

Robust Bayesian Inference under Partially Specified Priors*

Martin Fankhauser[†]

October 14, 2026

Abstract

Set identification leaves multiple structural parameter values observationally equivalent, often generating wide identified sets and correspondingly ambiguous economic conclusions. Bayesian analysis offers a way to reduce this ambiguity by using prior information to weight these values differently, but economic reasoning may only partially identify such prior beliefs. This paper develops a robust Bayesian framework that imposes only economically meaningful restrictions on selected prior marginals while remaining agnostic about otherwise unspecified prior features. In set-identified models, the resulting posterior bounds reduce to linear optimization problems and can be approximated by reweighting simulated structural draws. Applied to oil-market SVARs, we find that demand shocks robustly dominate supply shocks, while the stronger conclusion that supply shocks are nearly irrelevant requires much tighter prior discipline on short-run elasticities.

Keywords: Prior Sensitivity, Robust Bayes, Imprecise Probabilities, Marginally-specified priors, Oil Demand and Supply

JEL Codes: C11, C14, C18, C32, C52, E52, E58

*We thank Massimiliano Marcellino, Francesca Molinari, Raffaella Giacomini, Toru Kitagawa, Pamela Giustinelli, Geert Mesters, Giorgio Primiceri, Nicola Pavoni, Mikkel Plagborg-Møller, Lutz Kilian, Daniel Lewis, Efrem Castelnuovo, Amilcar Velez, Jiaming Huang and Tommaso Tornese for helpful discussions and comments. All mistakes are my own.

[†]PhD. Bocconi University

1 Introduction

In many empirical problems, researchers are only able to set identify a parameter of interest, so that ambiguity about its value persists even asymptotically. This may arise from incomplete data, such as missing covariates, or from incomplete models, including games with multiple equilibria and sign-restricted structural models. For a recent review, see Molinari (2020). The usual object of interest is the sharp identified set, that is, the smallest set of parameter values consistent with the data and the maintained model. In many applications, however, this set remains wide, leaving substantial ambiguity about the economic question the researcher ultimately seeks to answer. A natural question is therefore whether additional, but still weak and credible, information can be used to sharpen conclusions without abandoning the partial-identification principle of imposing only assumptions that can be credibly defended; see, e.g., Manski (2003).

Bayesian analysis provides a natural way to incorporate such additional information. Parameter values that are observationally equivalent receive the same likelihood, but need not receive the same prior probability. Prior information can therefore differentially weight values within the identified set and sharpen posterior conclusions beyond what is implied by the likelihood alone. This point is discussed in different forms in Moon and Schorfheide (2012), Baumeister and Hamilton (2015), and Giacomini and Kitagawa (2021). The difficulty is that a conventional Bayesian analysis requires a fully specified prior distribution, whereas the economic reasoning motivating the prior may determine only some of its features. For example, a researcher may be willing to assert that smaller values of a parameter are weakly more plausible than larger values, without having a defensible view about the exact shape, concentration, or tail behavior of its prior distribution. In this sense, economic reasoning may only partially identify prior beliefs, just as data may only partially identify structural parameters.

This distinction becomes especially salient in models with a high-dimensional structural parameter vector θ , where a conventional Bayesian analysis requires a full joint prior over all

of its elements. We instead start from economically interpretable functionals $S(\theta)$ about which the researcher has credible prior information, and let $T(\theta)$ denote the quantity she ultimately wishes to learn about. Rather than specifying a complete joint prior for θ , we restrict the class of admissible marginal distributions of $S(\theta)$, for example through probability bounds, moment restrictions, or qualitative shape restrictions, while leaving otherwise unspecified features of the structural prior unrestricted. This allows, for instance, a hard restriction on $S(\theta)$ to be weakened by assigning only a bounded probability to its violation, with the bound indexing the credibility attached to the restriction. Alternatively, one may impose only qualitative information, such as smaller deviations from a benchmark being weakly more plausible than larger ones, without selecting a particular parametric distribution. Because $S(\theta)$ and $T(\theta)$ are functions of the same underlying structural parameter, even weak prior information about $S(\theta)$ can sharpen inference about $T(\theta)$.

This paper develops inference based on classes of priors that encode such partial prior information. In set-identified models, the resulting robustness problem is substantially simplified by the structure of the likelihood. Conditional on the reduced-form parameter, observationally equivalent structural parameter values have the same likelihood, so the posterior distribution over the unidentified dimension coincides with the corresponding prior distribution. Posterior expectations are therefore linear in the structural prior. When the admissible prior information can be expressed through affine restrictions, which is the case for many economically meaningful restrictions, robust posterior bounds reduce to infinite-dimensional linear programs. We propose a simulation-based approximation to these population problems following the distributionally robust optimization literature (see Lam, Liu, and Singham (2024)): structural parameter values are drawn from a reference measure, and alternative admissible priors are represented by probability weights on the simulated draws. The resulting bounds are obtained from finite-dimensional linear programs, and we establish conditions under which they converge to their population counterparts.

Structural vector autoregressions identified through sign and inequality restrictions provide

a natural application. In the traditional sign-restriction approach following Uhlig (2005), economic restrictions can be interpreted as hard prior information: structural models violating the restrictions receive zero probability, while models satisfying them remain admissible. Bayesian implementations typically complement these restrictions with a particular prior over the remaining admissible structural models, such as the widely used Haar prior on the relevant orthogonal matrices; see, e.g., Arias and Waggoner (2024). Baumeister and Hamilton (2015) instead emphasize that economically meaningful information need not take the form of hard restrictions and advocate placing prior distributions directly on structural parameters such as elasticities. Our framework provides an intermediate possibility: economically meaningful restrictions can be softened without requiring the researcher to specify a complete parametric distribution for the structural parameters.

We revisit the importance of supply and demand shock contributions to economic activity and the real price of oil. Baumeister and Hamilton (2019) challenged the consensus view in the literature, established in Kilian (2009), Kilian and Murphy (2012), and Kilian and Murphy (2014), that supply shocks are relatively unimportant drivers of the latter two. A central point of disagreement concerns how strongly prior information about short-run oil-market elasticities should restrict the structural model. Starting from the conventional elasticity restrictions in Kilian and Murphy (2012), we allow a controlled prior probability of violating them and, in alternative specifications, impose only qualitative shape restrictions expressing that smaller elasticities are more plausible than larger ones. We then ask which conclusions about the importance of supply and demand shocks hold uniformly over all structural priors satisfying these weak restrictions. Varying the probability of violation traces a credibility frontier: as the elasticity assumptions are progressively weakened, we can distinguish conclusions that survive substantial disbelief in the restrictions from those that require much stronger prior discipline. This provides a direct way to assess how much weakening of the conventional identifying assumptions is necessary to obtain conclusions associated with Baumeister and Hamilton (2019).

The framework also gives an alternative interpretation to existing robustness approaches in Bayesian analysis. A partially specified structural prior involves two conceptually distinct choices: which distributions of the economically meaningful statistic $S(\theta)$ are compatible with the researcher’s prior information, and how a given distribution of $S(\theta)$ is completed into a joint prior for the remaining dimensions of θ . We formalize this decomposition in Section 3. An ordinary Bayesian analysis fixes both components and therefore selects a single precise weighting of the identified set. At the opposite extreme, Giacomini and Kitagawa (2021) allow the unrevisable structural prior to vary over all distributions supported on the identified set, corresponding to leaving both components unrestricted. Our baseline-free specification lies between these cases: economically defensible restrictions discipline the admissible distribution of $S(\theta)$, while features of the joint prior not implied by this information can remain unrestricted. This also distinguishes our approach from robust Bayesian methods that define prior ambiguity through distance from a particular reference prior, such as the Kullback–Leibler neighborhoods in Giacomini, Kitagawa, and Uhlig (2019) and Ho (2023). When the researcher does attach substantive meaning to a baseline joint prior, however, our framework also admits a complementary specification in which the baseline completion is preserved while the marginal distribution of $S(\theta)$ is varied. This connects the approach to the marginally specified priors of Kessler, Hoff, and Dunson (2015): rather than replacing one baseline marginal by one prespecified alternative, we allow a class of economically admissible marginals and retain the least-distortive baseline completion for each. When the admissible marginal class is itself defined through Kullback–Leibler distance from the baseline marginal, the resulting structural priors lie within the corresponding structural Kullback–Leibler neighborhood while restricting variation to changes in the distribution of $S(\theta)$.

The remainder of the paper is organized as follows. Section 2 illustrates the main idea in a simple supply–demand example and shows how partially specified prior information can sharpen inference without requiring a complete prior distribution. Section 3 develops

the general framework, characterizes the resulting robust posterior bounds, establishes a simulation-based finite-dimensional approximation, and discusses the relation to alternative specifications of prior ambiguity. Section 4 applies the framework to the oil-market SVAR and studies which conclusions about the relative importance of supply and demand shocks remain robust to progressively weaker prior information about short-run elasticities. Section 5 concludes.

2 Illustrative Example

Before outlining the general theory, let us illustrate the main contribution of the paper in a simple supply–demand example. Consider a static market with price p_t and quantity q_t . Supply and demand are

$$\begin{aligned} q_t &= \alpha_s p_t + u_t^s, \\ q_t &= \alpha_d p_t + u_t^d \end{aligned}$$

where α_s is the short-run supply elasticity and α_d is the short-run demand slope. The true values are given by $\alpha_s = 0.25$ and $\alpha_d = -0.50$. The shocks u_t^s and u_t^d shift supply and demand, respectively. The econometrician observes equilibrium prices and quantities. These data can reveal the joint distribution of (p_t, q_t) , but they do not in general point-identify the structural slopes (α_s, α_d) without further assumptions. Suppose we are willing to assume $\alpha_d < 0$ and $\alpha_s > 0$, so demand is downward sloping and supply is upward sloping.

For simplicity, assume the variables are demeaned and let

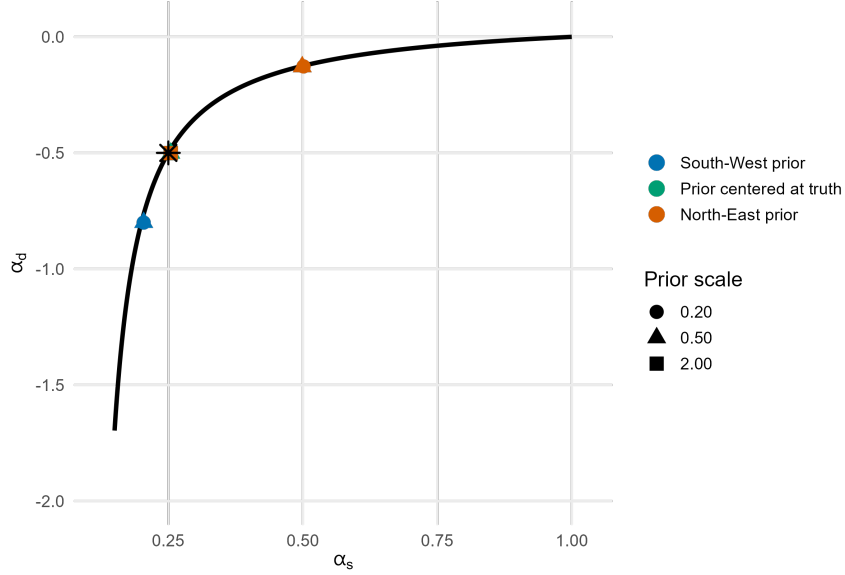
$$\Omega = \begin{pmatrix} \omega_{pp} & \omega_{pq} \\ \omega_{pq} & \omega_{qq} \end{pmatrix}$$

denote the population covariance matrix of equilibrium prices and quantities. Given (α_s, α_d) , the structural shocks are $u_t^s = q_t - \alpha_s p_t$ and $u_t^d = q_t - \alpha_d p_t$. Assume that the primitive supply and demand shocks are orthogonal, $\text{Cov}(u_t^s, u_t^d) = 0$. This is the static analogue of assuming a diagonal covariance matrix for structural shocks in SVARs. Orthogonality implies $0 = \omega_{qq} - (\alpha_s + \alpha_d)\omega_{pq} + \alpha_s \alpha_d \omega_{pp}$. Equivalently, for any admissible value of α_s ,

$$\alpha_d(\alpha_s) = \frac{\alpha_s \omega_{pq} - \omega_{qq}}{\alpha_s \omega_{pp} - \omega_{pq}}.$$

This curve is the identified set for the two elasticities: every point on it generates the same reduced-form covariance matrix Ω and is therefore observationally equivalent. Conditional on Ω , the likelihood does not choose among these points. Any posterior allocation of mass along the curve must therefore come from prior information. The pure sign restriction $\alpha_s > 0$ leaves a region in which α_s approaches ω_{pq}/ω_{pp} from above and the implied demand elasticity diverges. To keep the illustration focused on prior sensitivity inside a bounded identified set, we impose the maintained lower bound $\alpha_s > 0.15$ throughout. Without this tightening, the same mechanism would be present but visually and numerically more extreme. The identified curve is displayed in Figure 1 as the black solid line.

Figure 1: Identified Set for Elasticities



Suppose now that the researcher has additional, but deliberately soft, prior information about the elasticities. Suppose everyone agrees on the statements:

"Demand is unlikely to be more elastic than -0.5 "

and

"Supply is unlikely to be more elastic than 0.75 "

Following the spirit of Baumeister and Hamilton (2019), this information can be encoded by replacing additional hard restrictions with a precise prior distribution over the elasticities. Consider as a baseline a BH-style specification with independent truncated Student- t priors, $\alpha_s \sim t_3(0.50, 0.20^2)$ truncated to $[0.15, \infty)$ and $\alpha_d \sim t_3(-0.125, 0.20^2)$ truncated to $(-\infty, 0)$, where the truncation points correspond to the maintained sign restrictions imposed throughout the example. This is the north-east (NE) prior shown in Figure 1. It provides a natural precise encoding of the two qualitative statements: it assigns only $\Pr(\alpha_s > 0.75) \simeq 0.165$ and $\Pr(\alpha_d < -0.50) \simeq 0.111$ to their respective violations. The difficulty is that doing so necessarily specifies much more than these two economically meaningful probabilities. In

particular, it fixes a complete probability distribution over the identified set and thereby favors a particular observationally equivalent structural configuration. With a sample of 5,000 observations, the likelihood is highly informative about the reduced form but essentially uninformative for discriminating among points along the identified set. Consequently, as Figure 1 illustrates, the posterior mode remains close to the NE configuration selected by the prior rather than being pulled toward the true elasticities. This is not a peculiarity of the NE prior: the south-west and truth-centered specifications display the same persistence toward their respective prior locations. Only the prior centered at the truth yields a posterior mode at the truth, but this is tautological: we learn the truth only when we happen to already know it.

A natural response is to conduct conventional prior sensitivity analysis by making the prior increasingly diffuse, as for instance considered in Baumeister and Hamilton (2019). In a point-identified model, this would typically reduce the influence of a badly located prior and allow the likelihood to dominate. Here, however, increasing the common prior scale from 0.20 to 0.50 and 2.00 leaves the posterior mode close to the same observationally equivalent configuration, as shown by the nearly overlapping symbols for each prior location in Figure 1. More importantly, this exercise does not hold fixed the economic information that motivated the prior in the first place. For the NE specification, the violation probabilities increase from $(0.165, 0.111)$ at scale 0.20, to $(0.444, 0.430)$ at scale 0.50, and to $(0.806, 0.825)$ at scale 2.00 for $(\alpha_s > 0.75, \alpha_d < -0.50)$, respectively. Thus, making the prior less precise simultaneously changes the substantive beliefs being imposed: a specification motivated by the view that large supply elasticities and very elastic demand are unlikely eventually assigns most of its probability to precisely those events. Conventional scale sensitivity therefore confounds two conceptually distinct questions: how strongly the researcher wishes to commit to a particular precise prior location and which economic statements the prior is meant to encode.

This suggests a different form of sensitivity analysis: hold fixed the economically meaningful restrictions on these tail probabilities while allowing the otherwise unspecified features of the

prior to vary. In particular, consider encoding our beliefs in probability statements:

$$P(\alpha_d \geq -0.50) \geq 1 - \kappa \quad \text{and} \quad P(\alpha_s \leq 0.75) \geq 1 - \kappa$$

for different values of κ , i.e. different measures of subjective uncertainty about our prior beliefs.

We propose to bound the posterior quantities of interest over the class of all priors satisfying these probability statements. Importantly, this class retains only the economic content of the original prior information. For example, $\kappa = 0.10$ requires the researcher to assign at least 90% prior probability to demand being no more elastic than -0.50 and at least 90% prior probability to supply being no more elastic than 0.75 , but leaves the remaining shape and allocation of prior probability unrestricted. Varying κ therefore changes the confidence attached to the economic statements themselves rather than perturbing an otherwise arbitrary parametric prior.

The following table reports the resulting bounds on the posterior means of the two elasticities for different values of κ .

Table 1: Credal bounds for Elasticities

Restriction	κ	$E[\alpha_s Y]$	Width α_s	$E[\alpha_d Y]$	Width α_d
Full ambiguity	–	[0.150, 1.004]	100%	[-1.802, -0.000]	100%
Both tail restrictions	0.05	[0.249, 0.762]	60%	[-0.564, -0.038]	29%
Both tail restrictions	0.10	[0.244, 0.775]	62%	[-0.629, -0.036]	33%
Both tail restrictions	0.20	[0.233, 0.800]	66%	[-0.760, -0.032]	40%
Both tail restrictions	0.30	[0.223, 0.826]	71%	[-0.890, -0.028]	48%

Notes: The reported width is the width of the corresponding credal interval relative to the interval obtained under full ambiguity.

Table 1 shows that these weak probability statements can nevertheless be informative. Under full ambiguity, corresponding to the absence of any additional restriction on the prior, the posterior mean of the supply elasticity can range from 0.150 to 1.004, while that of the demand elasticity can range from -1.802 to approximately zero. Imposing only the two tail-probability restrictions substantially contracts these sets. For example, at $\kappa = 0.10$, the

admissible posterior means are $[0.244, 0.775]$ for supply and $[-0.629, -0.036]$ for demand, corresponding to only 62% and 33%, respectively, of their full-ambiguity widths.

The comparison also illustrates that weak prior information need not be equally informative about every object. In this example, the same two probability statements discipline the posterior mean of the demand elasticity considerably more strongly than that of the supply elasticity: across the values of κ considered, the demand bounds retain only 29–48% of their full-ambiguity width, compared with 60–71% for supply. Thus, rather than forcing a precise answer through an arbitrary completion of the prior, the credal analysis reveals which conclusions are actually supported by the combination of the data and the economic beliefs imposed. As κ increases and the statements are weakened, the bounds expand smoothly toward the full-ambiguity benchmark.

The conclusion of this example is not that inference under a single precise prior is inherently inappropriate in set-identified models. If a researcher can defend the full specification of a precise parametric prior, then using that prior is entirely coherent. Our point is that such knowledge is often unavailable. Researchers may instead be willing to defend only weaker statements about the distribution of structural parameters. For example, the belief that smaller values are more plausible than larger values can be represented by the class of weakly decreasing priors; absent further information, we argue that inference should reflect the entire class rather than an arbitrary selection of it. Likewise, as in the preceding example, a researcher may use prior information only to express distrust in a maintained economic restriction, without having a defensible view about the full shape of the prior. In such cases, the likelihood does not resolve uncertainty along the unidentified dimension, so posterior conclusions remain sensitive to how the prior completes the information that the researcher is actually willing to defend. Our proposal is therefore to represent that information directly through classes of distributions and to report the range of posterior conclusions consistent with them.

The preceding example is deliberately simple in another respect: the objects about which

prior information is imposed are themselves also the objects of interest. More generally, these need not coincide. Let $S(\theta)$ denote a feature of the structural parameter about which the researcher has credible prior information, while $T(\theta)$ denotes the posterior quantity ultimately of interest. Economic knowledge may then restrict the admissible distribution of $S(\theta)$, which in turn restricts the admissible distributions over the underlying structural parameter θ and thereby the posterior distribution of $T(\theta)$. Schematically, the information flows as

$$S(\theta) \implies \theta \implies T(\theta).$$

Thus, the purpose of the restrictions need not be to place direct prior bounds on the target itself which will often shrink ambiguity in a tautological way. They can instead encode credible knowledge about one feature of the model and ask what that knowledge implies for a different economic quantity of interest. The next section formalizes this construction.

3 Theory

We first set up the population posterior sensitivity problem under a class of well-defined prior restrictions. Thereafter, a finite Monte-Carlo approximation to the infinite dimensional linear program is introduced with associated convergence results. As our sensitivity analysis is with respect to the conditional unrevisable part of the prior the next subsection discusses implications for the joint reduced form and structural prior. We conclude with a discussion and relation to other proposals within the literature.

3.1 Set-up and Prior Decomposition

Let $\phi \in \Phi$ denote the reduced-form parameter and let $\theta \in \Theta$ denote the structural parameter, where $(\Phi, \mathcal{A}_\phi)$ and $(\Theta, \mathcal{A}_\theta)$ are standard Borel spaces. The two objects are linked by a measurable many-to-one function $g : \Theta \rightarrow \Phi$, $\phi = g(\theta)$, which defines the incompleteness of the structural model and thus the set-identification problem. We assume that the likelihood

depends only on ϕ .

Assumption 1. *Denote the observed data by Y and the likelihood function by $p(Y|\cdot)$. Then $p(Y|\phi, \theta) = p(Y|\phi)$.*

Suppose we have a set of assumptions which we are willing to assume with certainty, denoted by $\mathcal{A}_0$. These could be normalization assumptions as, e.g., in standard SVAR analysis (see for instance the unit effect normalization in Rubio-Ramirez, Waggoner, and Zha (2010)) or substantive but generally agreed upon assumptions (e.g. impact sign restrictions in the monetary policy example of Uhlig (2005)). Define the identified set as

$$\text{IS}_\theta(\phi) \equiv \{\theta \in \Theta : g(\theta) = \phi, \quad \mathcal{A}_0 \text{ holds}\}.$$

In the absence of further identifying assumptions, this is usually the object of interest for identification analysis in frequentist studies; for reviews see Molinari (2020) and Canay and Shaikh (2017) for inference.

For an ordinary Bayesian approach we need to specify a prior on both parameters which, under standard conditions, can be decomposed as

$$\pi(d\phi, d\theta) = \pi_{\theta|\phi}(d\theta|\phi)\pi_\phi(d\phi).$$

As in Giacomini, Kitagawa, and Uhlig (2019) and Giacomini and Kitagawa (2021), we focus on sensitivity with respect to the unrevisable component of the prior $\pi_{\theta|\phi}$, as this can never be updated by the data; see e.g. Giacomini and Kitagawa (2021). In general, one could also conduct sensitivity analysis with respect to the reduced-form prior as in Ho (2023).

As Moon and Schorfheide (2012) show, specifying a single precise prior for the structural parameter yields a posterior contracting within the identified set. Thus, from a frequentist point of view, a single-prior Bayesian will be too confident. However, from a Bayesian perspective, as long as we can defend the prior used there is no conceptual problem with

using a single prior. The issue is more that, from a substantive point of view, being able to specify an entire joint conditional prior for the structural parameter vector confidently seems implausible. For instance, modern SVAR applications have hundreds of structural parameters. Do we really claim to be able to specify a precise joint law for all of them? Even if so, are we comfortable with all implications such a prior carries for functionals of these parameters?

This paper starts from a different standpoint. Suppose we have knowledge about some implication a full joint conditional prior $\pi_{\theta|\phi}$ should have and leave everything else unspecified. Let $S : \Theta \rightarrow \mathbb{R}^{d_s}$ be some transformation of the full parameter vector. Let $\mathcal{M}_S(\phi)$ be the class of admissible distributions of S induced by $\pi_{\theta|\phi}$, i.e. $\mu = S_{\#}\pi_{\theta|\phi}$.

Many joint distributions $\pi_{\theta|\phi}$ can be compatible with a given marginal $\mu \in \mathcal{M}_S(\phi)$. Thus, even if $\mathcal{M}_S(\phi)$ is a singleton, a marginal prior must still be completed into a joint structural prior. To formalize this, for fixed ϕ define the feasible fiber associated with s as

$$F_{\phi}(s) \equiv \{\theta \in \text{IS}_{\theta}(\phi) : S(\theta) = s\}.$$

Since the parameter spaces are standard Borel, a joint structural prior can be disintegrated as

$$\pi_{\theta|\phi}(A) = \int \rho(A|s)\mu(ds), \quad A \in \mathcal{A}_{\theta},$$

where $\rho(\cdot|s)$ is a conditional distribution over the structural parameter given $S = s$. Compatibility with the identified set requires $\rho(F_{\phi}(s)|s) = 1$ for μ -almost every s . We refer to ρ as a completion of the marginal prior μ , and let $\mathcal{R}(\phi)$ denote the class of admissible completion kernels.

Example 1. Consider a two-dimensional structural parameter $\theta = (\theta_1, \theta_2)$ and suppose that, for a given ϕ , the identified set is

$$\text{IS}_{\theta}(\phi) = \{(\theta_1, \theta_2) : 0 \leq \theta_1 \leq 1, 0 \leq \theta_2 \leq 1 - \theta_1\}.$$

Suppose our prior knowledge concerns only $S(\theta) = \theta_1$, and is precise: $\mathcal{M}_S(\phi) = \{\mu^0\}$ with $\mu^0 = U[0, 1]$. Thus there is no ambiguity about the marginal prior for θ_1 . Nevertheless, this does not determine a joint prior for (θ_1, θ_2) .

For example, all of the following completion rules are compatible with the same marginal $\theta_1 \sim U[0, 1]$ and with the identified set:

$$\rho_1(d\theta_2|s) = \delta_0(d\theta_2),$$

$$\rho_2(d\theta_2|s) = U[0, 1 - s](d\theta_2),$$

$$\rho_3(d\theta_2|s) = \delta_{1-s}(d\theta_2).$$

They therefore generate three different joint structural priors despite agreeing exactly on the marginal distribution of S .

The distinction is consequential for other structural quantities. For instance, let $T(\theta) = \theta_1 + \theta_2$. Under the three completions above,

$$\mathbb{E}_{\pi_1}[T] = \frac{1}{2}, \quad \mathbb{E}_{\pi_2}[T] = \frac{3}{4}, \quad \mathbb{E}_{\pi_3}[T] = 1,$$

respectively.

Hence even a singleton class $\mathcal{M}_S(\phi)$ does not specify a unique structural prior, nor a unique prior for a target of interest. A completion specifies how prior mass is distributed over the remaining dimensions of the identified set conditional on the value of S .

Together, the class of admissible marginals $\mathcal{M}_S(\phi)$ and the completion class $\mathcal{R}(\phi)$ define the set of admissible joint priors for θ conditional on ϕ . For $\mu \in \mathcal{M}_S(\phi)$ and $\rho \in \mathcal{R}(\phi)$, define

$$\pi_{\mu, \rho}(A) \equiv \int \rho(A|s)\mu(ds), \quad A \in \mathcal{A}_\theta.$$

The corresponding class of admissible structural priors is

$$\Pi_{\mathcal{M},\mathcal{R}}(\phi) \equiv \{\pi_{\mu,\rho} : \mu \in \mathcal{M}_S(\phi), \rho \in \mathcal{R}(\phi)\}.$$

The framework nests several existing approaches as special cases.

Example 2. *If both the marginal and the completion are fixed,*

$$\mathcal{M}_S(\phi) = \{\mu^0\}, \quad \mathcal{R}(\phi) = \{\rho^0\},$$

then $\Pi_{\mathcal{M},\mathcal{R}}(\phi)$ is a singleton and we recover an ordinary Bayesian analysis with a fully specified conditional structural prior. The commonly used Haar prior in SVAR analysis is an example of such a construction: it specifies a unique joint prior and therefore implicitly fixes both the marginal distribution of any transformation S and its completion. See, for example, Arias and Waggoner (2024).

Example 3. *At the opposite extreme, Giacomini and Kitagawa (2021) leave the unrevisable conditional structural prior unrestricted over the identified set. In our notation, this corresponds to allowing both the marginal and the completion to vary freely:*

$$\mathcal{M}_S(\phi) = \mathcal{P}(S(\text{IS}_\theta(\phi))), \quad \mathcal{R}(\phi) = \mathcal{R}^{\text{full}}(\phi) \equiv \{\rho : \rho(F_\phi(s)|s) = 1 \text{ for all feasible } s\}.$$

Consequently, $\Pi_{\mathcal{M},\mathcal{R}}(\phi) = \mathcal{P}(\text{IS}_\theta(\phi))$.

Example 4. *For point-identified models Kessler, Hoff, and Dunson (2015) propose a marginally-specified prior construction. Starting from a baseline joint prior π^0 , they replace its marginal distribution for a particular functional S by a desired marginal μ^0 , while retaining the baseline conditional distribution given S . The resulting joint prior is the KL-closest prior to π^0 among all priors having marginal μ^0 . In our notation, this amounts to*

fixing

$$\mathcal{M}_S(\phi) = \{\mu^0\}, \quad \mathcal{R}(\phi) = \{\rho^0\},$$

where ρ^0 is the completion inherited from the baseline prior.

The last example naturally also connects our approach to sensitivity analysis with respect to some baseline prior as in Giacomini, Kitagawa, and Uhlig (2019) or Ho (2023). We will discuss this further below.

3.2 Population Problem

For some posterior target of interest $T : \Theta \rightarrow \mathbb{R}$ a robust Bayes approach given the class $\Pi(\phi)$ aims to find bounds on posterior expectations:

$$\underline{m}_T(\phi) \equiv \inf_{\pi \in \Pi(\phi)} \mathbb{E}_\pi[T \mid Y, \phi] \quad \bar{m}_T(\phi) \equiv \sup_{\pi \in \Pi(\phi)} \mathbb{E}_\pi[T \mid Y, \phi]. \quad (1)$$

This is the inner optimization for a decision problem under ambiguity as defined in Giacomini, Kitagawa, and Uhlig (2019), which establishes for this case, only the marginal distribution of $T(\theta)$ is relevant for the problem. Nevertheless, as both $S(\theta)$ and $T(\theta)$ depend on the common parameter vector θ , restricting the distribution of $S(\theta)$ can restrict the outcome marginal in a nontrivial way even if $\mathcal{R}(\phi)$ is left unrestricted.

Generically, finding these bounds is a complex task as first $\mathbb{E}_\pi[T \mid Y, \phi]$ is a non-linear function of the prior class elements π . Second, the elements themselves are functions and thus infinite dimensional objects. For the first issue, the defining set-identification property formalized by Assumption 1, which causes the underlying problem with precise priors, also massively

simplifies the robust problem. Under Assumption 1:

$$\begin{aligned}\mathbb{E}_\pi[T \mid Y, \phi] &= \frac{\int T(\theta) p(Y \mid \phi, \theta) \pi_{\theta|\phi}(d\theta)}{\int p(Y \mid \phi, \theta) \pi_{\theta|\phi}(d\theta)} \\ &= \frac{\int T(\theta) p(Y \mid \phi) \pi_{\theta|\phi}(d\theta)}{\int p(Y \mid \phi) \pi_{\theta|\phi}(d\theta)} \\ &= \int T(\theta) \pi_{\theta|\phi}(d\theta).\end{aligned}$$

or equivalently $\pi_{\theta|Y, \phi} = \pi_{\theta|\phi}$. In words, conditional on ϕ , posterior expectations with respect to the unrevisable structural prior coincide with prior expectations. Hence the objective in (1) is linear in $\pi_{\theta|\phi}$. If $\Pi(\phi)$ is defined by restrictions linear in $\pi_{\theta|\phi}$, (1) becomes an infinite dimensional linear program.

Proposition 1. *Fix ϕ and suppose $\Pi_{\mathcal{M}, \mathcal{R}}(\phi) \neq \emptyset$. Let $T : \Theta \rightarrow \mathbb{R}$ be bounded and measurable. Then*

$$\underline{m}_T(\phi) = \inf_{\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)} \int_{\Theta} T(\theta) \pi(d\theta) \quad \bar{m}_T(\phi) = \sup_{\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)} \int_{\Theta} T(\theta) \pi(d\theta)$$

If in addition $\Pi_{\mathcal{M}, \mathcal{R}}(\phi)$ admits an affine representation, i.e. for a potential infinite index set J , measurable functions a_j and constants b_j

$$\Pi_{\mathcal{M}, \mathcal{R}}(\phi) = \left\{ \pi \in \mathcal{P}(\text{IS}(\phi)) : \int a_j(\theta) \pi(d\theta) \leq b_j \quad \forall j \in J \right\}$$

then $\underline{m}_T(\phi)$ and $\bar{m}_T(\phi)$ are solutions to infinite dimensional linear programs over π .

Proposition 1 is useful because many forms of partial prior knowledge that arise naturally in applications admit an affine representation. We give several examples.

Example 5 (Moment restrictions). *Suppose prior information concerns moments of the economically meaningful transformation $S(\theta)$. For measurable functions $h_j : \mathbb{R}^{d_s} \rightarrow \mathbb{R}$,*

restrictions of the form

$$\underline{m}_j \leq \int h_j(S(\theta)) \pi(d\theta) \leq \overline{m}_j, \quad j = 1, \dots, J,$$

are affine in the structural prior π . Thus, although h_j may be nonlinear in the structural parameter, moment restrictions on $S(\theta)$ directly satisfy the affine representation in Proposition 1.

Example 6 (Probability restrictions). Suppose prior knowledge restricts the probability assigned to some measurable event $B \subseteq \mathbb{R}^{d_s}$. Since

$$P_\pi(S(\theta) \in B) = \int \mathbf{1}\{S(\theta) \in B\} \pi(d\theta),$$

restrictions of the form

$$\underline{p} \leq P_\pi(S(\theta) \in B) \leq \overline{p}$$

are affine in π . This includes, for example, bounds on the prior probability that an elasticity or another structural quantity lies above or below a specified threshold.

Example 7 (Shape restrictions on a marginal distribution). Affine representations can also accommodate nonparametric shape restrictions. Suppose a scalar transformation satisfies $S(\theta) \in [0, B]$ for all $\theta \in \text{IS}_\theta(\phi)$, and prior knowledge requires smaller values of S to be weakly more plausible than larger ones. Let $\mathcal{M}_S(\phi)$ consist of all distributions of the form

$$\mu(ds) = w_0 \delta_0(ds) + f(s) ds,$$

where $w_0 \in [0, 1]$ and $f : (0, B] \rightarrow [0, \infty)$ is weakly decreasing. The atom at zero makes this class weakly closed; it is the weak closure of the class of distributions on $[0, B]$ with a weakly

decreasing density. For integers $n \geq 0$ and $k \geq 1$, let

$$q_{n,k}(x) = \frac{d}{dx} \{x^{n+1}(1-x)^k\}, \quad x \in [0, 1].$$

Lemma 6 in Appendix B shows that $S_{\#}\pi \in \mathcal{M}_S(\phi)$ if and only if

$$\int q_{n,k}(S(\theta)/B) \pi(d\theta) \geq 0 \quad \text{for all } n \geq 0, k \geq 1.$$

With an unrestricted completion, $\Pi_{\mathcal{M},\mathcal{R}}(\phi)$ consists of all $\pi \in \mathcal{P}(\text{IS}_{\theta}(\phi))$ with $S_{\#}\pi \in \mathcal{M}_S(\phi)$, so it admits the affine representation of Proposition 1 with countably many restrictions. If S is continuous on the identified set, each restriction has a continuous integrand, as required for the finite approximation below. Monotonicity imposed only above a threshold admits an analogous representation; see Appendix B.

3.3 Finite Approximation

We approximate the infinite-dimensional problem by optimizing over probability measures supported on simulated structural parameters. The probability weights on these points are free decision variables. Under the conditions below, the resulting bounds converge to their population counterparts as the simulated grid grows.

Fix ϕ and write $K_{\phi} \equiv \text{IS}_{\theta}(\phi)$. For the approximation result, suppose the population class $\Pi(\phi) \equiv \Pi_{\mathcal{M},\mathcal{R}}(\phi)$ admits the representation

$$\Pi(\phi) = \left\{ \pi \in \mathcal{P}(K_{\phi}) : \int_{K_{\phi}} a_j(\theta) \pi(d\theta) \leq b_j \quad \forall j \in J, \quad \pi(G_{\ell}) \leq \kappa_{\ell} \quad \forall \ell \in L \right\}.$$

The first group consists of continuous restrictions. It can include moment restrictions and the shape restrictions characterized in Appendix B. The second group bounds the probabilities of specified events. Separating the two groups is useful because the simulation approximates continuous restrictions and event probabilities in different ways.

When J is infinite, the finite program imposes only the rows in a finite set $J_r \subseteq J$. All event restrictions are imposed at every stage. The corresponding population stage class is

$$\Pi^{(r)}(\phi) = \left\{ \pi \in \mathcal{P}(K_\phi) : \int_{K_\phi} a_j(\theta) \pi(d\theta) \leq b_j \quad \forall j \in J_r, \quad \pi(G_\ell) \leq \kappa_\ell \quad \forall \ell \in L \right\}.$$

The stage classes must become progressively more restrictive and jointly recover the population class. The row sets J_r themselves need not be nested.

Assumption 2. (K_ϕ, d) is a nonempty compact metric space.

Assumption 3. J is countable and L is finite. Each a_j belongs to $C(K_\phi)$ and $b_j \in \mathbb{R}$. Each G_ℓ is relatively open in K_ϕ , and $\kappa_\ell \in [0, 1]$.

Assumption 4. For every r , $J_r \subseteq J$ is finite. Moreover, $\Pi^{(r+1)}(\phi) \subseteq \Pi^{(r)}(\phi)$ and $\bigcap_{r \geq 1} \Pi^{(r)}(\phi) = \Pi(\phi)$.

The approximation theorem covers continuous targets $T \in C(K_\phi)$ and indicator targets $T = \mathbf{1}_A$ for relatively closed $A \subseteq K_\phi$. To describe where the simulated points must lie, form the finite partition $\mathcal{C}$ from the nonempty intersections of each G_ℓ or its complement in K_ϕ . For an indicator target, include A or its complement in these intersections as well. We call the resulting sets *cells*. Membership in every restriction event, and in A when applicable, is constant within a cell.

Fix a simulation measure $\nu \in \mathcal{P}(K_\phi)$ and let $\theta^{(1)}, \theta^{(2)}, \dots$ be draws from ν . Write $\Theta_N = \{\theta^{(1)}, \dots, \theta^{(N)}\}$. Every probability measure supported on this grid can be written as $\sum_{n=1}^N w_n \delta_{\theta^{(n)}}$ for some $w \in \Delta_N$. The simulation measure determines which structural points are available; the weights are chosen by the optimization. Its choice can affect the finite-sample bounds and the speed of convergence, even when it does not affect the limiting bounds.

Assumption 5. The draws $\theta^{(1)}, \theta^{(2)}, \dots$ are iid with law ν . For every cell $C \in \mathcal{C}$, every $x \in C$, and every $\epsilon > 0$,

$$\nu(B(x, \epsilon) \cap C) > 0.$$

Define the sampled stage class by $\Pi_{N,r}(\phi) \equiv \Pi^{(r)}(\phi) \cap \mathcal{P}(\Theta_N)$. The finite approximations to the population bounds in (1) are

$$\underline{m}_{T,N,r}(\phi) \equiv \inf_{\pi \in \Pi_{N,r}(\phi)} \mathbb{E}_\pi[T \mid Y, \phi], \quad \bar{m}_{T,N,r}(\phi) \equiv \sup_{\pi \in \Pi_{N,r}(\phi)} \mathbb{E}_\pi[T \mid Y, \phi]. \quad (2)$$

By Proposition 1, these objectives equal $\int_{K_\phi} T d\pi$. Consequently, both problems are finite linear programs in the weights w_n . The sampled measures satisfy the stage- r restrictions exactly; when J is infinite, they need not satisfy rows omitted from J_r .

The next condition ensures that small errors in continuous restrictions caused by approximating a population prior on the simulated grid can be repaired.

Assumption 6. *For every r , there exists $\pi_r^\circ \in \Pi^{(r)}(\phi)$ such that*

$$\eta_r \equiv \min_{j \in J_r} \left(b_j - \int_{K_\phi} a_j(\theta) \pi_r^\circ(d\theta) \right) > 0,$$

with $\eta_r = +\infty$ if $J_r = \emptyset$.

The following condition concerns only the lower bound for an indicator target when the stage classes do not eventually equal $\Pi(\phi)$.

Assumption 7. *For a relatively closed target event $A \subseteq K_\phi$,*

$$\inf_{\pi \in \Pi(\phi)} \pi(\text{int}_{K_\phi} A) = \inf_{\pi \in \Pi(\phi)} \pi(A).$$

Theorem 1. *Fix ϕ and suppose Assumptions 1- 6 hold. Let $T \in C(K_\phi)$, or let $T = \mathbf{1}_A$ for a relatively closed $A \subseteq K_\phi$. In the indicator case, suppose in addition that either Assumption 7 holds or $\Pi^{(r)}(\phi) = \Pi(\phi)$ for all sufficiently large r . Then there exists a deterministic nondecreasing sequence $r(N) \rightarrow \infty$ such that, almost surely,*

$$\underline{m}_{T,N,r(N)}(\phi) \longrightarrow \underline{m}_T(\phi), \quad \bar{m}_{T,N,r(N)}(\phi) \longrightarrow \bar{m}_T(\phi).$$

The theorem establishes the existence of a suitable schedule $r(N)$; it does not assert convergence for every schedule. If the population class has finitely many continuous rows, a single stage containing all of them suffices. The proof is given in Appendix A, and the representation of the shape restrictions is given in Appendix B.

When there are only event restrictions, the result is stronger for an indicator target. In that case no approximation of locations within a cell is needed: the restrictions and objective depend only on the total probability assigned to each cell.

Corollary 1. *Suppose $J = \emptyset$, $\Pi(\phi) \neq \emptyset$, and $T = \mathbf{1}_A$ for a relatively closed $A \subseteq K_\phi$. Suppose the draws are iid from ν and $\nu(C) > 0$ for every cell $C \in \mathcal{C}$. Define*

$$\mathcal{D} = \left\{ p \in \Delta(\mathcal{C}) : \sum_{C \subseteq G_\ell} p_C \leq \kappa_\ell \quad \forall \ell \in L \right\},$$

where $\Delta(\mathcal{C})$ is the simplex of cell masses. Then

$$\underline{m}_T(\phi) = \min_{p \in \mathcal{D}} \sum_{C \subseteq A} p_C, \quad \bar{m}_T(\phi) = \max_{p \in \mathcal{D}} \sum_{C \subseteq A} p_C.$$

Almost surely, both sampled bounds equal their population bounds for every sufficiently large N . Equality holds as soon as every nonempty cell contains a simulated point.

Thus, with three event restrictions and one indicator target, the tail-only population problem has at most $2^4 = 16$ cell masses. The corollary requires only positive simulation probability for each cell, a weaker condition than Assumption 5.

3.4 Implied Joint Prior

The previous results are stated conditional on a fixed reduced-form parameter ϕ . They also induce a corresponding class of joint priors and posteriors. Keeping the reduced-form

prior π_ϕ fixed, define

$$\Pi_{\mathcal{M},\mathcal{R}}^J \equiv \left\{ \pi(d\phi, d\theta) = \pi_\phi(d\phi) \pi_{\theta|\phi}(d\theta|\phi) : \pi_{\theta|\phi}(\cdot|\phi) \in \Pi_{\mathcal{M},\mathcal{R}}(\phi) \quad \pi_\phi\text{-a.s.} \right\},$$

where $\phi \mapsto \pi_{\theta|\phi}(A|\phi)$ is measurable for every $A \in \mathcal{A}_\theta$. Equivalently, for almost every ϕ ,

$$\pi_{\theta|\phi}(A|\phi) = \int \rho_\phi(A|s) \mu_\phi(ds), \quad \mu_\phi \in \mathcal{M}_S(\phi), \quad \rho_\phi \in \mathcal{R}(\phi).$$

All elements of $\Pi_{\mathcal{M},\mathcal{R}}^J$ therefore share the same reduced-form prior, but may differ in both their marginal beliefs about $S(\theta)$ and in how these beliefs are completed into a joint structural prior. In particular, the implied unconditional distributions of θ and S are

$$\begin{aligned} \pi_\theta(d\theta) &= \int_{\Phi} \pi_{\theta|\phi}(d\theta|\phi) \pi_\phi(d\phi), \\ \pi_S(ds) &= \int_{\Phi} \mu_\phi(ds) \pi_\phi(d\phi). \end{aligned}$$

The joint class is nonempty only if $\Pi_{\mathcal{M},\mathcal{R}}(\phi) \neq \emptyset$ for π_ϕ -almost every ϕ . This is a substantive requirement rather than a technicality. A nonempty identified set does not guarantee a nonempty prior class, because the restrictions defining $\mathcal{M}_S(\phi)$ must be satisfiable by distributions supported on $S(\text{IS}_\theta(\phi))$. For instance, with $\kappa = 0$ the probability restrictions in Section 4 admit no prior if no element of the identified set satisfies all elasticity bounds simultaneously. Similarly, every distribution in the monotone class of Example 7 assigns positive probability to every neighborhood of zero, so this class admits no prior if S is bounded away from zero on the identified set. Whether $\Pi_{\mathcal{M},\mathcal{R}}(\phi)$ is empty depends on ϕ and on the prior restrictions, but not on the data. Following Giacomini and Kitagawa (2021), who trim the reduced-form prior to reduced-form parameters with a nonempty identified set, we restrict π_ϕ to the (assumed measurable) set $\{\phi : \Pi_{\mathcal{M},\mathcal{R}}(\phi) \neq \emptyset\}$ and assume henceforth that $\Pi_{\mathcal{M},\mathcal{R}}(\phi) \neq \emptyset$ for π_ϕ -almost every ϕ . In the application, similar to their report of the posterior probability of a nonempty identified set, we report the number of posterior draws

of ϕ for which the prior class is empty.

Under Assumption 1, observing Y updates only the reduced-form marginal. Hence every admissible joint posterior takes the form

$$\pi(d\phi, d\theta|Y) = \pi_\phi(d\phi|Y)\pi_{\theta|\phi}(d\theta|\phi),$$

where $\pi_\phi(\cdot|Y)$ is common across all elements of the prior class and $\pi_{\theta|Y,\phi} = \pi_{\theta|\phi}$.

The joint class is rectangular across reduced-form values: apart from measurability, an admissible element of $\Pi_{\mathcal{M},\mathcal{R}}(\phi)$ can be selected separately for each ϕ . Hence, under standard measurable-selection conditions,

$$\underline{m}_T(Y) = \int_{\Phi} \underline{m}_T(\phi) \pi_\phi(d\phi|Y), \quad \overline{m}_T(Y) = \int_{\Phi} \overline{m}_T(\phi) \pi_\phi(d\phi|Y).$$

Thus the conditional optimization problems studied above aggregate directly over the posterior distribution of the reduced form.

Although these posterior bounds aggregate pointwise in ϕ , it is also useful to understand the unconditional implications of restrictions imposed on the conditional marginal distribution of S . The following result shows that two restrictions used below are preserved under mixing over the reduced form. Importantly, these conclusions depend only on the marginal class $\mathcal{M}_S(\phi)$ and therefore hold for any admissible completion class $\mathcal{R}(\phi)$.

Proposition 2. *Let $\pi \in \Pi_{\mathcal{M},\mathcal{R}}^J$ and denote $\mu_\phi = S_{\#}\pi_{\theta|\phi}$.*

1. *Suppose that for some measurable set $B \subseteq \mathbb{R}$ and $\kappa \in [0, 1]$, $\mu_\phi(B) \leq \kappa$ for π_ϕ -almost every ϕ . Then $\pi(S(\theta) \in B) \leq \kappa$.*
2. *Fix $c \in \mathbb{R}$ and let $Z = |S - c|$. Suppose that, for π_ϕ -almost every ϕ , the conditional distribution of Z is $w(\phi)\delta_0(dz) + f_{Z|\phi}(z) dz$, where $w(\phi) \in [0, 1]$ and $f_{Z|\phi}$ is supported on $[0, \bar{z}(\phi)]$ and non-increasing on $(0, \bar{z}(\phi)]$. Then the unconditional distribution of Z is*

$\bar{w} \delta_0(dz) + f_Z(z) dz$ with $\bar{w} = \int_{\Phi} w(\phi) \pi_{\phi}(d\phi)$ and

$$f_Z(z) = \int_{\Phi} f_{Z|\phi}(z) \pi_{\phi}(d\phi),$$

where $f_{Z|\phi}(z) = 0$ for $z > \bar{z}(\phi)$ and each $f_{Z|\phi}$ is taken right-continuous on $(0, \infty)$, and f_Z is non-increasing on $(0, \infty)$.

The same conclusions hold for the corresponding posterior distributions after replacing π_{ϕ} by $\pi_{\phi}(\cdot|Y)$.

In general, conditional restriction will not survive the aggregation, however, the type of restrictions in the preceding result are important, as they nest for instance the motivating example of modeling partial trust in some assumption as well as monotone violation from fixed bounds if the assumption consists of inequality constraints.

Sometimes conditional restrictions are more natural than unconditional ones. Consider for instance a missing data problem for which we observe the distribution of a target variable for some share of the population and not for the other one. Conditional on the realized distribution we might want to change our prior beliefs about the missing distribution.

3.5 Sensitivity with Partial Trust in Reference Measure

In the baseline-free specification above, the reference measure used for simulation serves only a computational role and, subject to the conditions of Theorem 1, does not affect the population robust bounds. This differs from approaches to Bayesian robustness that measure departures from a particular baseline prior, as in the KL-based sensitivity analysis of Giacomini, Kitagawa, and Uhlig (2019). At the same time, it is related to the marginally specified prior construction of Kessler, Hoff, and Dunson (2015), which modifies a specified marginal distribution while preserving the conditional structure of a baseline prior.

In some applications, however, the researcher may attach substantive meaning to a baseline conditional prior $\pi_{\theta|\phi}^0$ without being willing to accept all of its marginal implications for

S . The substantive baseline need not coincide with the measure used for simulation. Let $\pi_{S|\phi}^0 \equiv S_{\#}\pi_{\theta|\phi}^0$ and denote the baseline completion by ρ_{ϕ}^0 , so that

$$\pi_{\theta|\phi}^0(A) = \int \rho_{\phi}^0(A | s) \pi_{S|\phi}^0(ds), \quad A \in \mathcal{A}_{\theta}.$$

We consider the specialization $\mathcal{R}(\phi) = \{\rho_{\phi}^0\}$, while allowing $\mu \in \mathcal{M}_S(\phi)$. Thus ambiguity remains about the marginal distribution of S , but its completion is inherited from the baseline. Under absolute continuity, this is equivalent to reweighting the baseline by a function of S alone.

Proposition 3. *Fix $\phi \in \Phi$ and let $\pi_{\theta|\phi}^0$ be a baseline conditional prior supported on K_{ϕ} , with induced marginal $\pi_{S|\phi}^0$ and completion ρ_{ϕ}^0 . Suppose*

$$\emptyset \neq \mathcal{M}_S(\phi) \subseteq \{\mu \in \mathcal{P}(\mathbb{R}^{d_s}) : \mu \ll \pi_{S|\phi}^0\},$$

and set $\mathcal{R}(\phi) = \{\rho_{\phi}^0\}$. For every $\mu \in \mathcal{M}_S(\phi)$, the unique prior $\pi_{\theta|\phi}^{\mu} \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)$ with $S_{\#}\pi_{\theta|\phi}^{\mu} = \mu$ satisfies

$$\begin{aligned} \pi_{\theta|\phi}^{\mu}(A) &= \int \rho_{\phi}^0(A | s) \mu(ds), \quad A \in \mathcal{A}_{\theta}, \\ \pi_{\theta|\phi}^{\mu}(d\theta) &= \frac{d\mu}{d\pi_{S|\phi}^0}(S(\theta)) \pi_{\theta|\phi}^0(d\theta). \end{aligned}$$

Moreover,

$$\text{KL} \left(\pi_{\theta|\phi}^{\mu} \parallel \pi_{\theta|\phi}^0 \right) = \text{KL} \left(\mu \parallel \pi_{S|\phi}^0 \right).$$

If this divergence is finite, $\pi_{\theta|\phi}^{\mu}$ is the unique solution to

$$\min_{\substack{q_{\theta|\phi} \ll \pi_{\theta|\phi}^0 \\ S_{\#}q_{\theta|\phi} = \mu}} \text{KL} \left(q_{\theta|\phi} \parallel \pi_{\theta|\phi}^0 \right).$$

Absolute continuity ensures that the construction does not depend on the version of ρ_ϕ^0 chosen on $\pi_{S|\phi}^0$ -null sets. This is the domination condition imposed by Kessler, Hoff, and Dunson (2015). If $\mathcal{M}_S(\phi) = \{\mu^*\}$, Proposition 3 reduces to their marginally specified prior construction. Here, an entire class $\mathcal{M}_S(\phi)$ generates a corresponding class of structural priors, each retaining the same baseline completion. The KL chain rule explains the minimizing property: once the marginal is fixed, changing the completion adds a nonnegative conditional divergence.

Proposition 3 also clarifies the connection to Giacomini, Kitagawa, and Uhlig (2019). Let

$$\bar{\lambda}(\phi) = \sup_{\mu \in \mathcal{M}_S(\phi)} \text{KL}(\mu \parallel \pi_{S|\phi}^0).$$

If $\bar{\lambda}(\phi) < \infty$, every prior in $\Pi_{\mathcal{M}, \{\rho_\phi^0\}}(\phi)$ belongs to the structural KL ball of radius $\bar{\lambda}(\phi)$ around $\pi_{\theta|\phi}^0$. The converse need not hold: a structural KL ball also permits changes to the completion, whereas our construction fixes it. If $\mathcal{M}_S(\phi)$ is itself a KL ball of finite radius λ around $\pi_{S|\phi}^0$, then for bounded losses depending on θ only through S , the worst-case expected loss coincides with that over the structural KL ball of radius λ . This is the marginal-reduction result of Lemma 2.1 in Giacomini, Kitagawa, and Uhlig (2019). For targets T not determined by S , the two robust analyses need not coincide, because fixing the completion then restricts decision-relevant prior variation.

3.6 Full Ambiguity versus Partially Specified Priors

The distinction between our approach and the full ambiguity robust approach of Giacomini and Kitagawa (2021) can be best understood within the simple example proposed by Moon and Schorfheide (2012). This also provides an excellent abstract motivation for the current paper. Let $y_i = \phi + \varepsilon_i$ with $\varepsilon_i \sim N(0, 1)$ so the data identify ϕ . Suppose we know with certainty that the structural parameter θ satisfies $\phi \leq \theta \leq \phi + \lambda$ for known $\lambda > 0$. The sharp identified set, which is the main object of interest from a frequentist perspective, is

$$\text{IS}_\theta(\phi) = [\phi, \phi + \lambda].$$

A single-prior Bayesian specifies one conditional prior $\pi_{\theta|\phi}$ instead. Asymptotically, prior values outside $\text{IS}_\theta(\phi)$ are assigned zero mass, whereas the posterior inside the identified set is just proportional to the prior, see Moon and Schorfheide (2012) or Baumeister and Hamilton (2015). This is because the likelihood depends only on ϕ , thus this conditional prior is not updated by the data. For an in-depth discussion of whether the shape of the posterior inside the identified set can still be useful see Gustafson (2014). A full ambiguity-robust Bayesian in the sense of Giacomini and Kitagawa (2021) instead allows every conditional prior supported on the identified set. In the notation above, this corresponds to

$$\mathcal{M}_S(\phi) = \mathcal{P}(S(\text{IS}_\theta(\phi))), \quad \mathcal{R}(\phi) = \mathcal{R}^{\text{full}}(\phi),$$

and hence $\Pi_{\mathcal{M}, \mathcal{R}}(\phi) = \mathcal{P}([\phi, \phi + \lambda])$.

Giacomini and Kitagawa (2021) show that the corresponding set of posterior means coincides with the identified set. In particular, this implies that there exists $\pi_{\theta|\phi}^* \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)$ such that

$$\mathbb{E}_{\pi^*}[\theta \mid y, \phi] = \phi + \lambda.$$

This is only possible if $\pi_{\theta|\phi}^* = \delta_{\phi+\lambda}$. Such priors are formally admissible under full ambiguity, but may be substantively implausible if the researcher has additional prior knowledge.

As above, we might have such additional knowledge in terms of qualitative restrictions. For illustration, write $\theta = \phi + \lambda U$ with $U \in [0, 1]$ and restrict the density of U to be weakly decreasing:

$$\Pi_{\theta|\phi}^\downarrow = \left\{ \pi_{\theta|\phi} : \theta = \phi + \lambda U, P_U \ll \text{Leb}, \frac{dP_U}{du} \text{ is weakly decreasing on } [0, 1] \right\}.$$

This simply imposes a qualitative shape restriction on the conditional prior for the scalar structural parameter: values of θ closer to the lower endpoint ϕ are weakly more likely than

values closer to $\phi + \lambda$. Then the closure of the set of posterior means is

$$\left[\phi, \phi + \frac{\lambda}{2} \right].$$

A proof is provided in the Appendix. Thus, a qualitative shape restriction reduces ambiguity without forcing a fully specified prior distribution. Note that, as here the parameter is univariate there is no completion ambiguity.

In this case, the set of posteriors is shrunk purely by assumption. However, these qualitative shape restrictions can be motivated from a Bayesian perspective. For instance, assume θ captures the distance from the actual policy rate to the optimal one. A decision maker may regard small adjustments to the actual rate as *ex ante* more plausible or may use such a class as a regularizing ambiguity set around the benchmark. A monotone shape restriction as above captures this and can be seen as a soft tie-breaking rule. Independent of such a justification, the example should only be seen as illustrative. In richer models, $S(\theta)$ need not determine θ , so restrictions on $\mathcal{M}_S(\phi)$ interact with the completion class $\mathcal{R}(\phi)$ in determining the admissible distribution of the target $T(\theta)$.

4 Robust Conclusions in the Oil Market

The relative importance of supply and demand disturbances in driving oil-price fluctuations has been debated for almost two decades. The influential analysis of Kilian (2009) attributes much of the variation in real oil prices to demand-side disturbances rather than exogenous disruptions to oil production. Subsequent work by Kilian and Murphy (2012) and Kilian and Murphy (2014) reaches broadly similar conclusions while replacing recursive identification with sign restrictions and additional economically motivated bounds. A particularly important restriction is that the within-month price elasticity of oil supply cannot exceed 0.0258.

The treatment of this elasticity has subsequently become one of the central points of dis-

agreement in the literature. Baumeister and Hamilton (2019) argue that interpreting the 0.0258 value as a hard identifying restriction treats uncertain economic information as known with certainty. They instead place explicit prior distributions on economically meaningful structural parameters and obtain a larger role for supply disturbances. Caldara, Cavallo, and Iacoviello (2019), using external information on oil-market elasticities, similarly emphasize that relatively small changes in admissible supply and demand elasticities can materially change the decomposition of oil-price fluctuations. In contrast, Kilian (2022) argues that the evidence continues to support a very small within-month supply elasticity and that allowing too much prior mass on large elasticities removes important identifying information.

The disagreement is therefore not simply about whether prior information should be used. Rather, it concerns how strongly available economic knowledge about the elasticities should restrict the structural model. This is precisely the setting for the framework developed above. Instead of selecting one particular elasticity bound or one fully specified parametric prior, we ask which conclusions about oil-price fluctuations hold uniformly over classes of priors that encode only relatively weak features of this information.

We follow the four-variable monthly oil-market VAR of Kilian and Murphy (2014), containing global oil production, global real activity, the real price of oil, and changes in oil inventories. Except for the treatment of the elasticity restrictions, the data, sample period, transformations, lag length, reduced-form prior, shock normalizations, and maintained impact and dynamic restrictions, $\mathcal{A}_0$, are identical to those in Kilian and Murphy (2014). The maintained identifying assumptions consist of the standard impact and dynamic sign restrictions used to distinguish an oil supply shock, a flow-demand shock, a speculative-demand shock, and a residual shock. The exercise below leaves these restrictions unchanged and varies only the prior information imposed on the short-run elasticities.

For every admissible structural model, we compute two measures of the short-run supply elasticity. Let α_s^F denote the impact response of oil production relative to the impact response of the oil price following a flow-demand shock, and let α_s^S denote the analogous ratio following

a speculative-demand shock. Under the maintained sign restrictions both are non-negative. We also compute the impact price elasticity of demand in use, denoted ϵ_d , using the response of oil use, production net of inventory accumulation, to an oil supply shock.

The first prior class replaces the hard supply-elasticity bound by the probability restrictions

$$P_\pi(\alpha_s^F \leq 0.0258 \mid \phi) \geq 1 - \kappa, \quad (3)$$

$$P_\pi(\alpha_s^S \leq 0.0258 \mid \phi) \geq 1 - \kappa, \quad (4)$$

$$P_\pi(\epsilon_d \leq 0 \mid \phi) \geq 1 - \kappa. \quad (5)$$

Equivalently, the prior probability of each violation event, $\{\alpha_s^F > 0.0258\}$, $\{\alpha_s^S > 0.0258\}$, and $\{\epsilon_d > 0\}$, is at most κ ; these are the events G_ℓ of Section 3.3. For simplicity, the same value of κ is used for all three restrictions. Thus $\kappa = 0$ recovers hard upper restrictions on the two supply elasticities and rules out an upward-sloping demand curve, whereas increasing κ allows an increasing fraction of prior probability to violate these restrictions. Importantly, no particular distribution is specified within or outside these regions.

We consider two versions of this prior class in the main figure. The *Tail only* specification imposes only the tail events specified above. The *Tail + global monotonicity* specification additionally requires the marginal prior densities of $|\alpha_s^F|$, $|\alpha_s^S|$ and $|\epsilon_d|$ to be weakly decreasing, in the sense of Example 7. This expresses the qualitative belief that small elasticities are more plausible than large elasticities, while leaving their exact distribution unrestricted. The table below additionally reports an intermediate specification in which monotonicity is imposed only in the violation region.

The main question is not the exact posterior mean contribution of each shock, but whether the substantive ranking of supply and demand survives variation over the admissible prior class. Let $V_{p,h}^s(\theta)$, $V_{p,h}^f(\theta)$ and $V_{p,h}^{sp}(\theta)$ denote the h -month forecast-error variance shares of the real oil price attributable to the supply, flow-demand, and speculative-demand shocks, respectively. Define the combined demand contribution by $V_{p,h}^d(\theta) = V_{p,h}^f(\theta) + V_{p,h}^{sp}(\theta)$. The

FEVD denominator contains all four structural shocks, including the residual shock; the residual shock is simply not included in the demand aggregate.

The left panel of Figure 2 considers the event

$$A_D = \{V_{p,12}^d(\theta) \geq V_{p,12}^s(\theta)\}, \quad (6)$$

that is, that the two economically identified demand shocks jointly account for a larger share of the 12-month real-oil-price forecast error variance than the supply shock. For each κ , the figure reports

$$\underline{p}_D(\kappa) = \inf_{\pi \in \Pi_\kappa} P_\pi(A_D | Y), \quad \bar{p}_D(\kappa) = \sup_{\pi \in \Pi_\kappa} P_\pi(A_D | Y). \quad (7)$$

These are posterior probability bounds across *all* priors satisfying the stated elasticity restrictions. In particular, the lower bound answers how confident one can be in demand dominance under the prior in the class that is least favorable to that conclusion.

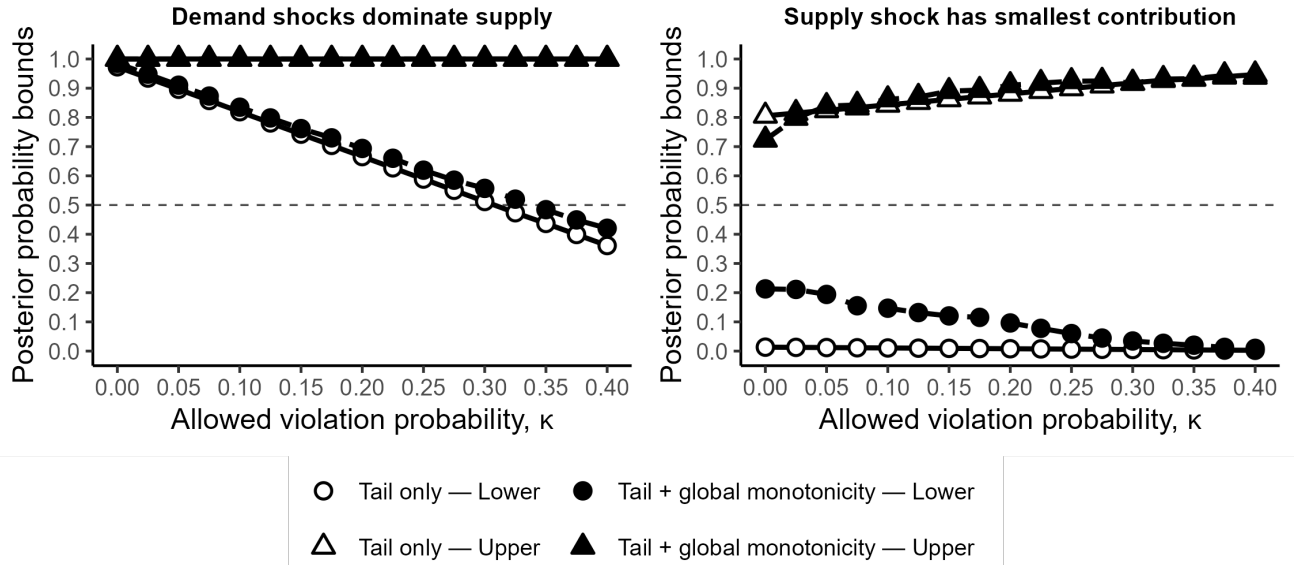


Figure 2: Robust posterior probability frontiers for real-oil-price FEVD rankings

Notes: The left panel reports lower and upper posterior probabilities that the combined contribution of the flow-demand and speculative-demand shocks to the 12-month forecast error variance of the real oil price exceeds the contribution of the supply shock. The right panel reports lower and upper posterior probabilities that the supply shock contributes less than each of the flow-demand and speculative-demand shocks individually. The residual shock is excluded from this ranking. “Tail only” imposes (3)–(5); “Tail + global monotonicity” additionally requires the marginal prior densities of the absolute elasticities to be weakly decreasing. The dashed horizontal line denotes probability one half.

The left panel delivers a sharp conclusion. When the elasticity restrictions are imposed without allowing violations, essentially the entire admissible prior class assigns very high posterior probability to demand dominance. More importantly, this conclusion does not depend on treating the elasticity bounds as literal support restrictions. Under the tail-only class, the worst-case posterior probability remains above one half even when as much as 30% prior probability is allowed to violate each elasticity restriction. Global monotonicity strengthens the result somewhat, but is not responsible for it. Thus the conclusion that demand-side disturbances jointly dominate the supply shock is robust to substantial disagreement about the exact elasticity prior.

This result should not be interpreted as saying that supply shocks are irrelevant. The right panel deliberately considers the stronger event

$$A_S = \left\{ V_{p,12}^s(\theta) \leq \min \left[V_{p,12}^f(\theta), V_{p,12}^{sp}(\theta) \right] \right\},$$

that the supply shock is smaller than *both* economically identified demand shocks individually. Here the data and weak prior information do not deliver a robust ranking. Even at $\kappa = 0$, the tail-only class contains priors assigning essentially zero posterior probability to this event as well as priors assigning probabilities above 80%. Global monotonicity narrows the range, but the lower probability remains far below one half. As κ increases, the bounds widen further. The contrast between the two panels is the main empirical payoff of the proposed approach. A single-prior analysis necessarily produces one probability for each event. Full set identification, by contrast, may leave both rankings largely unresolved. The intermediate prior classes used here distinguish the two claims. They provide strong evidence for the economically relevant aggregate statement that demand-side disturbances jointly dominate supply, while refusing to support the substantially stronger statement that the supply shock must be the least important individual disturbance.

Table 2 reports the underlying posterior-mean FEVD bounds. Unlike Figure 2, which directly

bounds posterior probabilities of economically meaningful ranking events, the table reports

$$\inf_{\pi \in \Pi_\kappa} E_\pi[V_{j,h}^r(\theta) \mid Y] \quad \text{and} \quad \sup_{\pi \in \Pi_\kappa} E_\pi[V_{j,h}^r(\theta) \mid Y]$$

for shock r , variable j , and horizons $h = 12, 24$.

Table 2: Credal bounds for forecast-error variance contributions

κ	Variable	Object	12 months			24 months		
			Tail only	Tail + monotone	Global monotone	Tail only	Tail + monotone	Global monotone
0.00	Real price	Supply	[0.07, 0.22]	[0.07, 0.22]	[0.10, 0.19]	[0.07, 0.16]	[0.07, 0.16]	[0.09, 0.15]
		Flow demand	[0.47, 0.81]	[0.47, 0.81]	[0.55, 0.75]	[0.51, 0.79]	[0.51, 0.79]	[0.58, 0.73]
		Speculative demand	[0.03, 0.33]	[0.03, 0.33]	[0.08, 0.26]	[0.05, 0.28]	[0.05, 0.28]	[0.09, 0.23]
		Total demand	[0.65, 0.88]	[0.65, 0.88]	[0.74, 0.86]	[0.70, 0.86]	[0.70, 0.86]	[0.76, 0.85]
0.00	Inventories	Supply	[0.08, 0.24]	[0.08, 0.24]	[0.15, 0.25]	[0.10, 0.23]	[0.10, 0.23]	[0.16, 0.23]
		Flow demand	[0.06, 0.23]	[0.06, 0.23]	[0.08, 0.16]	[0.12, 0.25]	[0.12, 0.25]	[0.13, 0.20]
		Speculative demand	[0.38, 0.75]	[0.38, 0.75]	[0.44, 0.63]	[0.36, 0.64]	[0.36, 0.64]	[0.41, 0.56]
		Total demand	[0.51, 0.83]	[0.51, 0.83]	[0.54, 0.72]	[0.53, 0.77]	[0.53, 0.77]	[0.56, 0.70]
0.15	Real price	Supply	[0.06, 0.27]	[0.06, 0.26]	[0.06, 0.23]	[0.06, 0.19]	[0.06, 0.19]	[0.06, 0.18]
		Flow demand	[0.35, 0.84]	[0.36, 0.83]	[0.38, 0.80]	[0.38, 0.80]	[0.39, 0.80]	[0.39, 0.76]
		Speculative demand	[0.02, 0.39]	[0.02, 0.38]	[0.05, 0.37]	[0.04, 0.33]	[0.04, 0.32]	[0.06, 0.31]
		Total demand	[0.49, 0.89]	[0.51, 0.89]	[0.56, 0.89]	[0.53, 0.88]	[0.55, 0.87]	[0.57, 0.87]
0.15	Inventories	Supply	[0.07, 0.33]	[0.07, 0.32]	[0.09, 0.31]	[0.09, 0.30]	[0.10, 0.29]	[0.11, 0.29]
		Flow demand	[0.06, 0.34]	[0.06, 0.33]	[0.07, 0.31]	[0.11, 0.33]	[0.11, 0.32]	[0.11, 0.31]
		Speculative demand	[0.31, 0.76]	[0.32, 0.76]	[0.32, 0.70]	[0.30, 0.65]	[0.31, 0.65]	[0.31, 0.61]
		Total demand	[0.43, 0.84]	[0.44, 0.84]	[0.43, 0.79]	[0.46, 0.78]	[0.47, 0.78]	[0.46, 0.75]
0.30	Real price	Supply	[0.05, 0.31]	[0.05, 0.30]	[0.06, 0.28]	[0.05, 0.22]	[0.05, 0.21]	[0.06, 0.21]
		Flow demand	[0.22, 0.86]	[0.25, 0.85]	[0.26, 0.83]	[0.25, 0.81]	[0.27, 0.81]	[0.28, 0.78]
		Speculative demand	[0.02, 0.44]	[0.02, 0.42]	[0.03, 0.41]	[0.03, 0.37]	[0.03, 0.35]	[0.04, 0.34]
		Total demand	[0.33, 0.91]	[0.37, 0.90]	[0.40, 0.90]	[0.36, 0.89]	[0.41, 0.88]	[0.42, 0.88]
0.30	Inventories	Supply	[0.06, 0.41]	[0.06, 0.39]	[0.08, 0.39]	[0.09, 0.37]	[0.09, 0.35]	[0.10, 0.35]
		Flow demand	[0.06, 0.45]	[0.06, 0.42]	[0.06, 0.40]	[0.10, 0.41]	[0.10, 0.39]	[0.10, 0.38]
		Speculative demand	[0.24, 0.77]	[0.26, 0.77]	[0.26, 0.73]	[0.24, 0.66]	[0.26, 0.65]	[0.26, 0.63]
		Total demand	[0.35, 0.85]	[0.37, 0.85]	[0.37, 0.82]	[0.39, 0.79]	[0.41, 0.79]	[0.41, 0.77]

Notes: Entries are lower and upper posterior expectations of FEVD shares across the corresponding prior class. “Total demand” is the sum of the flow-demand and speculative-demand shares. FEVD shares are computed relative to the variance generated by all four structural shocks, including the residual shock, although the residual contribution is not displayed. Tail only imposes (3)–(5). Tail + monotone additionally imposes decreasing density in the violation region. Global monotone imposes decreasing marginal densities in the absolute elasticities over their full support. All specifications maintain the impact and dynamic sign restrictions collected in $\mathcal{A}_0$.

The levels in Table 2 explain the probability result. At $\kappa = 0$, even the tail-only class bounds the posterior mean supply contribution to the 12-month real-price FEVD between 7% and 22%, whereas the combined demand contribution lies between 65% and 88%. Under global monotonicity these intervals tighten to 10–19% for supply and 74–86% for total demand. The corresponding 24-month bounds are similar. As κ increases, the intervals widen substantially, reflecting the growing ambiguity about the individual FEVD shares.

The table also illustrates why it is useful to distinguish aggregate demand dominance from a complete ranking of the shocks. The contribution of speculative demand is itself poorly

pinned down relative to flow demand, and the admissible ranges for individual shocks can be wide even when their combined contribution is tightly enough restricted to support the aggregate demand-dominance event.

Taken together, the results provide a qualification to both sides of the oil-market debate. They support the central conclusion of the demand-driven literature that demand-side disturbances are collectively more important than supply shocks for typical real-oil-price fluctuations. But this conclusion does not require treating the 0.0258 supply-elasticity bound as known with certainty. At the same time, the results do not support the stronger claim that supply shocks are negligible or invariably the least important disturbance. The robust conclusion is therefore one of *demand dominance*, not *supply irrelevance*.

5 Conclusion

This paper develops a robust Bayesian framework for structural models in which the researcher is willing to defend some implications of a prior distribution, but not a complete precise prior. The distinction matters particularly in set-identified models. Conditional on the reduced form, observationally equivalent structural parameters have the same likelihood, so the data cannot correct arbitrary features of the conditional structural prior. A precise prior therefore turns assumptions that may in fact be only qualitative into precise posterior probabilities that persist even with highly informative data.

The framework starts from the prior information the researcher is actually willing to defend. Let $S(\theta)$ denote an economically interpretable feature of the structural model and let $\mathcal{M}_S(\phi)$ denote a class of admissible conditional distributions for that feature. Rather than completing this information with an arbitrary joint prior for the entire structural parameter, we consider every conditional prior $\pi_{\theta|\phi}$ whose induced marginal satisfies $S_{\#}\pi_{\theta|\phi} \in \mathcal{M}_S(\phi)$. The class $\mathcal{M}_S(\phi)$ may encode tail probabilities, moments, shape restrictions, or combinations thereof. This permits statements such as “violations of this elasticity bound are unlikely” or “smaller

elasticities are more plausible than larger ones” to be imposed without additionally specifying probabilities that the underlying economic knowledge does not determine.

The statistic used to express prior information need not coincide with the object ultimately of interest. If $T(\theta)$ is the target, restrictions on the distribution of $S(\theta)$ restrict the admissible distributions of the common structural parameter θ and thereby the admissible posterior distributions of $T(\theta)$. This is important in structural applications, where prior knowledge may concern elasticities, signs, or narrative events, while the substantive question concerns variance decompositions, impulse responses, historical contributions, or policy statistics.

Set identification makes the resulting robust problem unusually tractable. Because the likelihood depends only on the reduced form, conditional on ϕ the posterior distribution of the unidentified structural parameter coincides with its prior. Posterior expectations are therefore linear in $\pi_{\theta|\phi}$. When $\mathcal{M}_S(\phi)$ is described by affine restrictions on probability measures, the lower and upper posterior expectations of T are solutions to infinite-dimensional linear programs. Under compactness and regularity conditions, the extrema exist. Moreover, the paper shows that these population bounds can be approximated consistently using a finite set of structural draws and optimization over their probability weights. The resulting finite-dimensional linear programs therefore provide a simulation-based implementation of the population robust Bayes problem.

The conditional construction also has a natural joint interpretation. Holding the reduced-form prior fixed generates a rectangular class of joint priors in which all elements share the same marginal law for ϕ but differ in the unrevisable conditional law of θ given ϕ . The conditional robust bounds consequently aggregate by integrating them with respect to the common reduced-form posterior. We give conditions under which the conditional prior features are inherited by the joint prior.

The general framework is deliberately baseline-free. The reference measure used to generate structural draws is computational and, under the support conditions of the approximation theorem, does not determine the population bounds. If instead a researcher attaches substan-

tive meaning to a baseline structural prior, the paper provides a complementary specialization. Restricting departures from the baseline to reweightings that depend only on $S(\theta)$ generates, for every admissible marginal distribution of S , the unique completion that preserves the baseline conditional distribution of θ given S . When the relevant divergence is finite, this completion is also the Kullback–Leibler closest joint prior to the baseline among all priors inducing that marginal. This links the framework to marginally specified priors and to KL-based robust Bayesian sensitivity analysis, while clarifying that distance from a baseline is only one possible way to define a prior class.

The oil-market application illustrates the empirical payoff. The disagreement in the literature is not whether elasticities contain identifying information, but how strongly uncertain information about those elasticities should restrict the model. Instead of treating conventional elasticity bounds as certain or replacing them by one fully specified parametric prior, we encode them through classes of distributions that allow a controlled probability of violation and, in some specifications, only the qualitative requirement that smaller elasticities are more plausible than larger ones. The resulting posterior probability frontiers distinguish two economic statements that a precise-prior analysis can easily blur together. The combined contribution of flow-demand and speculative-demand shocks to real-oil-price fluctuations robustly exceeds the contribution of the supply shock even under substantial weakening of the elasticity restrictions. By contrast, the stronger statement that the supply shock contributes less than each demand shock individually is not robust. The evidence therefore supports demand dominance, but not supply irrelevance.

The broader lesson is that ambiguity about a prior need not be represented either by selecting an arbitrary precise distribution or by allowing every distribution on the identified set. Economic knowledge often determines only some features of a prior. Those features can themselves be treated as the primitive identifying information. Doing so makes clear which posterior conclusions follow from the economic statements a researcher is prepared to defend and which conclusions arise only from an unnecessary completion of those statements into a

precise prior distribution.

References

- Arias J. Rubio-Ramirez, J. and D. Waggoner (2024). “Uniform Priors for Impulse Responses”. In: *forthcoming*.
- Baumeister, C. and J. Hamilton (2015). “Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information”. In: *Econometrica* 83.5, pp. 1963–1999.
- (2019). “Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks”. In: *American Economic Review* 109.5, pp. 1873–1910.
- Caldara, D., M. Cavallo, and M. Iacoviello (2019). “Oil Price Elasticities and Oil Price Fluctuations”. In: *Journal of Monetary Economics* 103, pp. 1–20.
- Canay, I. A. and A. M. Shaikh (2017). “Practical and Theoretical Advances in Inference for Partially Identified Models”. In: *Advances in Economics and Econometrics: Eleventh World Congress*. Vol. 2. Cambridge University Press, pp. 271–306.
- Giacomini, R. and T. Kitagawa (2021). “Robust Bayesian Inference for Set-Identified Models”. In: *Econometrica* 89, pp. 1519–1556.
- Giacomini, R., T. Kitagawa, and H. Uhlig (2019). “Estimation under Ambiguity”. In.
- Gustafson, P. (2014). “Bayesian inference in partially identified models: Is the shape of the posterior distribution useful?” In: *Electronic Journal of Statistics* 8, pp. 476–496.
- Ho, P. (2023). “Global Robust Bayesian Analysis in Large Models”. In: *Journal of Econometrics* 235, pp. 608–642.
- Kessler, D., P. Hoff, and D. Dunson (2015). “Marginally specified priors for non-parametric Bayesian estimation”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77, pp. 35–58.

- Kilian, L. (2009). “Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market”. In: *American Economic Review* 99.3, pp. 1053–69.
- (2022). “Facts and Fiction in Oil Market Modeling”. In: *Energy Economics* 110, p. 105973.
- Kilian, L. and D. Murphy (2012). “Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models”. In: *Journal of European Economic Association* 10.5, pp. 1166–88.
- (2014). “The Role of Inventories and Speculative Trading in the Global Market for Crude Oil”. In: *Journal of Applied Econometrics* 29.3, pp. 454–78.
- Lam, H., Z. Liu, and D. I. Singham (2024). “Shape-Constrained Distributional Optimization via Importance-Weighted Sample Average Approximation”. In: *arXiv preprint arXiv:2406.07825*.
- Manski, C. (2003). *Partial Identification of Probability Distributions*. Springer.
- Molinari, F. (2020). “Microeconometrics with Partial Identification”. In: *Handbook of Econometrics*. Vol. 7A. Elsevier, pp. 355–486.
- Moon, H. and F. Schorfheide (2012). “Bayesian and Frequentist Inference in Partially Identified Models”. In: *Econometrica* 80.2, pp. 755–782.
- Rubio-Ramirez, J., D. Waggoner, and T. Zha (2010). “Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference”. In: *The Review of Economic Studies* 77.2.
- Schennach, S. (2014). “Entropic Latent Variable Integration via Simulation”. In: *Econometrica* 82.1, pp. 345–385.
- Uhlig, H. (2005). “What are the effects of monetary policy on output? Results from an agnostic identification procedure”. In: *Journal of Monetary Economics* 52, pp. 381–419.

A Proofs

A.1 Proof of Proposition 1

Proof. Fix ϕ and let $\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)$. Under Assumption 1, the likelihood is constant in θ conditional on ϕ . Hence, for every ϕ for which the conditional posterior is well defined,

$$\begin{aligned}\pi(d\theta \mid Y, \phi) &= \frac{p(Y \mid \phi, \theta) \pi(d\theta)}{\int_{\Theta} p(Y \mid \phi, \tilde{\theta}) \pi(d\tilde{\theta})} \\ &= \frac{p(Y \mid \phi) \pi(d\theta)}{p(Y \mid \phi) \int_{\Theta} \pi(d\tilde{\theta})} = \pi(d\theta).\end{aligned}$$

Therefore,

$$\mathbb{E}_{\pi}[T \mid Y, \phi] = \int_{\Theta} T(\theta) \pi(d\theta).$$

Since T is bounded and measurable, this integral is finite for every $\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)$. Taking the infimum and supremum over the admissible prior class gives

$$\begin{aligned}\underline{m}_T(\phi) &= \inf_{\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)} \int_{\Theta} T(\theta) \pi(d\theta), \\ \overline{m}_T(\phi) &= \sup_{\pi \in \Pi_{\mathcal{M}, \mathcal{R}}(\phi)} \int_{\Theta} T(\theta) \pi(d\theta).\end{aligned}$$

Now suppose that the admissible prior class has the affine representation

$$\Pi_{\mathcal{M}, \mathcal{R}}(\phi) = \left\{ \pi \in \mathcal{P}(\text{IS}_{\theta}(\phi)) : \int a_j(\theta) \pi(d\theta) \leq b_j \quad \forall j \in J \right\}.$$

The map

$$\pi \mapsto \int_{\Theta} T(\theta) \pi(d\theta)$$

is linear in the decision variable π , and each constraint

$$\int_{\Theta} a_j(\theta) \pi(d\theta) \leq b_j$$

is affine in π . The probability-measure restrictions $\pi \geq 0$ and $\int_{\Theta} \pi(d\theta) = 1$ are likewise linear. Consequently, the lower and upper bounds are the optimal values of, respectively, an infinite-dimensional linear minimization and maximization problem over probability measures. $\square$

A.2 Auxiliary lemmas for Theorem 1

Fix ϕ throughout, write $K = \text{IS}_{\theta}(\phi)$, and suppress dependence on ϕ . Let $\mathcal{C}$ be the finite partition generated by the restriction events G_{ℓ} and, for a target $T = \mathbf{1}_A$, the target event A . For a bounded function $f : K \rightarrow \mathbb{R}$, write

$$\omega_f(t) = \sup\{|f(x) - f(y)| : x, y \in K, d(x, y) \leq t\}.$$

For the first N draws, define

$$h_N = \max_{C \in \mathcal{C}} \sup_{x \in C} \min_{\substack{1 \leq n \leq N \\ \theta^{(n)} \in C}} d(x, \theta^{(n)}), \quad \min \emptyset = +\infty.$$

The draws are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Because K is a compact metric space, $\mathcal{P}(K)$ is compact and metrizable under weak convergence. We use the Portmanteau theorem: if $\pi_k \Rightarrow \pi$, then $\limsup_k \pi_k(F) \leq \pi(F)$ for relatively closed $F \subseteq K$, and $\liminf_k \pi_k(G) \geq \pi(G)$ for relatively open $G \subseteq K$. Consequently, $\pi \mapsto \pi(F)$ is weakly upper semicontinuous, $\pi \mapsto \pi(G)$ is weakly lower semicontinuous, and $\pi \mapsto \int_K f d\pi$ is weakly continuous for $f \in C(K)$.

Lemma 1 (Closedness and nonemptiness). *Under Assumptions 2 and 3, each $\Pi^{(r)}(\phi)$ and $\Pi(\phi)$ is weakly compact. If also Assumptions 4 and 6 hold, then $\Pi(\phi) \neq \emptyset$.*

Proof. For every $j \in J$, the set $\{\pi : \int_K a_j d\pi \leq b_j\}$ is weakly closed because $a_j \in C(K)$, so that $\pi \mapsto \int_K a_j d\pi$ is weakly continuous by the definition of weak convergence. For every $\ell \in L$, the set $\{\pi : \pi(G_\ell) \leq \kappa_\ell\}$ is weakly closed because G_ℓ is relatively open, so that $\pi \mapsto \pi(G_\ell)$ is weakly lower semicontinuous by the Portmanteau theorem; since $\mathcal{P}(K)$ is metrizable, sequential lower semicontinuity suffices. Intersections of these closed sets are closed. Since K is a compact metric space, $\mathcal{P}(K)$ is weakly compact by Prokhorov's theorem, so both the population and stage classes, as closed subsets of $\mathcal{P}(K)$, are compact.

Assumption 6 makes every stage class nonempty. By Assumption 4, these compact classes are decreasing and their intersection is $\Pi(\phi)$; the finite intersection property therefore gives $\Pi(\phi) \neq \emptyset$. $\square$

Lemma 2 (Measurability of the finite-program values). *Under Assumptions 3, 4, and 5, for $T \in C(K)$ or $T = \mathbf{1}_A$ with A relatively closed, the bounds $\bar{m}_{T,N,r}(\phi)$ and $\underline{m}_{T,N,r}(\phi)$ are extended-real random variables for every N, r .*

Proof. Fix N, r , and let β_r collect the right-hand sides $(b_j)_{j \in J_r}$ and $(\kappa_\ell)_{\ell \in L}$. For $c \in \mathbb{R}^N$ and a matrix M with N columns and $|J_r| + |L|$ rows, set

$$V(c, M) = \max\{c^\top w : w \in \Delta_N, Mw \leq \beta_r\}, \quad V(c, M) = -\infty \text{ if infeasible.}$$

The map V is upper semicontinuous. Indeed, if $(c_k, M_k) \rightarrow (c, M)$ and $V(c_k, M_k) \geq \alpha$, choose a maximizer w_k for each k . A subsequence converges in the compact simplex Δ_N to some w . Passing to the limit gives $Mw \leq \beta_r$ and $c^\top w = \lim_k c_k^\top w_k \geq \alpha$. Thus every superlevel set $\{V \geq \alpha\}$ is closed, and V is Borel measurable.

The entries of the objective vector and restriction matrix are, respectively, $T(\theta^{(n)})$, $a_j(\theta^{(n)})$, and $\mathbf{1}_{G_\ell}(\theta^{(n)})$. They are measurable because the draws are random elements of K , the a_j are continuous, and both target types and the event indicators are Borel measurable. Therefore $\bar{m}_{T,N,r}(\phi) = V(c, M)$ is measurable, as is $\underline{m}_{T,N,r}(\phi) = -V(-c, M)$. $\square$

Lemma 3 (Cell covering radius). *Under Assumptions 2 and 5, there is $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that $h_N \rightarrow 0$ on Ω_0 .*

Proof. Fix $k \in \mathbb{N}$ and $C \in \mathcal{C}$. Cover the compact set K by finitely many balls $B(y_i, 1/(4k))$. For every ball meeting C , choose $x_i \in B(y_i, 1/(4k)) \cap C$. The corresponding balls $B(x_i, 1/(2k))$ cover C . By Assumption 5, each set $B(x_i, 1/(2k)) \cap C$ has positive probability p_i . Independence gives probability $(1 - p_i)^N \rightarrow 0$ that none of the first N draws lies in this set. Hence, almost surely, each of the finitely many sets contains a draw eventually.

Once they have all been hit, every $x \in C$ is within $1/k$ of a draw in C . The cells are finite in number, so almost surely $h_N \leq 1/k$ eventually. Intersect these measurable probability-one events over $k \in \mathbb{N}$ to obtain Ω_0 ; on this event $h_N \rightarrow 0$. $\square$

Lemma 4 (Cell-preserving projection). *Fix a realization of the draws and N with $h_N < \infty$. For x in cell C , let $\sigma_N(x)$ be a closest draw among $\{\theta^{(n)} : n \leq N, \theta^{(n)} \in C\}$, breaking ties by the smallest index. Then $\sigma_N : K \rightarrow \Theta_N$ is Borel measurable. For any $\pi \in \mathcal{P}(K)$, its pushforward $P_N = \pi \circ \sigma_N^{-1} \in \mathcal{P}(\Theta_N)$ satisfies:*

(P1) $\sigma_N(x)$ lies in the same cell as x ;

(P2) $d(x, \sigma_N(x)) \leq h_N$ for every $x \in K$;

(P3) $P_N(E) = \pi(E)$ for every union E of cells, including each G_ℓ and, for an indicator target, A ;

(P4) For every bounded Borel function $f : K \rightarrow \mathbb{R}$,

$$\left| \int_K f dP_N - \int_K f d\pi \right| \leq \omega_f(h_N).$$

Proof. For each cell C , put $I_C = \{n \leq N : \theta^{(n)} \in C\}$, which is nonempty because $h_N < \infty$.

For $n \in I_C$, define the set of points assigned *index* n by

$$H_{n,C} = \left\{ x \in C : d(x, \theta^{(n)}) < d(x, \theta^{(m)}) \text{ for all } m \in I_C, m < n; \right. \\ \left. d(x, \theta^{(n)}) \leq d(x, \theta^{(m)}) \text{ for all } m \in I_C, m > n \right\}.$$

These sets are Borel and partition C ; on $H_{n,C}$, $\sigma_N(x) = \theta^{(n)}$. Thus σ_N is Borel even when two draws have the same location. Properties (P1) and (P2) follow from its construction. For any union E of cells, (P1) implies $\sigma_N^{-1}(E) = E$, proving (P3). The restriction events and, when applicable, A are unions of cells by their definition. Finally, the pushforward identity and (P2) give

$$\left| \int_K f dP_N - \int_K f d\pi \right| = \left| \int_K (f \circ \sigma_N - f) d\pi \right| \leq \omega_f(h_N),$$

proving (P4). □

Lemma 5 (Deterministic diagonal). *Let $X_{N,r}$ be extended-real random variables and $x_r, x \in \mathbb{R}$. Suppose $X_{N,r} \rightarrow x_r$ almost surely as $N \rightarrow \infty$ for every fixed r , and $x_r \rightarrow x$. Then some deterministic nondecreasing $r(N) \rightarrow \infty$ satisfies $X_{N,r(N)} \rightarrow x$ almost surely. One schedule can serve finitely many such arrays simultaneously.*

Proof. Define $D_{M,r} = \sup_{N \geq M} |X_{N,r} - x_r|$, allowing $D_{M,r} = +\infty$. This is measurable, and for each fixed r , $D_{M,r} \rightarrow 0$ almost surely, hence in probability, as $M \rightarrow \infty$. Choose deterministic $N_1 < N_2 < \dots$ such that $\mathbb{P}(D_{N_r,r} > 1/r) \leq 2^{-r}$, and set

$$r(N) = \begin{cases} 1, & N < N_1, \\ \max\{r : N_r \leq N\}, & N \geq N_1. \end{cases}$$

By the Borel–Cantelli lemma, almost surely $D_{N_r,r} \leq 1/r$ for all sufficiently large r . Hence,

on that event and for all sufficiently large N ,

$$|X_{N,r(N)} - x| \leq D_{N_{r(N)},r(N)} + |x_{r(N)} - x| \leq \frac{1}{r(N)} + |x_{r(N)} - x| \longrightarrow 0.$$

For finitely many arrays, choose each N_r so that the sum of their exceptional probabilities is at most 2^{-r} . $\square$

A.3 Proof of Theorem 1

Proof. Fix ϕ , write $K = \text{IS}_\theta(\phi)$, and suppress dependence on ϕ where no confusion arises.

Let

$$\bar{m}^{(r)} = \sup_{\pi \in \Pi^{(r)}} \int_K T d\pi, \quad \underline{m}^{(r)} = \inf_{\pi \in \Pi^{(r)}} \int_K T d\pi.$$

Write $\bar{m}_{N,r}$ and $\underline{m}_{N,r}$ for the sampled bounds, and $\bar{m}$ and $\underline{m}$ for the population bounds.

Whenever $h_N < \infty$, set

$$e_{N,r} = \max_{j \in J_r} \omega_{a_j}(h_N), \quad \max \emptyset = 0,$$

and

$$\Delta_{N,r} = \begin{cases} \omega_T(h_N) + 2\|T\|_\infty \frac{e_{N,r}}{\eta_r}, & T \in C(K), \\ \frac{e_{N,r}}{\eta_r}, & T = \mathbf{1}_A. \end{cases}$$

We take $e_{N,r}/\eta_r = 0$ when $J_r = \emptyset$.

The argument proceeds in five steps. We first compare the sampled and population feasible sets at a fixed stage and bound the resulting error. We then pass to the limits in N and r , in that order, before using a diagonal argument to obtain the joint limit.

Step 1: Every $\pi \in \Pi_{N,r}$ is a probability measure supported on Θ_N that satisfies all stage- r restrictions. Hence $\Pi_{N,r} \subseteq \Pi^{(r)}$, and therefore

$$\bar{m}_{N,r} \leq \bar{m}^{(r)}, \quad \underline{m}_{N,r} \geq \underline{m}^{(r)}.$$

Step 2: Fix r , N , and a realization of the draws such that $h_N < \infty$ and $e := e_{N,r} \leq \eta_r$. Let $\pi \in \Pi^{(r)}$ be arbitrary, and let $\pi_r^\circ \in \Pi^{(r)}$ be the prior supplied by Assumption 6. Apply the cell-preserving projection of Lemma 4 to obtain

$$P_N = \pi \circ \sigma_N^{-1}, \quad P_N^\circ = \pi_r^\circ \circ \sigma_N^{-1}.$$

If $J_r \neq \emptyset$, put $t = e/\eta_r \in [0, 1]$. If $J_r = \emptyset$, put $t = 0$. In either case, define

$$Q_N = (1 - t)P_N + tP_N^\circ \in \mathcal{P}(\Theta_N).$$

Suppose first that $J_r \neq \emptyset$. For each $j \in J_r$, property (P4) of Lemma 4 and the Slater margin give

$$\int_K a_j dP_N \leq b_j + e, \quad \int_K a_j dP_N^\circ \leq b_j - \eta_r + e.$$

Consequently,

$$\int_K a_j dQ_N \leq (1 - t)(b_j + e) + t(b_j - \eta_r + e) = b_j + e - t\eta_r = b_j.$$

When $J_r = \emptyset$, there are no continuous rows to check.

For every event row, property (P3) gives

$$Q_N(G_\ell) = (1 - t)\pi(G_\ell) + t\pi_r^\circ(G_\ell) \leq \kappa_\ell,$$

because π and π_r° both belong to $\Pi^{(r)}$. Thus $Q_N \in \Pi_{N,r}$; in particular, the sampled stage

class is nonempty.

For the objective,

$$\begin{aligned} \int_K T dQ_N - \int_K T d\pi &= (1-t) \left(\int_K T dP_N - \int_K T d\pi \right) \\ &\quad + t \left(\int_K T dP_N^\circ - \int_K T d\pi \right). \end{aligned}$$

If T is continuous, (P4) bounds the absolute value of the first bracket by $\omega_T(h_N)$, while the second is bounded by $2\|T\|_\infty$. If $T = \mathbf{1}_A$, (P3) makes the first bracket zero, and the second equals $\pi_r^\circ(A) - \pi(A) \in [-1, 1]$. Hence, in both cases,

$$\left| \int_K T dQ_N - \int_K T d\pi \right| \leq \Delta_{N,r}.$$

Since Q_N is sampled-stage feasible, this inequality holds for a possibly different Q_N for every $\pi \in \Pi^{(r)}$. Taking the supremum and infimum over π , respectively, yields

$$\bar{m}_{N,r} \geq \bar{m}^{(r)} - \Delta_{N,r}, \quad \underline{m}_{N,r} \leq \underline{m}^{(r)} + \Delta_{N,r}.$$

Combining these inequalities with Step 1 gives

$$0 \leq \bar{m}^{(r)} - \bar{m}_{N,r} \leq \Delta_{N,r}, \quad 0 \leq \underline{m}_{N,r} - \underline{m}^{(r)} \leq \Delta_{N,r}.$$

Step 3: Let Ω_0 be the probability-one event in Lemma 3. On Ω_0 , $h_N \rightarrow 0$, so $h_N < \infty$ for all sufficiently large N . For each fixed r , the set J_r is finite and every a_j is uniformly continuous on the compact space K . Therefore $e_{N,r} \rightarrow 0$ on Ω_0 . Since $\eta_r > 0$ when $J_r \neq \emptyset$, the condition $e_{N,r} \leq \eta_r$ holds eventually.

For a continuous target, uniform continuity also gives $\omega_T(h_N) \rightarrow 0$. For an indicator target, $\Delta_{N,r} = e_{N,r}/\eta_r \rightarrow 0$ directly; projection preserves the probability of A exactly. The

error bound in Step 2 therefore implies, for every fixed r ,

$$\bar{m}_{N,r} \longrightarrow \bar{m}^{(r)}, \quad \underline{m}_{N,r} \longrightarrow \underline{m}^{(r)} \quad \text{almost surely as } N \rightarrow \infty.$$

The same event Ω_0 works for all fixed stages because the cell partition does not depend on r .

Step 4: Lemma 1 gives $\Pi \neq \emptyset$ and compactness of every stage class. By Assumption 4,

$$\Pi \subseteq \Pi^{(r+1)} \subseteq \Pi^{(r)}.$$

Thus $\bar{m}^{(r)}$ decreases toward a limit no smaller than $\bar{m}$, while $\underline{m}^{(r)}$ increases toward a limit no greater than $\underline{m}$.

Consider first the upper bound. The map $\Phi(\pi) = \int_K T d\pi$ is weakly continuous when $T \in C(K)$. When $T = \mathbf{1}_A$ with A relatively closed, it is weakly upper semicontinuous by Portmanteau. Hence it attains its maximum on each compact $\Pi^{(r)}$; choose $\pi_r \in \Pi^{(r)}$ with $\Phi(\pi_r) = \bar{m}^{(r)}$. Compactness of $\mathcal{P}(K)$ gives a weakly convergent subsequence $\pi_{r_k} \Rightarrow \pi^*$. For every fixed s , the nesting of the stage classes implies $\pi_{r_k} \in \Pi^{(s)}$ for all sufficiently large k . Because $\Pi^{(s)}$ is weakly closed,

$$\pi^* \in \bigcap_{s \geq 1} \Pi^{(s)} = \Pi.$$

Upper semicontinuity now gives

$$\lim_{r \rightarrow \infty} \bar{m}^{(r)} = \lim_{k \rightarrow \infty} \Phi(\pi_{r_k}) \leq \Phi(\pi^*) \leq \bar{m}.$$

Together with $\bar{m}^{(r)} \geq \bar{m}$, this proves $\bar{m}^{(r)} \downarrow \bar{m}$. If T is continuous, applying the same argument to $-T$ also proves $\underline{m}^{(r)} \uparrow \underline{m}$.

It remains to consider the lower bound when $T = \mathbf{1}_A$. Let $\underline{m}_\infty = \lim_r \underline{m}^{(r)} \leq \underline{m}$. Choose $\pi_r \in \Pi^{(r)}$ such that

$$\pi_r(A) \leq \underline{m}^{(r)} + \frac{1}{r}.$$

A subsequence converges weakly to some $\pi^* \in \Pi$ by the same compactness and nesting argument. Since $\text{int}_K A$ is relatively open, Portmanteau gives

$$\pi^*(\text{int}_K A) \leq \liminf_{k \rightarrow \infty} \pi_{r_k}(\text{int}_K A) \leq \liminf_{k \rightarrow \infty} \pi_{r_k}(A) \leq \underline{m}_\infty.$$

Consequently,

$$\inf_{\pi \in \Pi} \pi(\text{int}_K A) \leq \underline{m}_\infty \leq \underline{m}.$$

Assumption 7 makes the outer quantities equal, so $\underline{m}_\infty = \underline{m}$. The same conclusion holds without that assumption if $\Pi^{(r)} = \Pi$ for all sufficiently large r .

Step 5: By Lemma 2, the sampled bounds are extended-real random variables. Step 3 gives their almost sure limits for every fixed r , and Step 4 shows that those stage limits converge to the population bounds under the conditions of the theorem. Applying Lemma 5 simultaneously to the upper- and lower-bound arrays yields a deterministic nondecreasing sequence $r(N) \rightarrow \infty$ such that

$$\bar{m}_{N,r(N)} \longrightarrow \bar{m}, \quad \underline{m}_{N,r(N)} \longrightarrow \underline{m} \quad \text{almost surely.}$$

□

A.4 Proof of Corollary 1

Proof. Fix ϕ . For any $\pi \in \mathcal{P}(K)$, let $p_C = \pi(C)$ for each $C \in \mathcal{C}$. Because the cells partition K , the vector $p = (p_C)_{C \in \mathcal{C}}$ belongs to $\Delta(\mathcal{C})$. Each restriction event G_ℓ and the target event A is a union of cells, so

$$\pi(G_\ell) = \sum_{C \subseteq G_\ell} p_C, \quad \pi(A) = \sum_{C \subseteq A} p_C.$$

Since $J = \emptyset$, these are all the restrictions on π . Thus every population-feasible prior produces a vector $p \in \mathcal{D}$ with the objective value stated in the corollary.

Conversely, for any $p \in \mathcal{D}$, choose one point $x_C \in C$ from each nonempty cell and define

$$\pi_p = \sum_{C \in \mathcal{C}} p_C \delta_{x_C}.$$

This prior belongs to $\Pi(\phi)$ and assigns mass p_C to each cell. Hence every vector in $\mathcal{D}$ is attained by a population-feasible prior. Taking the minimum and maximum over $\mathcal{D}$ proves the cell-mass formulas.

Now suppose that every cell contains at least one of the first N simulated points. Choose such a point $\theta_C \in C$ for each cell. For every $p \in \mathcal{D}$, the sampled prior

$$\pi_{N,p} = \sum_{C \in \mathcal{C}} p_C \delta_{\theta_C}$$

has the same cell masses, satisfies every event restriction, and has objective value $\sum_{C \subseteq A} p_C$. Conversely, the cell masses of every sampled feasible prior belong to $\mathcal{D}$. The sampled and population programs therefore optimize over the same vectors p , so both their lower and upper values are equal.

Finally, for each cell C ,

$$\mathbb{P}(\theta^{(1)}, \dots, \theta^{(N)} \notin C) = (1 - \nu(C))^N \longrightarrow 0.$$

There are finitely many cells and $\nu(C) > 0$ for each one. Almost surely, all cells have therefore been hit by some finite random time N_0 . They remain represented for every $N \geq N_0$, giving the asserted eventual equality. Since $J = \emptyset$, the stage index r has no effect. $\square$

A.5 Proof of Proposition 2

Proof. For the first result, since $\mu_\phi = S_\# \pi_{\theta|\phi}$,

$$\begin{aligned} \pi(S(\theta) \in B) &= \int_{\Phi} \mu_\phi(B) \pi_\phi(d\phi) \\ &\leq \int_{\Phi} \kappa \pi_\phi(d\phi) = \kappa, \end{aligned}$$

where the inequality follows from $\mu_\phi(B) \leq \kappa$ for π_ϕ -almost every ϕ .

For the second result, let $\mu_{Z|\phi}$ denote the conditional distribution of Z given ϕ , extend $f_{Z|\phi}$ by zero outside $[0, \bar{z}(\phi)]$, and note that extension by zero preserves monotonicity on $(0, \infty)$. Since $f_{Z|\phi}$ is non-increasing on $(0, \infty)$, its averages over $(z, z + 1/m]$ converge to its right limit, so its right-continuous version is

$$f_{Z|\phi}(z) = \lim_{m \rightarrow \infty} m \mu_{Z|\phi}((z, z + 1/m]), \quad z > 0.$$

For fixed z and m , the map $\phi \mapsto \mu_{Z|\phi}((z, z + 1/m])$ is measurable because $\pi_{\theta|\phi}$ is a probability kernel, and for fixed ϕ it is right-continuous in z . It is therefore jointly measurable in (ϕ, z) , and so is its limit. The weight $w(\phi) = \mu_{Z|\phi}(\{0\})$ is measurable for the same reason. The unconditional distribution of Z is the mixture of its conditional distributions over ϕ . Hence, by Tonelli's theorem, for every Borel set $B \subseteq [0, \infty)$,

$$\pi(Z \in B) = \bar{w} \mathbf{1}_B(0) + \int_B f_Z(z) dz, \quad f_Z(z) = \int_{\Phi} f_{Z|\phi}(z) \pi_\phi(d\phi).$$

Moreover, for any $0 < z_1 \leq z_2$, $f_{Z|\phi}(z_1) \geq f_{Z|\phi}(z_2)$ for π_ϕ -almost every ϕ . Integrating this inequality over ϕ gives

$$\begin{aligned} f_Z(z_1) &= \int_{\Phi} f_{Z|\phi}(z_1) \pi_\phi(d\phi) \\ &\geq \int_{\Phi} f_{Z|\phi}(z_2) \pi_\phi(d\phi) = f_Z(z_2), \end{aligned}$$

so f_Z is non-increasing on $(0, \infty)$. It is finite there, because a non-increasing density satisfies $f_{Z|\phi}(z) \leq 1/z$.

Finally, under Assumption 1 the conditional structural distribution given ϕ is not updated by the data, so $\pi_{\theta|Y,\phi} = \pi_{\theta|\phi}$ and therefore the conditional marginal μ_ϕ of S , and the corresponding conditional distribution $\mu_{Z|\phi}$ of Z , are unchanged. The joint posterior only replaces the mixing distribution π_ϕ by $\pi_\phi(\cdot | Y)$. Repeating the preceding arguments with $\pi_\phi(d\phi | Y)$ therefore proves both posterior statements. $\square$

A.6 Proof of Proposition 3

Proof. Fix $\mu \in \mathcal{M}_S(\phi)$ and let $r_{\mu,\phi} = d\mu/d\pi_{S|\phi}^0$. Since $\mathcal{R}(\phi) = \{\rho_\phi^0\}$, the unique prior with marginal μ and baseline completion is

$$\pi_{\theta|\phi}^\mu(A) = \int \rho_\phi^0(A | s) \mu(ds), \quad A \in \mathcal{A}_\theta.$$

Uniqueness follows directly because both the marginal μ and the completion ρ_ϕ^0 are fixed.

Using $\mu(ds) = r_{\mu,\phi}(s) \pi_{S|\phi}^0(ds)$ and the disintegration of the baseline prior,

$$\begin{aligned} \pi_{\theta|\phi}^\mu(A) &= \int \rho_\phi^0(A | s) r_{\mu,\phi}(s) \pi_{S|\phi}^0(ds) \\ &= \int \mathbf{1}_A(\theta) r_{\mu,\phi}(S(\theta)) \pi_{\theta|\phi}^0(d\theta). \end{aligned}$$

Hence

$$\pi_{\theta|\phi}^\mu(d\theta) = \frac{d\mu}{d\pi_{S|\phi}^0}(S(\theta)) \pi_{\theta|\phi}^0(d\theta).$$

In particular, $\pi_{\theta|\phi}^\mu \ll \pi_{\theta|\phi}^0$, and therefore

$$\begin{aligned} \text{KL} \left(\pi_{\theta|\phi}^\mu \parallel \pi_{\theta|\phi}^0 \right) &= \int \log r_{\mu,\phi}(S(\theta)) \pi_{\theta|\phi}^\mu(d\theta) \\ &= \int \log r_{\mu,\phi}(s) \mu(ds) \\ &= \text{KL} \left(\mu \parallel \pi_{S|\phi}^0 \right). \end{aligned}$$

Finally, let $q_{\theta|\phi} \ll \pi_{\theta|\phi}^0$ satisfy $S_{\#}q_{\theta|\phi} = \mu$. By the chain rule for relative entropy,

$$\begin{aligned} \text{KL} (q_{\theta|\phi} \parallel \pi_{\theta|\phi}^0) &= \text{KL} (\mu \parallel \pi_{S|\phi}^0) \\ &\quad + \int \text{KL} (q_{\theta|S,\phi}(\cdot \mid s) \parallel \rho_\phi^0(\cdot \mid s)) \mu(ds). \end{aligned}$$

The second term is nonnegative and is zero for $\pi_{\theta|\phi}^\mu$, whose conditional distribution given $S = s$ is $\rho_\phi^0(\cdot \mid s)$. Therefore, whenever the marginal divergence is finite, $\pi_{\theta|\phi}^\mu$ minimizes KL divergence over all $q_{\theta|\phi} \ll \pi_{\theta|\phi}^0$ with marginal μ .

If another $q_{\theta|\phi}$ attains the same minimum, the conditional KL term must vanish for μ -almost every s , so

$$q_{\theta|S,\phi}(\cdot \mid s) = \rho_\phi^0(\cdot \mid s) \quad \mu\text{-a.s.}$$

Together with $S_{\#}q_{\theta|\phi} = \mu$, this implies $q_{\theta|\phi} = \pi_{\theta|\phi}^\mu$, establishing uniqueness. $\square$

A.7 Proof Toy Example

Proof. Under $\Pi_{\theta|\phi}^\downarrow$ we can write

$$\theta = \phi + \lambda U,$$

where $U \in [0, 1]$ has an absolutely continuous distribution with density f_U that is weakly decreasing on $[0, 1]$. Since the likelihood depends only on ϕ , the conditional posterior of θ

given ϕ coincides with the conditional prior, so the set of posterior means is the set of prior means

$$\left\{ \mathbb{E}[\theta|\phi] : \pi_{\theta|\phi} \in \Pi_{\theta|\phi}^\downarrow \right\} = \{ \phi + \lambda \mathbb{E}[U] : f_U \text{ weakly decreasing on } [0, 1] \}.$$

It therefore suffices to characterize the possible values of $\mathbb{E}[U]$. Let F_U denote the cdf of U . Since f_U is weakly decreasing,

$$F_U(u) = \int_0^u f_U(t) dt \geq u \int_0^1 f_U(t) dt = u \quad \text{for all } u \in [0, 1],$$

where the inequality follows because the average of a decreasing density over $[0, u]$ is at least its average over $[0, 1]$. Hence $F_U(u) \geq u$, i.e. U is first-order stochastically dominated by the uniform distribution on $[0, 1]$. Therefore

$$\mathbb{E}[U] = \int_0^1 (1 - F_U(u)) du \leq \int_0^1 (1 - u) du = \frac{1}{2}.$$

So every admissible posterior mean satisfies

$$\mathbb{E}[\theta|\phi] \leq \phi + \frac{\lambda}{2}.$$

This upper bound is attained by the uniform density $f_U(u) \equiv 1$, which belongs to $\Pi_{\theta|\phi}^\downarrow$.

For the lower bound, $\mathbb{E}[U] \geq 0$ always. Moreover, for $n \geq 1$, the decreasing densities

$$f_n(u) = n \mathbf{1}\{0 \leq u \leq 1/n\}$$

belong to $\Pi_{\theta|\phi}^\downarrow$ and satisfy

$$\mathbb{E}_n[U] = \int_0^{1/n} un du = \frac{1}{2n} \rightarrow 0.$$

Hence $\inf \mathbb{E}[U] = 0$, so the closure of the set of posterior means is

$$\left[\phi, \phi + \frac{\lambda}{2} \right].$$

If one enlarges $\Pi_{\theta|\phi}^\downarrow$ to include weak limits of such priors, then the set itself equals $[\phi, \phi + \lambda/2]$. $\square$

B Shape Restrictions

This appendix shows that the monotonicity restrictions of Example 7 and Section 4 can be represented by countably many affine restrictions with continuous integrands, via the Hausdorff moment problem. For integers $n \geq 0$ and $k \geq 1$, let $p_{n,k}(x) = x^{n+1}(1-x)^k$ on $[0, 1]$, so that $q_{n,k} = p'_{n,k}$ in Example 7. Note that $p_{n,k} \geq 0$, $p_{n,k}(0) = p_{n,k}(1) = 0$, and $q_{n,k}(0) = \mathbf{1}\{n = 0\}$.

Lemma 6. *For $\mu \in \mathcal{P}([0, 1])$ the following are equivalent:*

(a) $\mu(dx) = w_0 \delta_0(dx) + f(x) dx$ for some $w_0 \in [0, 1]$ and some weakly decreasing $f : (0, 1] \rightarrow [0, \infty)$;

(b) $\int q_{n,k} d\mu \geq 0$ for all $n \geq 0$ and $k \geq 1$.

In particular, the set of distributions satisfying (a) is weakly closed. It is the weak closure of the set of distributions on $[0, 1]$ with a weakly decreasing density.

Proof. (a) implies (b). Since f is weakly decreasing, for every $t \geq 0$ the set $\{x \in (0, 1] : f(x) > t\}$ is an interval with left endpoint 0 and some right endpoint $a(t) \in [0, 1]$. Because $\int_0^a q_{n,k}(x) dx = p_{n,k}(a)$, $q_{n,k}$ is bounded, and f is integrable, Fubini's theorem gives

$$\int_0^1 q_{n,k}(x) f(x) dx = \int_0^\infty \int_0^1 q_{n,k}(x) \mathbf{1}\{f(x) > t\} dx dt = \int_0^\infty p_{n,k}(a(t)) dt \geq 0.$$

Hence $\int q_{n,k} d\mu = w_0 q_{n,k}(0) + \int_0^1 q_{n,k}(x) f(x) dx \geq 0$.

(b) implies (a). Let $m_n = (n+1) \int x^n \mu(dx)$. Expanding $p_{n,k}$ gives

$$q_{n,k}(x) = \sum_{i=0}^k (-1)^i \binom{k}{i} (n+1+i) x^{n+i}, \quad \text{hence} \quad \int q_{n,k} d\mu = \sum_{i=0}^k (-1)^i \binom{k}{i} m_{n+i}.$$

By (b), these iterated differences are nonnegative for all $n \geq 0$ and $k \geq 1$, and for $k = 0$ they equal $m_n \geq 0$. Thus $(m_n)_{n \geq 0}$ is completely monotone, and Hausdorff's moment theorem gives a probability measure λ on $[0, 1]$ with $\int w^n \lambda(dw) = m_n$ for all n ; its total mass is $m_0 = 1$. Let $W \sim \lambda$ and let U be uniform on $[0, 1]$ and independent of W . Then $\mathbb{E}[(UW)^n] = m_n/(n+1) = \int x^n \mu(dx)$ for all n . Since distributions on $[0, 1]$ are determined by their moments, μ is the distribution of UW . For a Borel set $B \subseteq (0, 1]$, Tonelli's theorem therefore gives

$$\mu(B) = \mathbb{E} \left[\mathbf{1}\{W > 0\} \int_0^1 \mathbf{1}_B(uW) du \right] = \mathbb{E} \left[\mathbf{1}\{W > 0\} W^{-1} \int_0^W \mathbf{1}_B(x) dx \right] = \int_B f(x) dx,$$

with $f(x) = \mathbb{E}[W^{-1} \mathbf{1}\{W \geq x\}]$, which is weakly decreasing on $(0, 1]$. Moreover, $\mu(\{0\}) = \mathbb{P}(W = 0) =: w_0$.

Closure. By (b), the set in question is an intersection of the sets $\{\mu : \int q_{n,k} d\mu \geq 0\}$, which are weakly closed because each $q_{n,k}$ is continuous. It contains every distribution with a weakly decreasing density. Conversely, if μ satisfies (a), then $\mu_j(dx) = (w_0 j \mathbf{1}\{0 < x \leq 1/j\} + f(x)) dx$ has a weakly decreasing density and $\mu_j \Rightarrow \mu$ as $j \rightarrow \infty$. $\square$

Let $S : K_\phi \rightarrow [0, B]$ be continuous. Applying Lemma 6 to the distribution of S/B under π shows that $S_\# \pi \in \mathcal{M}_S(\phi)$ in Example 7 if and only if

$$\int q_{n,k}(S(\theta)/B) \pi(d\theta) \geq 0 \quad \text{for all } n \geq 0, k \geq 1.$$

These are countably many restrictions whose integrands are continuous on K_ϕ , as required by Assumption 3. Any increasing sequence of finite sets J_r with $\bigcup_r J_r = J$ then satisfies Assumption 4.

Monotonicity above a threshold is covered by the same lemma. Let $S : K_\phi \rightarrow (-\infty, B]$ be continuous, let $c < B$, and set $X = \max\{(S - c)/(B - c), 0\} \in [0, 1]$. The distribution of X on $(0, 1]$ is that of $(S - c)/(B - c)$ on $(0, 1]$, while all prior mass on $\{S \leq c\}$ is mapped to the atom at zero, which Lemma 6 leaves unrestricted. Hence the distribution of S restricted to $(c, B]$ has a weakly decreasing density if and only if $\int q_{n,k}(X(\theta)) \pi(d\theta) \geq 0$ for all $n \geq 0$ and $k \geq 1$. Again, the integrands are continuous on K_ϕ .