Estimation under Ambiguity (Very preliminary)

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

Abstract

To perform a Bayesian analysis for a set-identified model, two distinct approaches exist; the standard Bayesian inference that assumes a single prior for non-identified parameters, and the Bayesian inference for the identified set that assumes full ambiguity (multiple priors) for the parameters within their identified set. Both of the prior inputs considered by these two extreme approaches can often be a poor representation of the researcher's prior knowledge in practice. This paper fills this large gap between the two approaches by proposing a framework of multiple prior robust Bayes analysis that can simultaneously incorporate a probabilistic belief for the non-identified parameters and a misspecification concern about this belief. Our proposal introduces a *benchmark prior* representing the researcher's partially credible probabilistic belief for non-identified parameters, and *a set of priors* formed in its Kullback-Leibler (KL) neighborhood whose radius controls the degree of researcher's confidence put on the benchmark prior. We develop point estimation and interval estimation for the object of interest by minimizing the worst-case posterior risk over the resulting class of posteriors. We clarify that this minimax problem is analytically tractable and simple to solve numerically. We also derive analytical properties of the proposed robust Bayesian procedure in the limiting situations where the radius of KL neighborhood and/or the sample size are large.

[*] University College London, Department of Economics/Cemmap. Email: r.giacomini@ucl.ac.uk
[†] University College London, Department of Economics/Cemmap. Email: t.kitagawa@ucl.ac.uk
[‡] University of Chicago, Department of Economics. Email: huhlig@uchicago.edu

1 Introduction

Consider a parametric model with parameter vectors (θ, ϕ), where the likelihood is given by $l(x|\theta, \phi)$ and a sample is denoted by $X = x$. We consider a partially identified model (e.g., Poirier (1998), Moon and Schorfheide (2012)) where $\phi \in \Phi$ denotes the reduced-form parameters and θ denotes the auxiliary parameters that are non-identified under a given set of identifying assumptions but are necessary to pin down the value of a (scalar) object of interest $y = y(\theta, \phi) \in \mathbb{R}$. By the definition of reduced-form parameters, the value of the likelihood depends only on ϕ for every realization of X, equivalent to saying $X \perp \theta|\phi$. The domain of θ, on the other hand, is constrained by the imposed identifying assumptions and it depends on ϕ, and we refer to it as the identified set of θ denoted by $IS_\theta(\phi)$. The identified set of y is accordingly defined by the range of $y(\theta, \phi)$ when θ varies over $IS_\theta(\phi)$,

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

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

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

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

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

where $x_t = (\Delta w_t, \Delta n_t)$ with Δw_t and Δn_t the growth rates of wages and employment, respectively, $A = \begin{bmatrix} -\beta & 1 \\ -\alpha & 1 \end{bmatrix}$ with $\alpha \geq 0$ the short-run wage elasticity of supply and $\beta \leq 0$ the short-run wage elasticity of demand and u_t are shocks assumed to be i.i.d. $N(0, D)$ with $D = diag(d_1, d_2)$. The reduced form representation of the model is

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

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

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

[1] See Section 6.1 below for the transformation. If β is a parameter of interest, an alternative reparametrization allows us to transform the structural parameters into $(\beta, w_{11}, w_{12}, w_{22})$.

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

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

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

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

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

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

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

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

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

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

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

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

Generally, the identified set only collects all the admissible values of y given dogmatically imposed identifying assumptions and a knowledge of the distribution of observables (the reduced-form parameters). In some contexts, however, it may not be the case that the identifying assumptions imposed dogmatically exhaust all the available information that the researcher has. A more common situation is that the researcher has some form of additional but only partially credible assumptions about some underlying structural parameters or about the non-identified parameter θ based on economic theory, background knowledge of the problem, or empirical studies that use different data. From the standard Bayesian viewpoint, the recommendation is to incorporate this information into the analysis through specifying a prior distribution of (θ, ϕ) or that of the full structural parameters. For instance, in the case of Example 1.1, Baumeister and Hamilton (2015) propose a prior of the elasticity of supply α that draws on estimates obtained in microeconometric studies, and consider a Student's t density calibrated to assign 90% probability to the interval $\alpha \in (0.1, 2.2)$. Another example considered by Baumeister and Hamilton (2015) is a prior that incorporates long-run identifying restrictions in SVARs in a non-dogmatic way, as a way to capture the uncertainty one might have about the validity of this popular but controversial type of identifying restrictions. In the situations where additional informative prior information other than the identifying restrictions is not available, some Bayesian literature has recommended the use of the uniform prior as a representation of the indifference among θ's within the identified set. For example, in SVARs subject to sign restrictions (Uhlig (2005)) it is common to use the uniform distribution (the Haar measure) over the set of orthonormal matrices in (7) that satisfy the sign restrictions. For the entry game in Example 1.3, one of the prior specifications considered in Moon and Schorfheide (2012) is the uniform prior over the identified set of θ.

At the opposite end of the standard Bayesian spectrum, Kitagawa (2012) and Giacomini and Kitagawa (2015) advocate adopting a multiple-prior Bayesian approach when one has no

[2] See Section 6.2 below for concrete expressions of the transformation.

further information about θ besides a set of exact restrictions that can be used to characterize the identified set. While maintaining a single prior for ϕ, this set of priors consists of any conditional prior for θ given ϕ, $\pi_{\theta|\phi}$, supported on the identified set $IS_\theta(\phi)$. Kitagawa (2012) and Giacomini and Kitagawa (2015) propose to conduct a posterior bound analysis based on the resulting class of posteriors, that leads to an estimator for $IS_y(\phi)$ and an associated "robustified" credible region that asymptotically converge to the true identified set, which is the object of interest of frequentist inference. Being implicit about the ambiguity inherent in partial identification analysis, one can also consider posterior inference for the identified set as in Moon and Schorfheide (2011), Kline and Tamer (2013), and Liao and Simoni (2013), to obtain similar asymptotic equivalence between posterior inference and frequentist inference.

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

The main contribution of this paper is to fill the large gap between the single prior Bayesian approach and the fully ambiguous multiple prior Bayesian approach by proposing a method that can simultaneously incorporate a probabilistic belief for the non-identified parameters and a misspecification concern about this belief in a unified manner. Our idea is to replace the fully ambiguous beliefs for $\pi_{\theta|\phi}$ considered in Kitagawa (2012) and Giacomini and Kitagawa (2015) by a class of priors defined in a neighborhood of a *benchmark prior*. The benchmark prior $\pi^*_{\theta|\phi}$ represents the researcher's reasonable but partially credible prior knowledge about θ, and the class of priors formed around the benchmark prior captures ambiguity or misspecification concerns about the benchmark prior. The radius of the neighborhood prespecified by the researcher controls the degree of confidence put on the benchmark prior. We then propose point estimation and interval estimation for the object of interest y by minimizing the worst-case (minimax) posterior risk over the priors constrained to this neighborhood. Building on the robust control theory pioneered in operation research (Peterson, James, and Dupuis (2000)) and macroeconomics (Hansen and Sargent (2001)), we solve this constrained minimax problem via the unconstrained multiplier minimax formulation. Our paper makes the following unique contributions: (1) we clarify that the estimation for the partially identified parameter under vague prior knowledge can be formulated as a decision under ambiguity in the form considered in Hansen and Sargent (2001); (2) we provide an analytically tractable and numerically convenient way to solve the minimax estimation problem in general cases; (3).we give simple analytical solutions for the special cases of a quadratic and a check loss function and for the limit case when the shape of benchmark prior is irrelevant; (4) we derive the properties of our method in large samples.

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

2 Estimation as Statistical Decision under Ambiguity

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

Suppose that the decision maker can form a benchmark prior $\pi^*_{\theta|\phi}$ for the unrevisable part of the prior. This prior captures information about θ that is available before the model is brought to the data (see Section 5 for discussions on how to elicit a benchmark prior). Note that, if one were to impose a sufficient number of restrictions to point-identify y, this would amount to specifying the benchmark prior of y to be a point mass measure that selects one particular point from the identified set. With such point-mass prior, the posterior of ϕ induces a single posterior of y. Under partial identification, on the other hand, $\pi_{\theta|\phi}$ needs to be specified in order to have a single posterior distribution for y.

We consider a set of priors (ambiguous beliefs) in a neighborhood of $\pi^*_{\theta|\phi}$ - while maintaining a single prior of ϕ - and find the estimator of y that minimizes the worst-case posterior risk as the priors range over this neighborhood. Formally, define the Kullback-Leibler neighborhood of $\pi^*_{\theta|\phi}$ with radius $\lambda \in [0, \infty)$ as

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

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

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

which 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)). The main reason to define the neighborhood in terms of the KL-distance is its convexity property in $\pi_{\theta|\phi}$, which allows us to transform the constrained minimax problem in equation (13) below into the analytically more tractable unconstrained minimax problem in equation (14) below.

A bigger λ corresponds to a larger $\Pi^\lambda \left(\pi^*_{\theta|\phi}\right)$, and, in the extreme case, $\Pi^\infty(\pi^*_{\theta|\phi}) \equiv \lim_{\lambda \to \infty} \Pi^\lambda \left(\pi^*_{\theta|\phi}\right)$ contains any probability measure that is dominated by $\pi^*_{\theta|\phi}$, i.e., the benchmark prior becomes relevant only for determining the support of $\pi_{\theta|\phi}$ in the limiting situation of $\lambda \to \infty$. Note that $\Pi^\lambda \left(\pi^*_{\theta|\phi}\right)$ is defined for conditional priors at each ϕ so that the radius λ can be different over ϕ, although it is implicit in our notation. Indeed, in the multiplier minimax approach shown below, the implied set of priors for $\pi_{\theta|\phi}$ has the radius dependent on ϕ. It is important also to note that other than through the benchmark prior, the class of priors is not subject to any constraint that restricts the dependence of θ on 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 consider a point estimation problem where $\delta(X)$ is a scalar statistical decision function that maps data X to a space of actions and $h(\delta(X), y)$ is a loss function, such as the quadratic loss

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

or the check loss for the τ-th quantile

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

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

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

Provided that the decision maker faces ambiguous beliefs for $\pi_{\theta|\phi}$ in the form of multiple priors $\Pi^\lambda\left(\pi^*_{\theta|\phi}\right)$, we assume that the decision maker wishes to make a robust or conservative decision for y by minimizing the worst-case posterior risk over $\Pi^\lambda\left(\pi^*_{\theta|\phi}\right)$ given data $X = x$,

$$\textit{Constrained Minimax:} \quad \min_{\delta(x)} \int_{\Phi} \max_{\pi_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_\theta(\phi)} h(\delta(x), y(\theta, \phi)) d\pi_{\theta|\phi} \right] d\pi_{\phi|X} \tag{13}$$

Instead of working with the *constrained minimax* problem above, we consider the more analytically convenient *multiplier minimax* problem: for $\kappa > 0$,

$$\textit{Multiplier Minimax:} \quad \min_{\delta(x)} \int_{\Phi} \left[\max_{\pi_{\theta|\phi} \in \Pi^\infty\left(\pi^*_{\theta|\phi}\right)} \left\{ \int_{IS_\theta(\phi)} h(\delta(x), y(\theta, \phi)) d\pi_{\theta|\phi} - \kappa \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \right\} \right] d\pi_{\phi|X}. \tag{14}$$

The following well-known result from convex analysis[3] that shows equivalence between the two minimax problems.

Lemma 2.1 *Fix ϕ and $\delta(x)$. Assume $\pi^*_{\theta|\phi}(\theta)$ is nondegenerate and $\lambda > 0$. Let*

$$\mu_0 \equiv \max_{\pi_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)} \left[\int_{IS_\theta(\phi)} h(\delta(x), y(\theta, \phi)) d\pi_{\theta|\phi} \right]. \tag{15}$$

If $\mu_0 < \infty$, then there exists $\kappa \geq 0$ such that

$$\mu_0 = \max_{\pi_{\theta|\phi} \in \Pi^\infty\left(\pi^*_{\theta|\phi}\right)} \left\{ \int_{IS_\theta(\phi)} h(\delta(x), y(\theta, \phi)) d\pi_{\theta|\phi} - \kappa \left(\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) - \lambda \right) \right\}. \tag{16}$$

[3] See Lemma 2.2. in Peterson, James, and Dupuis (2000). These authors refer to David Luenberger's book (1969) "*Optimization by Vector Space Methods*" for a proof. Theorem 28.2 in Tyrrell Rockafeller's book (1970) "*Convex Analysis*" shows the same claim.

Furthermore, if $\pi^0_{\theta|\phi} \in \Pi^\lambda\left(\pi^*_{\theta|\phi}\right)$ *is a maximizer in (15),* $\pi^0_{\theta|\phi}$ *also maximizes (16) and satisfies* $\kappa\left(\mathcal{R}(\pi_{\theta|\phi}\|\pi^*_{\theta|\phi}) - \lambda\right) = 0$.

As is clear from the lemma, κ plays the role of the Lagrangian multiplier in a constrained optimization problem and thus it can be interpreted as the increase in the objective function associated with a relaxation of the constraint (a unit change in λ). Note that, since the constrained optimization problem depends on ϕ through $\pi^*_{\theta|\phi}$ and $IS_\theta(\phi)$, a value of κ that equalizes the two optimizations may depend on ϕ if λ is a constant independent of ϕ. One way to justify our fixed κ multiplier minimax analysis is therefore to think of the original constrained problem as having λ dependent on ϕ.[4]

3 Solving Multiplier Minimax Problem

The multiplier minimax problem of (14) has a convenient representation, as shown in the next theorem.

Theorem 3.1 *Assume* $h(\delta, y)$ *is bounded on* $IS_y(\phi)$, $\pi_\phi - a.s.$ *at every* δ. *The multiplier minimax problem (14) is then equivalent to*

$$\min_\delta \int_\Phi r_\kappa(\delta,\phi)\, d\pi_{\phi|X}, \tag{17}$$

$$\text{where} \quad r_\kappa(\delta,\phi) \equiv \kappa \ln\left(\int_{IS_\theta(\phi)} \exp\left\{h(\delta, y(\theta,\phi))/\kappa\right\} d\pi^*_{\theta|\phi}\right)$$

Proof. See Appendix A. ■

Note that the statement of the theorem is valid for any sample size and any realization of X. The obtained representation significantly simplifies the analytical investigation and the computation of the minimax decision, and we make use of it in the following sections. We can easily approximate the integrals in (17) using Monte Carlo draws of (θ,ϕ) sampled from the benchmark conditional prior $\pi^*_{\theta|\phi}$ and the posterior $\pi_{\phi|X}$. The minimization for $\delta(x)$ can

[4]Note that in the extreme situation where $\lambda = 0$ or $\lambda = \infty$, the optimal decision in the constrained minimax problem can be replicated by the multiplier minimax decision with κ independent of ϕ. When $\lambda = 0$, the optimal decision in the constrained minimax problem is reduced to the standard Bayes decision with a single posterior. This standard Bayes decision can be replicated by the multiplier minimax decision with $\kappa = \infty$, since if $\kappa = \infty$, the inner maximization in (14) always selects the benchmark prior. When $\lambda = \infty$, the constrained minimax problem is reduced to the unconstrained one, so that the multiplier minimax problem with $\kappa = 0$ coincides it.

be performed by a grid search using the approximated objective function. Section 6 applies this idea to some common applications.

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

Another advantage of expressing the multiplier minimax problem as in Theorem 3.1 is that it simplifies the investigation of the behavior of the optimal decision in large samples. 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) The posterior of ϕ is consistent for ϕ_0 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) $\mathcal{D}$ (the action space of δ), $\mathcal{Y}$ (the parameter space of y), and Φ (the parameter space of ϕ) are compact.

(iii) The loss function $h(\delta, y)$ is non-negative, bounded, and continuous in δ at every $(\delta, y) \in \mathcal{D} \times \mathcal{Y}$.

(iv) $IS_y(\phi)$ has a nonempty interior π_ϕ-a.s. and $IS_y(\phi_0)$ has a nonempty interior. The benchmark prior marginalized to y, $\pi^*_{y|\phi}$, is absolutely continuous with respect to the Lebesgue measure, and its density is differentiable in ϕ with bounded derivatives, $\left\| \frac{\partial}{\partial \phi} \left(\frac{d\pi^*_{y|\phi}}{dy} \right) \right\| \leq M < \infty$ at almost every $y \in IS_y(\phi)$, π_ϕ-a.s.

(v) $r_\kappa(\delta, \phi_0) \equiv \kappa \ln \left(\int_{IS_\theta(\phi_0)} \exp\left\{ h(\delta, y(\theta, \phi_0))/\kappa \right\} d\pi^*_{\theta|\phi_0} \right)$ 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. We do not present them here for brevity (see, e.g., Section 7.4 of Schervish (1995) for details about posterior consistency). Assumption 3.2 (iv) rules out point-identified models and assumes that the identified set has almost surely positive length.[5]

Under these regularity assumptions, we obtain the following asymptotic result.

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

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

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

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* $n \to \infty$*,* $\delta_\kappa\left(\hat{\phi}\right) \in \arg\min_{\delta \in \mathcal{D}} r_\kappa\left(\delta, \hat{\phi}\right)$ *converges in probability to* $\delta_\kappa(\phi_0)$ *as* $n \to \infty$*.*

Proof. See Appendix A. ∎

This theorem shows that the finite sample optimal minimax decision has a well-defined large sample limit which coincides with the optimal decision under the knowledge of the true value of ϕ. The theorem has a useful practical implication: When the sample size is moderate to large, so that the posterior distribution of ϕ is concentrated around its maximum likelihood estimator (MLE) $\hat{\phi}_{ML}$, one can well approximate the exact finite sample minimax decision by minimizing the "plug-in" objective function, where the averaging with respect to the posterior of ϕ in (17) is replaced by plugging $\hat{\phi}_{ML}$ in $r_\kappa(\delta, \phi)$. This will reduce computational cost of approximating the objective function since what we need in this case are only MCMC draws of θ (or y) from $\pi_{\theta|\hat{\phi}_{ML}}$ (or $\pi_{y|\hat{\phi}_{ML}}$).

4 Multiplier Minimax Estimation with Specific Loss Functions

This section presents further analytical results on the multiplier minimax decision problem for two common choices of loss function. In particular, we focus on the limiting situation of $\kappa \to 0$, i.e., the case when the decision maker faces extreme ambiguity.

In the case when $\kappa \to 0$, the choice of the benchmark conditional prior $\pi^*_{y|\phi}$ affects the optimal decision only through the support of the prior. We therefore impose the following regularity assumptions concerning the tail behavior of the benchmark conditional prior.

Assumption 4.1 *(i)* $IS_y(\phi)$ *has a nonempty interior* π_ϕ*-a.s. and the benchmark prior marginalized to y,* $\pi^*_{y|\phi}$*, is absolutely continuous with respect to the Lebesgue measure* π_ϕ*-a.s.*

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

(iii) There exist $\eta > 0$*,* $\alpha > 0$*, and* $b > 0$ *such that* $[y_*(\phi), y_*(\phi) + \eta) \subset IS_y(\phi)$ *and* $(y^*(\phi) - \eta, y^*(\phi)] \subset IS_y(\phi)$ *hold and the tails of* $\pi^*_{y|\phi}$ *near the boundary of the support satisfy*

$$\begin{aligned} \frac{d\pi^*_{y|\phi}}{dy}(y) &\geq b(y - y_*(\phi))^\alpha, \quad \forall y \in [y_*(\phi), y_*(\phi) + \eta) \quad \text{and} \\ \frac{d\pi^*_{y|\phi}}{dy}(y) &\geq b(y^*(\phi) - y)^\alpha, \quad \forall y \in (y^*(\phi) - \eta, y^*(\phi)], \quad \pi_\phi\text{-a.s.} \end{aligned}$$

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

Assumption 4.1 (i) rules out point-identified models as in Assumption 3.2 (i), though the current one is slightly weaker than Assumption 3.2 (i). Assumption 4.1 (ii) assumes that the benchmark conditional prior has bounded support, which automatically holds if the identified set $IS_y(\phi)$ is bounded. In particular, if the benchmark conditional prior supports the entire identified set, $[y_*(\phi), y^*(\phi)] = [\underline{y}(\phi), \bar{y}(\phi)]$ holds. Assumption 4.1 (iii) restricts the behavior of the benchmark conditional prior locally around the boundaries of the support. It requires the benchmark conditional prior to be bounded from below by a polynomial function with degree $\alpha > 0$ in a neighborhood of the support boundaries. When the density of the benchmark conditional prior is strictly positive at $y_*(\phi)$ and $y^*(\phi)$, then the polynomial lower bound conditions clearly holds. Assumption 4.1 (iv) implies that the support of the benchmark conditional prior varies continuously in ϕ. Assumption 4.1 is for example satisfied by the product of Student's t prior considered by Baumeister and Hamilton (2015), provided its support is bounded.

The next two theorems characterize the asymptotic behavior of the multiplier minimax decisions for the quadratic loss and the check loss. Theorem 4.2 concerns the limiting situation of $\kappa \to 0$ with a fixed sample size. Theorem 4.3 concerns the large sample asymptotics with $\kappa \to 0$.

Theorem 4.2 *Suppose Assumption 4.1 (i) - (iii) hold.*
(i) When $h(\delta, y) = (\delta - y)^2$,

$$\lim_{\kappa \to 0} \int_{\Phi} r_\kappa(\delta, \phi)\, d\pi_{\phi|X} = \int_{\Phi} \left[(\delta - y_*(\phi))^2 \vee (\delta - y^*(\phi))^2 \right] d\pi_{\phi|X}$$

holds whenever the right-hand side integral is finite.
(ii) When $h(\delta, y) = \rho_\tau(y - \delta)$,

$$\lim_{\kappa \to 0} \int_{\Phi} r_\kappa(\delta, \phi)\, d\pi_{\phi|X} = \int_{\Phi} \left[(1-\tau)(\delta - y_*(\phi)) \vee \tau\,(y^*(\phi) - \delta) \right] d\pi_{\phi|X}$$

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

Proof. See Appendix A. ■

Theorem 4.3 *Suppose Assumption 3.2 (i)-(ii) and Assumption 4.1 hold. Let*

$$\hat{\delta}_0 = \arg\min_{\delta \in \mathcal{D}} \left\{ \lim_{\kappa \to 0} \int_{\Phi} r_\kappa(\delta, \phi)\, d\pi_{\phi|X}(\phi) \right\}$$

be the multiplier minimax estimator in the limiting case $\kappa \to 0$.

(i) When $h(\delta, y) = (\delta - y)^2$, $\hat{\delta}_0 \to \frac{1}{2}(y_*(\phi_0) + y^*(\phi_0))$ *as the sample size* $n \to \infty$ *for almost every sampling sequence.*

(ii) When $h(\delta, y) = \rho_\tau(y - \delta)$, $\rho_\tau(u) = \tau u \cdot 1\{u > 0\} - (1 - \tau)u \cdot 1\{u < 0\}$, $\hat{\delta}_0 \to (1 - \tau) y_*(\phi_0) + \tau y^*(\phi_0)$ *as the sample size* $n \to \infty$ *for almost every sampling sequence.*

Proof. See Appendix A. ■

Theorem 4.2 shows that in the most ambiguous situation of $\kappa \to 0$, only the convex hull of the support of the benchmark prior, $[y_*(\phi), y^*(\phi)]$, matters for the optimal minimax decision as far as the tail condition of Assumption 3.2 holds, and the shape of $\pi^*_{y|\phi}$ is irrelevant for the minimax decision. This result is intuitive since smaller κ implies a larger class of priors and at the limit $\kappa \to 0$, any priors that share the support with the benchmark prior are included in the prior class.

Theorem 4.3 (i) shows that in the large sample situation, the minimax decision with the quadratic loss converges to the middle point of the boundary points of the support of the benchmark prior evaluated at the true reduced form parameters. When the benchmark prior supports the entire identified set, this means that the minimax decision at the limit is to report the central point of the true identified set. When the loss is the check function associated with the τ-th quantile, the minimax decision at the limit is given by the convex combination of the same boundary points with weights τ and $1 - \tau$. One useful implication of this result is that, in the case of the check loss, solving for the optimal δ can be seen as obtaining the robustified posterior τ-th quantile of y, and the optimal δ may be used to construct a robustified interval estimate for y that explicitly incorporates the ambiguous beliefs about the benchmark prior.

An implication of Proposition 4.3 is that in the case of the quantile check function the optimal estimator $\delta^*(\tau)$ always lies in the true identified set for any τ, even in the most conservative case, $\kappa \to 0$. This means that, if we use $[\delta^*(0.05), \delta^*(0.95)]$ as a robustified posterior credibility interval for y, this interval estimate will be asymptotically strictly narrower than the frequentist confidence interval for y, as $[\delta^*(0.05), \delta^*(0.95)]$ is contained in the true identified set asymptotically. This result is similar to the finding in Moon and Schorfheide (2012) for the single posterior Bayesian credible interval.

The asymptotic results of Theorems 4.2 and 4.3 assume that the benchmark prior is absolutely continuous with respect to the Lebesgue measure. We can instead consider a setting where the benchmark prior is given by a nondegenerate probability mass measure, which can naturally arise if the benchmark prior comes from a weighted combination of multiple point-identified models. This case leads to asymptotic results similar to Theorem 4.3. We present the formal analysis for this discrete benchmark prior setting in Appendix B.

5 Eliciting Benchmark Prior and κ

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 parameter that determines the degree of robustness κ. This section presents some practical recommendations on how to choose them.

5.1 Benchmark Prior

5.2 Robustness Parameter κ

6 Examples

We now illustrate the practical implementation of our method in two of the examples that we discussed in the introduction. Importantly, we show how in SVARs it is possible to incorporate prior information on non-identified parameters that is expressed as unconditional priors, which acknowledges the fact that it may not always be easy to specify a prior that is conditional on reduced form parameters.

6.1 Demand and Supply

Consider again the 2-variable SVAR(0) in example 1.1, where the object of interest is the elasticity of supply α. Suppose that the benchmark prior is specified by an *unconditional* prior distribution for the full structural parameter $\tilde{\theta} = (\alpha, \beta, d_1, d_2)$, whose probability *density* is denoted by

$$\pi_{\tilde{\theta}}\left(\alpha, \beta, d_1, d_2\right). \tag{18}$$

If the sign restrictions $\alpha \geq 0$, $\beta \leq 0$ are imposed through the support of $\pi_{\tilde{\theta}}(\cdot)$ and $\pi_{\tilde{\theta}}(\cdot)$ is specified as in Baumeister and Hamilton (2015), MCMC draws of $\tilde{\theta}$ from its posterior can be obtained easily from the sampling algorithm presented in Baumeister and Hamilton (2015).

Consider solving the multiplier minimax problem (17) with fixed sample size and $\kappa > 0$. In order to solve the problem, we have to figure out the benchmark conditional prior of α given ϕ and the posterior of ϕ induced by the prior of $\tilde{\theta}$, or at least we have to be able to draw α's from the benchmark conditional prior and draw ϕ's from the posterior. The benchmark conditional prior of α given ϕ can be derived by reparametrizing $\tilde{\theta}$ in terms of (α, ϕ). Since $\Omega = A^{-1}D\left(A^{-1}\right)'$, we have that

$$\omega_{11} = \frac{d_1 + d_2}{(\alpha - \beta)^2}, \quad \omega_{12} = \frac{\alpha d_1 + \beta d_2}{(\alpha - \beta)^2}, \quad \omega_{22} = \frac{\alpha^2 d_1 + \beta^2 d_2}{(\alpha - \beta)^2} \tag{19}$$

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

$$\begin{aligned}
\alpha &= \alpha, \\
\beta &= \frac{\alpha\omega_{12} - \omega_{22}}{\alpha\omega_{11} - \omega_{12}} \equiv \beta(\alpha, \phi), \\
d_1 &= \omega_{11}\left(\alpha - \frac{\alpha\omega_{12} - \omega_{22}}{\alpha\omega_{11} - \omega_{12}}\right)^2 - \alpha^2\omega_{11} + 2\alpha\omega_{12} - \omega_{22} \equiv d_1(\alpha, \phi), \\
d_2 &= \alpha^2\omega_{11} - 2\alpha\omega_{12} + \omega_{22} \equiv d_2(\alpha, \phi).
\end{aligned} \tag{20}$$

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

$$\frac{d\pi^*_{\alpha|\phi}}{d\alpha}(\alpha|\phi) \propto \pi_{\tilde{\theta}}(\alpha, \beta(\alpha, \phi), d_1(\alpha, \phi), d_2(\alpha, \phi)) \times |\det(J(\alpha, \phi))|, \tag{21}$$

where $J(\alpha, \phi)$ is the Jacobian of the mapping (20), and $|\det(\cdot)|$ is the absolute value of the determinant. This benchmark conditional prior supports the entire identified set $IS_\alpha(\phi)$ if $\pi_{\tilde{\theta}}(\cdot)$ supports any value of (α, β) satisfying the sign restrictions. An analytical expression of the posterior of ϕ could be obtained by integrating out α in the right hand side of (21). Even if the analytical expression of the posterior of ϕ were not easy to derive, it would be easy to obtain posterior draws of ϕ by transforming the posterior draws of θ according to $\Omega = A^{-1}D\left(A^{-1}\right)'$. We hereafter denote the posterior draws of ϕ by $(\phi_1, \ldots, \phi_M)$. An algorithm that approximates the objective function in (17) is as follows.

Algorithm 6.1 *($\kappa > 0$) Let posterior draws of ϕ, $(\phi_1, \ldots, \phi_M)$, be given.*

1. *For each $m = 1, \ldots, M$, we approximate $r_\kappa(\delta, \phi_m) = \ln \int_{IS_\alpha(\phi_m)} \exp\{h(\delta, \alpha)/\kappa\}\, d\pi^*_{\alpha|\phi}$ by importance sampling, i.e., draw N draws of α, $(\alpha_{m1}, \ldots, \alpha_{mN})$ from a proposal distribution (probability density) $\tilde{\pi}_{\alpha|\phi}(\alpha|\phi)$ (e.g., the uniform distribution on $IS_\alpha(\phi_m)$) and compute*

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

   where

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

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

$$\frac{1}{M}\sum_{m=1}^M \hat{r}_\kappa(\delta, \phi_m), \tag{22}$$

   and minimize it with respect to δ.

If the limiting case $\kappa \to 0$ is considered (either with a quadratic or check loss), Lemma 4.2 implies that Step 1 of this algorithm can be skipped and we can directly approximate the objective function to be minimized in δ by

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

for the quadratic loss case, where $[\underline{\alpha}(\phi_m), \bar{\alpha}(\phi_m)]$ is the identified set of α.

For the typical sample sizes in macroeconometric applications, it is simple to compute (22) and there will not be significant computational gain in employing the asymptotic results. Nevertheless, if one is interested in the large sample approximation, one can approximate the posterior of ϕ by a point mass at $\phi = \hat{\phi}_{ML}$, and replace the objective function (22) with $\hat{r}_\kappa\left(\delta, \hat{\phi}_{ML}\right)$.

In Algorithm (6.1), we consider that the ϕ's are drawn from the posterior of ϕ induced by the prior of θ specified in (18). If the prior of θ implies an informative prior of ϕ, then in finite samples, this can downplay the sample information for ϕ in the sense that the shape of the posterior of ϕ does not well represent the shape of the likelihood for ϕ due to the informativeness of the prior of ϕ. Since the motivation of our method is a concern that the prior for θ may be misspecified, one may not want to impose the restrictions on ϕ implied by the prior for θ but "let the data speak". These concerns might make the following hybrid approach attractive. Draw ϕ's from the posterior of ϕ obtained from a non-informative prior of ϕ (e.g., Jeffreys' prior), and use of the benchmark prior of θ specified in (18) only for the purpose of constructing the benchmark conditional prior (21). Note that the uninformative prior of ϕ combined with the benchmark conditional prior $\pi^*_\phi(\alpha)$ implied from (21) will not coincide with the benchmark prior (18).

We conclude this section by noting that it is straightforward to include the intercept and lags in the static simultaneous equation model we considered in this section. Consider a 2-variable SVAR with lags $L \geq 1$.

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

where $A_0 = \begin{bmatrix} -\beta & 1 \\ -\alpha & 1 \end{bmatrix}$ and D is as defined above. The reduced form VAR is

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

where $b = A_0^{-1}c$ and $B_l = A_0^{-1}A_l$. The reduced form parameters are $\phi = (\Omega, B)$, $B = (b, B_1, \ldots, B_L)$, and the full vector of structural parameters are $\tilde{\theta} = (\alpha, \beta, d_1, d_2, A)$, $A = (c, A_1, \ldots, A_L)$. The mapping between $(\alpha, \beta, d_1, d_2, A)$ and (α, ϕ) consists of those shown in (20) and

$$A = A_0(\alpha, \phi) B \equiv A(\alpha, \phi), \tag{23}$$

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

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

where $\pi_{\tilde{\theta}}(\alpha, \beta, d_1, d_2, A)$ is the benchmark prior of $\tilde{\theta}$ and $J(\alpha, \phi)$ is the Jacobian of the transformation (20) and (23). With this modification for the conditional benchmark prior, Algorithm 6.1 can be applied to solve the multiplier minimax problem for α.

6.2 Game Theoretic Model

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

$$\begin{aligned} \phi_{11} &= G(\beta_1 - \gamma_1)G(\beta_2 - \gamma_2), \\ \phi_{00} &= (1 - G(\beta_1))(1 - G(\beta_2)), \\ \phi_{10} &= G(\beta_1)[1 - G(\beta_2)] + G(\beta_1 - \gamma_1)[G(\beta_2) - G(\beta_2 - \gamma_2)] \\ &\quad + \psi[G(\beta_1) - G(\beta_1 - \gamma_1)][G(\beta_2) - G(\beta_2 - \gamma_2)]. \end{aligned} \tag{24}$$

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

As a benchmark prior $\pi_{\tilde{\theta}}(\beta_1, \gamma_1, \beta_2, \gamma_2, \psi)$, consider for example Priors 1 and 2 in Moon and Schorfheide (2012). Posterior draws of $\tilde{\theta}$ can be obtained by the Metropolis-Hastings Algorithm or its variant. Plug them into (24) the yields the posterior draws of ϕ. The transformation (24) offers the following one-to-one reparametrization mapping between $\tilde{\theta}$ and $(\beta_1, \gamma_1, \phi)$:

$$\begin{aligned} \beta_1 &= \beta_1, \\ \gamma_1 &= \gamma_1, \\ \beta_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) \equiv \beta_2(\beta_1, \phi), \\ \gamma_2 &= G^{-1}\left(1 - \frac{\phi_{00}}{1 - G(\beta_1)}\right) - G^{-1}\left(\frac{\phi_{11}}{G(\beta_1 - \gamma_1)}\right) \equiv \gamma_2(\beta_1, \gamma_1, \phi), \\ \psi &= \frac{[1 - G(\beta_1)][\phi_{10} + \phi_{11} - G(\beta_1 - \gamma_1)] + [G(\beta_1) - G(\beta_1 - \gamma_1)]\phi_{00}}{[G(\beta_1) - G(\beta_1 - \gamma_1)]\left[1 - G(\beta_1) - \phi_{00} - \frac{1 - G(\beta_1)}{G(\beta_1 - \gamma_1)}\phi_{11}\right]} \equiv \psi(\beta_1, \gamma_1, \phi). \end{aligned} \tag{25}$$

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 (25). 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

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

7 Concluding Remarks

Appendix

A Proofs

Proof of Theorem 3.1. Let ϕ and $\delta = \delta(x)$ be fixed. We first consider the case where $\pi^*_{\theta|\phi}$ is a discrete probability mass measure with m support points $(\theta_1, \ldots, \theta_m)$ in $IS_\theta(\phi)$. Since the KL-distance $\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi})$ is positive infinity unless $\pi_{\theta|\phi}$ is absolutely continuous with respect to $\pi^*_{\theta|\phi}$, we can restrict our search of the optimal $\pi_{\theta|\phi}$ to those whose support points are constrained to $(\theta_1, \ldots, \theta_m)$. Accordingly, let us denote a discrete $\pi_{\theta|\phi}$ and the discrete loss by

$$g_i \equiv \pi_{\theta|\phi}(\theta_i), \quad f_i \equiv \pi^*_{\theta|\phi}(\theta_i), \quad h_i = h(\delta, y(\theta_i, \phi)), \quad \text{for } i = 1, \ldots, m. \tag{26}$$

Then, the inner maximization problem of (14) can be written as

$$\max_{g_1, \cdots, g_m} \sum_{i=1}^m h_i g_i - \kappa \sum_{i=1}^m g_i \ln\left(\frac{g_i}{f_i}\right), \tag{27}$$

$$\text{s.t.} \quad \sum_{i=1}^m g_i = 1.$$

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

$$\begin{aligned} & h_i + \kappa \ln f_i - \kappa - \kappa \ln g_i - \zeta = 0 \\ \iff \quad & g_i = \frac{f_i \exp(h_i/\kappa)}{\exp(1+\zeta/\kappa)}. \end{aligned} \tag{28}$$

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

$$g_i^* = \frac{f_i \exp(h_i/\kappa)}{\sum_{j=1}^{m} f_j \exp(h_j/\kappa)}. \tag{29}$$

Plugging this back into the objective function, we obtain

$$\begin{aligned} & \kappa \sum_{i=1}^{m} \left[\frac{f_i \exp(h_i/\kappa)}{\sum_{j=1}^{n} f_i \exp(h_i/\kappa)} \ln \left(\sum_{j=1}^{m} f_j \exp(h_j/\kappa) \right) \right] \\ = \quad & \kappa \ln \left(\sum_{j=1}^{m} f_j \exp(h_j/\kappa) \right), \end{aligned} \tag{30}$$

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

We generalize the claim to arbitrary $\pi^*_{\theta|\phi}$. Based on an analogy to the optimal g_i obtained in (29), we guess that $\pi^0_{\theta|\phi} \in \Pi^\infty \left(\pi^*_{\theta|\phi} \right)$ maximizing $\left\{ \int_{IS_\theta(\phi)} h(\delta(x), y(\theta,\phi)) d\pi_{\theta|\phi} - \kappa \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \right\}$ satisfies

$$d\pi^0_{\theta|\phi}(\theta) = \frac{\exp(h(\delta, y(\theta,\phi))/\kappa)}{\int_{IS_\theta(\phi)} \exp(h(\delta, y(\theta,\phi))/\kappa) d\pi^*_{\theta|\phi}} \cdot d\pi^*_{\theta|\phi}(\theta) \tag{31}$$

for θ-a.e. Since $\exp(h(\delta, y(\theta,\phi))/\kappa) \in (0,\infty)$ for all $\theta \in IS_\theta(\phi)$ by assumption, (31) implies that $\pi^*_{\theta|\phi}$ is absolutely continuous with respect to $\pi^0_{\theta|\phi}$, and hence, any $\pi_{\theta|\phi}$ with $\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) < \infty$ is absolutely continuous with respect to $\pi^0_{\theta|\phi}$. Therefore, the objective function of the inner maximization can be rewritten as

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

Plugging in (31) leads to

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

Since $\mathcal{R}(\pi_{\theta|\phi} \| \pi^0_{\theta|\phi}) \geq 0$ for any $\pi_{\theta|\phi} \in \Pi^\infty \left(\pi^*_{\theta|\phi} \right)$ and equal to zero if and only if $\pi_{\theta|\phi} = \pi^0_{\theta|\phi}$ holds for almost every θ, $\pi^0_{\theta|\phi}$ defined in (31) solves the inner maximization problem, leading to

$$\max_{\pi_{\theta|\phi}\in\Pi^{\infty}\left(\pi^*_{\theta|\phi}\right)}\left\{\int_{IS_\theta(\phi)} h(\delta(x), y(\theta,\phi))d\pi_{\theta|\phi} - \kappa d(\pi_\phi(\theta), \pi^*_\phi(\theta))\right\} = \kappa \ln\left(\int e^{h(\delta(x),y(\theta,\phi))/\kappa} d\pi^*_{\theta|\phi}\right).$$

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

Proof of Theorem 3.3. (i) Fix $\delta \in \mathcal{D}$ and consider the finite sample objective function $\int_\Phi r_\kappa(\delta,\phi)\, d\pi_{\phi|X}$. Assumption 3.2 (ii) - (iv) imply that $r_\kappa(\delta,\phi)$ is bounded and continuous in ϕ. Hence, combined with the weak convergence of $\pi_{\phi|X}$ to the mass measure at $\phi = \phi_0$ implied by the assumption of posterior consistency for ϕ, $\int_\Phi r_\kappa(\delta,\phi)\, d\pi_{\phi|X} \to r_\kappa(\delta,\phi_0)$ as $n \to \infty$ for almost every sampling sequence. With Assumption 3.2 (ii) and (v), the conclusion follows if we can demonstrate that convergence $\int_\Phi r_\kappa(\delta,\phi)\, d\pi_{\phi|X} \to r_\kappa(\delta,\phi_0)$ is uniform over $\delta \in \mathcal{D}$. Since

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

we consider bounding $\sup_{\delta\in\mathcal{D}} |r_\kappa(\delta,\phi) - r_\kappa(\delta,\phi_0)|$ by a quantity converging to zero. By invoking Assumption 3.2 (iv) and noting that $r_\kappa(\delta,\phi)$ can be written as $\ln \int_{\mathcal{Y}} \exp\{h(\delta,y)/\kappa\}\, d\pi^*_{y|\phi}$, the following inequalities hold:

$$\begin{aligned}
|r_\kappa(\delta,\phi) - r_\kappa(\delta,\phi_0)| &\leq \kappa \sup_{\phi\in\Phi}\left|\frac{\int_{\mathcal{Y}} \exp\{h(\delta,y)/\kappa\} \frac{\partial}{\partial\phi'}\left(\frac{d\pi^*_{y|\phi}}{dy}\right) dy}{\int_{\mathcal{Y}} \exp\{h(\delta,y)/\kappa\}\, d\pi^*_{y|\phi}}\right| \|\phi - \phi_0\| \qquad (32) \\
&\leq \kappa M \sup_{\phi\in\Phi}\left|\int_{\mathcal{Y}} \exp\{h(\delta,y)/\kappa\}\, dy\right| \|\phi - \phi_0\| \\
&\leq \kappa M \exp\left\{\bar{h}/\kappa\right\} diam(\mathcal{Y})\, \|\phi - \phi_0\|,
\end{aligned}$$

where $\bar{h}$ is the upper bound of $h(\delta,y)$ on $(\delta,y) \in \mathcal{D}\times\mathcal{Y}$ and $diam(\mathcal{Y})$ is the diameter of the parameter space of y, which are both finite by Assumption 3.2 (ii). Hence,

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

for some constant $C < \infty$. By compactness of Φ and the posterior consistency of ϕ, $\int_\Phi \|\phi - \phi_0\|\, d\pi_{\phi|X} \to 0$ as $n \to \infty$ for almost every sampling sequence. This completes the proof of claim (i).

(ii) When $\hat{\phi} \to_p \phi_0$, the continuous mapping theorem implies $\left|r_\kappa\left(\delta,\hat{\phi}\right) - r_\kappa(\delta,\phi_0)\right| \to_p 0$ as $n \to \infty$ pointwise in δ. Hence, by applying the consistency theorem of the M-estimator (Theorem 2.1 in Newey and McFadden (1994)), the claim follows if we can extend this convergence to the uniform convergence in δ in probability. Following (32), this is indeed true, $\sup_{\delta\in\mathcal{D}}\left|r_\kappa\left(\delta,\hat{\phi}\right) - r_\kappa(\delta,\phi_0)\right| \leq C\left\|\hat{\phi} - \phi_0\right\| \to_p 0$ as $n \to \infty$. ■

Proof of Theorem 4.2. [6](i) Fix δ and let $h(\delta, y) = (\delta - y)^2$. We partition the parameter space Φ by

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

We write the objective function of Proposition 3 as

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

and we aim to derive the lower and upper bounds of each of the two terms that are shown to converge to the same limit. Note that for each $\phi \in \Phi_\delta^-$, $r_\kappa(\delta, \phi)$ can be bounded from below by

$$\begin{aligned} r_\kappa(\delta, \phi) &= \kappa \log \left\{ \exp\left(\frac{(\delta - y_*(\phi))^2}{\kappa} \right) \int_{IS_y(\phi)} \exp\left(-\frac{(2\delta - y_*(\phi) - y)(y - y_*(\phi))}{\kappa} \right) d\pi^*_{y|\phi} \right\} \\ &\geq \kappa \log \left\{ \exp\left(\frac{(\delta - y_*(\phi))^2}{\kappa} \right) \int_{y_*(\phi)}^{y_*(\phi)+\eta} \exp\left(-\frac{(2\delta - y_*(\phi) - y)(y - y_*(\phi))}{\kappa} \right) d\pi^*_{y|\phi} \right\} \\ &\geq (\delta - y_*(\phi))^2 + \kappa \log \int_{y_*(\phi)}^{y_*(\phi)+\eta} \exp\left(-\frac{c(\phi)(y - y_*(\phi))}{\kappa} \right) d\pi^*_{y|\phi}, \end{aligned}$$

where $c(\phi) \equiv 2(\delta - y_*(\phi)) \geq y^*(\phi) - y_*(\phi) > 0$ by Assumption 4.1 (ii) and $\phi \in \Phi_\delta^-$. Using Assumption 4.1 (iii), plugging in the polynomial lower bound for the density of $\pi^*_{y|\phi}$ on $y \in [y_*(\phi), y_*(\phi) + \eta)$ leads to

$$\begin{aligned} r_\kappa(\delta, \phi) &\geq (\delta - y_*(\phi))^2 + \kappa \log \left(b \int_{y_*(\phi)}^{y_*(\phi)+\eta} (y - y_*(\phi))^\alpha \exp\left(-\frac{c(\phi)(y - y_*(\phi))}{\kappa} \right) dy \right) \\ &= (\delta - y_*(\phi))^2 + \kappa \log \left(b\kappa^{\alpha+1} \int_0^{\eta/\kappa} z^\alpha \exp(-c(\phi) z)\, dz \right) \end{aligned} \tag{33}$$

Since $\lim_{\kappa \to 0} \int_0^{\eta/\kappa} z^\alpha \exp(-c(\phi) z)\, dy < \infty$ and $\lim_{\kappa \to 0} \kappa \log \kappa = 0$, we obtain

$$\liminf_{\kappa \to 0} r_\kappa(\delta, \phi) \geq (\delta - y_*(\phi))^2.$$

For the upper bound of $r_\kappa(\delta, \phi)$ on $\phi \in \Phi_\delta^-$, we have

$$\begin{aligned} r_\kappa(\delta, \phi) &\leq (\delta - y_*(\phi))^2 + \kappa \log \int_{IS_y(\phi)} \exp(0)\, d\pi^*_{y|\phi} \qquad (34) \\ &= (\delta - y_*(\phi))^2 \end{aligned}$$

[6] The proof given here is based on the proof of the Laplace's integral approximation method shown in Theorem 1 in Chapter II of Wong (1989).

for all κ, where we use $\exp\left(-\frac{(2\delta - y_*(\phi) - y)(y - y_*(\phi))}{\kappa}\right) \leq \exp(0)$ for all $y \in [y_*(\phi), y^*(\phi)]$ when $\phi \in \Phi_\delta^-$. It then holds

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

Hence, $\lim_{\kappa \to 0} r_\kappa(\delta, \phi) = (\delta - y_*(\phi))^2$ for $\phi \in \Phi_\delta^-$ pointwise.

Bounds for $r_\kappa(\delta, \phi)$ on $\phi \in \Phi_\delta^+$ follow similarly. For a lower bound, we have

$$\begin{aligned} r_\kappa(\delta, \phi) &\geq (\delta - y^*(\phi))^2 + \kappa \log \int_{y^*(\phi)-\eta}^{y^*(\phi)} \exp\left(-\frac{c(\phi)(y^*(\phi) - y)}{\kappa}\right) d\pi_{y|\phi}^* \\ &\geq (\delta - y^*(\phi))^2 + \kappa \log \left(b\kappa^{\alpha+1} \int_0^{\eta/\kappa} z^\alpha \exp(-c(\phi)z)\, dz\right) \\ &\to (\delta - y^*(\phi))^2 \quad \text{as } \kappa \to 0. \end{aligned}$$

For a upper bound, the same argument as in (34) applies to yield $r_\kappa(\delta, \phi) \leq (\delta - y^*(\phi))^2$ for all κ. Hence, $\lim_{\kappa \to 0} r_\kappa(\delta, \phi) = (\delta - y_*(\phi))^2$ for $\phi \in \Phi_\delta^+$ pointwise.

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

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

where the last line follows by noting that $(\delta - y_*(\phi))^2 \geq (\delta - y^*(\phi))^2$ for $\phi \in \Phi_\delta^-$ and vice versa for $\phi \in \Phi_\delta^+$.

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

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

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

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

For $\phi \in \Phi_\delta^-$, a lower bound of $r_\kappa(\delta, \phi)$ can be obtained as

$$
\begin{aligned}
r_\kappa(\delta,\phi) &= \kappa \log \left\{ \exp\left(\frac{(1-\tau)(\delta - y_*(\phi))}{\kappa} \right) \int_{IS_y(\phi)} \exp\left(\frac{\rho_\tau(y-\delta) - (1-\tau)(\delta - y_*(\phi))}{\kappa} \right) d\pi^*_{y|\phi} \right\} \\
&\geq (1-\tau)(\delta - y_*(\phi)) + \kappa \log \left\{ \int_{IS_y(\phi)} \exp\left(-\frac{(1-\tau)(y - y_*(\phi))}{\kappa} \right) d\pi^*_{y|\phi} \right\} \\
&\geq (1-\tau)(\delta - y_*(\phi)) + \kappa \log \left\{ b \int_{y_*(\phi)}^{y_*(\phi)+\eta} (y - y_*(\phi))^\alpha \exp\left(-\frac{(1-\tau)(y - y_*(\phi))}{\kappa} \right) dy \right\} \\
&\to (1-\tau)(\delta - y_*(\phi)) \quad \text{as } \kappa \to 0,
\end{aligned}
$$

where the second line follows by noting

$$
\begin{aligned}
& \rho_\tau(y-\delta) - (1-\tau)(\delta - y_*(\phi)) \\
=\ & [\tau y + (1-\tau) y_*(\phi) - \delta] 1\{y - \delta > 0\} \\
\geq\ & [-(1-\tau)(y - y_*(\phi)) + y - \delta] 1\{y - \delta > 0\} \\
\geq\ & -(1-\tau)(y - y_*(\phi)) 1\{y - \delta > 0\} \\
\geq\ & -(1-\tau)(y - y_*(\phi)) \quad \text{for all } y \in [y_*(\phi), y^*(\phi)],
\end{aligned}
$$

the third line follows by Assumption 4.1 (iii), and the convergence in the fourth line follows by the same reasoning as in (33). Also by noting $\rho_\tau(y-\delta) - (1-\tau)(\delta - y_*(\phi)) \leq 0$ for $\phi \in \Phi_\delta^-$, an upper bound of $r_\kappa(\delta,\phi)$ is given $(1-\tau)(\delta - y_*(\phi))$. Hence, we obtain $\lim_{\kappa\to 0} r_\kappa(\delta,\phi) = (1-\tau)(\delta - y_*(\phi))$ for $\phi \in \Phi_\delta^-$ pointwise. By the same argument (we omit the detail for brevity), it can be shown that $\lim_{\kappa\to 0} r_\kappa(\delta,\phi) = \tau(y^*(\phi) - \delta)$ for $\phi \in \Phi_\delta^+$. Hence, again by the dominated convergence theorem,

$$
\begin{aligned}
\lim_{\kappa\to 0} \int_\Phi r_\kappa(\delta,\phi) d\pi_{\phi|X} &= \int_{\Phi_\delta^-} (1-\tau)(\delta - y_*(\phi)) d\pi_{\phi|X} + \int_{\Phi_\delta^+} \tau(y^*(\phi) - \delta) d\pi_{\phi|X} \\
&= \int_\Phi [(1-\tau)(\delta - y_*(\phi)) \vee \tau(y^*(\phi) - \delta)] d\pi_{\phi|X}
\end{aligned}
$$

follows. This completes the proof. ■

Proof of Theorem 4.3. (i) Let $R_n(\delta) \equiv \lim_{\kappa\to 0} \int_\Phi r_\kappa(\delta,\phi) d\pi_{\phi|X}$, which, by Lemma 4.2 (i), is equal to $R_n(\delta) = \int_\Phi r_0(\delta,\phi) d\pi_{\phi|X}$ where $r_0(\delta,\phi) = (\delta - y_*(\phi))^2 \vee (\delta - y^*(\phi))^2$. Since the parameter space for y and the domain of δ are compact, $r_0(\delta,\phi)$ is a bounded function in ϕ. In addition, $y_*(\phi)$ and $y^*(\phi)$ 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

$$
\begin{aligned}
R_n(\delta) \to R_\infty(\delta) &\equiv \lim_{n\to\infty} \int_\Phi \left[(\delta - y_*(\phi))^2 \vee (\delta - y^*(\phi))^2 \right] d\pi_{\phi|X} \qquad (35) \\
&= (\delta - y_*(\phi_0))^2 \vee (\delta - y^*(\phi_0))^2
\end{aligned}
$$

pointwise in δ for almost every sampling sequence. Note that $R_\infty(\delta)$ is minimized uniquely at $\delta = \frac{1}{2}(y_*(\phi_0) + y^*(\phi_0))$. Hence, by an analogy to the argument of the convergence of M-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 $I(\phi) \equiv [y_*(\phi), y^*(\phi)]$ and note that $(\delta - y_*(\phi))^2 \vee (\delta - y^*(\phi))^2$ can be interpreted as the squared Hausdorff metric $[d_H(\delta, I(\phi))]^2$ between point $\{\delta\}$ and interval $I(\phi)$. Then

$$
\begin{aligned}
|R_n(\delta) - R_\infty(\delta)| &= \left| \int_\Phi \left([d_H(\delta, I(\phi))]^2 - [d_H(\delta, I(\phi_0))]^2 \right) d\pi_{\phi|X} \right| \qquad (36) \\
&\leq 2diam(\mathcal{Y}) \int_\Phi |d_H(\delta, I(\phi)) - d_H(\delta, I(\phi_0))| \, d\pi_{\phi|X} \\
&\leq 2diam(\mathcal{Y}) \int_\Phi d_H(I(\phi), I(\phi_0)) \, d\pi_{\phi|X},
\end{aligned}
$$

where $diam(\mathcal{Y}) < \infty$ is the diameter of the parameter space of y and the third line follows by the triangular inequality of a metric, $|d_H(\delta, I(\phi)) - d_H(\delta, I(\phi_0))| \leq d_H(I(\phi), I(\phi_0))$. Since $d_H(I(\phi), I(\phi_0))$ is bounded by the compactness assumption of the y space and is continuous at $\phi = \phi_0$ by 4.1 (iv), $\int_\Phi d_H(I(\phi), I(\phi_0)) \, d\pi_{\phi|X} \to 0$ as $\pi_{\phi|X}$ converges weakly to the point mass measure at $\phi = \phi_0$. This implies the uniform convergence of $R_n(\delta)$, $\sup_\delta |R_n(\delta) - R_\infty(\delta)| \to 0$ as $n \to \infty$.

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

$$
R_n(\delta) \to R_\infty(\delta) \equiv (1-\tau)(\delta - y_*(\phi_0)) \vee \tau(y^*(\phi_0) - \delta) = l(\delta, \phi_0), \qquad (37)
$$

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

$$
\begin{aligned}
\Phi_0^- &\equiv \{\phi \in \Phi : (1-\tau)y_*(\phi) + \tau y^*(\phi) \leq (1-\tau)y_*(\phi_0) + \tau y^*(\phi_0)\}, \qquad (38) \\
\Phi_0^+ &\equiv \{\phi \in \Phi : (1-\tau)y_*(\phi) + \tau y^*(\phi) > (1-\tau)y_*(\phi_0) + \tau y^*(\phi_0)\}.
\end{aligned}
$$

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

$$
l(\delta, \phi) - l(\delta, \phi_0) \qquad (39)
$$

$$
= \begin{cases} (1-\tau)[y_*(\phi_0) - y_*(\phi)], & \text{if } \delta \leq (1-\tau)y_*(\phi) + \tau y^*(\phi), \\ \tau[y^*(\phi) - y_*(\phi_0)] - [\delta - y_*(\phi_0)], & \text{if } (1-\tau)y_*(\phi) + \tau y^*(\phi) < \delta \leq (1-\tau)y_*(\phi_0) + \tau y^*(\phi_0), \\ \tau[y^*(\phi) - y^*(\phi_0)] & \text{if } \delta > (1-\tau)y_*(\phi_0) + \tau y^*(\phi_0). \end{cases} \qquad (40)
$$

By noting that in the second case in (40), 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 $|y_*(\phi) - y_*(\phi_0)| + |y^*(\phi) - y^*(\phi_0)|$. 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

$$\begin{aligned}\sup_{\delta}|R_n(\delta) - R_\infty(\delta)| &\leq \sup_{\delta}\int_{\Phi}|l(\delta,\phi) - l(\delta,\phi_0)|\, d\pi_{\phi|X} \\ &\leq \int_{\Phi}|y_*(\phi) - y_*(\phi_0)|\, d\pi_{\phi|X} + \int_{\Phi}|y^*(\phi) - y^*(\phi_0)|\, d\pi_{\phi|X},\end{aligned} \tag{41}$$

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

A.1 Asymptotic Analysis with Discrete Benchmark Prior

If the loss function $h(\delta, y)$ is differentiable with respect to δ at almost every y, the first order condition for the minimization problem (17) is obtained as

$$\int_{\Phi}\left[\int_{IS_y(\phi)}\frac{\partial}{\partial\delta}h(\delta,y)\left(\frac{\exp\{h(\delta,y)/\kappa\}}{\int_{IS_y(\phi)}\exp\{h(\delta,y)/\kappa\}\, d\pi^*_{y|\phi}}\right)d\pi^*_{y|\phi}\right]d\pi_{\phi|X} = 0. \tag{42}$$

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

$$\pi^*_\phi(y) = \sum_{m=1}^{M} w_m 1_{y_m(\phi)}(y), \quad w_m > 0 \quad \forall m, \quad \sum_{m=1}^{M} w_m = 1,$$

where the weights $(w_1, \ldots, w_M)$ specify benchmark credibility over each point-identified model. The set of conditional priors concerned in (14) consists of any mixture of these point mass measures,

$$\Pi^\infty_\phi\left(\pi^*_\phi\right) = \left\{\sum_{m=1}^{M} w'_m 1_{y_m(\phi)}(y) : \left(w'_1, \ldots, w'_M\right) \in \Delta_M\right\},$$

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

Denote $(y_1(\phi_0), \ldots, y_M(\phi_0))$ by $(y_1, \ldots, y_M)$ for short, and label the models according to $y_1 \leq y_2 \leq \cdots \leq y_M$. With a fixed $\kappa > 0$ and the degenerate posterior for ϕ, the first order condition (42) is simplified to

$$\frac{\sum_{m=1}^{M}(\delta - y_m) w_m \exp\left(\frac{(\delta - y_m)^2}{\kappa}\right)}{\sum_{m=1}^{M} w_m \exp\left(\frac{(\delta - y_m)^2}{\kappa}\right)} = 0, \tag{43}$$

where the denominator does not affect the solution. The next proposition shows that, as $\kappa \to 0$, optimal δ solving this first order condition converges to the mid-point of the two extreme point-identified models, $(y_1 + y_M)/2$.

Proposition A.1 *As $\kappa \to 0$, optimal δ that solves (43) converges to $(y_M + y_1)/2$.*[7]

Proof. (sketch) Rewrite the first-order condition (43) as

$$\sum_{m=1}^{M} (\delta - y_m) \exp\left(\frac{(\delta - y_m)^2}{\kappa} + \ln w_m \right) = 0. \tag{44}$$

As $\kappa \to 0$, the exponential term will shoot out to positive infinity, so in order for the first order condition to be solved for some δ at small κ, it must be the case that one exponential term diverge to negative infinity and another exponential term diverges to positive infinity at the same rate as $\kappa \to 0$. Along this reasoning, consider equalizing the exponential terms for the two extreme point-identified models, $m = 1$ and $m = M$,

$$\exp\left(\frac{(\delta - y_1)^2}{\kappa} + \ln w_1 \right) = \exp\left(\frac{(\delta - y_M)^2}{\kappa} + \ln w_M \right),$$

which gives

$$\delta^* = \frac{y_1 + y_M}{2} - \frac{\kappa}{2(y_M - y_1)} \ln\left(\frac{w_1}{w_M} \right).$$

Let $H = \frac{(\delta^* - y_1)^2}{\kappa} + \ln w_1 = \frac{(\delta^* - y_M)^2}{\kappa} + \ln w_M$. It can be shown that, for $m = 2, \ldots, (M-1)$,

$$\begin{aligned} & \frac{(\delta^* - y_m)^2}{\kappa} + \ln w_m - H \\ = \; & -\frac{(y_M - y_m)(y_m - y_1)}{\kappa} - \left(\frac{y_M - y_m}{y_M - y_1} \right) \ln\left(\frac{w_1}{w_M} \right) + \ln\left(\frac{w_m}{w_M} \right), \end{aligned}$$

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

We now show that δ^* constructed above satisfies the first order condition at the limit $\kappa \to 0$. Plug in δ^* into the left-hand side of (44) and divide it by $\exp(H)$,

$$(\delta^* - y_1) + (\delta^* - y_M) + \sum_{m=2}^{M-1} (\delta^* - y_m) \exp\left(\frac{(\delta^* - y_m)^2}{\kappa} + \ln w_m - H \right).$$

As $\kappa \to 0$, all the exponential terms in the summation converge to zero, and $(\delta^* - y_1) + (\delta^* - y_M) \to 0$ since $\delta^* \to \frac{y_1 + y_M}{2}$. That is, the optimal decision at the limit is given by the

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

mid-point decision $\frac{y_1+y_M}{2}$, and optimal δ that solves (43) should converge to $(y_1 + y_M)/2$. ■

For the check loss, we have the following results. Since $\rho_\tau(y-\delta)$ is differentiable in δ except for the kink point, we can still rely on the first order condition (42), as far as the solution exists (if a solution does not exist, it would probably imply that an optimum is occurring at a nondifferentiable point). When $\widetilde{\pi}(\phi)$ is a probability mass, we can ignore the denominator term in (42), so that the first order condition is simplified to

$$\int_{IS_{\phi_0}(y)} \left[(1-\tau)\exp\left\{ \frac{-(1-\tau)(y-\delta)}{\kappa} \right\} 1\{y<\delta\} - \tau \exp\left\{ \frac{\tau(y-\delta)}{\kappa} \right\} 1\{y>\delta\} \right] d\pi^*_{\phi_0}(y) = 0. \tag{45}$$

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

$$\begin{aligned} \delta &= \kappa \ln \left(\frac{\tau \int_\delta^\infty \exp\left(\frac{\tau y}{\kappa}\right) d\pi^*_{\phi_0}(y)}{(1-\tau)\int_{-\infty}^\delta \exp\left(\frac{-(1-\tau)y}{\kappa}\right) d\pi^*_{\phi_0}(y)} \right) \\ &\equiv f_\tau(\delta). \end{aligned} \tag{46}$$

If $\pi^*_{\phi_0}(y)$ does not involve any probability mass, then $f_\tau(\delta)$ is a continuous and weakly decreasing function in δ. Furthermore, $f_\tau(\delta)$ diverges to ∞ as δ approaches to the lower bound of $IS_y(\phi_0)$ and diverges to $-\infty$ as δ approaches to the upper bound of $IS_{\phi_0}(y)$. Therefore, the equation (46) has a unique solution for δ for every τ. We hereafter denote a unique root of $\delta = f_\tau(\delta)$ by $\delta^*(\tau)$ (if it exists).

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

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

Proof. Set $\pi^*_{\phi_0}(y) = (\underline{y} - \bar{y})^{-1} 1_{[\underline{y},\bar{y}]}(y)$ in (46), and solve for δ yields the result. ■

Proposition A.3 *Suppose that the benchmark prior is given by a mixture of two point masses at y_1 and y_2 with $y_1 < y_2$,*

$$\pi_{\phi_0}(y) = w1_{y_1}(y) + (1-w)1_{y_2}(y).$$

Then

$$\delta^*(\tau) = \max\left\{ y_1, \min\left\{ \begin{array}{c} (1-\tau)y_1 + \tau y_2 + \kappa \ln\left(\frac{(1-w)\tau}{w(1-\tau)}\right), \\ y_2 \end{array} \right\} \right\}.$$

This implies, as $\kappa \to 0$, $\delta^*(\tau) \to (1-\tau)y_1 + \tau y_2$.

Proof. For $\delta \in [y_1, y_2]$, the first order condition (45) can be written as

$$w(1-\tau)\exp\left\{\frac{(1-\tau)(\delta - y_1)}{\kappa}\right\} - (1-w)\tau\exp\left\{\frac{\tau(y_2-\delta)}{\kappa}\right\} = 0.$$

This solves for

$$\delta^*(\tau) = (1-\tau)\,y_1 + \tau y_2 + \kappa\ln\left(\frac{(1-w)\tau}{w(1-\tau)}\right).$$

Hence, if $\delta^*(\tau) \in [y_1, y_2]$, $\delta^*(\tau)$ is the optimum, and otherwise, it can be shown that either $\delta = y_1$ or $\delta = y_2$ becomes an optimum. Hence, the conclusion follows. ■

Proposition A.4 *Suppose that the benchmark prior is given by a mixture of M point masses at $y_1 < y_2 < \cdots < y_M$,*

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

Then, $\delta^*(\tau) \to (1-\tau)\,y_1 + \tau y_M$, as $\kappa \to 0$.

Proof. Let $\delta^* = (1-\tau)\,y_1 + \tau y_M$, and let $m^* \in \{1, \ldots, (M-1)\}$ be the index such that $y_i \leq \delta^*$ for all $i = 1, \ldots, m^*$, and $y_i > \delta^*$ for all $i = (m^*+1), \ldots, M$. Then, if $\delta^*(\tau) \in [y_{m^*}, y_{m^*+1}]$, the first order condition (46) should hold as

$$\begin{aligned}
\delta^*(\tau) &= \kappa\ln\left(\exp\left(\frac{\delta^*}{\kappa}\right)\left(\frac{\tau}{1-\tau}\right)\frac{\sum_{m=m^*+1}^{M} w_m \exp\left(\frac{\tau(y_m-y_M)}{\kappa}\right)}{\sum_{m=1}^{m^*} w_m \exp\left(\frac{(1-\tau)(y_1-y_m)}{\kappa}\right)}\right)\\
&= \delta^* + \kappa\ln\left(\left(\frac{\tau}{1-\tau}\right)\frac{w_M + \sum_{m=m^*+1}^{M-1} w_m \exp\left(\frac{\tau(y_m-y_M)}{\kappa}\right)}{w_1 + \sum_{m=2}^{m^*} w_m \exp\left(\frac{(1-\tau)(y_1-y_m)}{\kappa}\right)}\right)\\
&\to \delta^* \quad \text{as } \kappa \to 0.
\end{aligned}$$

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

References

Baumeister, C., and J. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *unpublished manuscript.*

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

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

Giacomini, R., and T. Kitagawa (2015): "Robust Inference about Partially-identified SVARs," *cemmap working paper, University College London.*

Hansen, L. P., and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review, AEA Papers and Proceedings*, 91(2), 60–66.

Kitagawa, T. (2012): "Estimation and Inference for Set-identified Parameters Using Posterior Lower Probabilities," *unpublished manuscript.*

Kline, B., and E. Tamer (2013): "Default Bayesian Inference in a Class of Partially Identified Models," *unpublished manuscript.*

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

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

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

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. Elsevier, Amsterdam, The Netherlands.

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(3), 398–412.

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

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

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.

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