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

Toru Kitagawa*
CeMMAP and Department of Economics, UCL

First Draft, July, 2010
This Draft, August, 2010

Abstract

This paper develops inference and statistical decision for set-identified parameters from the robust Bayes perspective. When the model is not identified, prior knowledge for the parameters is decomposed into the two components: the one that can be updated by data (revisable prior knowledge) and the one that never be updated (unrevisable prior knowledge). We accept a single prior distribution for the revisable and apply the Bayesian updating, while we introduce the unconstrained class of prior distributions for the unrevisable, and develop an inference procedure that is free from the unrevisable prior knowledge. We summarize the posterior uncertainty for the set-identified parameters by the posterior lower and upper probabilities (Dempster (1967, 1968)). We develop point estimation of the set-identified parameters based on the posterior gamma-minimax criterion, which is, in our context, equivalent to minimizing the Choquet expected loss with respect to the posterior upper probability. We also propose a use of the posterior lower probability to construct a posterior credible region for the set-identified parameters. Our framework offers a procedure to eliminate set-identified nuisance parameters, and yields an inference for the marginalized identified set. A simple numerical algorithm is provided for implementing the procedure, and the large sample property of the procedure is examined and compared with the existing frequentist procedures.

Keywords: Partial Identification, Bayesian Robustness, Lower and Upper Probabilities, Belief Function, Imprecise Probability, Gamma-minimax, Choquet Expected Loss, Random Set, Moment Inequality.

JEL Classification: C12, C15, C21.

*Email: t.kitagawa@ucl.ac.uk. I thank Andrew Chesher, Jean-Pierre Florens, Charles Manski, Adam Rosen, and Elie Tamer for valuable discussions and comments. All errors are mine.
1 Introduction

The conditional perspective claims that statistical inference and decision making should be conditional on what has been actually observed. The Bayesian method is one of the most commonly used procedures among those which stand on the conditional viewpoint. In the situation where the data (likelihood function) is not informative about the parameter of interest, then a specification of prior distribution may 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 due to the posterior sensitivity to prior, 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 discarding the conditionality principle, this paper develops a robust procedure of Bayesian inference and decision for set-identified models. Being different from the existing Bayesian analysis of the partially identified model (Bollinger and van Hasselt (2009), Gustafson (2009, 2010), Moon and Schorfheide (2009) and Liao and Jiang (2010)), our analysis aims at developing the posterior inference procedure that is insensitive to the specification of priors. When the parameters are not identified, or, more precisely, the likelihood is flat over some regions in the parameter space, a prior distribution for the parameters can be 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 in order to make posterior probabilistic judgement free from the unrevisable prior knowledge, we introduce the class of prior distributions into the analysis. Specifically, we accept a single probability distribution for the revisable prior knowledge, while we introduce the class of arbitrary prior distributions for the unrevisable prior knowledge. The Bayes rule is used to update the revisable prior knowledge, and the updated revisable prior and each unrevisable prior within the class generates a posterior of the set-identified parameters. So, as a result of introducing the class of priors for the unrevisable, we obtain the class of posteriors for the set-identified parameters.

We summarize these multiple posteriors by 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) in his fiducial argument of deriving the posterior probabilities, and Dempster (1968) extends the
idea to incorporate the partial prior knowledge into the Bayesian inference. Shafer (1976, 1982) employs the mathematical structure of the Dempster’s lower and upper probabilities to develop the belief function analysis: a system of probabilistic judgement and learning procedure that deal with both partial prior knowledge and imprecise or set-valued observations. Walley (1991) introduces the lower and upper probabilities into the Bayesian subjectivism to model the degree of prior ignorance and indecision in the decision problem. Our use of lower and upper probabilities is motivated by the robust Bayes analysis considered in DeRobertis and Hartigan (1981), Wasserman (1989, 1990), and Wasserman and Kadane (1990) where they define the prior class in terms of the lower and upper probabilities and consider updating rules of the lower and upper probabilities. These works, however, look at an identified model and it is not clear how to implement the lower and upper probability analysis to general non-identified models. The main contributions of this paper are, therefore, i) to clarify how an early idea of lower and upper probability analysis can fit to the recent issue on inference and decision for the partially identified models, and ii) to demonstrate that by designing the prior class to include arbitrary unrevisable prior knowledge, the posterior lower and upper probabilities corresponding to the prior class is analytically tractable and useful to conduct the robust inference and conservative decision in the partially identified model in econometrics.

With our specification of the prior class, the resulting posterior lower and upper probabilities are shown to be nonadditive measures. This implies that the posterior analysis based on the lower and upper probabilities must depart from the standard Bayesian posterior analysis using the posterior probability distribution. 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) an appropriately defined identified set, which is a posteriori random with its source of randomness coming from the posterior uncertainty for the identified component in the model. We consider a point estimator for the set-identified parameter based on the gamma-minimax criterion (Berger (1985)), which leads us to an action 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 has been limited to rather simple parametric models with a certain choice of prior class (see, e.g., Betro and Ruggeri (1992) and Vidakovic (2000)). In contrast, our specification of prior class gives a surprisingly simple analytical solution for the conditional gamma-minimax action in a general class of non-identified models. We show that the gamma-minimax action is obtained by minimizing the Choquet expected loss with respect to the posterior upper probability. The closed form expression of the point estimator is not available in general, but it can be easily computed with a help of the Markov Chain Monte Carlo method.
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 the highest probability region of the posterior lower probability. Since the posterior lower probability is not a probability measure so that the construction and interpretation of the posterior credible region differs from the standard Bayesian credible region constructed upon a posterior probability distribution. The interpretation of \( C_{\alpha^*} \), the volume minimizing posterior lower credible region with credibility \( \alpha \) is that \( C_{\alpha^*} \) is the smallest subset in the parameter space on which we place at least probability \( \alpha \) irrespective of the unrevisable prior knowledge. This paper provides an algorithm to construct such credible regions when the parameter of interest is a scalar. We analyze the asymptotic property of the volume minimizing posterior lower credible region, and compare it with the confidence regions constructed in the criterion function approach of Chernozhukov, Hong, and Tamer (2007).

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 derive the posterior distribution of the set-identified parameter, and we pin down revisable and unrevisable prior knowledge. In Section 4, we introduce the class of prior distributions for the unrevisable, and derive the posterior lower and upper probabilities. The point estimation problem with the multiple priors are examined in Section 5. In Section 6, we investigate how to construct the posterior credible region based on the posterior lower probability. Its large sample behaviour is examined and compared with a criterion function approach with an intervally identified parameter. Proofs and lemma are provided in Appendix A.

2 Likelihood and Set Identification

2.1 The General Framework

Let \((X, \mathcal{X})\) and \((\Theta, \mathcal{A})\) be measurable spaces of a sample and parameters, respectively. Our analytical framework up to Section 4 only requires the parameter space \( \Theta \) to be a Polish (complete separable metric) space, and it covers both a parametric model \( \Theta = \mathbb{R}^d, d < \infty \), and a nonparametric model where \( \Theta \) is a separable Banach space. We denote a probability measure on the product measurable space \((X \times \Theta, \mathcal{X} \otimes \mathcal{A})\) by \( \Pi \), and let \( \mu_\theta \) be a marginal probability distribution on the parameter space \((\Theta, \mathcal{A})\) referred to as a prior distribution for \( \theta \). Also, let \( P(\cdot|\theta) \) defined on \((X, \mathcal{X})\) be the likelihood, which is defined as a transition probability from \((\Theta, \mathcal{A})\) to \((X, \mathcal{X})\) that satisfies (i) \( P(\cdot|\theta) \) is a probability measure on \((X, \mathcal{X})\) for every \( \theta \in \Theta \), and (ii) \( P(X|\cdot) \) is \( \mathcal{A} \)-measurable for all \( X \in \mathcal{X} \). We assume that \( P(\cdot|\theta) \) has a Radon-Nikodym derivative \( p(x|\theta) \) at every \( \theta \in \Theta \) with respect to a \( \sigma \)-finite measure on
The parameter vector $\theta$ consists of the parameters that govern behaviors of economic agents as well as those that characterize the distribution of unobserved heterogeneities among the agents. Alternatively, in the context of missing data or the counterfactual causal model, $\theta$ would index the distribution of the underlying population outcomes or the potential outcomes. In any of these, the parameter $\theta$ should be distinguished from the parameters that are used solely to summarize the sampling distribution of data. The identification problem of $\theta$ arises in this context, and if multiple values of $\theta$ can generate the same distribution of data, then we claim that these $\theta$’s are observationally equivalent and identification of $\theta$ fails.

In terms of the likelihood function, the observational equivalence of $\theta = \theta'$ is equivalent to saying that the values of the likelihood function at $\theta$ and $\theta'$ are identical no matter what the observations are, i.e., \( p(x|\theta) = p(x|\theta') \) for every $x \in X$ (Rothenberg (1971), Drèze (1974), and Kadane (1974)). We represent the observational equivalence relation of $\theta$’s by a many-to-one and onto function $g : (\Theta, \mathcal{A}) \rightarrow (\Phi, \mathcal{B})$,

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

The equivalence relationship partitions the parameter space $\Theta$ into the equivalent classes on each of which the likelihood is "flat" irrespective of the observations, and $\phi = g(\theta)$ maps each of these equivalent classes to a point in another parameter space $\Phi$. 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 $\theta$ 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)).

Let $\mathcal{A}_0$ be the smallest sub-$\sigma$-algebra of $\mathcal{A}$ with respect to which the likelihood $p(x|\theta)$ is measurable for all $x \in X$. It is known that $\mathcal{A}_0$ is unique and coincides with the sub-$\sigma$-algebra in $\mathcal{A}$ induced by $g(\cdot)$ (See Lemma 2.7 of Picci (1977)). That is, in the presence of the sufficient parameter, the likelihood depends on $\theta$ only through an $\mathcal{A}_0$-measurable function $g(\theta)$ (Lemma 2.3.1 of Lehmann and Romano (2005)), and there exists a $\mathcal{B}$-measurable function $\hat{p}(x|\cdot)$ such that

\[ p(x|\theta) = \hat{p}(x|g(\theta)) \quad \forall x \in X \text{ and } \theta \in \Theta. \]  

(2.1)

Consider the inverse map of $g(\cdot)$ denoted by $\Gamma : \Phi \rightarrow \mathcal{A}_0$,

\[ \Gamma(\phi) = \{ \theta \in \Theta : g(\theta) = \phi \}. \]
Since $g(\theta)$ represents the observational equivalence, $\Gamma(\phi)$ and $\Gamma(\phi')$ for $\phi \neq \phi'$ are disjoint, and $\{\Gamma(\phi) ; \phi \in \Phi\}$ constitutes a partition of $\Theta$. In the structural model of econometrics, $\Gamma(\phi)$ can be seen as a set of observationally equivalent $\theta$’s that share the same value of the reduced form parameters. We assume throughout that $\Gamma(\phi)$ is nonempty for every $\phi \in \Phi$. 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 $g(\Theta)$, and in this case the model possesses the refutability property, i.e., $\Gamma(\phi)$ can be empty for some $\phi$. The above assumption, however, precludes the refutable model from our analysis.

We define the set-identification of $\theta$ and the identified set of $\theta$ as follows.

**Definition 2.1 (Set-identification and identified set of $\theta$)** (i) The model is set-identified if $\Gamma(\phi)$ is not a singleton for some $\phi \in \Phi$. Equivalently, the model is set-identified if the sufficient parameter $\sigma$-algebra $A_0$ is a proper sub-$\sigma$-algebra of $A$.

(ii) The inverse map of $g(\cdot)$, $\Gamma : \Phi \to A_0$ is called an identified set of $\theta$.

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 $\Gamma$ is seen as a multi-valued map from the sufficient parameter space $\Phi$ to the original parameter space $\Theta$. Note that the identification of $\theta$ only relies on the likelihood $p(x|\theta)$ and its definition does not change before and after observing data. In this sense, the Bayesians and frequentists share the same concept of identification (Kadane (1974)), and, furthermore, the identification argument does not require the hypothetical argument of availability of infinite number of observations.

In many of the set-identified models, the parameter of interest is given by a subvector or a transformation of $\theta$. We denote parameters of interest by $\eta = h(\theta)$ where $h : (\Theta, A) \to (H, D)$ be a measurable function of $\theta$. Elimination of the nuisance parameters in $\theta$ interprets $h(\cdot)$ as a coordinate projection of $\theta$. We define the identified set of $\eta$ simply by the projection of $\Gamma(\phi)$ onto $H$ through $h(\cdot)$.

**Definition 2.2 (Set-identification and identified set of $\eta$)** Let $H(\phi) \equiv \{h(\theta) : \theta \in \Gamma(\phi)\}$ be the image of $h(\cdot)$ with the domain given by $\Gamma(\phi)$. The parameter $\eta = h(\theta)$ is set-identified if $H(\phi)$ is not a singleton for some $\phi \in \Phi$. If it is the case, we call $H(\phi)$ the identified set of $\eta$.

The task of constructing the sharp bounds of $\eta$ in the partially identified model is essentially equivalent to finding the expression of $H(\phi)$ where $\phi$ is a parameter vector indexing the
sampling distribution of data. Below, we shall provide some examples of partially identified models where a closed form expression or a numerical method to compute \( H(\phi) \) is available.

### 2.2 Some Examples

To illustrate the formulation introduced above, we shall provide several examples, each of which the econometrics literatures have analyzed as a partially identified model.

**Example 2.1 (Missing data)** Let \( Y \in \{1, 0\} \) be the binary outcome. In data, some observations of \( Y \) is missing for some unknown reason. We indicate whether \( Y \) is observed or not by \( D \in \{1, 0\} \): \( D = 1 \) if \( Y \) is observed and \( D = 0 \) if \( Y \) is missing. Data is given by size \( N \) random sample \( x_N = \{(Y_i, D_i) : i = 1, \ldots, N\} \). Let us take the parameter vector \( \theta \) as the following probability masses,

\[
\begin{align*}
\theta_{11} &= \Pr(Y = 1, D = 1), & \theta_{01} &= \Pr(Y = 0, D = 1), \\theta_{10} &= \Pr(Y = 1, D = 0), & \theta_{00} &= \Pr(Y = 0, D = 0).
\end{align*}
\]

The parameter space for \( \theta \) is \( \Theta = \Delta^d_p \) where \( \Delta^d_p \) is the \( d \)-dimensional probability simplex. The likelihood of the observed data is given by

\[
p(x_N | \theta) = c(x_N)\theta_{11}^{n_{11}}\theta_{01}^{n_{01}}[\theta_{10} + \theta_{00}]^{n_{mis}}
\]

where \( n_{11} = \sum_{i=1}^{N} Y_i D_i, \ n_{01} = \sum_{i=1}^{N} (1 - Y_i) D_i, \ n_{mis} = \sum_{i=1}^{N} (1 - D_i), \) and \( c(x_N) \) be a constant that only depends on the observations. With \( \theta_{11} \) and \( \theta_{01} \) being fixed, any pairs of \( \theta_{10} \geq 0 \) and \( \theta_{00} \geq 0 \) whose sum is equal to \( 1 - \theta_{11} - \theta_{01} \) can generate the same probability law of data. That is, the likelihood function only depends on the triplet \((\phi_1, \phi_2, \phi_3) \equiv (\theta_{11}, \theta_{01}, \theta_{10} + \theta_{00}) = (\Pr(Y = 1, D = 1), \Pr(Y = 0, D = 1), \Pr(D = 0))\) irrespective of the observations \( x_N \). Hence, the sufficient parameters is obtained as

\[
\phi \equiv g(\theta) = (\theta_{11}, \theta_{01}, \theta_{10} + \theta_{00}) \equiv (\phi_1, \phi_2, \phi_3) \in \Delta^3_p,
\]

and its inverse map is written as

\[
\Gamma(\phi) = \{ \theta \in \Delta^4_p : \theta_{11} = \phi_1, \theta_{01} = \phi_2, \theta_{10} + \theta_{00} = \phi_3 \}.
\]

According to Definition 2.1, \( \theta \) is set-identified since \( \Gamma(\phi) \) is not always a singleton. a set valued map from \( \Delta^3_p \) to \( \Delta^4_p \). If the parameter of interest is the mean of \( Y \), \( \eta \equiv \Pr(Y = 1) = \theta_{11} + \theta_{10} \), then, since \( \theta_{11} = \phi_1 \) and \( 0 \leq \theta_{10} \leq \phi_3 \), the identified set of \( \eta \) is obtained by

\[
H(\phi) = [\phi_1, \phi_1 + \phi_3].
\]

This is identical to the sharp bounds of \( \Pr(Y = 1) \) given in Manski (1989).
Example 2.2 (Treatment Effect Model with a Randomized Experiment) Consider the Neyman-Rubin potential outcome model for inferring the causal effects in the randomized experiment setting. Let $D \in \{1, 0\}$ be an indicator for a binary treatment status, and $(Y_1, Y_0) \in \mathcal{Y} \times \mathcal{Y}$ be the pair of potential outcomes. Let $D$ be an indicator for a binary treatment status, and $(Y_1, Y_0)$ be the pair of potential outcomes. Let $Y$ be the actual observed outcomes $Y = D Y_1 + (1 - D) Y_0$, and data is a size $N$ random sample of $X = (Y, D)$ denoted by $x_N = \{(y_i, d_i) : i = 1, \ldots, N\}$. We consider the case of randomized treatment, so $D$ is independent of $(Y_1, Y_0)$. Therefore, the parameters in the model can be represented by $\theta = (f_{Y_1}, f_{Y_0}, p)$ where $f_{Y_1}$ represents a nonparametric density of joint distribution of $(Y_1, Y_0)$ and $p \equiv \Pr(D = 1)$.

The observed data likelihood in this example is written as 

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

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 obtained as $\phi = (f_{Y_1}, f_{Y_0}, p)$ 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 uniquely determined once $f_{Y_1}$ and $f_{Y_0}$ are given, so $H(\phi)$ of Definition 2.1 is a singleton for every $\phi$ and we conclude that ATE is identified.

Next, consider $\eta = F_{Y_1-Y_0}(0)$ the cumulative distribution function of the individual causal effects evaluated at zero. This can be a parameter of interest if the researcher wants to know how much fraction of the population can be benefited 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 the 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 obtained by the Makarov’s bounds (Makarov (1981)). Hence, $\eta$ is set-identified. For further details of bounding $F_{Y_1-Y_0}(-\cdot)$, see Heckman Smith, and Clemens (1997), Fan and Park (2009), and Firpo and Ridder (2009).

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

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

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

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 $\theta$ that is known to have a Bayesian justification (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 likelihood is written as

$$p(x_N|\theta) = \prod w_i(\theta)$$

where

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

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

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 that the data is only informative about. The identified set for $\theta$ is given by,

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

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

### 3 Posterior of $\theta$ in the Presence of Sufficient Parameters

In this section, we present the posterior of $\theta$ given a prior of $\theta$. Let $\mu_\phi$ be the marginal probability measure on the sufficient parameter space $(\Phi, \mathcal{B})$ induced by $\mu_\theta$ and $g(\cdot)$, i.e.,

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

\footnote{The Bayesian formulation of the moment inequality model shown here owes to Tony Lancaster, who suggested this to us in 2006.}
Let \( x \in X \) be sampled data. The Radon-Nykodim derivative of \( \Pi(\cdot \times A) \) with respect to \( dx \) is written as
\[
\frac{d\Pi(\cdot \times A)}{dx} = \int_\Theta p(x|\theta)1_A(\theta)d\mu_\theta = \int \mathbb{E}(p(x|\theta)1_A(\theta)|A_0)d\mu_\theta = \int \hat{p}(x|\theta)\mathbb{E}(1_A(\theta)|A_0)d\mu_\theta = \int \hat{p}(x|\theta)\mu_{\theta|\phi}(A|\phi)d\mu_\phi
\]
where \( \mathbb{E}(\cdot|A_0) \) is the conditional expectation with respect to the sufficient parameter \( \sigma \)-algebra \( A_0 \) and \( \mu_{\theta|\phi}(A|\phi) \) is the conditional distribution of \( \theta \) given \( \phi \) that satisfies
\[
\mu_\theta(A \cap \Gamma(B)) = \int_B \mu_{\theta|\phi}(A|\phi)d\mu_\phi \quad \text{for all} \quad A \in \mathcal{A}, \ B \in \mathcal{B}.
\]
The second line follows by the definition of the conditional expectation with respect to the sub-\( \sigma \)-algebra, and the third line follows since \( p(x|\theta) \) is \( A_0 \)-measurable and by (2.1). Similarly, the Radon-Nykodim derivative of the marginal sampling probability \( \Pi(X \times \Theta) \) with respect to \( dx \) is obtained as
\[
\frac{d\Pi(\cdot \times \Theta)}{dx} = \int \hat{p}(x|\phi)d\mu_\phi.
\]
Suppose that \( \mu_\phi \) has a dominating measure with the Radon-Nykodim derivative \( d\mu_\phi/d\phi = \tilde{\mu}_\phi \). Then, the posterior probability measure denoted by \( F_{\theta|X}(A), A \in \mathcal{A} \), is obtained as
\[
F_{\theta|X}(A) = \frac{\int_\Phi \mu_{\theta|\phi}(A|\phi)\hat{p}(x|\phi)d\phi}{\int_\Phi \hat{p}(x|\phi)d\mu_\phi} = \int_\Phi \mu_{\theta|\phi}(A|\phi)f_{\phi|X}(\phi|x)d\phi
\]  
(3.1)
where \( f_{\phi|X}(\phi|x) \) is the posterior density of \( \phi \),
\[
f_{\phi|X}(\phi|x) = \frac{\hat{p}(x|\phi)\tilde{\mu}_\phi(\phi)}{\int_\Phi \hat{p}(x|\phi)d\mu_\phi}.
\]
The expression of the posterior of \( \theta \) (3.1) highlights a fundamental feature of the set identified model: the posterior of \( \theta \) is the average of the conditional prior \( \mu_{\theta|\phi}(A|\phi) \) with respect to the posterior of the sufficient parameter \( f_{\phi|X}(\phi|x) \). Poirier (1998) obtained the same expression as above and he demonstrated that the data only allows us to revise belief on the sufficient parameters \( \phi \), while it does not for the conditional distribution of \( \theta \) given \( \phi \).
This result may be intuitively understood by the conditional independence implied by the parameter sufficiency. When the likelihood is endowed with the sufficient parameters, it holds (see, e.g., Florens et al. (1990))

\[ X \perp \theta \mid \phi. \]

That is, given the knowledge of the sufficient parameter \( \phi \), the observation does not convey any information on the parameter \( \theta \) because the likelihood is flat on \( \Gamma(\phi) \). It implies that the conditional distribution of \( \theta \) given \( X \) and \( \phi \) is identical to the conditional distribution of \( \theta \) given \( \phi \), and accordingly, the average of \( \mu_{\theta|\phi} \) with respect to the posterior of \( \phi \) yields the posterior of \( \theta \) given \( X = x \) as shown in (3.1).

4 Multiple-Prior Analysis and the Lower and Upper Probability

4.1 A Class of Priors

As shown above, the posterior distribution of \( \theta \) (3.1) is determined by the two components: \( \mu_{\theta|\phi} \) the conditional prior information of \( \theta \) given \( \phi \) and \( f_{\phi|X}(\phi|x) \) the posterior distribution of \( \phi \). We are able to revise the prior information on \( \phi \) by data so the posterior of \( \phi \) will be eventually dominated by the likelihood. In contrast, we are incapable of updating the conditional prior information of \( \theta \) given \( \phi \) due to the flat likelihood on \( \Gamma(\phi) \). In this sense, we can consider \( \mu_{\phi} \) to be the revisable prior knowledge and \( \mu_{\theta|\phi} \) to be the unrevisable prior knowledge.

If we want to represent the posterior uncertainty of \( \theta \) 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 \( \theta \), and this requires us to specify the unrevisable prior knowledge \( \mu_{\theta|\phi} \). If the researcher could justify his choice of \( \mu_{\theta|\phi} \) by any credible prior information, the standard Bayesian updating (3.1) would yield the valid posterior distribution of \( \theta \). From the robustness point of view, however, the statistical procedure that requires us to specify the unrevisable prior knowledge may be problematic especially when the researcher cannot translate his prior belief into a probability judgement about the parameter of interest, or her or his prior knowledge for the parameter is totally vacuous.

The partial identification analysis in econometrics seems to be motivated to obtain a robust conclusion that is free from empirically unverifiable assumptions. This paper attempts to pursue a similar spirit without departing from the conditionality principle by introducing
the class of prior distributions that represents arbitrary unrevisable prior knowledge.\textsuperscript{2}

Let $\mathcal{M}$ be the set of probability measures on $(\Theta, A)$ and $\mu_\phi$ be a prespecified prior on $(\Phi, B)$. We assume that $\mu_\phi$ is absolutely continuous with respect to a $\sigma$-finite measure on $(\Phi, B)$. Formally, the class of prior distributions of $\theta$ to be used is written as

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

In words, $\mathcal{M}(\mu_\phi)$ consists of the prior distributions of $\theta$ whose marginal for the sufficient parameters are identical to a prespecified $\mu_\phi$. That is, we accept a single prior distribution for the sufficient parameters $\phi$ while we allow the class to contain 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 $(\Theta, A)$. Note that the above class precludes improper prior of $\mu_\theta$ and $\mu_\phi$.

There are several reasons for considering this class. First, our goal is to make inference or 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 summarizing in a certain way the class of posteriors of $\theta$ induced by the prior class $\mathcal{M}(\mu_\phi)$. Specifically, we shall consider the posterior lower probability defined by $F_{\theta|X}(A) \equiv \inf_{\mu_\phi \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(A)$, for $A \in A$. This quantity is interpreted as that the posterior credibility for $\theta \in A$ is at least $F_{\theta|X}(A)$ irrespective of the unrevisable prior knowledge. Second, as we will show below, the prior class $\mathcal{M}(\mu_\phi)$ is analytically easy to work with and practically feasible to implement by employing the standard Bayesian computing of Markov Chain Monte Carlo.

As the final remark of this section, we comment that existing selection rules for "non-informative" or "recommended" prior for finite dimensional $\theta$ are not applicable if the model lacks identification. First of all, Jeffreys' general rule (Jeffereys (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 $\theta$ is nonsingular at almost every $\theta \in \Theta$ (see the examples of Section 2.2.)

In addition, the empirical Bayes type approach of choosing a prior for $\theta$ (Robbins (1951), Good (1965), Morris (1983), and Berger and Berliner (1986)) breaks down if the model involves the sufficient parameters. The empirical Bayes "robust" prior is obtained by finding a prior within the class that maximizes the marginal likelihood of data. That is, we choose

\textsuperscript{2}Bayesian analysis with multiple priors has been extensively studied in the literature of empirical Bayes and the robust Bayes analysis (see Robbins (1964), Good (1965), Blum and Rosenblatt (1967), Berger and Berliner (1982), Berger (1985), Huber (1973), Huber and Strassen (1973), Wasserman (1989, 1990) and the references therein). To my knowledge, however, none of the existing work has analysed the non- or partially identified model from the viewpoint of multiple prior Bayes analysis.
\( \mu \) so as to maximize

\[
m(x|\mu) = \int_{\Theta} p(x|\theta)d\mu
\]

over \( \mu \in \mathcal{M}(\mu_\phi) \). Unfortunately, if the likelihood has the sufficient parameters, the marginal distribution \( m(x|\mu) \) depends only on \( \mu_\phi \), since

\[
\int_{\Theta} p(x|\theta)d\mu = \int_{\phi} \hat{p}(x|\phi)d\mu_\phi \equiv m(x|\mu_\phi).
\]

Hence, the empirical Bayes approach fails to select a prior for \( \mu \) due to the lack of ability to compare the desirability of priors within \( \mathcal{M}(\mu_\phi) \).

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

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

This objective function is written as

\[
\int_{\phi} \log \left( \frac{\hat{p}(x|\phi)}{m(x|\mu_\phi)} \right) \cdot \frac{\hat{p}(x|\phi)}{m(x|\mu_\phi)} d\mu_\phi,
\]

and, therefore, it is only a function of \( \mu_\phi \) rather than \( \mu_\theta \). Hence, the conditional Kullback-Leibler distance does not offer a criterion of selecting \( \mu_\theta \) out of \( \mathcal{M}(\mu_\phi) \).

These prior selection rules are useful only for choosing a prior for the sufficient parameters \( \mu \), but not at all for selecting \( \mu_{\theta|\phi} \), or equivalently \( \mu_\theta \) within \( \mathcal{M}(\mu_\phi) \). This is another rationale for us to proceed the multiple prior analysis rather than the one with a single prior. In this paper, we will not discuss how to select \( \mu_\phi \), and treat \( \mu_\phi \) as given by the researcher.

### 4.2 Posterior Lower and Upper Probabilities

The prior class \( \mathcal{M}(\mu_\phi) \) results in yielding the class of posterior distributions of \( \theta \). In order to summarize the posterior class, we focus on the posterior lower probability \( F_{\theta|X}(\cdot) \) and the posterior upper probability \( F_{\theta|X}^*(\cdot) \), which are defined as, for \( A \in \mathcal{A} \),

\[
F_{\theta|X}(A) \equiv \inf_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(A),
\]

\[
F_{\theta|X}^*(A) \equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} F_{\theta|X}(A).
\]

For the lower and upper probabilities to be well defined, we assume the following regularity conditions.
Condition 4.1  (i). (measurability). For every $A \in \mathcal{A}$, $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\} \in \mathcal{B}$.

(ii). (closedness). $\Gamma(\phi)$ is a closed set in $\Theta$ for every $\phi \in \Phi$.

(iii) The likelihood $P(\cdot|\theta)$ is absolutely continuous with respect to a $\sigma$-finite measure on $(X, \mathcal{X})$ for every $\theta \in \Theta$, and the prior of $\phi$, $\mu_{\phi}$, is absolutely continuous with respect to a $\sigma$-finite measure on $(\Phi, \mathcal{B})$.

The first condition guarantees that the events $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}$, $\{\phi : \Gamma(\phi) \cap A = \emptyset\}$, and $\{\phi : \Gamma(\phi) \subset A\}$ are supported by a probability measure on $(\Phi, \mathcal{B})$.

The next theorem shows that the lower and upper probabilities defined above are equivalent to the posterior probabilities of the events $\{\phi : \Gamma(\phi) \subset A\}$ and $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}$, respectively, in terms of the posterior of $\phi$, $F_{\phi|X}(\cdot)$.

Theorem 4.1 Assume Condition 4.1.

(i) For each $A \in \mathcal{A}$,

$$F_{\phi|X}(A) = \int_{\Phi} 1\{\phi : \Gamma(\phi) \subset A\} f_{\phi|X}(\phi|x) d\phi = F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A\}), \tag{4.2}$$

$$F^*(A|x) = \int_{\Phi} 1\{\phi : \Gamma(\phi) \cap A \neq \emptyset\} f_{\phi|X}(\phi|x) d\phi = F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}), \tag{4.3}$$

where $F_{\phi|X}(B)$ is the posterior probability measure of $\phi$, $F_{\phi|X}(B) = \int_B f_{\phi|X}(\phi|x) d\phi$ for $B \in \mathcal{B}$.

(ii) Let

$$G_{\theta} = \left\{ G_{\theta} : G_{\theta} \text{ probability measure on } (\Theta, \mathcal{A}) , F_{\theta|X}(A) \leq G_{\theta}(A) \leq F^*_{\theta|X}(A) \text{ for every } A \in \mathcal{A} \right\}. \tag{4.4}$$

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

(iii) Let $\eta = h(\theta)$ be an $\mathcal{A}$-measurable transformation of $\theta$, $h : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$, and $H(\phi)$

---

In the theory of random closed set, the measurability is often assumed in terms of $\{\phi : \Gamma(\phi) \cap K \neq \emptyset\}$ for every compact set $K$ in $\Theta$. Since $\Theta$ is assumed to be a Polish space, this measurability condition is equivalent to the measurability of $\{\phi : \Gamma(\phi) \cap A \neq \emptyset\}$ for any subset $A \in \mathcal{A}$ (Theorem 2.3 of Molchanov (2005)).
be a projection of \( \Gamma(\phi) \) through \( h(\cdot) \), \( H(\phi) = h(\Gamma(\phi)) \). Define the posterior lower and upper probabilities of \( \eta \) by for each \( D \in \mathcal{D} \),

\[
F_{\eta|X^*}(D) = \inf_{\mu_\eta \in \mathcal{M}(\mu_\phi)} F_{\eta|X}(D) = \inf_{\mu_\eta \in \mathcal{M}(\mu_\phi)} F_{\eta|X}(h^{-1}(D)),
\]

\[
F_{\eta|X}^*(D) = \sup_{\mu_\eta \in \mathcal{M}(\mu_\phi)} F_{\eta|X}(D) = \sup_{\mu_\eta \in \mathcal{M}(\mu_\phi)} F_{\eta|X}(h^{-1}(D)).
\]

Then,

\[
F_{\eta|X^*}(D) = F_{\phi|X}(\{\phi : H(\phi) \subset D\}),
\]

\[
F_{\eta|X}^*(D) = F_{\phi|X}(\{\phi : H(\phi) \cap D \neq \emptyset\}).
\]

**Proof.** For a proof of (i) and (ii), see Appendix A. For a proof of (iii), see the equations (4.4) and (4.5) below. ■

The expression of \( F_{\eta|X^*}(A) \) given above implies that the lower probability on \( A \) is interpreted as the posterior probability that the random set \( \Gamma(\phi) \) is contained in subset \( A \subset \Theta \). On the other hand, the upper probability is interpreted as the posterior probability that the random set \( \Gamma(\phi) \) hits the subset \( A \). Our proof given in Appendix A is not restricted to finite dimensional \( \theta \), and the results hold even for infinite dimensional separable \( \Theta \). Intuitively speaking, the lower (upper) probability is attained by plugging a pointwise minimum (maximum) of \( \mu_{\theta|\phi}(A|\phi) \) into \( F_{\eta|X}(A) = \int_{\Phi} \mu_{\theta|\phi}(A|\phi)dF_{\phi|X} \). Since the class of priors \( \mathcal{M}(\mu_\phi) \) allows for arbitrary conditional distributions of \( \theta \) given \( \phi \) as far as its support is contained in \( \Gamma(\phi) \), for \( \phi \) with \( \Gamma(\phi) \subset A \), the infimum of \( \mu_{\theta|\phi}(A|\phi) \) is one and, for \( \phi \) with \( \Gamma(\phi) \) having some elements outside of \( A \), the infimum of \( \mu_{\theta|\phi}(A|\phi) \) is obtained as zero by letting the conditional measure \( \mu_{\theta|\phi}(\cdot|\phi) \) concentrating on these elements outside of \( A \). As a result, the lower probability becomes the posterior probabilities of the containment events \( F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A\}) \).

Conversely, for the upper probability \( F_{\eta|X}^*(A) \), the supremum of \( \mu_{\theta|\phi}(A|\phi) \) to be plugged into \( \int_{\Phi} \mu_{\theta|\phi}(A|\phi)dF_{\phi|X} \) is one whenever \( \Gamma(\phi) \) has a nonempty intersection with \( A \) and, it is zero otherwise. Hence, the upper probabilities can be written as the posterior probability of \( \Gamma(\phi) \) hitting \( A \).

From the expression of \( F_{\eta|X^*}(\cdot) \) and \( F_{\eta|X}^*(\cdot) \) given above, we can see that they do not necessarily satisfy additivity condition for measures, i.e., for disjoint subsets \( A_1 \) and \( A_2 \) in \( \mathcal{A} \),

\[
F_{\eta|X^*}(A_1 \cup A_2) = F_{\phi|X}(\{\phi : \Gamma(\phi) \subset (A_1 \cup A_2)\})
\]

\[
= F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A_1\}) + F_{\phi|X}(\{\phi : \Gamma(\phi) \subset A_2\})
\]

\[
+ F_{\phi|X}(\{\phi : \Gamma(\phi) \not\subset A_1, \Gamma(\phi) \not\subset A_2, \Gamma(\phi) \subset (A_1 \cup A_2)\})
\]

\[
\geq F_{\eta|X^*}(A_1) + F_{\eta|X^*}(A_2),
\]
\[
F_{\theta|X}(A_1 \cup A_2) = F_{\phi|X}(\{\phi : \Gamma(\phi) \cap (A_1 \cup A_2) \neq \emptyset\})
\]

\[
= F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A_1 \neq \emptyset\}) + F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A_2 \neq \emptyset\})
\]

\[
- F_{\phi|X}(\{\phi : \Gamma(\phi) \cap A_1 \neq \emptyset, \Gamma(\phi) \cap A_2 \neq \emptyset\})
\]

\[
\leq F_{\theta|X}^*(A_1) + F_{\theta|X}^*(A_2).
\]

In fact, the non-additivity of the lower and upper probabilities is a well-known fact (e.g., Dempster (1967)), and the lower and upper probabilities can be seen as capacities.\(^4\) If the model is identified, i.e., \(\Gamma(\phi)\) is a singleton \(f_{\phi|X}\)-almost surely, then \(F_{\theta|X}^*(\cdot) = F_{\theta|X}(\cdot)\) holds and the upper and lower probabilities become identical probability measures.

The second statement of the above theorem says that the class of posteriors induced by the prior class \(\mathcal{M}(\mu_\phi)\) exhausts all the probability measures lying between the lower and upper probabilities (c.f., Huber and Ronchetti (2009) calls this property as representability of the probability class by the lower and upper probabilities.). This property can be interpreted as convexity of the resulting posterior class, and useful to establish the equivalence between the upper (lower) expectation and the Choquet integral with respect to the upper (lower) probability (see Section 5).

The third statement of the theorem provides a procedure to transform or marginalize the lower and upper probabilities of \(\theta\). The expressions of \(F_{\theta|X}^*(D)\) and \(F_{\eta|X}^*(D)\) obtained there are simple and easy to interpret: the lower and upper probabilities of \(\eta = h(\theta)\) are the containment and hitting probabilities of the random set (induced by the posterior of \(\phi\)) obtained by projecting \(\Gamma(\phi)\) through \(h(\cdot)\). The marginalization rules of lower and upper probabilities are deduced as follows,

\[
F_{\eta|X}(D) = F_{\theta|X}(h^{-1}(D))
\]

\[
= F_{\phi|X}(\{\phi : \Gamma(\phi) \subset h^{-1}(D)\})
\]

\[
= F_{\phi|X}(\{\phi : H(\phi) \subset D\}).
\]

\(^4\)Interestingly, our specification of the prior class yields the lower (upper) probability that is monotone (alternating) of infinite order, which is not always the case depending on the class of priors (c.f., Huber and Strassen (1973), Wasserman (1989)). This finding is directly implied by the Choquet Theorem, stating that any set-function that is expressed by the containment (hitting) probability of a random closed set is the capacity of infinite monotone (alternating) order (see Molchanov (2005)).
and, similarly,
\[
F^*_{\eta|X}(D) = F^*_{\eta|X}(h^{-1}(D)) = F_{\phi|X}(\{\phi : \Gamma(\phi) \cap h^{-1}(D) \neq \emptyset\}) = F_{\phi|X}(\{\phi : H(\phi) \cap D \neq \emptyset\}).
\] (4.5)

Analogous to the lower and upper probabilities of \( \theta \), \( F^*_{\phi|X}(\cdot) \) and \( F^*_{\phi|X}(\cdot) \) are capacities on \((\mathcal{H}, \mathcal{D})\) (supadditive and subadditive measures respectively.)

5 Point Estimation of \( \eta = h(\theta) \) with Multiple Priors

In this section, we analyze point estimation of the parameter of interest \( \eta = h(\theta) \) from the perspective of a point decision problem with several criteria. In Section 5.1, we consider the conditional decision problem as the conditional gamma-minimax action problem (Berger (1985, p205), DasGupta and Studden (1989), Betro and Ruggeri (1992), and Vidakovic (2000)). In the multiple prior decision problem, the optimal decision rule a priori in general differs from the optimal action a posteriori. Hence, we will examine in Section 5.2 whether the conditional gamma-minimax action can be interpreted as the unconditional gamma-minimax decision (Kudo (1967), Berger (1985, p213-218), Vidakovic (2000)). In Section 5.3, we replace the gamma-minimax criterion with the gamma-minimax regret criterion (Berger (1985, p218), and Rios Insua, Ruggeri, and Vidakovic (1995)), and derive analytical properties of the gamma-minimax regret rule under the specification of quadratic loss.

5.1 Conditional Gamma-minimax Action

We first consider point estimation of the parameter of interest \( \eta = h(\theta) \in \mathcal{H} \) that is a posteriori optimal in the sense of minimizing the posterior gamma-minimax risk criterion. Let \( \delta(\cdot) \) be a decision rule that maps each \( x \in \mathbf{X} \) to the action space \( \mathcal{H}_a \subset \mathcal{H} \). In case that we consider the conditional decision given \( x \), the decision \( \delta(x) \) at the observed data \( x \) is usually called action \( a_x \in \mathcal{H}_a \). Given a particular action to be taken and \( \eta_0 \) being the true state of nature, a loss function \( L(\eta_0, a) : \mathcal{H} \times \mathcal{H}_a \rightarrow \mathbb{R}_+ \) yields how much cost the decision maker owes by taking such action.

Given a single posterior of \( \eta \), the Bayes action is defined by the action that minimizes the posterior risk,

\[
\rho(\mu_\theta, a) \equiv \int_\mathcal{H} L(\eta, a) dF_{\eta|X}(\eta)
\] (5.1)
where the first argument $\mu_\theta$ represents the dependence of the posterior of $\eta$ on the specification of prior on $(\Theta, \mathcal{A})$. Since our analysis involves the multiple posterior distributions of $\eta$, our decision criterion should involve the class of posterior risks $\{\rho(\mu_\theta, a) : \mu_\theta \in \mathcal{M}(\mu_\phi)\}$. In particular, we adopt the conditional gamma-minimax criterion, which defines the optimality of the action in terms of minimizing the posterior upper risk: the supremum of the posterior risk with respect to $\mu_\theta \in \mathcal{M}(\mu_\phi)$,

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

**Definition 5.1** The conditional gamma-minimax action $a^*_x$ is the action that minimizes the posterior upper risk, i.e.,

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

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 pure 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 5.1** (i) For each $a \in \mathcal{A}_x$, loss function $L(\eta, a)$ is $\mathcal{D}$-measurable and nonnegative. (ii) Given a prior for $\phi$, the upper posterior probability for $\theta$, $F^*_{\theta|X}(\cdot)$ obtained in (4.3) is regular, i.e., for each $A \in \mathcal{A}$,

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

The regularity of the upper posterior probability stated in Condition 5.1 (ii) is satisfied if no particular realizations of the random closed set $\Gamma(\phi)$ occurs with a strictly positive probability (see Graf (1980)). Under this condition, the posterior upper risk $\rho^*(\mu_\phi, a)$ is

---

5In the frequentist approach to the decision problem (Wald (1950), Manski (2004, 2008, 2009), and Stoye (2009)), the loss function is averaged with respect to the distribution of data so that the decision problem is to seek for a decision function $\delta(\cdot)$ that performs best on average under the hypothetical repeated sampling with taking into account all the possible state of nature.

6In the robust Bayes literature, the class of prior distributions is often notated as $\Gamma$, and this is why it is called (conditional) gamma-minimax criterion. Unfortunately, in the literature of belief function and the lower and upper probabilities, $\Gamma$ often denotes a set-valued mapping that generates the lower and upper probabilities. In this article, we adopted the notational convention of the latter, while still refer to our decision criterion as the gamma minimax criterion.
equivalently written as the Choquet expected loss with respect to the upper probability (see the proof of Proposition 5.1 below),

$$\rho^*(\mu, a) = \sup_{\eta \in \mathcal{M}(\mu)} \int L(\eta, a) dF_{\eta|X}^*(\eta) = \int L(\eta, a) dF_{\eta|X}^*(\eta),$$

(5.2)

where the integral with respect to the nonadditive measures $F_{\eta|X}$ is defined in the sense of Choquet integral: for a nonadditive measure $T(\cdot)$ on $(\mathcal{H}, \mathcal{D})$ and a nonnegative measurable function $l$ on $\mathcal{H}$, the Choquet integral is defined by

$$\int l(\eta) dT(\eta) = \int_0^\infty T(\{\eta : l(\eta) \geq t\}) dt.$$

Furthermore, the fact that the upper probability $F_{\eta|X}^*(\eta)$ represents the hitting probability of the random closed set $H(\phi)$ guarantees us to write the Choquet expected loss by the expectation of $\sup_{\eta \in H(\phi)} L(\eta, a)$ with respect to the posterior distribution of $\phi$.

**Proposition 5.1** Under Condition 4.1 and Condition 5.1, the posterior upper risk satisfies

$$\rho^*(\mu, a) = \int L(\eta, a) dF_{\eta|X}^*(\eta) = \int \sup_{\eta \in H(\phi)} L(\eta, a) f_{\phi|X}(\phi|x)d\phi.$$  \hspace{1cm} (5.3)

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

**Proof.** See Appendix A. \hfill $\blacksquare$

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 $\phi$ from its posterior $f_{\phi|X}(\phi|x)$. Let $\{\phi_s\}_{s=1}^S$ be $S$ random draws of $\phi$ from $f_{\phi|X}(\phi|x)$. Then, we can approximate $a_x^*$ by

$$\tilde{a}_x^* \equiv \arg\min_{a \in \mathcal{H}} \tilde{d}_S(a) \quad \text{where} \quad \tilde{d}_S(a) \equiv \frac{1}{S} \sum_{s=1}^S \sup_{\eta \in H(\phi_s)} L(\eta, a).$$

If $\tilde{d}_S(a)$ uniformly converges in probability to $d(a)$ as $S \to \infty$, then $\tilde{a}_x^*$ thus constructed should be a valid approximation for the conditional gamma-minimax decision rule under these assumptions.
5.2 Unconditional Gamma-minimax Decision.

In this section, we analyze the decision rule that minimizes the unconditional gamma-minimax criterion. Let \( r(\mu_\theta, \delta) \) be the Bayes risk of a decision rule \( \delta : X \to \mathcal{H}_a \) defined by

\[
r(\mu_\theta, \delta) = \int_{\Theta} \left[ \int_X L(\eta(\theta), \delta(x)) p(x|\theta) dx \right] d\mu_\theta.
\]

Given our prior class \( \mathcal{M}(\mu_\phi) \), the unconditional Gamma-minimax decision \( \delta^* \) is obtained by minimizing the supremum of the Bayes risk, called as the unconditional gamma-minimax criterion, \( r^*(\mu_\phi, \delta) \equiv \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} r(\mu_\theta, \delta) \).

**Definition 5.2** The unconditional gamma-minimax decision \( \delta^* \) is the decision rule that minimizes the unconditional gamma-minimax criterion \( r^*(\mu_\phi, \delta) \), i.e.,

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

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 sample, and either being unconditional or conditional on data does not generate any difference in 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 (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,

\[
r(\mu_\theta, \delta) = \int_X \rho(\mu_\theta, \delta(x)) m(x|\mu_\theta) dx.
\]

Given \( \delta \) and the class of priors, \( \mu_\theta \) that maximizes \( r(\mu_\theta, \delta) \) does not necessarily maximizes \( \rho(\mu_\theta, \delta(x)) \) since \( r(\mu_\theta, \delta) \) depends on \( \mu_\theta \) not only through \( \rho(\mu_\theta, \delta(x)) \) but also through the marginal distribution of data \( m(x|\mu_\theta) \). Recall, however, that in the non-identified model the marginal distribution of data depends only on \( \mu_\phi \) (see (4.1)), and therefore

\[
\sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} r(\mu_\theta, \delta) = \int_X \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \rho(\mu_\theta, \delta(x)) m(x|\mu_\phi) dx
\]

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

---

7Decision problems similar to above have been considered in Kudo (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 \( \eta \) can be viewed as a posterior probability distribution over the coarse collection of subsets \( \{ H(\phi) : \phi \in \Phi \} \subset \mathcal{H} \).
Proposition 5.2 \( \delta^*(x) = a^*_x, m(x|\mu_\phi) \)-almost surely.

**Proof.** A proof is sketched above. ■

Thus, for the nonidentified model with our specification of prior class, the dynamic inconsistency of the gamma-minimax decision problem does not exist.

### 5.3 Gamma-minimax Regret

As an alternative to the (posterior) gamma-minimax risk criterion considered above, it is natural to consider the gamma-minimax regret criterion with multiple priors (Rios-Insua et al (1995)), which is seen as an extension of the minimax regret criterion of Savage (1951) to the Bayes decision problem with multiple priors. For an ease of analysis, we consider the case where the parameter of interest \( \eta \) is a scalar and the loss function is specified to be 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 5.3** Define the lower bound of the posterior risk for a given \( \mu_\theta \) by \( \underline{\rho}(\mu_\theta) = \inf_{a \in \mathcal{H}} \rho(\mu_\theta, a) \). The posterior gamma-minimax regret action \( a^*_x \in \mathcal{H} \) solves

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

**Definition 5.4** Define the lower bound of the Bayes risk for a given \( \mu_\theta \) by \( \underline{r}(\mu_\theta) = \inf_\delta r(\mu_\theta, \delta) \). The unconditional gamma-minimax regret decision \( \delta^{reg} : X \to \mathcal{H}_a \) solves

\[
\inf_\delta \sup_{\mu_\theta \in \mathcal{M}(\mu_\phi)} \{ r(\mu_\theta, \delta) - \underline{r}(\mu_\theta) \}.
\]

Let \( L(\eta, a) \) be the quadratic loss, \( (\eta - a)^2 \). Then, for a given \( \mu_\theta \), the posterior risk \( \rho(\mu_\theta, a) \) is minimized at \( \hat{\eta}_{\mu_\theta} \) the posterior mean of \( \eta \). Therefore, the lower bound of the posterior risk is simply the posterior mean squared error, \( \underline{\rho}(\mu_\theta) = E_{\eta|X}( \left( \eta - \hat{\eta}_{\mu_\theta} \right)^2 ) \), and the posterior regret can be written as

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

21
Let \([\eta_x, \bar{\eta}_x]\) be the range of posterior mean of \(\eta\) when \(\mu_0\) varies over \(M(\mu_\phi)\), which is assumed to exist. Then, the posterior gamma-minimax regret is simplified to
\[
\sup_{\mu_0 \in M(\mu_\phi)} \{ \rho(\mu_0, a) - \rho(\mu_\phi) \} = \begin{cases} 
(\bar{\eta}_x - a)^2 & \text{for } a \leq \frac{\eta_x + \bar{\eta}_x}{2}, \\
(\eta_x - a)^2 & \text{for } a > \frac{\eta_x + \bar{\eta}_x}{2},
\end{cases}
\]
and, clearly, the posterior gamma-minimax regret is minimized at \(a = \frac{\eta_x + \bar{\eta}_x}{2}\). That is, the posterior gamma-minimax regret action is simply obtained as the mid point of \([\eta_x, \bar{\eta}_x]\).

Furthermore, due to a similar logic to the one of the previous subsection, we can also show that the unconditional gamma-minimax regret decision \(\delta^{reg}(x)\) is the same as the posterior gamma-minimax regret action \(a^{reg}_x = \frac{\eta_x + \bar{\eta}_x}{2}\) for almost all sample.

**Proposition 5.3** Let \(H \subset \mathcal{R}\) and \(L(\eta, a) = (\eta - a)^2\). Assume that the posterior variance of \(\eta\) is finite for every \(\mu_0 \in M(\mu_\phi)\). Let \(\eta(\phi) \equiv \inf \{ \eta : \eta \in H(\phi) \}\) and \(\bar{\eta}(\phi) \equiv \sup \{ \eta : \eta \in H(\phi) \}\).

(i) The posterior gamma-minimax regret action is
\[
a^{reg}_x = \frac{E_{\phi|x}(\eta(\phi)) + E_{\phi|x}(\bar{\eta}(\phi))}{2}.
\]

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

**Proof.** See Appendix A. □

Since the lower bound of the posterior risk \(\rho(\mu_0)\) in general depends on prior \(\mu_0\), the posterior gamma-minimax regret action \(a^{reg}_x\) differs from the posterior gamma-minimax action \(a_x^{ref}\) obtained in the previous section. To illustrate their difference, recall the conditional posterior gamma-minimax action of Proposition 5.1 with scalar \(\eta\) and the quadratic loss. Let \([\eta(\phi), \bar{\eta}(\phi)]\) be as defined in the above proposition, and let \(m(\phi) = (\eta(\phi) + \bar{\eta}(\phi))/2\) and \(r(\phi) = (\bar{\eta}(\phi) - \eta(\phi))/2 \geq 0\) 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
\[
E_{\phi|x} \left[ \sup_{\eta \in H(\phi)} (\eta - a)^2 \right] = E_{\phi|x} \left[ \max \{ (\eta(\phi) - a)^2, (\bar{\eta}(\phi) - a)^2 \} \right] = E_{\phi|x} \left[ (m(\phi) - a)^2 \right] + E_{\phi|x} \left[ r(\phi) |m(\phi) - a| \right] + E_{\phi|x} \left[ \left( \frac{r(\phi)}{2} \right)^2 \right].
\]
Note that \( a = E_{\phi|X}(m(\phi)) = a_{x}^{reg} \) is determined so as to minimize the first term, but it does not necessarily minimize the second term, and, therefore, \( a_{x} \) can differ from \( a_{x}^{reg} \). A sufficient condition for them to coincide is that \( r(\phi) \) and \( m(\phi) \) are a posteriori independent and the posterior marginal distribution of \( m(\phi) \) is symmetric. The gamma-minimax regret decision with the quadratic loss depends only on the distribution of \( m(\phi) \) the midpoint of \( H(\phi) \), while the gamma-minimax decision depends on the joint distribution of \( m(\phi) \) and \( r(\phi) \). For large sample, this difference disappears and, as shown in the next subsection, \( a_{x}^{reg} \) and \( a_{x} \) converge to the same action (see Appendix B for a discussion on the large sample behavior of the gamma-minimax decisions).

6 Set Estimation of \( \eta \)

Next, we discuss a use of the posterior lower probability of \( \eta \) to conduct a set estimation for \( \eta \). In the standard Bayesian inference, the posterior distribution of the parameter of interest is often summarized by the contour set of the posterior density of \( \eta \) with a prespecified credibility. In this section, we consider what would be an analogue of the posterior credible region if we want to summarize the posterior information of \( \eta \) by the posterior lower probability \( F_{\eta|x}(\cdot|x) \).

6.1 Posterior Lower Credible Region

Consider some set \( C \subset \mathcal{H} \) on which the posterior lower probability \( F_{\eta|x}(\cdot) \) is equal or greater than \( \alpha \),

\[
F_{\eta|x}(C) = F_{\phi|x}(H(\phi) \subset C) \geq \alpha.
\]

In words, \( C \) is interpreted as "the set on which the posterior credibility of \( \eta \) is at least \( \alpha \) 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 \( C \) 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. Among those \( C \), we focus on the volume minimizing posterior lower credible region with credibility \( \alpha \) defined by

\[
C_{\alpha} \equiv \arg \min_{C \subset C} \text{Vol}(C) \quad \text{s.t.} \quad F_{\phi|x}(H(\phi) \subset C) \geq \alpha,
\]
where $\text{Vol}(C)$ is the volume of subset $C$ in terms of the Lebesgue measure and $C$ is a family of subsets in $\mathcal{H}$ over which the volume minimizing credible region is searched.

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

Let $B_r(\eta_c)$ be a closed ball centered at $\eta_c \in \mathcal{H}$ with radius $r$. If $C$ is constrained to be the class of closed balls, the constrained minimization problem of (6.1) becomes

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

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

**Proposition 6.1** Let $\overline{d}_H : \mathcal{H} \times \mathcal{D} \to \mathbb{R}_+$ be

$$
\overline{d}_H(\eta_c, H(\phi)) \equiv \sup_{\eta \in H(\phi)} \{\|\eta - \eta\|\}.
$$

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

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

Then, the solution of the constrained minimization problem (6.2) is given by $(r^*_\alpha, \eta^*_c)$ where

$$
r^*_\alpha = r_\alpha(\eta^*_c) \quad \text{where} \quad \eta^*_c = \arg \min_{\eta_c \in \mathcal{H}} r_\alpha(\eta_c).
$$

**Proof.** See Appendix A. ■

Given random draws of $\phi$ from its posterior, it is straightforward to obtain an approximated volume minimizing credible region by applying the above proposition. Let $\{\phi_s : s = 1, \ldots, S\}$ be random draws of $\phi$ from its posterior. At each $\eta_c \in \mathcal{H}$, we first calculate $\hat{r}_\alpha(\eta_c)$ the empirical $\alpha$-th quantile of $\overline{d}_H(\eta_c, H(\phi))$ based on the simulated $\overline{d}_H(\eta_c, H(\phi_s))$, $s = 1, \ldots, S$. The obtained empirical $\alpha$-th quantile $\hat{r}_\alpha(\eta_c)$ should be a valid approximate for $r_\alpha(\eta_c)$, so we can approximate $(r^*_\alpha, \eta^*_c)$ by finding the minimizer and the minimized value of $\hat{r}_\alpha(\eta_c)$.
6.2 Asymptotic Property of the Posterior Lower Credible Region

In this section, we examine the large sample behavior of the volume minimizing credible region in a rather simple situation where \( \eta \) is a scalar and \( H(\phi) \) is a connected interval for almost all \( \phi \). In particular, we investigate a relationship between the volume minimizing posterior lower credible region and the frequentist confidence sets constructed by a level set of a criterion function (Chernozhukov, Hong, and Tamer (2007)), and characterize sufficient conditions under which they are approximately identical in large sample.

In what follows, we restrict our analysis to the case where the sufficient parameter space \( \Phi \) is finite dimensional (e.g., Example 2.1 and 2.3 in Section 2.1) and assume the posterior of \( \phi \) is consistent in the following sense.

**Definition 6.1 (Posterior consistency of \( \phi \))** Let \( B_\epsilon = \{ \phi : \| \phi - \phi_0 \| < \epsilon \} \) where \( \| \cdot \| \) is some metric in \( \Phi \). The posterior of \( \phi \) is consistent if, for every \( \epsilon > 0 \), \( \lim_{N \to \infty} F_{\phi|x^N}(B_\epsilon|x^N) = 1 \), \( \hat{p}(x^\infty|\phi_0) \)-almost surely.

For finite dimensional \( \phi \), sufficient conditions for the posterior consistency are (i) \( \mu_\phi \) puts a positive probability in the neighborhood of \( \phi_0 \), and (ii) the mode of \( \hat{p}(x^\infty|\phi) \) is consistent to \( \phi_0 \) (see Theorem 7.80 of Schervish (1995)).

Let \( l_N(\phi) \) be the log likelihood of \( \phi \), \( l_N(\phi) = \log \hat{p}(x^N|\phi) \), and \( \hat{\phi}_N \) be the maximum likelihood estimator (MLE) of \( \phi \). Under the regularity conditions (as indicated in Condition 6.1 below), a large sample approximation of the posterior of \( \phi \) is given by the multivariate normal distribution centering at MLE (see, e.g., Schervish (1995, Sec 7.4)). Let \( \hat{J}_N \) be the sample Fisher Information of \( l_N(\phi) \) evaluated at \( \hat{\phi}_N \), and consider the second order Taylor approximation of the log-likelihood and the first order approximation of the prior,

\[
\begin{align*}
l_N(\phi) - l_N(\hat{\phi}_N) &= -\frac{1}{2}(\phi - \hat{\phi}_N)' \hat{J}_N(\phi - \hat{\phi}_N) + o \left( \| \phi - \hat{\phi}_N \|^2 \right), \\
\mu_\phi(\phi) - \mu_\phi(\hat{\phi}_N) &= \mu'_\phi(\hat{\phi}_N)(\phi - \hat{\phi}_N) + o \left( \| \phi - \hat{\phi}_N \| \right),
\end{align*}
\]

where \( o(\cdot) \) specifies the reminder term of the expansion. Consider a local neighborhood of \( \hat{\phi}_N \) shrinking at the rate \( \hat{J}_N^{1/2} \) and let \( s = \hat{J}_N^{1/2}(\phi - \hat{\phi}_N) \propto O(1) \). Then,

\[
\begin{align*}
l_N(\hat{\phi}_N + \hat{J}_N^{-1/2} s) - l_N(\hat{\phi}_N) &= -\frac{1}{2} s' s + o \left( \| \hat{J}_N^{-1/2} s \|^2 \right), \\
\mu_\phi(\hat{\phi}_N + \hat{J}_N^{-1/2} s) - \mu_\phi(\hat{\phi}_N) &= \mu'_\phi(\hat{\phi}_N) \hat{J}_N^{-1/2} s + o \left( \| \hat{J}_N^{-1/2} s \| \right).
\end{align*}
\]
If the likelihood of \( \phi \) is regular, it holds \( \hat{J}_N = O(N) \) and, accordingly, we obtain
\[
\log f_{\phi|X^N}(\phi|x^N) - \log f_{\phi|X^N}(\phi_N|x^N) = -\frac{1}{2}s's + O(N^{-1/2})
\]
\[
\iff f_{\phi|X^N}(\phi|x^N) \propto \exp\left\{-\frac{1}{2}(\phi - \hat{\phi}_N)'\hat{J}_N(\phi - \hat{\phi}_N) + O(N^{-1/2})\right\}. \tag{6.3}
\]

This implies that the posterior of \( \phi \) in a local neighborhood of \( \hat{\phi}_N \) is approximated by the multivariate normal with mean \( \hat{\phi}_N \) and variance \( \hat{J}_N^{-1} \).

The asymptotic analysis of this section assumes the following conditions.

**Condition 6.1**

(i) \( H(\phi) \) is a nonempty, closed, and connected for almost all \( \phi \), i.e., \( H(\phi) = [\eta(\phi), \overline{\eta}(\phi)] \) with \( \overline{\eta}(\phi) \geq \eta(\phi) \).

(ii) \( \phi_0 \) lies in the interior of \( \Phi \) and \( r(\phi_0) \equiv (\overline{\eta}(\phi) - \eta(\phi))/2 > \epsilon > 0 \).

(iii) \( p(x^N|\phi) \) is twice continuously differentiable and \( N^{-1}\hat{J}_N \) converges to a full rank matrix, \( \hat{p}(x^\infty|\phi_0) \)-almost surely.

(iv) \( \mu_\phi(\phi) \) is bounded, continuously differentiable, and the posterior of \( \phi \) is consistent in the sense of Definition 6.1.

(v) The middle point and the radius of \( H(\phi) \) denoted by \( m(\phi) = (\overline{\eta}(\phi) + \eta(\phi))/2 \) and \( r(\phi) = (\overline{\eta}(\phi) - \eta(\phi))/2 \) are continuously differentiable in \( \phi \) and their derivatives are nonzero at \( \phi_0 \).

Condition (ii) through (iv) are standard in the regular likelihood model, but in the context of the set-identified model, a special attention should be paid to condition (ii). The positive radius condition \( r(\phi_0) > \epsilon > 0 \) that is often implied from \( \phi_0 \in \text{Int}(\Phi) \) (e.g., Example 2.1 in Section 2.2) precludes the case of \( H(\phi_0) \) being a singleton. Some literatures of frequentist inference for partially identified parameters has concerned the case of \( r(\phi_0) = 0 \) and they aim at constructing the confidence intervals for \( \eta \) that have uniformly (over \( \phi \)) valid asymptotic coverage probabilities regardless of \( r(\phi_0) = 0 \) or \( r(\phi_0) > 0 \) (Imbens and Manski (2004) and Stoye (2009)). Although this is a potentially important issue to link inference for the point-identified and set-identified model, our analysis does not consider such uniformity issue for the frequentist coverage property of our set estimator. Condition (v) is imposed in order to approximate the posterior of \( \left(m(\phi) - m(\hat{\phi}_N), r(\phi) - r(\hat{\phi}_N)\right)' \) by a bivariate normal distribution.

Note that, as illustrated in Example 1 of Chernozhukov, Hong, and Tamer (2007), we introduce a sample criterion function \( Q_N(\eta) \) that measures a distance from \( \eta \) to the nearest
point within the sample identified set, $[\eta(\hat{\phi}_N), \bar{\eta}(\hat{\phi}_N)]$. In terms of $m(\cdot)$ and $r(\cdot)$, the criterion function is written as

$$Q_N(\eta) = \left[ \eta - m(\hat{\phi}_N) - r(\hat{\phi}_N) \right]_+$$

where $[x]_+ = \max\{x, 0\}$. In the criterion function approach, a confidence region for the true identified set $H(\phi_0)$ with coverage $\alpha$ is given by a level set,

$$C(c_\alpha) \equiv \left\{ \eta \in \mathcal{H} : \sqrt{N}Q_N(\eta) \leq c_\alpha \right\},$$

where the cutoff level $c_\alpha$ is given by the $\alpha$-th quantile of the sampling distribution of $\sup_{\eta \in H(\phi_0)} \sqrt{N}Q_N(\eta)$.

The next proposition provides the condition that, for large $N$, the posterior lower credible region $C_{\alpha^*}$ of Proposition 6.1 coincides with the level set confidence set $C(c_\alpha)$ up to the first order approximation error.

**Proposition 6.2** Assume Condition 6.1, and let $(M, R) \sim \mathcal{N}(0, \Sigma)$ be bivariate normal random variables that almost surely approximates the posterior probability law of $m(\phi)$ and $r(\phi)$, i.e., for each $W \subset \mathcal{R}^2$

$$\lim_{N \to \infty} F_{\phi|X} \left( \sqrt{N} \left( \begin{array}{c} m(\phi) - m(\hat{\phi}_N) \\ r(\phi) - r(\hat{\phi}_N) \end{array} \right) \in W \right) = \Phi_2 \left( \Sigma^{-1/2} \left( \begin{array}{c} M \\ R \end{array} \right) \in W \right)$$

where $\Phi_2$ is the distribution function of the standard bivariate Gaussian. Let $\frac{1}{2} < \alpha < 1$.

(i) If $M$ and $R$ are independent, then the difference between $C_{\alpha^*}$ and $C(c_\alpha)$ vanishes at the order of $o(N^{-1/2})$, i.e., $\text{Leb}(C_{\alpha^*} \Delta C(c_\alpha)) = o(N^{-1/2})$, $p(x^{\infty}|\phi_0)$-almost surely.

(ii) If $M$ and $R$ are correlated, the width of $C_{\alpha^*}$ is shorter than the width of $C(c_\alpha)$ and their difference is at the order of $O(N^{-1/2})$, i.e., $\sqrt{N} (\text{Leb}(C(c_\alpha)) - \text{Leb}(C_{\alpha^*}))$ is positive and bounded for large $N$, $p(x^{\infty}|\phi_0)$-almost surely.

**Proof.** See Appendix A. ■

This proposition solely compares the width of the two set estimators with two different notions of confidence. The first result of the proposition provides a condition that the two

---

*Chernozhukov, Hong, and Tamer (2007) demonstrated validity of subsampling to estimate the critical value $c_\alpha$.\footnote{Chernozhukov, Hong, and Tamer (2007) demonstrated validity of subsampling to estimate the critical value $c_\alpha$.}
set estimators for $\eta$ coincide. The condition of $M$ and $R$ being independent is equivalent to that the lower and upper bounds of $H(\phi)$ shares the same posterior variance, $\text{Var}(\bar{\eta}(\phi)|x^N) = \text{Var}(\tilde{\eta}(\phi)|x^N)$. In this case, the criterion function based confidence region with the criterion given in (6.4) leads to the first order approximation of the posterior lower credible region. Otherwise, the criterion function based confidence region is wider by $O(N^{-1/2})$ than the posterior lower credible region.

7 Conclusion

This paper proposed a 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 introduced the class of prior distributions that contains any unrevisable prior knowledge, and summarized the posterior uncertainty for the model parameters in terms of the posterior lower and upper probabilities of the resulting posterior class. We showed that the posterior lower and upper probabilities can be interpreted as a posterior probability law of the identified set (Theorem 4.1). This robust-Bayes way of generating the random identified sets is novel in the literature of the partially identified model, and it provides a seamless link between the random set theory and the likelihood based inference for the set identified model (c.f. Beresteanu and Molinari (2009)).

We employed the gamma-minimax criterion to develop a conservative point decision rule that is insensitive to the unrevisable prior knowledge. We showed that, with our choice of prior class, it is feasible to obtain the gamma-minimax decision for the general class of set identified model. The objective function of the gamma-minimax criterion integrates both ambiguity for the set-identified parameters and the posterior uncertainty for the identified components in the model into the single objective function. Therefore, the sample size affects the optimal decision via the reduction of the posterior uncertainty for the identified parameters. In contrast to the standard gamma-minimax decision rule, a non-standard finding in our analysis is that the gamma-minimax optimal action is invariant no matter whether the optimal decision is calculated before and after observing data (Proposition 5.2).

The posterior lower probability is a nonadditive measure so that we cannot plot it as we do for the posterior probability densities. As a way to summarize it, we developed the way to compute the posterior lower credible region as an analogue to the posterior contour set in the standard Bayes procedure. We showed that for the case where the parameter of interest is a scalar, an algorithm to compute the posterior lower credible region is available and easy to implement. We also compared the large sample behavior of the volume minimizing posterior lower credible region with the frequentist confidence regions for the identified set
in the criterion function approach, and presented the sufficient condition under which these two confidence regions coincide.

In this article, we precluded observationally restrictive models and assumed throughout that the identified set is nonempty. If we would like to allow for the empty identified set, we should modify the definition of lower and upper probabilities to be the conditional lower and upper probabilities given the random identified set is nonempty.

Also, we left an analysis of the intersected identified set (Manski (2003)) out of scope of this paper, primarily because it is not clear what is a robust Bayes interpretation or justification for taking the intersection of multiple identified sets. Research on these topics is in progress.

**Appendix**

A Lemma and Proofs

Our proof of Theorem 4.2 uses the following two lemma. The first lemma says that, for a fixed subset $A \in \mathcal{A}$ and every $\mu_\theta \in M(\mu_\phi)$, the conditional probability $\mu_{\theta|\phi}(A|\phi)$ can be bounded above and below by the two indicator functions $1_{\{\Gamma(\phi) \cap A \neq \emptyset\}}(\phi)$ and $1_{\{\Gamma(\phi) \subset A\}}(\phi)$, respectively. The second lemma states that for each fixed subset $A \in \mathcal{A}$, we can construct the probability measures on the parameter space $(\Theta, \mathcal{A})$, $\mu_\theta^*$ and $\mu_{\theta*}$ that belong to the prior class $M(\mu_\phi)$ and achieve the upper bound and lower bound of the conditional probabilities obtained in the first lemma. As a corollary of these two lemma, we obtain the theorem.

**Lemma A.1** Assume Condition 4.1 (i). For each $A \in \mathcal{A}$,

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

holds $\mu_\phi$-almost surely for every $\mu_\theta \in M(\mu_\phi)$.

**Proof of Lemma A.1.** Fix $A \in \mathcal{A}$. For notational brevity, let $\Phi_1^A = \{\phi : \Gamma(\phi) \subset A\}$ and $\Phi_0^A = \{\phi : \Gamma(\phi) \cap A = \emptyset\}$. To prove the claim, it suffices to show

$$\int_B 1_{\Phi_1^A}(\phi) d\mu_\phi \leq \int_B \mu_{\theta|\phi}(A|\phi) d\mu_\phi \leq \int_B 1_{\Phi_0^A}(\phi) d\mu_\phi \quad \text{for every } \mu_\theta \in M(\mu_\phi) \text{ and } B \in \mathcal{B},$$

(A.1)
where \((\Phi_0^A)^c = \{ \phi : \Gamma(\phi) \cap A \neq \emptyset \}\). Let \(B \in \mathcal{B} \) and \(\mu_\theta \in \mathcal{M}(\mu_\phi)\). Consider

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

where the equality follows by the definition of conditional distribution. By the construction of \(\Phi_1^A, \Gamma(B \cap \Phi_1^A) \subset A\), and therefore,

\[
\mu_\theta(A \cap \Gamma(B \cap \Phi_1^A)) = \mu_\theta(\Gamma(B \cap \Phi_1^A)) = \mu_\phi(B \cap \Phi_1^A) = \int_B 1_{\Phi_1^A}(\phi)d\mu_\phi.
\]

Hence, the first inequality in (A.1) holds.

To show the other inequality, consider

\[
\int_B \mu_{\theta|\phi}(A|\phi)d\mu_\phi = \int_{B\cap \Phi_0^A} \mu_{\theta|\phi}(A|\phi)d\mu_\phi + \int_{B\cap (\Phi_0^A)^c} \mu_{\theta|\phi}(A|\phi)d\mu_\phi.
\]

By the construction of \(\Phi_0^A, A \cap \Gamma(B \cap \Phi_0^A) = \emptyset\), \(\int_{B\cap \Phi_0^A} \mu_{\theta|\phi}(A|\phi)d\mu_\phi = \mu_\theta(A \cap \Gamma(B \cap \Phi_0^A)) = 0\) implies

\[
\int_B \mu_{\theta|\phi}(A|\phi)d\mu_\phi = \int_{B\cap (\Phi_0^A)^c} \mu_{\theta|\phi}(A|\phi)d\mu_\phi = \int_B 1_{(\Phi_0^A)^c}(\phi)\mu_{\theta|\phi}(A|\phi)d\mu_\phi \leq \int_B 1_{(\Phi_0^A)^c}(\phi)d\mu_\phi
\]

where the inequality follows because \(\mu_{\theta|\phi}(A|\phi) \leq 1, \mu_\phi\)-almost surely. Thus, the second inequality in (A.1) is obtained, this completes the proof. 

**Lemma A.2** Assume Condition 4.1 (i). For each \(A \in \mathcal{A}\), there exist \(\mu_{\theta*} \in \mathcal{M}(\mu_\phi)\) and \(\mu_0^* \in \mathcal{M}(\mu_\phi)\) that achieve the lower and upper bounds of \(\mu_{\theta|\phi}(A|\phi)\) obtained in Lemma A.1.

**Proof of Lemma A.2.** Fix \(A \in \mathcal{A}\) and let

\[
\Phi_0^A = \{ \phi : \Gamma(\phi) \cap A = \emptyset \},
\]

\[
\Phi_1^A = \{ \phi : \Gamma(\phi) \subset A \},
\]

\[
\Phi_2^A = \{ \phi : \Gamma(\phi) \cap A \neq \emptyset, \Gamma(\phi) \cap A^c \neq \emptyset \},
\]

where each of them belongs to the sufficient parameter \(\sigma\)-algebra \(\mathcal{B}\) by Condition 4.1 (i). Note that \(\Phi_0^A, \Phi_1^A, \) and \(\Phi_2^A\) are mutually disjoint and constitute a measurable partition of \(\Phi\).
since $\Phi_2^A = \Phi \setminus (\Phi_0^A \cup \Phi_1^A)$. By the construction of the prior class $\mathcal{M}(\mu_\phi)$, $\mu_\theta \in \mathcal{M}(\mu_\phi)$ if and only if $\mu_\theta$ is a probability measure on $(\Theta, \mathcal{A})$ satisfying
\[
\mu_\phi(B) = \mu_\theta(\Gamma(B)) \quad \text{for all } B \in \mathcal{B}. \tag{A.2}
\]
Consider a probability measure $\mu_{\theta^*}$ on $(\Theta, \mathcal{A})$ that satisfies (A.2) and
\[
\mu_\phi(\Phi_2^A) = \mu_{\theta^*}(\Gamma(\Phi_2^A)) = \mu_{\theta^*}(\Gamma(\Phi_2^A) \cap A^c),
\]
whose existence is guaranteed since $\Gamma(\Phi_2^A) \cap A^c$ is never be empty by the construction of $\Phi_2^A$ as long as $\Phi_2^A$ is nonempty. This construction of $\mu_{\theta^*}$ states that the probability $\mu_\phi(\Phi_2^A)$ that $\mu_{\theta^*}$ puts on $\Gamma(\Phi_2^A)$ concentrates on its subset $\Gamma(\Phi_2^A) \cap A^c$. For an arbitrary $B \in \mathcal{B}$, consider
\[
\mu_{\theta^*}(A \cap \Gamma(B)) = \mu_{\theta^*}(A \cap \Gamma(B) \cap \Gamma(\Phi_0^A)) + \mu_{\theta^*}(A \cap \Gamma(B) \cap \Gamma(\Phi_1^A)) + \mu_{\theta^*}(A \cap \Gamma(B) \cap \Gamma(\Phi_2^A)).
\]
Recall that, by the construction of $\Phi_j^A$, $j = 0, 1, 2$, and $\mu_{\theta^*}$, $A \cap \Gamma(\Phi_0^A) = \emptyset$, $A \cap \Gamma(B) \cap \Gamma(\Phi_1^A) = \Gamma(B) \cap \Gamma(\Phi_1^A)$, and $\mu_{\theta^*}(A \cap \Gamma(B) \cap \Gamma(\Phi_2^A)) \leq \mu_{\theta^*}(A \cap \Gamma(\Phi_2^A)) = 0$. Hence,
\[
\mu_{\theta^*}(A \cap \Gamma(B)) = \mu_{\theta^*}(\Gamma(B) \cap \Gamma(\Phi_1^A)) = \int_B 1_{\Phi_1^A}(\phi) d\mu_\phi.
\]
Since $B \in \mathcal{B}$ is arbitrary, this implies that $\mu_{\theta^*}(A|\phi) = 1_{\Phi_1^A}(\phi)$, $\mu_\phi$-almost surely. Thus, $\mu_{\theta^*}$ achieves the lower bound obtained in Lemma A.1.

Following to the above argument, consider a probability measure $\mu_\theta^*$ on $(\Theta, \mathcal{A})$ that satisfies (A.2) and
\[
\mu_\phi(\Phi_2^A) = \mu_\theta^*(\Gamma(\Phi_2^A)) = \mu_\theta^*(\Gamma(\Phi_2^A) \cap A).
\]
Now, in contrast to $\mu_{\theta^*}$, the probability $\mu_\phi(\Phi_2^A)$ that $\mu_\theta^*$ puts on $\Gamma(\Phi_2^A)$ concentrates on its subset $\Gamma(\Phi_2^A) \cap A$. Note also that we can always construct such $\mu_\theta^*$ since by the definition of $\Phi_2^A$, $\Gamma(\Phi_2^A) \cap A$ is nonempty whenever $\Phi_2^A$ is nonempty.

For an arbitrary $B \in \mathcal{B}$, consider,
\[
\mu_\theta^*(A \cap \Gamma(B)) = \mu_\theta^*(A \cap \Gamma(B) \cap \Gamma(\Phi_0^A)) + \mu_\theta^*(A \cap \Gamma(B) \cap \Gamma(\Phi_1^A)) + \mu_\theta^*(A \cap \Gamma(B) \cap \Gamma(\Phi_2^A)) = \\
\mu_\theta^*(\Gamma(B) \cap \Gamma(\Phi_1^A)) + \mu_\theta^*(\Gamma(B) \cap \Gamma(\Phi_2^A)) = \\
\mu_\theta^*(\Gamma(B) \cap (\Gamma(\Phi_1^A) \cup \Gamma(\Phi_2^A))),
\]
where the second equality follows since $A \cap \Gamma(B) \cap \Gamma(\Phi_0^A) = \emptyset$, $A \cap \Gamma(B) \cap \Gamma(\Phi_1^A) = \Gamma(B) \cap \Gamma(\Phi_1^A)$, and $\mu_\theta^*(A \cap \Gamma(B) \cap \Gamma(\Phi_2^A)) = \mu_\theta^*(\Gamma(B) \cap \Gamma(\Phi_2^A))$. The third equality holds since $\Phi_1^A$
and $\Phi^A_1$ are disjoint. The above equality and $\Phi^A_1 \cup \Phi^A_2 = (\Phi^A_0)^C$ imply $\mu^*_\theta(A|\phi) = 1_{(\Phi^A_0)^C}(\phi)$, $\mu_\phi$-almost surely. Hence, $\mu^*_\theta$ attains the lower bound obtained in Lemma A.1. ■

**Proof of Theorem 4.2 (i).** Under the given assumptions, the posterior of $\theta$ 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 the monotonicity of Lebesgue integral, $F_{\theta|X}(A)$ is minimized and maximized by plugging the lower bound and upper bounds of $\mu_{\theta|\phi}(A|\phi)$ into the integrand. From Lemma A.1 and Lemma A.2, they are given by $1_{\{\Gamma(\phi) \subseteq A\}}(\phi)$ and $1_{\{\Gamma(\phi) \cap A \neq \emptyset\}}(\phi)$. Hence, the conclusion follows. ■

**Proof of Theorem 4.2 (ii).** For each $\mu_\theta \in \mathcal{M}(\mu_\phi)$ and $A \in \mathcal{A}$, $F_{\theta|X}(A) \leq F^*_{\theta|X}(A)$ holds 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 $\{ F_{\theta|X} : \mu_\theta \in \mathcal{M}(\mu_\phi) \}$.

To show the converse, recall Theorem 4.2 (i) and Condition 4.1 (ii) implying that the lower and upper probabilities are containment and capacity functional of the random closed set $\Gamma(\phi)$. Consequently, by applying the Selectionability Theorem of the random set (Molchanov (2005), Theorem 1.2.20), we can ascertain that for each $G_\theta \in \mathcal{G}_\theta$, there exists a $\Theta$-valued random variable $\xi(\phi)$, so called a selection of $\Gamma(\phi)$, such that $\xi(\phi) \in \Gamma(\phi)$ and $G_\theta(A) = F_{\phi|X}(\xi(\phi) \in A)$, $A \in \mathcal{A}$. With such selection $\xi(\phi)$, we consider a probability measure on $(\Theta, \mathcal{A})$, $\tilde{\mu}_\theta$, defined by

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

Note such $\tilde{\mu}_\theta$ belongs to $\mathcal{M}(\mu_\phi)$ since for each subset $B \in \mathcal{B}$ in the sufficient parameter space,

$$\tilde{\mu}_\theta(\Gamma(B)) = \mu_\phi(\{ \phi : \xi(\phi) \in \Gamma(B) \}) = \mu_\phi(B)$$

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

Since the conditional distribution for $\tilde{\mu}_\theta(A)$ given $\phi$ is $\tilde{\mu}_{\theta|\phi}(A|\phi) = 1_{\{\xi(\phi) \in A\}}(\phi)$, the posterior of $\theta$ generated from $\tilde{\mu}_\theta$ is, from (3.1),

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

32
Thus, we have shown that, for each $G_{\theta} \in \mathcal{G}_{\theta}$, there exists a prior $\tilde{\mu}_{\theta} \in \mathcal{M}(\mu_{\theta})$ with which the posterior coincides with the $G_{\theta}$. Hence, $\mathcal{G}_{\theta} \subset \{ F_{\theta|X}: \mu_{\theta} \in \mathcal{M}(\mu_{\theta}) \}$. ■

**Proof of Proposition 5.1.** Let $\mathcal{G}_{\theta}$ be the class of probability measures on $(\Theta, A)$ as considered in the proof of Theorem 4.2 (ii),

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

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 5.1 (ii)), then for any nonnegative measurable function $k: \Theta \rightarrow \mathcal{R}_{+}$

$$ \int k(\theta)dF_{\theta|X}^{*} = \sup_{G_{\theta} \in \mathcal{G}_{\theta}} \left\{ \int_{\Theta} k(\theta)dG_{\theta} \right\} \quad (A.3) $$

holds. Since $F_{\theta|X}^{*}(\cdot)$ is the capacity functional of the random closed set $\Gamma(\phi)$, $F_{\theta|X}^{*}(\cdot)$ is a subadditive capacity of infinite alternating order by the Choquet Theorem (see, e.g., Molchanov (2005, Sec. 1.1.2-1.1.3)). Furthermore, Theorem 4.1 (ii) implies that $\sup_{G_{\theta} \in \mathcal{G}_{\theta}} \{ \int_{\Theta} k(\theta)dG_{\theta} \}$ is equivalently written as $\sup_{\mu_{\theta} \in \mathcal{M}(\mu_{\theta})} \{ \int_{\Theta} k(\theta)dF_{\theta|X} \}$. Hence, setting $k(\theta) = L(h(\theta), a)$ in (A.3) leads to

$$ \int L(h(\theta), a)dF_{\theta|X}^{*} = \sup_{\mu_{\theta} \in \mathcal{M}(\mu_{\theta})} \left\{ \int_{\Theta} L(h(\theta), a)dF_{\theta|X} \right\} = \rho(\mu_{\theta}, a). \quad (A.4) $$

On the other hand, by the definition of Choquet integral and (??),

$$ \int L(\eta, a)dF_{\eta|X}^{*} = \int F_{\eta|X}^{*}(\{ \eta: L(\eta, a) \geq t \}) dt = \int F_{\eta|X}(\{ \theta: L(h(\theta), a) \}) dt = \int L(h(\theta), a)dF_{\theta|X}^{*}. \quad (A.5) $$

Combining (A.4) and (A.5) yields the first equality of the proposition.

Next, we show the second equality of the proposition. By the definition of the Choquet integral and Theorem 4.1 (iii),

$$ \int L(\eta, a)dF_{\eta|X}^{*}(\eta) = \int_{0}^{\infty} F_{\eta|X}^{*}(\{ \eta: L(\eta, a) \geq t \}) dt = \int_{0}^{\infty} F_{\eta|X}(\{ \phi: \{ \eta: L(\eta, a) \geq t \} \cap \phi(\phi) \neq \emptyset \}) dt. $$
Note that \( \{ \eta : L(\eta, a) \geq t \} \cap H(\phi) \neq \emptyset \) is true if and only if \( \{ \sup_{\eta \in H(\phi)} \{ L(\eta, a) \} \geq t \} \). Hence,

\[
\int L(\eta, a) dF^\eta_{|X}(\eta) = \int_0^\infty F^\eta_{|X} \left( \left\{ \phi : \sup_{\theta \in \Gamma(\phi)} \{ L(\theta, a) \} \geq t \right\} \right) d\eta.
\]

By noting that \( \sup_{\eta \in \Gamma(\phi)} \{ L(\eta, \delta) \} \) is nonnegative by Condition 5.1 (i), the Fubini's theorem yields

\[
\int L(\eta, a) dF^\eta_{|X}(\eta|x) = \int_0^\infty \int_\Phi 1_{\{ \sup_{\eta \in H(\phi)} \{ L(\eta, a) \} \geq t \}}(\phi) dF^\eta_{|X} x dt = \int_\Phi 1_{\{ \sup_{\eta \in H(\phi)} \{ L(\eta, a) \} \geq t \}}(x) dtdF^\eta_{|X} x = \int_\Phi \sup_{\eta \in H(\phi)} \{ L(\eta, a) \} dF^\eta_{|X}. \tag{A.6}
\]

**Proof of Proposition 5.3.**

(i) Given that the posterior variance of \( \eta \) is finite for \( \mu_\theta \in M(\mu_\phi) \), the posterior gamma-minimax regret is well defined and is obtained as

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

where \( \hat{\eta}_{\mu_\theta} \) is the posterior mean of \( \eta \) resulting from prior \( \mu_\theta \). Consider the bounds of \( \hat{\eta}_{\mu_\theta} \) when \( \mu_\theta \) varies over \( M(\mu_\phi) \),

\[
\begin{align*}
\underline{\eta}_x &= \inf_{\mu_\theta \in M(\mu_\phi)} \hat{\eta}_{\mu_\theta} = \inf_{\mu_\theta \in M(\mu_\phi)} \int \phi \phi(\theta) dF^\eta_{|X}, \\
\bar{\eta}_x &= \sup_{\mu_\theta \in M(\mu_\phi)} \hat{\eta}_{\mu_\theta} = \sup_{\mu_\theta \in M(\mu_\phi)} \int \phi \phi(\theta) dF^\eta_{|X}.
\end{align*}
\]

By the same argument used in obtaining (A.4), (A.5), and (A.6), \( \underline{\eta}_x \) and \( \bar{\eta}_x \) thus defined are equal to

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

which are assumed to be finite by the finite posterior variance assumption. Therefore, for a fixed \( a \in H_\alpha \), the posterior gamma-minimax regret is written as

\[
\sup_{\mu_\theta \in M(\mu_\phi)} \{ \rho(\mu_\theta, a) - \hat{\rho}(\mu_\theta) \} = \sup_{\mu_\theta \in M(\mu_\phi)} E_{|X}[ (a - \hat{\eta}_{\mu_\theta})^2 ]
\]

\[
= \begin{cases} 
(\bar{\eta}_x - a)^2 & \text{for } a \leq \frac{\underline{\eta}_x + \bar{\eta}_x}{2}, \\
(\underline{\eta}_x - a)^2 & \text{for } a > \frac{\underline{\eta}_x + \bar{\eta}_x}{2}.
\end{cases}
\]
This is clearly minimized at $\frac{\eta + \hat{\eta}}{2}$.

(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 $\eta$ with respect to the marginal distribution of data,

$$\underline{r}(\mu_\theta) = \inf_{\delta} \int_x \rho(\mu_\theta, \delta(x))m(x|\mu_\phi)dx$$

$$= \int_x \rho(\mu_\theta)m(x|\mu_\phi)dx$$

where $m(x|\mu_\phi)$ is the marginal distribution of data, $\int_\Theta \hat{p}(x|\phi)d\mu_\phi$. Therefore, the unconditional regret is written as

$$r(\mu_\theta, \delta) - \underline{r}(\mu_\theta) = \int_x \left[ \int_{\mathcal{H}} (\eta - \delta(x))^2 dF_{\eta|x} \right] m(x|\mu_\phi)dx - \int_x \rho(\mu_\theta)m(x|\mu_\phi)dx$$

$$= \int_x \left[ \rho(\mu_\theta, \delta(x)) - \underline{\rho}(\mu_\theta) \right] m(x|\mu_\phi)dx.$$ 

Since the marginal distribution of data does not depend on $\mu_\theta$ as long as $\mu_\phi$ is fixed, the unconditional gamma-minimax regret becomes

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

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

**Proof of Proposition 6.1.** Let $\eta_c$ be fixed and $B_r(\eta_c)$ be a closed ball centered at $\eta_c$ with radius $r$. The event $\{H(\phi) \subset B_r(\eta_c)\}$ happens if and only if $\left\{ \frac{d}{d_H}(\eta_c, H(\phi)) \leq r \right\}$. So, $r_\alpha(\eta_c) \equiv \inf \left\{ r : F_{\phi|x} \left( \left\{ \phi : \frac{d}{d_H}(\eta_c, H(\phi)) \leq r \right\} \right) \geq \alpha \right\}$ is the radius of the smallest closed ball centered at $\eta_c$ that contains the random set $H(\phi)$ with posterior lower probability $\alpha$. Therefore, finding the minimizer of $r_\alpha(\eta_c)$ over $\eta_c$ is equivalent to searching for the center of the smallest ball that contains $H(\phi)$ with posterior probability $\alpha$ and the minimum of $r_\alpha(\eta_c)$ corresponds to its radius. ■

**Proof of Proposition 6.2.** First, we derive the first-order approximation of the posterior lower credible region under the assumption that $M$ and $R$ are independent. Consider writing $\frac{d}{d_H}(\eta, H(\phi))$ in terms of $m(\cdot)$ and $r(\cdot)$,

$$\frac{d}{d_H}(\eta, H(\phi)) = |\eta - m(\phi)| + r(\phi)$$

$$= |\eta - (m(\phi) - m(\hat{\phi}_N)) - m(\hat{\phi}_N)| + \left[ r(\phi) - r(\hat{\phi}_N) \right] + r(\hat{\phi}_N)$$

$$= |\eta - N^{-1/2}M - m(\hat{\phi}_N)| + N^{-1/2}R + r(\hat{\phi}_N) + o(N^{-1/2}).$$
where the last expression implies the posterior distribution of \( \hat{d}_H(\eta, H(\phi)) \) is approximated by the posterior distribution of \( |\eta - N^{-1/2}M - m(\hat{\phi}_N)| + N^{-1/2}R + r(\hat{\phi}_N) \). Note that the posterior lower credible region has its center at the value of \( \eta \) that minimizes \( \alpha \)-th quantile of the posterior distribution of \( \hat{d}_H(\eta, H(\phi)) \) and its radius is given by that \( \alpha \)-th quantile. In order to find such value of \( \eta \), we look for \( \eta^* \) that satisfies
\[
F_{\phi|X}(\hat{d}_H(\eta^*, H(\phi)) \leq d) \geq F_{\phi|X}(\hat{d}_H(\eta, H(\phi)) \leq d) \quad \text{for every } \eta \text{ and } d > 0.
\]
For this goal, consider
\[
F_{\phi|X}(\hat{d}_H(\eta, H(\phi)) \leq d) = \Pr \left( \left| \eta - N^{-1/2}M - m(\hat{\phi}_N) \right| + N^{-1/2}R + r(\hat{\phi}_N) \leq d \right) + o(N^{-1/2})
\]
\[
= \int_{-\infty}^{\infty} \Pr \left( \left| \eta - N^{-1/2}M - m(\hat{\phi}_N) \right| \leq d - N^{-1/2}R - r(\hat{\phi}_N) \right) f_R(r)dr + o(N^{-1/2})
\]
where \( f_R \) is the marginal density of \( R \). Note that the last line of the above equations follows since \( R \) and \( M \) are assumed to be independent. Using the formula of the cumulative distribution function of \(|Z|\) with \( Z \sim N(\mu, \sigma^2)\),
\[
\Pr(|Z| \leq y) = \int_{\frac{y+\mu}{\sigma}}^{\infty} \psi(z)dz + \int_{-\infty}^{\frac{y-\mu}{\sigma}} \psi(z)dz,
\]
where \( \psi(\cdot) \) be the probability density function of univariate standard normal distribution, the first order condition of \( F_{\phi|X}(\hat{d}_H(\eta, H(\phi)) \leq d) \) with respect to \( \eta \) is written as
\[
\frac{\partial}{\partial \eta} F_{\phi|X}(\hat{d}_H(\eta, H(\phi)) \leq d) = \int_{-\infty}^{N^{1/2}(d-r(\hat{\phi}_N))} \left[ \psi \left( \frac{d - N^{-1/2}r - r(\hat{\phi}_N) + \left( \eta - m(\hat{\phi}_N) \right)}{\sigma_M/N^{1/2}} \right) - \psi \left( \frac{d - N^{-1/2}r - r(\hat{\phi}_N) - \left( \eta - m(\hat{\phi}_N) \right)}{\sigma_M/N^{1/2}} \right) \right] f_R(r)dr.
\]
where \( \sigma_M^2 = Var(M) \). From this first condition, we can claim that \( F_{\phi|X}(\hat{d}_H(\eta, H(\phi)) \leq d) \) is maximized at \( \eta = m(\hat{\phi}_N) \) for any \( d \), implying that the \( \alpha \)-quantile of the posterior distribution of \( \hat{d}_H(\eta, H(\phi)) \) is minimized at \( \eta = m(\hat{\phi}_N) \), and the minimized quantile is obtained as the \( \alpha \)-th quantile of \( |N^{-1/2}M| + N^{-1/2}R + r(\hat{\phi}_N) \).
On the other hand, if we implement the criterion function approach, we would focus on approximating the sampling distribution of the level set

\[
\sup_{\eta \in H(\phi_0)} Q_N(\eta) = \sup_{\eta \in [\bar{\eta}(\phi_0), \tilde{\eta}(\phi_0)]} \left[ \eta - (m(\hat{\phi}_N) - m(\phi_0)) - m(\phi_0) + [r(\phi_0) - r(\hat{\phi}_N)] - r(\phi_0) \right] + \\
= \sup_{\eta \in [\bar{\eta}(\phi_0), \tilde{\eta}(\phi_0)]} \left[ \eta - N^{-1/2} M - m(\phi_0) + N^{-1/2} R - r(\phi_0) \right] + o_p(N^{-1/2}) \]

where \( M \) and \( R \) approximate the sampling distribution of \( m(\hat{\phi}_N) - m(\phi_0) \) and \( r(\phi_0) - r(\hat{\phi}_N) \).

When \( M \geq 0 \), the supremum is attained at \( \eta = \bar{\eta}(\phi_0) \) and, therefore, \( \sup_{\eta \in H(\phi_0)} Q_N(\eta) = [N^{-1/2} M + N^{-1/2} R]_+ \). Similarly, when \( M < 0 \), the supremum is attained at \( \eta = \tilde{\eta}(\phi_0) \) and \( \sup_{\eta \in H(\phi_0)} Q_N(\eta) = [-N^{-1/2} M + N^{-1/2} R]_+ \). Therefore, for \( \alpha > 1/2 \), \( c_\alpha \) the \( \alpha \)-th quantile of \( \sup_{\eta \in H(\phi_0)} Q_N(\eta) \) is that of \( [N^{-1/2} M + N^{-1/2} R] \). Consequently, \( C_N(c_\alpha) \equiv \{ \eta \in H : Q_N(\eta) \leq c_\alpha \} \) can be seen as an interval centered at \( m(\hat{\phi}_N) \) with radius \( c_\alpha + r(\hat{\phi}_N) \), which coincides with the posterior lower credible region \( C_{\alpha^*} \). This completes the proof of the first statement of the proposition.

In order to show the second statement, consider the case that \( M \) and \( R \) are correlated.

Note that even when \( M \) and \( R \) are correlated, the way to obtain the cut-off value for the level set \( c_\alpha \) remains unchanged and \( c_\alpha \) is given by the \( \alpha \)-th quantile of \( [N^{-1/2} M + N^{-1/2} R] \). In contrast, we will show that the \( \alpha \)-quantile of the posterior distribution of \( \tilde{d}_H(\eta, H(\phi)) \) cannot be minimized at \( \eta = m(\hat{\phi}_N) \) in the presence of the correlation between \( M \) and \( R \), and therefore the duality between \( C_{\alpha^*} \) and \( C_N(c_\alpha) \) obtained in the previous case does not hold.

Let \( R|M \sim \mathcal{N}(\beta M, \sigma_{R|M}) \) be the conditional distribution of \( R \) given \( M \), and \( f_M \) be the marginal distribution of \( M \). Then, \( F_{\phi|X}(\tilde{d}_H(\eta, H(\phi)) \leq d) \) is written as

\[
\frac{\sqrt{N}}{\sigma_{R|M}} \left( \frac{d - r(\hat{\phi}_N)}{\sigma_{R|M}} \right) f_M(m) dm = \int_{-\infty}^{\infty} \Psi \left( \frac{\sqrt{N} \left[ (d - r(\hat{\phi}_N)) - \left( \eta - N^{-1/2} M - m(\hat{\phi}_N) \right) \right] - \beta m}{\sigma_{R|M}} \right) f_M(m) dm \\
+ \int_{-\infty}^{\infty} \Psi \left( \frac{\sqrt{N} \left[ (d - r(\hat{\phi}_N)) + \left( \eta - m(\hat{\phi}_N) \right) \right] - (1 + \beta) m}{\sigma_{R|M}} \right) f_M(m) dm,
\]

37
where $\Psi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The first derivative of $F_{\phi|X}(\bar{d}_H(\eta, H(\phi)) \leq d)$ with respect to $\eta$ evaluated at $\eta = m(\hat{\phi}_N)$ is

$$
\frac{\partial}{\partial \eta} F_{\phi|X}(\bar{d}_H(\eta, H(\phi)) \leq d) \bigg|_{\eta=m(\hat{\phi}_N)} = \sqrt{N} \sigma_{R|M} m(\hat{x}_{1|0}) f_M(m) dm.
$$

This indicates that the difference inside the brackets are nonzero $f_M$-almost surely if and only if $\beta = 0$ ($M$ and $R$ are independent). Therefore, $\frac{\partial}{\partial \eta} F_{\phi|X}(\bar{d}_H(\eta, H(\phi)) \leq d) \bigg|_{\eta=m(\hat{\phi}_N)} \neq 0$ for every $d > 0$ and $\eta = m(\hat{\phi}_N)$ does not minimize the quantile of the posterior distribution of $\bar{d}_H(\eta, H(\phi))$. Note that the radius of the criterion function based confidence region is equal to the $\alpha$-th quantile of $[N^{-1/2}M + N^{-1/2}R]$ plus $r(\hat{\phi}_N)$, which is equal to the $\alpha$-th quantile of $\bar{d}_H(m(\hat{\phi}_N), H(\phi))$, and it implies that the width of volume minimizing posterior lower credible region has a smaller radius. Thus, the conclusion follows.

**B Convergence of the Gamma-minimax Decisions**

In this appendix, the asymptotic behavior of the gamma-minimax decision rules analyzed in Section 5 is considered. Let $X^N \in (X^N, X^N)$ be the size $N$ observations generated from $\hat{p}(x^N|\phi_0)$ where $(X^N, X^N)$ is the product space of $N$ copies of $(X, X)$. We denote the infinite product space of $(X, X)$ by $(X^\infty, X^\infty)$ and a realization of $X^\infty$ by $x^\infty$.

The posterior consistency of $\phi$ in the sense of Definition 6.1 implies $F_{\phi|X^N}$ weakly converges to the point mass measure at $\phi_0$. Therefore, with additional regularity conditions, the posterior gamma-minimax regret criterion (the third expression of (5.3)) converges to $\sup_{\eta \in H(\phi_0)} L(\eta, a)$ so that the posterior gamma-minimax action should converge to $a^*_{x^\infty} \equiv \arg \inf_{a \in H_a} \left\{ \sup_{\eta \in H(\phi_0)} L(\eta, a) \right\}$, $p(x^\infty|\phi_0)$-almost surely.

**Proposition B.1** Assume the conditions of Proposition 5.1. If (i) the posterior of $\phi$ is consistent in the sense of Definition 6.1, (ii) action space $H_a$ is compact, (iii) $\sup_{\eta \in H(\phi_0)} L(\eta, a)$ is continuous in $a \in H_a$ and $a^*_{x^\infty} \equiv \arg \inf_{a \in H_a} \left\{ \sup_{\eta \in H(\phi_0)} L(\eta, a) \right\}$ uniquely exists, (iv)
there exists a subset \( \tilde{H} \subset \mathcal{H} \) such that \( H(\phi) \subset \tilde{H} \) at \( \phi = \phi_0 \) and \( \mu_{\phi} \)-almost every \( \phi \) and \( L(\eta,a) \) is bounded on \( \tilde{H} \times \mathcal{H}_a \), and (v) there exists \( \epsilon > 0 \) and a real-valued function \( u(\cdot) : \Phi \rightarrow \mathbb{R} \) such that \( u(\cdot) \) is continuous at \( \phi_0 \), \( u(\phi_0) = 0 \), \( u(\cdot) \) is bounded in the \( \epsilon \)-neighborhood of \( \phi_0 \), and

\[
\sup_{a \in \mathcal{H}_a} \left\| \sup_{\eta \in H(\phi)} L(\eta,a) - \sup_{\eta \in H(\phi_0)} L(\eta,a) \right\| \leq u(\phi) \quad \text{for all } \| \phi - \phi_0 \| < \epsilon.
\]

Then, the posterior gamma-minimax action \( a_{x,N}^* \) converges to \( a_{x,\infty}^* \), \( \hat{p}(x^\infty|\phi_0) \)-almost surely.

**Proof.** Let \( q_N(a) \equiv \int_\Phi \sup_{\eta \in H(\phi)} L(\eta,a) dF_{\eta|X^N} \) and \( q(a) \equiv \sup_{\eta \in H(\phi_0)} L(\eta,a) \). Under the given set of assumptions, we will first show \( q_N(a) \) uniformly converges to \( q(a) \), \( \hat{p}(x^\infty|\phi_0) \)-almost surely.

Let \( B_\epsilon \) be an \( \epsilon \)-neighborhood of \( \phi_0 \) as given in assumption (v), and let \( M < \infty \) be the upper bound of \( L(\eta,a) \) implied in the assumption (iv). The uniform bound of \( |q_N(a) - q(a)| \) is obtained as follows,

\[
\sup_{a \in \mathcal{H}_a} |q_N(a) - q(a)| = \sup_{a \in \mathcal{H}_a} \left| \int_\Phi \left( \sup_{\eta \in H(\phi)} L(\eta,a) - \sup_{\eta \in H(\phi_0)} L(\eta,a) \right) dF_{\eta|X^N} \right|
\leq \sup_{a \in \mathcal{H}_a} \int_{B_\epsilon} \left| \sup_{\eta \in H(\phi)} L(\eta,a) - \sup_{\eta \in H(\phi_0)} L(\eta,a) \right| dF_{\eta|X^N}
\]

where the last inequality follows by assumption (v) and the boundedness of the loss function (assumption (iv)). By the assumption of posterior consistency, \( \lim_{N \to \infty} F_{\eta|X^N}(B_\epsilon \cap \Phi) = 0 \) and \( \lim_{N \to \infty} \int_{B_\epsilon} u(\phi - \phi_0)dF_{\phi|X^N} = u(0) = 0 \), \( p(x^\infty|\phi_0) \)-almost surely. Therefore, \( q_N(a) \) uniformly converges to \( q(a) \) almost surely.

By the standard argument of consistency of the minimizers (see, e.g., Theorem 2.1 of Newey and McFadden (1994)), the uniform convergence property established above and assumptions (ii) and (iii) imply \( a_{x,N}^* \) converges to \( a_{x,\infty}^* \).

When the loss function is specified to be a monotonically increasing function of the metric \( \| \eta - a \| \), \( a_{x,\infty}^* \) is given by the center of the smallest circle that contains \( H(\phi_0) \). That is, for the case of convex \( H(\phi_0) \), the gamma-minimax action asymptotically coincides with a reasonable
midpoint of the true identified set. On the other hand, if $H(\phi_0)$ is not convex, $a^*_x$ may not be contained in $H(\phi_0)$. Under the conservatism represented by the gamma-minimax criterion, a topological property of the interior of $H(\phi_0)$ does not matter and only the action that minimizes the distance to the extremum point in $H(\phi_0)$ determines the optimal action.

Under the same set of assumptions, it can be shown that the gamma-minimax regret action also converges to $a^*_x$. Thus, for large sample, we obtain the same action no matter whether we minimize the posterior upper risk or posterior upper regret.

References


