cemmap
centre for microdata methods and practice

Estimation Under Ambiguity

Raffaella Giacomini
Toru Kitagawa
Harald Uhlig

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP24/19

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

An ESRC Research Centre

Estimation Under Ambiguity*

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

This draft: May 2019

Abstract

To perform Bayesian analysis of a partially identified structural model, two distinct approaches exist: standard Bayesian inference, which assumes a single prior for the structural parameters, including the non-identified ones; and multiple-prior Bayesian inference, which assumes full ambiguity for the non-identified parameters. 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 the large gap between the two approaches by proposing a multiple-prior Bayesian analysis that can simultaneously incorporate a probabilistic belief for the non-identified parameters and a concern about misspecification of 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 ambiguity." We obtain point estimators and optimal decisions involving non-identified parameters by solving a conditional gamma-minimax problem, which we show is analytically tractable and easy to solve numerically. We derive the remarkably simple analytical properties of the proposed procedure in the limiting situations where the radius of the KL neighborhood and/or the sample size are large. Our procedure can also be used to perform global sensitivity analysis.

*We would like to thank Stephane Bonhomme, Lars Hansen, Frank Kleibergen, Matthew Read 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

This paper develops a formal framework for robust Bayesian inference in partially identified structural models that accommodates a concern for misspecification of the researcher's prior knowledge. The framework can be used to perform global sensitivity analysis by constructing a class of priors in a neighborhood of a benchmark prior and obtaining the optimal posterior minimax decision (e.g., a point estimator) over this class.

The focus on partially identified models is a novel and natural starting point of the analysis, for two main reasons. First, prior misspecification is more of a concern under partial identification, because the effect of the prior does not disappear asymptotically, unlike in the point identified case (Poirier, 1998). Second, in partially identified models we can overcome a challenge that would arise under point identification. In that case, the priors in the class would generally correspond to different marginal likelihoods, and thus the minimax decision could be driven by a worst-case prior that fits the data very poorly. We believe that a more empirically useful characterization of sensitivity instead considers perturbations that preserve the marginal likelihood of the benchmark prior. This is readily accomplished under partial identification, where it is arguably natural to only perturb the unrevisable component of the prior.[1]

To motivate and illustrate our approach, consider the setting of a partially identified model. Following Poirier (1998) and Moon and Schorfheide (2012), suppose that the distribution of observables can be indexed by a vector of finite-dimensional reduced-form parameters $\phi \in \Phi$, but that knowledge of ϕ and additional a-priori restrictions fails to point identify the structural parameters $\theta \in \Theta$ and the *scalar* object of interest $\alpha = \alpha(\theta, \phi) \in \mathbb{R}$. We thus suppose that ϕ is identifiable (i.e., there are no $\phi, \phi' \in \Phi$, $\phi \neq \phi'$, that are observationally equivalent), while θ and α are not, even with the a-priori restrictions (which can depend on ϕ), denoted as $\theta \in \Theta_R(\phi) \subset \Theta$.

Let X be a random sample and x its realization. The value of the likelihood $l(x|\theta, \phi)$ depends only on ϕ for every realization of X, or, equivalently, $X \perp \theta | \phi$. We refer to the set of θ compatible with the value of ϕ and the a-priori restrictions as the identified set of θ, $IS_\theta(\phi)$. The identified set of α is accordingly defined by the values 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 $\mathbb{R}$.

[1] A way to partially overcome the challenge in point-identified models would be to shrink the class of priors, possibly adaptively to the data, for example by considering only priors with marginal likelihood above a certain threshold. We leave this extension for future work.

The following examples illustrate the set-up. Appendix D contains a further detailed illustration in the context of an entry game.

Example 1.1 (Supply and demand) *Suppose the object of interest α is a structural parameter in a system of simultaneous equations, e.g., a classical model of labor supply and demand:*

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

where: $x_t = (\Delta w_t, \Delta n_t)$ with Δw_t and Δn_t the growth rates of wages and employment, respectively; u_t are shocks assumed to be i.i.d. $N(0, D)$ with $D = diag(d_1, d_2)$; and $A = \begin{bmatrix} -\beta_d & 1 \\ -\beta_s & 1 \end{bmatrix}$ with β_s the short-run wage elasticity of supply and β_d the short-run wage elasticity of demand. The a-priori restrictions are $\beta_s \geq 0$ and $\beta_d \leq 0$. The reduced-form representation 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 reparameterized to $(\beta_s, w_{11}, w_{12}, w_{22})$.[2] *In our notation, θ can thus 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 Leamer, 1981):*

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

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

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

The reduced-form parameters are $\phi = (vec(B_1)', \ldots, vec(B_p)', vech(\Omega)')' \in \Phi$, where $vech(\Omega)$ is the vectorization of the lower triangular portion of Ω (see Lütkepohl, 1991), Φ is restricted to the set of ϕ such that the reduced-form VAR can be inverted into a $VMA(\infty)$ model:

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

[2] See Section 6.1 below for the transformation. If β_d is the parameter of interest, an alternative reparameterization transforms the structural parameters into $(\beta_d, w_{11}, w_{12}, w_{22})$.

The non-identified parameter is $\theta = vec(Q)$, where Q is the orthonormal matrix tranforming 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, capturing the effect on the i-th variable in x_{t+h} of a unit shock to the j-th element of u_t: $\alpha = e_i'C_h\Omega_{tr}Qe_j$, with e_i the $i-th$ column of I. The identified set of the $(i,j)-th$ impulse response in the absence of identifying restrictions is

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

where $\mathcal{O}$ is the space of orthonormal matrices. Additional a priori restrictions may be imposed, such as sign restrictions on the impulse responses (e.g., Uhlig, 2005).

The identified set collects all the admissible values of α that satisfy the imposed identifying assumptions, given the data. Often, however, the researcher has some form of additional but only partially credible assumptions about structural parameters based on economic theory, background knowledge, or empirical studies that use different data. Alternatively, she may wish to impose a-priori indifference among the parameters within the identified set. The standard Bayesian recommendation is to incorporate this information by specifying a prior for (θ, ϕ). For instance, in the case of Example 1.1, Baumeister and Hamilton (2015) propose a prior that draws on estimates of the elasticity parameters obtained in macroeconomic and microeconometric studies, and consider independent Student's t distributions calibrated to assign 90% probability to the intervals $\beta_s \in (0.1, 2.2)$, and $\beta_d \in (-2.2, -0.1)$. Another example in Baumeister and Hamilton (2015) is a prior that incorporates long-run identifying restrictions in SVARs non-dogmatically, as a way to capture uncertainty about these controversial restrictions. See also Baumeister and Hamilton (2018) and Baumeister and Hamilton (2019) for further applications of this approach. In situations where the researcher wants to impose indifference among values within the identified set, a uniform prior has often been recommended. 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. Other examples of the uniform prior appear in Moon and Schorfheide (2012) for an entry game and in Norets and Tang (2014) for a dynamic discrete choice model.

At the opposite end of the standard Bayesian spectrum, Giacomini and Kitagawa (2018) advocate adopting a fully ambiguous 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, the 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, which leads to an estimator for $IS_\alpha(\phi)$ with an associated "robust" credible region that asymptotically converges to the true identified set with a desired frequentist coverage.

The motivation for the methods we propose in this paper is the observation that the prior inputs considered by the two extreme approaches discussed above – a precise specification of a prior for (θ, ϕ), or full ambiguity about the conditional prior of θ given ϕ – could be a poor representation of the belief that the researcher possesses in a given application. For example, the prior specified by Baumeister and Hamilton (2015) in Example 1.1 builds on the set of plausible values of the elasticity parameters obtained by previous empirical studies. Such prior evidence, however, may not be sufficient for the researcher to be confident in the particular shape of the prior. At the same time, the fully ambiguous approach may not be attractive because it entirely discards available prior evidence for the elasticity parameters. In another example, a researcher who expresses indifference over values within the identified set by specifying a uniform prior for θ given ϕ may worry that this can induce unintentionally informative priors for α or other parameters (as discussed by Baumeister and Hamilton, 2015). On the other hand, full ambiguity may also be an unappealing representation of prior indifference, since, for example, it treats equally priors that support any value in the identified set and priors that are degenerate at extreme values in the identified set, which could appear less sensible.

The main contribution of this paper is to fill the large gap between the two extreme approaches by proposing a method that can simultaneously incorporate a probabilistic belief for structural parameters and a misspecification concern about this belief. Our idea is to replace the fully ambiguous beliefs considered in Giacomini and Kitagawa (2018) by a class of priors defined in a KL neighborhood of a *benchmark prior* for θ given ϕ. The benchmark prior represents the researcher's reasonable but partially credible prior knowledge about θ given ϕ, and the class of priors in the neighborhood captures ambiguity or misspecification concerns about the benchmark prior. The radius of the neighborhood is chosen by the researcher and controls the degree of confidence in the benchmark prior. We then obtain point estimators of α and other statistical decisions involving α by minimizing the worst-case (minimax) posterior expected loss with respect to the priors in the neighborhood. The proposed framework is useful for assessing the sensitivity of the posterior for α to perturbations of the unrevisable component of the benchmark prior. The fact that we perturb the prior for θ given ϕ, while keeping the prior for ϕ fixed, implies that all priors in the class share the same marginal likelihood. This ensures that posterior sensitivity is not driven by priors that poorly fit the data.

Our paper makes the following unique contributions: (1) we clarify that estimation of a nonidentified parameter under vague prior knowledge can be formulated as a decision under ambiguity, as considered in the literature on robust control methods (e.g., Hansen and Sargent, 2001); (2) we provide an analytically tractable and numerically convenient way to solve the 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 the benchmark prior is irrelevant; and (4) we derive the properties of our method in large samples.

1.1 Related Literature

The idea of introducing a set of priors to draw robust posterior inference goes back to the robust Bayesian analysis of Robbins (1951), whose basic premise is that the decision-maker cannot specify a unique prior distribution for the parameters due to limited prior knowledge or limited ability to elicit the prior. Good (1965) argues that the prior input that is 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, there is no consensus in the literature on how to update after observing the data. 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, considered in Jaffray (1992) and axiomatized by Pires (2002). Here we introduce a single prior for the reduced-form parameters and a class of priors for the non-identified parameters, which corresponds to the part of the prior distribution that is unrevisable by the data. Since any prior in the class leads to the same value of the marginal likelihood due to the single prior for the reduced-form parameters, we obtain the same set of posteriors no matter what updating rule we apply.

We perform minimax estimation by applying the minimax criterion to the set of posteriors, which is referred to as the conditional gamma-minimax criterion in the statistics literature (e.g., DasGupta and Studden, 1989; Betró and Ruggeri, 1992). This is distinguished from the (unconditional) gamma-minimax criterion where minimax is performed before observing the data (e.g., Manski, 1981; Berger, 1985; Chamberlain, 2000; Vidakovic, 2000). An analogue to gamma-minimax analysis in economic 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 considered various ways of constructing a prior class, including the class of bounded and unbounded variance priors (Chamberlain and Leamer, 1976; Leamer, 1982), the ϵ-contaminated class of priors

(Berger and Berliner, 1986), the class of priors built on a nonadditive lower probability (Wasserman, 1990), and the class of priors with a fixed marginal distribution (Lavine et al., 1991), to list a few. This paper focuses on partially-identified models where the sensitivity of the posterior remains present even in large samples due to the lack of identification. The class of priors proposed in this paper consists of those belonging to a specified Kullback-Leibler (KL) neighborhood around the benchmark prior. As shown in Lemma 2.2 below, the conditional gamma-minimax analysis with this class of priors is closely related to the 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 radius can replicate the class of priors considered in Giacomini and Kitagawa (2018).

Our procedure can be used for global sensitivity analysis. See Moreno (2000) and references therein for existing approaches in the statistics literature. For global sensitivity analysis in point-identified models, Ho (2019) considers a KL based class of priors similar to ours. His approach, if applied to set-identified models, would differ from ours in the following aspects. First, all priors in our prior class share a single prior for the reduced-form parameters, while this is not necessarily the case in Ho (2019). Allowing for multiple priors for the reduced-form parameters implies that a prior that fits the data poorly, i.e., that is far from the observed likelihood, will yield the worst-case posterior. Obtaining a large range of posteriors could thus be due to allowing for priors that are severely in conflict with the data, rather than being an indication of posterior non-robustness (i.e., the lack of information in the observed likelihood). In our approach, in contrast, all the posteriors in the class share the same value of the marginal likelihood. This allows us to assess posterior sensitivity while keeping the denominator of the Bayes rule constant. Second, our approaches differ in how to select the radius of the KL neighborhood. Ho (2019) recommends to set the radius so that the set of priors can span the posterior means around the Gaussian-approximated benchmark posterior. This approximation is reasonable only when the model is point-identified. In contrast, we propose to specify the radius by matching the set of prior means or other quantities for a parameter with available prior knowledge about it. We consider this empirically appealing, because a researcher often has access to prior knowledge in the form of inequalities or an interval for a parameter.

The robustness concern we address is about misspecification of the prior in a Bayesian setting. The frequentist approach to robustness concerns instead misspecification of the likelihood, identifying assumptions, moment conditions, or the distribution of unobservables. This type of sensitivity analysis is considered by, e.g., Andrews et al. (2017), Armstrong and Kolesár (2019), Bonhomme and Weidner (2018), Christensen and Connault (2019), Kitamura et al. (2013).

1.2 Roadmap

The remainder of the paper is organized as follows. Section 2 introduces the analytical framework and formulates the statistical decision problem with multiple priors localized around the benchmark prior. Section 3 solves the constrained posterior minimax problem for a general loss function. Section 4 applies the framework to global sensitivity analysis. For the quadratic and check loss functions, Section 5 analyzes point and interval estimation of the parameter of interest. Section 5 also considers two types of limiting situations: (1) the radius of the set of priors goes to infinity (fully ambiguous beliefs) and (2) the sample size goes to infinity. Section 6 discusses the implementation of the method with particular emphasis on how to elicit the benchmark prior and how to select the tuning parameter that governs the size of the prior class. Section 7 provides an empirical illustration of the method. Appendix A contains the proofs; Appendix B considers the asymptotic analysis for a discrete benchmark prior; Appendix C provides details about the construction of the benchmark prior in the empirical analysis and Appendix D discusses the approach in the contest of an entry game example.

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 structural parameter θ on Θ given the reduced-form parameter ϕ and π_{ϕ} is a marginal prior probability measure of ϕ. Imposing the additional a priori restrictions $\theta \in \Theta_R(\phi)$ implies that the support of $\pi_{\theta|\phi}$ is a subset of or all of $IS_{\theta}(\phi)$. Since $\alpha = \alpha(\theta, \phi)$ is a function of θ given ϕ, $\pi_{\theta|\phi}$ induces a conditional prior distribution of α given ϕ, the domain of which is a subset of or equal to $IS_{\alpha}(\phi)$ if the a priori restrictions are imposed. While sample data X are informative about ϕ and enable the researcher to update the prior π_{ϕ} to obtain the posterior $\pi_{\phi|X}$, the conditional prior $\pi_{\theta|\phi}$ (and hence $\pi_{\alpha|\phi}$) can never be updated by data and posterior inference for α remains sensitive to the choice of conditional prior no matter how large the sample size is. Therefore, misspecification of the unrevisable part of the prior $\pi_{\alpha|\phi}$ may be a major concern in conducting posterior inference for a decision-maker in practice.

Suppose that the decision-maker can form a benchmark prior $\pi^*_{\theta|\phi}$ and possibly imposes additional a priori restrictions $\theta \in \Theta_R(\phi)$, so that the support of $\pi^*_{\theta|\phi}$ is a subset of or equal to $IS_{\theta}(\phi)$. The benchmark prior 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). The benchmark prior for θ given ϕ induces a 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 reduce $\pi^*_{\alpha|\phi}$ to a point mass measure supported at the singleton identified set, and the posterior of ϕ would induce the single posterior of α. Generally, though, $\pi^*_{\theta|\phi}$ determines how the probabilistic belief is allocated within the identified non-singleton set $IS_\alpha(\phi)$.

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

Given some specification for the distance $\mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi})$ between two probability measures $\pi_{\theta|\phi}$ and $\pi^*_{\theta|\phi}$, a λ-neighborhood around the benchmark conditional prior at ϕ is the set

$$\Pi^\lambda\left(\pi^*_{\theta|\phi}\right) \equiv \left\{\pi_{\theta|\phi} : \mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi}) \leq \lambda\right\}. \tag{8}$$

For the specification of the distance $\mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi})$, and in line with Hansen and Sargent (2001) and a considerable literature, we choose the KL divergence from $\pi^*_{\theta|\phi}$ to $\pi_{\theta|\phi}$, or equivalently the relative entropy of $\pi_{\theta|\phi}$ relative to $\pi^*_{\theta|\phi}$, defined by

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

$\mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi})$ 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 is 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 in $\pi_{\theta|\phi}$, which guarantees that the constrained minimax problem (11) 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

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

Alternatively, one could form the KL neighborhood for the *unconditional* prior of (θ, ϕ) around its benchmark, as considered in Ho (2019). However, our approach to constructing priors simplifies the multiple-priors analysis both analytically and numerically.

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 unconditional prior $\pi_{\theta\phi}$. First, the major source of posterior sensitivity comes from $\pi_{\theta|\phi}$, because the effect of the prior for ϕ on the posterior is tempered by the likelihood. Our aim is to make estimation and inference robust to the prior input that can not be updated by the data. Second, allowing for multiple priors for ϕ, a prior that fits the data poorly (i.e., a prior that gives small marginal likelihood) will yield the worst-case posterior. Then, the range of posteriors would mix up the posterior sensitivity due to the lack of information in the likelihood (numerator in the Bayes formula) and the possibility that some priors in the class are severely in conflict with the data (denominator in the Bayes formula). Keeping π_ϕ fixed, in contrast, ensures that the marginal likelihood is common among the class so that the range of posteriors reflect purely the posterior sensitivity. Third, keeping π_ϕ fixed implies that the updating rules for the set of priors proposed in the literature on 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 (2012), all lead to the same set of posteriors. This means that the minimax decision after X is observed is invariant to the choice of the updating rule, which is not necessarily the case if one allows for multiple priors for ϕ.

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 interpretable scale. It is therefore challenging to translate the subjective notion of "credibility" of the benchmark prior into a choice of λ. Section 6 below proposes a practical way to elicit λ.

2.2 Posterior Minimax Decision

We first consider statistical decision problems in the presence of multiple priors and posteriors generated by $\Pi^\lambda(\pi^*_{\theta|\phi})$. We focus on point estimation of the scalar parameter of interest α. However, 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 the data X to a space of actions $\mathcal{D} \subset \mathbb{R}$, and let $h(\delta(X), \alpha)$ be a loss function. In the context of point estimation, the loss function can be, for instance, the quadratic loss

$$h(\delta(X), \alpha) = (\delta(X) - \alpha)^2 , \tag{9}$$

or the check loss for the τ-th quantile, $\tau \in (0,1)$,

$$\begin{aligned} h(\delta(X), \alpha) &= \rho_\tau (\alpha - \delta(X)) , \\ \rho_\tau(u) &= \tau u \cdot 1 \{u > 0\} - (1-\tau) u \cdot 1 \{u < 0\} . \end{aligned} \tag{10}$$

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

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

We assume an ambiguity-averse decision-maker reaches an optimal decision by applying the conditional gamma-minimax criterion, i.e., by choosing $\delta(x)$ to minimize the worst-case posterior expected loss when $\pi_{\theta|\phi}$ varies over $\Pi^\lambda(\pi^*_{\theta|\phi})$ for every $\phi \in \Phi$. We call this the *constrained posterior minimax problem*, formally given by

$$\begin{aligned} &\min_{\delta(x) \in \mathcal{D}} \max_{\pi_{\theta\phi} \in \Pi^\lambda_{\theta\phi}} \int_\Phi \left[\int_{IS_\theta(\phi)} h(\delta(x), \alpha(\theta, \phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X} \\ = &\min_{\delta(x) \in \mathcal{D}} \int_\Phi \max_{\pi_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)} \left[\int_\Theta h(\delta(x), \alpha(\theta, \phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X} . \end{aligned} \tag{11}$$

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 $\phi \in \Phi$. Note also that, since any $\pi_{\theta|\phi} \in \Pi^\lambda(\pi^*_{\theta|\phi})$ has the support contained in $IS_\theta(\phi)$, the region of integration with respect to θ can be extended from $IS_\theta(\phi)$ to the whole parameter space Θ without changing the value of the integral, so that

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

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

Since the loss function $h(\delta, \alpha(\theta, \phi))$ depends on θ only through the parameter of interest α, we can work with the set of priors for α given ϕ instead of θ given ϕ. Specifically, we consider 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 solve the following constrained posterior minimax problem:

$$\min_{\delta(x)\in\mathcal{D}} \int_{\Phi} \max_{\pi_{\alpha|\phi}\in\Pi^{\lambda}(\pi^*_{\alpha|\phi})} \left[\int_{IS_{\alpha}(\phi)} h(\delta(x),\alpha) d\pi_{\alpha|\phi}\right] d\pi_{\phi|X}. \tag{12}$$

$\Pi^{\lambda}(\pi^*_{\alpha|\phi})$ 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 the minimax problems (11) and (12) 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 for the entire non-identified parameter θ conditional on ϕ or only for the parameter of interest α conditional on ϕ while being agnostic about the conditional prior of $\theta|\alpha,\phi$, the constrained minimax problem supports the same decision as optimal, as far as a common λ is specified. This lemma therefore justifies ignoring ambiguity about the set-identified parameters other than α and focusing only on the set of priors of $\alpha|\phi$ that ultimately matter for the posterior expected loss.

A minimax problem closely related to the constrained posterior minimax problem formulated in (12) above is the *multiplier posterior minimax problem*:

$$\min_{\delta(x)\in\mathcal{D}} \int_{\Phi} \left[\max_{\pi_{\alpha|\phi}\in\Pi^{\infty}(\pi^*_{\alpha|\phi})} \left\{\int_{IS_{\alpha}(\phi)} h(\delta(x),\alpha) d\pi_{\alpha|\phi} - \kappa\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi})\right\}\right] d\pi_{\phi|X}, \tag{13}$$

where $\kappa \geq 0$ is a fixed constant. The next lemma, borrowed from the robust control literature, shows the relationship between the inner maximization problems in (12) and (13):

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*

$$r_\lambda(\delta, \phi) \equiv \max_{\pi_{\alpha|\phi} \in \Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left[\int_{IS_\alpha(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi}\right]. \tag{14}$$

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

$$r_\lambda(\delta, \phi) = \max_{\pi_{\alpha|\phi} \in \Pi^\infty(\pi^*_{\alpha|\phi})} \left\{\int_{IS_\alpha(\phi)} h(\delta, \alpha) d\pi_{\alpha|\phi} - \kappa_\lambda(\delta, \phi)\left(\mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) - \lambda\right)\right\}. \tag{15}$$

Furthermore, if $\pi^0_{\alpha|\phi} \in \Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)$ is a maximizer in (14), $\pi^0_{\alpha|\phi}$ also maximizes (15) 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 (14), whose value depends on λ. Furthermore, the $\kappa_\lambda(\delta, \phi)$ that makes the constrained optimization (14) and the unconstrained optimization (15) 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 starting from (13) with constant $\kappa > 0$ independent of ϕ and δ, an implied value of λ that equalizes (14) and (15) 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 posterior minimax problem with constant κ appears analytically and numerically simpler than the constrained posterior minimax problem with constant λ. However, its non-desirable feature is that the implied class of priors (or the radius of the KL neighborhood) is endogenously determined depending on what loss function one specifies. Since our robust Bayes analysis sets the set of priors as the primary input which is invariant to the choice of loss function, we focus on the constrained posterior minimax problem (12) with constant λ rather than the multiplier posterior minimax problem (13) with fixed κ. This approach is also consistent with the standard Bayesian global sensitivity analysis where the sets of posterior quantities are computed with the same set of priors no matter whether one focuses on posterior means or quantiles.

3 Solving the Constrained Posterior Minimax Problem

3.1 Finite-Sample Solution

The inner maximization in the constrained minimax problem of (12) has an analytical solution, as shown in the next theorem.

Theorem 3.1 *Assume at any $\delta \in \mathcal{D}$ and $\kappa > 0$, $\int_{IS_\alpha(\phi)} \exp(h(\delta,\alpha)/\kappa) d\pi^*_{\alpha|\phi} < \infty$ and the distribution of $h(\delta,\alpha)$ induced by $\alpha \sim \pi^*_{\alpha|\phi}$ is nondegenerate, π_ϕ-a.s. The constrained posterior minimax problem (12) is then equivalent to*

$$\min_{\delta \in \mathcal{D}} \int_\Phi r_\lambda(\delta,\phi)\, d\pi_{\phi|X}, \tag{16}$$

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

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

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. ■

This theorem is valid for any sample size and any realization of X. The benchmark prior $\pi^*_{\alpha|\phi}$ can be continuous, discrete, or their mixture as long as it generates stochastic variation in $h(\delta,\alpha)$ as assumed in the theorem. The obtained representation simplifies the analytical investigation of the minimax decision, and offers a simple way to approximate the objective function in (16) 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. Note also that the the worst-case prior $\pi^0_{\alpha|\phi}$ and the minimizer of the worst-case risk $r_\lambda(\delta,\phi)$ is invariant to monotonic affine transformations to the loss function.

3.2 Large-Sample Behavior

Investigating the large-sample approximation of the minimax decision can suggest further computational simplifications. 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) (Bounded loss) The loss function $h(\delta, \alpha)$ is bounded,

$$|h(\delta, \alpha)| \leq H < \infty,$$

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 benchmark prior means) There exists $G_0 \subset \Phi$ an open neighborhood of ϕ_0 such that for any $\kappa \in (0, \infty)$,

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

where $E^*_{\alpha|\phi}(\cdot)$ is the expectation with respect to the benchmark conditional prior $\pi^*_{\alpha|\phi}$.

(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). Boundedness of the loss function imposed in Assumption 3.2 (ii) can be implied by assuming, for instance, that $h(\delta,\alpha)$ 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 in 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 a 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, 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 a 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* $n \to \infty$*,* $\delta_\lambda(\hat{\phi}) \in \arg\min_{\delta\in\mathcal{D}} r_\lambda\left(\delta,\hat{\phi}\right)$ *converges in probability to* $\delta_\lambda(\phi_0)$ *as* $n \to \infty$*.*

Proof. See Appendix A. ■

[3] We treat the case of discrete benchmark prior in Appendix B, where the loss function is specified to be the quadratic or the check loss function.

Theorem 3.3 shows that the finite-sample constrained posterior minimax decision has a well-defined large-sample limit that coincides with the minimax 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 posterior minimax decision is the ambiguity of the unrevisable part of the prior given $\phi = \phi_0$. The second claim of the theorem has a useful practical implication: when the sample size is large, so that the posterior distribution of ϕ is concentrated around its maximum likelihood estimator (MLE) $\hat{\phi}_{ML}$, the finite-sample posterior minimax decision is well approximated by minimizing the "plug-in" objective function, where the averaging with respect to the posterior of ϕ in (16) is replaced by plugging $\hat{\phi}_{ML}$ in $r_\lambda(\delta, \phi)$. This reduces the computational cost of approximating the objective function since only Monte Carlo draws of θ (or α) from $\pi_{\theta|\hat{\phi}_{ML}}$ (or $\pi_{\alpha|\hat{\phi}_{ML}}$) are needed.

4 Set of Posteriors and Sensitivity Analysis

The analysis so far focused on the minimax decision with a given loss function $h(\delta, \alpha)$. The robust Bayes framework with multiple priors is also useful to computing the set of the posterior quantities (e.g., mean, median, probability of a hypothesis) as the prior varies over the KL neighborhood of the benchmark prior. These sets of posterior quantities can be used in global sensitivity analysis in which one can explore the robustness of posterior inference to perturbations of the benchmark prior.

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

$$E_{\alpha|X}(f(\alpha)) \in \left[\int_\Phi \min_{\pi_{\alpha|\phi}\in\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left(\int f(\alpha) d\pi_{\alpha|\phi}\right) d\pi_{\phi|X},\ \int_\Phi \max_{\pi_{\alpha|\phi}\in\Pi^\lambda\left(\pi^*_{\alpha|\phi}\right)} \left(\int f(\alpha) d\pi_{\alpha|\phi}\right) d\pi_{\phi|X}\right]. \tag{17}$$

The set of posterior means of α can be obtained by setting $f(\alpha) = \alpha$, and the set of posterior probabilities on a subset $B \subset \mathcal{B}$ is obtained by setting $f(\alpha) = 1\{\alpha \in B\}$. The set of posterior quantiles of α can be computed by inverting the set of the cumulative distribution functions (CDFs) of the posteriors of α, which corresponds to setting $f(\alpha) = 1\{\alpha \leq t\}$ in (17). The optimization problems to derive the bounds (17) are identical to the inner maximization in (12), with $h(\delta, \alpha)$ replaced by $f(\alpha)$ or $-f(\alpha)$.

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

$$\left[\int_{\Phi}\left(\int f(\alpha)d\pi^{\ell}_{\alpha|\phi}\right)d\pi_{\phi|X}, \int_{\Phi}\left(\int f(\alpha)d\pi^{u}_{\alpha|\phi}\right)d\pi_{\phi|X}\right], \tag{18}$$

where $\pi^{\ell}_{\alpha|\phi}$ and $\pi^{u}_{\alpha|\phi}$ are the exponentially tilted conditional priors solving the optimizations in (17):

$$\begin{aligned}
d\pi^{\ell}_{\alpha|\phi} &\equiv \frac{\exp\{-f(\alpha)/\kappa^{\ell}_{\lambda}(\phi)\}}{\int \exp\{-f(\alpha)/\kappa^{\ell}_{\lambda}(\phi)\}d\pi^{*}_{\alpha|\phi}} \cdot d\pi^{*}_{\alpha|\phi}, \\
d\pi^{u}_{\alpha|\phi} &\equiv \frac{\exp\{f(\alpha)/\kappa^{u}_{\lambda}(\phi)\}}{\int \exp\{f(\alpha)/\kappa^{u}_{\lambda}(\phi)\}d\pi^{*}_{\alpha|\phi}} \cdot d\pi^{*}_{\alpha|\phi}, \\
\kappa^{\ell}_{\lambda}(\phi) &\equiv \arg\min_{\kappa\geq 0}\left\{\kappa \ln \int \exp\left\{\frac{-f(\alpha)}{\kappa}\right\}d\pi^{*}_{\alpha|\phi} + \kappa\lambda\right\}, \\
\kappa^{u}_{\lambda}(\phi) &\equiv \arg\min_{\kappa\geq 0}\left\{\kappa \ln \int \exp\left\{\frac{f(\alpha)}{\kappa}\right\}d\pi^{*}_{\alpha|\phi} + \kappa\lambda\right\}.
\end{aligned} \tag{19}$$

The large-sample convergence of the worst-case risk shown in Theorem 3.3 immediately leads to the convergence of the set of posterior means of $f(\alpha)$. In addition to Assumptions 3.2 (i) and (v), if (ii), (iv), and (vi) hold in terms of $f(\alpha)$ instead of $h(\delta,\alpha)$, the set of posterior means (18) converges to

$$\left[\int f(\alpha)d\pi^{\ell}_{\alpha|\phi_0}, \int f(\alpha)d\pi^{u}_{\alpha|\phi_0}\right],$$

as $n\to\infty$, where $\pi^{\ell}_{\alpha|\phi_0}$ and $\pi^{u}_{\alpha|\phi_0}$ are the exponentially tilted conditional priors $\pi^{\ell}_{\alpha|\phi}$ and $\pi^{u}_{\alpha|\phi}$ conditional on $\phi=\phi_0$.

A robust Bayesian version of an interval estimator for α is the robust credible region $C_\gamma\subset\mathbb{R}$ with credibility $\gamma\in(0,1)$, defined by an interval satisfying

$$\inf_{\{\pi_{\alpha|\phi}\in\Pi^{\lambda}(\pi^{*}_{\alpha|\phi}):\phi\in\Phi\}} \pi_{\alpha|X}(C_\gamma)\geq\gamma.$$

C_γ can be interpreted as an interval estimate for α on which any posterior belonging to the posterior class assigns probability at least γ. While we might prefer the shortest interval among those satisfying this constraint (Giacomini and Kitagawa (2018)), one simple approach is an *equal-tailed robust credible region* formed by the lower bound of the $\frac{1-\gamma}{2}$-th quantile and the upper bound of the $\frac{1+\gamma}{2}$-th quantile,

$$C_{\gamma,ET}\equiv\left[\underline{F}^{-1}_{\alpha|X}\left(\frac{1-\gamma}{2}\right), \bar{F}^{-1}_{\alpha|X}\left(\frac{1+\gamma}{2}\right)\right],$$

where $\underline{F}_{\alpha|X}(t)$ and $\bar{F}_{\alpha|X}(t)$ are pointwise lower and upper bounds of the posterior CDFs, $\pi_{\alpha|X}(\{\alpha\leq t\})$.

Note that the set of prior means of $f(\alpha)$ is obtained similarly by replacing $\pi_{\phi|X}$ in (17) with the prior π_ϕ:

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

The set of prior means of α or another parameter is a useful object for the purpose of eliciting a reasonable value of λ in light of the researcher's partial prior knowledge for $f(\alpha)$. In Sections 6 and 7 below, we discuss and perform elicitation of λ using the set of prior means and quantiles of α.

5 Large-λ Asymptotics

This section analyzes how the robust Bayes procedures proposed in the previous section perform when the decision-maker has almost no confidence on the benchmark prior, or equivalently, possesses almost no prior belief for the set-identified parameters. We analyze such extreme ambiguity scenario by performing large-λ asymptotic analysis of the gamma-minimax point estimation and the set of posteriors. Our interest in this section is to analytically investigate as $\lambda \to \infty$, (i) whether or not the influence of the benchmark prior can vanish (ii) how the gamma-minimax estimator relates to the identified set, and (iii) whether or not the set of posteriors can recover the underlying identified set, i.e., consistent with the robust Bayes analysis of Giacomini and Kitagawa (2018) imposing the full ambiguity on the identified set.

Maintaining the finite-sample analysis of Section 3.1, we first focus on the limiting situation of $\lambda \to \infty$. We subsequently consider the large sample asymptotics in addition. For gamma-minimax point estimation, we consider two common choices of statistical loss, the quadratic loss and the check loss.

The sets $\Pi^\lambda \left(\pi^*_{\alpha|\phi} \right)$ are increasing with respect to λ in terms of inclusion, and its large λ limit can be defined by

$$\begin{aligned} \Pi^\infty(\pi^*_{\alpha|\phi}) &\equiv \bigcup_{\lambda>0} \Pi^\lambda \left(\pi^*_{\alpha|\phi} \right) \\ &= \left\{ \pi_{\alpha|\phi} : \mathcal{R}(\pi_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) < \infty \right\}, \end{aligned}$$

which contains any probability measure that is absolutely continuous with respect to $\pi^*_{\alpha|\phi}$. We can also consider the closure $\bar{\Pi}^\infty(\pi^*_{\alpha|\phi})$ of $\Pi^\infty(\pi^*_{\alpha|\phi})$ with respect to the weak*-topology, i.e. $\bar{\Pi}^\infty(\pi^*_{\alpha|\phi})$ is the set of all probability measures that are weak limits of probability measures in $\Pi^\infty(\pi^*_{\alpha|\phi})$. $\bar{\Pi}^\infty(\pi^*_{\alpha|\phi})$ contains all the probability measures whose support is the same as or contained in the support of $\pi^*_{\alpha|\phi}$.

5.1 Large-λ Asymptotics in Finite Samples

As $\lambda \to \infty$, the benchmark conditional prior $\pi^*_{\alpha|\phi}$ will affect the gamma-minimax decision only by determining the support of $\pi_{\alpha|\phi}$. To have a precise characterization of this claim and a 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 $\left\{\alpha : \frac{d\pi^*_{\alpha|\phi}}{d\alpha} > 0\right\}$. There exists $\bar{\alpha} < \infty$ such that $[\alpha_*(\phi), \alpha^*(\phi)] \subset [-\bar{\alpha}, \bar{\alpha}]$ holds for all $\phi \in \Phi$.*

(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 bounded support, which automatically holds if the identified set $IS_\alpha(\phi)$ is 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 a 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 (our notation leaves this implicit). 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. We notate $\max\{a, b\}$ by $a \vee b$.

Theorem 5.2 *Suppose Assumptions 3.2 (ii) and (iv), and 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 $h(\delta, \alpha) = \rho_{\tau}(\alpha - \delta)$,

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

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 tail condition of Assumption 5.1 (iii), the specific shape of $\pi^*_{\alpha|\phi}$ is irrelevant for the minimax decision. The intuition behind this result is that, as $\lambda \to \infty$, any prior with the same support as the benchmark prior is included in the prior class, and the worst-case conditional prior is the point-mass prior that assigns probability one to the furthest point from δ ($\alpha_*(\phi)$ or $\alpha^*(\phi)$).[4]

The large-λ asymptotics for the risk yield the following corollary concerning the set of posterior means.

Corollary 5.1 *Suppose Assumptions 5.1 (i) – (iii) hold. The set of posterior means of α, i.e., equation (18) with $f(\alpha) = \alpha$, converges to*

$$\left[E_{\phi|X}(\alpha_*(\phi)), E_{\phi|X}(\alpha^*(\phi)) \right],$$

[4]Such a point-mass prior is included in $\bar{\Pi}^{\infty}(\pi^*_{\alpha|\phi})$ (the weak*-closure of $\Pi^{\infty}(\pi^*_{\alpha|\phi})$) but not in $\Pi^{\infty}(\pi^*_{\alpha|\phi})$ itself, since the point-mass prior is not absolutely continuous with respect to $\pi^*_{\alpha|\phi}$ satisfying Assumption 5.1 (i).

as $\lambda \to \infty$. In particular, if $\alpha_*(\phi) = \inf_\alpha IS_\alpha(\phi)$ and $\alpha^*(\phi) = \sup_\alpha IS_\alpha(\phi)$ for π_ϕ-almost every ϕ, the set of posterior means coincides with that of Giacomini and Kitagawa (2018) in which the class of conditional priors $\pi_{\alpha|\phi}$ consists of any prior satisfying $\pi_{\alpha|\phi}(IS_\alpha(\phi)) = 1$ for all $\phi \in \Phi$.

This corollary implies that, in terms of the set of posterior means, the class of KL neighborhood priors with large λ can mimic Giacomini and Kitagawa's class of priors, which introduce full ambiguity within the identified set. Since the latter class of priors leads to posterior inference about the identified set, the robust Bayesian inference performed by our class of priors by varying $\lambda = 0$ to $\lambda \to \infty$ effectively bridges the gap between single-prior Bayesian inference and Bayesian analysis of the identified set as in Kline and Tamer (2016) and Chen et al. (2018).

5.2 Large-λ Asymptotics in Large Samples

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 large samples, the minimax decision under quadratic loss converges to the midpoint 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 midpoint 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 boundary points with weights τ and $1 - \tau$. Hence, 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, $\lambda \to 0$. This means that, if we use $[\delta^*(0.05), \delta^*(0.95)]$ as a posterior credibility interval for α, this interval estimate will be asymptotically strictly narrower than the frequentist confidence interval for α, because $[\delta^*(0.05), \delta^*(0.95)]$ 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.

One useful implication of Theorem 5.3 is that if the identified set $IS_\alpha(\phi)$ is a connected interval, the large-λ minimax estimators for the posterior mean and the τ-th posterior quantiles asymptotically agree with the Bayes estimator that assumes a uniform conditional prior for α. Viewing large-λ asymptotics as the lack of prior belief for the nonidentified parameter, this observation offers a novel justification for the use of the uniform prior for α given ϕ as a reference prior in the single-prior Bayesian approach.

Assuming additionally the continuity of $\alpha_*(\phi)$ and $\alpha^*(\phi)$ at ϕ_0, we obtain the large-sample version of Corollary 5.1.

Corollary 5.2 *Assume Assumptions 3.2 (i) and 5.1. Then, the large-λ set of posterior means of α converges to $[\alpha_*(\phi_0), \alpha^*(\phi_0)]$ as $n \to \infty$. Hence, if $IS_\alpha(\phi_0)$ is convex, and $\alpha_*(\phi_0) = \inf_\alpha IS_\alpha(\phi_0)$ and $\alpha^*(\phi_0) = \sup_\alpha IS_\alpha(\phi_0)$ hold, the large-λ set of posterior means of α converges to the true identified set.*

The latter statement of this corollary is analogous to the consistency of the set of posterior means shown in Giacomini and Kitagawa (2018), but in the case of the KL neighborhood class with large λ.

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 Theorems 5.2 and 5.3. We present a formal analysis of the discrete benchmark prior case in Appendix B.

6 Implementation

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. This section discusses how to choose these inputs carefully as well as computational methods to numerically approximate the worst-case posterior expected loss and the set of posteriors.

6.1 Constructing the Benchmark Prior

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

Let us denote the full vector of the structural parameters by $\tilde{\theta} = (\beta_s, \beta_d, d_1, d_2)$, and 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 a priori restrictions can be incorporated by trimming the support of $\pi^*_{\tilde{\theta}}$ such that $\pi^*_{\tilde{\theta}}(\{\beta_s \geq 0, \beta_d \leq 0\}) = 1$. Baumeister and Hamilton (2015) specify the prior for $\tilde{\theta}$ as the product of independent truncated Student's t distributions for (β_s, β_d) and independent inverse gamma distributions for $(d_1, d_2)|(\beta_s, \beta_d)$, with hyperparameters chosen through careful elicitation.

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

$$\omega_{11} = \frac{d_1 + d_2}{(\alpha - \beta)^2}, \quad \omega_{12} = \frac{\alpha d_1 + \beta d_2}{(\alpha - \beta)^2}, \quad \omega_{22} = \frac{\alpha^2 d_1 + \beta^2 d_2}{(\alpha - \beta)^2}, \tag{21}$$

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

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

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

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha|\phi) \propto \pi_{\tilde{\theta}}\left(\alpha, \beta\left(\alpha, \phi\right), d_1\left(\alpha, \phi\right), d_2\left(\alpha, \phi\right)\right) \times \left|\det\left(J\left(\alpha, \phi\right)\right)\right|, \tag{23}$$

where $J(\alpha, \phi)$ is the Jacobian of the mapping (22), and $|\det(\cdot)|$ is the absolute value of the determinant. This prior supports the entire identified set $IS_\alpha(\phi)$ if $\pi_{\tilde{\theta}}(\cdot)$ supports any value of (β_s, β_d) 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 (23) and by multiplying the reduced-form likelihood. When the analytical expression of the posterior of ϕ is not easy to derive or drawing ϕ directly from its posterior is challenging, an alternative is to obtain posterior draws of ϕ by transforming the posterior draws of $\tilde{\theta}$ according to $\Omega = A^{-1}D\left(A^{-1}\right)'$.

Since ϕ involves a nonlinear transformation of $\tilde{\theta}$, a diffuse prior for $\tilde{\theta}$ can imply an informative prior for ϕ. In finite samples, this can downplay the sample information for ϕ by distorting the shape of the likelihood. Given that our analysis can be motivated by concern about the robustness of posterior inference with respect to the choice of prior for $\tilde{\theta}$, one may not want to force the prior for ϕ to be informative as a result of this transformation. Such a concern might make the following hybrid approach attractive: the prior for $\tilde{\theta}$ is used solely to construct the benchmark conditional prior $\pi^*_{\alpha|\phi}$ via (23), while the prior for ϕ is separately specified from the prior for $\tilde{\theta}$ and is used to draw from the posterior of ϕ. Although the implied prior for $\tilde{\theta}$ resulting from this approach may be inconsistent with the prior initially specified for $\tilde{\theta}$, a benefit is that one can flexibly elicit a reasonable benchmark conditional prior with maintaining the noninformative prior for the reduced-form parameters.

6.2 Eliciting the Robustness Parameter λ

The radius of the KL neighborhood $\lambda \geq 0$ is an important prior input that directly controls the degree 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 about the benchmark prior. Since λ itself does not have an interpretable scale, it is necessary to map it into some prior quantity that the analyst can easily interpret and elicit.

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

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 the application, 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 set of prior means of $f(\tilde{\alpha})$ by applying expression (20). We then select λ that best matches with available but imprecise prior knowledge about $f(\tilde{\alpha})$. We illustrate this way of eliciting λ in the SVAR example of Section 7.

6.3 Computation

This section discusses how to compute the set of posterior means and posterior minimax decision. The algorithms below assume that the posterior draws of ϕ are given, and either α can be drawn from the benchmark conditional prior or its probability density can be evaluated up to a proportional constant.

We first present an algorithm that approximates the set of posterior means of $f(\alpha)$ shown in (18).

Algorithm 6.1 *Let M posterior draws of ϕ, $(\phi_1, \ldots, \phi_M)$, and benchmark conditional prior $\pi^*_{\alpha|\phi}$ be given. In Scenario 1 when we can draw α directly from the benchmark conditional prior $\pi^*_{\alpha|\phi}$, we follow Steps 1a and 1b. Otherwise, in Scenario 2 when the probability density of $\pi^*_{\alpha|\phi}$ can be evaluated up to a proportional constant, i.e., $\frac{d\pi^*_{\alpha|\phi}}{d\alpha} \propto cg(\alpha, \phi)$, $c > 0$, where $g(\cdot, \cdot)$ is known, we follow Steps 2a and 2b that implement importance sampling.*

Scenario 1

1a. For each $m = 1, \ldots, M$, obtain N independent draws of α, $\alpha_{mi} \sim \pi^*_{\alpha|\phi}$, $i = 1, \ldots, N$. We approximate the Lagrange multipliers $\kappa^\ell_\lambda(\phi_m)$ and $\kappa^u_\lambda(\phi_m)$ by solving the following optimizations:

$$\hat{\kappa}^\ell_\lambda(\phi_m) = \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{N} \exp\left\{ \frac{-f(\alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\},$$

$$\hat{\kappa}^u_\lambda(\phi_m) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{M} \exp\left\{ \frac{f(\alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\}.$$

1b. We approximate the posterior mean bounds of $f(\alpha)$ by

$$\left[\frac{1}{M} \sum_{m=1}^{M} \left(\frac{\sum_{i=1}^{N} f(\alpha_{mi}) \exp\left\{ \frac{-f(\alpha_{mi})}{\hat{\kappa}^\ell_\lambda(\phi_m)} \right\}}{\sum_{i=1}^{N} \exp\left\{ \frac{-f(\alpha_{mi})}{\hat{\kappa}^\ell_\lambda(\phi_m)} \right\}} \right), \frac{1}{M} \sum_{m=1}^{M} \left(\frac{\sum_{i=1}^{N} f(\alpha_{mi}) \exp\left\{ \frac{f(\alpha_{mi})}{\hat{\kappa}^u_\lambda(\phi_m)} \right\}}{\sum_{i=1}^{N} \exp\left\{ \frac{f(\alpha_{mi})}{\hat{\kappa}^u_\lambda(\phi_m)} \right\}} \right) \right].$$

Scenario 2

2a. *For each $m = 1, \ldots, M$, obtain N independent draws of α, $(\alpha_{m1}, \ldots, \alpha_{mN})$, from a proposal distribution $\tilde{\pi}_{\alpha|\phi}(\alpha|\phi_m)$ (e.g., the uniform distribution on $IS_\alpha(\phi_m)$). We approximate the Lagrange multipliers $\kappa_\lambda^\ell(\phi_m)$ and $\kappa_\lambda^u(\phi_m)$ by solving the following optimizations:*

$$\kappa_\lambda^\ell(\phi_m) = \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{N} w_{mi} \exp\left\{ \frac{-f(\alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\},$$
$$\kappa_\lambda^u(\phi_m) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{N} w_{mi} \exp\left\{ \frac{f(\alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\},$$

where

$$w_{mi} \equiv \frac{g(\alpha_{mi}, \phi_m)}{\frac{d\tilde{\pi}_{\alpha|\phi}(\alpha_{mi}, \phi_m)}{d\alpha}}, \quad i = 1, \ldots, N.$$

2b. *We approximate the posterior mean bounds of $f(\alpha)$ by*

$$\left[\frac{1}{M} \sum_{m=1}^{M} \left(\frac{\sum_{i=1}^{N} w_{mi} f(\alpha_{mi}) \exp\left\{ \frac{-f(\alpha_{mi})}{\hat{\kappa}_\lambda^\ell(\phi_m)} \right\}}{\sum_{i=1}^{N} w_{mi} \exp\left\{ \frac{-f(\alpha_{mi})}{\hat{\kappa}_\lambda^\ell(\phi_m)} \right\}} \right), \frac{1}{M} \sum_{m=1}^{M} \left(\frac{\sum_{i=1}^{N} w_{mi} f(\alpha_{mi}) \exp\left\{ \frac{f(\alpha_{mi})}{\hat{\kappa}_\lambda^u(\phi_m)} \right\}}{\sum_{i=1}^{N} w_{mi} \exp\left\{ \frac{f(\alpha_{mi})}{\hat{\kappa}_\lambda^u(\phi_m)} \right\}} \right) \right].$$

For computation of the posterior gamma-minimax estimator, Algorithm 6.1 is used to approximate the worst-case risk of decision δ, i.e., the objective function in (16).

Algorithm 6.2 *Let M posterior draws of ϕ, $(\phi_1, \ldots, \phi_M)$, and benchmark conditional prior $\pi^*_{\alpha|\phi}$ be given. Let $h(\delta, \alpha)$ be the loss function, and Scenarios 1 and 2 be as defined in Algorithm 6.1. We approximate the posterior gamma minimax estimator by a minimizer of $\frac{1}{M} \sum_{m=1}^{M} \hat{r}_\lambda(\delta, \phi_m)$ in δ, where $\hat{r}_\lambda(\delta, \phi_m)$ is computed from Step 1c or 2c below:*

Scenario 1

1c. *Following Step 1a of Algorithm 6.1, let N draws of α from $\pi^*_{\alpha|\phi}$, $(\alpha_1, \ldots, \alpha_N)$, be given. We approximate the Lagrange multipliers $\kappa_\lambda(\delta, \phi_m)$ by*

$$\hat{\kappa}_\lambda(\delta, \phi_m) = \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{N} \exp\left\{ \frac{h(\delta, \alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\}.$$

Then, compute

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

Scenario 2

2c. *Following Step 2a of Algorithm 6.1, let N draws of α, $(\alpha_{m1}, \ldots, \alpha_{mN})$, be drawn from a proposal distribution $\tilde{\pi}_{\alpha|\phi}(\alpha|\phi)$ (e.g., the uniform distribution on $IS_\alpha(\phi_m)$) and let $(w_{m1}, \ldots, w_{mN})$ be the importance weights as defined in Algorithm 6.1. We approximate the Lagrange multipliers $\kappa_\lambda(\delta, \phi_m)$ by*

$$\hat{\kappa}_\lambda(\delta, \phi_m) = \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(N^{-1} \sum_{i=1}^{N} w_{mi} \exp \left\{ \frac{h(\delta, \alpha_{mi})}{\kappa} \right\} \right) + \kappa\lambda \right\}.$$

Then, compute

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

The optimization for the Lagrange multiplier in these algorithms is a convex program so that a gradient-based algorithm can quickly solve the optimizations. For optimizing $\frac{1}{M}\sum_{m=1}^{M} \hat{r}_\lambda(\delta, \phi_m)$ in δ, we can use the gradient-based optimization algorithm by keeping the draws of α same in evaluating $\hat{r}_\lambda(\delta, \phi_m)$ if the loss is differentiable in δ. Even when the loss is non-differentiable, since δ is a scalar valued, it is also not computationally costly to perform brute-force grid search.

In the limiting case $\lambda \to \infty$, with the quadratic loss (or a check loss), Theorem 5.2 implies that Algorithm 6.2 can be skipped and we can directly approximate the worst-case risk 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 $[\underline{\alpha}(\phi_m), \bar{\alpha}(\phi_m)]$ are 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 replacing the approximated worst-case risk $\frac{1}{M}\sum_{m=1}^{M} \hat{r}_\lambda(\delta, \phi_m)$ with $\hat{r}_\lambda\left(\delta, \hat{\phi}_{ML}\right)$, where $\hat{\phi}_{ML}$ is the MLE for ϕ. This large-sample approximation further simplifies the computation.

7 Empirical Example

This section applies our procedure to the dynamic labor supply and demand model analysed in Baumeister and Hamilton (2015). Our motivation of this empirical application are three-folds. First, we want to explore the robustness of the results in Baumeister and Hamilton

(2015) to their choice of prior for the structural parameters by computing the set of posterior quantities. Second, we want to construct robust point estimators for the elasticity parameters with different KL-neighborhoods. Third, we want to illustrate how to specify a benchmark prior and the radius parameter λ for the KL-neighborhood.

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, $x_t = (\Delta w_t, \Delta n_t)$. The observations are quarterly and run from t=1970:Q1 to 2014:Q4. Following the convention of time-series analysis, here we denote the sample size by T instead of n.

7.1 Specification and Parameterization

The model is a bivariate SVAR with $L = 8$ lags and a constant,

$$A_0 x_t = c + \sum_{l=1}^{L} A_l x_{t-l} + u_t, \qquad u_t \sim_{iid} \mathcal{N}(0, D), \ t = 1, \ldots, T,$$

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

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

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

We set the supply elasticity as the parameter of interest, $\alpha = \beta_s$. The mapping between $\tilde{\theta}$ and (α, ϕ) is given by (22) and

$$A = A_0(\alpha, \phi) B \equiv A(\alpha, \phi), \tag{24}$$

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 for α given ϕ is

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha|\phi) \propto \pi_{\tilde{\theta}}(\alpha, \beta_d(\alpha, \phi), d_1(\alpha, \phi), d_2(\alpha, \phi), A(\alpha, \phi)) \times |\det(J(\alpha, \phi))|, \tag{25}$$

where $\pi_{\tilde{\theta}}(\beta_s, \beta_d, d_1, d_2, A)$ is a prior distribution for $\tilde{\theta}$ that induces the benchmark conditional prior for $\alpha|\phi$ and the single prior for ϕ, and $J(\alpha, \phi)$ is the Jacobian of the transformations (22) and (24).

7.2 Benchmark Prior

We construct the benchmark conditional prior $\pi^*_{\alpha|\phi}$ by setting the prior for $\tilde{\theta}$ to the one used in Baumeister and Hamilton (2015) and applying formula (25). Decomposing a prior for $\tilde{\theta}$ as

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

Baumeister and Hamilton (2015) recommend to elicit each of the components 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. For completeness, Appendix C presents the specific choice of hyperparameters.

1. Elicitation of $\pi_{(\beta_s,\beta_d)}$: Independent truncated t-distributions are used as priors for β_s and β_d, where the truncations incorporate the dogmatic sign restrictions that with prior probability one, $\beta_s \geq 0$ and $\beta_d \leq 0$. Their hyperparameters are chosen based on meta-analysis of microeconomic and macroeconomic studies that estimate the labor supply and demand elasticities. Specifically, Baumeister and Hamilton (2015) identify that most of these estimates fall in the interval $\beta_s \in [0.1, 2.2]$ and $\beta_d \in [-2.2, -0.1]$, and they accordingly choose the hyperparameters of the t-distribution so that $\pi_{\beta_s}([0.1, 2.2]) = 0.9$ and $\pi_{\beta_d}([-2.2, -0.1]) = 0.9$.

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

3. Elicitation of $\pi_{A|(d_1,d_2,\beta_s,\beta_d)}$: Since the reduced-form VAR coefficients satisfy $B = A_0^{-1}A$, elicitation of the conditional prior for A given $(\beta_s, \beta_d, d_1, d_2)$ 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 the proposals in these works, Baumeister and Hamilton (2015) specify a prior for B as a multivariate normal with prior means corresponding to $(\Delta w_t, \Delta n_t)$ being independent random walk processes. The prior is diffuse for short-lag coefficients and becomes tighter at longer lags.

In this elicitation process, available prior evidence suggests only vague restrictions on the prior and certainly is not precise enough to pin down the exact shape of a prior distribution.

For instance, the elicitation of a prior for β_s and β_d relies on a belief that some large amount of prior probability should be assigned to the ranges identified by the meta-analysis, but it cannot pin down the shape of the prior to a t-distribution. In the step of eliciting the conditional priors for (d_1, d_2) and A, the available prior knowledge can reasonably be translated into the location and scale of the prior distributions, but the shape of the prior is chosen for computational convenience. In addition, the independence of the priors invoked in each step 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 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 concern about the robustness of posterior. The situation is worse in set-identified models, since robustness concerns are magnified due to the lack of identification.

7.3 Results

Figure 1 summarizes the set of posterior means for each choice of $\lambda \in \{0, 0.1, 0.5, 1, 2, 8\}$. To compute them, we approximate the integration with respect to $\pi_{\phi|X}$ by plugging in the maximum likelihood estimator of ϕ, as justified by the large-sample result of Theorem 3.3 (ii). That is, these plots summarize the set of posteriors spanned by $\Pi^{\lambda}(\pi^*_{\alpha|\hat{\phi}_{ML}})$ for each λ =0.1, 0.5, 1, 2, 8. The top-left panel corresponds to the single prior/posterior case ($\lambda = 0$), which corresponds to the benchmark conditional prior $\pi^*_{\alpha|\phi}$ given $\phi = \hat{\phi}_{MLE}$. The benchmark prior (truncated t-distribution) results in the asymmetric posterior (black solid density) that has the heavy right-tail and concentrates near the origin. The identified set estimate $IS_{\alpha}(\hat{\phi}_{ML})$, on the other hand, spans 0 to 5.2 (the red horizontal segment in the plots). The concentration of the posterior toward the lower bound of the identified set is driven by the specification of the prior, since the likelihood function is flat with respect to α over the identified set. This motivates exploring the robustness of posterior inference to the choice of conditional prior for α. To this end, we conduct a global sensitivity analysis by varying λ and computing the set of posterior means.

The other five panels in Figure 1 show the set of posterior means produced by our procedures. As we increase λ, the set of posterior means indeed expands, and its upper bound reaches the middle point of the identified set at $\lambda = 2$. A larger choice of λ such as $\lambda = 8$ leads to a set of posterior means that is nearly identical to the estimate of the identified set, which is consistent with the large-λ asymptotic result of Corollary 5.1.

The black square in each of the panels in Figure 1 shows the conditional gamma-minimax estimator for α under the quadratic loss. The gamma-minimax estimator of α is larger than the posterior mean under the benchmark conditional prior. This is because the set of posteriors include those that put substantially more probability masses toward the upper bound of the identified set. In the large-λ case of $\lambda = 8$, the gamma-minimax estimator coincides with the midpoint of the identified set, as predicted by Theorem 5.3.

In Figure 2, we perform robust Bayes analyses for the posterior median instead of the mean, where the choices of λ are the same as in Figure 1. The gamma-minimax estimators reported in Figure 2 are obtained under the absolute loss (i.e., check loss with $\tau = 0.5$). Due to the asymmetry of the benchmark posterior, the posterior median under the benchmark lies closer to the origin than the posterior mean. The sets of posterior medians, however, are overall similar to the sets of posterior means of Figure 1. In case of $\lambda = 8$, we do not see any notable differences between Figures 1 and 2.

Figure 3 plots the posteriors that attain the lower and upper bounds of the set of posterior means for λ =0.1, 0.5, 1 and 2 shown in Figure 1. The posterior attaining the lower bound (blue dashed density) concentrates toward zero more, and more sharply as λ increases, while the posterior attaining the upper bound (red dashed density) becomes more spread out and resembles the uniform distribution on the identified set at $\lambda = 2$.

7.4 Choice of λ

To apply the procedure for eliciting λ presented in Section 6.2, we focus on assessing how much the marginal prior for α can vary with λ. Figure 4 presents the set of prior means for α for each choice of $\lambda \in \{0.1, 0.5, 1, 2\}$. It also plots the marginal priors of α attaining the lower and upper bounds of the prior mean. Our recommended elicitation procedure is to find a value of λ under which the sets of the prior means or other features of the distribution match with the decision-maker's vague prior knowledge. Furthermore, as done in Figure 4, plotting the priors attaining the lower and upper bounds of the set helps the elicitation process, since under a choice of λ that can well represent the ambiguous beliefs of the decision-maker, the decision-maker should be indifferent between the extreme priors attaining the bounds and the benchmark prior. For instance, if the decision-maker believes that the prior mean of α is between $[0.5, 1.5]$, and the extreme priors plotted by the blue and red densities in the top right panel of Figure 4 appear as plausible as the benchmark prior, his reasonable choice of λ would be 0.5.

To further summarize the set of prior distributions spanned by λ, Figure 5 plots the pointwise CDF bounds of the prior for α for each choice of $\lambda \in \{0.1, 0.5, 2\}$. Inverting the CDF bounds at level τ gives the set of the τ-th quantile. For instance, $\lambda = 0.5$ spans the prior median of α from 0.34 to 1.28 with the median of the benchmark prior being 0.80. The set of prior quantiles can also be useful for eliciting λ, though it should be noted that the CDF bounds of Figure 5 are valid only pointwise and no single prior in the class can attain the upper or lower bound uniformly in α.[5]

Based on the meta-analysis revealing that the supply elasticity estimates previously reported in the literature vary between 0.1 and 2.2, Baumeister and Hamilton (2015) specify the hyperparameters in the prior for α so as to assign prior probability 90% on $\{\alpha \in [0.1, 2.2]\}$. To perform our robust Bayesian analysis, we may want to focus on the range of prior probabilities on $\{\alpha \in [0.1, 2.2]\}$ in order to elicit λ. Specifically, we specify a lower bound of $\pi_\alpha([0.1, 2.2])$ and tune λ to match with this lower bound. For this purpose, Figure 6 plots the lower prior probability on $\{\alpha \in [0.1, 2.2]\}$,

$$p \equiv \int_\Phi \left[\inf_{\pi_{\alpha|\phi} \in \Pi^\lambda(\pi^*_{\alpha|\phi})} \pi_{\alpha|\phi}([0.1, 2.2]) \right] d\pi_\phi, \tag{26}$$

over $\lambda \in [0, 8]$. For instance, at each $\lambda = 0.1$, 0.5, 1, and 2, the corresponding prior lower probability is computed as $p = 0.81$, 0.66, 0.53, and 0.37, respectively.

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. (2017). 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 unconditional gamma-minimax decision, less is

[5] This claim follows from the fact that priors maximizing or minimizing $\pi_\alpha(\{\alpha \le t\})$ are obtained by setting $f(\alpha) = 1\{\alpha \le t\}$ in the construction of $\pi^u_{\alpha|\phi}$ and $\pi^\ell_{\alpha|\phi}$ in (19) and plugging them into (20). Hence, varying t alters the priors attaining the bounds of $\pi_\alpha(\{\alpha \le t\})$.

known about statistical admissibility of the conditional one. As DasGupta and Studden (1989) argue, the conditional gamma-minimax can often lead to a reasonable estimator with good frequentist performance. Further decision-theoretic justifications for the conditional gamma-minimax decision, including its statistical admissibility in set-identified models, remain open questions.

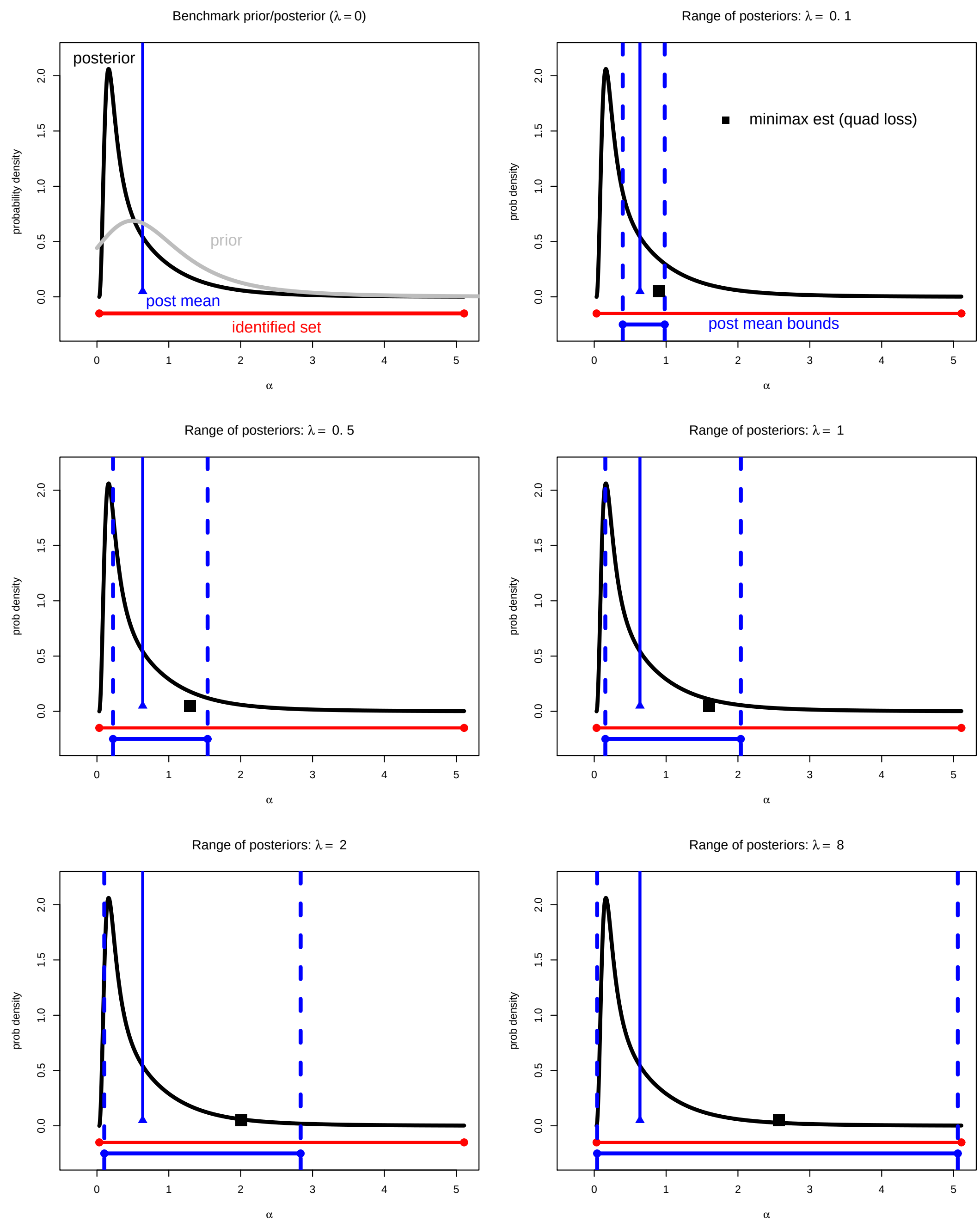

Figure 1: **Range of posterior means for supply elasticity:**
The top-left panel shows the prior and posterior for α obtained in Baumeister and Hamilton (2015). We treat it as the benchmark prior and posterior in the rest of the panels. The red horizontal segment shows $IS_\alpha(\hat{\phi}_{ML})$. The blue vertical solid line is the posterior mean at the benchmark. The blue vertical dashed lines show the set of posterior means. The black square indicates the value of the conditional gamma-minimax estimator under quadratic loss.

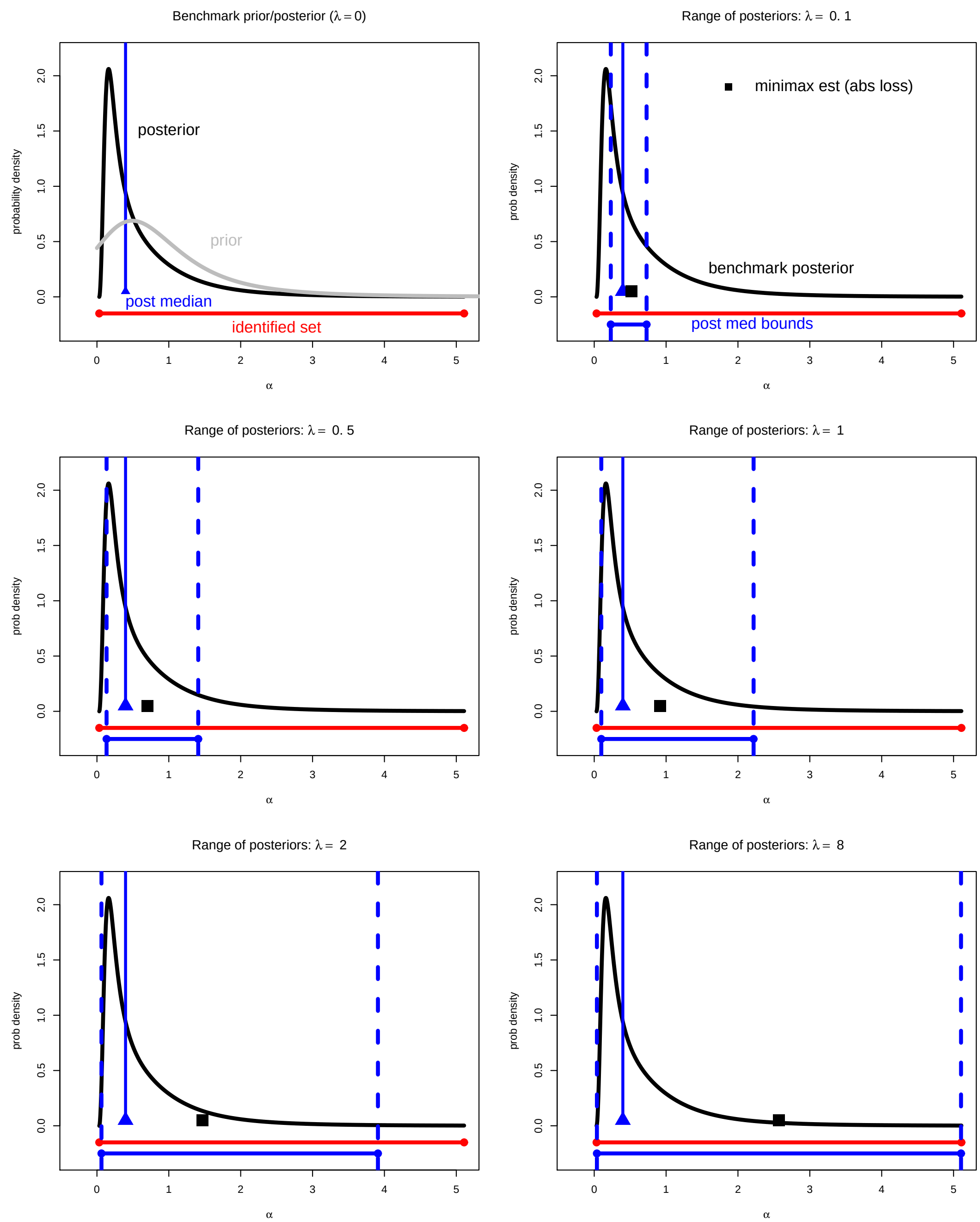

Figure 2: **Range of posterior medians for supply elasticity:**
The top-left panel shows the prior and posterior for α obtained in Baumeister and Hamilton (2015). The blue vertical solid line is the posterior median at the benchmark. The blue vertical dashed lines show the set of posterior medians. The black square indicates the value of the conditional gamma-minimax estimator under absolute loss.

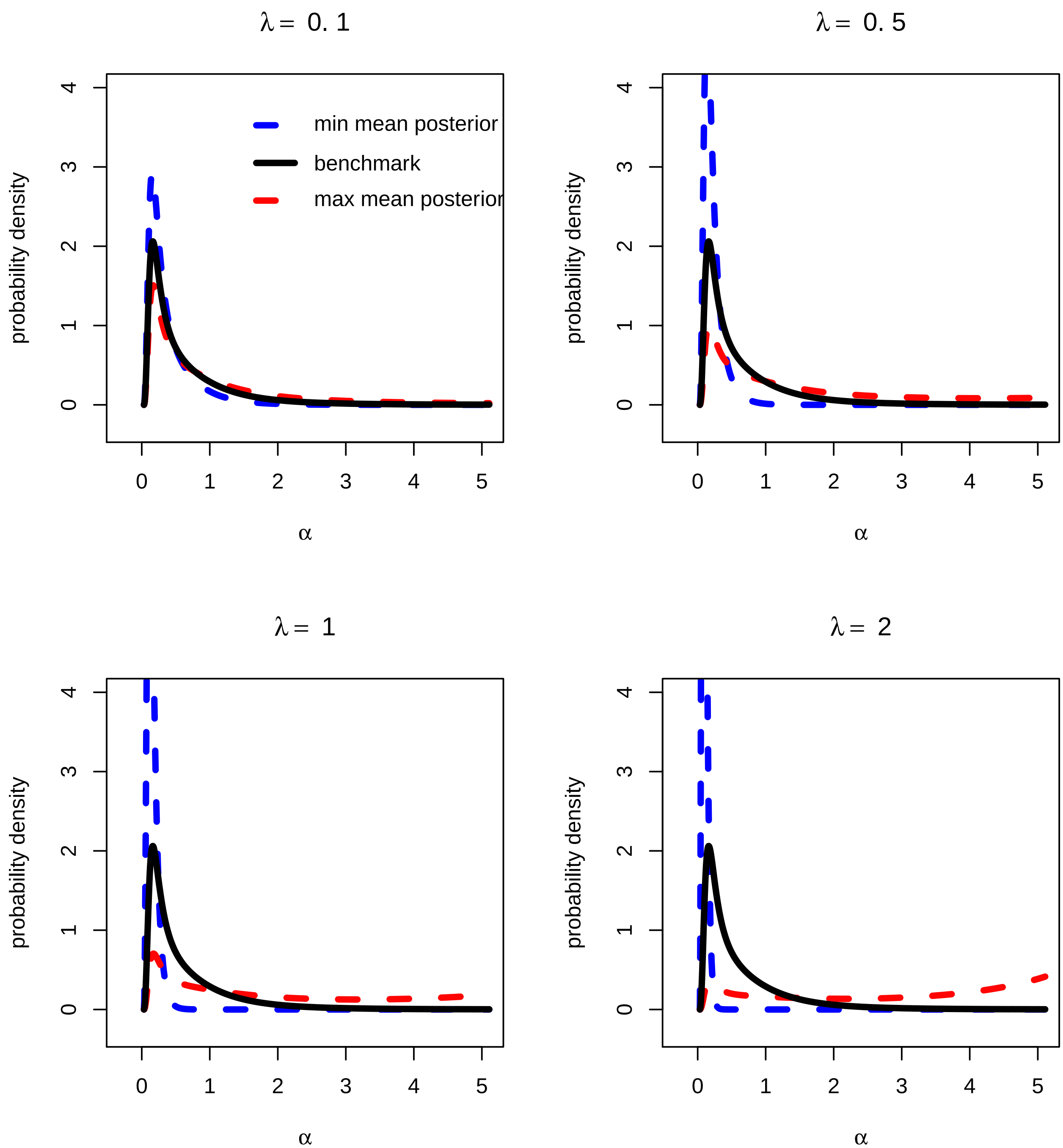

Figure 3: **Posteriors attaining the posterior mean bounds:**
The density drawn by the black solid line is the benchmark posterior. The blue and red dashed densities are the posteriors of α that attain the lower and upper bounds of the posterior mean shown in Figure 1, respectively.

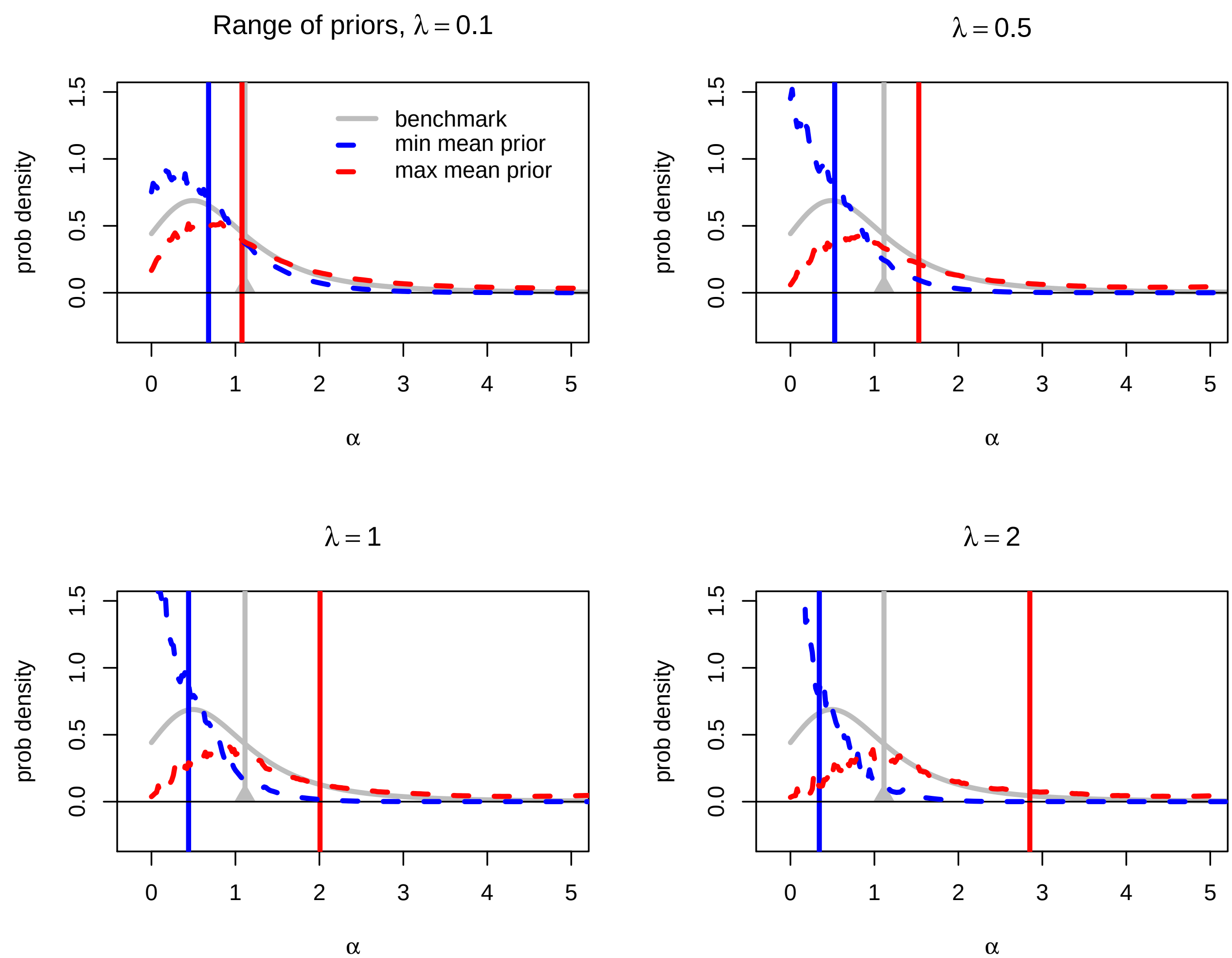

Figure 4: **Prior mean bounds and priors attaining the bounds:**
The benchmark prior is the truncated t-distribution used in Baumeister and Hamilton (2015). The vertical grey line with the grey triangle plots the mean of the benchmark prior. The blue and red vertical lines show the lower and upper bounds of the prior mean of α, and the blue and red densities are prior distributions attaining the prior mean bounds.

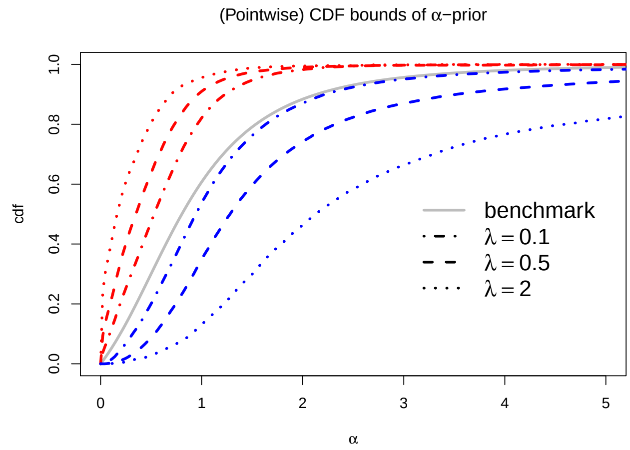

Figure 5: **Pointwise bounds of the α-prior cdf:**
The benchmark prior is the truncated t-distribution used in Baumeister and Hamilton (2015). The red and blue curves plot the upper and lower bounds of $\pi_\alpha(\alpha \le t)$, $t \in [0, 5]$ over $\pi_{\alpha|\phi} \in \Pi^\lambda(\pi^*_{\alpha|\phi})$, respectively, for $\lambda \in \{0.1, 0.5, 2\}$.

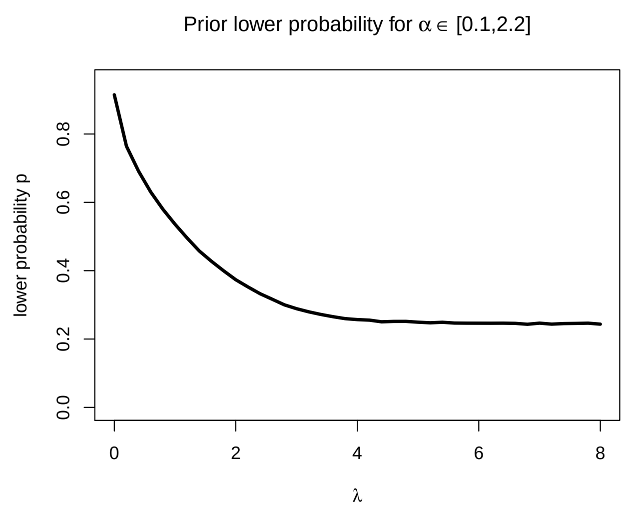

Figure 6: **Prior lower probabilities on $\alpha \in [0.1, 2.2]$:**
The benchmark prior is the truncated t-distribution used in Baumeister and Hamilton (2015). Plotting the relationship of λ and p defined in equation (26).

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*

$$\mathcal{R}(\pi_{\alpha|\phi}\|\pi^*_{\alpha|\phi}) = -\int_{IS_\alpha(\phi)} \mathcal{R}(\pi_{\theta|\alpha\phi}\|\pi^*_{\theta|\alpha\phi})d\pi_{\alpha|\phi} + \mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi}), \tag{27}$$

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 $\frac{d\pi_{\theta|\phi}}{d\theta}$ and $\frac{d\pi_{\alpha|\phi}}{d\alpha}$. Note they satisfy $\frac{d\pi_{\alpha|\phi}(\alpha)}{d\alpha} = \int_{\Theta(\alpha,\phi)} \frac{d\pi_{\theta|\phi}(\theta)}{d\theta} d\theta$.

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

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

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

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

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

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 benchmark conditional prior $\pi^*_{\theta|\phi}$. Since thus-constructed $\pi^0_{\theta|\phi}$ shares the conditional distribution of θ given (α, ϕ) with $\pi^*_{\theta|\phi}$, Lemma A.1 implies $\mathcal{R}(\pi^0_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \mathcal{R}(\pi^0_{\alpha|\phi} \| \pi^*_{\alpha|\phi}) \leq \lambda$. Hence, $\pi^0_{\theta|\phi} \in \Pi^{\lambda}\left(\pi^*_{\theta|\phi}\right)$ holds, and this implies

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

■

A.2 Proof of Theorem 3.1

Proof of Theorem 3.1. Let ϕ and $\delta = \delta(x)$ be fixed. Let $\kappa_{\lambda}\left(\delta, \phi\right)$ be as defined in Lemma 2.2. Since $\kappa_{\lambda}\left(\delta, \phi\right)$ does not depend on $\pi_{\alpha|\phi}$, we treat $\kappa^* \equiv \kappa_{\lambda}\left(\delta, \phi\right)$ as a constant, and we focus on solving the inner maximization problem in the multiplier minimax problem (13).

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 receive 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

$$g_i \equiv \pi_{\alpha|\phi}\left(\alpha_i\right), \quad f_i \equiv \pi^*_{\alpha|\phi}\left(\alpha_i\right), \;\; h_i \equiv h(\delta, \alpha_i), \quad \text{for } i = 1, \ldots, m. \tag{28}$$

Then, the inner maximization problem of (13) can be written as

$$\max_{g_1,\dots,g_m} \sum_{i=1}^m h_i g_i - \kappa^* \sum_{i=1}^m g_i \ln\left(\frac{g_i}{f_i}\right), \tag{29}$$
$$\text{s.t.} \qquad \sum_{i=1}^m g_i = 1.$$

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

$$h_i + \kappa^* \ln f_i - \kappa^* - \kappa^* \ln g_i - \xi = 0, \quad i = 1, \dots, m. \tag{30}$$

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

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

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

$$g_i^* = \frac{f_i \exp(h_i/\kappa^*)}{\sum_{j=1}^m f_j \exp(h_j/\kappa^*)}. \tag{31}$$

Plugging this back into the objective function, we obtain

$$\kappa^* \sum_{i=1}^m \left[\frac{f_i \exp(h_i/\kappa^*)}{\sum_{j=1}^m f_i \exp(h_i/\kappa^*)} \ln\left(\sum_{j=1}^m f_j \exp(h_j/\kappa^*) \right) \right] \tag{32}$$
$$= \kappa^* \ln\left(\sum_{j=1}^m f_j \exp(h_j/\kappa^*) \right),$$

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

We generalize the claim to arbitrary $\pi^*_{\alpha|\phi}$. Based on the optimal g_i obtained in (31), 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

$$d\pi^0_{\alpha|\phi} = \frac{\exp(h(\delta, \alpha)/\kappa^*)}{\int_{IS_\alpha(\phi)} \exp(h(\delta, \alpha)/\kappa^*) d\pi^*_{\alpha|\phi}} \cdot d\pi^*_{\alpha|\phi}, \qquad \alpha\text{ -a.e.} \tag{33}$$

with $\kappa^* > 0$. Since $\exp(h(\delta, \alpha)/\kappa^*)$ is integrable with respect to $\pi^*_{\alpha|\phi}$, $\exp(h(\delta, \alpha)/\kappa^*) \in (0, \infty)$ holds, $\pi^*_{\alpha|\phi}$-a.s. Equation (33) then implies that $\pi^*_{\alpha|\phi}$ is absolutely continuous with respect to $\pi^0_{\alpha|\phi}$, and 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

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

Plugging in (33) 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 (33) solves uniquely (up to α-a.e.) the inner maximization problem. Hence, analogous to the discrete benchmark prior case, it holds

$$\max_{\pi_{\alpha|\phi}\in\Pi^{\infty}\left(\pi^{*}_{\alpha|\phi}\right)}\left\{\int_{IS_{\alpha}(\phi)}h(\delta,\alpha)d\pi_{\alpha|\phi}-\kappa^{*}\mathcal{R}(\pi_{\alpha|\phi}\|\pi^{0}_{\alpha|\phi})\right\}=\kappa^{*}\ln\left(\int_{IS_{\alpha}(\phi)}\exp\left(\frac{h(\delta,\alpha)}{\kappa^{*}}\right)d\pi^{*}_{\alpha|\phi}\right). \tag{34}$$

By Lemma 2.2, $\pi^{0}_{\alpha|\phi}(\alpha)$ derived in (33) solves the inner maximization problem of (11). Hence, the value function is given by

$$\begin{aligned}\max_{\pi_{\alpha|\phi}\in\Pi^{\lambda}\left(\pi^{*}_{\alpha|\phi}\right)}\left\{\int_{IS_{\alpha}(\phi)}h(\delta,\alpha)d\pi_{\alpha|\phi}\right\} &= \int_{IS_{\alpha}(\phi)}h(\delta,\alpha)d\pi^{0}_{\alpha|\phi}\\ &= \int_{IS_{\alpha}(\phi)}\frac{h(\delta,\alpha)\exp(h(\delta,\alpha)/\kappa^{*})}{\int_{IS_{\alpha}(\phi)}\exp(h(\delta,\alpha)/\kappa^{*})d\pi^{*}_{\alpha|\phi}}d\pi^{*}_{\alpha|\phi}. \end{aligned}\tag{35}$$

Also, by the Kuhn-Tucker slackness condition stated in Lemma 2.2, $\kappa^{*}>0$ implies $\lambda=\mathcal{R}(\pi^{0}_{\alpha|\phi}\|\pi^{*}_{\alpha|\phi})$. It then translates to the following condition for κ^{*}:

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

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 the second derivative of $f_{\lambda}(\kappa)$ in κ equals to the variance of $h\left(\delta,\alpha\right)$ with $\alpha\sim\pi^{0}_{\alpha|\phi}$, which is strictly positive by the imposed nondegeneracy assumption of $h\left(\delta,\alpha\right)$. Hence, $f_{\lambda}(\kappa)$ is strictly convex. Therefore, κ^{*} solving the first-order condition is unique and strictly positive.

The conclusion follows by integrating (35) with respect to $\pi_{\phi|X}$. ■

A.3 Proof of Theorem 3.3

The following lemmas A.3 – A.6 are used to prove Theorem 3.3. Since the support of $\pi^*_{\alpha|\phi}$ is contained in $IS_\alpha(\phi)$ and any $\pi_{\alpha|\phi} \in \Pi^\lambda(\pi^*_{\alpha|\phi})$ is absolutely continuous with respect to $\pi^*_{\alpha|\phi}$, for any measurable function $f(\alpha)$, it holds $\int_{IS_\alpha(\phi)} f(\alpha)d\pi_{\alpha|\phi} = \int_{\mathbb{R}} f(\alpha)d\pi_{\alpha|\phi}$ for any $\pi_{\alpha|\phi} \in \Pi^\lambda(\pi^*_{\alpha|\phi})$. In the proofs to follow, we do not explicitly specify the limit of integration when the integration of $f(\alpha)$ is over $IS_\alpha(\phi)$.

Lemma A.3 *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.3. Let $h = h(\delta,\alpha)$. By Markov's inequality and Assumption 3.2 (iv),

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

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 $a > 0$,

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

In order for Assumption 3.2 (iv) to hold, we require

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

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,

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

would hold for all large ν. ■

Lemma A.4 *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*

$$\kappa_\lambda(\delta, \phi) \leq \frac{2H}{\lambda},$$

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.4. We first show the upper bound. Recall $\kappa_\lambda(\delta, \phi)$ solves (see equation (36))

$$\lambda = -\ln\left(\int \exp\left(\frac{h(\delta, \alpha)}{\kappa}\right) d\pi^*_{\alpha|\phi}\right) + E^0_{\alpha|\phi}\left[\frac{h(\delta, \alpha)}{\kappa}\right], \tag{40}$$

where $E^0_{\alpha|\phi}(\cdot)$ is the expectation with respect to the exponentially tilted conditional prior $\pi^0_{\alpha|\phi}$. Boundedness of $h(\delta, \alpha)$ implies that the first term in equation (40) is bounded from above by H/κ and the second term can be also bounded from above by H/κ. Hence, we have

$$\lambda \leq \frac{2H}{\kappa_\lambda(\delta, \phi)}.$$

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

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

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

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

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

$$\begin{aligned}
&0 \geq E^*_{\alpha|\phi}(W) \geq -\frac{1}{\kappa^*} c_1, \qquad (43)\\
&c_1 \equiv H - \inf_{\delta \in \mathcal{D}, \phi \in G_0} E^*_{\alpha|\phi}(h(\delta, \alpha)) \geq 0
\end{aligned}$$

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.3 (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.5 *Under Assumption 3.2 (ii) and (iv), we have*

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

where $Var^0_{\alpha|\phi}(\cdot)$ is the variance with respect to the worst-case (exponentially tilted) conditional prior $\pi^0_{\alpha|\phi}$ shown in Theorem 3.1.

Proof of Lemma A.5. Let $c > 0$ be the constant defined in Assumption 3.2 (iv), and $h = h(\delta, \alpha)$. By Markov's inequality,

$$\begin{aligned} Var^0_{\alpha|\phi}(h) &\geq cE^0_{\alpha|\phi}\left(1\left\{(h - E^0_{\alpha|\phi}(h))^2 \geq c\right\}\right) \\ &\geq c \cdot \exp\left(-\frac{H}{C_1(\lambda)}\right) E^*_{\alpha|\phi}\left(1\left\{(h - E^0_{\alpha|\phi}(h))^2 \geq c\right\}\right) \\ &\geq c\epsilon \cdot \exp\left(-\frac{H}{C_1(\lambda)}\right), \end{aligned}$$

where the second inequality follows by the lower bound of $\kappa_\lambda(\delta, \phi)$ shown in Lemma A.4 and $E^0_{\alpha|\phi}(f(\alpha)) \geq \exp\left(-\frac{H}{C_1(\lambda)}\right) E^*_{\alpha|\phi}(f(\alpha))$ for any nonnegative random variables $f(\alpha)$. ■

Lemma A.6 *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.6. By the mean value theorem, 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 implicit function theorem to $\kappa_\lambda(\delta, \phi)$ defined as the solution to

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

Since $|\partial g / \partial \kappa| = Var^0_{\alpha|\phi}(h(\delta, \alpha) / \kappa_\lambda(\delta, \phi))$, we obtain

$$\sup_{\delta \in \mathcal{D}, \phi \in G_0} \left\| \frac{\partial \kappa_\lambda(\delta, \phi)}{\partial \phi} \right\| \leq \left(\frac{H}{\lambda} \right)^2 \frac{\sup_{\delta \in \mathcal{D}, \phi \in G_0} \|\partial g / \partial \phi\|}{\inf_{\delta \in \mathcal{D}, \phi \in G_0} Var^0_{\alpha|\phi}(h(\delta, \alpha))}, \tag{46}$$

where the differentiability of g with respect to ϕ requires Assumption 3.2 (vi). By Lemma A.5, the variance lower bound in the denominator is bounded away from zero. For the numerator, we have

$$\begin{aligned}
\left\| \frac{\partial g}{\partial \phi} \right\| &\leq \frac{\left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}(\exp(h/\kappa)) \right\|}{E^*_{\alpha|\phi}(\exp(h/\kappa))} + \frac{\left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}[(h/\kappa) \cdot \exp(h/\kappa)] \right\|}{E^*_{\alpha|\phi}(\exp(h/\kappa))} + \frac{\left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}(h/\kappa) \right\| \cdot E^*_{\alpha|\phi}[h/\kappa \cdot \exp(h/\kappa)]}{[E^*_{\alpha|\phi}(\exp(h/\kappa))]^2} \\
&\leq \left\{ \exp \left(\frac{H}{C_1(\lambda)} \right) + \frac{H}{C_1(\lambda)} \exp \left(\frac{3H}{C_1(\lambda)} \right) \right\} \cdot \sup_{\delta \in \mathcal{D}, \phi \in G_0} \left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi}(h) \right\| \\
&\quad + \exp \left(\frac{H}{C_1(\lambda)} \right) \cdot \sup_{\delta \in \mathcal{D}, \phi \in G_0, \kappa \in [C_1(\lambda), H/\lambda]} \left\| \frac{\partial}{\partial \phi} E^*_{\alpha|\phi} \left[\left(\frac{h}{\kappa} \right) \cdot \exp \left(\frac{h}{\kappa} \right) \right] \right\| \\
&\equiv C_2(\lambda) < \infty,
\end{aligned}$$

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

Proof of Theorem 3.3. Since the posterior minimax decision is invariant to an additive constant to the loss function, we assume without loss of generality that the loss function is nonnegative.

(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 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).

For this goal, let

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

where $s_{\lambda}(\delta, \phi) \geq 1$ follows by the normalization of the loss function to being nonnegative. Since

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

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 doing so results in confusion.

By Lemma 2.2 and equation (34) 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

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

By noting $\ln(x) \leq x - 1$, Lemma A.4, and $s_{\lambda}(\phi) \geq 1$, we have

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

Combining equations (47) and (48), and applying Lemma A.6, we obtain for $\phi \in G_0$,

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

where $C_3(\lambda) = \left[\lambda + \frac{H}{C_1(\lambda)} + \frac{H^2}{\lambda C_1(\lambda)} \exp\left(\frac{H}{C_1(\lambda)}\right)\right] C_2(\lambda)$. Thus,

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

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 $\delta \in \mathcal{D}$, 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.

Since Assumption 5.1 (i) and (ii) imply Assumption 3.2 (ii) and (iv), we can apply Lemma A.4. It implies that as $\lambda \to \infty$, $\kappa_\lambda(\delta, \phi) \to 0$ at every (δ, ϕ). Hence, to assess the pointwise convergence behavior of $r_\lambda(\delta, \phi)$ as $\lambda \to \infty$ at each (δ, ϕ), it suffices to analyze 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

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

We proceed by showing 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 at an exponential rate 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 at an exponential rate 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 $\sum_{k=1}^{\infty} \tilde{a}_k z^k = \left(-c(\phi)z + z^2\right)\left(\sum_{k=0}^{\infty} a_k z^k\right)$. 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,

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

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

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

Since $r_\kappa\left(\delta, \phi\right)$ 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

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

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

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

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

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

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

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

Proof of Theorem 5.3. (i) Let $r_\kappa(\delta, \phi)$ be defined as in the proof of Theorem 5.2. Since $\lambda \to \infty$ asymtptotics imply $\kappa \to 0$ asymptotics, we consider working with $R_n\left(\delta\right) \equiv$

$\lim_{\kappa\to 0}\int_\Phi r_\kappa(\delta,\phi)\,d\pi_{\phi|X}$, 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

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

where $\mathrm{diam}(\mathcal{D})<\infty$ is the diameter of the action space 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 Assumption 5.1 (ii) and continuous at $\phi=\phi_0$ by Assumption 5.1 (iv), it holds that $\int_\Phi d_H(I(\phi),I(\phi_0))\,d\pi_{\phi|X}\to 0$ as $\pi_{\phi|X}$ converges weakly to the point mass measure at $\phi=\phi_0$. This implies the uniform convergence of $R_n(\delta)$, i.e., $\sup_\delta|R_n(\delta)-R_\infty(\delta)|\to 0$ 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

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

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

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

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

$$l(\delta,\phi) - l(\delta,\phi_0) \tag{54}$$
$$= \begin{cases} (1-\tau)\left[\alpha_*(\phi_0) - \alpha_*(\phi)\right], & \text{if } \delta \leq (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi), \\ \tau\left[\alpha^*(\phi) - \alpha_*(\phi_0)\right] - \left[\delta - \alpha_*(\phi_0)\right], & \text{if } (1-\tau)\alpha_*(\phi) + \tau\alpha^*(\phi) < \delta \leq (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0), \\ \tau\left[\alpha^*(\phi) - \alpha^*(\phi_0)\right] & \text{if } \delta > (1-\tau)\alpha_*(\phi_0) + \tau\alpha^*(\phi_0). \end{cases}$$

By noting that in the second case in (54), 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

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

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

B Asymptotic Analysis with Discrete Benchmark Prior

This appendix modifies the large-λ asymptotic analysis of Section 5 in the main text to allow the benchmark prior $\pi^*_{\alpha|\phi}$ to be discrete. In particular, suppose that the benchmark conditional prior is a mixture of a finite number of multiple probability masses. Such a benchmark prior can arise if the benchmark model corresponds to Bayesian model averaging over observationally equivalent point-identified models, e.g., just-identified SVARs differing in terms of the causal ordering assumptions. See, e.g., Giacomini et al. (2017) for the analysis of Bayesian model averaging over the observationally equivalent candidate models. An alternative setting that yields a discrete benchmark conditional prior is a locally-identified (but not globally-identified) structural model in which knowledge of the reduced-form parameters can pin down the structural parameters up to a discrete set of values, i.e., $IS_\theta(\phi)$ is a set with a finite number of elements. See Bacchiocchi and Kitagawa (2019) for a robust Bayesian approach to inference on locally-identified SVARs.

Given the reduced-form parameter ϕ, we denote the discrete set of support points of $\pi^*_{\alpha|\phi}$ by $\{\alpha_1(\phi), \ldots, \alpha_{M(\phi)}(\phi)\}$, where $M(\phi) < \infty$ is the number of support points that can depend on ϕ. We represent the benchmark prior as a discrete measure with those support points

$$\pi^*_{\alpha|\phi}(\alpha) = \sum_{m=1}^{M(\phi)} w_m(\phi) \cdot 1_{\alpha_m(\phi)}(\alpha), \;\; w_m(\phi) > 0 \;\; \forall m, \;\; \sum_{m=1}^{M(\phi)} w_m(\phi) = 1, \tag{55}$$

where $1_{\alpha'}(\alpha)$ is an indicator function for $\alpha = \alpha'$. We accordingly define $\alpha_*(\phi) \equiv \min_{1 \le m \le M(\phi)} \alpha_m(\phi)$ and $\alpha^*(\phi) \equiv \max_{1 \le m \le M(\phi)} \alpha_m(\phi)$ for the discrete benchmark prior case.

In the context of the Bayesian model averaging over observationally equivalent models (where $M(\phi) = \bar{M}$ should be independent of ϕ), the probability weights $(w_1(\phi), \ldots, w_{\bar{M}}(\phi))$ specify benchmark credibility over each of $\bar{M}$ point-identified models. Our robust Bayesian analysis applied to the model averaging setting concerns ambiguity in the initial allocation of the model weights.

To accommodate the discrete benchmark prior, we replace Assumption 5.1 in the main text with the following.

Assumption B.1 *At ϕ_0 the true value of the reduced-form parameters, $\alpha_*(\phi)$ and $\alpha^*(\phi)$ are continuous at $\phi = \phi_0$.*

With the discrete benchmark conditional prior, Theorem B.2 below shows large λ and large n asymptotic results for the conditional minimax decision, which is analogous to Theorems 5.2 and 5.3 in the main text covering the case with the continuous benchmark prior.

Theorem B.2 *Assume that the benchmark conditional prior $\pi^*_{\alpha|\phi}$ is given in the form of (55). Suppose Assumption 3.2 (ii) and (iv) hold.*

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

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

holds whenever the right-hand side integral is finite.

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

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

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

Theorem B.3 *Assume that the benchmark conditional prior $\pi^*_{\alpha|\phi}$ is given in the form of (55). Suppose Assumption 3.2 and Assumption B.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 of Theorem B.2. (i) Fix δ and ϕ. For short notation, denote $\left(\alpha_1(\phi), \ldots, \alpha_{M(\phi)}(\phi)\right)$ by $(\alpha_1, \ldots, \alpha_M)$. By Lemma A.4, $\lambda \to \infty$ implies $\kappa_\lambda(\delta, \phi) \to 0$. Hence, similarly to equation (51) in the proof of Theorem 5.2, the pointwise limit $\lim_{\lambda \to \infty} r_\lambda(\delta, \phi)$ can be obtained by

$$\lim_{\kappa \to 0} \frac{\sum_m (\delta - \alpha_m)^2 \exp\left\{\frac{(\delta - \alpha_m)^2}{\kappa}\right\} w_m}{\sum_m \exp\left\{\frac{(\delta - \alpha_m)^2}{\kappa}\right\} w_m}.$$

Let $\alpha^\star \equiv \arg\max_{\{\alpha_1, \ldots, \alpha_M\}} (\delta - \alpha_m)^2$, $\mathcal{M}^* = \arg\max_m (\delta - \alpha_m)^2$, and $w^* = \sum_{m \in \mathcal{M}^*} w_m > 0$. Then,

$$\frac{\sum_m (\delta - \alpha_m)^2 \exp\left\{\frac{(\delta - \alpha_m)^2}{\kappa}\right\} w_m}{\sum_m \exp\left\{\frac{(\delta - \alpha_m)^2}{\kappa}\right\} w_m} = \frac{(\delta - \alpha^\star)^2 w^* + \sum_{m \notin \mathcal{M}^*} (\delta - \alpha_m)^2 \exp\left\{-\frac{(\delta - \alpha^\star)^2 - (\delta - \alpha_m)^2}{\kappa}\right\} w_m}{w^* + \sum_{m \notin \mathcal{M}^*} \exp\left\{-\frac{(\delta - \alpha^\star)^2 - (\delta - \alpha_m)^2}{\kappa}\right\} w_m}$$

$$\to (\delta - \alpha^\star)^2 = (\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2,$$

as $\kappa \to 0$. The dominated convergence theorem leads to the conclusion of (i). The proof of (ii) proceeds similarly, and we omit a proof for brevity. ■

Proof of Theorem B.3. (i) Posterior consistency of ϕ, compactness of the parameter space of α (Assumption 3.2 (i) and (ii)), and Assumption B.1 imply the convergence of $R_n(\delta)$ to $R_\infty(\delta)$, as shown in equation (53) in the proof of Theorem 5.3. Repeating the argument of the proof of Theorem 5.3, this convergence can be shown to be uniform in δ. Hence, $\hat{\delta}_\infty \to \arg\min_\delta (\delta - \alpha_*(\phi))^2 \vee (\delta - \alpha^*(\phi))^2 = \frac{1}{2}(\alpha_*(\phi_0) + \alpha^*(\phi_0))$ holds.

The claim of (ii) can be shown similarly. ■

C The Benchmark Prior Specification in the SVAR application

This section provides the precise construction of the benchmark prior used in our empirical application of Section 7.

Let $f_t(x; c, \sigma, \nu)$ be the pdf of Student's t-distribution with location c, scale σ, and degree of freedom ν. The prior distribution of (β_s, β_d) are independent t-distributions truncated by the sign constraints $\beta_s \geq 0$ and $\beta_d \leq 0$:

$$\frac{\partial^2 \pi_{(\beta_s,\beta_d)}}{\partial\beta_d \partial\beta_s} = \frac{f_t(\beta_s; c_s, \sigma, \nu)}{1 - \int_0^\infty f_t(\beta_s; c_s, \sigma, \nu) d\beta_s} \cdot \frac{f_t(\beta_d; c_d, \sigma, \nu)}{\int_{-\infty}^0 f_t(\beta_d; c_d, \sigma, \nu) d\beta_d},$$

where $c_s = 0.6$, $c_d = -0.6$, $\sigma = 0.6$, and $\nu = 3$.

Regarding the conditional prior of the structural variances (d_1, d_2) given (β_s, β_d), we specify independent independent inverse gamma distributions with shape parameter κ_i and scale parameter τ_i, i=1,2. We set the shape parameters to be common across the distributions, with $\kappa_1 = \kappa_2 = 2$. Let $\hat{\epsilon}_t = (\hat{\epsilon}_{dt}, \hat{\epsilon}_{st})'$ be the residuals of the least-square estimation of the reduced-form VAR, and let $\hat{\Omega} = (T-8)^{-1} \sum_{t=9}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$. We set scale parameters τ_1 and τ_2 to $\tau_i = \kappa_i a_i' \hat{\Omega} a_i$, where a_i' is the i-th row vector of A_0.

Next, we specify a conditional prior for the remaining structural coefficients, $A = (c, A_1, \ldots, A_L)$, given $(\beta_s, \beta_d, d_1, d_2)$. Let b_i, $i = 1, 2$, be the i-th row vector of A with length $(2L+1)$. We specify the prior for b_1 and b_2 to be independent multivariate Gaussian, and denote b_i's mean vector and variance-covariance matrix by m_i and M_i, respectively. We set $m_i' = (0, a_i', \mathbf{0}')$, and let M_i be the diagonal matrix whose j-th element, $j = 1, \ldots, (2L+1)$, corresponds to j-th element of the following vector

$$v_3 = \eta_0^2 \begin{pmatrix} \eta_1^2 \\ v_1 \otimes v_2 \end{pmatrix},$$

where $\eta_0 = 0.2$, $\eta_1 = 100$, $v_1 = (1/(1^2), 1/(2^2), \ldots, 1/(L^2))'$, and v_2 is the vector of diagonal elements of $\hat{\Omega}$.

D Entry Game Example

As a microeconometric application, consider the two-player entry game in **?** used as the illustrating example in Moon and Schorfheide (2012). Let $\pi_{ij}^M = \beta_j + \epsilon_{ij}$, $j = 1, 2$, be firm j's profit in market i, $i \in \{1, \ldots, n\}$, if firm j is monopolistic in market i, $and \pi_{ij}^D = \beta_j - \gamma_j + \epsilon_{ij}$ be firm j's profit in market i if the competing firm also enters the market i (duopolistic). ϵ_{ij} represents components of firm j's profits in market i that are known by the firms but not observed by the econometrician. We assume $(\epsilon_{i1}, \epsilon_{i2})$ follow $\mathcal{N}(0, I_2)$ independently and identically over i. We restrict our analysis to the pure strategy Nash equilibrium, and assume that the decisions are strategic substitutes, so $\gamma_1, \gamma_2 \geq 0$. The data consist of iid observations on entry decisions of the two firms. The non-redundant set of reduced-form parameters are $\phi = (\phi_{11}, \phi_{00}, \phi_{10})$, which are, respectively, the probabilities of observing a duopoly, no entry, or the monopoly 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 equilibria 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 values of $(\epsilon_{i1}, \epsilon_{i2})$ leading to multiplicity of equilibria. Let the parameter of interest be $\alpha = \gamma_1$, the substitution effect for firm 1 from firm 2 entering. The vector of full structural parameters augmented by the equilibrium selection parameter ψ is $(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$, with the additional a priori restriction $\gamma_1, \gamma_2 \geq 0$. This parameter vector can be reparametrized into $(\beta_1, \gamma_1, \phi_{11}, \phi_{00}, \phi_{10})$.[6] 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

$$IS_\theta(\phi) = \left\{ (\beta_1, \gamma_1) : \gamma_1 \geq 0, \min_{\beta_2 \in \mathbb{R}^2, \gamma_2 \geq 0, \psi \in [0,1]} \|\phi - \phi(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)\| = 0 \right\}, \tag{56}$$

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 $\alpha = \gamma_1$.

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

$$\begin{aligned} \phi_{11} &= G(\beta_1 - \gamma_1) G(\beta_2 - \gamma_2), \\ \phi_{00} &= (1 - G(\beta_1))(1 - G(\beta_2)), \\ \phi_{10} &= G(\beta_1)\left[1 - G(\beta_2)\right] + G(\beta_1 - \gamma_1)\left[G(\beta_2) - G(\beta_2 - \gamma_2)\right] \\ &\quad + \psi\left[G(\beta_1) - G(\beta_1 - \gamma_1)\right]\left[G(\beta_2) - G(\beta_2 - \gamma_2)\right]. \end{aligned} \tag{57}$$

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 variants. Plugging them into (56) the yields the posterior draws of ϕ. The transformation

[6]See Appendix D below for concrete expressions of this transformation.

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

$$
\begin{aligned}
\beta_1 &= \beta_1, \\
\gamma_1 &= \gamma_1, \\
\beta_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) \equiv \beta_2(\beta_1, \phi), \\
\gamma_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) - G^{-1}\left(\frac{\phi_{11}}{G(\beta_1 - \gamma_1)}\right) \equiv \gamma_2(\beta_1, \gamma_1, \phi), \\
\psi &= \frac{[1 - G(\beta_1)]\,[\phi_{10} + \phi_{11} - G(\beta_1 - \gamma_1)] + [G(\beta_1) - G(\beta_1 - \gamma_1)]\,\phi_{00}}{[G(\beta_1) - G(\beta_1 - \gamma_1)]\left[1 - G(\beta_1) - \phi_{00} - \frac{1 - G(\beta_1)}{G(\beta_1 - \gamma_1)}\phi_{11}\right]} \equiv \psi(\beta_1, \gamma_1, \phi).
\end{aligned}
\tag{58}
$$

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 (57). Solving for the multiplier minimax estimator for γ_1 follows similar steps to those in Algorithm 6.1, 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

Andrews, I., M. Gentzkow, and J. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132, 1553–1592.

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

Bacchiocchi, E. and T. Kitagawa (2019): "Locally (but not Globally) Identified SVARs," *unpublished manuscript.*

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

——— (2018): "Inference in Structural Vector Autoregressions When the Identifying Assumptions are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations," *Journal of Monetary Economics*, 100, 48–65.

——— (2019): "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 109, 1873–1910.

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.

Chen, X., T. Christensen, and E. Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965–2018.

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 (2017): "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.

Lütkepohl, H. (1991): *Introduction to Multiple Times Series*, Springer.

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 (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.

Norets, A. and X. Tang (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, 81, 1229–1262.

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102, 2410–2436.

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 Reivews*, 39, 9.

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.