cemmap

centre for microdata methods and practice

Inference and decision for set identified parameters using posterior lower and upper probabilities

Toru Kitagawa

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP16/11

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Inference and Decision for Set Identified Parameters Using Posterior Lower and Upper Probabilities

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

April, 2011

Abstract

This paper develops inference and statistical decision for set-identified parameters from the robust Bayes perspective. When a model is set-identified, prior knowledge for model parameters is decomposed into two parts: the one that can be updated by data (revisable prior knowledge) and the one that never be updated (unrevisable prior knowledge.) We introduce a class of prior distributions that shares a single prior distribution for the revisable, but allows for arbitrary prior distributions for the unrevisable. A posterior inference procedure proposed in this paper operates on the resulting class of posteriors by focusing on the posterior lower and upper probabilities. We analyze point estimation of the set-identified parameters with applying the gamma-minimax criterion. We propose a robustified posterior credible region for the set-identified parameters by focusing on a contour set of the posterior lower probability. Our framework offers a procedure to eliminate set-identified nuisance parameters, and yields inference for the marginalized identified set. For an interval identified parameter case, we establish asymptotic equivalence of the lower probability inference to frequentist inference for the identified set.

Keywords: Partial Identification, Bayesian Robustness, Belief Function, Imprecise Probability, Gamma-minimax, Random Set.

JEL Classification: C12, C15, C21.

*Email: t.kitagawa@ucl.ac.uk. I thank Andrew Chesher, Siddhartha Chib, Jean-Pierre Florens, Charles Manski, Adam Rosen, and Elie Tamer for valuable discussions and comments. I also thank the seminar participants at Academia Sinica Taiwan, EC^2 Conference 2010 in Toulouse, MEG 2010, Simon Frasor University, and the University of British Columbia for helpful comments. All remaining errors are mine. 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 *conditionality principle* claims that statistical inference and decision making should be conditional on what has been actually observed. The Bayesian method is one of the inference procedures that stand on the conditional viewpoint. In a situation where the data (likelihood function) is not informative about the parameter of interest, a specification of prior distribution can have significant influence to the posterior distribution, and the subsequent posterior analysis may largely depend upon the researcher's prior knowledge. We encounter such situation in the partially identified model put forward by the sequence of seminal work by Manski (1989, 1990, 2003, 2008). Accordingly, some may claim that in the absence of credible prior information the Bayesian analysis is less suitable in partially identified model especially when the goal of analysis is to obtain the conclusion that is robust to empirically unverifiable assumptions.

To remedy this lack of robustness, but without giving up the conditionality principle, this paper develops a robust Bayesian procedure of inference and decision for partially identified models. When the parameters are not identified, or, more precisely, the likelihood is flat over some region in the parameter space, a prior distribution of the parameters is decomposed into two components: the one that can be updated by data (*revisable prior knowledge*) and the one that never be updated by data (*unrevisable prior knowledge*). We claim that the lack of posterior robustness is due to the unrevisable prior knowledge, and aim to make posterior probabilistic judgement free from it. For this purpose, we introduce a class of prior distributions that shares a single prior distribution for the revisable, but allows for arbitrary prior distributions for the unrevisable. We use the Bayes rule to update each prior in the class in order to form the class of posteriors. The posterior inference proposed in this paper operates on thus constructed class of posteriors by focusing on its lower and upper envelopes, so called the *posterior lower and upper probabilities.*

For each subset in the parameter space, the posterior lower and upper probabilities are defined by the lowest and highest probabilities allocated on the subset among those multiple posteriors. The lower and upper probabilities originate from Dempster (1966, 1967a, 1967b, 1968) in his fiducial argument of drawing posterior inference without specifying a prior distribution. Shafer (1976, 1982) extends the Dempster's analysis to develop the belief function analysis: a system of probabilistic judgement and learning that can deal with both partial prior knowledge and set-valued observations. Walley (1991) introduces the lower and upper probabilities into the Bayesian subjectivism in order to model the degree of prior ignorance and coherent indecision of a decision maker. Our use of lower and upper probabilities is motivated by the robust Bayes analysis considered in DeRoberts and Hartigan (1981), Wasserman (1989, 1990), and Wasserman and Kadane (1990). These authors consider a class of priors as a set of probability distributions bounded below and above by the lower and upper probabilities, and derive the updating rules of the lower and upper probabilities so as to obtain the posterior class. In econometrics, the pioneering work of using multiple priors are Chamberlain and Leamer (1976) and Leamer (1982), who obtain the bounds for the posterior mean of regression coefficients when a prior varies over a certain class. All of these previous works do not explicitly consider non-identified models, and let a prior class be formed by the researcher's imprecise prior knowledge or simply by analytical convenience. This paper, in contrast, focuses on non-identified models, and proposes an unambiguous way to construct the prior class with which the posterior inference embodies the notion of robustness pursued in Manski's partial identification analysis. The main contributions of this paper are therefore (i) to clarify how the early idea of lower and upper probability analysis can fit to the recent issue on inference and decision for partially identified parameters, and (ii) to demonstrate that by designing the prior class to include arbitrary unrevisable prior knowledge, we can obtain analytically tractable posterior lower and upper probabilities, and they are useful for robust posterior inference and decision in the partially identified models in econometrics.

The inference procedure proposed in this paper retains some of coherent features in Bayesian inference: inference and decision are made conditional on data without involving large sample approximation. On the other hand, unlike the standard Bayesian procedure, our procedure does not need a complete specification of a prior distribution, and this makes our inference different from the existing Bayesian analysis of set-identified parameters (Bollinger and van Hasselt (2009), Gustafson (2009, 2010), Moon and Schorfheide (2010), and Liao and Jiang (2010)). This paper does not provide normative discussion on whether we should proceed with a single prior or multiple priors for inferring the non-identified parameters. We believe judgement on this should depend on to what extent the researcher wants to make inference "robust" and in which manner he wants to incorporate the attitude for being "agnostic" about the non-identified parameters into statistical inference. This paper clarifies that the multiple prior approach provides one way to make the posterior inference explicit about such notion of robustness and agnosticism inherited in the partial identification analysis.

When the model is not identified, the resulting posterior lower and upper probabilities become *supadditive and subadditive measures* respectively. Specifically, we show that the posterior lower (upper) probabilities on a subset in the parameter space is given by the posterior probability that the subset contains (hits) the identified set, which is a posteriori random sets with its source of randomness coming from the posterior probability distribution of the identified components in the model. Beresteanu and Molinari (2008) and Bereseanu, Molchanov, and Molinari (2010) show usefulness and wide applicability of the random set theory to the partially identified model by viewing an observation as a random set and defining the true identified set by its Aumann expectation. Galichon and Henry (2006) proposed a use of Choquet capacity in defining and inferring the identified set in their framework of incomplete model. We obtain the identified set as a random object whose probability law is represented by the posterior lower and upper probabilities, and in this sense, our analysis conceptually differs from these frequentist ways of obtaining the identified set.

From the statistical decision viewpoint, the robust Bayes formulation offers a convenient framework to formulate statistical decision. We define a point estimator for the set-identified parameter by adopting the gamma-minimax criterion (Berger (1985)), which leads us to reporting an estimate that minimizes the posterior risk formed under the most pessimistic prior within the class. The gamma-minimax decision problem often becomes challenging and its analysis is often limited to rather simple parametric models with a certain choice of prior class (see, e.g., Betro and Ruggeri (1992) and Vidakovic (2000)). Our specification of prior class, however, offers a simple solution for the gamma-minimax decision problem in a general class of non-identified models. The closed form expression of the point estimator is not available in general, but it can be computed by applying the standard Bayesian computing such as Markov Chain Monte Carlo.

As a summary of the posterior uncertainty of the set-identified parameters, we develop a procedure to construct a posterior credible region by focusing on contour sets of the posterior lower probability. We propose *the posterior lower credible region* with credibility α, which is defined as the smallest subset in the parameter space on which we place *at least probability* α *irrespective of the unrevisable prior knowledge*. An algorithm to construct such credible regions is provided for interval identified parameter cases. We also analyze the large sample property of the volume minimizing posterior lower credible region, and establish Bernstein-von Mises type asymptotic equivalence between the lower probability inference and frequentist inference for the identified set.

1.1 An Overview of the Main Results

To illustrate the main results of this paper, we present how to implement our posterior inference procedure for the mean of a binary random variable when some observations are missing. Let $Y \in \{1, 0\}$ be a binary outcome of interest and let $D \in \{1, 0\}$ be an indicator of whether Y being observed ($D = 1$) or not ($D = 0$). Data is given by a size N random sample, $x^N = \{(Y_i \cdot D_i, D_i) : i = 1, \ldots, N\}$.

The starting point of our analysis is to specify a parameter vector θ that can determine the distribution of data as well as the parameter of interest. Here, θ can be specified by (θ_{yd}) where $\theta_{yd} \equiv \Pr(Y = y, D = d)$, $y = 1, 0$, $d = 1, 0$. The observed data likelihood for θ is written as

$$p(x^N|\theta) = \theta_{11}^{n_{11}} \theta_{01}^{n_{01}} \left[\theta_{10} + \theta_{00}\right]^{n_{mis}}$$

where $n_{11} = \sum_{i=1}^N Y_i D_i$, $n_{01} = \sum_{i=1}^N (1 - Y_i) D_i$, $n_{mis} = \sum_{i=1}^N (1 - D_i)$. Clearly, this likelihood function depends on θ only through the three probability masses, $\phi = (\phi_{11}, \phi_{01}, \phi_{mis}) \equiv$

$(\theta_{11}, \theta_{01}, \theta_{10} + \theta_{00})$, implying that the likelihood for θ has flat regions and each of them can be expressed by a set-valued map of ϕ,

$$\Gamma(\phi) \equiv \{\theta : \theta_{11} = \phi_{11},\ \theta_{01} = \phi_{01},\ \theta_{10} + \theta_{00} = \phi_{mis}\}.$$

Note that the mean of Y is written as a function of θ, $\eta \equiv \Pr(Y = 1) = \theta_{11} + \theta_{10}$, and the identified set of η notated by $H(\phi)$ is constructed as the range of η when the domain of θ is $\Gamma(\phi)$. In the missing data model, $H(\phi)$ becomes the bounds of $\Pr(Y = 1)$ obtained in Manski (1989),

$$H(\phi) = [\phi_{11}, \phi_{11} + \phi_{mis}].$$

In order to obtain a posterior distribution for η in the standard Bayes procedure, we *have to* specify a prior distribution for θ. Due to the fact that the likelihood for θ has flat regions, some part of prior cannot be updated, and, as a result, the posterior inference for η is sensitive to a specification of prior no matter how large the sample size is.

To make the posterior inference free from such sensitivity issue, we consider the following posterior inference procedure. First, we specify a prior distribution for ϕ, preferably, a reasonably noninformative prior such as Jeffreys's prior.[1] Next, we introduce a class of prior distributions of θ that consists of any probability distributions of θ that are compatible with the prior of ϕ and the set valued mapping $\Gamma(\phi)$. Every prior of θ in the class is updated by the Bayes rule and is marginalized to η. As a result, we obtain the class of posteriors for η.

We summarize the class of posteriors of η by their lower and upper envelopes, and use them for point estimation and set estimation of η. Specifically, our lower probability inference proceeds as follows.

1. Specify a reasonably noninformative prior for ϕ and update it by the Bayes rule.
2. Let $\{\phi_s : s = 1, \ldots, S\}$ be random draws of ϕ from the posterior. The mean and median of the posterior lower probability of η can be approximated by
$$\arg\min_a \frac{1}{S}\sum_{s=1}^{S} \sup_{\eta \in H(\phi_s)} (a - \eta)^2 \quad \text{and} \quad \arg\min_a \frac{1}{S}\sum_{s=1}^{S} \sup_{\eta \in H(\phi_s)} |a - \eta|,$$
respectively.
3. The posterior lower credible region of η at credibility level α, which can be interpreted as a contour set of the posterior lower probability of η, is constructed by the smallest interval that contains $H(\phi)$ with posterior probability α.

[1] See Kass and Wasserman (1996) for a survey on "reasonably" noninformative priors.

We verify that the point estimator of η in Step 2 yields the optimal decision under the gamma-minimax criterion and the posterior lower credible region of η in Step 3 asymptotically coincides with the frequentist confidence intervals for the identified set.

1.2 Plan of the Paper

The rest of the paper is organized as follows. In Section 2, we introduce the likelihood based framework of the set identified model. In Section 3, we discuss which part of prior can be seen as unrevisable prior knowledge. We introduce the class of prior distributions that represents arbitrary unrevisable prior knowledge, and derive the posterior lower and upper probabilities. The point estimation problem with the multiple priors are examined in Section 4. In Section 5, we investigate how to construct the posterior credible region based on the posterior lower probability. Its large sample behavior is also examined in an interval identified parameter case. Proofs and lemma are provided in Appendix A.

2 Likelihood and Set Identification

2.1 The General Framework

Let $(\mathbf{X}, \mathcal{X})$ and $(\Theta, \mathcal{A})$ be measurable spaces of a sample $X \in \mathbf{X}$ and a parameter vector $\theta \in \Theta$, respectively. Our analytical framework up to Section 4 only requires the parameter space Θ to be a Polish (complete separable metric) space, and it covers both a parametric model $\Theta = \mathcal{R}^d$, $d < \infty$, and a nonparametric model where Θ is a separable Banach space. We make the sample size implicit in our notation until the large sample analysis in Section 5. Let μ_θ be a marginal probability distribution on the parameter space $(\Theta, \mathcal{A})$ referred to as a *prior distribution* for θ. We assume that the conditional distribution of X given θ has the probability density $p(x|\theta)$ at every $\theta \in \Theta$ with respect to a σ-finite measure on $(\mathbf{X}, \mathcal{X})$. We call $p(x|\theta)$ the *likelihood* for θ.

Parameter vector θ consists of parameters that determine behaviors of economic agents as well as those that characterize the distribution of unobserved heterogeneities in the population. In the context of missing data or the counterfactual causal model, θ indexes the distribution of the underlying population outcomes or the potential outcomes (see Example 2.1 and 2.2 below). In all of these cases, the parameter θ should be distinguished from the parameters that are used solely to index the sampling distribution of observations. The identification problem of θ arises in this context: If multiple values of θ can generate the same distribution of data, then we claim that these θ's are *observationally equivalent* and identification of θ fails. In terms of the likelihood function, the observational equivalence of θ and $\theta' \neq \theta$ means that the values of the likelihood at θ and θ' are equal for every possible observations, i.e., $p(x|\theta) = p(x|\theta')$ for every $x \in \mathbf{X}$ (Rothenberg (1971), Drèze (1974), and Kadane (1974)). We represent the observational equivalence relation of θ's by a many-to-one function $g : (\Theta, \mathcal{A}) \rightarrow (\Phi, \mathcal{B})$,

$$g(\theta) = g(\theta') \text{ if and only if } p(x|\theta) = p(x|\theta') \text{ for all } x \in \mathbf{X}.$$

The equivalence relationship partitions the parameter space Θ into equivalent classes on each of which the likelihood of θ is "flat" irrespective of observations, and $\phi = g(\theta)$ maps each of these equivalent classes to a point in another parameter space Φ. In the language of structural model in econometrics (Hurwicz (1950) and Koopman and Reiersol (1950)), $\phi = g(\theta)$ is interpreted as the reduced form parameter that carries all the information for the structural parameters θ through the value of the likelihood function. In the literature of Bayesian statistics, $\phi = g(\theta)$ is referred to as the *minimally sufficient parameters (sufficient parameters* for short), and the range space of $g(\cdot)$, $(\Phi, \mathcal{B})$, is called the *sufficient parameter space* (Barankin (1960), Dawid (1979), Florens and Mouchart (1977), Florens, Mouchart, and Rolin (1990), and Picci (1977)).[2]

In the presence of sufficient parameters, the likelihood depends on θ only through the function $g(\theta)$ and there exists a $\mathcal{B}$-measurable function $\hat{p}(x|\cdot)$ such that

$$p(x|\theta) = \hat{p}(x|g(\theta)) \quad \forall x \in \mathbf{X} \text{ and } \theta \in \Theta, \tag{2.1}$$

holds (Lemma 2.3.1 of Lehmann and Romano (2005)).

Denote the inverse image of $g(\cdot)$ by Γ,

$$\Gamma(\phi) = \{\theta \in \Theta : g(\theta) = \phi\}.$$

Since $g(\theta)$ is many-to-one, $\Gamma(\phi)$ and $\Gamma(\phi')$ for $\phi \neq \phi'$ are disjoint, and $\{\Gamma(\phi)\ ;\ \phi \in \Phi\}$ constitutes a partition of Θ. In the structural model of econometrics, $\Gamma(\phi)$ can be seen as a set of observationally equivalent θ's that share the same value of the reduced form parameter ϕ. We assume that $g(\cdot)$ is onto, $g(\Theta) = \Phi$, so that $\Gamma(\phi)$ is assumed to be nonempty for every $\phi \in \Phi$.[3]

We define set identification of θ and the identified set of θ as follows.

Definition 2.1 (Identified set of θ) *(i) The identified set of θ is defined by the inverse image of $g(\cdot)$, $\Gamma(\phi)$.*

(ii) The model is point-identified at $\phi \in \Phi$, if $\Gamma(\phi)$ is a singleton. The model is set-identified at ϕ, if $\Gamma(\phi)$ is not a singleton.

Our definition of set identification given above is in fact a paraphrase of the classical definition of non-identification of the structural model. The identified set $\Gamma(\cdot)$ is seen as a multi-valued map

[2] The sufficient parameter space is unique up to one-to-one transformation (Picci (1977)).

[3] In an observationally restrictive model in the sense of Koopman and Reiersol (1950), $\hat{p}(x|\cdot)$ the likelihood function for the sufficient parameters is well defined for a domain larger than $g(\Theta)$, e.g., see Example 2.3 in Section 2.2. In this case the model possesses the refutability property, and $\Gamma(\phi)$ can be empty for some $\phi \in \Phi$. We do not consider such refutable model in this paper.

from the reduced form parameter to the structural parameter space Θ, and point-identification means this mapping is a single-valued map. The identification of θ only relies on the likelihood $p(x|\theta)$ and its definition does not change before and after observing data. Furthermore, the notion of identification stated above does not require a hypothetical argument of availability of infinite number of observations.

In the set-identified model, the parameter of interest is typically a subvector or a transformation of θ. We denote parameters of interest by a measurable transformation of θ, $\eta = h(\theta)$, $h : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$. We define the identified set of η by the projection of $\Gamma(\phi)$ onto $\mathcal{H}$ through $h(\cdot)$.

Definition 2.2 (Identified set of η) *(i) The identified set of η is defined by $H(\phi) \equiv \{h(\theta) : \theta \in \Gamma(\phi)\}$.*
(ii) The parameter $\eta = h(\theta)$ is point-identified at ϕ if $H(\phi)$ is a singleton, while η is set-identified at ϕ if $H(\phi)$ is not a singleton.

The task of constructing the sharp bounds of η in the partially identified model is equivalent to finding the expression of $H(\phi)$. As seen in the examples given below, it is often the case in partially-identified models that set-identification is an almost sure event, i.e., $\{\phi : H(\phi) \text{ is a singleton}\}$ is a measure zero set in Φ. This contrasts with so called weakly identified models where point-identification is an almost sure event in Φ. Although our framework of lower and upper probabilities do not distinguish these two cases, our analytical development focuses on partially identified models rather than weakly identified models.

2.2 Examples

We now give some examples, both to illustrate the above concepts and notations and to provide a concrete focus for later development.

Example 2.1 (Bounding Distribution of Causal Effects) *Consider the Neyman-Rubin potential outcome model with a randomized experiment. Let $D \in \{1, 0\}$ be an indicator of a binary treatment status, and $(Y_1, Y_0) \in \mathcal{Y} \times \mathcal{Y}$ be a pair of treated and control outcomes. Let Y be an observed outcome $Y = DY_1 + (1-D)Y_0$. Data is a size N random sample of $X = (Y, D)$ denoted by $x^N = \{(y_i, d_i) : i = 1, \ldots, N\}$. We assume D is independent of (Y_1, Y_0). Accordingly, parameters in the model can be specified by $\theta = (f_{Y_1,Y_0}, p)$ where f_{Y_1,Y_0} represents the joint probability density of (Y_1, Y_0) and $p \equiv \Pr(D = 1)$.*

The observed data likelihood is

$$p(x^N|\theta) = p^{n_1} (1-p)^{n_0} \prod_{i=1}^{N} [f_{Y_1}(y_i)]^{d_i} [f_{Y_0}(y_i)]^{1-d_i}$$

where f_{Y_1} *and* f_{Y_0} *are the marginal distributions of* Y_1 *and* Y_0 *respectively, and* $n_1 = \sum_{i=1}^{N} d_i$ *and* $n_0 = N - n_1$. *It can be seen that the likelihood is a function of* f_{Y_1}, f_{Y_0}, *and* p, *so that the sufficient parameters are* $\phi = (f_{Y_1}, f_{Y_0}, p)$ *and* $g : \theta \mapsto \phi$ *maps the joint distribution* f_{Y_1,Y_0} *to each marginal of* Y_1 *and* Y_0. *The identified set* $\Gamma(\phi)$ *is written as*

$$\Gamma(\phi) = \left\{ (f_{Y_1,Y_0}, p) : \int f_{Y_1,Y_0} dy_0 = f_{Y_1}, \ \int f_{Y_1,Y_0} dy_1 = f_{Y_0} \right\}.$$

Consider the average treatment effect (ATE), $E(Y_1) - E(Y_0)$, *as a parameter of interest. Clearly, ATE is pinned down by* f_{Y_1} *and* f_{Y_0}, *so* $H(\phi)$ *of Definition 2.1 is a singleton for every* ϕ, *implying that ATE is identified.*

On the other hand, $\eta = F_{Y_1 - Y_0}(0)$ *the cumulative distribution function of the individual causal effects evaluated at zero can be a parameter of interest if the researcher wants to know how much fraction of the population can be benefitted from the treatment. Now,* $H(\phi)$ *is defined as the range of* $F_{Y_1 - Y_0}(0)$ *under the constraint that the joint distribution* f_{Y_1,Y_0} *has fixed marginals* (f_{Y_1}, f_{Y_0}). *It is known that* $H(\phi)$ *is typically an interval and the closed-form expression of* $H(\phi)$ *is given by the Makarov's bounds (Makarov (1981)). Hence,* η *is set-identified. For further discussions on bounding* $F_{Y_1 - Y_0}(\cdot)$, *see Heckman Smith, and Clemens (1997), Fan and Park (2010), and Firpo and Ridder (2009)).*

Example 2.2 (Bounding ATE by Linear Programming) *Consider the treatment effect model with incompliance and a binary instrument* $Z \in \{1, 0\}$ *as considered in Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). Assume that the treatment status and the outcome of interest are both binary. Let* $Y \in \mathcal{Y}$ *be the observed outcome and* (Y_1, Y_0) *be the potential outcomes as defined in Example 2.1. We denote by* $(D_1, D_0) \in \{1, 0\}^2$ *the potential selection responses to the instrument with the observed treatment status* $D = ZD_1 + (1 - Z)D_0$. *Data consists of a random sample of* (Y_i, D_i, Z_i). *Following Imbens and Angrist (1994), we partition the population into the four subpopulations defined in terms of potential treatment selection responses,*

$$T_i = \begin{cases} c & \text{if } D_{1i} = 1 \text{ and } D_{0i} = 0 & : \text{complier}, \\ a & \text{if } D_{1i} = D_{0i} = 1 & : \text{always-taker}, \\ n & \text{if } D_{1i} = D_{0i} = 0 & : \text{never-taker}, \\ d & \text{if } D_{1i} = 0 \text{ and } D_{0i} = 1 & : \text{defier}. \end{cases}$$

where T_i *is the indicator for the type of selection responses that is latent since we cannot observe both* (D_{1i}, D_{0i}). *]*

We assume a randomized instrument, $Z \perp (Y_1, Y_0, D_1, D_0)$. *Then, the distribution of observ-*

ables and the distribution potential outcomes satisfy the following equalities: for $y \in \{1,0\}$,

$$\begin{aligned}
\Pr(Y=y,D=1|Z=1) &= \Pr(Y_1=y,T=c)+\Pr(Y_1=y,T=a), \qquad (2.2)\\
\Pr(Y=y,D=1|Z=0) &= \Pr(Y_1=y,T=d)+\Pr(Y_1=y,T=a),\\
\Pr(Y=y,D=0|Z=1) &= \Pr(Y_0=y,T=d)+\Pr(Y_1=y,T=n),\\
\Pr(Y=y,D=0|Z=0) &= \Pr(Y_0=y,T=c)+\Pr(Y_1=y,T=n).
\end{aligned}$$

With ignoring the marginal distribution of Z, a full parameter vector of the model can be specified by a joint distribution of (Y_1,Y_0,T) that consists of 16 probability masses,

$$\theta = \left(\Pr(Y_1=y,Y_0=y',T=t): y=1,0, \quad y'=1,0, \quad t=c,n,a,d\right).$$

Let ATE be the parameter of interest,

$$\begin{aligned}
\eta &\equiv E(Y_1-Y_0)=\sum_{t=c,n,a,d}\left[\Pr(Y_1=1,T=t)-\Pr(Y_0=1,T=t)\right]\\
&= \sum_{t=c,n,a,d}\sum_{y=1,0}\left[\Pr(Y_1=1,Y_0=y,T=t)+\Pr(Y_1=y,Y_0=1,T=t)\right]\\
&\equiv h(\theta).
\end{aligned}$$

The likelihood conditional on Z depends on θ only through the distribution of (Y,D), so the sufficient parameter vector consists of 8 probability masses,

$$\phi = \left(\Pr(Y=y,D=d|Z=z): y=1,0, \quad d=1,0, \quad z=1,0\right).$$

The set of equations given in (2.2) characterizes the observationally equivalent set of distributions of (Y_1,Y_0,T), $\Gamma(\phi)$, when the distribution of data is given at ϕ. Balke and Pearl (1997) derive the identified set of ATE, $H(\phi)=h(\Gamma(\phi))$, by maximizing or minimizing $h(\theta)$ subject to θ being in the probability simplex and satisfying the constraints (2.2). Since the objective function and the constraints are all linear, this optimization can be solved by linear programming and, consequently, $H(\phi)$ is obtained as the connected intervals whose lower and upper bound are the achieved minimum and maximum of the linear optimization (see Balke and Pearl (1997) for a closed-form expression of the bounds and Kitagawa (2009) for an extension to general support of Y).

Note that, In this model, special attention is needed to the sufficient parameter space Φ in order to ensure that the identified set of θ, $\Gamma(\phi)$, is nonempty. Pearl (1995) shows that the distribution of data is compatible with the instrument exogeneity condition, $Z \perp (Y_1,Y_0,D_1,D_0)$, if and only if

$$\max_d \sum_y \max_z \left\{\Pr(Y=y,D=d)|Z=z\right\} \leq 1. \qquad (2.3)$$

This implies that in order to guarantee $\Gamma(\phi) \neq \emptyset$, a prior distribution for ϕ must be supported only on the distributions of data that fulfills (2.3).

Example 2.3 (Linear Moment Inequality Model) *Consider the model where the parameter of interest $\eta \in \mathcal{H}$ is characterized by linear moment inequalities,*

$$E(m(X) - A\eta) \geq 0,$$

where the parameter space $\mathcal{H}$ is a subset of $\mathcal{R}^L$, $m(X)$ is a J-dimensional vector of known functions of an observation, and A is a $J \times L$ known constant matrix. By augmenting the J-dimesional parameter $\lambda \in [0, \infty)^J$, these moment inequalities can be written as the J-moment equalities,[4]

$$E(m(X) - A\eta - \lambda) = 0.$$

We let the full parameter vector $\theta = (\eta, \lambda) \in \mathcal{H} \times [0, \infty)^J$.

To obtain a likelihood function for the moment equality model, we employ the exponentially tilted empirical likelihood for θ that is considered in the Bayesian context in Schennach (2005). Let x^N be a size N random sample of observations and let $g(\theta) = A\eta + \lambda$. If the convex hull of $\cup_i \{m(x_i) - g(\theta)\}$ contains the origin, then, the exponentially tilted empirical likelihood is written as

$$p(x^N|\theta) = \prod w_i(\theta)$$

where

$$w_i(\theta) = \frac{\exp\{\gamma(g(\theta))'(m(x_i) - g(\theta))\}}{\sum_{i=1}^N \exp\{\gamma(g(\theta))'(m(x_i) - g(\theta))\}},$$

$$\gamma(g(\theta)) = \arg\min_{\gamma \in R_+^J} \left\{ \sum_{i=1}^N \exp\{\gamma'(m(x_i) - g(\theta))\} \right\}.$$

Thus, the parameter $\theta = (\eta, \lambda)$ enters in the likelihood only through $g(\theta) = A\eta + \lambda$, so we take $\phi = g(\theta)$ as the sufficient parameters. The identified set for θ is given by,

$$\Gamma(\phi) = \left\{(\eta, \lambda) \in \mathcal{H} \times [0, \infty)^L : A\eta + \lambda = \phi\right\}$$

If we consider the coordinate projection of $\Gamma(\phi)$ onto $\mathcal{H}$, we obtain $H(\phi)$ the identified set for η.

Example 2.4 (Bounding Regression Coefficients with Errors in Regressors) *One of early developments of the partially identified model appears in the linear regression model with errors*

[4] The Bayesian formulation of the moment inequality model shown here owes to Tony Lancaster, who suggested this to us in 2006.

in regressors. Frisch (1934) considers the single regressor case and Klepper and Leamer (1984)) extend it to the case with multiple regressors. These authors derive the identified sets of the regression coefficients under a weak set of assumptions on errors in variables.[5] *Here, we present how to formulate the analysis of Klepper and Leamer within our framework. For simplicity, consider a case where there are only two regressors.*
Let

$$Y = X_1^* \beta_1 + X_2^* \beta_2 + \epsilon$$

be the regression equation of Y onto two regressors (X_1^*, X_2^*), and the parameter of interest here is $\beta = (\beta_1, \beta_2)'$. The means of Y and (X_1^*, X_2^*) can be normalized to zero without affecting the construction of the identified set. We assume normality of ϵ and (X_1^*, X_2^*) with variance σ_ϵ^2 and Σ. In data, precise measurement of (X_1^*, X_2^*) is not available, while their noisy measurement (X_1, X_2) is available. We introduce the classical measurement error assumption, $X_1 = X_1^* + v_1$ and $X_2 = X_2^* + v_2$ with $(v_1, v_2)' \sim \mathcal{N}(0, V)$ independent of (ϵ, X_1^*, X_2^*) and V is a diagonal variance matrix.[6] *The model parameter is specified as $\theta \equiv \left(\beta, \sigma_\epsilon^2, \Sigma, V\right)$ and the distribution of data is written in terms of θ as*

$$\begin{pmatrix} Y \\ X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_\epsilon^2 + \beta'\Sigma\beta & \beta'\Sigma \\ \Sigma\beta & \Sigma + V \end{pmatrix}\right). \tag{2.4}$$

With a random sample of (Y, X_1, X_2), the value of likelihood function varies only with the second moments of (Y, X_1, X_2), so the reduce form parameter ϕ consists of the second moments of (Y, X_1, X_2), $\phi \equiv \left(s_y^2, r, \Omega\right)$ where $s_y^2 = E\left(Y^2\right)$, $r = E\left(Y, (X_1, X_2)'\right)$, and $\Omega = E\left[(X_1, X_2)'(X_1, X_2)\right]$. From (2.4), the relationship between the reduced form parameter ϕ and the model parameter θ are obtained as

$$\begin{aligned} s_y^2 &= \sigma_\epsilon^2 + \beta'\Sigma\beta, \\ r &\equiv \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \equiv \Sigma\beta, \\ \Omega &\equiv \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix} \equiv \Sigma + V. \end{aligned}$$

With fixing the left hand side variable $\phi = \left(s_y^2, r, \Omega\right)$, these equalities defines $\Gamma(\phi)$ the identified set of $\theta = \left(\beta, \sigma_\epsilon^2, \Sigma, V\right)$. In particular, the identified set for β consists of the values of β for each

[5] They refer to the identified set as "the set of maximum likelihood estimates."

[6] As Klepper and Leamer (1984) noted, the normality assumptions of ϵ and (v_1, v_2) are not necessarily as long as we are concerned with identifying the coefficients using the first and second moments of the data.

of which there exist nonnegative σ_ϵ^2 *and positive semidefinite matrices for* Σ *and* V *satisfying the above equations.*

Let $b^1 = \left(b_1^1, b_2^1\right)$ *be the coefficient vector of linear projection of* Y *onto* (X_1, X_2), b^2 *be the coefficient vector for* (X_1, X_2) *obtained by regressing* X_1 *onto* (Y, X_2) *and solving for* Y, *and* b^3 *be obtained by regressing* X_2 *onto* (Y, X_1) *and solving for* Y. *Note that all* b^1 *through* b^3 *are expressed in terms of the reduced form parameters* $\phi = \left(s_y^2, r, N\right)$. *The elegant analysis in Klepper and Leamer (1984, Theorem 1 - 3) shows that if* b^1, b^2, *and* b^3 *all lie in the same orthant the identified set of* β *is given by their convex hull, and otherwise, it is an unbounded set. Specifically, if our interest is in the first coefficient* β_1, *application of Theorem 1-4 in in Klepper and Leamer (1984) provides the following closed-form expression of the marginalized identified set for* β_1,

$$\begin{cases} \min\left\{b_1^1, b_1^2, b_1^3\right\} \leq \beta_1 \leq \max\left\{b_1^1, b_1^2, b_1^3\right\} & \text{if } sgn\left(b_1^1\right) sgn\left(b_2^1\right) = sgn\left(r_1 r_2 - s_y^2 \sigma_{12}\right) \\ (-\infty, \infty) & \text{otherwise} \end{cases}$$

This identified set can be unbounded depending on the distribution of data. The inferential framework of this paper can deal with unbounded identified sets like this.

3 Multiple-Prior Analysis and the Lower and Upper Probability

3.1 Posterior of θ in the Presence of Sufficient Parameters

Let μ_θ be a prior of θ and μ_ϕ be the marginal probability measure on the sufficient parameter space $(\Phi, \mathcal{B})$ implied by μ_θ and $g(\cdot)$, i.e.,

$$\mu_\phi(B) = \mu_\theta(\Gamma(B)) \quad \text{for all } B \in \mathcal{B}.$$

Let $x \in \mathbf{X}$ be sampled data, and assume that μ_ϕ has a dominating measure with respect to which its Radon-Nykodim derivative exists, $d\mu_\phi / d\phi = \tilde{\mu}_\phi$. We denote the indicator function on set A by $1_A(\cdot)$. The posterior distribution of θ denoted by $F_{\theta|X}(\cdot)$ is obtained as, for $A \in \mathcal{A}$,

$$\begin{aligned} F_{\theta|X}(A) &= \frac{\int_\Theta p(x|\theta) 1_A(\theta) d\mu_\theta}{\int_\Theta p(x|\theta) d\mu_\theta} \\ &= \frac{\int_\Phi E[p(x|\theta) 1_A(\theta)|\phi] d\mu_\phi}{\int_\Phi E[p(x|\theta)|\phi] d\mu_\phi} \\ &= \int_\Phi E(1_A(\theta)|\phi) \frac{\hat{p}(x|\phi)\tilde{\mu}_\phi(\phi)}{\int_\Phi \hat{p}(x|\phi) d\mu_\phi} d\phi \\ &= \int_\Phi \mu_{\theta|\phi}(A|\phi) f_{\phi|X}(\phi) d\phi, \end{aligned} \tag{3.1}$$

where the second equality follows by the definition of conditional expectation and the transformation of random variables from θ to ϕ, and the third equality follows because $p(x|\theta) = \hat{p}\,(x|\phi)$. $\mu_{\theta|\phi}(A|\phi)$ denotes the conditional distribution of θ given ϕ that satisfies

$$\mu_{\theta}(A \cap \Gamma(B)) = \int_B \mu_{\theta|\phi}(A|\phi) d\mu_{\phi} \quad \text{for all} \;\; A \in \mathcal{A}, \;\; B \in \mathcal{B}.$$

$f_{\phi|X}(\phi)$ is the posterior density of ϕ,

$$f_{\phi|X}(\phi) = \frac{\hat{p}(x|\phi)\tilde{\mu}_{\phi}(\phi)}{\int_{\Phi} \hat{p}(x|\phi) d\mu_{\phi}}.$$

The expression of the posterior of θ (3.1) highlights the fundamental feature of the non- or set-identified model: the posterior of θ is the average of the conditional prior $\mu_{\theta|\phi}(A|\phi)$ weighted by the posterior density of the sufficient parameter $f_{\phi|X}(\phi)$. This result is well known in the literature (Barankin (60), Florens and Mouchart (74), Picci (77), Dawid (1979), Poirier (1998), among many others), and it can be interpreted that the data only allows us to revise belief on the sufficient parameters ϕ, while it does not for the conditional prior of θ given ϕ.

3.2 The Class of Unrevisable Prior Knowledge

The posterior distribution of θ given in (3.1) implies that we are incapable of updating the conditional prior information of θ given ϕ due to the flat likelihood on the equivalence classes $\Gamma(\phi) \subset \Theta$. Accordingly, we can consider the prior information marginalized to the sufficient parameter μ_{ϕ} as the *revisable prior knowledge* and the conditional prior of θ given ϕ, $\mu_{\theta|\phi}$, as the *unrevisable prior knowledge.*

If we want to obtain posterior uncertainty of θ in the form of a probability distribution on $(\Theta, \mathcal{A})$ as desired in the Bayesian paradigm, we need to have a single prior distribution of θ, and this requires us to specify the unrevisable prior knowledge $\mu_{\theta|\phi}$. No matter how we specify $\mu_{\theta|\phi}$, the specification of $\mu_{\theta|\phi}$ always carries an assumption on θ in the form of a probability distribution, and no specification of $\mu_{\theta|\phi}$ is able to represent the lack of prior knowledge. If the researcher could justify his choice of $\mu_{\theta|\phi}$ by any credible prior information, the standard Bayesian updating (3.1) would yield a valid posterior distribution of θ. From the robustness point of view, however, a statistical procedure that requires us to specify the unrevisable prior knowledge may be less desirable when the researcher cannot or is not willing to translate his vague or vacuous prior belief into a probability distribution about θ.

One way to deal with such lack of prior knowledge in statistical inference is to introduce a class of prior distributions in the Bayesian framework. Let $\mathcal{M}$ be the set of probability measures on $(\Theta, \mathcal{A})$ and μ_{ϕ} be a prior on $(\Phi, \mathcal{B})$ specified by the researcher. Consider the class of prior distributions of θ defined by

$$\mathcal{M}(\mu_\phi) = \left\{ \mu_\theta \in \mathcal{M} : \mu_\theta(\Gamma(B)) = \mu_\phi(B) \text{ for every } B \in \mathcal{B} \right\}.$$

$\mathcal{M}(\mu_\phi)$ consists of prior distributions of θ whose marginal for the sufficient parameters coincides with the prespecified μ_ϕ. In other words, we accept specifying a single prior distribution for the sufficient parameters ϕ while we allow for arbitrary conditional priors $\mu_{\theta|\phi}$ as far as $\mu_\theta(\cdot) = \int_\Phi \mu_{\theta|\phi}(\cdot|\phi) d\mu_\phi$ is a probability measure on $(\Theta, \mathcal{A})$.

There are several reasons for considering this class. First, our goal is to make inference and decision on the parameter of interest essentially the same for any of empirically unrevisable assumptions. The given choice of prior class contains any $\mu_{\theta|\phi}$ so that we can achieve it by considering an inference procedure that operates on the class of posteriors induced by the prior class $\mathcal{M}(\mu_\phi)$. Second, as we shall show below, if we have a closed-form expression of $H(\phi)$ or at least are able to compute $H(\phi)$, the prior class $\mathcal{M}(\mu_\phi)$ yields analytically tractable form of the posterior lower probability for η and enables us to implement the lower probability inference by the standard Bayesian computing of Markov Chain Monte Carlo.

As the final remark of this section, we comment that existing selection rules for "noninformative" or "recommended" prior for finite dimensional θ are not applicable if the model lacks identification. First of all, Jeffreys's general rule (Jeffreys (1961)), which takes the prior density to be proportional to the square root of the determinant of the Fisher information matrix, is not well defined if the information matrix for θ is nonsigular at almost every $\theta \in \Theta$.

The empirical Bayes type approach of choosing a prior for θ (Robbins (1964), Good (1965), Morris (1983), and Berger and Berliner (1986)) breaks down if the model lacks identification. To see why, consider finding a prior within the class that maximizes the marginal likelihood of data, i.e., we want to choose μ_θ so as to maximize

$$m(x|\mu_\theta) = \int_\Theta p(x|\theta) d\mu_\theta$$

in $\mu_\theta \in \mathcal{M}(\mu_\phi)$. If the likelihood involves sufficient parameters, the marginal distribution $m(x|\mu_\theta)$ depends only on μ_ϕ, because

$$\int_\Theta p(x|\theta) d\mu_\theta = \int_\Phi \hat{p}(x|\phi) d\mu_\phi \equiv m(x|\mu_\phi). \tag{3.2}$$

Hence, the empirical Bayes approach fails to pick a prior for θ out of $\mathcal{M}(\mu_\phi)$.

It is also worth noting that we cannot obtain the reference prior of Bernardo (1979), which is selected to maximize the conditional Kullback-Leibler distance between the posterior density $f_{\theta|X}(\theta)$ and a prior density $d\mu_\theta(\theta)$,

$$\int_\Theta \log \left(\frac{f_{\theta|X}(\theta)}{d\mu_\theta(\theta)} \right) dF_{\theta|X}(\theta).$$

It can be shown that this objective function again depends only on μ_ϕ.

These prior selection rules are useful only for choosing a prior for the sufficient parameters μ_ϕ, but not at all for selecting μ_θ out of $\mathcal{M}(\mu_\phi)$. In the analysis given below, we shall not discuss how to select μ_ϕ, and treat μ_ϕ as given. As discussed already, the existing prior selection rule is potentially useful to define objective or noninformative μ_ϕ. Moreover, the influence of μ_ϕ to the posterior of ϕ diminishes as the sample size increases, so the sensitivity issue of the posterior of ϕ is not so severe when the sample size is moderate or large.

3.3 Posterior Lower and Upper Probabilities

The prior class $\mathcal{M}(\mu_\phi)$ results in yielding the class of posterior distributions of θ. We summarize the posterior class by the posterior lower probability $F_{\theta|X*}(\cdot)$ and the posterior upper probability $F^*_{\theta|X}(\cdot)$, which are defined as, for $A \in \mathcal{A}$,

$$\begin{aligned} F_{\theta|X*}(A) &\equiv \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(A), \\ F^*_{\theta|X}(A) &\equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(A). \end{aligned}$$

Note that the posterior lower probability and the upper probability have a conjugate property, $F_{\theta|X*}(A) = 1 - F^*_{\theta|X}(A^c)$, so it suffices to focus on one of them in deriving their analytical form.

For the lower and upper probabilities to be well defined, we assume the following regularity conditions.

Condition 3.1 *(i) A prior for ϕ, μ_ϕ, is a proper probability measure on $(\Phi, \mathcal{B})$, and it is absolutely continuous with respect to a σ-finite measure on $(\Phi, \mathcal{B})$.*
(ii) $g : (\Theta, \mathcal{A}) \to (\Phi, \mathcal{B})$ is measurable and its inverse image $\Gamma(\phi)$ is a closed set in Θ, μ_ϕ-almost every $\phi \in \Phi$.
(iii) $h : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$ is measurable and $H(\phi) = h(\Gamma(\phi))$ is a closed set in $\mathcal{H}$, μ_ϕ-almost every $\phi \in \Phi$.

The first condition is imposed in order for the prior class $\mathcal{M}(\mu_\phi)$ to be analytically tractable. Although the inference procedure proposed in this paper can be implemented as long as the posterior of ϕ is proper, we do not know how to accommodate an improper prior for ϕ in our development of the analytical results. The second and the third conditions are required for identified sets $\Gamma(\phi)$ and $H(\phi)$ to be interpreted as random closed sets in Θ and $\mathcal{H}$ induced by a probability law on $(\Phi, \mathcal{B})$. A sufficient condition for closedness of $\Gamma(\phi)$ is, for instance, continuity of $g(\cdot)$. Under these conditions, we can guarantee that for each $A \in \mathcal{A}$, $\{\phi : \Gamma(\phi) \cap A = \emptyset\}$ and $\{\phi : \Gamma(\phi) \subset A\}$ are supported by a probability measure on $(\Phi, \mathcal{B})$.

Theorem 3.1 *Assume Condition 3.1.*
(i) For each $A \in \mathcal{A}$,

$$F_{\theta|X*}(A) = F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A\}), \tag{3.3}$$
$$F^*_{\theta|X}(A) = F_{\phi|X}\left(\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}\right), \tag{3.4}$$

where $F_{\phi|X}(B)$ is the posterior probability measure of ϕ, $F_{\phi|X}(B) = \int_B f_{\phi|X}(\phi)d\phi$ for $B \in \mathcal{B}$.
(ii) Define the posterior lower and upper probabilities of $\eta = h(\theta)$ by for each $D \in \mathcal{D}$,

$$F_{\eta|X*}(D) \equiv \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(h^{-1}(D)),$$
$$F^*_{\eta|X}(D) \equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(h^{-1}(D)).$$

They are given by

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

Proof. For a proof of (i), see Appendix A. For a proof of (ii), see equation (3.5) below. ■

The expression of $F_{\theta|X*}(A)$ given above implies that the posterior lower probability on A calculates the probability that the random set $\Gamma(\phi)$ is contained in subset A in terms of the posterior probability law of ϕ. On the other hand, the upper probability is interpreted as the posterior probability that the random set $\Gamma(\phi)$ hits the subset A. Since the probability law of random sets is uniquely characterized by the containment probability as in (3.3) or the hitting probability as in (3.4), these results show that the posterior lower and upper probabilities for θ and η with our specification of the prior class represents the posterior probability law of the identified sets of θ and η.

The result of Theorem 3.1 (i) can be seen as a special case of Wasserman (1990)'s general construction of the posterior lower and upper probabilities. In the general set-up of Wasserman (1990), the posterior lower and upper probabilities do not necessarily represent the distribution of random sets. The reason that we can a posteriori obtain probability laws of the random sets is due to the special way of constructing the prior class. Our proof given in Appendix A is not restricted to finite dimensional θ, and the results hold even for infinite dimensional Θ (e.g., Example 2.1) as long as Θ is a complete separable metric space.

As is well known in the literature, the lower and upper probabilities of a set of probability measure do not necessarily satisfy additivity, i.e., for disjoint subsets A_1 and A_2 in $\mathcal{A}$, $F_{\theta|X*}$ is supadditive, $F_{\theta|X*}(A_1 \cup A_2) \geq F_{\theta|X*}(A_1) + F_{\theta|X*}(A_2)$, and $F^*_{\theta|X}$ is subadditive $F^*_{\theta|X}(A_1 \cup A_2) \leq F^*_{\theta|X}(A_1) + F^*_{\theta|X}(A_2)$. Note that if the model is identified in the sense of

$\Gamma(\phi)$ being a singleton $f_{\phi|X}$-almost surely, then $F_{\theta|X*}(\cdot) = F^*_{\theta|X}(\cdot)$ holds and they become an identical probability measure.

The second statement of the theorem provides a procedure to transform or marginalize the lower and upper probabilities of θ into the ones of the parameter of interest η. The expressions of $F_{\eta|X*}(D)$ and $F^*_{\eta|X}(D)$ are simple and easy to interpret: the lower and upper probabilities of $\eta = h(\theta)$ are the containment and hitting probabilities of the random sets obtained by projecting $\Gamma(\phi)$ through $h(\cdot)$. This marginalization rule of the lower probability follows from

$$\begin{aligned} F_{\eta|X*}(D) &= F_{\theta|X*}(h^{-1}(D)) \\ &= F_{\phi|X}(\{\phi : \Gamma(\phi) \subset h^{-1}(D)\}) \\ &= F_{\phi|X}(\{\phi : H(\phi) \subset D\}). \end{aligned} \tag{3.5}$$

and, the marginalization rule for the upper probability is obtained similarly. Analogous to the lower and upper probabilities of θ, $F_{\eta|X*}(\cdot)$ and $F^*_{\eta|X}(\cdot)$ are non-additive measures on $(\mathcal{H}, \mathcal{D})$ (supadditive and subadditive measures respectively) if the model lacks identification of η.

4 Point Decision of $\eta = h(\theta)$ with Multiple Priors

In the standard Bayes posterior analysis, the mean or median of a posterior distribution of η is often reported as a point estimate of η. If we summarize posterior information of η in terms of its posterior lower and upper probabilities instead of a posterior distribution, how should we construct a reasonable point estimator for η?

Being motivated by such question, we consider the point decision problem for the parameter of interest $\eta = h(\theta)$ under several risk criteria. In Section 4.1, we consider the decision problem under the *posterior gamma-minimax criterion* (Berger (1985, p205)). In the presence of multiple priors, it is known that a priori optimal decision may differ from a posteriori optimal action (dynamic inconsistency problem, Vidakovic (2000).) Hence, we examine in Section 4.2 whether the posterior gamma-minimax action obtained in Section 4.1 coincides with the *unconditional* gamma-minimax decision (Kudo (1967), Berger (1985, p213-218), Vidakovic (2000)).

4.1 Posterior Gamma-minimax Action

We first consider point estimator of $\eta = h(\theta) \in \mathcal{H}$ that is *a posteriori* optimal in the sense of the posterior gamma-minimax criterion. Let $a \in \mathcal{H}_a$ be an action where $\mathcal{H}_a \subset \mathcal{H}$ is an action space. Here, action is interpreted as reporting a particular point estimate for η. Given action a to be taken and η_0 being the true state of nature, a loss function $L(\eta_0, a) : \mathcal{H} \times \mathcal{H}_a \to \mathcal{R}_+$ yields how much cost the decision maker owes by taking such action.

Given μ_θ, prior for θ, the *posterior risk is defined by,*

$$\rho(\mu_\theta, a) \equiv \int_{\mathcal{H}} L(\eta, a) dF_{\eta|X}(\eta) \tag{4.1}$$

where the first argument μ_θ represents the dependence of the posterior of η on the specification of prior on $(\Theta, \mathcal{A})$. Our posterior analysis involves the multiple posterior distributions of η, so the class of posterior risks $\left\{\rho(\mu_\theta, a) : \mu_\theta \in \mathcal{M}(\mu_\phi)\right\}$ is considered. The *posterior gamma-minimax criterion*[7] ranks actions in terms of the worst case posterior risk (upper posterior risk),

$$\rho^*(\mu_\phi, a) \equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \rho(\mu_\theta, a).$$

Definition 4.1 *A posterior gamma-minimax action a_x^* is an action that minimizes the upper posterior risk, i.e.,*

$$\rho^*(\mu_\phi, a_x^*) = \inf_{a \in \mathcal{H}_a} \rho^*(\mu_\phi, a) = \inf_{a \in \mathcal{H}_a} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \rho(\mu_\theta, a).$$

The gamma-minimax decision approach involves a favor for a conservative action that guards against the least favorable prior within the class, and it can be seen as a compromise of the Bayesian decision principle and the minimax decision principle. To establish an analytical result for the gamma-minimax action, we introduce the following regularity conditions.

Condition 4.1 *(i) For each $a \in \mathcal{H}_a$, loss function $L(\eta, a)$ is $\mathcal{D}$-measurable and nonnegative.*
(ii) Given a prior for ϕ, the upper posterior probability for θ, $F^*_{\theta|X}(\cdot)$ obtained in Theorem 3.1 is regular, i.e., for each $A \in \mathcal{A}$,

$$F^*_{\theta|X}(A) = \sup\left\{F^*_{\theta|X}(K) : K \subset A,\ K \text{ compact}\right\} = \inf\left\{F^*_{\theta|X}(G) : A \subset G,\ G \text{ open}\right\}.$$

Regularity of the upper posterior probability stated in Condition 4.1 (ii) is satisfied if no particular realizations of the random closed set $\Gamma(\phi)$ occurs with a strictly positive probability (see Graf (1980)), which, in turn, demands μ_ϕ not to have a probability mass. Under these conditions, the next proposition shows that the upper posterior risk $\rho^*(\mu_\phi, a)$ is written in a more analytically tractable form.

[7] In the robust Bayes literature, the class of prior distributions is often notated as Γ, and this is why it is called (conditional) gamma-minimax criterion. Unfortunately, in the literature of belief function and the lower and upper probabilities, Γ often denotes a set-valued mapping that generates the lower and upper probabilities. In this article, we adopt the notational convention of the latter, while still refer to the decision criterion as the gamma minimax criterion.

Proposition 4.1 *Under Condition 3.1 and Condition 4.1, the upper posterior risk satisfies*

$$\rho^*(\mu_\phi, a) = \int L(\eta, a) dF^*_{\eta|X}(\eta) = \int_\Phi \sup_{\eta \in H(\phi)} L(\eta, a) f_{\phi|X}(\phi|x) d\phi, \tag{4.2}$$

where $\int L(\eta, a) dF^*_{\eta|X}(\eta)$ *is in the sense of Choquet integral, i.e.,*

$$\int L(\eta, a) dF^*_{\eta|X}(\eta) = \int_0^\infty F_{\eta|X}\left(\{\eta \in \mathcal{H} : L(\eta, a) \geq t\}\right) dt.$$

Accordingly, a^*_x *exists if and only if* $E_{\phi|X}\left(\sup_{\eta \in H(\phi)} L(\eta, a)\right)$ *has a minimizer in* $a \in \mathcal{H}_a$.

Proof. See Appendix A. ■

The third expression in (4.2) shows that the posterior gamma-minimax criterion is written as the expectation of the worst-case loss function $\sup_{\eta \in H(\phi)} L(\eta, a)$ with respect to the posterior of ϕ. The supremum part stems from ambiguity of η: given ϕ, what the researcher knows about η is only that it lies within the identified set $H(\phi)$, so the researcher forms loss by supposing that the worst case would happen. On the other hand, the expectation in ϕ represents posterior uncertainty of the identified set $H(\phi)$. The gamma minimax criterion with class of priors $\mathcal{M}(\mu_\phi)$ combines such ambiguity of η with posterior uncertainty of the identified set to yield a single objective function to be minimized.

Although a closed form expression of a^*_x is not in general available, this proposition suggests a simple numerical algorithm to approximate a^*_x using a random sample of ϕ from its posterior $f_{\phi|X}(\phi|x)$. Let $\{\phi_s\}_{s=1}^S$ be S random draws of ϕ from posterior $f_{\phi|X}(\phi)$. We will approximate a^*_x by

$$\hat{a}^*_x \equiv \arg\min_{a \in \mathcal{H}_a} \frac{1}{S} \sum_{s=1}^S \sup_{\eta \in H(\phi_s)} L(\eta, a).$$

4.2 Gamma-minimax Decision

In this section, we consider a decision rule when an optimal decision is made in prior to observing data. Let $\delta(\cdot)$ be a decision rule that maps each $x \in \mathbf{X}$ to the action space $\mathcal{H}_a \subset \mathcal{H}$ and let Δ be the space of decisions (the set of functions: $\mathbf{X} \to \mathcal{H}_a$.) The Bayes risk is defined as usual,

$$r(\mu_\theta, \delta) = \int_\Theta \left[\int_{\mathcal{X}} L(h(\theta), \delta(x)) p(x|\theta) dx \right] d\mu_\theta. \tag{4.3}$$

Given the prior class $\mathcal{M}(\mu_\phi)$, the *gamma-minimax criterion* ranks decisions in terms of the supremum of the Bayes risk, $r^*(\mu_\phi, \delta) \equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} r(\mu_\theta, \delta)$. Accordingly, the optimal decision under this criterion is defined as follows.[8]

Definition 4.2 *A gamma-minimax decision $\delta^* \in \Delta$ is a decision rule that minimizes the upper Bayes risk, i.e.,*

$$r^*(\mu_\phi, \delta^*) = \inf_{\delta \in \Delta} r^*(\mu_\phi, \delta) = \inf_{\delta \in \Delta} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} r(\mu_\theta, \delta).$$

In the standard Bayes decision problem with a single prior, the Bayes rule that minimizes $r(\mu_\theta, \delta)$ coincides with the posterior Bayes action for every possible data $x \in \mathbf{X}$, and either being unconditional or conditional on data does not matter for the actual action to be taken. With multiple priors, however, the decision rule that minimizes $r^*(\mu_\phi, \delta)$ in general does not coincide with the conditional gamma minimax action (see, e.g., Betro and Ruggeri (1992)). This phenomenon can be easily understood by writing the Bayes risk as the average of the posterior risk with respect to the marginal distribution of data: interchanging the order of integrations in (4.3) gives

$$r(\mu_\theta, \delta) = \int_{\mathbf{X}} \rho(\mu_\theta, \delta(x)) m(x|\mu_\theta) dx. \tag{4.4}$$

Given $\delta \in \Delta$ and the class of priors, μ_θ that maximizes $r(\mu_\theta, \delta)$ does not necessarily maximizes the posterior risk $\rho(\mu_\theta, \delta(x))$ since the Bayes risk $r(\mu_\theta, \delta)$ depends on μ_θ not only through the posterior risk $\rho(\mu_\theta, \delta(x))$ but also through the marginal distribution of data $m(x|\mu_\theta)$. Recall, however, that the marginal distribution of data depends only on μ_ϕ (see (3.2)) and our class of priors admits only one specification of μ_ϕ. As a result, the supremum of the Bayes risk can be written as the supremum of the posterior risk averaged by $m(x|\mu_\theta)$,

$$\sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} r(\mu_\theta, \delta) = \int_{\mathbf{X}} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \rho(\mu_\theta, \delta(x)) m(x|\mu_\phi) dx.$$

Hence, the unconditional gamma-minimax decision that minimizes the left hand side should coincide with the posterior gamma-minimax action at $m(x|\mu_\phi)$-almost every x.

Proposition 4.2 (No Dynamic Inconsistency) *If (4.4) holds, $\delta^*(x) = a_x^*$, $m(x|\mu_\phi)$-almost surely.*

[8] A decision criterion similar to the one considered here appears in Kudō (1967), Manski (1981), and Lambert and Duncan (1986). These literatures considered the model where the subjective probability distribution on the state of nature can be elicited only up to the class of coarse subsets of the parameter space. Our decision problem shares a similar feature to theirs since the posterior upper probability of η can be viewed as a posterior probability distribution over the coarse collection of subsets $\{H(\phi) : \phi \in \Phi\} \subset \mathcal{H}$.

Proof. A proof is given above. ■

As an alternative to the (posterior) gamma-minimax criterion considered above, it is possible to consider the gamma-minimax *regret* criterion (Berger (1985, p218) and Rios Insua, Ruggeri, and Vidakovic (1995)), which is seen as an extension of the minimax regret criterion of Savage (1951) to the multiple prior Bayes context. In Appendix B, we provide some analytical results of the gamma minimax regret analysis where the parameter of interest η is a scalar and the loss function is quadratic, $L(\eta, a) = (\eta - a)^2$, and demonstrate that the gamma minimax regret decision does not differ much from the gamma minimax decision derived above, and they coincide with a large sample.

5 Set Estimation of η

This section discusses how to use the posterior lower probability of η to conduct set estimation of η. In the standard Bayesian inference, set estimation is done by reporting contour sets of the posterior probability density of η. If the posterior information for η is summarized by the lower and upper probabilities, how should we conduct set estimation of η?

5.1 Posterior Lower Credible Region

Consider subset $C_\alpha \subset \mathcal{H}$ such that the posterior lower probability $F_{\eta|X*}(C_\alpha)$ is equal or greater than α,

$$F_{\eta|X*}(C_\alpha) = F_{\phi|X}(H(\phi) \subset C_\alpha)) \geq \alpha.$$

In words, C_α is interpreted as "a set on which the posterior credibility of η is *at least α irrespective of the unrevisable prior knowledge.*" If we drop the italicized part from this statement, we obtain the usual interpretation of the posterior credible region, so C_α defined in this way seems to be a natural way to extend the Bayesian posterior credible region to our analysis of the posterior lower probabilities. Analogous to the Bayesian posterior credible region, there are also multiple C_α's that satisfies this requirement.[9] In what follows, we focus on constructing the smallest posterior credible region among these C_α's, which we refer to it as a *posterior lower credible region* with

[9] For instance, given a posterior credibility region of ϕ with credibility α, say B_α, the subset in $\mathcal{H}$ defined by $\cup_{\phi \in B_\alpha} H(\phi)$ can be also interpreted as C_α. Moon and Schorfheide (2010) interpret thus constructed credible region as a Bayesian credibility set for the *identified set*. In our analysis, our object of interest is parameter η and the posterior lower credible region C_α can be interpreted as a robustified posterior credible region for η.

credibility α,

$$C_{\alpha*} \equiv \arg\min_{C \in \mathcal{C}} Leb(C) \tag{5.1}$$
$$\text{s.t.} \quad F_{\phi|X}(H(\phi) \subset C)) \geq \alpha,$$

where $Leb(C)$ is the volume of subset C in terms of the Lebesgue measure and $\mathcal{C}$ is a family of subsets in $\mathcal{H}$ over which the volume minimizing lower credible region is searched.

Finding $C_{\alpha*}$ is difficult if η is multi-dimensional and no restriction is placed on the class of subsets $\mathcal{C}$. In what follows, we restrict $\mathcal{C}$ to the class of closed balls and propose a method to calculate $C_{\alpha*}$. Note that, for scalar η, the class of closed balls contains any connected intervals, and even when η is a vector, we can construct the marginal posterior lower credible region for each element in η based upon the projected identified set of η.

Let $B_r(\eta_c)$ be a closed ball in $\mathcal{H}$ centered at $\eta_c \in \mathcal{H}$ with radius r. If $\mathcal{C}$ is constrained to the class of closed balls, the constrained minimization problem of (5.1) is reduced to

$$\min_{r,\eta_c} r \tag{5.2}$$
$$\text{s.t.} \qquad F_{\phi|X}\left(H(\phi) \subset B_r(\eta_c)\right) \geq \alpha.$$

This optimization problem can be solved by focusing on the α-th quantiles of the posterior distribution of the directed Hausdorff distance measured from $\eta_c \in \mathcal{H}$ to a random set $H(\phi)$.

Proposition 5.1 *Let* $\overrightarrow{d} : \mathcal{H} \times \mathcal{D} \to \mathcal{R}_+$ *be*

$$\overrightarrow{d}(\eta_c, H(\phi)) \equiv \sup_{\eta \in H(\phi)} \left\{ \|\eta_c - \eta\| \right\}.$$

For each $\eta_c \in \mathcal{H}$*, let* $r_\alpha(\eta_c)$ *be the* α*-th quantile of the distribution of* $\overrightarrow{d}(\eta_c, H(\phi))$ *induced by the posterior distribution of* ϕ*, i.e.,*

$$r_\alpha(\eta_c) \equiv \inf\left\{ r : F_{\phi|X}\left(\left\{\phi : \overrightarrow{d}(\eta_c, H(\phi)) \leq r\right\}\right) \geq \alpha \right\}.$$

Then, the solution of the constrained minimization problem (5.2) is given by (r_α^*, η_c^*) *where*

$$r_\alpha^* = r_\alpha(\eta_c^*) \quad \textit{where} \quad \eta_c^* = \arg\min_{\eta_c \in \mathcal{H}} r_\alpha(\eta_c).$$

Proof. See Appendix A. ■

Given random draws of ϕ from its posterior, this proposition suggests a straightforward way to compute an approximate of the posterior lower credible region. Let $\{\phi_s : s = 1, \ldots, S\}$ be random draws of ϕ from its posterior. At each $\eta_c \in \mathcal{H}$, we first calculate $\hat{r}_\alpha(\eta_c)$ the empirical α-th quantile of $\overrightarrow{d}(\eta_c, H(\phi))$ based on the simulated $\overrightarrow{d}(\eta_c, H(\phi_s))$, $s = 1, \ldots, S$. Provided that the posterior distribution of $\overrightarrow{d}(\eta_c, H(\phi))$ is continuous, the obtained empirical α-th quantile $\hat{r}_\alpha(\eta_c)$ should be a valid approximate for the underlying α-th quantile $r_\alpha(\eta_c)$, so we can approximate (r_α^*, η_c^*) by searching the minimizer and the minimized value of $\hat{r}_\alpha(\eta_c)$.

5.2 Asymptotic Properties of the Posterior Lower Probability

This section examines the large sample behavior of the posterior lower probability and the posterior lower credible region $C_{\alpha*}$ constructed above. We make the sample size explicit in our notation: X^N denote size N observations generated from $\hat{p}(x^N|\phi_0)$ where ϕ_0 denotes the sufficient parameter value that corresponds to the true data generating process. A limiting sampling sequence is denoted by X^∞ and its distribution is denoted by $\hat{p}(x^\infty|\phi_0)$. We denote the maximum likelihood estimator for ϕ by $\hat{\phi}$.

For ease of analysis, we restrict our analysis to the case where the sufficient parameter space Φ is finite dimensional (e.g., Example 2.2 - 2.4.) and we assume the following two sets of the regularity conditions.

Condition 5.1 *(i) $H(\phi_0)$ is a bounded subset of $\mathcal{H}$.*

(ii) Let $d_H(\cdot,\cdot)$ be the Hausdorff metric defined by

$$d_H(D_1, D_2) = \max\left\{ \sup_{\eta_1 \in D_1} d(\eta_1, D_2), \sup_{\eta_2 \in D_2} d(\eta_2, D_1) \right\}, \qquad D_1, D_2 \subset \mathcal{H},$$

where $d(\eta, D) = \inf_{\eta' \in D} \|\eta - \eta'\|$. The identified set $H(\cdot)$ is continuous at ϕ_0 in terms of the Hausdorff metric, i.e., for arbitrary $\epsilon > 0$, there exists an open neighborhood G of ϕ_0 such that $d_H(H(\phi), H(\phi_0)) < \epsilon$ holds for every $\phi \in G$.

(iii) Posterior of ϕ is consistent in the sense that for every open neighborhood G of ϕ_0, $\lim_{N\to\infty} F_{\phi|X^N}(N) = 1$ holds $\hat{p}(x^\infty|\phi_0)$-almost surely.

Condition 5.2 *(i) The parameter of interest η is a scalar and the identified set $H(\phi)$ is μ_ϕ-almost surely a nonempty and connected interval, $H(\phi) = [\eta_l(\phi), \eta_u(\phi)]$, $-\infty \leq \eta_l(\phi) \leq \eta_u(\phi) \leq \infty$. $H(\phi_0) = [\eta_l(\phi_0), \eta_u(\phi_0)]$ is a bounded interval.*

(ii) Define random variables $L_N(\phi) = \sqrt{N}\left(\eta_l(\phi) - \eta_l(\hat{\phi})\right)$ and $U_N(\phi) = \sqrt{N}\left(\eta_u(\phi) - \eta_u(\hat{\phi})\right)$ whose distribution is induced by the posterior distribution of ϕ. There exists bivariate random variables (L, U) distributed according to the bivariate normal with mean zero and variance-covariance Σ such that, for each W a subset in $\mathcal{R}^2$,

$$F_{\phi|X^N}((L_N(\phi), U_N(\phi)) \in W) \to \Pr((L, U) \in W)$$

in probability under $\hat{p}(x^{\infty}|\phi_0)$*.*

(iii) Let (L,U) *be the bivariate normal random variables as defined in (ii). The sampling distribution of* $\sqrt{N}\left(\eta_l(\phi_0)-\eta_l(\hat{\phi})\right)$ *and* $\sqrt{N}\left(\eta_u(\phi_0)-\eta_u(\hat{\phi})\right)$ *converges in distribution to the probability distribution of* (L,U)*.*

Condition 5.1 (i)-(iii) are used to establish consistency of the posterior lower probability and these conditions allow for multi-dimensional η. Condition 5.1 (ii) impose continuity of the set-valued mapping $H(\cdot)$ in terms of the Hausdorff metric, In case of an interval identified case, i.e., $H(\phi)=[\eta_l(\phi),\eta_u(\phi)]$, this condition is implied by continuity of $\eta_l(\phi)$ and $\eta_u(\phi)$ in ϕ. If $H(\phi_0)$ is a singleton, this continuity condition is equivalent to upper-hemicontinuity of $H(\cdot)$ at ϕ_0. Condition 5.1 (iii) states posterior consistency for ϕ in the form of Theorem 7.80 in Schervish (1995). While posterior consistency for ϕ requires the set of higher level conditions for the likelihood of ϕ, we omit to list it up here for the sake of brevity (see, e.g., Section 7.4 of Schervish (1995).)

Condition 5.2 (i)-(iii) are used to demonstrate that the volume minimizing posterior lower credible region has the correct coverage probability of $H(\phi_0)$. Condition 5.2 (i) assumes the parameter of interest η to be a scalar and its identified set is convex. Condition 5.2 (ii) and (iii) imply that the estimator for the lower and upper bounds of $H(\phi)$ satisfy the Bernstein-von Mises' type of asymptotic equivalence between Bayesian estimation and maximum likelihood estimation. These asymptotic normality conditions can be implied from more primitive assumptions: (i) regularity of the likelihood of ϕ, (ii) μ_ϕ puts a positive probability on every open neighborhood of ϕ_0 and μ_ϕ's density is smooth at ϕ_0, and (iii) applicability of the delta method to $\eta_l(\cdot)$ and $\eta_u(\cdot)$ given asymptotic normality of $\sqrt{N}\left(\phi-\hat{\phi}\right)$.[10]

Note that the asymptotic normality conditions of Condition 5.2 (ii) and (iii) preclude possible values of ϕ_0 where the sampling distribution of the point estimators of the lower and upper bounds is non-Gaussian. For instance, in the class of models where $\eta_l(\cdot)$ and $\eta_u(\cdot)$ involve the maximum or minimum operation such as in the intersection bound analysis (Manski (1990)), the asymptotic distribution of the maximum likelihood estimator for $\eta_l(\cdot)$ and $\eta_u(\cdot)$ is not normal when the arguments in these maximum or minimum happen to be equal at ϕ_0. See Chernozhukov, Lee, and Rosen (2009) for more examples and a general frequentist treatment in such situation.

The next proposition provides the large sample property of the posterior lower probability of η.

[10] For more rigorous statements of the former two of these three assumptions, see, e.g., Schervish (1995, Sec. 7.4).

Proposition 5.2 *(i) Assume Condition 3.1 and 5.1. Let* $d_H(H(\phi), H(\phi_0))$ *be the Hausdorff distance between* $H(\phi)$ *and* $H(\phi_0)$ *as defined in Condition 5.1 (ii). Then, for every* $\epsilon > 0$,

$$\lim_{N\to\infty} F_{\phi|X^N}(\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\}) = 0, \qquad \hat{p}(x^\infty|\phi_0)\text{-almost surely.}$$

(ii) Assume Condition 3.1 and 5.2. Let $C_{\alpha*}$ *be the volume minimizing posterior lower credible region as defined above.* $C_{\alpha*}$ *can be interpreted as the frequentist confidence interval for the true identified set* $H(\phi_0)$ *with the asymptotic coverage coverage probability* α, *i.e.,*

$$\lim_{N\to\infty} P_{X^N|\phi}(H(\phi_0) \subset C_{\alpha*}|\phi_0) \geq \alpha,$$

where $P_{X^N|\phi}(\cdot|\phi_0)$ *is the probability law of sample* X^N *when* $\phi = \phi_0$.

Proof. See Appendix A. ■

The first statement in the proposition establishes posterior consistency of the lower probability in the sense that the random sets $H(\phi)$ represented by the posterior lower probability converges to the deterministic set $H(\phi_0)$ in terms of the Hausdorff metric.

The second result shows that, for interval identified η, asymptotic equivalence holds between the posterior lower probability inference and the frequentist inference for the *identified set* considered in Horowitz and Manski (2000), Chernozhukov, Hong, and Tamer (2007), and Romano and Shaikh (2010). Note that the posterior lower credible region $C_{\alpha*}$ cannot be interpreted as the frequentist confidence interval for the *parameter of interest* considered in Imbens and Manski (2004). We conjecture that a multiple-prior Bayesian analogue of the frequentist confidence intervals for η with coverage α can be obtained by forming the union of the posterior credible regions with credibility α over each posterior generated from $\mathcal{M}(\mu_\phi)$. Bayesian interpretation of such union of the multiple credible regions, however, does not seem as lucid as that of the posterior lower credible region.

It is worth noting that the asymptotic coverage probability presented above is in the sense of pointwise asymptotics rather than the asymptotic uniform coverage probability over ϕ_0. The literatures has stressed importance of the uniform coverage property in order to ensure that the confidence intervals can have the correct coverage probability in the finite sample situation(Imbens and Manski (2004), Romano and Shaikh (2010), Stoye (2010), and Andrew and Soares (2010).). We do not know whether or not our posterior inference procedure attains uniformly valid coverage probability, and leave it for future research.

6 Conclusion

This paper proposes a framework of robust Bayes analysis for the set-identified models in econometrics. In order to obtain statistical inference and decision procedure that is insensitive to the empirically unrevisable prior knowledge, we introduce the class of prior distributions and propose the use of the posterior lower and upper probabilities for posterior inference. We demonstrate that the posterior lower and upper probabilities corresponding to prior class $\mathcal{M}\left(\mu_\phi\right)$ can be interpreted as the posterior probability law of the identified set (Theorem 3.1). This robust-Bayes way of generating the identified set as a random object is novel in the literature of the partially identified model, and it provides a seamless link between the random set theory and a likelihood based statistical inference in the set identified model.

We employ the gamma-minimax criterion to develop a conservative point decision for the set-identified parameter. The objective function of the gamma-minimax criterion integrates ambiguity associated with set-identification and posterior uncertainty for the identified set into the single objective function. We clarify it is feasible to calculate the gamma-minimax decision as long as we can simulate the identified sets $H(\phi)$ from the posterior of ϕ. Being different from the standard gamma-minimax decision rule, a non-standard finding with our specification of prior class is that the gamma-minimax optimal action is invariant no matter whether the optimal decision is computed before and after observing data (Proposition 5.2), which is potentially useful insight for extending the analysis to a sequential decision problem.

The posterior lower probability is a nonadditive measure so that one drawback of the lower probability analysis is that we cannot plot it as we do for the posterior probability densities. As a way to visualize it, we develop the way to compute the posterior lower credible region as an analogue to the posterior credible region in the standard Bayes procedure. We show that, for the interval identified parameter case, the posterior lower credible region with credibility α can be easily computed and can be interpreted as asymptotically valid frequentist confidence intervals for the identified set with coverage α.

In this article, we preclude observationally restrictive models and assume throughout that the identified set is nonempty. If we would like to incorporate possibility of obtaining empty identified set, we would expand the sufficient parameter space Φ to a larger one as long as the likelihood $\hat{p}(x|\phi)$ is well defined, and modify the definition of the lower and upper probabilities of θ to be the conditional lower and upper probabilities given $\Gamma(\phi) \neq \emptyset$. Research on applicability of the lower probability analysis to this context is in progress (Kitagawa (2011)).

Appendix

A Lemma and Proofs

In this appendix, we first demonstrate that set-valued mappings $\Gamma(\phi)$ and $H(\phi)$ defined in the main text are random sets (measurable set valued mappings) induced by a probability measure on $(\Phi, \mathcal{B})$.

Lemma A.1 *Assume* $(\Theta, \mathcal{A})$ *and* $(\Phi, \mathcal{B})$ *are complete separable metric spaces. Under Condition 3.1,* $\Gamma(\phi)$ *and* $H(\phi)$ *are random closed sets induced by a probability measure on* $(\Phi, \mathcal{B})$*, i.e.,* $\Gamma(\phi)$ *and* $H(\phi)$ *are closed and for* $A \in \mathcal{A}$ *and* $D \in \mathcal{H}$*,*

$$\begin{aligned} \{\phi : \Gamma(\phi) \cap A \neq \emptyset\} &\in \mathcal{B} \quad \textit{for} \quad A \in \mathcal{A}, \\ \{\phi : H(\phi) \cap D \neq \emptyset\} &\in \mathcal{B} \quad \textit{for} \quad D \in \mathcal{H}. \end{aligned}$$

Proof. Closedness of $\Gamma(\phi)$ and $H(\phi)$ is implied directly from Condition 3.1 (ii)-(iii). To prove measurability of $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}$, we use Theorem 2.6 in Molchanov, which states that, given $(\Theta, \mathcal{A})$ as Polish, $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\} \in \mathcal{B}$ holds if and only if $\{\phi : \theta \in \Gamma(\phi)\} \in \mathcal{B}$ is true for every $\theta \in \Theta$. Since $\Gamma(\phi)$ is an inverse image of the many-to-one and onto mapping, $g : \Theta \to \Phi$, a unique value of $\phi \in \Phi$ exists for each $\theta \in \Theta$, and $\{\phi\} \in \mathcal{B}$ since Φ is a metric space. Hence, $\{\phi : \theta \in \Gamma(\phi)\} \in \mathcal{B}$ holds.

To verify measurability of $\{\phi : H(\phi) \cap D \neq \emptyset\}$, we note

$$\{\phi : H(\phi) \cap D \neq \emptyset\} = \left\{\phi : \Gamma(\phi) \cap h^{-1}(D) \neq \emptyset\right\}.$$

Since $h^{-1}(D) \in \mathcal{A}$ by measurability of h (Condition 3.1 (iii)), the first statement of this lemma implies $\{\phi : H(\phi) \cap D \neq \emptyset\} \in \mathcal{B}$. ■

A.1 Proof of Theorem 3.1

Given measurability $\Gamma(\phi)$ and $H(\phi)$ as proved in Lemma A.1, our proof of Theorem 3.1 utilizes the following two lemma. The first lemma says that, for a fixed subset $A \in \mathcal{A}$ in the parameter space of θ and every $\mu_\theta \in \mathcal{M}(\mu_\phi)$, the conditional probability $\mu_{\theta|\phi}(A|\phi)$ can be bounded below by the indicator function $1_{\{\Gamma(\phi) \cap A \neq \emptyset\}}(\phi)$. The second lemma shows that for each fixed subset $A \in \mathcal{A}$, we can construct a probability measure on $(\Theta, \mathcal{A})$ that belongs to the prior class $\mathcal{M}(\mu_\phi)$ and achieves the lower bound of the conditional probability obtained in the first lemma. Theorem 4.1 follows as a corollary of these two lemma.

Lemma A.2 *Assume Condition 3.1 and let $A \in \mathcal{A}$ be an arbitrary fixed subset of Θ. For every $\mu_\theta \in \mathcal{M}(\mu_\phi)$,*

$$1_{\{\Gamma(\phi)\subset A\}}(\phi) \leq \mu_{\theta|\phi}(A|\phi)$$

holds μ_ϕ-almost surely.

Proof. For the given subset A, define $\Phi_1^A = \{\phi : \Gamma(\phi) \subset A\} \subset \Phi$. Note that, by Lemma A.1, Φ_1^A belongs to the sufficient parameter σ-algebra $\mathcal{B}$. To prove the claim, it suffices to show

$$\int_B 1_{\Phi_1^A}(\phi) d\mu_\phi \leq \int_B \mu_{\theta|\phi}(A|\phi) d\mu_\phi \tag{A.1}$$

for every $\mu_\theta \in \mathcal{M}(\mu_\phi)$ and $B \in \mathcal{B}$.

Consider

$$\int_B \mu_{\theta|\phi}(A|\phi) d\mu_\phi \geq \int_{B\cap\Phi_1^A} \mu_{\theta|\phi}(A|\phi) d\mu_\phi = \mu_\theta(A \cap \Gamma(B \cap \Phi_1^A)).$$

By the construction of Φ_1^A, $\Gamma(B \cap \Phi_1^A) \subset A$ holds, so

$$\begin{aligned} \mu_\theta(A \cap \Gamma(B \cap \Phi_1^A)) &= \mu_\theta(\Gamma(B \cap \Phi_1^A)) \\ &= \mu_\phi(B \cap \Phi_1^A) \\ &= \int_B 1_{\Phi_1^A}(\phi) d\mu_\phi. \end{aligned}$$

Thus, the inequality (A.1) is proven. ■

Lemma A.3 *Assume Condition 3.1. For each $A \in \mathcal{A}$, there exists $\mu_{\theta*} \in \mathcal{M}(\mu_\phi)$ whose conditional distribution $\mu_{\theta*|\phi}$ achieves the lower bound of $\mu_{\theta|\phi}(A|\phi)$ obtained in Lemma A.2, μ_ϕ-almost surely.*

Proof. Fix subset $A \in \mathcal{A}$ throughout the proof. Consider partitioning the sufficient parameter space Φ into three based on the relationship between $\Gamma(\phi)$ and A,

$$\begin{aligned} \Phi_0^A &= \{\phi : \Gamma(\phi) \cap A = \emptyset\}, \\ \Phi_1^A &= \{\phi : \Gamma(\phi) \subset A\}, \\ \Phi_2^A &= \{\phi : \Gamma(\phi) \cap A \neq \emptyset \text{ and } \Gamma(\phi) \cap A^c \neq \emptyset\}, \end{aligned}$$

where each of them belongs to the sufficient parameter σ-algebra $\mathcal{B}$ by Lemma A.1. Note that Φ_0^A, Φ_1^A, and Φ_2^A are mutually disjoint and constitute a partition of Φ.

Now, define a function $\xi^A(\cdot)$ that maps Φ_2^A to Θ such that $\xi^A(\phi) \in [\Gamma(\phi) \cap A^c]$ holds for μ_ϕ-almost every $\phi \in \Phi_2^A$. Note existence of such $\xi^A(\phi)$ is guaranteed by the construction of Φ_2^A

and Theorem 2.13 in Chap.1 of Molchanov (2005). Let us pick a probability measure from the prior class, $\mu_\theta \in \mathcal{M}\left(\mu_\phi\right)$, and construct another measure $\mu_{\theta*}$ by

$$\mu_{\theta*}\left(\tilde{A}\right) = \mu_\theta\left(\tilde{A} \cap \Gamma\left(\Phi_0^A\right)\right) + \mu_\theta\left(\tilde{A} \cap \Gamma\left(\Phi_1^A\right)\right) + \mu_\phi\left(\left\{\xi^A(\phi) \in \tilde{A}\right\} \cap \Phi_2^A\right), \qquad \tilde{A} \in \mathcal{A}.$$

It can be checked that $\mu_{\theta*}$ thus constructed is is a probability measure on $(\Theta, \mathcal{A})$, i.e., $\mu_{\theta*}$ satisfies $\mu_{\theta*}(\emptyset) = 0$, $\mu_{\theta*}(\Theta) = 1$, and countable additivity. Furthermore, $\mu_{\theta*}$ belongs to $\mathcal{M}\left(\mu_\phi\right)$ because for $B \in \mathcal{B}$,

$$\begin{aligned}
\mu_{\theta*}(\Gamma(B)) &= \mu_\theta\left(\Gamma(B) \cap \Gamma\left(\Phi_0^A\right)\right) + \mu_\theta\left(\Gamma(B) \cap \Gamma\left(\Phi_1^A\right)\right) + \mu_\phi\left(\left\{\xi^A(\phi) \in \Gamma(B)\right\} \cap \Phi_2^A\right) \\
&= \mu_\phi\left(B \cap \Phi_0^A\right) + \mu_\phi\left(B \cap \Phi_1^A\right) + \mu_\phi\left(B \cap \Phi_2^A\right) \\
&= \mu_\phi(B)
\end{aligned}$$

where the second line follows because $\xi(\phi) \in \Gamma(\phi)$ holds for almost every $\phi \in \Phi_2^A$ and $\Gamma(\phi)$'s are disjoint.

With thus constructed $\mu_{\theta*}$ and an arbitrary subset $B \in \mathcal{B}$, consider

$$\begin{aligned}
\mu_{\theta*}(A \cap \Gamma(B)) &= \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_0^A)) + \mu_\theta(A \cap \Gamma(B) \cap \Gamma(\Phi_1^A)) \\
&\quad + \mu_\phi\left(\left\{\xi^A(\phi) \in [A \cap \Gamma(B)]\right\} \cap \Phi_2^A\right).
\end{aligned}$$

Here, by the construction of $\left\{\Phi_j^A\right\}_{j=1,2,3}$ and $\xi^A(\phi)$, we have $A \cap \Gamma(\Phi_0^A) = \emptyset$, $\Gamma(\Phi_1^A) \subset A$, and $\mu_\phi\left(\left\{\xi^A(\phi) \in [A \cap \Gamma(B)]\right\} \cap \Phi_2^A\right) = 0$. Accordingly, we obtain,

$$\begin{aligned}
\mu_{\theta*}(A \cap \Gamma(B)) &= \mu_\theta(\Gamma(B) \cap \Gamma(\Phi_1^A)) \\
&= \mu_\phi\left(B \cap \Phi_1^A\right) \\
&= \int_B 1_{\Phi_1^A}(\phi) d\mu_\phi.
\end{aligned}$$

Since $B \in \mathcal{B}$ is arbitrary, this implies that $\mu_{\theta*}(A|\phi) = 1_{\Phi_1^A}(\phi)$, μ_ϕ-almost surely. Thus, $\mu_{\theta*}$ achieves the lower bound obtained in Lemma A.2. ■

Proof of Theorem 3.1 (i). Under the given assumptions, the posterior of θ is given by (see equation (3.1))

$$F_{\theta|X}(A) = \int_\Phi \mu_{\theta|\phi}(A|\phi) f_{\phi|X}(\phi|x) d\phi.$$

Since $f_{\phi|X}(\phi|x) \geq 0$ almost surely and monotonicity of the integral, $F_{\theta|X}(A)$ is minimized by plugging the lower bound bound of $\mu_{\theta|\phi}(A|\phi)$ into the integrand. From Lemma A.2 and Lemma A.3, it is given by $1_{\{\Gamma(\phi) \subset A\}}(\phi)$, so

$$F_{\theta|X*}(A) = \int_\Phi 1_{\{\Gamma(\phi) \subset A\}}(\phi) f_{\phi|X}(\phi|x) d\phi = F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A\}).$$

The posterior upper probability is obtained by its conjugacy with the lower probability,

$$F^{*}_{\theta|X}(A) = 1 - F_{\theta|X*}(A^c) = F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}).$$

■

A.2 Proof of Proposition 4.1

The next lemma is used to prove Proposition 4.1. It says that the class of posteriors induced by prior class $\mathcal{M}(\mu_\phi)$ exhausts all the probability measures lying uniformly between the posterior lower and upper probabilities. In the terminology of Huber (1973), this property is called representability of the class of probability measures by the lower and upper probabilities. Wasserman and Kadane (1990) calls it as that the posterior class is closed with respect to majorization.

Lemma A.4 *Assume Condition 3.1. Let*

$$\mathcal{G}_\theta = \left\{ G_\theta : G_\theta \text{ probability measure on } (\Theta, \mathcal{A}), \, F_{\theta|X*}(A) \leq G_\theta(A) \leq F^{*}_{\theta|X}(A) \text{ for every } A \in \mathcal{A} \right\}.$$

. *Then,* $\mathcal{G}_\theta = \left\{ F_{\theta|X} : F_{\theta|X} \text{ posterior distribution on } (\Theta, \mathcal{A}) \text{ induced by some } \mu_\theta \in \mathcal{M}(\mu_\phi) \right\}$.

Proof of Lemma A.4. For each $\mu_\theta \in \mathcal{M}(\mu_\phi)$, $F_{\theta|X*}(A) \leq F_{\theta|X}(A) \leq F^{*}_{\theta|X}(A)$ holds for every $A \in \mathcal{A}$ by the definition of the lower and upper probabilities. Hence,

$$\mathcal{G}_\theta = \left\{ G_\theta : F_{\theta|X*}(A) \leq G_\theta(A) \leq F^{*}_{\theta|X}(A) \text{ for every } A \in \mathcal{A} \right\}$$

contains $\left\{ F_{\theta|X} : \mu_\theta \in \mathcal{M}(\mu_\phi) \right\}$.

To show the converse, recall Theorem 3.1 (i) that shows that the lower and upper probabilities are containment and capacity functional of the random closed set $\Gamma(\phi)$. As a result, by applying the Selectionability Theorem of the random set (Molchanov (2005), Theorem 1.2.20), it holds that for each $G_\theta \in \mathcal{G}_\theta$, there exists a Θ-valued random variable $\xi(\phi)$, so called a selection of $\Gamma(\phi)$, such that $\xi(\phi) \in \Gamma(\phi)$ holds for every $\phi \in \Phi$ and $G_\theta(A) = F_{\phi|X}(\xi(\phi) \in A)$, $A \in \mathcal{A}$.

Let $G_\theta \in \mathcal{G}_\theta$ be a fixed arbitrary distribution in $\mathcal{G}_\theta$. Let $\xi(\phi)$ be the associated selection of $\Gamma(\phi)$ and let μ^{ξ}_θ be the probability distribution of such $\xi(\phi)$ induced by the prior of ϕ,

$$\mu^{\xi}_\theta(A) = \mu_\phi \left(\{\phi : \xi(\phi) \in A\} \right).$$

Note such μ^{ξ}_θ belongs to$\mathcal{M}(\mu_\phi)$ since for each subset $B \in \mathcal{B}$ in the sufficient parameter space,

$$\begin{aligned} \mu^{\xi}_\theta(\Gamma(B)) &= \mu_\phi \left(\{\phi : \xi(\phi) \in \Gamma(B)\} \right) \\ &= \mu_\phi(B) \end{aligned}$$

where the second equality holds because $\{\Gamma(\phi) : \phi \in B\}$ are mutually disjoint and $\xi(\phi) \in \Gamma(\phi)$ for every ϕ.

Since the conditional distribution for $\mu_\theta^\xi(A)$ given ϕ, $A \in \mathcal{A}$, is $\mu_{\theta|\phi}^\xi(A|\phi) = 1_{\{\xi(\phi)\in A\}}(\phi)$, the posterior distribution of θ generated from μ_θ^ξ is, by (3.1),

$$\begin{aligned}\tilde{F}_{\theta|X}(A) &= \int 1_{\{\xi(\phi)\in A\}}(\phi) f_{\phi|X}(\phi|x)d\phi \\ &= F_{\phi|X}(\xi(\phi) \in A) \\ &= G_\theta(A).\end{aligned}$$

Thus, we have shown that, for each $G_\theta \in \mathcal{G}_\theta$, there exists a prior $\mu_\theta^\xi \in \mathcal{M}(\mu_\phi)$ with which the posterior of θ coincides with the G_θ. Hence, $\mathcal{G}_\theta \subset \left\{F_{\theta|X} : \mu_\theta \in \mathcal{M}(\mu_\phi)\right\}$. ■

Proof of Proposition 4.1. Let $\mathcal{G}_\theta$ be the class of probability measures on $(\Theta, \mathcal{A})$ as defined in Lemma A.4. Graf (1980, Proposition 2.3) showed that if $F_{\theta|X}^*(\cdot)$ is a subadditive alternating capacity of order two and it is regular (Condition 4.1 (ii)), then for any nonnegative measurable function $k : \Theta \to \mathcal{R}_+$

$$\int k(\theta)dF_{\theta|X}^* = \sup_{G_\theta\in\mathcal{G}_\theta}\left\{\int_\Theta k(\theta)dG_\theta\right\} \tag{A.2}$$

holds. Since $F_{\theta|X}^*(\cdot)$ is the capacity functional of the random closed set $\Gamma(\phi)$, the Choquet Theorem (see, e.g. ,Molchanov (2005, Sec. 1.1.2-1.1.3)) ensures that $F_{\theta|X}^*(\cdot)$ is a subadditive capacity of infinite alternating order. Furthermore, Lemma A.4 implies that $\sup_{G_\theta\in\mathcal{G}_\theta}\left\{\int_\Theta k(\theta)dG_\theta\right\}$ is equivalent to $\sup_{\mu_\theta\in\mathcal{M}(\mu_\phi)}\left\{\int_\Theta k(\theta)dF_{\theta|X}\right\}$. Hence, setting $k(\theta) = L(h(\theta), a)$ in (A.2) leads to

$$\int L(h(\theta), a)dF_{\theta|X}^* = \sup_{\mu_\theta\in\mathcal{M}(\mu_\phi)}\left\{\int_\Theta L(h(\theta), a)dF_{\theta|X}\right\} = \rho(\mu_\theta, a). \tag{A.3}$$

On the other hand, by the definition of Choquet integral and Theorem 4.1 (ii),

$$\begin{aligned}\int L(\eta, a)dF_{\eta|X}^* &= \int F_{\eta|X}^*(\{\eta : L(\eta, a) \geq t\})dt \\ &= \int F_{\theta|X}^*\left(\{\theta : L(h(\theta), a)\}\right)dt \\ &= \int L(h(\theta), a)dF_{\theta|X}^*.\end{aligned} \tag{A.4}$$

Combining (A.3) and (A.4) yields the first equality of the proposition.

Next, we show the second equality of the proposition. By Theorem 4.1 (ii),

$$\begin{aligned}\int L(\eta, a)dF_{\eta|X}^*(\eta) &= \int_0^\infty F_{\eta|X}^*\left(\{\eta : L(\eta, a) \geq t\}\right)dt \\ &= \int_0^\infty F_{\phi|X}\left(\{\phi : \{\eta : L(\eta, a) \geq t\} \cap H(\phi) \neq \emptyset\}\right)dt.\end{aligned}$$

Note that $\{\eta : L(\eta, a) \geq t\} \cap H(\phi) \neq \emptyset$ is true if and only if $\left\{\sup_{\eta \in H(\phi)} \{L(\eta, a)\} \geq t\right\}$, so we obtain

$$\int L(\eta, a) dF^*_{\eta|X}(\eta) = \int_0^\infty F_{\phi|X}\left(\left\{\phi : \sup_{\eta \in H(\phi)} \{L(\eta, a)\} \geq t\right\}\right) dt.$$

By interchanging the order of integrations by Tonelli's theorem, we obtain

$$\begin{aligned}
\int L(\eta, a) dF^*_{\eta}(\eta|x) &= \int_0^\infty \int_\Phi 1_{\left\{\sup_{\eta \in H(\phi)}\{L(\eta,a)\} \geq t\right\}}(\phi) dF_{\phi|X} dt \\
&= \int_\Phi \int_0^\infty 1_{\left\{\sup_{\eta \in H(\phi)}\{L(\eta,a)\} \geq t\right\}}(t) dt dF_{\phi|X} \\
&= \int_\Phi \sup_{\eta \in H(\phi)} \{L(\eta, a)\} dF_{\phi|X}. \qquad \text{(A.5)}
\end{aligned}$$

■

A.3 Proof of Proposition 5.1 and 5.2

Proof of Proposition 5.1. Let $\eta_c \in \mathcal{H}$ be fixed and $B_r(\eta_c)$ be a closed ball centered at η_c with radius r. The event $\{H(\phi) \subset B_r(\eta_c)\}$ happens if and only if $\left\{\overrightarrow{d}(\eta_c, H(\phi)) \leq r\right\}$. So, $r_\alpha(\eta_c) \equiv \inf\left\{r : F_{\phi|X}\left(\left\{\phi : \overrightarrow{d}(\eta_c, H(\phi)) \leq r\right\}\right) \geq \alpha\right\}$ is the radius of the smallest closed ball centered at η_c that contains random sets $H(\phi)$ with posterior probability at least α. Therefore, finding the minimizer of $r_\alpha(\eta_c)$ over η_c is equivalent to searching for the center of the smallest ball that contains $H(\phi)$ with posterior probability α, and the attained minimum of $r_\alpha(\eta_c)$ gives the radius of the smallest ball. ■

Proof of Proposition 5.2 (i). Let $\epsilon > 0$ be arbitrary. By Condition 5.1 (ii), there exists an open neighborhood G of ϕ_0 such that $d_H(H(\phi), H(\phi_0)) < \epsilon$ holds for every $\phi \in N$. Consider

$$\begin{aligned}
F_{\phi|X^N}(\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\}) &= F_{\phi|X^N}(\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\} \cap G) \\
&\quad + F_{\phi|X^N}(\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\} \cap G^c) \\
&\leq F_{\phi|X^N}(G^c)
\end{aligned}$$

where the last line follows because $\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\} \cap G = \emptyset$ by the construction of G. Posterior consistency of Condition 5.1 (iii) yields $\lim_{N\to\infty} F_{\phi|X^N}(G^c) = 0$, $p(x^\infty|\phi_0)$-a.s., so $\lim_{N\to\infty} F_{\phi|X^N}(\{\phi : d_H(H(\phi), H(\phi_0)) > \epsilon\}) = 0$ holds $p(x^\infty|\phi_0)$-a.s. ■

Proof of Proposition 5.2 (ii). By denoting a connected interval as $[l, u]$, we write the optimization problem for obtaining $C_{\alpha *}$ as

$$\begin{aligned}
&\min_{l \leq u} [u - l] \\
\text{s.t.} \quad &F_{\phi|X^N}(l \leq \eta_l(\phi) \ \text{ and } \ \eta_u(\phi) \leq u) \geq \alpha.
\end{aligned}$$

In terms of the random variables $L_N(\phi) = \sqrt{N}\left(\eta_l(\phi) - \eta_l(\hat{\phi})\right)$ and $U_N(\phi) = \sqrt{N}\left(\eta_u(\phi) - \eta_u(\hat{\phi})\right)$, the above constraint becomes

$$F_{\phi|X^N}\left(\sqrt{N}\left(l - \eta_l(\hat{\phi})\right) \leq L_N(\phi) \quad \text{and} \quad U_N(\phi) \leq \sqrt{N}u - \eta_u(\hat{\phi})\right) \geq \alpha.$$

Therefore, by denoting $c_l = -\sqrt{N}\left(l - \eta_l(\hat{\phi})\right)$ and $c_u = \sqrt{N}\left(u - \eta_u(\hat{\phi})\right)$, we can formulate the optimization problem as

$$\begin{aligned} & \min\left[c_l + c_u\right] \\ \text{s.t.} \quad & F_{\phi|X^N}\left(-c_l \leq L_N(\phi) \quad \text{and} \quad U_N(\phi) \leq c_u\right) \geq \alpha, \end{aligned}$$

and denote its solution by $(\hat{c}_l, \hat{c}_u)$. The expression of the posterior lower credible region is obtained as $C_{\alpha *} = \left[\eta_l(\hat{\phi}) - \frac{\hat{c}_l}{\sqrt{N}}, \eta_u(\hat{\phi}) + \frac{\hat{c}_u}{\sqrt{N}}\right]$. Under Condition 5.2 (ii), $(L_N(\phi), U_N(\phi))$ converges in distribution to bivariate normal for almost every X^∞, so the fact that their limiting distribution has the probability density implies that, for almost every X^∞, $(\hat{c}_l, \hat{c}_u)$ converges to

$$\begin{aligned} (c_l^*, c_u^*) \quad \equiv \quad & \arg\min\left[c_l + c_u\right] \\ \text{s.t.} \quad & \Pr\left(-c_l \leq L \quad \text{and} \quad U \leq c_u\right) \geq \alpha. \end{aligned}$$

With thus constructed $C_{\alpha *}$, consider the coverage probability for the true identified set

$$\begin{aligned} & P_{X^N|\phi}\left(H(\phi_0) \subset C_{\alpha *}|\phi_0\right) \\ = \; & P_{X^N|\phi}\left(\eta_l\left(\hat{\phi}\right) - \frac{\hat{c}_l}{\sqrt{N}} \leq \eta_l(\phi_0) \quad \text{and} \quad \eta_u(\phi_0) \leq \eta_u(\hat{\phi}) + \frac{\hat{c}_u}{\sqrt{N}} \,\middle|\, \phi_0\right) \\ = \; & P_{X^N|\phi}\left(-\hat{c}_l \leq \sqrt{N}\left(\eta_l(\phi_0) - \eta_l(\hat{\phi})\right) \quad \text{and} \quad \sqrt{N}\left(\eta_u(\phi_0) - \eta_u(\hat{\phi})\right) \leq \hat{c}_u \,\middle|\, \phi_0\right). \end{aligned}$$

By Condition 5.2 (iii), the sampling distribution of $\sqrt{N}\left(\eta_l(\phi_0) - \eta_l(\hat{\phi})\right)$ and $\sqrt{N}\left(\eta_u(\phi_0) - \eta_u(\hat{\phi})\right)$ converges in distribution to (L, U), and, as we demonstrated above, $(\hat{c}_l, \hat{c}_u)$ converges to (c_l^*, c_u^*) for almost every sampling sequence. Therefore,

$$\begin{aligned} & \lim_{N\to\infty} P_{X^N|\phi}\left(H(\phi_0) \subset C_{\alpha *}|\phi_0\right) \\ = \; & \lim_{N\to\infty} P_{X^N|\phi}\left(-\hat{c}_l \leq \sqrt{N}\left(\eta_l(\phi_0) - \eta_l(\hat{\phi})\right) \quad \text{and} \quad \sqrt{N}\left(\eta_u(\phi_0) - \eta_u(\hat{\phi})\right) \leq \hat{c}_u \,\middle|\, \phi_0\right) \\ = \; & \Pr\left(-c_l^* \leq L \quad \text{and} \quad U \leq c_u^*\right) \geq \alpha \end{aligned}$$

holds. ■

B Gamma Minimax Regret Decision

In this appendix, we consider the gamma-minimax *regret* criterion as an alternative to the (posterior) gamma-minimax criterion considered in Section 4. In order to keep analytical tractability, we consider the case where the parameter of interest η is a scalar and the loss function is quadratic, $L(\eta, a) = (\eta - a)^2$.

The statistical decision under the conditional and unconditional gamma-minimax regret criterion are set up as follows.

Definition B.1 *Define the lower bound of the posterior risk given μ_θ by $\underline{\rho}(\mu_\theta) = \inf_{a \in \mathcal{H}} \rho(\mu_\theta, a)$. The posterior gamma-minimax regret action $a_x^{reg} \in \mathcal{H}$ solves,*

$$\inf_{a \in \mathcal{H}_a} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \left\{ \rho(\mu_\theta, a) - \underline{\rho}(\mu_\theta) \right\}.$$

Definition B.2 *Define the lower bound of the Bayes risk given μ_θ by $\underline{r}(\mu_\theta) = \inf_{\delta \in \Delta} r(\mu_\theta, \delta)$. The gamma-minimax regret decision $\delta^{reg} : \mathbf{X} \to \mathcal{H}_a$ solves*

$$\inf_{\delta \in \Delta} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \left\{ r(\mu_\theta, \delta) - \underline{r}(\mu_\theta) \right\}.$$

Since the loss function is quadratic, the posterior risk $\rho(\mu_\theta, a)$ for given μ_θ is minimized at $\hat{\eta}_{\mu_\theta}$ the posterior mean of η if it exists. Therefore, the lower bound of the posterior risk is simply the posterior variance, $\underline{\rho}(\mu_\theta) = E_{\eta|X}(\left(\eta - \hat{\eta}_{\mu_\theta}\right)^2)$, and the posterior regret can be written as

$$\begin{aligned} \rho(\mu_\theta, a) - \underline{\rho}(\mu_\theta) &= E_{\eta|X}\left[(\eta - a)^2 - \left(\eta - \hat{\eta}_{\mu_\theta}\right)^2 \right] \\ &= E_{\eta|X}\left[(a - \hat{\eta}_{\mu_\theta})^2 \right]. \end{aligned}$$

Let $\left[\underline{\eta}_x, \bar{\eta}_x\right]$ be the range of posterior mean of η when μ_θ varies over the prior class $M(\mu_\phi)$, which we assume is bounded, which can be implied if $H(\phi)$ is bounded, μ_ϕ-almost surely. Then, the posterior gamma-minimax regret is simplified to

$$\sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \left\{ \rho(\mu_\theta, a) - \underline{\rho}(\mu_\theta) \right\} = \begin{cases} (\bar{\eta}_x - a)^2 & \text{for } a \leq \frac{\underline{\eta}_x + \bar{\eta}_x}{2}, \\ \left(\underline{\eta}_x - a\right)^2 & \text{for } a > \frac{\underline{\eta}_x + \bar{\eta}_x}{2}, \end{cases} \tag{B.1}$$

and, therefore, the posterior gamma-minimax regret is minimized at $a = \frac{\underline{\eta}_x + \bar{\eta}_x}{2}$. That is, the posterior gamma-minimax regret action is simply obtained as the mid point of $\left[\underline{\eta}_x, \bar{\eta}_x\right]$, which is qualitatively similar to the local asymptotic minimax regret decision analyzed in Song (2009).

Proposition B.1 *Let $\mathcal{H} \subset \mathcal{R}$ and $L(\eta, a) = (\eta - a)^2$. Assume that the posterior variance of η is finite for every $\mu_\theta \in \mathcal{M}(\mu_\phi)$. Let $\underline{\eta}(\phi) \equiv \inf\{\eta : \eta \in H(\phi)\}$ and $\bar{\eta}(\phi) \equiv \sup\{\eta : \eta \in H(\phi)\}$.*
(i) The posterior gamma-minimax regret action is

$$a_x^{reg} = \frac{E_{\phi|X}(\underline{\eta}(\phi)) + E_{\phi|X}(\bar{\eta}(\phi))}{2}.$$

(ii) The unconditional gamma-minimax regret decision with the quadratic loss satisfies $\delta^{reg}(x) = a_x^{reg}$, $m(x|\mu_\phi)$*-almost surely.*

Proof. (i) Given that the posterior variance of η is finite for $\mu_\theta \in \mathcal{M}(\mu_\phi)$, the posterior gamma-minimax regret is well-defined and it is given by

$$\rho(\mu_\theta, a) - \underline{\rho}(\mu_\theta) = E_{\eta|X}\left[(a - \hat{\eta}_{\mu_\theta})^2\right]$$

where $\hat{\eta}_{\mu_\theta}$ is the posterior mean of η when prior of θ is μ_θ. Consider the bounds of $\hat{\eta}_{\mu_\theta}$ when μ_θ varies over $\mathcal{M}(\mu_\phi)$,

$$\begin{aligned} \underline{\eta}_x &\equiv \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \hat{\eta}_{\mu_\theta} = \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \int_\Theta h(\theta) dF_{\theta|X}, \\ \bar{\eta}_x &\equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \hat{\eta}_{\mu_\theta} = \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \int_\Theta h(\theta) dF_{\theta|X}. \end{aligned}$$

By the same argument as used in obtaining (A.3), (A.4), and (A.5), $\underline{\eta}_x$ and $\bar{\eta}_x$ thus defined are equal to

$$\begin{aligned} \underline{\eta}_x &= E_{\phi|X}\left[\inf\{\eta : \eta \in H(\phi)\}\right], \\ \bar{\eta}_x &= E_{\phi|X}\left[\sup\{\eta : \eta \in H(\phi)\}\right], \end{aligned}$$

which are assumed to be finite by the finite posterior variance assumption. Then, as already discussed in (B.1), the posterior gamma-minimax regret action is obtained as $\frac{\underline{\eta}_x + \bar{\eta}_x}{2}$.

(ii) Note that the lower bound of the Bayes risk when $\mu_\theta \in \mathcal{M}(\mu_\phi)$ is written as the average of the posterior variance of η with respect to the marginal distribution of data,

$$\begin{aligned} \underline{r}(\mu_\theta) &= \inf_{\delta \in \Delta} \int_{\mathbf{X}} \rho(\mu_\theta, \delta(x)) m(x|\mu_\phi) dx \\ &= \int_{\mathbf{X}} \underline{\rho}(\mu_\theta) m(x|\mu_\phi) dx. \end{aligned}$$

Therefore, the unconditional regret is written as

$$\begin{aligned} r(\mu_\theta, \delta) - \underline{r}(\mu_\theta) &= \int_{\mathbf{X}} \rho(\mu_\theta, \delta(x)) m(x|\mu_\phi) dx - \int_{\mathbf{X}} \underline{\rho}(\mu_\theta) m(x|\mu_\phi) dx \\ &= \int_{\mathbf{X}} \left[\rho(\mu_\theta, \delta(x)) - \underline{\rho}(\mu_\theta)\right] m(x|\mu_\phi) dx. \end{aligned}$$

Since the marginal distribution of data does not depend on μ_θ once μ_ϕ is fixed, the unconditional gamma-minimax regret can be written as the average of the posterior gamma minimax regret with respect to $m(x|\mu_\phi)$,

$$\sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \{r(\mu_\theta, \delta) - \underline{r}(\mu_\theta)\} = \int_{\mathbf{X}} \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \left\{\rho(\mu_\theta, a) - \underline{\rho}(\mu_\theta)\right\} m(x|\mu_\phi) dx.$$

This implies that the optimal gamma-minimax regret decision $\delta^{reg}(x)$ coincides with the posterior gamma-minimax regret action a_x^{reg}, $m(x|\mu_\phi)$-almost surely. ■

Since the lower bound of the posterior risk $\underline{\rho}(\mu_\theta)$ in general depends on prior μ_θ, the posterior gamma-minimax regret action a_x^{reg} differs from the posterior gamma-minimax action a_x^* obtained in Section 4.1. To illustrate the difference, consider the posterior gamma-minimax action of Proposition 4.1 when η is a scalar and the loss is quadratic. Let $[\underline{\eta}(\phi), \bar{\eta}(\phi)]$ be as defined in Proposition B.1, and let $m(\phi) = (\underline{\eta}(\phi) + \bar{\eta}(\phi))/2$ and $r(\phi) = (\bar{\eta}(\phi) - \underline{\eta}(\phi))/2 \geq 0$ be the midpoint and the radius of the smallest interval that contains $H(\phi)$. The objective function to be minimized in the posterior gamma-minimax decision problem can be written as

$$
\begin{aligned}
E_{\phi|X}\left[\sup_{\eta \in H(\phi)} (\eta - a)^2\right] &= E_{\phi|X}\left[\max\left\{(\underline{\eta}(\phi) - a)^2, (\bar{\eta}(\phi) - a)^2\right\}\right] \\
&= E_{\phi|X}\left[(m(\phi) - a)^2\right] + E_{\phi|X}\left[r(\phi)\left|m(\phi) - a\right|\right] + E_{\phi|X}\left[\left(\frac{r(\phi)}{2}\right)^2\right].
\end{aligned}
$$

Note that a_x^{reg} minimizes the first term, but it does not necessarily minimize the second term, and, therefore, a_x^{reg} can in general differ from the gamma minimax action. The gamma-minimax regret decision with the quadratic loss depends only on the distribution of $m(\phi)$, while the gamma-minimax decision depends on the joint distribution of $m(\phi)$ and $r(\phi)$. For large sample, this difference disappears and a_x^{reg} and a_x^* converge to the same action. In general, when the loss function is specified to be a monotonically increasing function of a metric $\|\eta - a\|$, a_x^{reg} and a_x^* converges to the center of the smallest circle that contains $H(\phi_0)$.

References

[1] Andrews, D.W.K. and G.Soares (2010) "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-157.

[2] Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996) "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association,* 91: 444 - 472.

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

[4] Barankin, E.W. (1960) "Sufficient Parameters: Solution of the Minimal Dimensionality Problem," *Annals of the Institute of Mathematical Statistics*, 12, 91-118.

[5] Beresteanu, A. and F. Molinari (2008) "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763-814.

[6] Beresteanu, A., I.S. Molchanov, and F. Molinari (2010) "Partial ldentification Using Random Sets Theory," unpublished manuscript.

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

[8] Berger, J.O. and L.M. Berliner (1986) "Robust Bayes and Empirical Bayes Analysis with ϵ-contaminated Priors." *Annals of Statistics*, 14, 461-486.

[9] Bernardo, J.M. (1979) "Reference Posterior Distributions for Bayesian Inference." *Journal of the Royal Statistical Society, Series B*, 41, No.1, 113-147.

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

[11] Bollinger, C.R. and M. van Hasselt (2009) "A Bayesian Analysis of Binary Misclassification: Inference in Partially Identified Models." unpublished manuscript, The University of Kentucky and The University of Western Ontario.

[12] Chamberlain, G. and E.E. Leamer (1976) "Matrix Weighted Averages and Posterior Bounds" *Journal of the Royal Statistical Society, Series B*, 38, 73-84.

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

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

[15] DeRoberts, L. and J.A. Hartigan (1981) "Bayesian Inference Using Intervals of Measures," *The Annals of Statistics*, 9, No.2, 235-244.

[16] Dempster, A.P. (1966) "New Methods for Reasoning Towards Posterior Distributions Based on Sample Data." *Annals of Mathematical Statistics* 37, 355-374.

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

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

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

[20] Drèze, J.H. (1974) "Bayesian Theory of identification in Simultaneous Equations Models." 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] Fan Y. and S. Park (2010) "Sharp Bounds on the Distribution of the Treatment Effects and Their Statistical Inference," *Econometric Theory*, 26, 931-951.

[22] Firpo, S. and G. Ridder (2009) "Bounds on Functionals of the Distribution of Treatment Effects," IEPR Working Paper 08.09.

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

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

[25] Frisch, R. (1934) *Statistical Confluence Analysis by Means of Complete Regression Systems*. University Institute for Economics, Oslo, Norway.

[26] Galichon, A. and M. Henry (2006) "Inference in Incomplete Models." unpublished manuscript, Ecole Polytechnique and Universite de Montreal.

[27] Good, I.J. (1965) *The Estimation of Probabilities*. MIT Press, Cambridge, Massachusetts.

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

[29] Gustafson, P. (2009) "What Are the Limits of Posterior Distributions Arising from Nonidentified Models, and Why Should We Care?" *Journal of the American Statistical Association*, Vol. 104, No. 488, 1682-1695.

[30] Gustafson, P. (2010) "Bayesian Inference for Partially Identified Models." *The International Journal of Biostatistics,* Vol. 6, Issue 2.

[31] Heckman, J. J., J. Smith, and N. Clements (1997) "Making the Most out of Programme Evaluation and Social Experiments Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

[32] Horowitz, J., and C.F. Manski (2000) "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

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

[34] Hurwicz, L. (1950) "Generalization of the Concept of Identification" *Statistical Inference in Dynamic Economic Models, Cowles Commision Monograph 10.* John WIley and Sons, New York.

[35] Imbens, G. W. and J. D. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

[36] Imbens, G. W. and C.F. Manski (2004) "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

[37] Jeffreys, H. (1961) *Theory of Probability, Third edition.* Oxford University Press, Oxford, UK.

[38] 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.

[39] Kass, R.E. and L. Wasserman (1996) "The Selection of Prior Distributions by Formal Rules." *Journal of the American Statistical Association*, Vol. 92, No. 435, 1343-1369.

[40] Kitagawa, T. (2009) "Identification Region of the Potential Outcome Distributions under Instrument Independence." *Cemmap Working paper,* University College London.

[41] Kitagawa, T. (2011) "Robust Bayes Inference and Intersection Bounds: Pessimistic Inference with a Class of Optimistic Priors," *unpublished manuscript.* University College London.

[42] Klepper, S. and E.E. Leamer (1984) "Consistent Sets of Estimates for Regressions with Errors in All Variables." *Econometrica*, Vol.52, No.1. 163-183.

[43] Kudō, H. (1967) "On Partial Prior Information and the Property of Parametric Sufficiency." in L. Le Cam and J. Neyman eds. *Proceedings the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 251-265.

[44] Koopmans and Riersol (1950) "The Identification of Structural Characteristics," *Annals of Mathematical Statistics* 21, 165-181.

[45] Lambert, D. and G.T. Duncan (1986) "Single-parameter Inference Based on Partial Prior Information" *The Canadian Journal of Statistics*, Vol. 14, No. 4, 297-305.

[46] Leamer, E.E. (1982) "Sets of Posterior Means with Bounded Variance Priors" *Econometrica* 50, 725-736.

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

[48] Liao, Y. and W. Jiang (2010) "Bayesian Analysis in Moment Inequality Models." *The Annals of Statistics,* 38, 1, 275-316.

[49] Makarov, G. (1981) "Estimates for the Distribution Function of a Sum of Two Random Variables when the Marginal Distributions are Fixed," *Theory of Probability and its Applications*, 26, 803-806.

[50] Manski, C.F. (1981) "Learning and Decision Making When Subjective Probabilities Have Subjective Domains," *The Annals of Statistics*, Vol.9, No.1, 59-65.

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

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

[53] Manski, C.F. (2000) "Identification Problems and Decision under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice," *Journal of Econometrics* 95, 415-442.

[54] Manski, C.F. (2002) "Treatment Choice under Ambiguity Induced by Inferential Problems," *Journal of Statistical Planning and Inference* 105, 67-82.

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

[56] Manski, C.F. (2008) *Identification for Prediction and Decision*, Harvard University Press, Cambridge, Massachusetts.

[57] Manski, C.F. (2009) "Diversified Treatment under Ambiguity," *International Economic Review* 50, 1013-1041.

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

[59] Moon, H.R. and F. Schorfheide (2010) "Bayesian and Frequentist Inference in Partially Identified Models." manuscript. University of Sourthern California and University of Pennsylvania.

[60] Morris, C. (1983) "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, Vol. 78, 47-65.

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

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

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

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

[65] Robbins, H. (1964) "The empirical Bayes Approach to Statistical Decision Problems." *Annals of Mathematical Statistics* 35, 1-20.

[66] Romano, J.P. and A.M. Shaikh (2010) "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-211.

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

[68] Savage, L. (1951) "The Theory of Statistical Decision." *Journal of the American Statistical Association*, Vol. 46, 55-67.

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

[70] Shafer, G. (1982) "Belief Functions and Parametric Models." *Journal of the Royal Statistical Society, Series B*, 44, No.3, 322-352.

[71] Schennach, S.M. (2005) "Bayesian Exponentially Tilted Empirical Likelihood." *Biometrika*, 92, 1, 31-46.

[72] Schervish, M.J. (1995) *Theory of Statistics*. Springer-Verlag, New York.

[73] Song, K. (2009) "Point Decisions for Interval Identified Parameters," unpublished manuscript. University of Pennsylvania.

[74] Stoye, J. (2010) "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299-1315.

[75] 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.

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

[77] Wasserman, L.A. (1989) "A Robust Bayesian Interpretation of Likelihood Regions," *The Annals of Statistics*, 17, 3, 1387-1393.

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

[79] Wasserman, L.A. and J.B. Kadane (1990) "Bayes Theorem for Choquet Capacities," *The Annals of Statistics*, 18, 3, 1328-1339.