Estimation under Ambiguity*

Raffaella Giacomini[†], Toru Kitagawa[‡], and Harald Uhlig[§]

This draft: April 2019

Abstract

To perform a Bayesian analysis for a set-identified model, two distinct approaches exist; the standard Bayesian inference that assumes a single prior for non-identified parameters, and the Bayesian inference for the identified set that assumes full ambiguity (multiple priors) for the parameters within their identified set. Both of the prior inputs considered by these two extreme approaches can often be a poor representation of the researcher's prior knowledge in practice. This paper fills this large gap between the two approaches by proposing a framework of multiple prior robust Bayes analysis that can simultaneously incorporate a probabilistic belief for the non-identified parameters and a misspecification concern about this belief. Our proposal introduces a *benchmark prior* representing the researcher's partially credible probabilistic belief for non-identified parameters, and *a set of priors* formed in its Kullback-Leibler (KL) neighborhood whose radius controls the degree of researcher's confidence put on the benchmark prior. We propose point estimation and optimal decision involving set-identified parameters by minimizing the worst-case posterior expected loss, i.e., solving conditional gamma-minimax. We clarify that the conditional gamma-minimax problem is analytically tractable and simple to solve numerically. We also derive analytical properties of the proposed robust Bayesian procedure in the limiting situations where the radius of KL neighborhood and/or the sample size are large. Our procedure can be also used to deliver the range of posterior quantities including the mean, quantiles, and the probability, based on which one can perform global sensitivity analysis.

*We would like to thank Lars Hansen, Frank Kleibergen, and several seminar and conference participants for their valuable comments. We gratefully acknowledge financial support from ERC grants (numbers 536284 and 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001).

[†]University College London, Department of Economics/Cemmap. Email: r.giacomini@ucl.ac.uk

[‡]University College London, Department of Economics/Cemmap. Email: t.kitagawa@ucl.ac.uk

[§]University of Chicago, Department of Economics. Email: huhlig@uchicago.edu

1 Introduction

Consider a partially identified model where the distribution of observables can be indexed by a vector of finite dimensional reduced-form parameters $\phi \in \Phi \subset \mathbb{R}^{dim(\phi)}$, $dim(\phi) < \infty$, but knowledge of ϕ fails to pin down the values of underlying structural parameters and the object of interest. By the definition of the reduced-form parameters of the structural modelling in econometrics (equivalent to minimally sufficient parameters in statistics), ϕ is identifiable (i.e., there is no $\phi, \phi' \in \Phi$, $\phi \neq \phi'$, that are observationally equivalent).

Following the parametrizations used in Poirier (1998) and Moon and Schorfheide (2012) for set-identified models, let $\theta \in \Theta$ denote the auxiliary parameters that are non-identified even with knowledge of ϕ and a set of identifying assumptions. On the other hand, θ is necessary in order to pin down the value of a (scalar) object of interest $\alpha = \alpha(\theta, \phi) \in \mathbb{R}$. We denote a sample by X and the realized one by x. Since ϕ are the reduced-form parameters, the value of the likelihood $l(x|\theta, \phi)$ depends only on ϕ for every realization of X, equivalent to saying $X \perp \theta|\phi$. The domain of θ, on the other hand, can be constrained by the value of ϕ and/or the imposed identifying assumptions that can depend on ϕ. We refer to the set of θ logically compatible with the conditioned value of ϕ and the imposed identifying assumption as the identified set of θ denoted by $IS_\theta(\phi)$. The identified set of α is accordingly defined by the range of $\alpha(\theta, \phi)$ when θ varies over $IS_\theta(\phi)$,

$$IS_\alpha(\phi) \equiv \{\alpha(\theta, \phi) : \theta \in IS_\theta(\phi)\}, \tag{1}$$

which can be viewed as a set-valued map from ϕ to $\mathbb{R}$.

We will focus on several leading examples of this set-up in the paper and use them to illustrate our methods.

Example 1.1 (Supply and demand) *Suppose the object of interest α is a structural parameter in a system of simultaneous equations. For example, consider a static version of the model of labor supply and demand analyzed by Baumeister and Hamilton (2015):*

$$Ax_t = u_t, \tag{2}$$

where $x_t = (\Delta w_t, \Delta n_t)$ with Δw_t and Δn_t the growth rates of wages and employment, respectively, $A = \begin{bmatrix} -\beta_d & 1 \\ -\beta_s & 1 \end{bmatrix}$ with $\beta_s \geq 0$ the short-run wage elasticity of supply and $\beta_d \leq 0$ the short-run wage elasticity of demand and u_t are shocks assumed to be i.i.d. $N(0, D)$ with

$D = diag(d_1, d_2)$. *The reduced form representation of the model is*

$$x_t = \varepsilon_t, \tag{3}$$

with $E(\varepsilon_t \varepsilon_t') = \Omega = A^{-1} D \left(A^{-1}\right)'$. *The reduced form parameters are* $\phi = (w_{11}, w_{12}, w_{22})'$, *with* w_{ij} *the* $(i,j)-th$ *element of* Ω. *Let* β_s *be the parameter of interest. The full vector of structural parameters is* $(\beta_s, \beta_d, d_1, d_2)'$, *which can be reparametrized to* $(\beta_s, w_{11}, w_{12}, w_{22})$.[1] *Accordingly, in our notation,* θ *can be set to* β_s, *and the object of interest* α *is* $\theta = \beta_s$ *itself. The identified set of* α *when* $w_{12} > 0$ *can be obtained as (see, e.g., Baumeister and Hamilton (2015)):*

$$IS_\alpha(\phi) = \{\alpha : w_{12}/w_{11} \leq \alpha \leq w_{22}/w_{12}\}. \tag{4}$$

Example 1.2 (Impulse response analysis) *Suppose the object of interest is an impulse-response in a general partially identified structural vector autoregression (SVAR) for a zero mean vector* x_t*:*

$$A_0 x_t = \sum_{j=1}^{p} A_j x_{t-j} + u_t, \tag{5}$$

where u_t *is* $i.i.d.N(0, I)$, *with* I *the identity matrix. The reduced form VAR representation is*

$$x_t = \sum_{j=1}^{p} B_j x_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim \mathcal{N}(0, \Omega),$$

The reduced form parameters are $\phi = (vec(B_1)', \ldots, vec(B_p)', w_{11}, w_{12}, w_{22})' \in \Phi$, *with* Φ *restricted to the set of* ϕ *such that the reduced form VAR can be inverted into a* $VMA(\infty)$ *model:*

$$x_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}. \tag{6}$$

The non-identified parameter is $\theta = (vec(Q)')'$, *where* Q *is the orthonormal rotation matrix that transforms the reduced form residuals into structural shocks (i.e.,* $u_t = Q' \Omega_{tr}^{-1} \varepsilon_t$, *where* Ω_{tr} *is the Cholesky factor from the factorization* $\Omega = \Omega_{tr} \Omega_{tr}'$*). The object of interest is the* $(i,j)-th$ *impulse response at horizon* h, *which captures the effect on the* i*-th variable in* x_{t+h} *of a unit shock to the* j*-th element of* u_t *and is given by* $\alpha = e_i' C_h \Omega_{tr} Q e_j$, *with* e_j *the* $j-th$ *column of the identity matrix. The identified set of the* $(i,j)-th$ *impulse response in the absence of any identifying restrictions is*

$$IS_\alpha(\phi) = \{\alpha = e_i' C_h \Omega_{tr} Q e_j : Q \in \mathcal{O}\}, \tag{7}$$

where $\mathcal{O}$ *is the space of orthonormal matrices.*

[1] See Section **??** below for the transformation. If β_d is a parameter of interest, an alternative reparametrization allows us to transform the structural parameters into $(\beta_d, w_{11}, w_{12}, w_{22})$.

Example 1.3 (Entry game) *As a microeconometric application, consider the two-player entry game in Bresnahan and Reiss (1991) used as the illustrating example in Moon and Schorfheide (2012). Let* $\pi_{ij}^M = \beta_j + \epsilon_{ij}$, $j = 1, 2$, *be the profit of firm* j *if firm* j *is monopolistic in market* $i \in \{1, \ldots, n\}$, *and* $\pi_{ij}^D = \beta_j - \gamma_j + \epsilon_{ij}$ *be firm* j*'s profit if the competing firm also enters the market* i *(duopolistic). The* ϵ_{ij}*'s capture unobservable (to the econometrician) profit components of firm* j *in market* i *and they are known to the players, and we assume* $(\epsilon_{i1}, \epsilon_{i2}) \sim \mathcal{N}(0, I_2)$. *We restrict our analysis to the pure strategy Nash equilibrium, and assume that the game is strategic substitute,* $\gamma_1, \gamma_2 \geq 0$. *The data consist of iid observations on entry decisions of the two firms. The non-redundant set of reduced form parameters are* $\phi = (\phi_{11}, \phi_{00}, \phi_{10})$, *the probabilities of observing a duopoly, no entry, or the entry of firm 1. This game has multiple equilibria depending on* $(\epsilon_{i1}, \epsilon_{i2})$*; the monopoly of firm 1 and the monopoly of firm 2 are pure strategy Nash equilibrium if* $\epsilon_{i1} \in [-\beta_1, -\beta_1 + \gamma_1]$ *and* $\epsilon_{i2} \in [-\beta_2, -\beta_2 + \gamma_2]$. *Let* $\psi \in [0, 1]$ *be a parameter for an equilibrium selection rule representing the probability that the monopoly of firm 1 is selected given* $(\epsilon_{i1}, \epsilon_{i2})$ *leading to multiplicity of equilibria. Let the parameter of interest be* $\alpha = \gamma_1$, *the substitution effect for firm 1 from the firm 2 entry. The vector of full structural parameters augmented by the equilibrium selection parameter* ψ *is* $(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$, *and they can be reparametrized into* $(\beta_1, \gamma_1, \phi_{11}, \phi_{00}, \phi_{10})$.[2] *Hence, in our notation,* θ *can be set to* $\theta = (\beta_1, \gamma_1)$ *and* $\alpha = \gamma_1$. *The identified set for* θ *does not have a convenient closed-form, but it can be expressed implicitly as*

$$IS_\theta(\phi) = \left\{ (\beta_1, \gamma_1) : \gamma_1 \geq 0, \min_{\beta_2 \in \mathbb{R}^2, \gamma_2 \geq 0, \psi \in [0,1]} \| \phi - \phi(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi) \| = 0 \right\}, \tag{8}$$

where $\phi(\cdot)$ *is the map from structural parameters* $(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$ *to reduced-form parameters* ϕ. *Projecting* $IS_\theta(\phi)$ *to the* γ_1*-coordinate gives the identified set for* $\alpha = \gamma_1$.

Generally, the identified set only collects all the admissible values of α that meets the imposed identifying assumptions given knowledge of the distribution of observables (the reduced-form parameters). In some contexts, however, it may not be the case that the identifying assumptions imposed dogmatically exhaust all the available information that the researcher has. A more common situation is that the researcher has some form of additional but only partially credible assumptions about some underlying structural parameters or about the non-identified parameter θ based on economic theory, background knowledge of the problem, or empirical studies that use different data. From the standard Bayesian viewpoint, the recommendation is

[2] See Section C below for concrete expressions of the transformation.

to incorporate this information into the analysis through specifying a prior distribution of (θ, ϕ) or that of the full structural parameters. For instance, in the case of Example 1.1, Baumeister and Hamilton (2015) propose a prior of the elasticity of supply β_s that draws on estimates obtained in microeconometric studies, and consider a Student's t density calibrated to assign 90% probability to the interval $\beta_s \in (0.1, 2.2)$. Another example considered by Baumeister and Hamilton (2015) is a prior that incorporates long-run identifying restrictions in SVARs in a non-dogmatic way, as a way to capture the uncertainty one might have about the validity of this popular but controversial type of identifying restrictions. In the situations where additional informative prior information other than the identifying restrictions is not available, some Bayesian literature has recommended the use of the uniform prior as a representation of the indifference among θ's within the identified set. For example, in SVARs subject to sign restrictions (Uhlig (2005)) it is common to use the uniform distribution (the Haar measure) over the set of orthonormal matrices in (7) that satisfy the sign restrictions. For the entry game in Example 1.3, one of the prior specifications considered in Moon and Schorfheide (2012) is the uniform prior over the identified set of θ.

At the opposite end of the standard Bayesian spectrum, Giacomini and Kitagawa (2018) advocate adopting a multiple-prior Bayesian approach when one has no further information about θ besides a set of exact restrictions that can be used to characterize the identified set. While maintaining a single prior for ϕ, this set of priors consists of any conditional prior for θ given ϕ, $\pi_{\theta|\phi}$, supported on the identified set $IS_\theta(\phi)$. Giacomini and Kitagawa (2018) propose to conduct a posterior bound analysis based on the resulting class of posteriors, that leads to an estimator for $IS_\alpha(\phi)$ and an associated "robust" credible region that asymptotically converge to the true identified set, which is the object of interest of frequentist inference. Being implicit about the ambiguity inherent in partial identification analysis, one can also consider posterior inference for the identified set as in Moon and Schorfheide (2011), Kline and Tamer (2016), and Liao and Simoni (2013), to obtain similar asymptotic equivalence between posterior inference and frequentist inference.

The motivation for the methods that we propose in this paper is the observation that both types of prior inputs considered by the two extreme approaches discussed above - a precise specification of $\pi_{\theta|\phi}$ or full ambiguity about $\pi_{\theta|\phi}$ - could be a poor representation of the belief that the researcher actually possesses in a given application. For example, the Student's t prior specified by Baumeister and Hamilton (2015) in Example 1.1 builds on the plausible values of α found in microeconometric studies, but such prior evidence may not to be sufficient for the researcher to be confident in the particular shape of the prior. At the same time, the researcher may not want to entirely discard such available prior evidence for α and take the fully ambiguous approach. Further, a researcher who is indifferent over values of θ within its identified set may be concerned about the fact that even a uniform prior on $IS_\theta(\phi)$ can cause unintentionally informative prior for α or other parameters. Full ambiguity for $\pi_{\theta|\phi}$ may also not be appealing, if, for instance, a prior that is degenerate at an extreme value in $IS_\theta(\phi)$ appears less sensible than a non-degenerate prior that supports any θ in the identified set. The existing approaches to inference in partially identified models lack a formal and convenient framework that enables one to incorporate any "vague" prior knowledge for the non-identified parameters that the researcher possesses and is willing to exploit.

The main contribution of this paper is to fill the large gap between the single prior Bayesian approach and the fully ambiguous multiple prior Bayesian approach by proposing a method that can simultaneously incorporate a probabilistic belief for the non-identified parameters and a misspecification concern about this belief in a unified manner. Our idea is to replace the fully ambiguous beliefs for $\pi_{\theta|\phi}$ considered in Giacomini and Kitagawa (2018) by a class of priors defined in a neighborhood of a *benchmark prior*. The benchmark prior $\pi^*_{\theta|\phi}$ represents the researcher's reasonable but partially credible prior knowledge about θ, and the class of priors formed around the benchmark prior captures ambiguity or misspecification concerns about the benchmark prior. The radius of the neighborhood prespecified by the researcher controls the degree of confidence put on the benchmark prior. We then propose point estimation and interval estimation for the object of interest α by minimizing the worst-case (minimax) posterior expected loss with respect to the priors constrained to this neighborhood.

Our paper makes the following unique contributions: (1) we clarify that the estimation for the partially identified parameter under vague prior knowledge can be formulated as a decision under ambiguity in the form considered in Hansen and Sargent (2001); (2) we provide an analytically tractable and numerically convenient way to solve the conditional gamma-minimax estimation problem in general cases; (3).we give simple analytical solutions for the special cases of a quadratic and a check loss function and for the limit case when the shape of benchmark prior is irrelevant; (4) we derive the properties of our method in large samples.

1.1 Related Literature

Approaches of introducing a set of priors to draw robust posterior inference go back to the robust Bayesian analysis of Robbins (1951), where the basic premise is that the decision maker cannot specify a unique prior distribution for the parameters due to the limited prior knowledge or limited ability to elicit the prior. Good (1965) argues that prior input easier to elicit in practice is a class of priors rather than a single prior. When the class of priors is used as prior input, however, how to update the class in light of the data is short of consensus. One extreme is the Type-II maximum likelihood (empirical Bayes) updating rule of Good (1965) and Gilboa and Schmeidler (1993), while the other extreme is what Gilboa and Marinacci (2016) call the full Bayesian updating rule. See Jaffray (1992) and Pires (2002). We introduce a single prior for the reduced-form parameters and a class of priors for the unrevisable part of prior. Since any prior in the class leads to the same value of marginal likelihood, we obtain the same set of posteriors no matter what updating rule we apply.

We perform minimax estimation/decision by applying the minimax criterion to the set of *posteriors*, which is referred to as the conditional gamma-minimax criterion in the statistics literature; see, e.g., DasGupta and Studden (1989), and Betró and Ruggeri (1992). The conditional gamma-minimax criterion is distinguished from the (unconditional) gamma-minimax criterion where minimax is performed a priori observing the data. See, e.g., Manski (1981), Berger (1985), Chamberlain (2000), and Vidakovic (2000). An analogue to the gamma-minimax analysis in the economics decision theory is the maximin expected utility theory axiomatized by Gilboa and Schmeidler (1989).

The existing gamma-minimax analyses focus on identified models and have consider various ways of constructing a prior class, including the class of bounded and unbounded variance priors (Chamberlain and Leamer (1976) and Leamer (1982)), ϵ-contaminated class of priors (Berger and Berliner (1986)), and class priors built on a nonadditive lower probability (Wasserman (1990)), a class of fixed marginal priors (Lavine et al. (1991)), to list a few. This paper focuses on a class of set-identified models where the posterior remains to be sensitive even in large samples due to the lack of identification. A class of priors proposed in this paper consists of those belonging to a specified KL-neighborhood around the benchmark prior. As shown in Lemma 2.2 below, conditional gamma-minimax with such class of priors is closely related to multiplier minimax problem considered in Peterson et al. (2000) and Hansen and Sargent (2001). When the benchmark prior covers the entire identified set, the KL-class of priors with an arbitrarily large neighborhood can replicate the class of priors considered in Giacomini and Kitagawa (2018).

Our analysis of obtaining the set of posterior means, probabilities, quantiles, etc. can be used for global sensitivity analysis, where the goal of analysis is to summarize the sensitivity of the posterior to a choice of prior. See Moreno (2000) and references therein. For global sensitivity analysis, Ho (2019) also considers a KL-based class of priors similar to ours. Our approach differs from his approach if applied to the set-identified models in the following aspects. First, any priors in our prior class shares a prior for the reduced-form parameters, while it is not necessarily the case in Ho (2019). Second, Ho (2019) recommends to set the radius of the KL-neighborhood by reverse engineering in reference to a Gaussian approximation of the posterior. We in contrast argue to pin down the radius of KL-neighborhood by directly eliciting the range of prior means of a parameter for which we have reasonable prior knowledge in the form of an interval.

The concern of robustness tackled in our approach is about misspecification of the prior distribution in the Bayesian setting. In contrast, the frequentist approach to robustness typically concerns misspecification in the likelihood, moment conditions, or a specification of the distribution of unobservables. Estimators that are less sensitive to such misspecification and/or how to assess sensitivity thereof are analyzed in Armstrong and Kolesár (2019), Bonhomme and Weidner (2018), Christensen and Connault (2019), Kitamura et al. (2013), among others.

1.2 Roadmap

The remainder of the paper is organized as follows. In Section 2, we introduce the analytical framework and formulate the statistical decision problem with the multiple priors localized around the benchmark prior. Section 3 solves the multiplier minimax problem with a general loss function. With the quadratic and check loss functions, Section 4 analyzes point and interval estimations of the parameter of interest. Section 4 also considers the two types of limiting situations: (1) the radius of the set of priors diverges to infinity (fully ambiguous beliefs) and (2) the sample size goes to infinity. Section 5 discusses how to elicit the benchmark prior and how to set up the tuning parameter that governs the size of the prior class. In Section 6, we provide one empirical and one numerical examples.

2 Estimation as Statistical Decision under Ambiguity

2.1 Setting up the Set of Priors

The starting point of the analysis is to express a joint prior of (θ, ϕ) by $\pi_{\theta|\phi}\pi_{\phi}$, where $\pi_{\theta|\phi}$ is a conditional prior probability measure of the non-identified parameter θ given the reduced form parameter ϕ and π_{ϕ} is a marginal prior probability measure of ϕ. Since $\alpha = \alpha(\theta, \phi)$ is a function of θ given ϕ being fixed, $\pi_{\theta|\phi}$ induces a conditional prior distribution of α given ϕ, $\pi_{\alpha|\phi}$. The set of identifying assumptions imposed characterizes $IS_{\alpha}(\phi)$, the range of α values that is consistent with the given reduced-form parameter value ϕ and the identifying restrictions. Any prior for (θ, ϕ) that satisfies the imposed identifying assumptions with probability one has the support of $\pi_{\alpha|\phi}$ contained in the identified set $IS_\alpha(\phi)$, i.e., $\pi_{\alpha|\phi}(\alpha \in IS_\alpha(\phi)) = 1$, for all $\phi \in \Phi$. A sample X is informative about ϕ so that π_ϕ can be updated by data to obtain a posterior $\pi_{\phi|X}$, whereas the conditional prior $\pi_{\theta|\phi}$ (and hence $\pi_{\alpha|\phi}$) can never be updated by data and the posterior inference for α remains sensitive to the choice of conditional prior no matter how large the sample size is. Therefore, for the decision maker who is aware of these facts, misspecification of the unrevisable part of the prior $\pi_{\alpha|\phi}$ becomes a major concern in conducting posterior inference.

Suppose that the decision maker can form a benchmark prior $\pi^*_{\theta|\phi}$ for the unrevisable part of the prior. This conditional prior for θ given ϕ captures information about θ that is available before the model is brought to the data (see Section 6 for discussions on how to elicit a benchmark prior). This benchmark prior induces the benchmark prior for α given ϕ, denoted by $\pi^*_{\alpha|\phi}$. If one were to impose a sufficient number of restrictions to point-identify α, this would amount to have the benchmark prior of α given ϕ to be the point mass measure supported at the singleton identified set, and having the posterior of ϕ induces the single posterior of α. In contrast, under set-identification, $\pi^*_{\theta|\phi}$ determines how the benchmark probabilistic belief is allocated within the identified set $IS_\alpha(\phi)$ if non-singleton.

We consider a set of priors (ambiguous beliefs) in a neighborhood of $\pi^*_{\theta|\phi}$ - while maintaining a single prior of ϕ - and find the estimator of α that minimizes the worst-case posterior risk as the priors range over this neighborhood.

For the construction of a neighborhood around the benckmark conditional prior, we consider the Kullback-Leibler neighborhood of $\pi^*_{\theta|\phi}$ with radius $\lambda \in [0, \infty)$:

$$\Pi^\lambda\left(\pi^*_{\theta|\phi}\right) \equiv \left\{\pi_{\theta|\phi} : \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \leq \lambda\right\}, \tag{9}$$

where $\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \geq 0$ is the Kullback-Leibler divergence (KL-divergence) from $\pi^*_{\theta|\phi}$ to $\pi_{\theta|\phi}$, or equivalently the relative entropy of $\pi_{\theta|\phi}$ relative to $\pi^*_{\theta|\phi}$:

$$\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \int_{IS_\theta(\phi)} \ln\left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}}\right) d\pi_{\theta|\phi}.$$

$\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi})$ is finite if and only if $\pi_{\theta|\phi}$ is absolutely continuous with respect to $\pi^*_{\theta|\phi}$. Otherwise, we define $\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \infty$ following the convention. As is well known in information theory, $\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) = 0$ if and only if $\pi_{\theta|\phi} = \pi^*_{\theta|\phi}$ (see, e.g., Lemma 1.4.1 in Dupuis and Ellis (1997)). Since the support of the benchmark prior $\pi^*_{\theta|\phi}$ coincides with or contained by $IS_\theta(\phi)$, any $\pi_{\theta|\phi}$ belonging to $\Pi^\lambda(\pi^*_{\theta|\phi})$ satisfies $\pi_{\theta|\phi}(IS_\theta(\phi)) = 1$.

An analytically attractive property of the KL-divergence is its convexity property in $\pi_{\theta|\phi}$, which guarantees that the constrained minimax problem (13) below has a unique solution under mild regularity conditions. Note that the KL-neighborhood is constructed at each $\phi \in \Phi$ independently, and no constraint is imposed to restrict the priors in $\Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ across different values of ϕ, i.e., fixing $\pi_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ at one value of ϕ does not restrict feasible priors in $\Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ for the remaining values of ϕ. We denote the class of *joint* priors of (θ, ϕ) formed by selecting $\pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi})$ for each $\phi \in \Phi$ by

$$\Pi^{\lambda}_{\theta\phi} \equiv \left\{\pi_{\theta\phi} = \pi_{\theta|\phi}\pi_{\phi} : \pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi}), \ \ \forall \phi \in \Phi\right\}.$$

This way of constructing priors simplifies our multiple-prior analysis both analytically and numerically, and it is what we pursue in this paper. Alternatively, one could consider to form the KL-neighborhood for the *unconditional* prior of (θ, ϕ) around its benchmark, as considered in Ho (2019).

In the class of partially identified models we consider, there are several reasons why we prefer to introduce ambiguity to the unrevisable part of the prior $\pi_{\theta|\phi}$ rather than to the fully unconditional prior $\pi_{\theta\phi}$. First, the major source of posterior sensitivity comes from $\pi_{\theta|\phi}$, and our aim is to make estimation and inference robust to the prior input that can not be updated by the data. Second, if we allow for multiple priors also for ϕ, we potentially distort the posterior information about the identified parameter by allowing for a prior of π_{ϕ} that fits poorly to the data, i.e., π_{ϕ} far from the observed likelihood. Keeping π_{ϕ} fixed, we can ensure that any posteriors equally fits the data, i.e., the value of the marginal likelihood is kept fixed. Third, if we keep π_{ϕ} fixed, the updating rules of for the set of priors proposed in the decision theory under ambiguity, including, for instance, the full Bayesian updating rule axiomatized by Pires (2004), the maximum likelihood updating rule axiomatized by Gilboa and Schmeidler (1993), and the hypothesis-testing updating rule axiomatized by Ortoleva (2004), all lead to the same set of posteriors. This means that the minimax decision decided after X is observed is invariant to the choice of the updating rule, while it is not necessarily the case if we allow multiple priors for ϕ.

$\Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ is increasing in λ and, $\Pi^{\infty}(\pi^*_{\theta|\phi}) \equiv \lim_{\lambda\to\infty} \sup \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ contains any probability measure that is dominated by $\pi^*_{\theta|\phi}$, i.e., the benchmark prior becomes relevant only for determining the support of $\pi_{\theta|\phi}$ in the limiting situation of $\lambda \to \infty$. The radius λ is the scalar choice parameter that represents the researcher's degree of credibility placed on the benchmark prior. Since our construction of the prior class is pointwise at each $\phi \in \Phi$, the radius λ could in principle differ across ϕ, but we set λ to a positive constant independent of ϕ in order to simplify the analysis and its elicitation. The radius parameter λ itself does not have an easily interpretative scale. It is therefore hard to translate the subjective notion of "credibility" on the benchmark prior into a proper choice of λ. Section 6 below proposes a practical way to elicit λ in reference to a parameter for which partial prior knowledge is available.

2.2 Posterior Minimax Decision

We first consider statistical decision problems in the presence of multiple priors and posteriors formulated by $\Pi^{\lambda}(\pi^*_{\theta|\phi})$. Specifically, we focus on point estimation problem for the parameter of interest α, while the framework and the main results shown below can be applied to other statistical decision problems including interval estimation and statistical treatment choice (Manski (2004)).

Let $\delta(X)$ be a statistical decision function that maps data X to a space of actions $\mathcal{D} \subset \mathbb{R}$, and $h(\delta(X), \alpha)$ is a loss function. In point-estimation of scalar parameter of interest, the loss function can be, for instance, the quadratic loss

$$h(\delta(X), \alpha) = (\delta(X) - \alpha)^2 , \tag{10}$$

or the check loss for the τ-th quantile

$$\begin{aligned} h(\delta(X), \alpha) &= \rho_\tau (\alpha - \delta(X)) \qquad (11) \\ \rho_\tau(u) &= \tau u \cdot 1\{u > 0\} - (1-\tau) u \cdot 1\{u < 0\} . \end{aligned}$$

Given a conditional prior $\pi_{\theta|\phi}$ and the single posterior for ϕ, the posterior expected loss is given by

$$\int_{\Phi} \left[\int_{IS_\theta(\phi)} h(\delta(x), \alpha(\theta, \phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X} . \tag{12}$$

Provided that the decision maker faces ambiguous beliefs for $\pi_{\theta|\phi}$ in the form of multiple priors $\Pi^{\lambda}(\pi^*_{\theta|\phi})$, we assume that the decision maker reaches an optimal decision by applying the conditional gamma-minimax criterion, i.e., minimizes in $\delta(x)$ the worst-case posterior expected loss when $\pi_{\theta|\phi}$ varies over $\Pi^{\lambda}(\pi^*_{\theta|\phi})$ for every $\phi \in \Phi$,

Constrained posterior minimax:

$$
\begin{aligned}
&\min_{\delta(x)\in\mathcal{D}} \max_{\pi_{\theta\phi}\in\Pi^{\lambda}_{\theta\phi}} \int_{\Phi} \left[\int_{IS_\theta(\phi)} h(\delta(x), \alpha(\theta,\phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X} \\
&= \min_{\delta(x)\in\mathcal{D}} \int_{\Phi} \max_{\pi_{\theta|\phi}\in\Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_\theta(\phi)} h(\delta(x), \alpha(\theta,\phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X}
\end{aligned} \tag{13}
$$

where the equality follows by noting that the class of joint priors $\Pi^{\lambda}_{\theta\phi}$ is formed by an independent selection of $\pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi})$ at each $\phi \in \Phi$. Since any $\pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi})$ has the support contained in $IS_\theta(\phi)$, the region of the integration with respect to θ can be extended to the whole parameter space Θ without changing its value;

$$
\int_{IS_\theta(\phi)} h(\delta(x), \alpha(\theta,\phi)) d\pi_{\theta|\phi} = \int h(\delta(x), \alpha(\theta,\phi)) d\pi_{\theta|\phi}
$$

for $\pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi})$.

Since the loss function $h(\delta, \alpha(\theta,\phi))$ depends on θ only through the parameter of interest α, we may want to work with the set of priors for α given ϕ instead of that of θ given ϕ. Specifically, we form the set of priors for $\pi_{\alpha|\phi}$ by forming the KL-neighborhood around $\pi^*_{\alpha|\phi}$ the benchmark conditional prior for α given ϕ constructed by marginalizing $\pi^*_{\theta|\phi}$ to α,

$$
\Pi^{\lambda}(\pi^*_{\alpha|\phi}) = \left\{ \pi_{\alpha|\phi} : \mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \lambda \right\},
$$

and consider the following constrained posterior minimax problem:

$$
\min_{\delta(x)\in\mathcal{D}} \int_{\Phi} \max_{\pi_{\alpha|\phi}\in\Pi^{\lambda}(\pi^*_{\alpha|\phi})} \left[\int_{IS_\alpha(\phi)} h(\delta(x), \alpha) d\pi_{\alpha|\phi} \right] d\pi_{\phi|X}. \tag{14}
$$

$\Pi^{\lambda}(\pi^*_{\alpha|\phi})$ nests and is generally larger than the set of priors formed by $\alpha|\phi$ -marginals of $\pi_{\theta|\phi} \in \Pi^{\lambda}(\pi^*_{\theta|\phi})$, as shown in Lemma A.1 in Appendix A. Nevertheless, the next lemma implies that (13) and (14) lead to the same solution.

Lemma 2.1 *Fix $\phi \in \Phi$ and $\delta \in \mathbb{R}$, and let $\lambda \geq 0$ be given. For any measurable loss function $h(\delta, \alpha(\theta,\phi))$, it holds*

$$
\max_{\pi_{\theta|\phi}\in\Pi^{\lambda}(\pi^*_{\theta|\phi})} \left[\int_{IS_\theta(\phi)} h(\delta, \alpha(\theta,\phi)) d\pi_{\theta|\phi} \right] = \max_{\pi_{\alpha|\phi}\in\Pi^{\lambda}(\pi^*_{\alpha|\phi})} \left[\int_{IS_\alpha(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi} \right].
$$

Proof. See Appendix A. ■

This lemma implies that no matter whether we introduce ambiguity to the entire non-identified parameterss θ conditional on ϕ or only to the parameter of interest α conditional on ϕ with being agnostic about the conditional prior of $\theta|\alpha, \phi$, the constrained minimax problem supports the same decision as optimal, as far as the common λ is specified. This lemma therefore justifies us to ignore ambiguity or robustness concern about the set-identified parameters other than α and to focus only on the the set of priors of $\alpha|\phi$ which ultimately matters for the posterior expected loss. A useful insight of this lemma is that, given π_ϕ, elicitation of λ can be done by focusing on a set-identified parameter for which partial prior knowledge is available in the form of the range of prior probabilities or the range of means. This parameter used to elicit λ can be different from the parameter of interest α, while it is useful to tune up the amount of ambiguity to be incorporated into the robust estimation and inference for α. We discuss further detail about elicitation of λ in Section 6 and illustrate in the empirical application of Section 7.

A minimax problem closely related to the constrained posterior minimax problem formulated in (14) above is the *multiplier minimax* problem:

Multiplier minimax:

$$\min_{\delta(x)\in\mathcal{D}} \int_\Phi \left[\max_{\pi_{\alpha|\phi}\in\Pi^\infty(\pi^*_{\alpha|\phi})} \left\{ \int_{IS_\alpha(\phi)} h(\delta(x),\alpha) d\pi_{\alpha|\phi} - \kappa \mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \right\} \right] d\pi_{\phi|X}, \tag{15}$$

where $\kappa \geq 0$ is a fixed constant. The next lemma shows the relationship between the inner maxization problems in (14) and (15):

Lemma 2.2 *(Lemma 2.2. in Peterson et al. (2000), Hansen and Sargent (2001)) Fix $\delta \in \mathcal{D}$ and let $\lambda > 0$. Define*

$$r_\lambda(\delta, \phi) \equiv \max_{\pi_{\alpha|\phi}\in\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_\alpha(\phi)} h(\delta,\alpha) d\pi_{\alpha|\phi} \right]. \tag{16}$$

If $r_\lambda(\delta,\phi) < \infty$, then there exist $\kappa_\lambda(\delta,\phi) \geq 0$ such that

$$r_\lambda(\delta, \phi) = \max_{\pi_{\alpha|\phi}\in\Pi^\infty(\pi^*_{\alpha|\phi})} \left\{ \int_{IS_\alpha(\phi)} h(\delta,\alpha) d\pi_{\alpha|\phi} - \kappa_\lambda(\delta,\phi) \left(\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) - \lambda \right) \right\}. \tag{17}$$

Furthermore, if $\pi^0_{\alpha|\phi} \in \Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)$ is a maximizer in (16), $\pi^0_{\alpha|\phi}$ also maximizes (17) and satisfies

$$\kappa_\lambda(\delta,\phi) \left(\mathcal{R}(\pi^0_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) - \lambda \right) = 0.$$

In this lemma, $\kappa_\lambda(\delta, \phi)$ is interpreted as the Lagrangian multiplier in the constrained optimization problem (16), whose value depends on λ. Furthermore, $\kappa_\lambda(\delta, \phi)$ that makes the constrained optimization (16) and the unconstrained optimization (17) equivalent depends on ϕ and δ through $\pi^*_{\alpha|\phi}$ and the loss function $h(\delta, \alpha)$ (See Theorem 3.1 below). Conversely, if we formulate the robust decision problem with starting from (15) with constant $\kappa > 0$ independent of ϕ and δ, an implied value of λ that equalizes (16) and (17) depends on ϕ and δ, i.e., the radii of the implied sets of priors vary across ϕ and depend on the loss function $h(\delta, \alpha)$. The multiplier minimax problem with constant κ appears analytically and numerically simpler than the constrained minimax problem with constant λ, whereas a non-desirable feature is that the implied class of priors (the radius of KL-neighborhood) is endogenously determined depending on what loss function one specifies. Since our robust Bayes analysis sets the set of priors (ambiguous belief) as the primary input which is invariant to the choice of loss function, we consider the constrained minimax problem (14) with constant λ rather than the multiplier minimax problem (15) with fixed κ. Such approach is also consistent with the standard Bayesian global sensitivity analysis where the ranges of posterior quantities are computed with the same set of priors no matter whether one focuses on the range of posterior means or quantiles. Nevertheless, the relationship shown in Lemma 2.2 is useful for examining the analytical properties of the constrained minimax problem, as we shall shown in the next section.

3 Solving Constrained Posterior Minimax Problem

3.1 Finite Sample Solution

The inner maximization in the constrained minimax problem of (14) has an analytical solution, as shown in the next theorem.

Theorem 3.1 *Assume for any $\delta \in \mathcal{D}$, $h(\delta, \alpha)$ is bounded on $IS_\alpha(\phi)$, $\pi_\phi - a.s.$, and the distribution of $h(\delta, \alpha)$ induced by $\alpha \sim \pi^*_{\alpha|\phi}$ is nondegenerate. The constrained posterior minimax problem (14) is then equivalent to*

$$\min_{\delta \in \mathcal{D}} \int_\Phi r_\lambda(\delta, \phi)\, d\pi_{\phi|X}, \tag{18}$$

$$\text{where} \qquad r_\lambda(\delta, \phi) = \int_{IS_\alpha(\phi)} h(\delta, \alpha) d\pi^0_{\alpha|\phi},$$

$$d\pi^0_{\alpha|\phi} = \frac{\exp\{h(\delta, \alpha)/\kappa_\lambda(\delta, \phi)\}}{\int_{IS_\alpha(\phi)} \exp\{h(\delta, \alpha)/\kappa_\lambda(\delta, \phi)\}\, d\pi^*_{\alpha|\phi}} \cdot d\pi^*_{\alpha|\phi},$$

and $\kappa_\lambda(\delta,\phi) > 0$ is the unique solution to

$$\min_{\kappa \geq 0} \left\{ \kappa \ln \int_{IS_\alpha(\phi)} \exp\left\{ \frac{h(\delta,\alpha)}{\kappa} \right\} d\pi^*_{\alpha|\phi} + \kappa\lambda \right\}.$$

Proof. See Appendix A. ■

Note that the statement of the theorem is valid for any sample size and any realization of X. The obtained representation simplifies the analytical investigation and the computation of the minimax decision, and we make use of it in the following sections. We can easily approximate the integrals of the objective function in (18) using Monte Carlo draws of (α,ϕ) sampled from the benchmark conditional prior $\pi^*_{\alpha|\phi}$ and the posterior $\pi_{\phi|X}$. The minimization for δ can be performed, for instance, by a grid search using the Monte Carlo approximated objective function.

3.2 Large Sample Behavior of the Posterior Minimax Decision

Investigating the large sample approximation of the optimal decision can suggest further computational simplification. Let n denote the sample size and $\phi_0 \in \Phi$ be the value of ϕ that generated the data (the true value of ϕ). To establish asymptotic convergence of the minimax optimal decision, we impose the following set of regularity assumptions.

Assumption 3.2

(i) (Posterior consistency) The posterior of ϕ is consistent to ϕ_0 almost surely, in the sense that for any open neighborhood G of ϕ_0, $\pi_{\phi|X}(G) \to 1$ as $n \to \infty$ for almost every sampling sequence.

(ii) (Nonnegative and bounded loss) The loss function $h(\delta,\alpha)$ is nonnegative and bounded,

$$0 \leq h(\delta,\alpha) \leq M < \infty,$$

holds for every $(\delta,\alpha) \in \mathcal{D} \times \mathcal{A}$.

(iii) (Compact action space) $\mathcal{D}$ the action space of δ is compact.

(iv) *(Nondegeneracy of loss) There exists $G_0 \subset \Phi$ an open neighborhood of ϕ_0 and positive constants $c > 0$ and $\epsilon > 0$ such that*

$$\pi^*_{\alpha|\phi}\left(\left\{\left(h(\delta,\alpha) - \tilde{h}\right)^2 \geq c\right\}\right) \geq \epsilon$$

holds for all $\delta \in \mathcal{D}$, $\tilde{h} \in \mathbb{R}$, and $\phi \in G_0$.

(v) *(Continuity of $\pi_{\alpha|\phi}$) The benchmark prior satisfies*

$$\left\|\pi^*_{\alpha|\phi} - \pi^*_{\alpha|\phi_0}\right\|_{TV} \equiv \sup_{B \in \mathcal{B}} \left|\pi^*_{\alpha|\phi}(B) - \pi^*_{\alpha|\phi_0}(B)\right| \to 0$$

as $\phi \to \phi_0$, where $\mathcal{B}$ is the class of measurable subsets in $\mathcal{A}$.

(vi) *(Differentiability of prior means) For each $\kappa \in (0, \infty)$,*

$$\sup_{\delta \in \mathcal{D}, \phi \in G_0} \left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}(h(\delta,\alpha)) \right\| < \infty,$$
$$\sup_{\delta \in \mathcal{D}, \phi \in G_0} \left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}\left(h(\delta,\alpha)\exp\left(\frac{h(\delta,\alpha)}{\kappa}\right)\right) \right\| < \infty.$$

(vii) *(Continuity of the worst-case loss and uniqueness of minimax action) $r_\lambda(\delta, \phi_0)$ defined in Lemma 2.2 and shown in Theorem 3.1 is continuous in δ and has a unique minimizer in δ.*

Assumption 3.2 (i) assumes that the posterior of ϕ is well-behaved and the true ϕ_0 can be estimated consistently in the Bayesian sense. The posterior consistency of ϕ can be ensured by imposing higher level assumptions on the likelihood of ϕ and the prior for ϕ. We do not present them here for brevity (see, e.g., Section 7.4 of Schervish (1995) for details about posterior consistency). The boundedness of the loss function imposed in Assumption 3.2 (ii) can be implied if, for instance, we assume $h(\delta, \alpha)$ is continuous and $\mathcal{D}$ and $\mathcal{A}$ are compact. The nondegeneracy condition of the benchmark conditional prior stated in Assumption 3.2 (iv) requires that $IS_\alpha(\phi)$ is non-singleton at ϕ_0 and on its neighborhood G_0 since otherwise $\pi^*_{\alpha|\phi}$ supported only on $IS_\alpha(\phi)$ must be a Dirac measure at the point-identified value of α. Assumption 3.2 (v) says that the benchmark conditional prior for α given ϕ is continuous at ϕ_0 in the total variation distance sense. When $\pi_{\alpha|\phi}$ supports the entire identified set $IS_\alpha(\phi)$, this assumption requires that $IS_\alpha(\phi)$ is a continuous correspondence at ϕ_0. This assumption also requires that any measures dominating $\pi^*_{\alpha|\phi_0}$ have to dominate $\pi^*_{\alpha|\phi}$ for ϕ in a neighborhood of ϕ_0, as otherwise $\left\|\pi^*_{\alpha|\phi} - \pi^*_{\alpha|\phi_0}\right\|_{TV} = 1$ holds for some $\phi \to \phi_0$. It hence rules out the cases such as (1) $IS_\alpha(\phi_0)$ is a singleton (i.e., $\pi^*_{\alpha|\phi_0}$ is the Dirac measure) while $IS_\alpha(\phi)$ has the nonempty interior with continuously distributed $\pi^*_{\alpha|\phi}$ for ϕ's in a neighborhood of ϕ_0, and (2) $\pi^*_{\alpha|\phi_0}$ and $\pi^*_{\alpha|\phi}$, $\phi \in G_0$ are discrete measures with different support points.[3] In addition, the differentiability of the prior means Assumption 3.2 (vi) imposes smoothness of the conditional average loss functions with respect to ϕ. Assumption 3.2 (vii) assumes that conditional on the true reduced-form parameter value $\phi = \phi_0$, the constrained minimax objective function is continuous in the action and has the unique optimal action.

Under these regularity assumptions, we obtain the following asymptotic result about convergence of the constrained posterior minimax decision.

Theorem 3.3 *(i) Let $\hat{\delta}_\lambda \in \arg\min_{\delta \in \mathcal{D}} \int_\Phi r_\lambda(\delta, \phi)\, d\pi_{\phi|X}$. Under Assumption 3.2,*

$$\hat{\delta}_\lambda \to \delta_\lambda(\phi_0) \equiv \arg\min_{\delta \in \mathcal{D}} r_\lambda(\delta, \phi_0),$$

as $n \to \infty$ for almost every sampling sequence.

(ii) Furthermore, for any $\hat{\phi}$ such that $\left\|\hat{\phi} - \phi_0\right\| \to_p 0$ as $n \to \infty$, $\delta_\lambda\left(\hat{\phi}\right) \in \arg\min_{\delta \in \mathcal{D}} r_\lambda\left(\delta, \hat{\phi}\right)$ converges in probability to $\delta_\lambda(\phi_0)$ as $n \to \infty$.

Proof. See Appendix A. ■

This theorem shows that the finite sample conditional gamma-minimax decision has a well-defined large sample limit that coincides with the optimal decision under the knowledge of the true value of ϕ. In other words, the posterior uncertainty of the reduced-form parameters vanishes in large samples and what matters asymptotically for the conditional gamma-minimax decision is the ambiguity of the unrevisable part of the prior (the conditional prior of α given ϕ) minimax posterior uncertainty. The theorem has a useful practical implication: when the sample size is moderate to large so that the posterior distribution of ϕ is concentrated around its maximum likelihood estimator (MLE) $\hat{\phi}_{ML}$, one can well approximate the exact finite sample minimax decision by minimizing the "plug-in" objective function, where the averaging with

[3] When a benchmark prior $\pi^*_{\alpha|\phi}$ is a probability mass measure selecting a point from $IS_y(\phi)$ for every ϕ (i.e., the benchmark prior delivers an additional restriction that makes the model point-identified), the optimal $\delta(x)$ is given by the Bayes action with respect to the single posterior of y induced by such benchmark prior irrespective of the value of κ. This implies that robust estimation via the multiplier minimax approach is not effective if the benchmark prior is chosen based on a point-identifying restriction.

respect to the posterior of ϕ in (18) is replaced by plugging $\hat{\phi}_{ML}$ in $r_\lambda(\delta, \phi)$. This will reduce computational cost of approximating the objective function since what we need in this case are only MCMC draws of θ (or α) from $\pi_{\theta|\hat{\phi}_{ML}}$ (or $\pi_{\alpha|\hat{\phi}_{ML}}$).

4 Range of Posteriors and Sensitivity Analysis

Our analysis so far focuses on minimax decision (estimation) with a given loss function $h(\delta, \alpha)$. The robust Bayes framework with multiple priors is also useful to compute the range of the posterior quantities (mean, median, probability) when the prior varies over the KL-neighborhood of the benchmark prior. The range of posterior quantities offers a useful tool for global sensitivity analysis. For instance, by reporting the range of posterior means, we can learn what posterior results are robust to what size of perturbations specified by λ to the prior input.

Given the posterior for ϕ and the class of priors $\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)$, the range of posterior means of $f(\alpha)$ is defined as

$$\left[\int_\Phi \min_{\pi_{\alpha|\phi} \in \Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left(\int f(\alpha) d\pi_{\alpha|\phi}\right) d\pi_{\phi|X}, \int_\Phi \max_{\pi_{\alpha|\phi} \in \Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left(\int f(\alpha) d\pi_{\alpha|\phi}\right) d\pi_{\phi|X}\right]. \quad (19)$$

For instance, the range of posterior means of α is obtained by setting $f(\alpha) = \alpha$, and the range of posterior probabilities on subset $B \subset \mathcal{B}$ is obtained by setting $f(\alpha) = 1\{\alpha \in B\}$. The range of posterior quantiles of α can be computed by inverting the range of cumulative distribution functions of the posterior of α. The optimization problems to derive the bounds (19) are identical to the inner maximization problems of the constrained minimax problem (14) with $h(\delta, \alpha)$ replaced by $f(\alpha)$ or $-f(\alpha)$.

Applying the expression of the worst-case expected loss shown in Theorem 3.1 to the current context, we obtain analytical expressions for the posterior ranges. By replacing $h(\delta, \alpha)$ in $r_\lambda(\delta, \phi)$ defined in equation (18) with $f(\alpha)$ or $-f(\alpha)$, the range of posterior means of $f(\alpha)$ can be expressed as

$$\left[\int_\Phi \left(\int f(\alpha) d\pi^\ell_{\alpha|\phi}\right) d\pi_{\phi|X}, \int_\Phi \left(\int f(\alpha) d\pi^u_{\alpha|\phi}\right) d\pi_{\phi|X}\right], \quad (20)$$

where $\pi^\ell_{\alpha|\phi}$ and $\pi^u_{\alpha|\phi}$ are the exponentially tilted conditional priors obtained as the solutions to the optimizations in (19):

$$d\pi^{\ell}_{\alpha|\phi} \equiv \frac{\exp\{-f(\alpha)/\kappa^{\ell}(\phi)\}}{\int \exp\{-f(\alpha)/\kappa^{\ell}(\phi)\}d\pi^*_{\alpha|\phi}} \cdot d\pi^*_{\alpha|\phi},$$
$$d\pi^{u}_{\alpha|\phi} \equiv \frac{\exp\{f(\alpha)/\kappa^{u}(\phi)\}}{\int \exp\{f(\alpha)/\kappa^{u}(\phi)\}d\pi^*_{\alpha|\phi}} \cdot d\pi^*_{\alpha|\phi},$$
$$\kappa^{\ell}(\phi) \equiv \arg\min_{\kappa\geq 0} \left\{\kappa \ln \int \exp\left\{\frac{-f(\alpha)}{\kappa}\right\} d\pi^*_{\alpha|\phi} + \kappa\lambda\right\},$$
$$\kappa^{u}(\phi) \equiv \arg\min_{\kappa\geq 0} \left\{\kappa \ln \int \exp\left\{\frac{f(\alpha)}{\kappa}\right\} d\pi^*_{\alpha|\phi} + \kappa\lambda\right\}$$

As we explain more details in Section 7 below, it is feasible to compute these quantities using the Monte Carlo draws of (α, ϕ) sampled from the benchmark conditional prior $\pi^*_{\alpha|\phi}$ and the posterior of ϕ.

Note that the range of prior means of $f(\alpha)$ is obtained similarly by replacing $\pi_{\phi|X}$ in (19) with the prior π_{ϕ}.

$$\left[\int_{\Phi}\left(\int f(\alpha)d\pi^{\ell}_{\alpha|\phi}\right)d\pi_{\phi}, \int_{\Phi}\left(\int f(\alpha)d\pi^{u}_{\alpha|\phi}\right)d\pi_{\phi}\right], \tag{21}$$

The range of prior mean or posterior for α (or other parameter) is a useful object for the purpose of eliciting a reasonable value of λ in light of the partial prior knowledge that the researcher has for $f(\alpha)$. In Section 6 below, we discuss how to elicit λ using the range of prior means of α.

5 Minimax Estimation with Large λ

This section presents further analytical results on the posterior minimax analysis for two common choices of statistical loss, the quadratic loss and the check loss. Maintaing the finite sample situation of Section 3.1, we first focus on the limiting situation of $\lambda \to \infty$, i.e., the case when the decision maker faces extreme ambiguity, and argues how the minimax estimator for α relates to the posterior distribution of the identified set $IS_{\alpha}(\phi)$, $\phi \sim \pi_{\phi|X}$.

5.1 Large λ Asymptotics in Finite Sample

Informally speaking, as $\lambda \to \infty$, the benchmark conditional prior $\pi^*_{\alpha|\phi}$ will affect the minimax decision only through its support, since $\Pi^{\infty}(\pi^*_{\alpha|\phi})$ consists of any prior that shares the same support as $\pi^*_{\alpha|\phi}$. To have precise characterization of this claim and formal investigation of the limiting behavior of $\hat{\delta}_\lambda$ as $\lambda \to \infty$, we impose the following regularity assumptions restricting the topological properties of $IS_\alpha(\phi)$ and the tail behavior of $\pi^*_{\alpha|\phi}$.

Assumption 5.1

(i) *$IS_\alpha(\phi)$ has a nonempty interior π_ϕ-a.s. and the benchmark prior marginalized to α, $\pi^*_{\alpha|\phi}$, is absolutely continuous with respect to the Lebesgue measure π_ϕ-a.s.*

(ii) *Let $[\alpha_*(\phi), \alpha^*(\phi)]$ be the convex hull of $\left\{\alpha : \frac{d\pi^*_{\alpha|\phi}}{d\alpha} > 0\right\}$. Assume $[\alpha_*(\phi), \alpha^*(\phi)]$ is a bounded interval, π_ϕ-a.s.*

(iii) *At π_ϕ-almost every ϕ, there exist $\eta > 0$ such that $[\alpha_*(\phi), \alpha_*(\phi) + \eta) \subset IS_\alpha(\phi)$ and $(\alpha^*(\phi) - \eta, \alpha^*(\phi)] \subset IS_\alpha(\phi)$ hold and the probability density function of $\pi^*_{\alpha|\phi}$ is real analytic on $[\alpha_*(\phi), \alpha_*(\phi) + \eta)$ and on $(\alpha^*(\phi) - \eta, \alpha^*(\phi)]$, i.e.,*

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha) = \sum_{k=0}^{\infty} a_k(\alpha - \alpha_*(\phi))^k \quad \text{for } \alpha \in [\alpha_*(\phi), \alpha_*(\phi) + \eta),$$
$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha) = \sum_{k=0}^{\infty} b_k(\alpha^*(\phi) - \alpha)^k \quad \text{for } \alpha \in (\alpha^*(\phi) - \eta, \alpha^*(\phi)].$$

(iv) *Let ϕ_0 be the true value of the reduced form parameters. Assume $\alpha_*(\phi)$ and $\alpha^*(\phi)$ are continuous in ϕ at $\phi = \phi_0$.*

Assumption 5.1 (i) rules out point-identified models, as in Assumption 3.2 (iv) and (vi). Assumption 5.1 (ii) assumes that the benchmark conditional prior has the bounded support π_ϕ-almost surely, which automatically holds if the identified set $IS_\alpha(\phi)$ is π_ϕ-almost surely bounded. In particular, if the benchmark conditional prior supports the entire identified set, i.e., $[\alpha_*(\phi), \alpha^*(\phi)]$ is the convex hull of $IS_\alpha(\phi)$, Assumption 5.1 (iii) imposes a mild restriction on the behavior of the benchmark conditional prior locally around the boundary points of the support. It requires that the benchmark conditional prior can be represented as polynomial series in a neighborhood of the support boundaries, where the neighborhood parameter η and the series coefficients can depend on the conditioning value of the reduced-form parameters ϕ. Assumption 5.1 (iv) will be imposed only in the large sample asymptotics of the next subsection. It implies that the support of the benchmark conditional prior varies continuously in ϕ.

The next theorem characterizes the asymptotic behavior of the conditional gamma-minimax decisions for the cases of quadratic loss and check loss, in the limiting situation of $\lambda \to \infty$ with a fixed sample size.

Theorem 5.2 *Suppose Assumption 5.1 (i) - (iii) hold.*

(i) When $h(\delta, \alpha) = (\delta - \alpha)^2$,

$$\lim_{\lambda \to \infty} \int_\Phi r_\lambda(\delta, \phi)\, d\pi_{\phi|X} = \int_\Phi \left[(\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2 \right] d\pi_{\phi|X}$$

holds whenever the right-hand side integral is finite.

(ii) When $h(\delta, \alpha) = \rho_\tau(\alpha - \delta)$,

$$\lim_{\lambda \to \infty} \int_\Phi r_\lambda(\delta, \phi)\, d\pi_{\phi|X} = \int_\Phi [(1-\tau)(\delta - \alpha_*(\phi)) \vee \tau(\alpha^*(\phi) - \delta)]\, d\pi_{\phi|X}$$

holds, whenever the right-hand side integral is finite.

Proof. See Appendix A. ■

Theorem 5.2 shows that in the most ambiguous situation of $\lambda \to \infty$, only the boundaries of the support of the benchmark prior, $[\alpha_*(\phi), \alpha^*(\phi)]$, and the posterior of ϕ matter for the conditional gamma-minimax decision. Other than the tails, the specific shape of $\pi^*_{\alpha|\phi}$ is irrelevant for the minimax decision. This result is intuitive since larger λ implies a larger class of priors and at the limit $\lambda \to \infty$, any priors that share the support with the benchmark prior are included in the prior class, and the worst-case conditional prior in the limit is reduced to the point-mass prior that assignes probability one to $\alpha_*(\phi)$ or $\alpha^*(\phi)$, the furthest point from δ.

5.2 Large Sample Analysis with Large λ

The next theorem 5.3 concerns the large sample ($n \to \infty$) asymptotics with large λ.

Theorem 5.3 *Suppose Assumption 3.2 and Assumption 5.1 hold. Let*

$$\hat{\delta}_\infty = \arg\min_{\delta \in \mathcal{D}} \left\{ \lim_{\lambda \to \infty} \int_\Phi r_\lambda(\delta, \phi)\, d\pi_{\phi|X}(\phi) \right\}$$

be the conditional gamma-minimax estimator in the limiting case $\lambda \to \infty$.

(i) When $h(\delta, \alpha) = (\delta - \alpha)^2$, $\hat{\delta}_\infty \to \frac{1}{2}(\alpha_*(\phi_0) + \alpha^*(\phi_0))$ *as the sample size* $n \to \infty$ *for almost every sampling sequence.*

(ii) When $h(\delta, \alpha) = \rho_\tau(\alpha - \delta)$, $\rho_\tau(u) = \tau u \cdot 1\{u > 0\} - (1-\tau) u \cdot 1\{u < 0\}$, $\hat{\delta}_\infty \to (1-\tau)\alpha_*(\phi_0) + \tau \alpha^*(\phi_0)$ *as the sample size* $n \to \infty$ *for almost every sampling sequence.*

Proof. See Appendix A. ■

Theorem 5.3 (i) shows that in the large sample situation, the minimax decision with the quadratic loss converges to the middle point of the boundary points of the support of the benchmark prior evaluated at the true reduced form parameters. When the benchmark prior supports the entire identified set, this means that the minimax decision at the limit is to report the central point of the true identified set. When the loss is the check function associated with the τ-th quantile, the minimax decision at the limit is given by the convex combination of the same boundary points with weights τ and $1-\tau$. One useful implication of this result is that, in case of the check loss, solving for the optimal δ can be seen as obtaining the robust posterior τ-th quantile of α, and the optimal δ may be used to construct a robust interval estimate for α that explicitly incorporates the ambiguous beliefs about the benchmark prior.

Another implication of Theorem 5.3 (ii) is that the limit of the minimax quantile estimator $\delta^*(\tau)$ always lies in the true identified set for any τ, even in the most conservative case, $\kappa \to 0$. This means that, if we use $[\delta^*(0.05), \delta^*(0.95)]$ as a robustified posterior credibility interval for α, this interval estimate will be asymptotically strictly narrower than the frequentist confidence interval for α, as $[\delta^*(0.05), \delta^*(0.95)]$ is contained in the true identified set asymptotically. This result is similar to the finding in Moon and Schorfheide (2012) for the single posterior Bayesian credible interval.

The asymptotic results of Theorems 5.2 and 5.3 assume that the benchmark prior is absolutely continuous with respect to the Lebesgue measure. We can instead consider a setting where the benchmark prior is given by a nondegenerate probability mass measure, which can naturally arise if the benchmark prior comes from a weighted combination of multiple point-identified models. This case leads to asymptotic results similar to Theorem 5.3. We present a formal analysis in such discrete benchmark prior setting in Appendix B.

6 Eliciting Benchmark Prior and λ

To implement our robust estimation and inference procedures, the key inputs that the researcher has to specify are the benchmark conditional prior $\pi^*_{\theta|\phi}$ and the radius parameter for the KL-neighborhood. To be specific, we focus on the simultaneous equation model of demand and supply introduced in Example 1.1, and illustrate how to choose them in that context.

6.1 Benchmark Prior

Our construction of the prior class takes the conditional prior of θ given ϕ as the benchmark prior $\pi^*_{\theta|\phi}$. The benchmark prior should represent or be implied by a probabilistic belief that is reasonable and most credible. Depending on what parametrization facilitates elicitation process, we can form the benchmark prior directly through the conditional prior $\theta|\phi$ and the single prior of ϕ, or alternatively construct a prior for $\tilde{\theta}$, where $\tilde{\theta}$ is a one-to-one parametrization of (θ, ϕ) and it is typically a vector of structural parameters in our examples. In what follows, we apply the latter approach to the simultaneous equation model of Example 1.1, as we think eliciting the benchmark prior for demand and supply elasticities is easier than the former approach of eliciting the reduced form parameters and the conditional prior of $\theta|\phi$, separately.

Let us denote the full vector of the structural parameters by $\tilde{\theta} = (\beta_s, \beta_d, d_1, d_2)$, and denote its prior probability density by $\frac{d\pi^*_{\tilde{\theta}}}{d\tilde{\theta}}(\beta_s, \beta_d, d_1, d_2)$. As in Leamer (1981) and Baumeister and Hamilton (2015), it is natural to impose the sign restrictions for the slopes of supply and demand equations, $\beta_s \geq 0$, $\beta_d \leq 0$. These restrictions can be incorporated by trimming the support of $\pi^*_{\tilde{\theta}}$ such that $\pi^*_{\tilde{\theta}}(\{\beta_s \geq 0, \beta_d \leq 0\}) = 1$. In the context of aggregate labor supply and demand, Baumeister and Hamilton (2015) specify the prior of $\tilde{\theta}$ as the product of the independent truncated Student t distributions for (β_s, β_d) and the independent inverse gamma distributions for $(d_1, d_2)|(\beta_s, \beta_d)$, of which the hyperparameters are chosen through careful elicitation.

Having specified the prior for $\tilde{\theta}$ and setting $\alpha = \beta_s$ as a parameter of interest, let us discuss how to obtain the benchmark conditional prior of α given $\phi = (\omega_{11}, \omega_{12}, \omega_{22})$. The benchmark conditional prior of α given ϕ can be derived by reparametrizing $\tilde{\theta}$ to (α, ϕ). Since $\Omega = A^{-1}D\left(A^{-1}\right)'$, we have that

$$\omega_{11} = \frac{d_1 + d_2}{(\alpha - \beta)^2}, \quad \omega_{12} = \frac{\alpha d_1 + \beta d_2}{(\alpha - \beta)^2}, \quad \omega_{22} = \frac{\alpha^2 d_1 + \beta^2 d_2}{(\alpha - \beta)^2} \tag{22}$$

which implies the following mapping between $(\alpha, \beta, d_1, d_2)$ and $(\alpha, \omega_{11}, \omega_{12}, \omega_{22})$:

$$\begin{aligned} \alpha &= \alpha, \\ \beta &= \frac{\alpha\omega_{12} - \omega_{22}}{\alpha\omega_{11} - \omega_{12}} \equiv \beta(\alpha, \phi), \\ d_1 &= \omega_{11}\left(\alpha - \frac{\alpha\omega_{12} - \omega_{22}}{\alpha\omega_{11} - \omega_{12}}\right)^2 - \alpha^2\omega_{11} + 2\alpha\omega_{12} - \omega_{22} \equiv d_1(\alpha, \phi), \\ d_2 &= \alpha^2\omega_{11} - 2\alpha\omega_{12} + \omega_{22} \equiv d_2(\alpha, \phi). \end{aligned} \tag{23}$$

Since the conditional prior $\pi_{\alpha|\phi}$ is proportional to the the joint prior of (α, ϕ), the benchmark conditional prior $\pi^*_{\alpha|\phi}$ satisfies

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha|\phi) \propto \pi_{\tilde{\theta}}\left(\alpha, \beta\left(\alpha, \phi\right), d_1\left(\alpha, \phi\right), d_2\left(\alpha, \phi\right)\right) \times \left|\det\left(J\left(\alpha, \phi\right)\right)\right|, \tag{24}$$

where $J(\alpha, \phi)$ is the Jacobian of the mapping (23), and $|\det(\cdot)|$ is the absolute value of the determinant. This benchmark conditional prior supports the entire identified set $IS_\alpha(\phi)$ if $\pi_{\tilde{\theta}}(\cdot)$ supports any value of (α, β) satisfying the sign restrictions. An analytical expression of the posterior of ϕ could be obtained by integrating out α in the right-hand side of (24). Even if the analytical expression of the posterior of ϕ were not easy to derive, it would be easy to obtain posterior draws of ϕ by transforming the posterior draws of $\tilde{\theta}$ according to $\Omega = A^{-1}D\left(A^{-1}\right)'$.

If ϕ involves nonlinear transformations of $\tilde{\theta}$, a diffuse prior of $\tilde{\theta}$ can imply an informative prior of ϕ. If so, in finite samples, this can downplay the sample information for ϕ by distorting the shape of the likelihood of ϕ. Given that our analysis can be motivated by a robustness concern about the choice of prior of $\tilde{\theta}$, one may not want to force the prior of ϕ to be informative as a result of transformation. Such concern might make the following hybrid approach attractive. Draw ϕ's from the posterior of ϕ obtained from a non-informative prior of ϕ (e.g., Jeffreys' prior) that is not compatible with the prior of $\tilde{\theta}$. On the other hand, the prior of $\tilde{\theta}$ is used solely to construct the benchmark conditional prior of $\pi^*_{\alpha|\phi}$ via (24).

6.2 Robustness Parameter λ

The radius of the KL-neighborhood $\lambda \geq 0$ is an important prior input that directly controls the size of ambiguity in the robust Bayes analysis. Its elicitation should be based on the degree of confidence or fear of misspecification that the analyst has on the benchmark prior. Since λ itself does not have interpretative scale, it is necessary to map it into some prior quantity that the analyst can easily interpret and work with for eliciting λ.

Along this line, we propose to elicit λ by mapping it into the range of *prior* means and finding the value of λ such that the implied prior range matches best to available (partial) prior knowledge. Thanks to the invariance of λ with respect to reparametrization or marginalization of as shown in Lemma 2.1, we can focus on one or a few of the subset parameters in $\tilde{\theta}$ or transformations of (θ, ϕ) to match the range of priors.

To be specific, let $\tilde{\alpha} = \tilde{\alpha}(\theta, \phi)$ be a scalar parameter for which the analyst can feasibly assess the range of its prior beliefs. Depending on applications, it can be different from the parameter of interest $\alpha = \alpha(\theta, \phi)$. Given the benchmark conditional prior $\pi_{\tilde{\alpha}|\phi}$ and the single prior π_ϕ, for each candidate choice of λ, we can compute the range of prior means of $f(\tilde{\alpha})$

by applying the expression (21). We then select λ that seems to best match with available but imprecise prior knowledge about $f(\tilde{\alpha})$. We implement this way of eliciting λ in the SVAR example of Section 7.

7 Empirical Example

This section applies our robust Bayesian procedure to the bivariate SVAR for the dynamic labor supply and demand in the US analysed in Baumeister and Hamilton (2015). We use the data available in the supplementary material of Baumeister and Hamilton (2015). The endogenous variables are the growth rate of total US employment Δn_t and the growth rate of hourly real compensation Δw_t, $x_t = (\Delta n_t, \Delta w_t)$. The observations are quarterly, and the sample covers $t =$1970:Q1–2014:Q4.

7.1 Specification and Parametrization

The model is a bivariate SVAR of with eight lags $L = 8$.

$$A_0 x_t = c + \sum_{l=1}^{L} A_l x_{t-1} + u_t. \quad u_t \sim \mathcal{N}(0, D),$$

where $A_0 = \begin{bmatrix} -\beta_d & 1 \\ -\beta_s & 1 \end{bmatrix}$ and $D = diag(d_1, d_2)$ as defined in Example 1.1. The reduced form VAR is

$$x_t = b + \sum_{l=1}^{L} B_l x_{t-1} + \epsilon_t,$$

where $b = A_0^{-1} c$ and $B_l = A_0^{-1} A_l$. The reduced form parameters are $\phi = (\Omega, B)$, $B = (b, B_1, \ldots, B_L)$, and the full vector of structural parameters are $(\beta_s, \beta_d, d_1, d_2, A)$, $A = (c, A_1, \ldots, A_L)$.

We set the supply elasticity as the parameter of interest, $\alpha = \beta_s$, so that we express the full vector of structural parameters as $\tilde{\theta} = (\alpha, \beta_d, d_1, d_2, A)$. The mapping between $\tilde{\theta}$ and (α, ϕ) consists of those shown in (23) and

$$A = A_0(\alpha, \phi) B \equiv A(\alpha, \phi), \tag{25}$$

where $A_0(\alpha, \phi) = \begin{bmatrix} -\beta_d(\alpha, \phi) & 1 \\ -\alpha & 1 \end{bmatrix}$. Hence, if the benchmark prior is specified in terms of $\tilde{\theta}$, the conditional benchmark prior of α given ϕ is given by

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha|\phi) \propto \pi_{\tilde{\theta}}(\alpha, \beta_d(\alpha, \phi), d_1(\alpha, \phi), d_2(\alpha, \phi), A(\alpha, \phi)) \times |\det(J(\alpha, \phi))|, \tag{26}$$

where $\pi_{\tilde{\theta}}(\alpha, \beta_d, d_1, d_2, A)$ is a prior distribution for $\tilde{\theta}$ that induces the benchmark conditional prior of $\alpha|\phi$ and the single prior for ϕ, and $J(\alpha, \phi)$ is the Jacobian of the transformations (23) and (25).

7.2 Benchmark Prior

We construct the benchmark conditional prior $\pi^*_{\alpha|\phi}$ by setting the prior of $\tilde{\theta}$ to the one used in Baumeister and Hamilton (2015) and applying the formula (26). Decomposing a prior of $\tilde{\theta}$ as

$$\pi_{\tilde{\theta}} = \pi_{(\alpha,\beta_d)} \cdot \pi_{(d_1,d_2)|(\alpha,\beta_d)} \cdot \pi_{A|(d_1,d_2,\alpha,\beta_d)},$$

Baumeister and Hamilton (2015) recommends to elicit each of the components carefully by spelling out the class of structural models that the researcher has in mind and/or referring to the existing studies providing prior evidence about these parameters. In the current context of labor supply and demand, the prior elicitation process of Baumeister and Hamilton (2015) is summarized as follows

1. Elicitation of $\pi_{(\alpha,\beta_d)}$: Independent truncated t-distributions are used as priors for α and β_d, where the truncations incorporate the dogmatic form of sign restrictions; with prior probability one, $\alpha \geq 0$ and $\beta_d \leq 0$. Their hyperparameters are chosen based on meta-analysis of microeconomic and macroeconomic studies estimating the labor supply and demand elasticities. Specifically, the meta-analysis of Baumeister and Hamilton (2015) identifies interval $\alpha \in [0.1, 2.2]$ and $\beta_d \in [-2.2, -0.1]$, and they accordingly tune the hyperparameters of the t-distribution so as to meet $\pi_\alpha([0.1, 2.2]) = 0.9$ and $\pi_{\beta_d}([-2.2, -0.1]) = 0.9$.

2. Elicitation of $\pi_{(d_1,d_2)|(\alpha,\beta_d)}$: Independent natural conjugate priors (inverse gamma family) are specified for d_1 and d_2. To reflect the scale of the data in the choice of hyperparameters, set the prior means to $A_0\hat{\Omega}A_0'$, where $\hat{\Omega}$ is the maximum likelihood estimate of the reduced-form error variances $E(\epsilon_t\epsilon_t')$.

3. Elicitation of $\pi_{A|(d_1,d_2,\alpha,\beta_d)}$: Since the reduced-form VAR coefficients satisfy $B = A_0^{-1}A$, elicitation of the conditional prior of A given $(\alpha, \beta_d, d_1, d_2)$ can be facilitated by available prior knowledge about the reduced-form VAR coefficients. Prior choice for the reduced-form VAR parameters is well studied in the literature as in Doan et al. (1984) and Sims and Zha (1998). Building on their proposals, Baumeister and Hamilton (2015) specifies a prior of B as a multivariate normal with prior means corresponding to $(\Delta n_t, \Delta w_t)$ being

the independent random walk processes. The prior variances are diffuse for short lagged coefficients and become tighter for the longer lagged coefficients.

As is evident from this description elicitation process, available prior evidence or knowledge suggests only vague restrictions to the prior and certainly is not precise enough to pin down the exact shape of a prior distribution. For instance, the elicitation of α and β_d relies on a belief that some large amount of prior probability should be assigned to the ranges identified by the meta-study, but it cannot pin down the shape of the priors to the t-distributions. In the step of eliciting (d_1, d_2) and A, the available prior knowledge suggest reasonable location and scale of the prior distributions, but the reasoning for determining the shape of the prior comes from computational convenience. In addition, the independence of the priors invoked in all of the steps is convenient and simplifies the construction of the prior but is not innocuous. It is important to be aware that prior independence among the parameters does not represent the lack of knowledge about the prior dependence among them.

The issues raised here apply to many contexts of Bayesian analysis, and they can be a source of robustness concerns about the posterior results. The situation is worse in the set-identified models, since robustness concerns are magnified due to the lack of identification. Our robust Bayes proposal can accommodate such robustness concern by forming the benchmark conditional prior of $\pi^*_{\alpha|\phi}$ from the elicited prior of $\tilde{\theta}$ following (26), setting a KL-neighborhood around it, and performing robust Bayes inference and/or a conditional gamma-minimax decision that is robust to misspecification of the unrevisable part of the prior within the KL-neighborhood.

7.3 Computation

The prior of $\tilde{\theta}$ constructed in the manner of induces the prior of ϕ, while the prior of $\tilde{\theta}$ specified in Baumeister and Hamilton (2015) does not lead to a conjugate prior of ϕ in the reduced-form VAR. Obtaining an analytical expression of the posterior of ϕ is therefore not simple, while the posterior draws of ϕ are easy to obtain by transforming the posterior draws of $\tilde{\theta}$ obtained, for instance, by the random-walk Metropolis-Hastings algorithm as in Baumeister and Hamilton (2015). We hereafter denote the posterior draws of ϕ by $(\phi_1, \ldots, \phi_M)$.

An algorithm that approximates the objective function in (18) is as follows.

Algorithm 7.1 *Let posterior draws of ϕ, $(\phi_1, \ldots, \phi_M)$, be given.*

1. *For each $m = 1, \ldots, M$, we approximate $r_\lambda(\delta, \phi_m) = \ln \int_{IS_\alpha(\phi_m)} \exp\{h(\delta, \alpha)/\kappa(\delta, \phi)\} d\pi^*_{\alpha|\phi}$ by importance sampling, i.e., draw N draws of α, $(\alpha_{m1}, \ldots, \alpha_{mN})$ from a proposal distri-*

bution (probability density) $\tilde{\pi}_{\alpha|\phi}(\alpha|\phi)$ *(e.g., the uniform distribution on* $IS_\alpha(\phi_m)$*) and compute*

$$\hat{r}_\lambda(\delta, \phi_m) = \frac{\sum_{i=1}^{N} w(\alpha_{mi}, \phi_m) h(\delta, \alpha_m) \exp\{h(\delta, \alpha_{mi})/\kappa(\delta, \phi_m)\}}{\sum_{i=1}^{N} w(\alpha_{mi}, \phi_m) \exp\{h(\delta, \alpha_{mi})/\kappa(\delta, \phi_m)\}},$$

where

$$w(\alpha_{mi}, \phi_m) = \frac{\frac{d\pi_{\tilde{\theta}}}{d\alpha}(\alpha_{mi}, \beta_d(\alpha_{mi}, \phi_m), d_1(\alpha_{mi}, \phi_m), d_2(\alpha_{mi}, \phi_m)) \times |\det(J(\alpha_{mi}, \phi_m))|}{\frac{d\tilde{\pi}_{\alpha|\phi}}{d\alpha}(\alpha_{mi}|\phi_m)}.$$

2. *We then approximate the objective function of the multiplier minimax problem by*

$$\frac{1}{M}\sum_{m=1}^{M} \hat{r}_\lambda(\delta, \phi_m), \tag{27}$$

and minimize it with respect to δ*.*

If the limiting case $\lambda \to \infty$ is considered (either with a quadratic or check loss), Theorem 5.2 implies that Step 1 of Algorithm 7.1 can be skipped and we can directly approximate the objective function to be minimized in δ by

$$\frac{1}{M}\sum_{m=1}^{M} \left[(\delta - \underline{\alpha}(\phi_m))^2 \vee (\delta - \bar{\alpha}(\phi_m))^2\right]$$

for the quadratic loss case, where $[\underline{\alpha}(\phi_m), \bar{\alpha}(\phi_m)]$ is the lower and upper bounds of the identified set of α if $\pi_{\alpha|\phi}$ supports the entire identified set.

If one is interested in the large sample approximation of the worst-case posterior expected loss, Theorem 3.3 (ii) justifies to replace the objective function (27) with $\hat{r}_\lambda\left(\delta, \hat{\phi}_{ML}\right)$, where $\hat{\phi}_{ML}$ is the maximum likelihood estimator of ϕ.

8 Concluding Remarks

This paper proposes a robust Bayes analysis in the class of set-identified models. The class of priors considered is formed by the KL-neighborhood of a benchmark prior. This way of constructing the class of prior distinguishes the current paper from Giacomini and Kitagawa (2018) and Giacomini et al. (2018). We show how to formulate and solve the conditional gamma-minimax problem, and investigate its analytical properties in finite and large samples.

We illustrate a use of our robust Bayes methods in the SVAR anaysis of Baumeister and Hamilton (2015).

When performing the gamma-minimax analysis, there is no consensus about whether we should condition on the data or not. We perform the conditional gamma-minimax analysis mainly due to analytical and computational tractability, and we do not intend to settle this open question. In fact, compared with the unconditonal gamma-minimax decision, less is known about statistical admissibility of the conditional one. As DasGupta and Studden (1989) argues, the conditional gamma-minimax can often lead to a reasonable estimator with good frequentist performance. It remains to be seen further decision-theoretic justification of the conditional gamma-minimax decision such as admissibility in the current setting of the set-identified models.

Appendix

A Lemmas and Proofs

This appendix collects lemmas and proofs that are omitted from the main text.

A.1 Proof of Lemma 2.1

The next set of lemmas are used to prove Lemma 2.1. Lemma A.1 derives a general formula that links the KL-distance of the probability distributions for θ and the KL-distance of the probability distributions for the transformation of θ to lower-dimensional parameter α. Lemma A.2 shows the inclusion relationship between the KL-neighborhood of $\pi^*_{\alpha|\phi}$ and the projection of the KL-neighborhood of $\pi^*_{\theta|\alpha}$ onto the space of α-marginals. Lemma 2.1 in the main text then follows as a corollary of these two lemmas.

Lemma A.1 *Given ϕ, let $\pi^*_{\alpha|\phi}$ be the marginal distribution for α induced from $\pi^*_{\theta|\phi}$ that has a dominating measure, and $\pi_{\alpha|\phi}$ be the marginal distribution for α induced from $\pi_{\theta|\phi}$. It holds*

$$\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) = -\int_{IS_\alpha(\phi)} \mathcal{R}(\pi_{\theta|\alpha\phi}\|\pi^*_{\theta|\alpha\phi})d\pi_{\alpha|\phi} + \mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi}), \tag{28}$$

where $\pi_{\theta|\alpha\phi}$ is the conditional distribution of θ given (α,ϕ) whose support is contained in $\Theta(\alpha,\phi) \equiv \{\theta\in\Theta : \alpha = \alpha(\theta,\phi)\}$, and $\mathcal{R}(\pi_{\theta|\alpha\phi}\|\pi^*_{\theta|\alpha\phi}) = \int_{\Theta(\alpha,\phi)} \ln\left(\frac{d\pi_{\theta|\alpha\phi}}{d\pi^*_{\theta|\alpha\phi}}\right) d\pi_{\theta|\alpha\phi} \geq 0$. Accordingly, $\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) \leq \mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi})$ holds. In particular, $\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) = \mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi})$ if and only if $\pi_{\theta|\alpha\phi} = \pi^*_{\theta|\alpha\phi}$, $\pi^*_{\alpha|\phi}$-almost surely.

Proof. We denote the densities of $\pi_{\theta|\phi}$ and its α-marginal distribution $\pi_{\alpha|\phi}$ (with respect to their dominating measures) by $\frac{d\pi_{\alpha|\phi}(\alpha)}{d\alpha}$ and $\frac{d\pi_{\theta|\phi}(\alpha)}{d\theta}$. Note they satisfy $\frac{d\pi_{\alpha|\phi}(\alpha)}{d\alpha} = \int_{\Theta(\alpha,\phi)} \frac{d\pi_{\theta|\phi}(\alpha)}{d\theta} d\theta$. Hence,

$$
\begin{aligned}
\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) &= \int_{IS_\alpha(\phi)} \ln\left(\frac{d\pi_{\alpha|\phi}}{d\pi^*_{\alpha|\phi}}\right)\left(\int_{\Theta(\alpha,\phi)} d\pi_{\theta|\phi}\right) d\alpha \\
&= \int_{IS_\alpha(\phi)} \left[\int_{\Theta(\alpha,\phi)} \ln\left(\frac{d\pi_{\alpha|\phi}}{d\pi^*_{\alpha|\phi}}\right) d\pi_{\theta|\phi}\right] d\alpha \\
&= \int_{IS_\alpha(\phi)} \int_{\Theta(\alpha,\phi)} \left[\ln\left(\frac{d\pi_{\alpha|\phi}}{d\pi_{\theta|\phi}} \cdot \frac{d\pi^*_{\theta|\phi}}{d\pi^*_{\alpha|\phi}}\right) + \ln\left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}}\right)\right] d\pi_{\theta|\phi} d\alpha \\
&= \int_{IS_\alpha(\phi)} \left[\int_{\Theta(\alpha,\phi)} \left[\ln\left(\frac{d\pi_{\alpha|\phi}}{d\pi_{\theta|\phi}} \cdot \frac{d\pi^*_{\theta|\phi}}{d\pi^*_{\alpha|\phi}}\right)\right] d\pi_{\theta|\alpha\phi}\right] d\pi_{\alpha|\phi} + \int_{IS_\theta(\phi)} \ln\left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}}\right) d\pi_{\theta|\phi} \\
&= \int_{IS_\alpha(\phi)} \left[\int_{\Theta(\alpha,\phi)} \left[\ln\left(\frac{d\pi_{\alpha|\phi}/d\alpha}{d\pi_{\theta|\phi}/d\theta} \cdot \frac{d\pi^*_{\theta|\phi}/d\theta}{d\pi^*_{\alpha|\phi}/d\alpha}\right)\right] d\pi_{\theta|\alpha\phi}\right] d\pi_{\alpha|\phi} + \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}),
\end{aligned}
$$

where the second term in the fourth line uses $\int_{IS_\alpha(\phi)} \int_{\Theta(\alpha,\phi)} f(\theta) d\pi_{\theta|\phi} d\alpha = \int_{IS_\theta(\phi)} f(\theta) d\pi_{\theta|\phi}$ for any measurable function $f(\theta)$. Since

$$
\frac{d\pi_{\theta|\alpha\phi}}{d\theta} = \left(\int_{\Theta(\alpha,\phi)} \frac{d\pi_{\theta|\phi}}{d\theta} d\theta\right)^{-1} \left(\frac{d\pi_{\theta|\phi}}{d\theta}\right) = \left(\frac{d\pi_{\alpha|\phi}}{d\alpha}\right)^{-1} \left(\frac{d\pi_{\theta|\phi}}{d\theta}\right)
$$

holds for $\theta \in \Theta(\alpha, \phi)$, we obtain

$$
\begin{aligned}
\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) &= -\int_{IS_\alpha(\phi)} \int_{\Theta(\alpha,\phi)} \ln\left(\frac{d\pi_{\theta|\alpha\phi}}{d\pi^*_{\theta|\alpha\phi}}\right) d\pi_{\theta|\alpha\phi} d\pi_{\alpha|\phi} + \int_{IS_\theta(\phi)} \ln\left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}}\right) d\pi_{\theta|\phi} \\
&= -\int_{IS_\alpha(\phi)} \mathcal{R}(\pi_{\theta|\alpha\phi} \| \pi^*_{\theta|\alpha\phi}) d\pi_{\alpha|\phi} + \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}).
\end{aligned}
$$

Since $\mathcal{R}(\pi_{\theta|\alpha\phi} \| \pi^*_{\theta|\alpha\phi}) \geq 0$, $\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi})$ holds. This inequality holds with equality if and only if $\int_{IS_\alpha(\phi)} \mathcal{R}(\pi_{\theta|\alpha\phi} \| \pi^*_{\theta|\alpha\phi}) d\pi_{\alpha|\phi} = 0$. That is, $\pi_{\theta|\alpha\phi} = \pi^*_{\theta|\alpha\phi}$ for $\pi_{\alpha|\phi}$-almost surely, or equivalently $\pi^*_{\alpha|\phi}$-almost surely as $\pi_{\alpha|\phi}$ is dominated by $\pi^*_{\alpha|\phi}$. ■

Lemma A.2 *Let ϕ and $\lambda \geq 0$ be given. Consider the set of α-marginal distributions constructed by marginalizing $\pi_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)$ to α,*

$$
\bar{\Pi}^\lambda \equiv \left\{\pi_{\alpha|\phi} : \pi_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)\right\}.
$$

On the other hand, for α-marginal of $\pi^*_{\theta|\phi}$, $\pi^*_{\alpha|\phi}$, define its KL-neighborhood with radius λ,

$$
\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right) = \left\{\pi_{\alpha|\phi} : \mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \lambda\right\}.
$$

Then, $\bar{\Pi}^{\lambda} \subset \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)$.

Proof. By Lemma A.1, $\pi_{\alpha|\phi} \in \bar{\Pi}^{\lambda}$ implies $\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \lambda$. Hence, $\bar{\Pi}^{\lambda} \subset \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)$. ■

Lemma A.3 *Let ϕ, δ, and $\lambda \geq 0$ be given. For any loss function $h(\delta, \alpha(\theta, \phi))$, it holds*

$$\max_{\pi_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_{\theta}(\phi)} h(\delta, \alpha(\theta, \phi)) d\pi_{\theta|\phi}\right] = \max_{\pi_{\alpha|\phi} \in \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_{\alpha}(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right].$$

Proof of Lemma 2.1. At fixed ϕ, $h(\delta, \alpha(\theta, \phi))$ depends on θ only through $\alpha(\cdot, \phi)$. Hence,

$$\begin{aligned} \max_{\pi_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_{\theta}(\phi)} h(\delta, \alpha(\theta, \phi)) d\pi_{\theta|\phi}\right] &= \max_{\pi_{\alpha|\phi} \in \bar{\Pi}^{\lambda}} \left[\int_{IS_{\theta}(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right] \\ &\leq \max_{\pi_{\alpha|\phi} \in \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_{\alpha}(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right], \end{aligned}$$

where the inequality follows by Lemma A.2. To show the reverse inequality, let $\pi^0_{\alpha|\phi}$ be a solution of $\max_{\pi_{\alpha|\phi} \in \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_{\alpha}(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right]$ and construct the conditional distribution of θ given ϕ by

$$\pi^0_{\theta|\phi} = \int_{IS_{\alpha}(\phi)} \pi^*_{\theta|\alpha\phi} d\pi^0_{\alpha|\phi},$$

where $\pi^*_{\theta|\alpha\phi}$ is the conditional distribution of θ given (α, ϕ) induced by $\pi^*_{\theta|\phi}$. Clearly, $\pi^0_{\theta|\phi}$ shares the conditional distribution of θ given (α, ϕ) with $\pi^*_{\theta|\phi}$, so that by Lemma A.1, $\mathcal{R}(\pi^0_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \mathcal{R}(\pi^0_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \lambda$. Hence, $\pi^0_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ holds, and this implies

$$\max_{\pi_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_{\theta}(\phi)} h(\delta, \alpha(\theta, \phi)) d\pi_{\theta|\phi}\right] \geq \max_{\pi_{\alpha|\phi} \in \Pi^{\lambda}\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_{\alpha}(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right].$$

■

A.2 Proof of Theorem 3.1

Proof of Theorem 3.1. Let ϕ and $\delta = \delta(x)$ be fixed. Let $\kappa_{\lambda}(\delta, \phi)$ be as defined in Lemma 2.2. Since $\kappa_{\lambda}(\delta, \phi)$ does not depend on $\pi_{\alpha|\phi}$, we treat $\kappa^* \equiv \kappa_{\lambda}(\delta, \phi)$ as a constant, and let us focus on solving the inner maximization problem in the multiplier minimax problem (15).

We first consider the case where $\pi^*_{\alpha|\phi}$ is a discrete probability mass measure with m support points $(\alpha_1, \ldots, \alpha_m)$ in $IS_{\alpha}(\phi)$. Since the KL-distance $\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi})$ is positive infinity unless

$\pi_{\alpha|\phi}$ is absolutely continuous with respect to $\pi^*_{\alpha|\phi}$, we can restrict our search of the optimal $\pi_{\alpha|\phi}$ to those whose support points (the set of points that receiving positive probabilities according to $\pi^*_{\alpha|\phi}$) are constrained to $(\alpha_1, \ldots, \alpha_m)$. Accordingly, let us denote a discrete $\pi_{\alpha|\phi}$ and the discrete loss by

$$g_i \equiv \pi_{\alpha|\phi}(\alpha_i), \quad f_i \equiv \pi^*_{\alpha|\phi}(\alpha_i), \;\; h_i = h(\delta, \alpha_i), \quad \text{for } i = 1, \ldots, m. \tag{29}$$

Then, the inner maximization problem of (15) can be written as

$$\begin{aligned} & \max_{g_1,\ldots,g_m} \sum_{i=1}^{m} h_i g_i - \kappa^* \sum_{i=1}^{m} g_i \ln\left(\frac{g_i}{f_i}\right), \qquad (30) \\ \text{s.t.} \qquad & \sum_{i=1}^{m} g_i = 1. \end{aligned}$$

With the Lagrangian multiplier ξ, the first order conditions in g_i are obtained as

$$h_i + \kappa^* \ln f_i - \kappa^* - \kappa^* \ln g_i - \xi = 0. \tag{31}$$

If $\kappa^* = 0$, $h_i = \xi$ for all i, which contradicts the assumption of non-degeneracy of $h(\delta, \alpha)$. Hence, $\kappa^* > 0$. Accordingly, these first order conditions lead to

$$g_i = \frac{f_i \exp\left(h_i/\kappa^*\right)}{\exp(1 + \zeta/\kappa^*)}.$$

$\sum_{j=1}^{m} g_j = 1$ pins down $\exp(1 + \xi/\kappa^*) = \sum_{j=1}^{m} f_j \exp(h_j/\kappa^*)$, so the optimal g_i is obtained as

$$g_i^* = \frac{f_i \exp\left(h_i/\kappa^*\right)}{\sum_{j=1}^{m} f_j \exp(h_j/\kappa^*)}. \tag{32}$$

Plugging this back into the objective function, we obtain

$$\begin{aligned} & \kappa^* \sum_{i=1}^{m} \left[\frac{f_i \exp(h_i/\kappa^*)}{\sum_{j=1}^{n} f_i \exp(h_i/\kappa^*)} \ln\left(\sum_{j=1}^{m} f_j \exp(h_j/\kappa^*)\right)\right] \qquad (33) \\ = \;\; & \kappa^* \ln\left(\sum_{j=1}^{m} f_j \exp(h_j/\kappa^*)\right), \end{aligned}$$

which is equivalent to $\kappa^* \ln\left(\int_{IS_\alpha(\phi)} \exp\left(h(\delta(x), \alpha)/\kappa^*\right) d\pi^*_{\alpha|\phi}\right)$ with discrete $\pi^*_{\alpha|\phi}$.

We generalize the claim to arbitrary $\pi^*_{\alpha|\phi}$. Based on an analogy to the optimal g_i obtained in (32), we guess that $\pi^0_{\alpha|\phi} \in \Pi^\infty\left(\pi^*_{\alpha|\phi}\right)$ maximizing $\left\{\int_{IS_\alpha(\phi)} h(\delta(x), \alpha) d\pi_{\alpha|\phi} - \kappa^* \mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi})\right\}$ satisfies

$$d\pi^0_{\alpha|\phi} = \frac{\exp(h(\delta, \alpha)/\kappa^*)}{\int_{IS_\alpha(\phi)} \exp(h(\delta, \alpha)/\kappa^*) d\pi^*_{\alpha|\phi}} \cdot d\pi^*_{\alpha|\phi}, \qquad \alpha\text{ -a.e.} \tag{34}$$

with $\kappa^* > 0$. Since $\exp(h(\delta,\alpha)/\kappa^*) \in (0,\infty)$ for all $\alpha \in IS_\alpha(\phi)$ and $\delta \in \mathcal{D}$ by the assumption, (34) implies that $\pi^*_{\alpha|\phi}$ is absolutely continuous with respect to $\pi^0_{\alpha|\phi}$, and hence, any $\pi_{\alpha|\phi}$ with $\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) < \infty$ is absolutely continuous with respect to $\pi^0_{\alpha|\phi}$. Therefore, the objective function can be rewritten as

$$\begin{aligned} & \int_{IS_\alpha(\phi)} h(\delta,\alpha)d\pi_{\alpha|\phi} - \kappa^*\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) \\ = \; & \int_{IS_\alpha(\phi)} h(\delta,\alpha)d\pi_{\alpha|\phi} - \kappa^*\mathcal{R}(\pi_{\alpha|\phi}\|\pi^0_{\alpha|\phi}) - \kappa^* \int_{IS_\alpha(\phi)} \log\left(\frac{d\pi^0_{\alpha|\phi}}{d\pi^*_{\alpha|\phi}}\right) d\pi_{\alpha|\phi}. \end{aligned}$$

Plugging in (34) leads to

$$-\kappa^*\mathcal{R}(\pi_{\alpha|\phi}\|\pi^0_{\alpha|\phi}) + \int_{IS_\alpha(\phi)} \exp(h(\delta,\alpha)/\kappa^*)d\pi^*_{\alpha|\phi}.$$

Since $\mathcal{R}(\pi_{\alpha|\phi}\|\pi^0_{\alpha|\phi}) \geq 0$ for any $\pi_{\alpha|\phi} \in \Pi^\infty\left(\pi^*_{\alpha|\phi}\right)$ and equal to zero if and only if $\pi_{\alpha|\phi} = \pi^0_{\alpha|\phi}$ holds for almost every α, $\pi^0_{\alpha|\phi}$ defined in (34) solves uniquely (up to α-a.e.) the inner maximization problem. Hence, it holds

$$\max_{\pi_{\alpha|\phi}\in\Pi^\infty\left(\pi^*_{\alpha|\phi}\right)} \left\{\int_{IS_\alpha(\phi)} h(\delta,\alpha)d\pi_{\alpha|\phi} - \kappa^*\mathcal{R}(\pi_{\alpha|\phi}\|\pi^0_{\alpha|\phi})\right\} = \kappa^* \ln\left(\int_{IS_\alpha(\phi)} \exp\left(\frac{h(\delta,\alpha)^*}{\kappa}\right) d\pi^*_{\alpha|\phi}\right). \tag{35}$$

By Lemma 2.2, $\pi^0_{\alpha|\phi}(\alpha)$ derived in (34) solves the inner maximization problem of (13). Hence, the value function is given by

$$\begin{aligned} \max_{\pi_{\alpha|\phi}\in\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left\{\int_{IS_\alpha(\phi)} h(\delta,\alpha)d\pi_{\alpha|\phi}\right\} &= \int_{IS_\alpha(\phi)} h(\delta,\alpha)d\pi^0_{\alpha|\phi} \\ &= \int_{IS_\alpha(\phi)} \frac{h(\delta,\alpha)\exp(h(\delta,\alpha)/\kappa^*)}{\int_{IS_\alpha(\phi)}\exp(h(\delta,\alpha)/\kappa^*)d\pi^*_{\alpha|\phi}} d\pi^*_{\alpha|\phi}. \end{aligned}$$

Also, by the Kuhn-Tucker condition stated in Lemma 2.2, $\lambda = \mathcal{R}(\pi^0_{\alpha|\phi}\|\pi^*_{\alpha|\phi})$ holds, which leads to the following condition for κ^*:

$$f'_\lambda(\kappa) \equiv \lambda + \ln\left(\int_{IS_\alpha(\phi)} \exp\left(h(\delta,\alpha)/\kappa^*\right) d\pi^*_{\alpha|\phi}\right) - \frac{\int_{IS_\alpha(\phi)} h(\delta,\alpha)\exp(h(\delta,\alpha)/\kappa^*)d\pi^*_{\alpha|\phi}}{\kappa^*\int_{IS_\alpha(\phi)}\exp(h(\delta,\alpha)/\kappa^*)d\pi^*_{\alpha|\phi}} = 0. \tag{36}$$

Note that this condition is obtained as the first-order condition of

$$f_\lambda(\kappa) \equiv \kappa\ln\left(\int_{IS_\alpha(\phi)} \exp\left(h(\delta,\alpha)/\kappa\right) d\pi^*_{\alpha|\phi}\right) + \kappa\lambda$$

with respect to κ. Note that $\lim_{\kappa\to 0} f'_\lambda(\kappa) = -\infty$ and $\lim_{\kappa\to\infty} f'_\lambda(\kappa) = \lambda > 0$. Furthermore, it can be shown that its second derivative of $f_\lambda(\kappa)$ in κ equals to the variance of $h(\delta,\alpha)$ with $\alpha \sim \pi^0_{\alpha|\phi}$, which is strictly positive by the imposed nondegeneracy assumption of $h(\delta,\alpha)$. Hence, $f_\lambda(\kappa)$ is strictly convex. Therefore, κ^* solving the first-order condition is unique and strictly positive.

The conclusion follows by integrating this value function with respect to $\pi_{\phi|X}$. ■

A.3 Proof of Theorem 3.3

The following lemmas A.4 - A.7 are used to prove Theorem 3.3.

Lemma A.4 *Under Assumption 3.2 (iv), we have*

$$(i) \quad \inf_{\delta\in\mathcal{D},\phi\in G_0} Var^*_{\alpha|\phi}(h(\delta,\alpha)) > 0$$

$$(ii) \quad \inf_{\delta\in\mathcal{D},\phi\in G_0} E^*_{\alpha|\phi}\left[\{h(\delta,\alpha) - E^*_{\alpha|\phi}(h(\delta,\alpha))\}^2 \cdot 1\{h(\delta,\alpha) - E^*_{\alpha|\phi}(h(\delta,\alpha)) \geq 0\}\right] > 0,$$

where $E^*_{\alpha|\phi}(\cdot)$ and $Var^*_{\alpha|\phi}(\cdot)$ are the mean and variance with respect to the benchmark conditional prior $\pi^*_{\alpha|\phi}$

Proof of Lemma A.4. Let $h = h(\delta,\alpha)$. By Markov's inequality and Assumption 3.2 (iv),

$$Var^*_{\alpha|\phi}(h) \geq c\pi^*_{\alpha|\phi}\left(\left\{(h - E^*_{\alpha|\phi}(h))^2 \geq c\right\}\right) = c\epsilon > 0.$$

This proves the first inequality.

To show the second inequality, suppose it is false. Then, there exists a sequence, (δ^ν,ϕ^ν), $\nu = 1,2,\ldots$, such that

$$\lim_{\nu\to\infty} E^*_{\alpha|\phi^\nu}\left[\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha))\}^2 \cdot 1\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha)) \geq 0\}\right] = 0.$$

By Markov's inequality, this means for any $a > 0$,

$$\lim_{\nu\to\infty} \pi^*_{\alpha|\phi^\nu}\left(\left\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha)) \geq a\right\}\right) = 0. \tag{37}$$

Assumption 3.2 (iv) then requires that

$$\lim_{\nu\to\infty} \pi^*_{\alpha|\phi^\nu}\left(\left\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha)) \leq -c\right\}\right) \geq \epsilon. \tag{38}$$

Equations (37) and (38) contradict $E^*_{\alpha|\phi^\nu}\left[h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha))\right] = 0$ for any ν, since if (37) and (38) were true,

$$
\begin{aligned}
& E^*_{\alpha|\phi^\nu}\left[h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha))\right] \\
\leq & \int_0^\infty \pi^*_{\alpha|\phi^\nu}\left(\left\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha)) \geq a\right\}\right) da - c\pi^*_{\alpha|\phi^\nu}\left(\left\{h(\delta^\nu,\alpha) - E^*_{\alpha|\phi^\nu}(h(\delta^\nu,\alpha)) \leq -c\right\}\right) \\
\leq & -c\epsilon/2 < 0 \qquad (39)
\end{aligned}
$$

would hold for all large ν. ■

Lemma A.5 *Suppose Assumption 3.2 (ii) and (iv) hold, and let $\lambda > 0$ be given. Let $\kappa_\lambda(\delta,\alpha)$ be the Lagrange multiplier defined in Lemma 2.2. We have*

$$\kappa_\lambda(\delta,\phi) \leq \frac{M}{\lambda},$$

for all $\delta \in \mathcal{D}$ and $\phi \in \Phi$, and

$$0 < C_1(\lambda) \leq \kappa_\lambda(\delta,\phi)$$

for all $\delta \in \mathcal{D}$ and $\phi \in G_0$, where $C_1(\lambda)$ is a positive constant that depends on λ but does not depend on δ and ϕ.

Proof of Lemma A.5. We first show the upper bound. Recall $\kappa_\lambda(\delta,\phi)$ solves (see equation (36))

$$\lambda = -\ln\left(\int \exp\left(\frac{h(\delta,\alpha)}{\kappa}\right) d\pi^*_{\alpha|\phi}\right) + E^o_{\alpha|\phi}\left[\frac{h(\delta,\alpha)}{\kappa}\right], \qquad (40)$$

where $E^o_{\alpha|\phi}(\cdot)$ is the expectation with respect to the exponentially tilted conditional prior $\pi^o_{\alpha|\phi}$. By noting that the nonnegativity of $h(\delta,\alpha)$ implies first term in equation (40) is negative and the boundedness $h(\delta,\alpha) \leq M$ implies the second term can be bounded by M/κ, we have

$$\lambda \leq \frac{M}{\kappa_\lambda(\delta,\phi)}.$$

This leads to the upper bound.

To show the lower bound, let $\kappa^* = \kappa_\lambda(\delta,\phi)$ be a short-hand notation for the solution of (40). Define

$$W \equiv \frac{h(\delta,\alpha)}{\kappa^*} - \ln\left(\int \exp\left(\frac{h(\delta,\alpha)}{\kappa^*}\right) d\pi^*_{\alpha|\phi}\right). \qquad (41)$$

By rewriting equation (40), we obtain the following inequality:

$$
\begin{aligned}
\lambda &= E^o_{\alpha|\phi}(W) = E^*_{\alpha|\phi}(W\exp(W)) \\
&\geq E^*_{\alpha|\phi}[W(1+W)\cdot 1\{W \geq 0\}] + E^*_{\alpha|\phi}[W \cdot 1\{W < 0\}] \\
&= E^*_{\alpha|\phi}(W) + E^*_{\alpha|\phi}(W^2 \cdot 1\{W \geq 0\}), \qquad (42)
\end{aligned}
$$

where the inequality holds by $e^x \geq 1+x$ for $x \geq 0$ and $e^x \leq 1$ for $x < 0$. By Jensen's inequality applied to $\ln\left(\int \exp\left(\frac{h(\delta,\alpha)}{\kappa}\right) d\pi^*_{\alpha|\phi}\right)$ and the nondgeneracy assumption of Assumption 3.2 (ii), we have

$$0 > E^*_{\alpha|\phi}(W) \geq -\frac{1}{\kappa^*}c_1, \tag{43}$$

$$c_1 \equiv M - \inf_{\delta\in\mathcal{D},\phi\in G_0} E^*_{\alpha|\phi}(h(\delta,\alpha)) > 0.$$

For the second term in (42),

$$E^*_{\alpha|\phi}\left[W^2 \cdot 1\{W \geq 0\}\right] \geq \frac{1}{(\kappa^*)^2} E^*_{\alpha|\phi}\left[\tilde{W}^2 \cdot 1\{\tilde{W} \geq 0\}\right] \geq \frac{1}{(\kappa^*)^2}c_2 > 0, \tag{44}$$

$$\tilde{W} \equiv h(\delta,\alpha) - E^*_{\alpha|\phi}(h(\delta,\alpha)),$$

$$c_2 \equiv \inf_{\delta\in\mathcal{D},\phi\in G_0} E^*_{\alpha|\phi}\left[\{h(\delta,\alpha) - E^*_{\alpha|\phi}(h(\delta,\alpha))\}^2 \cdot 1\{h(\delta,\alpha) - E^*_{\alpha|\phi}(h(\delta,\alpha)) \geq 0\}\right] > 0,$$

where the first inequality follows since $W \geq \tilde{W}/\kappa^*$ holds for any α, and the positivity of c_2 follows by Lemma A.4 (ii).

Combining (42), (43), and (44), we obtain

$$\lambda \geq -\frac{1}{\kappa^*}c_1 + \frac{1}{(\kappa^*)^2}c_2.$$

Solving this inequality for κ^* leads to

$$\kappa^* \geq \frac{-c_1 + \sqrt{c_1^2 + 4\lambda c_2}}{2\lambda} \equiv C_1(\lambda) > 0.$$

■

Lemma A.6 *Under Assumption 3.2 (ii) and (iv), we have*

$$\inf_{\delta\in\mathcal{D},\phi\in G_0} Var^o_{\alpha|\phi}(h(\delta,\alpha)) \geq c\epsilon \cdot \exp\left(-\frac{M}{C_1(\lambda)}\right) > 0,$$

where $Var^o_{\alpha|\phi}(\cdot)$ is the variance with respect to the worst-case conditional prior $\pi^o_{\alpha|\phi}$ shown in Theorem 3.1.

Proof of Lemma A.6. Let $c > 0$ be the constant defined in Assumption 3.2 (iv), and $h = h(\delta,\alpha)$ By Markov's inequality,

$$\begin{aligned} Var^o_{\alpha|\phi}(h(\delta,\alpha)) &\geq cE^o_{\alpha|\phi}\left(1\left\{(h - E^o_{\alpha|\phi}(h))^2 \geq c\right\}\right) \\ &\geq c \cdot \exp\left(-\frac{M}{C_1(\lambda)}\right) E^*_{\alpha|\phi}\left(1\left\{(h - E^o_{\alpha|\phi}(h))^2 \geq c\right\}\right) \\ &\geq c\epsilon \cdot \exp\left(-\frac{M}{C_1(\lambda)}\right), \end{aligned}$$

where the second inequality follows by the lower bound of $\kappa_\lambda(\delta,\phi)$ shown in Lemma A.5 and $E^o_{\alpha|\phi}(f(\alpha)) \geq \exp\left(-\frac{M}{C_1(\lambda)}\right) E^*_{\alpha|\phi}(f(\alpha))$ for any nonnegative random variables $f(\alpha)$. ■

Lemma A.7 *Suppose Assumption 3.2 (ii), (iv), and (vi) hold. Then,*

$$|\kappa_\lambda(\delta,\phi) - \kappa_\lambda(\delta,\phi_0)| \leq C_2(\lambda)\|\phi - \phi_0\| \tag{45}$$

holds for all $\delta \in \mathcal{D}$ and $\phi \in G_0$, where $0 \leq C_2(\lambda) < \infty$ is a constant that depends on $\lambda > 0$ but does not depend on δ and ϕ.

Proof of Lemma A.7. By the mean value expansion of $\kappa_\lambda(\delta,\phi)$, we have for $\phi \in G_0$,

$$|\kappa_\lambda(\delta,\phi) - \kappa_\lambda(\delta,\phi_0)| \leq \sup_{\delta\in\mathcal{D},\phi\in G_0}\left\{\left\|\frac{\partial\kappa_\lambda(\delta,\phi)}{\partial\phi}\right\|\right\} \cdot \|\phi - \phi_0\|.$$

Hence, it suffices to find $C_2(\lambda)$ that satisfies $\sup_{\delta\in\mathcal{D},\phi\in G_0}\left\|\frac{\partial\kappa_\lambda(\delta,\phi)}{\partial\phi}\right\| \leq C_2(\lambda) < \infty$.

We apply the derivative formula of the implicit function to $\kappa_\lambda(\delta,\phi)$ defined as the solution to

$$g(\delta,\kappa,\phi) \equiv \lambda + \ln\left(\int \exp\left(\frac{h(\delta,\alpha)}{\kappa}\right) d\pi^*_{\alpha|\phi}\right) - E^o_{\alpha|\phi}\left[\frac{h(\delta,\alpha)}{\kappa}\right] = 0.$$

Since $|\partial g/\partial\kappa| = Var^o_{\alpha|\phi}(h(\delta,\alpha)/\kappa_\lambda(\delta,\phi))$, we obtain

$$\sup_{\delta\in\mathcal{D},\phi\in G_0}\left\|\frac{\partial\kappa_\lambda(\delta,\phi)}{\partial\phi}\right\| \leq \left(\frac{M}{\lambda}\right)^2 \frac{\sup_{\delta\in\mathcal{D},\phi\in G_0}\|\partial g/\partial\phi\|}{\inf_{\delta\in\mathcal{D},\phi\in G_0} Var^o_{\alpha|\phi}(h(\delta,\alpha))}, \tag{46}$$

where the differentiability of g with respect to ϕ requires Assumption 3.2 (vi). By Lemma A.6, the variance lower bound in the denominator is bounded away from zero. For the numerator, we have

$$\begin{aligned}
\left\|\frac{\partial g}{\partial\phi}\right\| &\leq \frac{\left\|\frac{\partial}{\partial\phi}E^*_{\alpha|\phi}(h/\kappa)\right\|}{E^*_{\alpha|\phi}(\exp(h/\kappa))} + \frac{\left\|\frac{\partial}{\partial\phi}E^*_{\alpha|\phi}[(h/\kappa)\cdot\exp(h/\kappa)]\right\|}{E^*_{\alpha|\phi}(\exp(h/\kappa))} + \frac{\left\|\frac{\partial}{\partial\phi}E^*_{\alpha|\phi}(h/\kappa)\right\| \cdot E^*_{\alpha|\phi}[h/\kappa\cdot\exp(h/\kappa)]}{[E^*_{\alpha|\phi}(\exp(h/\kappa))]^2} \\
&\leq \left(\frac{M}{C_1(\lambda)} + \frac{M^2}{C_1(\lambda)^2}\exp\left(\frac{M}{C_1(\lambda)}\right)\right) \cdot \sup_{\delta\in\mathcal{D},\phi\in G_0}\left\|\frac{\partial}{\partial\phi}E^*_{\alpha|\phi}(h)\right\| \\
&\quad + \sup_{\delta\in\mathcal{D},\phi\in G_0,\kappa\in[C_1(\lambda),M/\lambda]}\left\|\frac{\partial}{\partial\phi}E^*_{\alpha|\phi}\left[\left(\frac{h}{\kappa}\right)\cdot\exp\left(\frac{h}{\kappa}\right)\right]\right\| \\
&\equiv C_2(\lambda) < \infty,
\end{aligned}$$

where the second inequality follows by noting $E^*_{\alpha|\phi}(\exp(h/\kappa)) \geq 1$ and $E^*_{\alpha|\phi}[h/\kappa\cdot\exp(h/\kappa)] \leq (M/C_1(\lambda))\exp(M/C_1(\lambda))$, and the third inequality follows from Assumption 3.2 (vi). ■

Proof of Theorem 3.3. (i) By Assumption 3.2 (iii) and (vii) and the consistency theorem of the extremum estimator (Theorem 2.1 in Newey and McFadden (1994)), a minimizer of the the finite sample objective function $\int_{\Phi} r_{\lambda}(\delta, \phi)\, d\pi_{\phi|X}$ converges to $\delta_{\lambda}(\phi_0)$ almost surely (in probability) if $\int_{\Phi} r_{\lambda}(\cdot, \phi)\, d\pi_{\phi|X}$ converges to $r_{\lambda}(\cdot, \phi_0)$ uniformly almost surely (in probability).

To this goal, let

$$s_{\lambda}(\delta, \phi) \equiv \int \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\delta, \phi)}\right) d\pi^{*}_{\alpha|\phi} \in \left[1, \exp\left(\frac{M}{\kappa_{\lambda}(\delta, \phi)}\right)\right].$$

Since

$$\sup_{\delta \in \mathcal{D}} \left| \int_{\Phi} r_{\lambda}(\delta, \phi)\, d\pi_{\phi|X} - r_{\lambda}(\delta, \phi_0) \right| \leq \int_{\Phi} \sup_{\delta \in \mathcal{D}} |r_{\lambda}(\delta, \phi) - r_{\lambda}(\delta, \phi_0)|\, d\pi_{\phi|X},$$

we consider bounding $\sup_{\delta \in \mathcal{D}} |r_{\lambda}(\delta, \phi) - r_{\lambda}(\delta, \phi_0)|$ for $\phi \in G_0$. In what follows, we omit the arguments δ from r_{λ}, s_{λ}, and κ_{λ} unless confusion arises.

By Lemma 2.2 and equation (35) in the proof of Theorem 3.1, $r_{\lambda}(\phi)$ can be expressed as

$$r_{\lambda}(\phi) = \kappa_{\lambda}(\phi) \ln s_{\lambda}(\phi) + \kappa_{\lambda}(\phi)\lambda.$$

Hence, we have

$$\begin{aligned} |r_{\lambda}(\phi) - r_{\lambda}(\phi_0)| &= \kappa_{\lambda}(\phi) \ln s_{\lambda}(\phi) - \kappa_{\lambda}(\phi_0) \ln s_{\lambda}(\phi_0) + (\kappa_{\lambda}(\phi) - \kappa_{\lambda}(\phi_0))\lambda \\ &\leq \kappa_{\lambda}(\phi) |\ln s_{\lambda}(\phi) - \ln s_{\lambda}(\phi_0)| + |\kappa_{\lambda}(\phi) - \kappa_{\lambda}(\phi_0)| \ln s_{\lambda}(\phi) \\ &\quad + |\kappa_{\lambda}(\phi) - \kappa_{\lambda}(\phi_0)|\, \lambda. \end{aligned} \tag{47}$$

By noting $\ln(x) \leq x - 1$, Lemma A.5, and $s_{\lambda}(\phi) \geq 1$, we have

$$\begin{aligned} &\kappa_{\lambda}(\phi) | \ln s_{\lambda}(\phi) - \ln s_{\lambda}(\phi_0)| \\ \leq &\frac{M}{\lambda} \cdot \frac{|s_{\lambda}(\phi) - s_{\lambda}(\phi_0)|}{s_{\lambda}(\phi) \wedge s_{\lambda}(\phi_0)} \\ = &\frac{M}{\lambda} \left| \int \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi)}\right) d\pi^{*}_{\alpha|\phi} - \int \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi_0)}\right) d\pi^{*}_{\alpha|\phi_0} \right| \\ \leq &\frac{M}{\lambda} \int \left| \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi)}\right) - \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi)}\right) \right| d\pi^{*}_{\alpha|\phi} + \int \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi_0)}\right) |d\pi^{*}_{\alpha|\phi_0} - d\pi_{\alpha|\phi}| \\ \leq &\frac{M}{\lambda} \int \exp\left(\frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi)}\right) \left| \frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi)} - \frac{h(\delta, \alpha)}{\kappa_{\lambda}(\phi_0)} \right| d\pi^{*}_{\alpha|\phi} + \frac{M}{C_1(\lambda)} \|\pi^{*}_{\alpha|\phi_0} - \pi_{\alpha|\phi}\|_{TV} \\ \leq &\frac{M^2}{\lambda C_1(\lambda)} \exp\left(\frac{M}{C_1(\lambda)}\right) |\kappa_{\lambda}(\phi) - \kappa_{\lambda}(\phi_0)| + \frac{M}{C_1(\lambda)} \|\pi^{*}_{\alpha|\phi} - \pi_{\alpha|\phi_0}\|_{TV} \end{aligned} \tag{48}$$

Combining equations (47) and (48), and applying Lemma A.7, we obtain for $\phi \in G_0$,

$$\sup_{\delta \in \mathcal{D}} |r_{\kappa}(\delta, \phi) - r_{\kappa}(\delta, \phi_0)| \leq \frac{M}{C_1(\lambda)} \left\| \pi^{*}_{\alpha|\phi} - \pi^{*}_{\alpha|\phi_0} \right\|_{TV} + C_3(\lambda) \|\phi - \phi_0\|, \tag{49}$$

where $C_3(\lambda) = \lambda + \frac{M}{C_1(\lambda)} + \frac{M^2}{\lambda C_1(\lambda)} \exp\left(\frac{M}{C_1(\lambda)}\right)$. Thus,

$$
\begin{aligned}
\int_{\Phi} \sup_{\delta \in \mathcal{D}} \left| r_{\lambda}(\delta, \phi) - r_{\lambda}(\delta, \phi_0) \right| d\pi_{\phi|X} &\leq \int_{G_0} \sup_{\delta \in \mathcal{D}} \left| r_{\lambda}(\delta, \phi) - r_{\lambda}(\delta, \phi_0) \right| d\pi_{\phi|X} + 2M\pi_{\phi|X}(G_0^c) \\
&\leq \frac{M}{C_1(\lambda)} \int_{G_0} \| \pi^*_{\alpha|\phi} - \pi^*_{\alpha|\phi_0} \|_{TV} d\pi_{\phi|X} \\
&\quad + C_3(\lambda) \int_{G_0} \| \phi - \phi_0 \| d\pi_{\phi|X} + 2M\pi_{\phi|X}(G_0^c). \qquad (50)
\end{aligned}
$$

The almost sure posterior consistency of $\pi_{\phi|X}$ in Assumption 3.2 (i) implies $\pi_{\phi|X}(G_0^c) \to 0$ as $n \to \infty$. Also, viewing $\|\pi^*_{\alpha|\phi} - \pi_{\alpha|\phi_0}\|_{TV}$ and $\|\phi - \phi_0\|$ as continuous functions of ϕ (Assumption 3.2 (v)), the continuous mapping theorem implies the other two terms in the right-hand side of (50) converge to zero as $n \to \infty$ almost surely. This completes the proof of claim (i).

(ii) When $\hat{\phi} \to_p \phi_0$, the continuous mapping theorem and (49) imply that $\left| r_{\kappa}\left(\delta, \hat{\phi}\right) - r_{\kappa}(\delta, \phi_0) \right| \to_p 0$ as $n \to \infty$ uniformly over δ. By the consistency theorem of the extremum estimator (Theorem 2.1 in Newey and McFadden (1994)), the claim follows. ■

Proof of Theorem 5.2. Fixing $\delta \in \mathcal{D}$, let us partition the reduced-form parameter space Φ by

$$
\begin{aligned}
\Phi_{\delta}^{+} &= \left\{ \phi \in \Phi : \frac{\alpha_*(\phi) + \alpha^*(\phi)}{2} \geq \delta \right\}, \\
\Phi_{\delta}^{-} &= \left\{ \phi \in \Phi : \frac{\alpha_*(\phi) + \alpha^*(\phi)}{2} < \delta \right\}.
\end{aligned}
$$

We write the objective function of Theorem 3.1 as

$$
\int_{\Phi_{\delta}^{-}} r_{\lambda}(\delta, \phi) \, d\pi_{\phi|X} + \int_{\Phi_{\delta}^{+}} r_{\lambda}(\delta, \phi) \, d\pi_{\phi|X},
$$

and aim to derive the limits of each of the two terms.

By Lemma A.5, as $\lambda \to \infty$, we have $\kappa_{\lambda}(\delta, \phi) \to 0$ at every (δ, ϕ). Hence, to assess the point-wise convergence behavior of $r_{\lambda}(\delta, \phi)$ as $\lambda \to \infty$ at each (δ, ϕ), it suffices to analyzing the limit behavior with respect to $\kappa \to 0$ of

$$
r_{\kappa}(\delta, \phi) \equiv \frac{\int (\delta - \alpha)^2 \exp\left\{ \frac{(\delta - \alpha)^2}{\kappa} \right\} d\pi^*_{\alpha|\phi}}{\int \exp\left\{ \frac{(\delta - \alpha)^2}{\kappa} \right\} d\pi^*_{\alpha|\phi}}.
$$

For $\phi \in \Phi_{\delta}^{-}$, we rewrite $r_{\kappa}(\delta, \phi)$ as

$$
r_{\kappa}(\delta, \phi) = (\delta - \alpha_*(\phi))^2 + \frac{\int [(\delta - \alpha)^2 - (\delta - \alpha_*(\phi))^2] \exp\left\{ -\frac{(\delta - \alpha_*(\phi))^2 - (\delta - \alpha)^2}{\kappa} \right\} d\pi^*_{\alpha|\phi}}{\int \exp\left\{ -\frac{(\delta - \alpha_*(\phi))^2 - (\delta - \alpha)^2}{\kappa} \right\} d\pi^*_{\alpha|\phi}}, \qquad (51)
$$

and shows that the second term in the right-hand side converges to zero.

For the denominator, let $c(\phi) = 2(\delta - \alpha_*(\phi)) > 0$ and note

$$
\begin{aligned}
&\int \exp\left\{-\frac{(\delta-\alpha_*(\phi))^2 - (\delta-\alpha)^2}{\kappa}\right\} d\pi^*_{\alpha|\phi} \\
=&\int \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{\alpha|\phi} \\
=&\int_{\alpha_*(\phi)}^{\alpha_*(\phi)+\eta} \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{\alpha|\phi} \\
&+\int_{\alpha_*(\phi)+\eta}^{\alpha^*(\phi)} \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{\alpha|\phi} \\
=&\int_0^{\eta} \left(\sum_{k=1}^{\infty} a_k z^k\right) \exp\left\{-\frac{c(\phi)z - z^2}{\kappa}\right\} dz \\
&+\int_{\alpha_*(\phi)+\eta}^{\alpha^*(\phi)} \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{\alpha|\phi}, \qquad (52)
\end{aligned}
$$

where the third equality uses Assumption 5.1 (iii). The integrand of the second term in (52) converges exponentially fast to zero as $\kappa \to 0$ at every $\alpha \in [\alpha_*(\phi) + \eta, \alpha^*(\phi)]$. Hence, by the dominated convergence theorem, the second term in (52) converges exponentially fast to zero as $\kappa \to 0$. We apply the general Laplace approximation (see, e.g., Theorem 1 in Chapter 2 of Wong (1989)) to the first term in (52). Let $k^* \geq 0$ be the least nonnegative integer k such that $a_k \neq 0$. Then, the leading term in the Laplace approximation is given by

$$
\int_0^{\eta} \left(\sum_{k=0}^{\infty} a_k z^k\right) \exp\left\{-\frac{c(\phi)z - z^2}{\kappa}\right\} dz = \Gamma(k^*+1)\left(\frac{a_{k^*}}{c(\phi)^{k^*+1}}\right)\kappa^{k^*+1} + o(\kappa^{k^*+1}).
$$

As for the numerator of the second term in the right-hand side of (51),

$$
\begin{aligned}
&\int [(\delta-\alpha)^2 - (\delta-\alpha_*(\phi))^2] \exp\left\{-\frac{(\delta-\alpha_*(\phi))^2 - (\delta-\alpha)^2}{\kappa}\right\} d\pi^*_{\alpha|\phi} \\
=&\int [-c(\phi)(\alpha-\alpha_*(\phi)) + (\alpha-\alpha_*(\phi))^2] \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{*|\phi} \\
=&\int_{\alpha_*(\phi)}^{\alpha_*(\phi)+\eta} [-c(\phi)(\alpha-\alpha_*(\phi)) + (\alpha-\alpha_*(\phi))^2] \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{*|\phi} \\
&+\int_{\alpha_*(\phi)+\eta}^{\alpha^*(\phi)} [-c(\phi)(\alpha-\alpha_*(\phi)) + (\alpha-\alpha_*(\phi))^2] \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{*|\phi} \\
=&\int_0^{\eta} \left(\sum_{k=1}^{\infty} \tilde{a}_k z^k\right) \exp\left\{-\frac{c(\phi)z - z^2}{\kappa}\right\} dz \\
&+\int_{\alpha_*(\phi)+\eta}^{\alpha^*(\phi)} [-c(\phi)(\alpha-\alpha_*(\phi)) + (\alpha-\alpha_*(\phi))^2] \exp\left\{-\frac{c(\phi)(\alpha-\alpha_*(\phi)) - (\alpha-\alpha_*(\phi))^2}{\kappa}\right\} d\pi^*_{\alpha|\phi}
\end{aligned}
$$

where $\sum_{k=1}^{\infty} \tilde{a}_k z^k = (-c(\phi)z + z^2) \left(\sum_{k=0}^{\infty} a_k z^k\right)$. Similarly to the previous argument, the second term in the right-hand converges to zero exponentially fast as $\kappa \to 0$ by the dominated convergence theorem. Regarding the first-term, the Laplace approximation yields

$$\int_0^{\eta} \left(\sum_{k=1}^{\infty} \tilde{a}_k z^k\right) \exp\left\{-\frac{c(\phi)z - z^2}{\kappa}\right\} dz = \Gamma(k^* + 2) \left(-\frac{a_{k^*}}{c(\phi)^{k^*+1}}\right) \kappa^{k^*+2} + o(\kappa^{k^*+2}).$$

Combining the arguments, the second term in the right-hand side of (51) is $O(\kappa)$. Hence,

$$\lim_{\kappa \to 0} r_\kappa (\delta, \phi) = (\delta - \alpha_* (\phi))^2 .$$

for $\phi \in \Phi_\delta^-$ pointwise.

The limit for $r_\kappa (\delta, \phi)$ on $\phi \in \Phi_\delta^+$ can be obtained similarly, $\lim_{\kappa \to 0} r_\kappa (\delta, \phi) = (\delta - \alpha^* (\phi))^2$, and we omit the detailed proof for brevity.

Since $r_\kappa (\delta, \phi)$ has an integrable envelope (e.g., $(\delta - \alpha_*(\phi))^2$ on $\phi \in \Phi_\delta^-$ and $(\delta - \alpha^*(\phi))^2$ on $\phi \in \Phi_\delta^+$), the dominated convergence theorem leads to

$$\begin{aligned} \lim_{\kappa \to 0} \int_\Phi r_\kappa (\delta, \phi) \, d\pi_{\phi|X} &= \int_{\Phi_\delta^-} \lim_{\kappa \to 0} r_\kappa (\delta, \phi) \, d\pi_{\phi|X} + \int_{\Phi_\delta^+} \lim_{\kappa \to 0} r_\kappa (\delta, \phi) \, d\pi_{\phi|X} \\ &= \int_{\Phi_\delta^-} (\delta - \alpha_* (\phi))^2 \, d\pi_{\phi|X} + \int_{\Phi_\delta^+} (\delta - \alpha^* (\phi))^2 \, d\pi_{\phi|X} \\ &= \int_\Phi \left((\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2\right) d\pi_{\phi|X}, \end{aligned}$$

where the last line follows by noting that $(\delta - \alpha_* (\phi))^2 \geq (\delta - \alpha^* (\phi))^2$ holds for $\phi \in \Phi_\delta^-$ and the reverse inequality holds for $\phi \in \Phi_\delta^+$.

(ii) Fix δ and set $h(\delta, \alpha) = \rho_\tau (\alpha - \delta)$. Partition the parameter space Φ by

$$\begin{aligned} \Phi_\delta^+ &= \{\phi \in \Phi : (1 - \tau)\alpha_*(\phi) + \tau\alpha^*(\phi) \geq \delta\}, \\ \Phi_\delta^- &= \{\phi \in \Phi : (1 - \tau)\alpha_*(\phi) + \tau\alpha^*(\phi) < \delta\}, \end{aligned}$$

and write $\int_\Phi r_\kappa (\delta, \phi) \, d\pi_{\phi|X}$ as

$$\int_{\Phi_\delta^-} r_\kappa (\delta, \phi) \, d\pi_{\phi|X} + \int_{\Phi_\delta^+} r_\kappa (\delta, \phi) \, d\pi_{\phi|X}.$$

We then repeat the proof techniques used in part (i). We omit the details for brevity. ■

Proof of Theorem 5.3. (i) Let $r_\kappa(\delta, \phi)$ as defined in the proof of Theorem 5.2. Since $\lambda \to \infty$ asymtptotics implies $\kappa \to 0$ asymptotics, we consider working with $R_n (\delta) \equiv \lim_{\kappa \to 0} \int_\Phi r_\kappa (\delta, \phi) \, d\pi_{\phi|X}$, which is equal to $R_n(\delta) = \int_\Phi r_0(\delta,\phi)\, d\pi_{\phi|X}$ where $r_0(\delta,\phi) = (\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2$. Since the parameter space for α and the domain of δ are compact, $r_0(\delta,\phi)$ is a bounded function in ϕ. In addition, $\alpha_*(\phi)$ and $\alpha^*(\phi)$ are assumed to be continuous at $\phi = \phi_0$, so $r_0(\delta,\phi)$ is continuous at $\phi = \phi_0$. Hence, the weak convergence of $\pi_{\phi|X}$ to the point mass measure implies the convergence in mean

$$\begin{aligned} R_n(\delta) \to R_\infty(\delta) &\equiv \lim_{n\to\infty} \int_\Phi \left[(\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2 \right] d\pi_{\phi|X} \\ &= (\delta - \alpha_*(\phi_0))^2 \vee (\delta - \alpha^*(\phi_0))^2 \end{aligned} \tag{53}$$

pointwise in δ for almost every sampling sequence. Note that $R_\infty(\delta)$ is minimized uniquely at $\delta = \frac{1}{2}(\alpha_*(\phi_0) + \alpha^*(\phi_0))$. Hence, by an analogy to the argument of the convergence of extremum-estimators (see, e.g., Newey and McFadden (1994)), the conclusion follows if the convergence of $R_n(\delta)$ to $R_\infty(\delta)$ is uniform in δ. To show this is the case, define $I(\phi) \equiv [\alpha_*(\phi), \alpha^*(\phi)]$ and note that $(\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2$ can be interpreted as the squared Hausdorff metric $[d_H(\delta, I(\phi))]^2$ between point $\{\delta\}$ and interval $I(\phi)$. Then

$$\begin{aligned} |R_n(\delta) - R_\infty(\delta)| &= \left| \int_\Phi \left([d_H(\delta, I(\phi))]^2 - [d_H(\delta, I(\phi_0))]^2 \right) d\pi_{\phi|X} \right| \\ &\leq 2diam(\alpha) \int_\Phi |d_H(\delta, I(\phi)) - d_H(\delta, I(\phi_0))|\, d\pi_{\phi|X} \\ &\leq 2diam(\alpha) \int_\Phi d_H(I(\phi), I(\phi_0))\, d\pi_{\phi|X}, \end{aligned} \tag{54}$$

where $diam(\alpha) < \infty$ is the diameter of the parameter space of α and the third line follows by the triangular inequality of a metric, $|d_H(\delta, I(\phi)) - d_H(\delta, I(\phi_0))| \leq d_H(I(\phi), I(\phi_0))$. Since $d_H(I(\phi), I(\phi_0))$ is bounded by the compactness assumption of the α space and is continuous at $\phi = \phi_0$ by Assumption 5.1 (iv), $\int_\Phi d_H(I(\phi), I(\phi_0))\, d\pi_{\phi|X} \to 0$ as $\pi_{\phi|X}$ converges weakly to the point mass measure at $\phi = \phi_0$. This implies the uniform convergence of $R_n(\delta)$, $\sup_\delta |R_n(\delta) - R_\infty(\delta)| \to 0$ as $n \to \infty$.

We now prove (ii). Let $l(\delta,\phi) \equiv (1-\tau)(\delta - \alpha_*(\phi)) \vee \tau(\alpha^*(\phi) - \delta)$. Similarly to the quadratic loss case shown above, we have

$$R_n(\delta) \to R_\infty(\delta) \equiv (1-\tau)(\delta - \alpha_*(\phi_0)) \vee \tau(\alpha^*(\phi_0) - \delta) = l(\delta, \phi_0), \tag{55}$$

which is minimized uniquely at $\delta = (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0)$. Hence, the conclusion follows if $\sup_\delta |R_n(\delta) - R_\infty(\delta)| \to 0$ is proven. To show this uniform convergence, define

$$\begin{aligned} \Phi_0^- &\equiv \{\phi \in \Phi : (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi) \leq (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0)\}, \\ \Phi_0^+ &\equiv \{\phi \in \Phi : (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi) > (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0)\}. \end{aligned} \tag{56}$$

On $\phi \in \Phi_0^-$, $l(\delta, \phi) - l(\delta, \phi_0)$ can be expressed as

$$l(\delta, \phi) - l(\delta, \phi_0) \tag{57}$$

$$= \begin{cases} (1-\tau)\left[\alpha_*(\phi_0) - \alpha_*(\phi)\right], & \text{if } \delta \leq (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi), \\ \tau\left[\alpha^*(\phi) - \alpha_*(\phi_0)\right] - \left[\delta - \alpha_*(\phi_0)\right], & \text{if } (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi) < \delta \leq (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0), \\ \tau\left[\alpha^*(\phi) - \alpha^*(\phi_0)\right] & \text{if } \delta > (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0). \end{cases} \tag{58}$$

By noting that in the second case in (58), the absolute value of $l(\delta, \phi) - l(\delta, \phi_0)$ is maximized at either of the boundary values of δ, it can be shown that $|l(\delta, \phi) - l(\delta, \phi_0)|$ can be bounded from above by $|\alpha_*(\phi) - \alpha_*(\phi_0)| + |\alpha^*(\phi) - \alpha^*(\phi_0)|$. Symmetrically, on $\phi \in \Phi_0^+$, $|l(\delta, \phi) - l(\delta, \phi_0)|$ can be bounded from above by the same upper bound. Hence, $\sup_\delta |R_n(\delta) - R_\infty(\delta)|$ can be bounded by

$$\begin{aligned} \sup_\delta |R_n(\delta) - R_\infty(\delta)| &\leq \sup_\delta \int_\Phi |l(\delta, \phi) - l(\delta, \phi_0)|\, d\pi_{\phi|X} \\ &\leq \int_\Phi |\alpha_*(\phi) - \alpha_*(\phi_0)|\, d\pi_{\phi|X} + \int_\Phi |\alpha^*(\phi) - \alpha^*(\phi_0)|\, d\pi_{\phi|X}, \end{aligned} \tag{59}$$

which converges to zero by the weak convergence of $\pi_{\phi|X}$, compactness of α space, and continuity of $\alpha_*(\phi)$ and $\alpha^*(\phi)$ at $\phi = \phi_0$. This completes the proof. ■

B Asymptotic Analysis with Discrete Benchmark Prior

If the loss function $h(\delta, \alpha)$ is differentiable with respect to δ at almost every α, the first order condition for the minimization problem (18) is obtained as

$$\int_\Phi \left[\int_{IS_\alpha(\phi)} \frac{\partial}{\partial \delta} h(\delta, \alpha) \left(\frac{\exp\{h(\delta, \alpha)/\kappa\}}{\int_{IS_\alpha(\phi)} \exp\{h(\delta, \alpha)/\kappa\}\, d\pi^*_{\alpha|\phi}}\right) d\pi^*_{\alpha|\phi}\right] d\pi_{\phi|X} = 0. \tag{60}$$

Suppose the benchmark conditional prior is a mixture of multiple probability masses (multiple point-identifying models). These point-identifying models are indexed by $m = 1, \ldots, M$, and they differ in the sense that each model selects a different point in the identified set. Denote the selection of α resulting from model m by $\alpha_m(\phi) \in IS_\phi(\alpha)$. A benchmark prior is given by a particular mixture of these point mass measures,

$$\pi^*_\phi(\alpha) = \sum_{m=1}^{M} w_m 1_{\alpha_m(\phi)}(\alpha), \;\; w_m > 0 \;\; \forall m, \;\; \sum_{m=1}^{M} w_m = 1,$$

where the weights $(w_1, \ldots, w_M)$ specify benchmark credibility over each point-identified model. The set of conditional priors concerned in (15) consists of any mixture of these point mass measures,

$$\Pi_{\phi}^{\infty}\left(\pi_{\phi}^{*}\right)=\left\{\sum_{m=1}^{M} w_m' 1_{\alpha_m(\phi)}(\alpha):\left(w_1',\ldots,w_M'\right)\in\Delta_M\right\},$$

where Δ_M is the probability simplex in $\mathcal{R}^M$.

Denote $(\alpha_1(\phi_0),\ldots,\alpha_M(\phi_0))$ by $(\alpha_1,\ldots,\alpha_M)$ for short, and label the models according to $\alpha_1\le\alpha_2\le\cdots\le\alpha_M$. With a fixed $\kappa>0$ and the degenerate posterior for ϕ, the first order condition (60) is simplified to

$$\frac{\sum_{m=1}^{M}(\delta-\alpha_m)w_m\exp\left(\frac{(\delta-\alpha_m)^2}{\kappa}\right)}{\sum_{m=1}^{M}w_m\exp\left(\frac{(\delta-\alpha_m)^2}{\kappa}\right)}=0, \tag{61}$$

where the denominator does not affect the solution. The next proposition shows that, as $\kappa\to0$, optimal δ solving this first order condition converges to the mid-point of the two extreme point-identified models, $(\alpha_1+\alpha_M)/2$.

Proposition B.1 *As $\kappa\to0$, optimal δ that solves (61) converges to* $(\alpha_M+\alpha_1)/2$.[4]

Proof. (sketch) Rewrite the first-order condition (61) as

$$\sum_{m=1}^{M}(\delta-\alpha_m)\exp\left(\frac{(\delta-\alpha_m)^2}{\kappa}+\ln w_m\right)=0. \tag{62}$$

As $\kappa\to0$, the exponential term will shoot out to positive infinity, so in order for the first order condition to be solved for some δ at small κ, it must be the case that one exponential term diverge to negative infinity and another exponential term diverges to positive infinity at the same rate as $\kappa\to0$. Along this reasoning, consider equalizing the exponential terms for the two extreme point-identified models, $m=1$ and $m=M$,

$$\exp\left(\frac{(\delta-\alpha_1)^2}{\kappa}+\ln w_1\right)=\exp\left(\frac{(\delta-\alpha_M)^2}{\kappa}+\ln w_M\right),$$

which gives

$$\delta^*=\frac{\alpha_1+\alpha_M}{2}-\frac{\kappa}{2(\alpha_M-\alpha_1)}\ln\left(\frac{w_1}{w_M}\right).$$

[4] In our Whiteboard 07, we obtain the result similar to this proposition for the case of two point-identified models ($M=2$) with benchmark weights $w_1^*\to\frac{1}{2}$ and $w_2^*\to\frac{1}{2}$. The current proposition extends it to the case with more than two models and allows for arbitrary benchmark weights.

Let $H = \frac{(\delta^* - \alpha_1)^2}{\kappa} + \ln w_1 = \frac{(\delta^* - \alpha_M)^2}{\kappa} + \ln w_M$. It can be shown that, for $m = 2, \ldots, (M-1)$,

$$\begin{aligned} & \frac{(\delta^* - \alpha_m)^2}{\kappa} + \ln w_m - H \\ = \; & -\frac{(\alpha_M - \alpha_m)(\alpha_m - \alpha_1)}{\kappa} - \left(\frac{\alpha_M - \alpha_m}{\alpha_M - \alpha_1} \right) \ln \left(\frac{w_1}{w_M} \right) + \ln \left(\frac{w_m}{w_M} \right), \end{aligned}$$

which diverges to negative infinity as $\kappa \to 0$.

We now show that δ^* constructed above satisfies the first order condition at the limit $\kappa \to 0$. Plug in δ^* into the left-hand side of (62) and divide it by $\exp(H)$,

$$(\delta^* - \alpha_1) + (\delta^* - \alpha_M) + \sum_{m=2}^{M-1} (\delta^* - \alpha_m) \exp\left(\frac{(\delta^* - \alpha_m)^2}{\kappa} + \ln w_m - H \right).$$

As $\kappa \to 0$, all the exponential terms in the summation converge to zero, and $(\delta^* - \alpha_1) + (\delta^* - \alpha_M) \to 0$ since $\delta^* \to \frac{\alpha_1 + \alpha_M}{2}$. That is, the optimal decision at the limit is given by the mid-point decision $\frac{\alpha_1 + \alpha_M}{2}$, and optimal δ that solves (61) should converge to $(\alpha_1 + \alpha_M)/2$. ■

For the check loss, we have the following results. Since $\rho_\tau(\alpha - \delta)$ is differentiable in δ except for the kink point, we can still rely on the first order condition (60), as far as the solution exists (if a solution does not exist, it would probably imply that an optimum is occurring at a nondifferentiable point). When $\widetilde{\pi}(\phi)$ is a probability mass, we can ignore the denominator term in (60), so that the first order condition is simplified to

$$\int_{IS_{\phi_0}(\alpha)} \left[(1-\tau) \exp\left\{ \frac{-(1-\tau)(\alpha - \delta)}{\kappa} \right\} 1\{\alpha < \delta\} - \tau \exp\left\{ \frac{\tau(\alpha - \delta)}{\kappa} \right\} 1\{\alpha > \delta\} \right] d\pi^*_{\phi_0}(\alpha) = 0. \tag{63}$$

By noting that the δ's appearing in the exponential terms can be factored out from the integral, we can rewrite this first-order condition as

$$\begin{aligned} \delta &= \kappa \ln \left(\frac{\tau \int_\delta^\infty \exp\left(\frac{\tau\alpha}{\kappa}\right) d\pi^*_{\phi_0}(\alpha)}{(1-\tau) \int_{-\infty}^\delta \exp\left(\frac{-(1-\tau)\alpha}{\kappa}\right) d\pi^*_{\phi_0}(\alpha)} \right) \\ &\equiv f_\tau(\delta). \end{aligned} \tag{64}$$

If $\pi^*_{\phi_0}(\alpha)$ does not involve any probability mass, then $f_\tau(\delta)$ is a continuous and weakly decreasing function in δ. Furthermore, $f_\tau(\delta)$ diverges to ∞ as δ approaches to the lower bound of $IS_\alpha(\phi_0)$ and diverges to $-\infty$ as δ approaches to the upper bound of $IS_{\phi_0}(\alpha)$. Therefore, the equation (64) has a unique solution for δ for every τ. We hereafter denote a unique root of $\delta = f_\tau(\delta)$ by $\delta^*(\tau)$ (if it exists).

The next sequence of propositions solve for $\delta^*(\tau)$ for various choices of benchmark prior $\pi^*_{\phi_0}(\alpha)$.

Proposition B.2 *Suppose* $IS_{\phi_0}(\alpha) = [\underline{y}, \alpha]$ *and* $\pi^*_{\phi_0}(\alpha)$ *is uniform on* $[\underline{y}, \alpha]$. *Then,* $\delta^*(\tau) = (1-\tau)\underline{y} + \tau\alpha$ *for all* $\tau \in (0,1)$. *Note that* κ *does not appear in* $\delta^*(\tau)$. *This result implies that* $\delta^*(\tau)$ *coincides with the* τ*-th quantile of* $\pi^*_{\phi_0}(\alpha)$.

Proof. Set $\pi^*_{\phi_0}(\alpha) = (\underline{y} - \alpha)^{-1} 1_{[\underline{y},\alpha]}(\alpha)$ in (64), and solve for δ yields the result. ■

Proposition B.3 *Suppose that the benchmark prior is given by a mixture of two point masses at* α_1 *and* α_2 *with* $\alpha_1 < \alpha_2$,

$$\pi_{\phi_0}(\alpha) = w 1_{\alpha_1}(\alpha) + (1-w) 1_{\alpha_2}(\alpha).$$

Then

$$\delta^*(\tau) = \max\left\{ \alpha_1, \min\left\{ \begin{array}{c} (1-\tau)\alpha_1 + \tau\alpha_2 + \kappa \ln\left(\frac{(1-w)\tau}{w(1-\tau)}\right), \\ \alpha_2 \end{array} \right\} \right\}.$$

This implies, as $\kappa \to 0$, $\delta^*(\tau) \to (1-\tau)\alpha_1 + \tau\alpha_2$.

Proof. For $\delta \in [\alpha_1, \alpha_2]$, the first order condition (63) can be written as

$$w(1-\tau)\exp\left\{\frac{(1-\tau)(\delta-\alpha_1)}{\kappa}\right\} - (1-w)\tau \exp\left\{\frac{\tau(\alpha_2-\delta)}{\kappa}\right\} = 0.$$

This solves for

$$\delta^*(\tau) = (1-\tau)\alpha_1 + \tau\alpha_2 + \kappa \ln\left(\frac{(1-w)\tau}{w(1-\tau)}\right).$$

Hence, if $\delta^*(\tau) \in [\alpha_1, \alpha_2]$, $\delta^*(\tau)$ is the optimum, and otherwise, it can be shown that either $\delta = \alpha_1$ or $\delta = \alpha_2$ becomes an optimum. Hence, the conclusion follows. ■

Proposition B.4 *Suppose that the benchmark prior is given by a mixture of* M *point masses at* $\alpha_1 < \alpha_2 < \cdots < \alpha_M$,

$$\pi^*_{\phi_0}(\alpha) = \sum_{m=1}^{M} w_m 1_{\alpha_m}(\alpha).$$

Then, $\delta^*(\tau) \to (1-\tau)\alpha_1 + \tau\alpha_M$, *as* $\kappa \to 0$.

Proof. Let $\delta^* = (1-\tau)\,\alpha_1 + \tau\alpha_M$, and let $m^* \in \{1, \ldots, (M-1)\}$ be the index such that $\alpha_i \leq \delta^*$ for all $i = 1, \ldots, m^*$, and $\alpha_i > \delta^*$ for all $i = (m^*+1), \ldots, M$. Then, if $\delta^*(\tau) \in [\alpha_{m^*}, \alpha_{m^*+1}]$, the first order condition (64) should hold as

$$\begin{aligned}
\delta^*(\tau) &= \kappa \ln \left(\exp\left(\frac{\delta^*}{\kappa}\right) \left(\frac{\tau}{1-\tau}\right) \frac{\sum_{m=m^*+1}^{M} w_m \exp\left(\frac{\tau(\alpha_m - \alpha_M)}{\kappa}\right)}{\sum_{m=1}^{m^*} w_m \exp\left(\frac{(1-\tau)(\alpha_1-\alpha_m)}{\kappa}\right)} \right) \\
&= \delta^* + \kappa \ln \left(\left(\frac{\tau}{1-\tau}\right) \frac{w_M + \sum_{m=m^*+1}^{M-1} w_m \exp\left(\frac{\tau(\alpha_m - \alpha_M)}{\kappa}\right)}{w_1 + \sum_{m=2}^{m^*} w_m \exp\left(\frac{(1-\tau)(\alpha_1-\alpha_m)}{\kappa}\right)} \right) \\
&\to \delta^* \quad \text{as } \kappa \to 0.
\end{aligned}$$

Since $\delta^* \in [\alpha_{m^*}, \alpha_{m^*+1}]$ by the construction of m^*, $\delta^*(\tau) \in [\alpha_{m^*}, \alpha_{m^*+1}]$ should hold for κ small enough. ∎

C Game Theoretic Model

For the entry game considered in example 1.3, the reduced form parameters ϕ relates to the full structural parameter $\tilde{\theta} = (\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$ by

$$\begin{aligned}
\phi_{11} &= G(\beta_1 - \gamma_1)G(\beta_2 - \gamma_2), \\
\phi_{00} &= (1 - G(\beta_1))(1 - G(\beta_2)), \\
\phi_{10} &= G(\beta_1)\left[1 - G(\beta_2)\right] + G(\beta_1 - \gamma_1)\left[G(\beta_2) - G(\beta_2 - \gamma_2)\right] \\
&\quad + \psi\left[G(\beta_1) - G(\beta_1 - \gamma_1)\right]\left[G(\beta_2) - G(\beta_2 - \gamma_2)\right].
\end{aligned} \tag{65}$$

where $G(\cdot)$ is the cdf of the standard normal distribution.

As a benchmark prior $\pi_{\tilde{\theta}}(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$, consider for example Priors 1 and 2 in Moon and Schorfheide (2012). Posterior draws of $\tilde{\theta}$ can be obtained by the Metropolis-Hastings Algorithm or its variant. Plug them into (65) the yields the posterior draws of ϕ. The transformation

(65) offers the following one-to-one reparametrization mapping between $\tilde{\theta}$ and $(\beta_1, \gamma_1, \phi)$:

$$\begin{aligned}
\beta_1 &= \beta_1, \\
\gamma_1 &= \gamma_1, \\
\beta_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) \equiv \beta_2(\beta_1, \phi), \\
\gamma_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) - G^{-1}\left(\frac{\phi_{11}}{G(\beta_1 - \gamma_1)}\right) \equiv \gamma_2(\beta_1, \gamma_1, \phi), \\
\psi &= \frac{[1 - G(\beta_1)]\,[\phi_{10} + \phi_{11} - G(\beta_1 - \gamma_1)] + [G(\beta_1) - G(\beta_1 - \gamma_1)]\,\phi_{00}}{[G(\beta_1) - G(\beta_1 - \gamma_1)]\left[1 - G(\beta_1) - \phi_{00} - \frac{1 - G(\beta_1)}{G(\beta_1 - \gamma_1)}\phi_{11}\right]} \equiv \psi(\beta_1, \gamma_1, \phi).
\end{aligned} \tag{66}$$

As in the SVAR example above, the conditional benchmark prior for $\theta = (\beta_1, \gamma_1)$ given ϕ satisfies

$$\pi_{\theta|\phi}(\beta_1, \gamma_1) \propto \pi_{\tilde{\theta}}(\beta_1, \gamma_1, \beta_2(\beta_1, \phi), \gamma_2(\beta_1, \gamma_1, \phi), \psi(\beta_1, \gamma_1, \phi)) \times |\det(J(\beta_1, \gamma_1, \phi))|,$$

where $J(\beta_1, \gamma_1, \phi)$ is the Jacobian of the transformation shown in (66). Solving for the multiplier minimax estimator for γ_1 follows similar steps to those in Algorithm 7.1, except for a slight change in Step 1. Now, in the importance sampling step given a draw of ϕ, we draw (β_1, γ_1) jointly from a proposal distribution $\tilde{\pi}_{\theta|\phi}(\beta_1, \gamma_1)$ even though the object of interest is γ_1 only. That is, to approximate $r_\kappa(\delta, \phi) = \ln \int_{IS_{\gamma_1}(\phi)} \exp\{h(\delta, \gamma_1)/\kappa\}\, d\pi^*_{\gamma_1|\phi}$, we draw N draws of (β_1, γ_1), from a proposal distribution $\tilde{\pi}_{\theta|\phi}(\beta_1, \gamma_1)$ (e.g., a diffuse bivariate normal truncated to $\gamma_1 \geq 0$) and compute

$$\hat{r}_\kappa(\delta, \phi_m) = \ln\left[\frac{\sum_{i=1}^N w(\beta_{1i}, \gamma_{1i}, \phi)\exp\{h(\delta, \gamma_{1i})/\kappa\}}{\sum_{i=1}^N w(\beta_{1i}, \gamma_{1i}, \phi)}\right],$$

where

$$w(\beta_1, \gamma_1, \phi) = \frac{\pi_{\tilde{\theta}}(\beta_1, \gamma_1, \beta_2(\beta_1, \phi), \gamma_2(\beta_1, \gamma_1, \phi), \psi(\beta_1, \gamma_1, \phi)) \times |\det(J(\beta_1, \gamma_1, \phi))|}{\tilde{\pi}_{\theta|\phi}(\beta_1, \gamma_1)}.$$

References

Armstrong, T. and M. Kolesár (2019): "Sensitivity Analysis using Approximate Moment Condition Models," *unpublished manuscript.*

Baumeister, C. and J. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

Berger, J. (1985): *Statistical Decision Theory and Bayesian Analysis*, New York, NY: Springer-Verlag, 2nd ed.

Berger, J. and L. Berliner (1986): "Robust Bayes and Empirical Bayes Analysis with ϵ-contaminated Priors," *The Annals of Statistics*, 14, 461–486.

Betró, B. and F. Ruggeri (1992): "Conditional Γ-minimax Actions Under Convex Losses," *Communications in Statistics, Part A - Theory and Methods*, 21, 1051–1066.

Bonhomme, S. and M. Weidner (2018): "Minimizing Sensitivity to Model Misspecification," *cemmap working paper 59/18*.

Bresnahan, T. and P. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57–81.

Chamberlain, G. (2000): "Econometric Applications of Maxmin Expected Utility," *Journal of Applied Econometrics*, 15, 625–644.

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

Christensen, T. and B. Connault (2019): "Counterfactual Sensitivity and Robustness," *unpublished manuscript.*

DasGupta, A. and W. Studden (1989): "Frequentist Behavior of Robust Bayes Estimates of Normal Means," *Statistics and Decisions*, 7, 333–361.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Dupuis, P. and R. S. Ellis (1997): *A Weak Convergence Approach to the Theory of Large Deviations*, New York: Wiley.

Giacomini, R. and T. Kitagawa (2018): "Robust Bayesian Inference for Set-identified Models," *Cemmap working paper.*

Giacomini, R., T. Kitagawa, and A. Volpicella (2018): "Uncertain Identification," *Cemmap working paper.*

Gilboa, I. and M. Marinacci (2016): "Ambiguity and Bayesian Paradigm," in *Readings in Formal Epsitemology*, ed. by H. Arló-Costa, V. Hendricks, and J. Benthem, Springer, vol. 1, 385–439.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility With Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

——— (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59, 33–49.

Good, I. (1965): *The Estimation of Probabilities*, MIT Press.

Hansen, L. P. and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review, AEA Papers and Proceedings*, 91, 60–66.

Ho, P. (2019): "Global Robust Bayesian Analysis in Large Models," *unpublished manuscript.*

Jaffray, Y. (1992): "Bayesian Updating and Belief Functions," *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 1144–1152.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81, 1185–1201.

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

Lavine, M., L. Wasserman, and R. Wolpert (1991): "Bayesian Inference with Specified Prior Marginals," *Journal of the American Statistical Association*, 86, 964–971.

Leamer, E. (1981): "Is It a Supply Curve, or Is It a Demand Curve: Partial Identification through Inequality Constraints," *Review of Economics and Statistics*, 63, 319–327.

——— (1982): "Sets of Posterior Means with Bounded Variance Priors," *Econometrica*, 50, 725–736.

Liao, Y. and A. Simoni (2013): "Semi-parametric Bayesian Partially Identified Models based on Support Function," *unpublished manuscript.*

Manski, C. (1981): "Learning and Decision Making When Subjective Probabilities Have Subjective Domains," *Annals of Statistics*, 9, 59–65.

Manski, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

Moon, H. and F. Schorfheide (2011): "Bayesian and Frequentist Inference in Partially Identified Models," *NBER working paper*.

——— (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

Moreno, E. (2000): "Global Bayeisan Robustness for Some Classes of Prior Distributions," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics.

Newey, W. K. and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics Volume 4*, ed. by R. F. Engle and D. L. McFadden, Amsterdam, The Netherlands: Elsevier.

Peterson, I. R., M. R. James, and P. Dupuis (2000): "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints," *ISSS Transactions on Automatic Control*, 45, 398–412.

Pires, C. (2002): "A Rule for Updating Ambiguous Beliefs," *Theory and Decision*, 33, 137–152.

Poirier, D. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14, 483–509.

Robbins, H. (1951): "Asymptotically Sub-minimax Solutions to Compound Statistical Decision Problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*.

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

Sims, C. and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Reiwews*, 39, 949–968.

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Vidakovic, B. (2000): "Γ-minimax: A Paradigm for Conservative Robust Bayesians," in *Robust Bayesian Analysis*, ed. by D. R. Insua and F. Ruggeri, Springer, Lecture Notes in Statistics.

Wasserman, L. (1990): "Prior Envelopes Based on Belief Functions," *The Annals of Statistics*, 18, 454–464.

Wong, R. (1989): *Asymptotic Approximations of Integrals*, New York: Wiley.