Default Priors in Econometrics

Nicholas G. Polson
Booth School of Business, University of Chicago
ngp@chicagobooth.edu

August 2026

Abstract

Econometricians estimate high-dimensional parameters and report low-dimensional nonlinear functionals of them: an elasticity, a long-run multiplier, an impulse response, an indirect effect, a variance share. A prior that is flat in the natural parameterisation is generally not flat in the reported quantity, and where identification is weak the discrepancy does not vanish with the sample size. This paper catalogues default priors for econometrics under one criterion, taken from Bhadra, Datta, Polson and Willard (2016): *regular variation*. Regular variation is closed under the sums, maxima, products, ratios and rational functions from which econometric quantities are built, and is therefore the property that makes flatness survive the map from parameter to estimand.

We develop a tail calculus for induced priors and prove five results. The prior induced on a ratio of coefficients is Cauchy tailed with constant $p_2(0)\mathbb{E}|\theta_1|$, so diffusing the first-stage prior makes the implied prior on the elasticity *tighter* and manufactures identification. On a product, a coordinate spike at the origin compounds: the induced density grows like $\log^3$ where a bounded prior gives log. A half-Cauchy prior on a global scale induces exactly the arcsine, or Beta$(1/2, 1/2)$, prior on the population R^2, which recovers the horseshoe law as a statement about model fit and identifies R2–D2 as its reparameterisation. Any prior with positive density at the unit root implies an infinite prior mean half-life and multiplier. And the induced prior on a maximum keeps the coordinate tail index, which a purely global scale destroys.

We then catalogue some thirty priors in current use, classify each by behaviour at the origin and in the tail, and trace the consequences through regression, instrumental variables, large VARs, factor models, time-varying parameter and stochastic volatility models, discrete choice and causal machine learning. Two numerical illustrations and a set of practical recommendations close the paper.

Keywords: default priors; regular variation; global–local shrinkage; horseshoe; induced priors; weak instruments; Bayesian VAR; R^2; half-life.

1 Introduction

1.1 The problem

The Bayesian econometrician specifies $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ and reports a posterior. The parameter $\boldsymbol{\theta}$ is high dimensional: the coefficients of a vector autoregression, the loadings of a factor model, the first-stage coefficients on a hundred instruments. The quantity of economic interest is a scalar,

$$\psi = \psi(\boldsymbol{\theta}),$$

and it is nonlinear: a long-run multiplier, a half-life, a variance share, an elasticity written as a ratio, an indirect effect written as a product. Each is a many-to-one map from a large $\boldsymbol{\theta}$ to a small ψ.

The prior is judged on ψ, not on $\boldsymbol{\theta}$. A change of variables carries a Jacobian, and integrating out nuisance coordinates concentrates mass in ways unrelated to the economics. Efron (1973), in his discussion of Dawid, Stone and Zidek (1973), made the point in the normal means model $(y_i \mid \theta_i) \sim \mathcal{N}(\theta_i, 1)$, $i \leq p$. Take $\theta_i \sim \mathcal{N}(0, A)$ with A large and let $\psi = \sum_i \theta_i^2$. With $p = 100$ and $\sum_i y_i^2 = 200$, the minimum variance unbiased estimate is $\hat{\psi} = 100$ with standard error 25; the posterior mean is 300. The prior is flat in $\boldsymbol{\theta}$ and grotesque in ψ, because the induced prior $p(\psi) \propto \psi^{p/2-1} e^{-\psi/2A}$ is repelled from the origin when p is large.

This is the generic situation in econometrics, where p is large by construction and the reported number is always a functional. A twenty-variable VAR with twelve lags has $p = 4800$ coefficients and delivers an impulse response. A first stage with a hundred and fifty instruments delivers one elasticity.

1.2 The criterion

Bhadra, Datta, Polson and Willard (2016) proposed assessing whether a prior deserves the name *default* by the property of *regular variation*: $f \in \mathrm{RV}_\alpha$ if $f(\lambda x)/f(x) \to \lambda^\alpha$ for all $\lambda > 0$ as $x \to \infty$, equivalently $f(x) = x^\alpha \ell(x)$ with ℓ slowly varying. Two facts make it the right criterion here.

First, regular variation is closed under exactly the operations that build econometric functionals: sums, maxima, products, ratios, and rational functions with positive coefficients (Bingham, Goldie and Teugels, 1989, Prop. 1.5.7). A regularly varying prior on coordinates induces a regularly varying prior on ψ; flatness survives the map. Gaussian tails are rapidly varying and enjoy no such closure, so the induced prior must be recomputed case by case, which nobody does.

Second, regular variation of the prior relative to the likelihood delivers tail robustness (Dawid, 1973; O'Hagan, 1979; Andrade and O'Hagan, 2006): with polynomial prior tails and a location-normal likelihood, $(\theta \mid y) \rightsquigarrow \mathcal{N}(y, 1)$ as $|y| \to \infty$. When the data speak the prior stands aside.

Regular variation alone is not enough. Econometric problems are sparse or approximately sparse, and a spherically symmetric heavy-tailed prior such as a multivariate t is tail robust but leaves the noise coordinates unshrunk, so functionals that aggregate over coordinates are badly overstated. What is required is a prior that is simultaneously *tail robust* — regularly varying at infinity — and *sparse robust* — unbounded at the origin. Global–local scale mixtures (Polson and Scott, 2010, 2012),

$$(\theta_i \mid \lambda_i, \tau) \sim \mathcal{N}(0, \lambda_i^2 \tau^2), \qquad \lambda_i \sim p(\lambda_i), \qquad \tau \sim p(\tau), \tag{1}$$

are built to have both. The horseshoe, $\lambda_i \sim \mathcal{C}^+(0,1)$, $\tau \sim \mathcal{C}^+(0,1)$ (Carvalho, Polson and Scott, 2010), is the canonical member.

1.3 Contributions

Section 2 sets out the criterion and formalises what it means for a prior to be default for a class of functionals. Section 3 develops a tail calculus for induced priors and proves five results, each with a direct econometric reading.

(i) *Ratios.* For independent coordinates the density of $\psi = \theta_1/\theta_2$ satisfies $p_\psi(t) \sim p_2(0)\, \mathbb{E}|\theta_1|\, t^{-2}$: the tail of the estimand is governed by the denominator's prior density *at the origin*, so diffusing the first-stage prior tightens the implied prior on the elasticity and manufactures identification where the data provide none (Theorem 3.2).

(ii) *Products.* A coordinate spike compounds across channels. The induced density on an indirect effect grows like $\log^3(1/|t|)$ under global–local priors against $\log(1/|t|)$ under any prior bounded at the origin (Proposition 3.5).

(iii) *Variance shares.* A half-Cauchy prior on a global scale induces exactly the arcsine law $\mathrm{Beta}(1/2, 1/2)$ on the population R^2 as $p \to \infty$. The horseshoe's defining $\mathrm{Beta}(1/2, 1/2)$ shrinkage law is, read in the other direction, a statement about model fit, and it identifies R2–D2 as a reparameterisation rather than a rival (Theorem 3.6).

(iv) *Persistence.* If the prior has positive continuous density at the unit root, the induced priors on the long-run multiplier and on the half-life are both RV_{-2} and have infinite mean (Proposition 3.8).

(v) *Maxima.* The induced prior on $\max_i \theta_i$ keeps the coordinate tail index, but only if the coordinates are heavy tailed given the global scale. A purely global scale destroys the property: this is Efron's second pathology and the failure mode of the Minnesota prior (Proposition 3.9).

Section 4 is the catalogue: some thirty priors, classified by tail and origin behaviour. Sections 5 and 6 trace the consequences through the model classes of applied econometrics. Section 7 gives two numerical illustrations, Section 8 the recommendations, Section 9 a conclusion.

2 Regular variation and default priors

2.1 Closure

Definition 2.1. A measurable $f : \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at infinity with index α, written $f \in \mathrm{RV}_\alpha$, if $\lim_{x\to\infty} f(\lambda x)/f(x) = \lambda^\alpha$ for all $\lambda > 0$; slowly varying if $\alpha = 0$; rapidly varying if the limit is 0 or ∞. A random variable Z has a regularly varying right tail of index $\alpha > 0$ if $\bar{F}_Z(x) = \mathbb{P}(Z > x) = x^{-\alpha}\ell(x)$ with ℓ slowly varying, written $Z \in \mathrm{RV}(\alpha)$.

We use tail index α for random variables and index $-\alpha - 1$ for the corresponding density, the two being related by Karamata's theorem.

Proposition 2.2 (Closure; Bingham et al., 1989, Prop. 1.5.7)**.** (i) $f \in \mathrm{RV}_\alpha \Rightarrow f^\rho \in \mathrm{RV}_{\rho\alpha}$. (ii) $f_i \in \mathrm{RV}_{\alpha_i}$, $i = 1, 2 \Rightarrow f_1 + f_2 \in \mathrm{RV}_{\max(\alpha_1,\alpha_2)}$. (iii) *If* $f_i \in \mathrm{RV}_{\alpha_i}$ *and* $r(x_1, \ldots, x_k)$ *is a rational function with positive coefficients, then* $r(f_1, \ldots, f_k) \in \mathrm{RV}_\alpha$ *for some* α.

Part (iii) is the workhorse. Nearly every reported econometric quantity is a rational function of the underlying coefficients: the cumulative impulse response $\boldsymbol{c}^\top(\boldsymbol{I} - \boldsymbol{A}_1 - \cdots - \boldsymbol{A}_q)^{-1}\boldsymbol{b}$ is a ratio of polynomials in the autoregressive coefficients; a forecast error variance share is a ratio of quadratic forms; the Wald estimand is a ratio of linear forms.

2.2 Scale mixtures

The construction of regularly varying priors runs through normal scale mixtures.

Theorem 2.3 (Barndorff-Nielsen, Kent and Sørensen, 1982)**.** *Let* $p(\theta) = \int (2\pi v)^{-1/2} e^{-\theta^2/2v} dF(v)$ *with mixing density* $f(v) = e^{-\psi^+ v} v^{a-1} L(v)$ *as* $v \to \infty$, *where* $\psi^+ = \sup\{s : \int e^{su} F(du) < \infty\}$ *and* L *is slowly varying. Then as* $\theta \to \infty$,

$$p(\theta) \sim \begin{cases} (2\pi)^{-1/2} 2^{1/2-a} \Gamma\big(\tfrac{1}{2} - a\big)\, |\theta|^{2a-1} L(\theta^2), & \psi^+ = 0, \\ (2\psi^+)^{-a/2} L(|\theta|) |\theta|^{a-1} \exp\{-(2\psi^+)^{1/2}|\theta|\}, & \psi^+ > 0. \end{cases}$$

The dichotomy is the single most useful diagnostic in the catalogue below. If the mixing density for the variance is polynomially tailed, so $\psi^+ = 0$, the marginal on the coefficient is polynomially tailed and regularly varying. If the mixing density has any exponential decay, $\psi^+ > 0$, the marginal inherits an exponential tail and is not regularly varying, however appealing the mixing density looks. This one test separates horseshoe, Student-t, generalized double Pareto, R2–D2 and Zellner–Siow from ridge, lasso, normal–gamma, Dirichlet–Laplace and penalised complexity. Applied to (1) with $p(\lambda_i^2) = K(\lambda_i^2)^{-a-1}L(\lambda_i^2)$, Theorem 2.3 holds with index $-a$ in the mixing density and yields

$$p(\theta_i) \propto |\theta_i|^{-2a-1}L(\theta_i^2), \qquad |\theta_i| \to \infty, \tag{2}$$

so that $\theta_i \in \mathrm{RV}(2a)$: the tail index of the coefficient is twice that of the local variance.

Remark 2.4. Two checks fix (2). For $\lambda_i^2 \sim \mathrm{IG}(\nu/2, \nu/2)$ the marginal is exactly Student-t_ν, whose density decays as $|\theta|^{-\nu-1}$; here $a = \nu/2$ and (2) gives $|\theta|^{-\nu-1}$. For the horseshoe, $\lambda \sim \mathcal{C}^+(0,1)$ gives $p(\lambda^2) = \pi^{-1}(\lambda^2)^{-1/2}(1+\lambda^2)^{-1} \sim \pi^{-1}(\lambda^2)^{-3/2}$, so $a = 1/2$ and $p(\theta) \sim c\,\theta^{-2}$ with $c = 2/(\pi\sqrt{2\pi}) = (2/\pi^3)^{1/2}$. The same constant governs the origin: $p(\theta) \sim c\log(1/|\theta|)$ as $\theta \to 0$. The horseshoe is thus pinned at both ends by a single number, and c reappears in Propositions 3.3 and 3.5 below.

2.3 What it means to be default

We can now state the criterion the rest of the paper applies. Write p_ψ for the prior induced on $\psi(\boldsymbol{\theta})$ by $p(\boldsymbol{\theta})$.

Definition 2.5 (RV-default)**.** Let $\mathcal{G}$ be a class of measurable functionals $\psi : \mathbb{R}^p \to \mathbb{R}$. A proper prior $p(\boldsymbol{\theta})$ is *default for* $\mathcal{G}$ if for every $\psi \in \mathcal{G}$ the induced prior p_ψ is (a) proper, (b) regularly varying at infinity with finite index, and (c) bounded away from zero on a neighbourhood of every attainable value of ψ, including boundary values at which ψ is not identified.

Condition (b) is tail robustness in the estimand: it guarantees, by Dawid (1973), that the posterior for ψ follows the integrated likelihood when the data are informative. Condition (c) is the requirement that the prior not resolve by fiat a question the data cannot answer; it is what fails when a prior is repelled from the origin, and it is the condition that does the work in weakly identified problems. Condition (a) rules out the reference and Jeffreys routes in hierarchies and in model comparison.

Proposition 2.6. *Let $\theta_1, \dots, \theta_p$ be independent with common marginal $p(\theta_i)$ satisfying $p(\theta_i) = |\theta_i|^{-\alpha-1}\ell(|\theta_i|)$ as $|\theta_i| \to \infty$ for some $\alpha > 0$ and slowly varying ℓ, with $\int p = 1$ and $\liminf_{\theta\to 0} p(\theta) > 0$. Let $\mathcal{G}$ be the class of rational functions of $\boldsymbol{\theta}$ with positive coefficients. Then* (a) *and* (b) *of Definition 2.5 hold for every $\psi \in \mathcal{G}$.*

Proof. Properness is immediate since ψ is a measurable function of a proper random vector. For (b), the hypothesis states $\theta_i \in \mathrm{RV}(\alpha)$; Proposition 2.2(iii) transfers regular variation through any rational function with positive coefficients. □

Condition (c) is not implied in general and must be checked functional by functional. That is what the next section does for the functionals that matter in econometrics, and in each case the condition turns out to be exactly the requirement that the coordinate prior have an unbounded — or at least strictly positive — density at the origin. The combination of (b) and (c), tail robustness and sparse robustness, is what singles out the global–local class.

Table 1: Efron's four problems and their econometric counterparts.

Problem	Functional	Econometric counterparts
Sum of squares	$\psi = \sum_i \theta_i^2$	R^2; forecast error variance decompositions; long-run variance; portfolio variance $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$; signal-to-noise ratio in state space models; Wald and J statistics.
Maximum	$\psi = \max_i \theta_i$	Peak impulse response; best fund or strategy in a cross-section; maximum Sharpe ratio; largest subgroup treatment effect; dominant eigenvalue of a companion matrix.
Product	$\psi = \theta_1 \theta_2$	Indirect and mediation effects; two-stage pass-through; interaction effects; loading times factor variance.
Ratio	$\psi = \theta_1 / \theta_2$	Elasticities; the Wald IV estimand and LATE; long-run multiplier $\sum_j \beta_j / (1 - \sum_j \alpha_j)$; half-life $\log(1/2)/\log \rho$; willingness to pay; velocity.

3 A tail calculus for econometric functionals

Efron's four problems — sum of squares, maximum, product, ratio — are not curiosities. Each is an econometric functional in disguise, as Table 1 records. This section computes the induced priors.

3.1 Tail arithmetic

Proposition 3.1. *Let $\theta_1, \ldots, \theta_p$ be independent and identically distributed with $\theta_i \in \mathrm{RV}(\alpha)$, $\alpha > 0$. Then, as $x \to \infty$,*

(i) maximum:	$\mathbb{P}\big(\max_i \theta_i > x\big) \sim p\, \bar{F}(x)$,	$\max_i \theta_i \in \mathrm{RV}(\alpha)$;
(ii) sum:	$\mathbb{P}\big(\sum_i \theta_i > x\big) \sim p\, \bar{F}(x)$,	$\sum_i \theta_i \in \mathrm{RV}(\alpha)$;
(iii) sum of squares:		$\sum_i \theta_i^2 \in \mathrm{RV}(\alpha/2)$;
(iv) product:		$\theta_1 \theta_2 \in \mathrm{RV}(\alpha)$ *up to a factor* $\log x$.

Proof. (i) $\mathbb{P}(\max_i \theta_i \le x) = F(x)^p$, so $1 - F(x)^p \sim p\bar{F}(x)$. (ii) Regularly varying tails with $\alpha > 0$ are subexponential, for which $\mathbb{P}(\sum_i \theta_i > x) \sim p\bar{F}(x)$ (Bingham et al., 1989, §1.3). (iii) $\mathbb{P}(\theta_i^2 > x) = \bar{F}(\sqrt{x}) + F(-\sqrt{x})$, which is $\mathrm{RV}(\alpha/2)$ by Proposition 2.2(i); apply (ii). (iv) By Breiman's lemma, if $\mathbb{E}|\theta_2|^{\alpha+\epsilon} < \infty$ then $\mathbb{P}(\theta_1\theta_2 > x) \sim \mathbb{E}[\theta_2^\alpha]\bar{F}(x)$; when both factors have the same index the moment condition fails and the tail acquires a $\log x$ factor, which is slowly varying, so the index is unchanged. □

Two consequences deserve emphasis. The tail index is *preserved* by maxima and sums and *halved* by squaring. For the horseshoe, $\alpha = 1$ by Remark 2.4, so $\sum_i \theta_i^2 \in \mathrm{RV}(1/2)$: the induced prior on a variance share has infinite mean and a tail $\psi^{-1/2}$. Corollary 3.7 below confirms this from an exact calculation.

3.2 Ratios: the Fieller–Creasy problem and weak instruments

The Wald estimand $\psi = \pi_1/\pi_2$, with π_1 the reduced form and π_2 the first stage, is Fieller's problem (Fieller, 1940; Creasy, 1954). The following is the central result of this section.

Theorem 3.2 (Ratio)**.** *Let $\theta_1 \perp\!\!\!\perp \theta_2$ have densities p_1, p_2 with p_2 bounded and continuous at the origin, and let $\mathbb{E}|\theta_1| < \infty$. Then the density of $\psi = \theta_1/\theta_2$ satisfies*

$$p_\psi(t) \;=\; t^{-2}\int |u|\, p_1(u)\, p_2(u/t)\, du \;\sim\; p_2(0)\, \mathbb{E}|\theta_1|\; t^{-2}, \qquad t \to \infty, \tag{3}$$

so that $p_\psi \in \mathrm{RV}_{-2}$ and $\psi \in \mathrm{RV}(1)$ whenever $p_2(0) > 0$. If $p_2(0) = 0$ then $p_\psi(t) = o(t^{-2})$; if $\mathbb{E}|\theta_1| = \infty$ then $t^2 p_\psi(t) \to \infty$.

Proof. The ratio density is $p_\psi(t) = \int |v| p_1(tv) p_2(v)\, dv$, since $\theta_1 = \psi\theta_2$ has Jacobian $|\theta_2|$. Substituting $u = tv$ gives the exact identity in (3). As p_2 is bounded and $\mathbb{E}|\theta_1| < \infty$, the integrand is dominated by $\|p_2\|_\infty |u| p_1(u)$, which is integrable; $p_2(u/t) \to p_2(0)$ pointwise by continuity, and dominated convergence gives the limit. The two edge cases follow by the same argument and by Fatou's lemma respectively. □

Four remarks. First, the result is a statement about the *denominator*: the tail weight of the estimand is set by how much prior mass sits near $\pi_2 = 0$. Making the prior on π_1 heavier changes the constant, not the rate.

Second, $p_2(0)\mathbb{E}|\theta_1|$ is a natural measure of prior information about identification. A Gaussian prior $\pi_2 \sim \mathcal{N}(0, A)$ has $p_2(0) = (2\pi A)^{-1/2} \to 0$ as $A \to \infty$: making the first-stage prior *more* diffuse makes the induced prior on the estimand *lighter* tailed, hence more confident. This is Efron's phenomenon in its sharpest econometric form. A prior assigning negligible mass to a neighbourhood of zero reports a well-identified elasticity from data that identify nothing.

Third, the reference prior for this problem, $p(\psi) \propto (1+\psi^2)^{-1/2}$ (Liseo, 1993), is RV_{-1} and improper. Theorem 3.2 says any proper prior with $p_2(0) \in (0, \infty)$ and $\mathbb{E}|\theta_1| < \infty$ gives RV_{-2}: the proper default is one power lighter than the reference answer, which is the price of properness. The horseshoe violates both hypotheses — $p_2(0) = \infty$ and $\mathbb{E}|\theta_1| = \infty$ — and the following makes precise where it lands.

Proposition 3.3 (Horseshoe ratio)**.** *Let $\theta_1 \perp\!\!\!\perp \theta_2$ be standard horseshoe, so $p_i(\theta) \sim c\,\theta^{-2}$ as $|\theta| \to \infty$ and $p_i(\theta) \sim c \log(1/|\theta|)$ as $\theta \to 0$, with $c = (2/\pi^3)^{1/2}$. Then, as $t \to \infty$,*

$$p_{\theta_1/\theta_2}(t) \;\sim\; c^2\, t^{-2} \log^2 t.$$

Proof. By symmetry $p_\psi(t) = 2\int_0^\infty v\, p_1(tv) p_2(v)\, dv$. Split at $v = 1/t$ and $v = 1$. On $(1/t, 1)$, $p_1(tv) \sim c(tv)^{-2}$ and $p_2(v) \sim c\log(1/v)$, so with $u = \log(1/v)$ the contribution is $c^2 t^{-2}\int_0^{\log t} u\, du = c^2 t^{-2}\log^2 t/2$. On $(0, 1/t)$ both factors are logarithmic and the contribution is $O(t^{-2}\log t)$; on $(1, \infty)$ both are quadratic and the contribution is $O(t^{-2})$. Doubling the dominant term gives the result. □

The index is still -2; the horseshoe buys a slowly varying $\log^2 t$ of extra tail weight over a prior with bounded density at the origin. Numerically, $t^2 p_\psi(t)$ equals 3.217 at $t = 10^3$ against $c^2 \log^2 t = 3.078$, and 5.235 at $t = 10^4$ against 5.472.

Fourth, weak identification is by definition the regime in which the likelihood does not overwhelm the prior, so the choice of prior is decisive there and immaterial elsewhere. Table 5 shows exactly this pattern.

Corollary 3.4 (Identification neutrality)**.** *Under the hypotheses of Theorem 3.2, condition* (c) *of Definition 2.5 holds for $\psi = \theta_1/\theta_2$ if and only if p_2 is bounded away from zero on a neighbourhood of the origin. Priors that vanish at $\pi_2 = 0$ — including any prior placing π_2 on a set bounded away from zero, and, in the diffuse limit, the Gaussian — fail it.*

3.3 Products: mediation and two-stage pass-through

Where the ratio problem is about the tail, the product problem is about the origin. An indirect effect $\psi = \theta_1\theta_2$ is zero whenever either channel is inactive, so the honest report under weak evidence is a posterior massed at zero, and what matters is how fast the induced prior accumulates there.

Proposition 3.5 (Product)**.** *Let $\theta_1 \perp\!\!\!\perp \theta_2$ have symmetric densities and let $\psi = \theta_1\theta_2$, so that $p_\psi(t) = 2\int_0^\infty p_1(t/v)p_2(v)v^{-1}dv$. As $t \to 0$:*

(i) *if p_i is bounded and continuous at the origin with $p_i(0) = b_i > 0$, then $p_\psi(t) \sim 2b_1b_2\log(1/|t|)$;*

(ii) *if $p_i(\theta) \sim c\log(1/|\theta|)$, as for the horseshoe with $c = (2/\pi^3)^{1/2}$, then $p_\psi(t) \sim \frac{1}{3}c^2\log^3(1/|t|)$.*

Proof. Write $L = \log(1/t)$ and $u = \log(1/v)$, so $du = -dv/v$ and $v \in (t, 1)$ corresponds to $u \in (0, L)$; this range carries the mass as $t \to 0$. In case (i), $p_1(t/v) \to b_1$ and $p_2(v) \to b_2$ uniformly on $u \in (\epsilon L, (1-\epsilon)L)$, so the integral is $b_1b_2\int_0^L du = b_1b_2L$. In case (ii), $p_1(t/v) \sim c(L-u)$ and $p_2(v) \sim cu$, so the integral is $c^2\int_0^L (L-u)u\,du = c^2L^3/6$. Multiply by 2 in each case. □

The gap is two powers of log. Numerically, at $t = 10^{-7}$ the horseshoe product density is 94.6 against $\frac{1}{3}c^2L^3 = 90.0$, and against $\pi^{-1}L = 5.1$ for a pair of standard normals. The spike at the origin is what buys the concentration, and it compounds multiplicatively across the two channels: a mediation analysis with two weakly identified stages is precisely where the choice of prior does the most work. Table 5 below records the consequence — a threefold narrower interquartile range for the indirect effect, and a posterior median of exactly zero where the Gaussian prior reports a nonzero effect.

3.4 Variance shares: the arcsine law for R^2

Efron's first problem is the variance share. Let $\theta_i \mid \tau \sim \mathcal{N}(0, \tau^2)$ independently, $i \le p$, with $\tau \sim \mathcal{C}^+(0, 1)$, and let

$$\psi = \sum_{i=1}^{p} \theta_i^2, \qquad R^2 = \frac{\psi}{\psi + p},$$

so that R^2 is the share of variation attributable to signal when each coordinate carries unit noise variance. The following is exact in its first step and asymptotic only in its second.

Theorem 3.6 (Arcsine law)**.** *Let $\kappa = (1+\tau^2)^{-1}$. If $\tau \sim \mathcal{C}^+(0,1)$ then $\kappa \sim \mathrm{Beta}(1/2, 1/2)$ exactly. Consequently $R^2 \Rightarrow \mathrm{Beta}(1/2, 1/2)$ as $p \to \infty$, with density*

$$p(r) = \frac{1}{\pi}\, r^{-1/2}(1-r)^{-1/2}, \qquad 0 < r < 1.$$

Proof. Write $s = \tau^2$. If $\tau \sim \mathcal{C}^+(0,1)$ then $p(s) = \pi^{-1}s^{-1/2}(1+s)^{-1}$. The map $s \mapsto r = s/(1+s)$ has $s = r/(1-r)$ and $ds/dr = (1-r)^{-2}$, so

$$p(r) = \frac{1}{\pi}\left(\frac{r}{1-r}\right)^{-1/2}(1-r)\,(1-r)^{-2} = \frac{1}{\pi}r^{-1/2}(1-r)^{-1/2},$$

which is $\mathrm{Beta}(1/2, 1/2)$; since $\kappa = 1 - r$ and the arcsine density is symmetric about $1/2$, κ has the same law. For the second claim, $\psi \overset{d}{=} \tau^2\chi_p^2$ with $\tau \perp\!\!\!\perp \chi_p^2$, so $\psi/p = \tau^2(\chi_p^2/p)$ and $\chi_p^2/p \to 1$ in probability. By Slutsky and the continuous mapping theorem applied to $x \mapsto x/(1+x)$, $R^2 = (\psi/p)/(1+\psi/p) \Rightarrow \tau^2/(1+\tau^2)$, whose law is $\mathrm{Beta}(1/2, 1/2)$ by the first part. □

Corollary 3.7. *Under the hypotheses of Theorem 3.6, the limiting induced prior on $\psi = \sum_i \theta_i^2$ is $p(\psi) \propto \psi^{-1/2}(\psi + p)^{-1}$, which is $\mathrm{RV}_{-3/2}$, so $\psi \in \mathrm{RV}(1/2)$ and $\mathbb{E}[\psi] = \infty$.*

Proof. Invert $r = \psi/(\psi+p)$, with $dr/d\psi = p(\psi+p)^{-2}$, in the arcsine density: $p(\psi) = \pi^{-1}\{\psi/(\psi+p)\}^{-1/2}\{p/(\psi+p)\}^{-1/2} \cdot p(\psi+p)^{-2} = \pi^{-1}\sqrt{p}\,\psi^{-1/2}(\psi+p)^{-1}$. □

The density is strictly positive at $\psi = 0$, which is condition (c) of Definition 2.5 and the reason the posterior can report a variance share of zero. Bhadra et al. (2016) obtain the same expression by Laplace approximation; Theorem 3.6 shows it is exact in the limit and identifies the law behind it.

The arcsine law is the horseshoe's namesake: Carvalho, Polson and Scott (2010) named the prior for the $\mathrm{Beta}(1/2, 1/2)$ density of the shrinkage weight $\kappa_i = (1 + \lambda_i^2\tau^2)^{-1}$, whose two poles at 0 and 1 express indifference between no shrinkage and total shrinkage. Theorem 3.6 reads that fact in the other direction and in the econometrician's units: the half-Cauchy global scale is the prior that is agnostic about model fit, placing the arcsine law on R^2. Three points follow.

First, the theorem supplies the missing justification for the half-Cauchy default on a global scale. The usual argument (Gelman, 2006; Polson and Scott, 2012b) is that the half-Cauchy behaves well at both ends of the scale. Theorem 3.6 says something stronger and checkable: it is the prior that makes R^2 arcsine. The convergence is fast — at $p = 50$ the simulated quartiles of R^2 are $(0.142, 0.494, 0.852)$ against arcsine values $(0.146, 0.500, 0.854)$.

Second, this places the R2–D2 prior of Zhang et al. (2022) in the catalogue exactly. That prior starts from a $\mathrm{Beta}(a, b)$ prior on R^2 and induces the implied prior on coefficients; Theorem 3.6 shows the half-Cauchy global scale is the special case $a = b = 1/2$. The R2–D2 construction is therefore not an alternative to global–local shrinkage but a reparameterisation of it in the units the economist reports, which is exactly the recommendation of Section 8.

Third, the contrast with the diffuse Gaussian is now explicit. Under $\theta_i \sim \mathcal{N}(0, A)$ with A fixed, $R^2 \to A/(1+A)$ degenerately as $p \to \infty$: a point mass, and at $A = 300$ a point mass at 0.997. That is Efron's 300, restated.

3.5 Persistence: long-run multipliers and half-lives

Consider $y_t = \rho y_{t-1} + \varepsilon_t$ with ρ given a prior with density p_ρ continuous at 1. The reported quantities are the long-run multiplier $M = (1-\rho)^{-1}$ and the half-life $H = \log(1/2)/\log\rho$.

Proposition 3.8. *If $p_\rho(1) = c \in (0, \infty)$ then, as $t \to \infty$,*

$$p_M(t) \sim c\,t^{-2}, \qquad p_H(t) \sim c \log 2\; t^{-2},$$

so $M, H \in \mathrm{RV}(1)$ and $\mathbb{E}[M] = \mathbb{E}[H] = \infty$ under the prior.

Proof. For M: $\rho = 1 - 1/t$, $|d\rho/dt| = t^{-2}$, so $p_M(t) = p_\rho(1-1/t)t^{-2} \sim ct^{-2}$. For H: $\rho = 2^{-1/t}$, $|d\rho/dt| = (\log 2)t^{-2}e^{-(\log 2)/t} \sim (\log 2)t^{-2}$, and $p_H(t) = p_\rho(2^{-1/t})|d\rho/dt| \sim c\log 2\, t^{-2}$. In both cases $\int^\infty t \cdot t^{-2}dt$ diverges. □

The proposition is elementary and its content is not. Any prior on ρ that does not exclude the unit root implies an infinite prior mean half-life and an infinite prior mean multiplier. Reporting a posterior mean half-life is meaningful only because the likelihood truncates the tail, and how much it truncates depends on the sample size and on the local-to-unity parameter. Two priors on ρ that are visually indistinguishable can differ by an order of magnitude in the implied prior median half-life, because that median is governed by p_ρ on a shrinking neighbourhood of 1.

This also explains why the unit root debate (Sims and Uhlig, 1991; Phillips, 1991) was conducted at cross purposes. Phillips's objection to the flat prior was that the Jeffreys prior for this model is far from flat near unity, so a flat prior is informative *against* a unit root. The functional-side statement is complementary and sharper: whatever the prior on ρ, it is the value of its density in a neighbourhood of 1 that determines the entire tail of the reported persistence measure.

3.6 Maxima: why a global scale is not enough

Proposition 3.9. *Let* $(\theta_i \mid \lambda_i, \tau) \sim \mathcal{N}(0, \lambda_i^2\tau^2)$ *as in* (1). (i) *If* $p(\lambda_i^2) \propto (\lambda_i^2)^{-a-1}L(\lambda_i^2)$ *with* $a > 0$, *then by* (2) *the conditional marginal satisfies* $\theta_i \mid \tau \in \mathrm{RV}(2a)$, *and by Proposition 3.1(i)* $\max_i \theta_i \mid \tau \in \mathrm{RV}(2a)$ *with* $\mathbb{P}(\max_i \theta_i > x \mid \tau) \sim p\,\bar{F}_\tau(x)$*: the tail index is that of a single coordinate, independent of* p. (ii) *If instead* $\lambda_i \equiv 1$, *so shrinkage is purely global, the coordinates are Gaussian given* τ, $\mathbb{P}(\max_i \theta_i > x \mid \tau) \le p\exp(-x^2/2\tau^2)$, *and the conditional tail is rapidly varying.*

The proof is immediate from the union bound and the Gaussian tail bound. The econometric content is Efron's second example. Under (ii) the marginal on $\max_i \theta_i$ is heavy tailed only through the mixing over τ, and the posterior for τ is determined by all p coordinates jointly. With $p-1$ coordinates at zero and one at 10, that posterior concentrates near zero and the single large coefficient is shrunk along with the noise: Section 7 finds a posterior mean of 4.81 for a coefficient the data place at 9.47. Under (i) the local λ_i decouples the coordinates and the large one escapes.

This is the failure mode of the Minnesota prior in a large VAR, of the James–Stein estimator, and of any procedure that selects one penalty by cross-validation and applies it uniformly. Loosening the global scale does not repair it, because loosening breaks the variance share. The two functionals pull in opposite directions and only local scales serve both.

4 A catalogue of default priors

Table 2 covers priors on coefficients, Table 3 priors on scales. For each we record the mixing representation, the marginal tail, behaviour at the origin, and a verdict against Definition 2.5.

4.1 Reading the catalogue

The g-prior family is a microcosm. Zellner's g-prior with fixed g is Gaussian and rapidly varying. Its two defects — Bartlett's paradox, in which $g \to \infty$ drives the Bayes factor to the null regardless of the data, and the information paradox, in which the Bayes factor is bounded as the F statistic diverges — are the same failure at infinity. The repairs of Zellner and Siow (1980) and Liang et al. (2008) are to mix over g with a polynomially tailed density, producing a Cauchy-like regularly varying marginal. Zellner–Siow is the horseshoe idea avant la lettre, restricted to the global scale, and Theorem 2.3 explains why the same fix cures both paradoxes.

A spike is not a default. Normal–gamma, Dirichlet–Laplace, spike-and-slab lasso and penalised complexity priors deliver sparsity and all attenuate large coefficients, since by Theorem 2.3 exponential mixing implies exponential marginal tails. For the horseshoe the posterior mean satisfies $\mathbb{E}(\theta \mid y) \to y$; for the Laplace the bias $|\mathbb{E}(\theta \mid y) - y| \approx \sqrt{2}/\tau$ does not vanish (Carvalho et al., 2010). If the object of interest is the large coefficient — and in econometrics it usually is — this is the wrong trade. Simpson et al. (2017) note that the exponential tail is the one setting in which their construction is problematic.

Table 2: Priors on coefficients. "RV" is regular variation of the marginal at infinity; "spike" is an unbounded density at the origin. A default requires both.

Prior	Mixing / form	Marginal tail $p(\theta)$	RV	Spike	Verdict
Flat, $p(\theta) \propto 1$	improper	constant	—	no	fails for $\psi(\boldsymbol{\theta})$
Jeffreys	$\lvert I(\theta)\rvert^{1/2}$	model dependent	—	no	marginalisation paradox
Ridge / $\mathcal{N}(0, A)$	degenerate at A	$e^{-\theta^2/2A}$	no	no	not default
Zellner g-prior	$\mathcal{N}(0, g\sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1})$	Gaussian	no	no	Bartlett, information paradox
Zellner–Siow	$g \sim \mathrm{IG}(1/2, 1/2)$	mv Cauchy; marginal $\lvert\theta\rvert^{-2}$	yes	no	default for testing
Hyper-g	$g/(1+g) \sim \mathrm{Be}(1, a/2-1)$	polynomial	yes	no	default for testing
Minnesota	$\mathcal{N}(0, \lambda^2/\ell^2)$, ℓ = lag	Gaussian	no	no	pure global; fails maxima
Hierarchical Minnesota	λ^2 given a prior	Gaussian given λ	no[†]	no	better; still no local
Laplace / Bayesian lasso	$\lambda_i^2 \sim \mathrm{Exp}$	$e^{-\lvert\theta\rvert/b}$	no	no	attenuates large signals
Bayesian bridge ($\alpha < 1$)	stable mixing	$e^{-\lvert\theta\rvert^\alpha}$	no	yes	subexponential, not RV
Normal–gamma	$\lambda_i^2 \sim \mathrm{Ga}(a, b)$	Bessel, $e^{-c\lvert\theta\rvert}$	no	yes ($a < 1/2$)	spike but light tail
Dirichlet–Laplace	Dirichlet × Laplace	exponential	no	yes	spike but light tail
Normal–inverse-Gaussian	inverse Gaussian	exponential	no	yes	spike but light tail
Student-t_ν	$\lambda_i^2 \sim \mathrm{IG}(\nu/2, \nu/2)$	$\lvert\theta\rvert^{-(\nu+1)}$	yes	no	tail robust, not sparse
Cauchy	$\nu = 1$	$\lvert\theta\rvert^{-2}$	yes	no	tail robust, not sparse
Generalized double Pareto	Laplace–gamma	$\lvert\theta\rvert^{-(\alpha+1)}$	yes	yes	default
Three-parameter beta	$\mathrm{TPB}(a, b, \phi)$	$\lvert\theta\rvert^{-2b-1}$	yes	yes ($a \le 1/2$)	default
Horseshoe	$\lambda_i \sim \mathcal{C}^+(0, \tau)$	$\lvert\theta\rvert^{-2}$	yes	yes (log)	**recommended default**
Horseshoe$^+$	$\lambda_i \sim \mathcal{C}^+(0, \tau\eta_i)$	$\lvert\theta\rvert^{-2}\log\lvert\theta\rvert$	yes	yes ($\log^2$)	ultra-sparse default
R2–D2	beta-prime on R^2	polynomial	yes	yes	default; interpretable
Horseshoe-like	closed-form variant	$\lvert\theta\rvert^{-2}$	yes	yes	default; fast
Spike-and-slab, normal slab	$\pi\delta_0 + (1-\pi)\mathcal{N}$	Gaussian	no	atom	slab tail too light
Spike-and-slab, Cauchy slab	$\pi\delta_0 + (1-\pi)\mathcal{C}^+$	$\lvert\theta\rvert^{-2}$	yes	atom	default; combinatorial
Spike-and-slab lasso	two Laplace scales	exponential	no	yes	attenuates large signals

[†] Conditionally Gaussian; regular variation of the marginal on θ requires a polynomially tailed prior on λ^2, which standard implementations do not use. See Section 5.3.

Table 3: Priors on scale and variance parameters. For scales the binding failure is at the origin, not at infinity: a prior repelled from zero manufactures variation that is not in the data.

Prior	Form	Tail	Origin	Verdict
$p(\sigma) \propto 1/\sigma$	Jeffreys	σ^{-1}	σ^{-1}	improper; fine for one scale
Inverse-gamma(ϵ, ϵ)	$v^{-1-\epsilon}e^{-\epsilon/v}$	$\mathrm{RV}_{-1-\epsilon}$	repelled from 0	**avoid** in hierarchies
Inverse-gamma(a, b) fixed	conjugate	RV_{-1-a}	repelled from 0	convenient, not default
Half-Cauchy $\mathcal{C}^+(0, A)$	$2/\{\pi A(1 + (\sigma/A)^2)\}$	RV_{-2}	finite, positive	**recommended default**
Half-t_ν	ν small	$\mathrm{RV}_{-\nu-1}$	finite, positive	default; $\nu \in \{1, 2\}$
Uniform on σ	$p(\sigma) \propto 1$ on $(0, U)$	truncated	finite	acceptable; U arbitrary
Uniform on σ^2	$p(v) \propto 1$	—	finite	improper posterior risk
Penalised complexity	$\lambda\sigma^{-1/2}e^{-\lambda\sqrt{\sigma}}$	exponential	spike	good at 0, light tail
Half-Cauchy squared	horseshoe$^+$ scale	$\mathrm{RV}_{-2}\log$	stronger spike	ultra-sparse
Beta-prime	$\mathrm{BP}(a, b)$	RV_{-b-1}	tunable	flexible default

Heavy tails alone are not a default either. Independent Cauchy or multivariate t priors are tail robust and fail condition (c) of Definition 2.5 for aggregate functionals: they do not shrink noise, so by Proposition 3.1(ii) the $p-1$ noise coordinates accumulate and the variance share is overstated.

Only the intersection works. Horseshoe, horseshoe$^+$, horseshoe-like, generalized double Pareto, three-parameter beta, R2–D2 and Cauchy-slab spike-and-slab pass both tests. All are normal scale mixtures, hence all are tractable by the same data augmentation.

5 High-dimensional econometrics

5.1 Regression and predictive regressions

For $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with p comparable to or exceeding n, the recommendation is a global–local prior with the global scale learned. Econometric designs are strongly correlated — macro predictors, characteristics in the cross-section of returns, lags of a persistent series — and this matters: ridge is optimal against a dense truth in an orthogonal design and can be badly suboptimal against a sparse truth in a correlated one. Bhadra et al. (2019b) characterise the prediction risk of horseshoe regression and show it dominates ridge across a wide range of sparsity and signal strength.

In return predictability the reported quantity is not $\boldsymbol{\beta}$ but the implied R^2 or the certainty-equivalent gain from timing. Theorem 3.6 says the half-Cauchy global scale already implies an arcsine prior on R^2; if the applied question concerns a small R^2, as it does in that literature, then the arcsine, with its pole at zero, is a defensible default and its Beta(a, b) generalisations let the researcher say more.

5.2 Instrumental variables with many instruments

Many instruments is a shrinkage problem in the first stage (Chamberlain and Imbens, 2004). Hahn, He and Lopes (2018) shrink through a factor model; Belloni et al. (2012) select by lasso. Theorem 3.2 makes the requirements sharp, and they are in tension:

(i) shrink the irrelevant instruments hard, or the estimand inherits many-instrument bias;

(ii) do *not* shrink the relevant ones, since attenuating π_2 inflates π_1/π_2;

(iii) keep $p_2(0) > 0$, so that genuine non-identification is reported rather than resolved.

Requirement (ii) is tail robustness and rules out lasso and normal–gamma first stages. Requirement (iii) is Corollary 3.4 and rules out t and Cauchy priors, and a fortiori Gaussian ones, without a spike. Only global–local priors with regularly varying tails satisfy all three. The argument is about the estimand, not about estimation of $\boldsymbol{\pi}$, which is why it does not follow from the usual optimality results.

5.3 Large Bayesian vector autoregressions

The Minnesota prior (Doan et al., 1984; Litterman, 1986) is, in this language, pure global shrinkage with a deterministic decay schedule: coefficient (i, j, ℓ) receives prior variance proportional to λ^2/ℓ^2. It is conditionally Gaussian and rapidly varying, so Proposition 3.9(ii) applies.

Two developments improved it, and both are Efron examples. First, Giannone, Lenza and Primiceri (2015) give the tightness λ a prior and integrate it out. This is the repair of the *first* example: by Theorem 3.6, giving the global scale a heavy-tailed prior converts the induced prior on the variance share from a degenerate point mass into the arcsine law. That is why the hierarchical treatment improves forecasts so reliably, and Bańbura et al. (2010) document the same phenomenon through the sensitivity of forecast performance to tightness.

Second, and less widely adopted, is the addition of local scales. Proposition 3.9 warns that no single λ, however chosen, can simultaneously shrink the many irrelevant coefficients of a large VAR and leave the few large ones alone; in practice this appears as over-smoothed impulse responses and understated persistence. The literature has moved this way — Huber and Feldkircher (2019) with normal–gamma, Korobilis (2013) with variable selection, Chan (2021) with adaptive hierarchical Minnesota priors — but the catalogue suggests going further, to a Minnesota-structured global–local prior,

$$(\beta_{ij\ell} \mid \lambda_{ij\ell}, \tau) \sim \mathcal{N}\left(0, \frac{\lambda_{ij\ell}^2 \tau^2 s_i^2}{\ell^2 s_j^2}\right), \qquad \lambda_{ij\ell} \sim \mathcal{C}^+(0,1), \quad \tau \sim \mathcal{C}^+(0,1), \tag{4}$$

which keeps the economically motivated decay, learns the tightness, and adds regularly varying local scales. The normal–gamma choice achieves the spike but, by Theorem 2.3, not the tail; replacing gamma mixing by half-Cauchy mixing costs nothing computationally and delivers both.

This matters because the reported object is never $\boldsymbol{\beta}$. It is $\boldsymbol{e}_i^\top \boldsymbol{\Phi}_h \boldsymbol{e}_j$, or $\boldsymbol{c}^\top (\boldsymbol{I} - \sum_\ell \boldsymbol{A}_\ell)^{-1} \boldsymbol{b}$, or a variance share — rational functions to which Proposition 2.2(iii) applies and Gaussianity does not. In the univariate case Proposition 3.8 is exact: the induced prior on the multiplier is Cauchy tailed.

5.4 Structural identification

A distinct instance arises in structural VARs identified by sign restrictions. Baumeister and Hamilton (2015) showed that the conventional Haar prior over orthogonal rotations, apparently uninformative, induces a sharply informative prior on structural impulse responses and elasticities that does not vanish asymptotically because the model is set identified. This is Efron's problem in its purest form. The robust-Bayes response of Giacomini and Kitagawa (2021), reporting the range of posteriors over all priors consistent with the restrictions, is complementary rather than

competing: the default-prior contribution is to insist that whatever prior is placed on the reduced form, the induced prior on the reported response be simulated, inspected, and checked for regular variation.

5.5 Covariance, factors, time variation and panels

For $\boldsymbol{\Sigma}$ or $\boldsymbol{\Sigma}^{-1}$ in high dimension the reported quantities — portfolio weights $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, the efficient frontier, network centrality — are again functionals. The inverse Wishart is the matrix analogue of inverse-gamma(ϵ, ϵ): one degrees-of-freedom parameter governs every eigenvalue, and it is repelled from singularity. Global–local shrinkage on loadings in $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{\Omega}$ respects the fact that a few loadings are large and lets the number of factors be determined by shrinkage rather than by a discrete choice.

In $\beta_t = \beta_{t-1} + \eta_t$, $\eta_t \sim \mathcal{N}(0, \omega^2)$, the parameter ω^2 governs whether there is time variation at all. Inverse-gamma(ϵ, ϵ) is repelled from $\omega^2 = 0$ and reports time variation whether or not it is present, which is then given an economic interpretation (Jacquier et al., 1994; Kim et al., 1998; Primiceri, 2005). The non-centred parameterisation of Frühwirth-Schnatter and Wagner (2010), $\beta_t = \beta + \omega\tilde{\beta}_t$, places the prior on ω rather than ω^2 and admits $\omega = 0$; with a half-Cauchy or horseshoe prior on ω this is a genuine default that can report no time variation when that is the truth. This is the clearest case in econometrics where the default prior changes the qualitative conclusion, and it is entirely a failure at the origin. The same reasoning applies to unit effects in panels, where the functional of interest is often the dispersion, the maximum, or a quantile of the effects.

6 Nonlinear and structural models

Discrete choice. Pólya–Gamma augmentation (Polson, Scott and Windle, 2013) renders the conditional posterior Gaussian and makes any normal scale mixture prior, horseshoe included, immediately implementable. The functionals are sharper than in the linear case: marginal effects, predicted shares, and willingness to pay, which is a ratio with the price coefficient in the denominator. Theorem 3.2 applies verbatim, and reporting willingness to pay from a Gaussian-prior posterior without examining the induced prior on the ratio is not defensible.

Structural models. Conventional DSGE priors — beta on $(0, 1)$ parameters, gamma on positive ones (Del Negro and Schorfheide, 2008; Herbst and Schorfheide, 2015) — are rapidly varying and therefore compete with the likelihood. In weakly identified structural models, which is most of them, the posterior for a reported quantity may be substantially prior driven, and the standard diagnostic of comparing prior and posterior marginals for individual parameters does not detect it because the problem lives in the functional. Welfare costs, multipliers and variance decompositions are functionals; the induced prior should be simulated and inspected, which is cheap. For partially identified models (Moon and Schorfheide, 2012) the prior on non-identified directions never washes out, so the question is unavoidable rather than asymptotically irrelevant.

Causal inference and machine learning. BART (Chipman et al., 2010) and Bayesian causal forests (Hahn et al., 2020) are the most successful default priors in econometrics, and instructively so: the regularisation is placed on the *function*, in the units of the outcome, and Bayesian causal forests parameterise directly in terms of the prognostic and treatment-effect functions. That is the "put the prior on the reported quantity" principle. By contrast, regularising a nuisance function with an exponentially tailed penalty and then forming a ratio or difference inherits an attenuation that

Table 4: Posterior summaries for two functionals of a 100-dimensional normal mean vector with one coefficient equal to 10 and the rest zero. Truth: $\sum_i \theta_i^2 = 100$, $\max_i \theta_i = 10$. Data: $\sum_i y_i^2 = 207.3$, $\max_i y_i = 9.47$.

	$\psi_1 = \sum_i \theta_i^2$				$\psi_2 = \max_i \theta_i$	
Prior	Q_1	Median	Q_3	Mean	Mean	S.d.
Horseshoe (global–local)	87.0	100.8	116.0	102.1	9.24	1.00
Half-Cauchy, global only	87.2	104.0	121.7	104.9	4.81	0.98
Bayesian lasso	109.3	124.9	141.6	126.1	7.96	1.02
Diffuse $\mathcal{N}(0, 300)$	283.9	305.2	327.5	306.0	9.46	1.00

must be undone by orthogonalisation (Chernozhukov et al., 2018); the Bayesian analogue is to use a prior that does not attenuate in the first place. For neural network predictors (Polson and Sokolov, 2017) weight priors are almost universally Gaussian; heavy-tailed weight priors change the infinite-width limit from a Gaussian to a stable process and give sparser representations, and since the reported quantity is typically an average marginal effect, the catalogue argues for them.

7 Two illustrations

7.1 Variance shares and extreme coefficients

Let $(y_i \mid \theta_i) \sim \mathcal{N}(\theta_i, 1)$, $i \le 100$, with $\theta_1 = 10$ and $\theta_i = 0$ otherwise: one large coefficient in ninety-nine zeros, the canonical sparse econometric configuration. The realised data have $\sum_i y_i^2 = 207.3$ and $\max_i y_i = 9.47$. We report posteriors for $\psi_1 = \sum_i \theta_i^2$, true value 100, and $\psi_2 = \max_i \theta_i$, true value 10, under four priors: horseshoe; a pure-global half-Cauchy, which is the Minnesota structure in reduced form; the Bayesian lasso with the penalty learned; and a diffuse $\mathcal{N}(0, 300)$. Posteriors are computed by Gibbs sampling, 20,000 draws after 5,000 burn-in.

Table 4 makes the argument in four rows. The diffuse Gaussian, the prior that looks most uninformative, overstates the variance share threefold: Efron's 300, and the degenerate limit of Theorem 3.6. The pure-global prior fixes the variance share but reports the largest coefficient as 4.81 when the data say 9.47, which is Proposition 3.9(ii) with numbers attached. The Bayesian lasso, the most popular sparsity prior in applied econometrics, is biased in both directions at once — overstating the share by a quarter and attenuating the maximum by a fifth — because it has a spike and an exponential tail. Only the global–local prior is close on both functionals simultaneously, and no tuning of a single global scale can achieve this because the two functionals pull in opposite directions.

7.2 Ratios and products under weak identification

The second illustration is the bivariate reduced form $(y_1, y_2)^\top \sim \mathcal{N}((\theta_1, \theta_2)^\top, \sigma^2 \boldsymbol{I})$ with $\sigma = 0.1$, corresponding to $n = 100$. We report posteriors for the ratio θ_1/θ_2, the Wald estimand, and the product $\theta_1\theta_2$, the indirect effect, in three configurations: strong first stage, weak first stage with first-stage t statistic one, and complete non-identification. Posteriors are computed exactly on a fine grid, so there is no Monte Carlo error.

Three lessons. When the first stage is strong all four priors agree, as they should: the likelihood dominates. When identification is weak they separate on the product even while agreeing on the ratio; the horseshoe interquartile range for the indirect effect is a third of the Gaussian one and

Table 5: Posterior quartiles for θ_1/θ_2 and for $\theta_1\theta_2$ (entries $\times 1000$), and posterior concentration near the origin. Reduced-form standard error $\sigma = 0.1$.

	θ_1/θ_2			$1000\,\theta_1\theta_2$			
Prior	Q_1	Median	Q_3	Q_1	Median	Q_3	$\mathbb{P}(\lVert\boldsymbol{\theta}\rVert < 0.25 \mid \boldsymbol{y})$
Strong first stage, $\boldsymbol{y} = (0.40, 1.00)$*, true* $\psi = 0.4$							
$\mathcal{N}(0, 100)$	0.329	0.400	0.475	325.6	396.3	470.4	0.000
Laplace	0.322	0.394	0.469	312.7	382.7	455.8	0.000
Cauchy	0.326	0.397	0.472	316.3	386.0	458.6	0.000
Horseshoe	0.312	0.385	0.461	301.6	372.4	446.2	0.000
Weak first stage, $\boldsymbol{y} = (0.04, 0.10)$*, first-stage* $t = 1$							
$\mathcal{N}(0, 100)$	−0.400	0.241	0.968	−2.70	1.60	9.90	0.865
Laplace	−0.424	0.222	0.967	−2.50	1.25	8.55	0.887
Cauchy	−0.409	0.238	0.970	−2.70	1.50	9.50	0.872
Horseshoe	−0.393	0.043	0.944	−0.92	0.00	3.30	0.939
No identification, $\boldsymbol{y} = (0.00, 0.00)$							
$\mathcal{N}(0, 100)$	−1.000	0.000	1.000	−3.60	0.00	3.60	0.956
Laplace	−1.000	0.000	1.000	−3.30	0.00	3.30	0.964
Cauchy	−1.000	0.000	1.000	−3.60	0.00	3.60	0.958
Horseshoe	−1.000	0.000	1.000	−1.10	0.00	1.10	0.983

its median is zero rather than 1.6×10^{-3}. A researcher using a diffuse Gaussian prior would report an indirect effect centred away from zero; the honest answer is zero. Under complete non-identification all priors correctly report ratio quartiles at ± 1 — the ratio of two zeros is undefined and the posterior says so — but the product posteriors again differ threefold in spread.

The pattern is the one Theorem 3.2 predicts. Differences between priors are invisible where the data are informative and decisive where they are not, which is exactly where a default prior earns its name.

8 Practical recommendations

1. Put the prior on the reported quantity. If the paper reports an R^2, a half-life, an elasticity or an impulse response, specify the prior in those units and induce the prior on the coefficients. This eliminates the Jacobian problem by construction. Theorem 3.6 shows the half-Cauchy global scale already does this for R^2; R2–D2 and Bayesian causal forests generalise it.

2. Otherwise, simulate the induced prior and look at it. Draw $\boldsymbol{\theta}$ from the prior, compute $\psi(\boldsymbol{\theta})$, plot the histogram. This costs nothing and would have prevented every pathology in this paper. It belongs alongside convergence diagnostics as a reporting requirement.

3. Use global–local priors with regularly varying tails for coefficients. Horseshoe as baseline; horseshoe$^+$ when ultra-sparse; R2–D2 when an interpretable global scale is wanted; generalized double Pareto or three-parameter beta when a tunable tail index is wanted.

4. Use half-Cauchy or half-t on scales, never inverse-gamma(ϵ, ϵ). In state space models use the non-centred parameterisation so that a zero innovation variance is attainable.

5. Learn the global scale. Fixing Minnesota tightness, or g, is the direct cause of Efron's first pathology and of Bartlett's paradox.

Table 6: A default toolkit for Bayesian econometrics.

Setting	Recommended default	Failure mode avoided
Regression, p large	Horseshoe or R2–D2 on $\boldsymbol{\beta}$; half-Cauchy on σ	upward bias in R^2; attenuation of large coefficients
Predictive regression	Prior on R^2, induced on $\boldsymbol{\beta}$	Jacobian distortion of the reported quantity
IV, many instruments	Global–local first stage; half-Cauchy global	many-instrument bias; attenuated denominator
IV, weak instruments	$p_2(0) > 0$ and regularly varying	manufactured identification
Large VAR	Minnesota-structured global–local, (4)	over-smoothed responses; halved large coefficients
Structural VAR	Simulate induced prior on responses; report robust range	informative Haar prior on rotations
Covariance / factor	Global–local on loadings	inverse Wishart eigenvalue distortion
TVP and SV models	Non-centred; half-Cauchy on ω	manufactured time variation
Panel / hierarchical	Half-Cauchy on effect dispersion	inverse-gamma repulsion from zero
Discrete choice	Pólya–Gamma with horseshoe	Fieller problem in willingness to pay
Testing / selection	Zellner–Siow or hyper-g	Bartlett and information paradoxes
Causal / ML	Prior on the treatment-effect function	regularisation bias in the estimand

6. Do not stop at the global scale. Proposition 3.9 and Table 4: one parameter cannot serve both the variance share and the largest coefficient. For VARs, keep the Minnesota structure inside a global–local hierarchy as in (4).

7. Do not confuse a spike with a default. Normal–gamma, Dirichlet–Laplace, Bayesian lasso, spike-and-slab lasso and penalised complexity priors all attenuate large coefficients.

8. Do not confuse heavy tails with a default either. Independent Cauchy and multivariate t priors do not shrink noise, and aggregate functionals are overstated in high dimension.

9. In ratio problems, keep mass at the denominator's origin. By Theorem 3.2 the constant $p_2(0)\mathbb{E}|\theta_1|$ governs the whole tail of the estimand. Diffusing the first-stage prior tightens the implied prior on the elasticity, which is the opposite of the intent.

10. For testing, use mixtures of g-priors. Zellner–Siow or hyper-g; both are regularly varying and both resolve the information paradox. Never a fixed g.

11. Check properness. Reference and Jeffreys priors are frequently improper, acceptable for estimation in a single well-identified block and unacceptable in hierarchies or for model comparison. The recommended global–local priors are proper by construction — a practical advantage over the reference route, which additionally requires recomputation under reparameterisation and depends on a parameter ordering that economics rarely supplies.

12. Report sensitivity in the functional. Varying a hyperparameter and showing the coefficient posteriors barely move can be entirely uninformative, because the problem lives in p_ψ. Vary the prior and report the effect on ψ.

9 Discussion

The argument compresses to one sentence: in econometrics the prior is judged on the functional, not on the parameter, and regular variation is the property that makes flatness survive the map from one to the other.

The criterion is more modest than Bernardo's reference prior programme, deliberately. Reference priors formally maximise the divergence between prior and posterior and are in that sense optimal, but they are cumbersome or impossible in the models econometricians use, must be recomputed under reparameterisation, depend on an ordering economics rarely supplies, and are frequently improper, which rules them out for model comparison. Regularly varying global–local priors are none of these things: proper, computable by standard augmentation in every model class considered here, and adequate to Efron's four problems. They are a practical compromise between Bernardo's formal framework and flat priors.

Three directions remain. First, the theory here is tail based; posterior concentration results for econometric functionals — impulse responses, multipliers, variance shares — under global–local priors are largely unwritten, though van der Pas et al. (2014) and Bhadra et al. (2019b) provide the template. Second, Theorem 3.6 invites a converse: which priors on the global scale induce a $\mathrm{Beta}(a, b)$ law on R^2, and what does the choice of (a, b) mean for an economist who has a view about how much of the variation is explicable? Third, and most practically, the induced-prior diagnostic of recommendation 2 could be automated in standard Bayesian econometric software. There is no technical obstacle, and that it is not routine is why Efron's fifty-year-old warning still bites.

Acknowledgements. This paper draws throughout on joint work with Anindya Bhadra, Jyotishka Datta and Brandon Willard, and on conversations with Vadim Sokolov and James Scott.

References

Andrade, J. A. A. and O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Analysis* **1**, 169–188.

Bańbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics* **25**, 71–92.

Barndorff-Nielsen, O. E., Kent, J. T. and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review* **50**, 145–159.

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica* **83**, 1963–1999.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80**, 2369–2429.

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association* **84**, 200–207.

Bhadra, A., Datta, J., Polson, N. G. and Willard, B. (2016). Default Bayesian analysis with global–local shrinkage priors. *Biometrika* **103**, 955–969.

Bhadra, A., Datta, J., Polson, N. G. and Willard, B. (2019a). Lasso meets horseshoe: a survey. *Statistical Science* **34**, 405–427.

Bhadra, A., Datta, J., Li, Y., Polson, N. G. and Willard, B. (2019b). Prediction risk for the horseshoe regression. *Journal of Machine Learning Research* **20**(78), 1–39.

Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association* **110**, 1479–1490.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation.* Cambridge University Press.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480.

Chamberlain, G. and Imbens, G. (2004). Random effects estimators with many instrumental variables. *Econometrica* **72**, 295–306.

Chan, J. C. C. (2021). Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting* **37**, 1212–1226.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21**, C1–C68.

Chipman, H. A., George, E. I. and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics* **4**, 266–298.

Creasy, M. A. (1954). Limits for the ratio of means. *Journal of the Royal Statistical Society, Series B* **16**, 186–194.

Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60**, 664–667.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *Journal of the Royal Statistical Society, Series B* **35**, 189–233.

Del Negro, M. and Schorfheide, F. (2008). Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics* **55**, 1191–1208.

Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3**, 1–100.

Efron, B. (1973). Discussion of "Marginalization paradoxes in Bayesian and structural inference". *Journal of the Royal Statistical Society, Series B* **35**, 219–221.

Fieller, E. C. (1940). The biological standardization of insulin. *Supplement to the Journal of the Royal Statistical Society* **7**, 1–64.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics* **154**, 85–100.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

Giacomini, R. and Kitagawa, T. (2021). Robust Bayesian inference for set-identified models. *Econometrica* **89**, 1519–1556.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics* **97**, 436–451.

Good, I. J. (1995). Dual density functions. *Journal of Statistical Computation and Simulation* **52**, 193–194.

Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* **5**, 171–188.

Hahn, P. R., He, J. and Lopes, H. F. (2018). Bayesian factor model shrinkage for linear IV regression with many instruments. *Journal of Business & Economic Statistics* **36**, 278–287.

Hahn, P. R., Murray, J. S. and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference. *Bayesian Analysis* **15**, 965–1056.

Herbst, E. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics* **37**, 27–39.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* **12**, 371–389.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**, 361–393.

Kleibergen, F. and Zivot, E. (2003). Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics* **114**, 29–72.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics* **28**, 204–230.

Liang, F., Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association* **103**, 410–423.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295–304.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: five years of experience. *Journal*

of Business & Economic Statistics **4**, 25–38.

Moon, H. R. and Schorfheide, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica* **80**, 755–782.

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica* **81**, 1805–1849.

O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society, Series B* **41**, 358–367.

Phillips, P. C. B. (1991). To criticize the critics: an objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* **6**, 333–364.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: sparse Bayesian regularization and prediction. In *Bayesian Statistics 9*, 501–538. Oxford University Press.

Polson, N. G. and Scott, J. G. (2012a). Local shrinkage rules, Lévy processes and regularized regression. *Journal of the Royal Statistical Society, Series B* **74**, 287–311.

Polson, N. G. and Scott, J. G. (2012b). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* **7**, 887–902.

Polson, N. G. and Sokolov, V. (2017). Deep learning: a Bayesian perspective. *Bayesian Analysis* **12**, 1275–1304.

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association* **108**, 1339–1349.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* **72**, 821–852.

Ročková, V. and George, E. I. (2018). The spike-and-slab lasso. *Journal of the American Statistical Association* **113**, 431–444.

Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical Science* **32**, 1–28.

Sims, C. A. and Uhlig, H. (1991). Understanding unit rooters: a helicopter tour. *Econometrica* **59**, 1591–1599.

van der Pas, S. L., Kleijn, B. J. K. and van der Vaart, A. W. (2014). The horseshoe estimator: posterior concentration around nearly black vectors. *Electronic Journal of Statistics* **8**, 2585–2618.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In *Bayesian Statistics*, 585–603. Valencia University Press.

Zhang, Y. D., Naughton, B. P., Bondell, H. D. and Reich, B. J. (2022). Bayesian regression using a prior on the model fit: the R2–D2 shrinkage prior. *Journal of the American Statistical Association* **117**, 862–874.