_2(\phi_2)$ and $H_\lambda(\phi)$ be as defined in (5.1). Assume $\mu_\phi(\{\phi : H_\cap(\phi) \neq \emptyset\}) > 0$. The lower probabilities of $F_{\tau|X}(T)$, $T \subset [0,1]$, when posterior $F_{\eta|X}$*

varies over $\mathcal{F}^*_{\eta|X}$ *and* $\mathcal{F}_{\eta|X}$ *are*

$$\begin{aligned} F^{DS}_{\tau|X*}(T) &\equiv \inf\left\{F_{\tau|X}(T) : F_{\eta|X} \in \mathcal{F}^*_{\eta|X}\right\} \\ &= F_{\phi|H_\cap(\phi)\neq\emptyset,X}\left(H_\cap(\phi) \subset T\right) F_{\phi|X}\left(H_\cap(\phi) \neq \emptyset\right) \\ &\quad + F_{\phi|H_\cap(\phi)=\emptyset,X}\left(H_\lambda(\phi) \subset T\right) F_{\phi|X}\left(H_\cap(\phi) = \emptyset\right), \\ F^{FB}_{\tau|X*}(T) &\equiv \inf\left\{F_{\tau|X}(T) : F_{\eta|X} \in \mathcal{F}_{\eta|X}\right\} \\ &= F_{\phi|X}\left(H_\lambda(\phi) \subset T\right), \end{aligned}$$

respectively.

(ii) $F^{DS}_{\tau|X*}(\cdot) \geq F^{FB}_{\tau|X*}(\cdot)$ *holds for every sample* X.

Proof. See Appendix A. ■

The expression of $F^{DS}_{\tau|X*}(\cdot)$ given in (i) shows that when the researcher imposes the exogeneity restriction by implementing the Dempster-Shafer updating rule, then $F^{DS}_{\tau|X*}(T)$, the resulting lower probability for τ, is obtained as a mixture of the two containment functionals of random sets: the containment functional of the nonempty intersection bounds $H_\cap(\phi)$ and the containment functional of the averaged bounds $H_\lambda(\phi)$, where the mixture weight corresponds to the probability of having nonempty intersections and empty intersections, respectively. Intuition behind this result can be described as follows. When $H_\cap(\phi) \neq \emptyset$, meaning that empirical evidence does not contradict the exogeneity restriction, we then form a belief that the exogeneity restriction is true and τ must be contained in the intersection bounds $H_\cap(\phi)$, whereas, we would form a belief that the exogeneity restriction is wrong and τ is contained in $H_\lambda(\phi)$ only when the empirical evidence implies violation of the exogeneity restriction ($H_\cap(\phi) = \emptyset$). The weighted average of these two ways to form a belief for τ is represented by $F^{DS}_{\tau|X*}(\cdot)$ obtained in the above theorem.

The expression of the lower probability of $F_{\tau|X}(\cdot)$ using all the beliefs in $\mathcal{F}_{\eta|X}$ instead of the reduced class $\mathcal{F}^*_{\eta|X}$ results in the ambiguous beliefs for τ that we would obtain by ignoring any information of the instrument. This contrast between $F^{DS}_{\tau|X*}(T)$ and $F^{FB}_{\tau|X*}(T)$ combined with the result of (ii) show that, when we summarize posterior beliefs for τ in terms of the lower probability of $F_{\tau|X}$ rather than $F_{\tau|X,D_A}$, the instrument helps increase informativeness of statistical inference only through the reduction of $\mathcal{F}_{\eta|X}$ to a smaller class $\mathcal{F}^*_{\eta|X}$.

By noting that $F^{DS}_{\tau|X*}(\cdot)$ is seen as a containment functional of the mixture of random sets, $H_\cap(\phi)$ and $H_\lambda(\phi)$, we can use the procedure introduced in Section 4 for computing the posterior lower credible region and the gamma minimax action. Let $\{\phi_s\}_{s=1}^S$ be S random draws of ϕ from the posterior $F_{\phi|X}(\phi)$. For each draw of ϕ, construct

$$H^*(\phi_s) = \begin{cases} H_\cap(\phi_s) & \text{if} \quad H_\cap(\phi_s) \neq \emptyset \\ H_\lambda(\phi_s) & \text{if} \quad H_\cap(\phi_s) = \emptyset \end{cases} \qquad s = 1, \ldots, S. \tag{5.2}$$

By replacing the simulated intersection bounds $\{H_\cap(\phi_{s^*}) : s^* = 1, \ldots, S^*\}$ with thus-generated random sets $\{H^*(\phi_s) : s = 1, \ldots, S\}$ in the algorithm of Section 4, we can obtain approximates of the lower credible regions based on $F^{DS}_{\tau|X*}(\cdot)$ and an associated gamma minimax action for τ.

6 Conclusion

From the multiple prior Bayes perspective, this paper proposes inference and decision for a partially identified parameter whose identified set is constructed by intersecting the two identified sets. The focus of our analysis is what kind of robust Bayes framework can justify the operation of taking the intersection of the two identified sets even in the finite sample situation. By treating "imposing an assumption" as "updating ambiguous belief," we implement the Dempster-Shafer updating rule and the full Bayesian updating rule and derive the lower probabilities of the updated class of beliefs for each case. A comparison between them shows that the Dempster-Shafer updating rule yields a posterior probability law of the intersected identified sets, while the full Bayesian updating rule does not. This leads us to a claim that the Dempster-Shafer updating rule somewhat mimics a naive implementation of the analogy principle in the context of the intersection bound analysis.

It is worth noting that, being different from Chernozhukov, Lee, and Rosen (2009), our lower probability inference does not raise any concern about the bias correction for the lower and upper bound estimators. In case that the true identified sets to be intersected are close each other, the sample analogue of the lower and upper bounds of the intersected identified sets tend to be estimated with an inward bias, and the bias correction is therefore needed in order to ensure the correct frequentist coverage. Our set estimation output (the volume minimizing lower credible region) does not yield any of such bias correction arguments, and, as a result, the volume minimizing posterior credible region can lead to a smaller frequentist coverage than the desired nominal coverage probability. This indicates that the degree of belief represented by the lower probability of a class of posteriors updated by the Dempster-Shafer rule is fundamentally different from the frequentist coverage probability for the true identified set.

Appendix

A Lemma and Proofs

A.1 Theorem 3.1

A proof of Theorem 3.1 given below applies the formulae of the Dempster-Shafer updating rule and the full Bayesian updating rule to the class of probability measures $\mathcal{F}_{\eta|X}$ constructed in (2.2). See Denneberg (1994) for an excellent review and proofs for these formulae.

Proof of Theorem 3.1. Statement (i) is a corollary of Theorem 2.1 (i),

$$\begin{aligned} F^*_{\eta|X}(D_A) &= F_{\phi|X}(\{\phi : H(\phi) \cap D_A \neq \emptyset\}) \\ &= F_{\phi|X}(\{\phi : \exists \eta_0 \in [0,1] \text{ s.t. } \eta_0 \in H(\phi_1) \text{ and } \eta_0 \in H(\phi_2)\}) \\ &= F_{\phi|X}(\{\phi : H_\cap(\phi) \neq \emptyset\}). \end{aligned}$$

As for the lower probability,

$$\begin{aligned} F_{\eta|X*}(D_A) &= F_{\phi|X}(\{\phi : H(\phi) \subset D_A\}) \\ &= F_{\eta|X*}(D_A) = F_{\phi|X}(\{\phi : H_1(\phi_1) \text{ and } H_2(\phi_2) \text{ are singletons and } H_1(\phi_1) = H_2(\phi_2)\}), \end{aligned}$$

where the second line follows because a rectangular set $H(\phi) = H_1(\phi_1) \times H_2(\phi_2)$ in $[0,1]^2$ is contained in the 45-degree line if and only if $H_1(\phi_1)$ and $H_2(\phi_2)$ are singletons and $H_1(\phi_1) = H_2(\phi_2)$.

(ii) Since $F^*_{\eta|B}(\cdot)$ is submodular (Theorem 2.1 (ii)) and $F^*_{\eta|X}(D_A) > 0$ is assumed, Theorem 3.4 in Denneberg (1994) applies, and the upper probability of $\mathcal{F}^{DS}_{\tau|X,D_A}$, defined by $F^{DS*}_{\tau|X,D_A}(\cdot) = \sup\left\{\frac{F_{\eta|X}(t^{-1}(\cdot)\cap D_A)}{F_{\eta|X}(D_A)} : F_{\eta|X} \in \mathcal{F}^*_{\eta|X}\right\}$, is obtained by

$$F^{DS*}_{\tau|X,D_A}(\cdot) = \frac{F^*_{\eta|X}\left(t^{-1}(\cdot) \cap D_A\right)}{F^*_{\eta|X}(D_A)}.$$

Note by Theorem 2.1 (i), for $T \subset [0,1]$,

$$\begin{aligned} F^*_{\eta|X}\left(t^{-1}(T) \cap D_A\right) &= F_{\phi|X}\left(H(\phi) \cap t^{-1}(T) \cap D_A \neq \emptyset\right) \\ &= F_{\phi|X}(t(H(\phi) \cap D_A) \cap T \neq \emptyset) \\ &= F_{\phi|X}(H_\cap(\phi) \cap T \neq \emptyset), \end{aligned}$$

where the third line follows by noting $t(H(\phi) \cap D_A) = H_\cap(\phi)$ holds. Combining this with the statement of (i) yields

$$\begin{aligned} F^{DS*}_{\tau|X,D_A}(T) &= \frac{F_{\phi|X}(H_\cap(\phi) \cap T \neq \emptyset)}{F_{\phi|X}(H_\cap(\phi) \neq \emptyset)} \\ &= F_{\phi|H_\cap(\phi)\neq\emptyset,X}(H_\cap(\phi) \cap T \neq \emptyset). \end{aligned}$$

Conjugation of the upper and lower probabilities shows

$$\begin{aligned} F^{DS}_{\tau|X,D_A*}(T) &= 1 - F^{DS*}_{\tau|X,D_A}(T^c) \\ &= 1 - F_{\phi|H_\cap(\phi)\neq\emptyset,X}(H_\cap(\phi)\cap T^c \neq \emptyset) \\ &= F_{\phi|H_\cap(\phi)\neq\emptyset,X}(H_\cap(\phi)\subset T). \end{aligned}$$

(iii) Given submodularity of $F^*_{\eta|B}(\cdot)$ and $F_{\eta|X*}(D_A) > 0$, Proposition 2.1 and Theorem 2.4 in Denneberg (1994) yield a closed form expression of the upper probability of $\mathcal{F}^{FB}_{\tau|X,D_A}$,

$$\begin{aligned} F^{FB*}_{\tau|X,D_A}(\cdot) &\equiv \sup\left\{\frac{F_{\eta|X}\left(t^{-1}(\cdot)\cap D_A\right)}{F_{\eta|X}(D_A)} : F_{\eta|X}\in\mathcal{F}_{\eta|X}\right\} \\ &= \frac{F^*_{\eta|X}\left(t^{-1}(\cdot)\cap D_A\right)}{F^*_{\eta|X}(t^{-1}(\cdot)\cap D_A) + F_{\eta|X*}([t^{-1}(\cdot)]^c\cap D_A)}. \end{aligned}$$

By conjugation of upper and lower probability and Theorem 2.1 (i), for $T\subset[0,1]$,

$$\begin{aligned} F^{FB}_{\tau|X,D_A*}(T) &= 1 - F^{FB*}_{\tau|X,D_A}(T^c) \\ &= \frac{F_{\eta|X*}\left(\left[t^{-1}(T^c)\right]^c\cap D_A\right)}{F^*_{\eta|X}(t^{-1}(T^c)\cap D_A) + F_{\eta|X*}([t^{-1}(T^c)]^c\cap D_A)} \\ &= \frac{F_{\eta|X*}\left(t^{-1}(T)\cap D_A\right)}{F^*_{\eta|X}(t^{-1}(T^c)\cap D_A) + F_{\eta|X*}(t^{-1}(T)\cap D_A)} \\ &= \frac{F_{\phi|X}\left(H(\phi)\subset\left[t^{-1}(T)\cap D_A\right]\right)}{F_{\phi|X}(H_\cap(\phi)\cap T^c\neq\emptyset) + F_{\phi|X}(H(\phi)\subset[t^{-1}(T)\cap D_A])}. \end{aligned}$$

Note $F_{\phi|X}\left(H(\phi)\subset\left[t^{-1}(T)\cap D_A\right]\right) = F_{\phi|X}(H_1(\phi_1) = H_2(\phi_2) = \{\tau\}\subset T)$, and therefore the conclusion follows.

(iv) is obvious since $\mathcal{F}^*_{\eta|X}\subset\mathcal{F}_{\eta|X}$. ■

A.2 Theorem 5.1

For a proof of Theorem 5.1, we introduce the following notations.

$$\mathcal{M}^*(\mu_\phi) \equiv \left\{\mu_\theta\in\mathcal{M}(\mu_\phi) : F_{\eta|X}\in\arg\max_{F_{\eta|X}\in\mathcal{F}_{\eta|X}} F_{\eta|X}(D_A)\right\}.$$

In words, $\mathcal{M}^*(\mu_\phi)$ is a set of priors for θ that belongs to $\mathcal{M}(\mu_\phi)$ and the implied posterior for η puts maximal probability on $D_A = \{\eta : \eta_1 = \eta_2\}$. $\mathcal{M}^*(\mu_\phi)$ can be also written as

$$\mathcal{M}^*(\mu_\phi) = \arg\max\left\{F_{\theta|X}(J) : \mu_\theta\in\mathcal{M}(\mu_\phi)\right\},$$

where $J\equiv h^{-1}(D_A) = \bigcup_{\eta_0\in[0,1]}\left[h_1^{-1}(\eta_0)\times h_1^{-1}(\eta_0)\right]\subset\Theta$. This subclass appears to depend on data X by its construction, while it actually does not because $\max\left\{F_{\theta|X}(J) : \mu_\theta\in\mathcal{M}(\mu_\phi)\right\}$

is attained by specifying conditional prior of $\theta|\phi$ as $\mu_{\theta|\phi} = 1\{\Gamma(\phi) \cap J \neq \emptyset\}$. (Lemma A.3 of Kitagawa (2011)). Note also that $\{\phi \in \Phi : H_{\cap}(\phi) \neq \emptyset\}$ is equivalent to $\{\phi \in \Phi : \Gamma(\phi) \cap J \neq \emptyset\}$, because

$$[\Gamma_1(\phi_1) \times \Gamma_2(\phi_2)] \cap \bigcup_{\eta_0 \in [0,1]} \left[h_1^{-1}(\eta_0) \times h_1^{-1}(\eta_0)\right] \neq \emptyset$$

occurs if and only if there exist some $\eta_0 \in [0,1]$ such that

$$[\Gamma_1(\phi_1) \times \Gamma_2(\phi_2)] \cap \left[h_1^{-1}(\eta_0) \times h_2^{-1}(\eta_0)\right] \neq \emptyset,$$

and, by noting $H_j(\phi_j) = h_j(\Gamma_j(\phi_j))$, this is also equivalent to that there exist some $\eta_0 \in [0,1]$ such that

$$\eta_0 \in H_1(\phi_1) \text{ and } \eta_0 \in H_2(\phi_2).$$

To prove the theorem, we introduce the following notations, for A measurable subset in Θ,

$$\begin{aligned} \Phi_A &= \{\phi : \Gamma(\phi) \cap A \neq \emptyset\}, \\ \Phi_A^c &= \{\phi : \Gamma(\phi) \cap A = \emptyset\}, \end{aligned}$$

where $\{\Phi_A, \Phi_A^c\}$ partitions Φ. In particular, if $A = J$, we have by the above argument

$$\begin{aligned} \Phi_J &= \{\phi : \Gamma(\phi) \cap J \neq \emptyset\} = \{\phi : H_{\cap}(\phi) \neq \emptyset\}, \\ \Phi_J^c &= \{\phi : \Gamma(\phi) \cap J = \emptyset\} = \{\phi \in \Phi : H_{\cap}(\phi) = \emptyset\}. \end{aligned}$$

We shall prove Theorem 5.1 using the following two lemma.

Lemma A.1 *If probability measure on Θ, μ_θ, belongs to class of priors $\mathcal{M}^*(\mu_\phi)$, then*

$$\mu_\theta(\Gamma(\Phi_J) \cap J^c) = 0.$$

Proof. By Lemma A.3 and Theorem 3.1 of Kitagawa (2011), the upper bound of $F_{\theta|X}(J)$ when a prior varies over $\mathcal{M}(\mu_\phi)$ is attained ($\mu_\theta \in \mathcal{M}(\mu_\phi)$) if and only if μ_θ has conditional prior distribution

$$\mu_\theta(J|\phi) = 1\{\Gamma(\phi) \cap J \neq \emptyset\} = 1_{\Phi_J}(\phi), \quad \mu_\phi\text{-almost surely.} \tag{A.1}$$

Therefore, if $\mu_\theta \in \mathcal{M}^*(\mu_\phi)$,

$$\begin{aligned} \mu_\theta(\Gamma(\Phi_J) \cap J^c) &= \mu_\theta(\Gamma(\Phi_J)) - \mu_\theta(\Gamma(\Phi_J) \cap J) \\ &= \int_{\Phi_j} \left[1 - \mu_{\theta|\phi}(J|\phi)\right] d\mu_\phi \\ &= \int_{\Phi} \left[1_{\Phi_J}(\phi) - 1_{\Phi_J}(\phi)\right] d\mu_\phi \\ &= 0. \end{aligned}$$

■

Lemma A.2 *Let A be a measurable subset of Θ.*

$$\begin{aligned}\sup_{\mu_\theta \in \mathcal{M}^*(\mu_\phi)} F_{\theta|X}(A) &= F_{\phi|X}(\{\phi : \Gamma(\phi) \cap J \cap A \neq \emptyset\} \cap \Phi_J) \\ &\quad + F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A \neq \emptyset\} \cap \Phi_J^c)\end{aligned}$$

Proof. Fix $A \in \mathcal{A}$ throughout the proof. We first show an inequality: for every $\mu_\theta \in \mathcal{M}^*(\mu_\phi)$

$$\mu_\theta(A|\phi) \leq 1_{\Phi_{A\cap J}}(\phi)\, 1_{\Phi_J}(\phi) + 1_{\Phi_A}(\phi)\, 1_{\Phi_J^c}(\phi) \tag{A.2}$$

holds μ_ϕ-almost surely. To show this, let $B \in \mathcal{B}$ be an arbitrary subset in Φ, and consider

$$\begin{aligned}\int_B \mu_\theta(A|\phi)\, d\mu_\phi &= \mu_\theta(A \cap \Gamma(B)) \\ &= \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J})) \\ &\quad + \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J}^c)) \\ &\quad + \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_J^c) \cap \Gamma(\Phi_A)) \\ &\quad + \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_J^c) \cap \Gamma(\Phi_A^c)).\end{aligned}$$

The first term is bounded above by $\mu_\theta(\Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J}))$. As for the second term, by Lemma A.1,

$$\begin{aligned}& \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J}^c)) \\ =\ & \mu_\theta([A \cap J] \cap \Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J}^c)) \\ =\ & 0 \quad (\because\ [A \cap J] \cap \Gamma(\Phi_{A\cap J}^c) = \emptyset).\end{aligned}$$

The third term is bounded above by $\mu_\theta(\Gamma(B) \cap \Gamma(\Phi_J^c) \cap \Gamma(\Phi_A))$ and the fourth term is zero because $A \cap \Gamma(\Phi_A^c) = \emptyset$. Hence,

$$\begin{aligned}\int_B \mu_\theta(A|\phi)\, d\mu_\phi &\leq \mu_\theta(\Gamma(B) \cap \Gamma(\Phi_J) \cap \Gamma(\Phi_{A\cap J})) + \mu_\theta(\Gamma(B) \cap \Gamma(\Phi_J^c) \cap \Gamma(\Phi_A)) \\ &= \int_B \left[1_{\Phi_{A\cap J}}(\phi)\, 1_{\Phi_J}(\phi) + 1_{\Phi_A}(\phi)\, 1_{\Phi_J^c}(\phi)\right] d\mu_\phi.\end{aligned}$$

Since $B \in \mathcal{B}$ is arbitrary, we obtain inequality (A.2).

In the next step, we show that there exists $\mu_\theta^* \in \mathcal{M}^*(\mu_\phi)$ that attains the upper bound given in (A.2). To construct μ_θ^*, let $\xi_A(\phi)$ be a θ-valued function defined on $[\Phi_J^c \cap \Phi_A]$ such that $\xi_A(\phi) \in [\Gamma(\phi) \cap A]$, μ_ϕ-almost every $\phi \in [\Phi_J^c \cap \Phi_A]$. Similarly, let $\xi_{A\cap J}(\phi)$ be a θ-valued function defined on $[\Phi_J \cap \Phi_{A\cap J}]$ such that $\xi_{A\cap J}(\phi) \in [\Gamma(\phi) \cap A \cap J]$ holds for μ_ϕ-almost every $\phi \in [\Phi_J \cap \Phi_{A\cap J}]$. Such $\xi_A(\phi)$ exists since $[\Gamma(\phi) \cap A]$ is nonempty whenever $\phi \in [\Phi_J^c \cap \Phi_A]$. Similarly, $\xi_{A\cap J}(\phi)$ exists since $[\Gamma(\phi) \cap A \cap J]$ is nonempty whenever $\phi \in [\Phi_J \cap \Phi_{A\cap J}]$ (Theorem

2.13 of Molchanov (2005)). Let μ_θ be a probability measure belonging to $\mathcal{M}^*\left(\mu_\phi\right)$, and define μ_θ^* a probability measure on Θ by, for $\tilde{A} \subset \Theta$,

$$\begin{aligned}
\mu_\theta^*\left(\tilde{A}\right) &= \overbrace{\mu_\theta\left(\tilde{A} \cap \Gamma\left(\Phi_J^c\right) \cap \Gamma\left(\Phi_A^c\right)\right)}^{(i)} + \overbrace{\mu_\phi\left(\left\{\xi_A\left(\phi\right) \in \tilde{A}\right\} \cap \Phi_J^c \cap \Phi_A\right)}^{(ii)} \\
&\quad + \underbrace{\mu_\theta\left(\tilde{A} \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right)\right)}_{(iii)} \\
&\quad + \underbrace{\mu_\phi\left(\left\{\xi_{A\cap A_J}\left(\phi\right) \in \tilde{A}\right\} \cap \Phi_J \cap \Phi_{A\cap J}\right)}_{(iv)}. \qquad \text{(A.3)}
\end{aligned}$$

Thus constructed $\mu_\theta^*\left(\cdot\right)$ belongs to $\mathcal{M}^*\left(\mu_\phi\right)$. To check this, we will show $\mu_\theta^*\left(J|\phi\right) = 1\left\{\Gamma\left(\phi\right) \cap J \neq \emptyset\right\} = 1_{\Phi_J}\left(\phi\right)$ because this is a necessary and sufficient condition for $\mu_\theta^* \in \mathcal{M}\left(\mu_\phi\right)$ (Lemma A.3 and Theorem 3.1 of Kitagawa (2011)). For $B \subset \Phi$, let $\tilde{A} = \Gamma\left(B\right) \cap J$. We have $(i) = 0$ because $J \cap \Gamma\left(\Phi_J^c\right) = \emptyset$. As for (ii), by the definition of Φ_J^c, no $\phi \in \left[\Phi_J^c \cap \Phi_A\right]$ satisfies $\xi_A\left(\phi\right) \in J$. So, (ii) is zero. For (iii), by Lemma A.1 and $\mu_\theta \in \mathcal{M}^*\left(\mu_\phi\right)$, it holds

$$\begin{aligned}
\mu_\theta\left(\Gamma\left(B\right) \cap J \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right)\right) &= \mu_\theta\left(\Gamma\left(B\right) \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right)\right) \\
&= \mu_\phi\left(B \cap \Phi_J \cap \Phi_{A\cap J}^c\right),
\end{aligned}$$

where $\Gamma\left(B\right) \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right) = \Gamma\left(B \cap \Phi_J \cap \Phi_{A\cap J}^c\right)$ holds because $\left\{\Gamma\left(\phi\right) : \phi \in \Phi\right\}$ is a partition of Θ. The fourth term (iv) becomes $\mu_\phi\left(B \cap \Phi_J \cap \Phi_{A\cap J}\right)$ by the construction of $\xi_{A\cap J}\left(\phi\right)$. By combining all these, we have

$$\begin{aligned}
\mu_\theta^*\left(J \cap \Gamma\left(B\right)\right) &= \mu_\phi\left(B \cap \Phi_J \cap \Phi_{A\cap J}^c\right) + \mu_\phi\left(B \cap \Phi_J \cap \Phi_{A\cap J}\right) \\
&= \mu_\phi\left(B \cap \Phi_J\right) = \int_B 1_{\Phi_J}\left(\phi\right) d\mu_\phi,
\end{aligned}$$

implying the desired claim, $\mu_\theta^*\left(J|\phi\right) = 1_{\Phi_{\eta_1=\eta_2}}\left(\phi\right)$, μ_ϕ-almost surely. Hence, μ_θ^* constructed in (A.3) belongs to $\mathcal{M}^{opt}\left(\mu_\phi\right)$.

In the final step, we will show that μ_θ^* constructed in (A.3) achieves the upper bound of (A.2). Let us set $\tilde{A} = \Gamma\left(B\right) \cap A$ for $B \subset \Phi$. The first term (i) returns zero because $A \cap \Gamma\left(\Phi_A^c\right) = \emptyset$. Regarding (ii), we note that by the construction of $\xi_A\left(\phi\right)$, we have $\left\{\xi_A\left(\phi\right) \in \left[\Gamma\left(B\right) \cap A\right]\right\}$ if and only if $\left\{\phi \in \left[B \cap \Phi_J^c \cap \Phi_A\right]\right\}$, leading to $(ii) = \mu_\phi\left(B \cap \Phi_J \cap \Phi_A\right)$. The third term (iii) is zero by Lemma A.1, because

$$\begin{aligned}
&\mu_\theta\left(\Gamma\left(B\right) \cap A \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right)\right) \\
= \; &\mu_\theta\left(\Gamma\left(B\right) \cap A \cap J \cap \Gamma\left(\Phi_J\right) \cap \Gamma\left(\Phi_{A\cap J}^c\right)\right) \\
= \; &0 \qquad \left(\because A \cap J \cap \Gamma\left(\Phi_{A\cap J}^c\right) = \emptyset\right).
\end{aligned}$$

As for the fourth term (iv), $\{\xi_{A\cap J}(\phi) \in [\Gamma(B) \cap A]\}$ if and only if $\{\phi \in [B \cap \Phi_J \cap \Phi_{A\cap J}]\}$. Hence, by summing up these, we obtain

$$\mu_\theta^*(\Gamma(B) \cap A) = \mu_\phi(B \cap \Phi_J^c \cap \Phi_A) + \mu_\phi(B \cap \Phi_J \cap \Phi_{A\cap J}),$$

implying that μ_θ^* have conditional distribution,

$$\mu_\theta^*(A|\phi) = 1_{\Phi_{A\cap J}}(\phi)\, 1_{\Phi_J}(\phi) + 1_{\Phi_A}(\phi)\, 1_{\Phi_J^c}(\phi),$$

μ_ϕ-almost surely, which achieves the upper bound given in (A.2).

Recall the expression of the posterior $F_{\theta|X}(A) = \int_\Phi \mu_{\theta|\phi}(A|\phi)\, dF_{\phi|X}$ (Equation (2.1)). The above argument has shown that the attainable upper bound of $\mu_{\theta|\phi}(A|\phi)$ when $\mu_\theta \in \mathcal{M}^*(\mu_\phi)$ is (A.2). Therefore,

$$\begin{aligned}
\sup_{\mu_\theta \in \mathcal{M}^*(\mu_\phi)} F_{\theta|X}(A) &= \int_\Phi \left[1_{\Phi_{A\cap J}}(\phi)\, 1_{\Phi_J}(\phi) + 1_{\Phi_A}(\phi)\, 1_{\Phi_J^c}(\phi)\right] dF_{\phi|X} \\
&= F_{\phi|X}(\{\phi : \Gamma(\phi) \cap J \cap A \neq \emptyset\} \cap \Phi_J) \\
&\quad + F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A \neq \emptyset\} \cap \Phi_J^c).
\end{aligned}$$

■

Proof of Theorem 5.1. We note that the range of $\tau = \tau(\theta) \equiv \lambda h_1(\theta_1) + (1-\lambda)\, h_2(\theta_2)$ when the domain is $\theta \in \Gamma(\phi)$ is given by $H_\lambda(\phi)$, where $\tau(\cdot) : \Theta \to [0,1]$ and denote $\tau^{-1}(\cdot)$ be its inverse image. Using this result, we first derive $F_{\tau|X*}^{FB}(\cdot)$. By applying Theorem 3.1 (ii) in Kitagawa (2011), the posterior lower probability of τ when prior μ_θ varies over $\mathcal{M}(\mu_\phi)$ is obtained as

$$\begin{aligned}
F_{\tau|X*}^{FB}(\cdot) &= \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\tau|X}(T) = F_{\phi|X}(\{\tau(\Gamma(\phi)) \subset T\}) \\
&= F_{\phi|X}(\{H_\lambda(\phi) \subset T\}).
\end{aligned}$$

Next, consider the upper probability under the class of priors $\mathcal{M}^*(\mu_\phi)$. By applying Lemma A.2, it is given by

$$\begin{aligned}
&\sup_{\mu_\theta \in \mathcal{M}^*(\mu_\phi)} F_{\tau|X}(T) = \sup_{\mu_\theta \in \mathcal{M}^*(\mu_\phi)} F_{\theta|X}\left(\tau^{-1}(T)\right) \\
&= F_{\phi|X}\left(\left\{\phi : \Gamma(\phi) \cap J \cap \tau^{-1}(T) \neq \emptyset\right\} \cap \Phi_J\right) + F_{\phi|X}\left(\left\{\phi : \Gamma(\phi) \cap \tau^{-1}(T) \neq \emptyset\right\} \cap \Phi_J^c\right) \\
&= F_{\phi|X}(\{\phi : \tau(\Gamma(\phi) \cap J) \cap T \neq \emptyset\} \cap \Phi_J) + F_{\phi|X}(\{\phi : \tau(\Gamma(\phi)) \cap T \neq \emptyset\} \cap \Phi_J^c).
\end{aligned}$$

Since $\tau(\Gamma(\phi) \cap J) = t(H_\cap(\phi)) = H_\cap(\phi)$, the first term in the sum is $F_{\phi|X}(\{\phi : H_\cap(\phi) \cap T \neq \emptyset\} \cap \Phi_J)$. As for the second term, $F_{\phi|X}(\{\phi : \tau(\Gamma(\phi)) \cap T \neq \emptyset\} \cap \Phi_J^c) = F_{\phi|X}(\{\phi : H_\lambda(\phi) \cap T \neq \emptyset\} \cap \Phi_J^c)$.

Therefore, the upper probability becomes

$$\begin{aligned}\sup_{\mu_\theta\in\mathcal{M}^*(\mu_\phi)} F_{\tau|X}(T) &= F_{\phi|X}(\{\phi : H_\cap(\phi)\cap T\neq\emptyset\}\cap\{H_\cap(\phi)\neq\emptyset\}) \\ &\quad +F_{\phi|X}(\{H_\lambda(\phi)\cap T\neq\emptyset\}\cap\{H_\cap(\phi)=\emptyset\}) \\ &= F_{\phi|H_\cap(\phi)\neq\emptyset,X}(\{H_\cap(\phi)\cap T\neq\emptyset\}|H_\cap(\phi)\neq\emptyset)\,F_{\phi|X}(\{H_\cap(\phi)\neq\emptyset\}) \\ &\quad +F_{\phi|H_\cap(\phi)=\emptyset,X}(\{H_\lambda(\phi)\cap T\neq\emptyset\}|H_\cap(\phi)=\emptyset)\,F_{\phi|X}(\{H_\cap(\phi)=\emptyset\}).\end{aligned}$$

The lower probability $F^{DS}_{\tau|X*}(T)$ follows by duality with the upper probability. ■

References

[1] Barankin, E.W. (1961), "Sufficient Parameters: Solution of the Minimal Dimensionality Problem," *Annals of the Institute of Mathematical Statistics*, 12, 91-118.

[2] Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis, Second edition.* Springer-Verlag.

[3] Betro, B. and Ruggeri, F. (1992) "Conditional Γ-minimax actions under Convex Losses." *Communications in Statistics, Part A - Theory and Methods*, 21, 1051-1066.

[4] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323-363.

[5] Chernozhukov, V., S. Lee, and A. Rosen (2009), "Intersection Bounds: Estimation and Inference," *Cemmap working paper.* MIT and University College London.

[6] Dawid, A.P. (1979), "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society, Series B*, 41, No.1, 1-31.

[7] Dempster, A.P. (1967a) "Upper and Lower Probabilities Induced by a Multivalued Mapping." *Annals of Mathematical Statistics* 38, 325-339.

[8] Dempster, A.P. (1967b) "Upper and Lower Probability Inferences Based on a Sample from a Finite Univariate Population." *Biometrika,* 54, 515-528.

[9] Dempster, A.P. (1968) "A Generalization of Bayesian Inference." *The Journal of Royal Statistical Society*, B30, 205-247.

[10] Denneberg, D. (1994) "Conditioning (Updating) Non-additive Measures," *Annals of Operations Research,* 52, 21-42.

[11] Epstein, L.G. and M. Schneider (2003) "Recursive Multiple-priors," *Journal of Economic Theory*, 113, 1, 1-31.

[12] Fagin, R. and J.Y. Halpern (1991), "A New Approach to Updating Beliefs," in the *Proceedings of 6th Conference on Uncertainty in Artificial Intelligence*, 6 347-374.

[13] Florens, J,-P. and M. Mouchart (1977), "Reduction of Bayesian Experiments." CORE Discussion Paper 7737, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

[14] Florens, J,-P., M. Mouchart, and J.-M. Rolin (1990), *Elements of Bayesian Statistics*, Marcel Dekker, New York.

[15] Gilboa, I. and M. Marinacci (2011), "Ambiguity and the Bayesian Paradigm," Working Paper, Università Bocconi.

[16] Gilboa, I. and D.Schmeidler (1993), "Updating Ambiguous Beliefs," *The Journal of Economic Theory*, 59, 33-49.

[17] Graf, S. (1980) "A Radon-Nikodym Theorem for Capacities." *Journal für die Reine und Angewandte Mathematik*, 320, 194-214.

[18] Huber, P.J. (1973) "The Use of Choquet Capacities in Statistics." *Bulletin of International Statistical Institute*. 45, 181-191.

[19] Jaffray, Y. (1992) "Bayesian Updating and Belief Functions," *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 5, 1144-1152.

[20] Kadane, J.B. (1974) "The Role of Identification in Bayesian Theory." In S.E. Fienberg and A. Zellner eds. *Studies in Bayesian Econometrics and Statistics, In Honor of Leonard J. Savage*, pp. 175-191. North-Holland, Amsterdam.

[21] Kitagawa, T. (2011) "Inference and Decision for Set Identified Parameters Using Posterior Lower Probability," cemmap working paper, University College London.

[22] Lehmann, E.L. and J. P. Romano (2005), *Testing Statistical Hypotheses, 2nd edition*, Springer Science+Business Media, New York.

[23] Manski, C.F. (1988) *Analog Estimation Methods in Econometrics.* Chapman and Hall, New York.

[24] Manski, C.F. (1989) "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24, No. 3, 343-360.

[25] Manski, C.F. (1990) "Nonparametric Bounds on Treatment Effects" *The American Economic Review*, Vol. 80, No. 2, 319-323.

[26] Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.

[27] Manski, C.F. and J.F. Pepper (2000), "Monotone Instrumental Variables, with an Application to the Returns to Schooling," *Econometrica*, 68, 4, 997-1022.

[28] I. Molchanov. (2005), *Theory of Random Sets.* Springer-Verlag, New York.

[29] Picci, G. (1977), "Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem." *SIAM Journal on Applied Mathematics*, 33. 383-398.

[30] Pires, C.P. (2002), "A Rule for Updating Ambiguous Beliefs," *Theory and Decision*, 33, 137-152.

[31] Poirier, D.J. (1998) "Revising Beliefs in Nonidentified Models" *Econometric Theory* 14, 483-509.

[32] Rios Insua, D., F. Ruggeri, and B. Vidakovic (1995) "Some Results on Posterior Regret Γ-minimax estimation." *Statistics and Decisions*, 13, 315-331.

[33] Rothenberg, T.J. (1971) "Identification in Parametric Models," *Econometrica*, 39, No.3, 577-591.

[34] Shafer, G. (1976) *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ.

[35] Vidakovic, B. (2000) "Γ-minimax: A Paradigm for Conservative Robust Bayesians." In D. Rios Insua and F. Ruggeri eds. *Robust Bayesian Analysis, Lecture Notes in Statistics.* 152, 241-259. Springer, New York.

[36] Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, UK.

[37] Wasserman, L.A. (1990) "Prior Envelopes Based on Belief Functions," *The Annals of Statistics*, 18, 1, 454-464.