

Online Appendix for Closed-Form Augmentation for Locally Robust Bayesian Inference

Seung Hyun Kim

Contents

A	A Parametric BvM Theorem	2
A.1	Assumptions and statement	2
A.2	Proof	3
B	Delta Method for TV Distance	4
C	Formal Derivation of SVMA Augmentation	7
D	Model-Implied Choice Probabilities in DDC Models	10
E	Details on the Entry-Exit Model Simulation	12
E.1	Details on Baseline and Augmented Likelihood Derivations	12
E.1.1	Likelihood of the Enlarged Model	12
E.1.2	The Augmented Likelihood	14
E.2	Prior Selection	16
E.3	Monte Carlo Design	18
E.4	Sampler Diagnostics	20
F	Details on Krusell-Smith Model Simulation	20
F.1	Numerical Differentiation for KS Model	20
F.2	Monte Carlo Design	24
F.3	Sampler Diagnostics	26
F.4	Robustness Checks for Misspecification Magnitude	27

$|\cdot|$ denotes the euclidean norm, and $\|\cdot\|$ the trace norm on matrix spaces. $\mathbb{N}_+$ is the set of all natural numbers. $\mathbf{i}$ denotes $\sqrt{-1}$.

A A Parametric BvM Theorem

This section gives a parametric Bernstein–von Mises (BvM) result under an integrated, rather than uniform-in-probability, local asymptotic normality condition. The proof adapts Kleijn and van der Vaart (2012), Theorem 2.1.

A.1 Assumptions and statement

We state the assumptions for a triangular array $\{X_{T,t}\}_{1 \leq t \leq n_T, T \in \mathbb{N}_+}$ of random vectors, where each $X_{T,t}$ takes values in the measurable space $(E, \mathcal{E})$ for generality. Let $\mathbf{X}_T = (X_{T,1}, \dots, X_{T,n_T})$ have distribution $P_{T,\theta}$, with strictly positive density $f_T(\cdot \mid \theta)$ relative to a σ -finite measure on E^{n_T} , where $\theta \in \Theta \subset \mathbb{R}^k$. Write $\Pi_T(\cdot \mid \mathbf{X}_T)$ for the posterior associated with Lebesgue prior density q_T , and let θ^* be the parameter relative to which probability statements are made (e.g. some baseline calibration). Assume that Θ is open and convex, $f_T(x \mid \cdot)$ is continuous at θ^* almost everywhere, and the posterior is well-defined. We impose the following conditions.

Assumption A.1.1 (Local flatness of the prior). *For all sufficiently large T , $q_T(\theta^*) > 0$ and, for every compact $K \subset \mathbb{R}^k$,*

$$\sup_{h \in K} \left| \log \frac{q_T(\theta^* + \delta_T h)}{q_T(\theta^*)} \right| \rightarrow 0.$$

Assumption A.1.2 (Integrated LAN). *There exist nonsingular matrices $\delta_T \rightarrow 0$, a positive definite matrix V_0 , and measurable vectors $\Delta_T^*(\mathbf{X}_T) = O_{P_{T,\theta^*}}(1)$ such that, writing*

$$D_T(h, x) = \log \frac{f_T(x \mid \theta^* + \delta_T h)}{f_T(x \mid \theta^*)} - h' \Delta_T^*(x) + \frac{1}{2} h' V_0 h,$$

we have, for every compact $K \subset \mathbb{R}^k$,

$$\sup_{h \in K} \mathbb{E}_{T,\theta^*} [|D_T(h, \mathbf{X}_T)|] \rightarrow 0.$$

Assumption A.1.3 (Locally bounded likelihood ratios). *For every compact $K \subset \mathbb{R}^k$,*

$$\sup_{h \in K} \left| \log \frac{f_T(\mathbf{X}_T \mid \theta^* + \delta_T h)}{f_T(\mathbf{X}_T \mid \theta^*)} \right| = O_{P_T, \theta^*}(1).$$

Assumption A.1.4 (Posterior concentration). *For every deterministic $M_T \uparrow \infty$,*

$$\mathbb{E}_{T, \theta^*} \left[\Pi_T \left(\left\{ \theta \mid |\delta_T^{-1}(\theta - \theta^*)| > M_T \right\} \mid \mathbf{X}_T \right) \right] \rightarrow 0$$

Let $\tilde{\Pi}_T$ denote the posterior distribution of $h = \delta_T^{-1}(\theta - \theta^*)$, and put

$$\Phi_T(\cdot \mid x) = \mathcal{N}(V_0^{-1} \Delta_T^*(x), V_0^{-1}).$$

The assumptions on densities, priors, and parameter space above replace the corresponding preliminary regularity assumptions in the longer formulation.

Theorem A.1.1 (Bernstein–von Mises theorem). *Under the preceding assumptions,*

$$d_{TV}(\tilde{\Pi}_T(\cdot \mid \mathbf{X}_T), \Phi_T(\cdot \mid \mathbf{X}_T)) = o_{P_T, \theta^*}(1).$$

A.2 Proof

Fix a compact set $K \subset \mathbb{R}^k$ of positive Lebesgue measure. Write $\tilde{\Pi}_T^K$ and Φ_T^K for the corresponding conditional distributions restricted to K (defined arbitrarily if the posterior gives K zero mass). Their densities on the event of positive posterior mass are proportional, respectively, to

$$f_T(x \mid \theta^* + \delta_T h) q_T(\theta^* + \delta_T h) \mathbf{1}_K(h), \quad \exp\{h' \Delta_T^*(x) - \frac{1}{2} h' V_0 h\} \mathbf{1}_K(h).$$

For $h, g \in K$, the ratio of these two unnormalized densities at g relative to h is

$$\exp\{D_T(g, x) - D_T(h, x)\} \frac{q_T(\theta^* + \delta_T g)}{q_T(\theta^* + \delta_T h)}.$$

Consequently, Jensen's inequality and $(1 - e^u)^+ \leq |u|$ give

$$\begin{aligned} d_{TV}(\tilde{\Pi}_T^K, \Phi_T^K) &\leq \iint_{K \times K} \left[|D_T(g, x)| + |D_T(h, x)| + r_T(K) \right] \Phi_T^K(dg \mid x) \tilde{\Pi}_T^K(dh \mid x), \\ r_T(K) &:= 2 \sup_{h \in K} \left| \log \frac{q_T(\theta^* + \delta_T h)}{q_T(\theta^*)} \right| \rightarrow 0. \end{aligned}$$

Here and below the dependence on x in the first line is understood.

To integrate this bound under P_{T,θ^*} , fix $\epsilon > 0$. Since $\Delta_T^*(\mathbf{X}_T) = O_p(1)$, there is an $L < \infty$ for which $P_{T,\theta^*}(|\Delta_T^*| > L) \leq \epsilon$ for all large T . On $\{|\Delta_T^*| \leq L\}$, the normalized Gaussian density Φ_T^K is bounded above by a finite constant $A_{K,L}$. Moreover, the local likelihood-ratio bound and prior flatness imply that the oscillation of the log posterior density over K is $O_p(1)$. Hence, for a constant $C_{K,\epsilon} < \infty$, the normalized posterior density $\tilde{\Pi}_T^K$ is bounded by $C_{K,\epsilon}$ outside an event of probability at most ϵ for all large T . On their intersection $\mathcal{G}_T$,

$$\begin{aligned} \mathbb{E}_{T,\theta^*} [\mathbf{1}_{\mathcal{G}_T} d_{TV}(\tilde{\Pi}_T^K, \Phi_T^K)] &\leq (A_{K,L} + C_{K,\epsilon}) \int_K \mathbb{E}_{T,\theta^*} [|D_T(h, \mathbf{X}_T)|] dh + r_T(K) + O(\epsilon) \\ &\leq (A_{K,L} + C_{K,\epsilon}) \text{vol}(K) \times \sup_{h \in K} \mathbb{E}_{T,\theta^*} [|D_T(h, \mathbf{X}_T)|] + r_T(K) + O(\epsilon). \end{aligned}$$

Because total variation is bounded by one, the exceptional events contribute at most $O(\epsilon)$. Integrated LAN therefore yields

$$\mathbb{E}_{T,\theta^*} [\mathbf{1}_{\{\tilde{\Pi}_T(K|\mathbf{X}_T) > 0\}} d_{TV}(\tilde{\Pi}_T^K, \Phi_T^K)] \rightarrow 0$$

after letting $T \rightarrow \infty$ and then $\epsilon \downarrow 0$.

Finally choose closed balls $K_M = \{h \mid |h| \leq M\}$. Posterior concentration gives $\lim_{M \rightarrow \infty} \limsup_T \mathbb{E}_{T,\theta^*} [\tilde{\Pi}_T(K_M^c \mid \mathbf{X}_T)] = 0$. Since the Gaussian means $V_0^{-1} \Delta_T^*(\mathbf{X}_T)$ are stochastically bounded and their covariance is fixed, $\lim_{M \rightarrow \infty} \limsup_T \mathbb{E}_{T,\theta^*} [\Phi_T(K_M^c \mid \mathbf{X}_T)] = 0$ as well. The triangle inequality for the restrictions to K_M now gives

$$d_{TV}(\tilde{\Pi}_T, \Phi_T) \leq d_{TV}(\tilde{\Pi}_T^{K_M}, \Phi_T^{K_M}) + \tilde{\Pi}_T(K_M^c \mid \mathbf{X}_T) + \Phi_T(K_M^c \mid \mathbf{X}_T)$$

on the event of positive posterior mass on K_M , whose probability tends to one as $M \rightarrow \infty$. Taking expectations, then $T \rightarrow \infty$ and $M \rightarrow \infty$, establishes convergence in mean and hence in probability. This is the compact-set argument of Kleijn and van der Vaart (2012), adapted to integrated LAN.

B Delta Method for TV Distance

This section proves Theorem 1. The key point is that, for a locally invertible smooth transformation, convergence of Gaussian densities under a rescaled nonlinear map holds in total variation, not merely weakly.

Lemma B.1 (Contraction of total variation). *For probability measures μ, ν and a measurable map g ,*

$$d_{TV}(\mu \circ g^{-1}, \nu \circ g^{-1}) \leq d_{TV}(\mu, \nu).$$

Proof. For every measurable A , the difference of the pushforward probabilities at A equals $\mu(g^{-1}A) - \nu(g^{-1}A)$; take the supremum over A . $\square$

Lemma B.2 (Total-variation delta method for Gaussian laws). *Let G be continuously differentiable in a neighborhood of $0 \in \mathbb{R}^k$, with $G(O_{k \times 1}) = O_{k \times 1}$ and nonsingular derivative $J = DG(0)$. Define $g_T(z) = \sqrt{T} G(z/\sqrt{T})$. If Σ is positive definite and $m_T = O_p(1)$, then, conditionally on the random mean m_T ,*

$$d_{TV}\left(\mathcal{N}(m_T, \Sigma) \circ g_T^{-1}, \mathcal{N}(m_T, \Sigma) \circ J^{-1}\right) = o_p(1).$$

The same conclusion holds for a T -dependent local map with uniformly convergent derivatives on every fixed compact neighborhood of zero.

Proof. Choose a closed ball K_R such that $\sup_T P(|m_T| > R_0)$ is arbitrarily small and the Gaussian probability of K_R^c is uniformly small whenever $|m_T| \leq R_0$. By the inverse function theorem, there is a neighborhood U of zero on which G is a diffeomorphism. For large T , g_T is a diffeomorphism on a fixed bounded open neighborhood N of K_R , and

$$\sup_{z \in N} \|Dg_T(z) - J\| \rightarrow 0, \quad \sup_{z \in N} |g_T(z) - Jz| \rightarrow 0.$$

In particular, $g_T^{-1}(y) - J^{-1}y \rightarrow 0$ uniformly for $y \in g_T(N)$, and the inverse Jacobian determinants converge uniformly there. On a fixed bounded region containing $g_T(N)$, the change-of-variables formula expresses the density of the transformed Gaussian as

$$p_{T,m}(y) = \frac{\phi_\Sigma(g_T^{-1}(y) - m)}{|\det Dg_T(g_T^{-1}(y))|}, \quad y \in g_T(N),$$

where ϕ_Σ is the $\mathcal{N}(0, \Sigma)$ density. Its linear counterpart is $\phi_\Sigma(J^{-1}y - m)/|\det J|$. Uniform continuity of the Gaussian density and the preceding Jacobian convergence imply that the integral of their absolute difference over the common interior tends to zero, uniformly over $|m| \leq R_0$. To handle the possibly different boundaries, choose an inner compact ball K_R and a slightly larger open ball N ; for large T , $JK_R \subset g_T(N)$, so the omitted mass is bounded by the Gaussian probability of K_R^c . The same bound controls the mass of both transformed distributions outside the common region. Let $T \rightarrow \infty$, then $R \rightarrow \infty$ and $R_0 \rightarrow \infty$. This proves the assertion. $\square$

Proof of Theorem 1. Write $\vartheta = (\nu, \delta) \in \mathbb{R}^{k_A}$, where $k_A = k_\nu + k_\delta$. Under the assumptions of

Theorem 1, the augmented posterior of $\sqrt{T}(\vartheta - \vartheta_0)$ satisfies the parametric BvM approximation

$$d_{TV}(\Pi_T(\cdot | \mathbf{X}_T), \Phi_T(\cdot | \mathbf{X}_T)) = o_{P_{T,\theta^*}}(1), \quad \Phi_T = \mathcal{N}\left((\mathbb{I}_*^{\mathcal{A}})^{-1} \frac{\dot{\ell}_{T,\vartheta}^{\mathcal{A}}(\vartheta_0 | \mathbf{X}_T)}{\sqrt{T}}, (\mathbb{I}_*^{\mathcal{A}})^{-1}\right).$$

The normal mean is $O_{P_{T,\theta^*}}(1)$ by the assumed local likelihood expansion.

The derivative $\dot{H}_*^{\mathcal{A}}$ has full row rank $k_H \leq k_A$. After reordering the coordinates if necessary, its first k_H columns are linearly independent. Extend the target map to a square map

$$\mathbf{H}(\vartheta) = \begin{pmatrix} H(\nu, \eta^*) + D_\zeta H(\nu, \eta^*) \mathbf{B}_0 \delta \\ \vartheta_{k_H+1:k_A} \end{pmatrix}.$$

Its derivative $J = D\mathbf{H}(\vartheta_0)$ is nonsingular, since its top k_H rows equal $\dot{H}_*^{\mathcal{A}}$ and its remaining rows select the remaining coordinates. The inverse function theorem therefore applies in a neighborhood of ϑ_0 . Set

$$g_T(z) = \sqrt{T} \left\{ \mathbf{H}(\vartheta_0 + z/\sqrt{T}) - \mathbf{H}(\vartheta_0) \right\}.$$

By contraction of total variation (Lemma B.1) and the Gaussian delta method (Lemma B.2),

$$d_{TV}(\Pi_T \circ g_T^{-1}, \Phi_T \circ J^{-1}) \leq d_{TV}(\Pi_T, \Phi_T) + d_{TV}(\Phi_T \circ g_T^{-1}, \Phi_T \circ J^{-1}) = o_{P_{T,\theta^*}}(1).$$

Projecting onto the first k_H coordinates, again using contraction, yields the desired BvM approximation for $\sqrt{T}\{H^{\mathcal{A}}(\nu, \delta) - H(\theta^*)\}$, with normal mean and covariance

$$\dot{H}_*^{\mathcal{A}}(\mathbb{I}_*^{\mathcal{A}})^{-1} \frac{\dot{\ell}_{T,\vartheta}^{\mathcal{A}}(\vartheta_0 | \mathbf{X}_T)}{\sqrt{T}}, \quad \dot{H}_*^{\mathcal{A}}(\mathbb{I}_*^{\mathcal{A}})^{-1} \dot{H}_*^{\mathcal{A}\prime}.$$

Finally, under the local misspecification $\eta_0 = \eta^* + h/\sqrt{T}$, the assumed differentiability of H with respect to η gives

$$a_T := \sqrt{T}\{H(\theta_0) - H(\theta^*)\} \longrightarrow a := \dot{H}_{*,\eta} h.$$

The posterior centered at $H(\theta_0)$ and its proposed Gaussian approximation are translations by $-a_T$ and $-a$, respectively, of the two distributions just compared. Total variation is invariant under a common translation. Moreover, the total-variation distance between a nondegenerate Gaussian and its translation by $a_T - a$ tends to zero as $a_T \rightarrow a$ (by L^1 continuity of translations of its density). Applying the triangle inequality proves the stated limit under P_{T,θ^*} . The assumed contiguity of P_{T,θ_0} with respect to P_{T,θ^*} now yields the result under P_{T,θ_0} . $\square$

C Formal Derivation of SVMA Augmentation

Here we formally derive the augmentation for the SVMA model. Under an SVMA DGP, each $\{X_{T,t}\}_{t \in \mathbb{Z}}$ is generated as

$$\mathbf{X}_{T,t} = \mu(\theta) + \sum_{h=0}^{\infty} \Psi_{T,h}(\theta) \epsilon_{t-h}, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N} [O_{n_\epsilon \times 1}, I_{n_\epsilon}]$$

where $\mu(\nu, \eta)$ is an n -dimensional vector, each $\Psi_{T,h}(\nu, \eta)$ is an $n \times n_\epsilon$ matrix, and $n \leq n_\epsilon$. Suppose that, for each $\theta \in \Theta$, the MA coefficients are absolutely summable:

$$\sum_{h=0}^{\infty} \|\Psi_{T,h}(\theta)\| < +\infty.$$

Then, we can immediately see that the spectral density $f_T(\cdot; \theta)$ of $\{X_{T,t}\}_{t \in \mathbb{Z}}$ under the parameter θ is

$$f_T(\omega; \theta) = \frac{1}{2\pi} \Psi_T(e^{i\omega} \mid \theta) \Psi_T(e^{i\omega} \mid \theta)^* \quad \forall \omega \in (-\pi, \pi]$$

where

$$\Psi_T(z \mid \theta) = \sum_{h=0}^{\infty} \Psi_{T,h}(\theta) z^h \quad \forall z \in \mathbb{C}, |z| \leq 1.$$

Then, the DGP of $\mathbf{X}_T$ can be written as

$$\mathbf{X}_T; \theta \sim \mathcal{N} \left[\iota_T \otimes \mu(\theta), \mathbf{V}_T [f_T(\cdot; \theta)] \right].$$

where $\mathbf{V}_T [\cdot]$ is the block Toeplitz operator. Denote $\Sigma_T = \mathbf{V}_T [f_T(\cdot; \theta^*)]$, or the covariance matrix of $\mathbf{X}_T$ under θ^* .

Suppose that each $\Psi_{T,h}(\theta)$ is continuously differentiable on a neighborhood $\mathcal{N}$ of $\theta^* = (\nu_0, \eta^*)$, so that

$$\begin{aligned} s_{T,j}(\omega; \theta) &:= \frac{\partial f_T(\omega; \theta)}{\partial \theta_j} = \frac{1}{2\pi} \left[\sum_{h=0}^{\infty} \frac{\partial \Psi_{T,h}(\theta)}{\partial \theta_j} e^{i\omega h} \right] \Psi_T(e^{i\omega} \mid \theta)^* + \frac{1}{2\pi} \Psi_T(e^{i\omega} \mid \theta) \left[\sum_{h=0}^{\infty} \frac{\partial \Psi_{T,h}(\theta)}{\partial \theta_j} e^{i\omega h} \right]^* \\ &= \frac{1}{2\pi} \{ \Psi_T(e^{i\omega} \mid \theta) + D_{\theta_j} \Psi_T(e^{i\omega} \mid \theta) \} \{ \Psi_T(e^{i\omega} \mid \theta) + D_{\theta_j} \Psi_T(e^{i\omega} \mid \theta) \}^* \end{aligned}$$

$$-\frac{1}{2\pi}\Psi_T(e^{i\omega} \mid \theta)\Psi_T(e^{i\omega} \mid \theta)^* - \frac{1}{2\pi}D_{\theta_j}\Psi_T(e^{i\omega}; \theta) \left(D_{\theta_j}\Psi_T(e^{i\omega}; \theta)\right)^*$$

for any $\theta \in \mathcal{N}$ and $\omega \in (-\pi, \pi]$, where

$$D_{\theta_j}\Psi_T(z \mid \theta) = \sum_{h=0}^{\infty} \frac{\partial \Psi_{T,h}(\theta)}{\partial \theta_j} z^h \quad \forall z \in \mathbb{C}, \quad |z| \leq 1.$$

By the additivity of the Toeplitz operator $\mathbf{V}_T(\cdot)$, this decomposition implies that $\mathbf{V}_T[s_{T,j}(\cdot; \theta)]$ can be written as

$$\mathbf{V}_T[s_{T,j}(\cdot; \theta)] = \mathbf{V}_T[f_T^{\Sigma(j)}(\cdot; \theta)] - \mathbf{V}_T[f_T^{D(j)}(\cdot; \theta)] - \mathbf{V}_T[f_T(\cdot; \theta)]$$

where $f_T^{\Sigma(j)}(\cdot; \theta)$ and $f_T^{D(j)}(\cdot; \theta)$ are the spectra corresponding to the SVMA(∞) processes

$$\sum_{h=0}^{\infty} \left[\Psi_{T,h}(\theta) + \frac{\partial \Psi_{T,h}(\theta)}{\partial \theta_j} \right] \epsilon_{t-h}$$

and

$$\sum_{h=0}^{\infty} \frac{\partial \Psi_{T,h}(\theta)}{\partial \theta_j} \epsilon_{t-h}.$$

Clearly, each $\frac{\partial f_T(\omega; \theta)}{\partial \theta_j}$ is Hermitian valued, and the mapping $\omega \mapsto \frac{\partial f_T(\omega; \theta)}{\partial \theta_j}$ is also continuous and thus Borel measurable for each θ . Defining

$$s_{T,j}(\omega) = s_{T,j}(\omega; \theta^*) \quad \forall 1 \leq j \leq k,$$

we can see that $s_{T,1}, \dots, s_{T,k} : (-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$ are measurable Hermitian valued functions. Now all that is needed to form the augmented model is to find the augmented means and $\mu^{\mathcal{A}}(\vartheta)$ and spectral density $f_T^{\mathcal{A}}(\cdot; \vartheta)$ that satisfy conditions (4.2) and (4.3) in the main paper.

A simple construction for $\mu^{\mathcal{A}}(\cdot)$ that satisfies the derivative condition is a linear expansion of $\mu(\cdot)$ around η^* :

$$\mu^{\mathcal{A}}(\nu, \delta) = \mu(\nu, \eta^*) + \frac{\partial \mu(\nu, \eta^*)}{\partial \zeta'} \mathbf{B}_0 \delta.$$

A similar construction works for the spectral density because it is a quadratic form. Specifically,

we can let

$$f_T^{\mathcal{A}}(\omega; \nu, \delta) = \frac{1}{2\pi} \left\{ \Psi_T(e^{i\omega} \mid \nu, \eta^*) + \sum_{j=1}^{k_\delta} \delta_j \left(\sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \nu, \eta^*) \right) \right\} \\ \times \left\{ \Psi_T(e^{i\omega} \mid \nu, \eta^*) + \sum_{j=1}^{k_\delta} \delta_j \left(\sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \nu, \eta^*) \right) \right\}^*.$$

The derivatives of $f_T^{\mathcal{A}}(\omega; \nu, \delta)$ with respect to ν are left unchanged compared to the baseline model because we evaluate the derivative at $\delta_0 = O_{k_\delta \times 1}$. Meanwhile, for any $1 \leq j \leq k_\delta$,

$$\frac{\partial f_T^{\mathcal{A}}(\omega; \nu_0, \delta_0)}{\partial \delta_j} = \frac{1}{2\pi} \left\{ \left(\sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \theta^*) \right) \Psi_T(e^{i\omega} \mid \theta^*)^* \right. \\ \left. + \Psi_T(e^{i\omega} \mid \theta^*) \left(\sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \theta^*) \right)^* \right\} \\ = \sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} \times \frac{1}{2\pi} \left\{ D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \theta^*) \Psi_T(e^{i\omega} \mid \theta^*)^* + \Psi_T(e^{i\omega} \mid \theta^*) \left(D_{\zeta_\ell} \Psi_T(e^{i\omega} \mid \theta^*) \right)^* \right\} \\ = \sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} \frac{\partial f_T(\omega; \theta^*)}{\partial \zeta_\ell}$$

so that we obtain exactly the same derivatives that we desired.

The final step is to recognize that the augmented spectral density $f_T^{\mathcal{A}}(\cdot; \nu, \delta)$ itself corresponds to an absolutely summable SVMA(∞) process for values of (ν, δ) contained in a small enough neighborhood of (ν_0, δ_0) . To see this, note that

$$\Psi_T(e^{i\omega} \mid \nu, \eta^*) + \sum_{j=1}^{k_\delta} \delta_j \left(\sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} D_{\zeta_\ell} \Psi_T(e^{i\omega}; \nu, \eta_0) \right) = \sum_{h=0}^{\infty} \left[\Psi_{T,h}(\nu, \eta_0) + \sum_{j=1}^{k_\delta} \delta_j \sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} \frac{\partial \Psi_{T,h}(\nu, \eta_0)}{\partial \zeta_\ell} \right] e^{i\omega h}.$$

Defining

$$\Psi_{T,h}^{\mathcal{A}}(\nu, \delta) = \Psi_{T,h}(\nu, \eta^*) + \sum_{j=1}^{k_\delta} \delta_j \sum_{\ell=1}^{k_\zeta} \mathbf{B}_{0,\ell j} \frac{\partial \Psi_{T,h}(\nu, \eta^*)}{\partial \zeta_\ell},$$

the coefficients $\{\Psi_{T,h}^{\mathcal{A}}(\nu, \delta)\}_{h \in \mathbb{N}}$ are absolutely summable for large enough T if δ is small enough, since the coefficients of the original model are absolutely summable on a neighborhood of (ν_0, η^*) .

it follows that

$$f_T^{\mathcal{A}}(\omega; \nu, \delta) = \frac{1}{2\pi} \left(\sum_{h=0}^{\infty} \Psi_{T,h}^{\mathcal{A}}(\nu, \delta) e^{i\omega h} \right) \left(\sum_{h=0}^{\infty} \Psi_{T,h}^{\mathcal{A}}(\nu, \delta) e^{i\omega h} \right)^*,$$

which is exactly the spectral density corresponding to the model

$$X_{T,t} = \mu^{\mathcal{A}}(\nu, \delta) + \sum_{h=0}^{\infty} \Psi_{T,h}^{\mathcal{A}}(\nu, \delta) \epsilon_{t-h}, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}[O_{n_\epsilon}, I_{n_\epsilon}].$$

D Model-Implied Choice Probabilities in DDC Models

We consider a model where a single agent faces a discrete time infinite-horizon problem where the choice set $\mathcal{D}$ consists of J distinct options, labeled 1 to J . The state vector s can be decomposed as $s = (x, \epsilon)$. x is the endogenous state that is Markovian, and so depends only on the value of x and the agent's choice in the immediate past. ϵ is the exogenous choice variable that is iid given the endogenous state, and whose distribution at each period depends only on the contemporaneous endogenous state¹. We let $f(x' | x, d)$ be the probability of observing x' tomorrow given x, d today. $g(x, \cdot)$ denotes the distribution of the exogenous state given today's endogenous state x . $u(s, d)$ is the flow utility of the agent given a pair (s, d) of the state and control, and here we assume that $u(s, d)$ is additively separable in the exogenous state, so that $u(s, d) = \tilde{u}(x, d) + \epsilon(d)$, where $\epsilon(d)$ is the d th coordinate of the random vector ϵ .

The value function $V(\cdot)$ satisfies the Bellman equation

$$V(s) = \max_{d \in \mathcal{D}} \left[h(x, d)' g(\theta_u) + \epsilon(d) + \rho \sum_{x' \in \mathcal{X}} V_\sigma(x') f(x' | x, d; \theta_f) \right],$$

where, following Aguirregabiria and Mira (2002), we define the smoothed Bellman operator $V_\sigma : \mathcal{X} \rightarrow \mathbb{R}$ as $V_\sigma(x) = \int_{\mathcal{E}} V(x, \epsilon) g(x, d\epsilon)$. $\rho \in (0, 1)$ is the discount factor that is fixed at a predetermined value because it is generally unidentified.

Letting $d^*(x, \epsilon)$ be the optimal policy function, the conditional choice probabilities are defined as $P_d(x) = \mathbb{P}(d^*(x, \epsilon) = d | x)$ for each $(x, d) \in \mathcal{X} \times \mathcal{D}$. If the endogenous state space is continuous, we discretize it using a grid of M points, $\mathcal{X}_g = \{x_1, \dots, x_M\}$; if the endogenous state space is already discrete, then we let M be its cardinality. Stacking the smoothed Bellman operator and the CCPs by the discretized endogenous state as $V_\sigma, P_d \in \mathbb{R}^{M \times 1}$, it is shown in Aguirregabiria and Mira (2002)

¹This type of decomposition of the state vector originates from the conditional independence assumption in Rust (1987) and is standard in the literature.

that an approximate solution to the dynamic programming problem is

$$V_\sigma = \sum_{d \in \mathcal{D}} P_d \odot \{u_d + e(d, P) + \rho F_d V_\sigma\},$$

where the (i, j) th element of $F_d \in \mathbb{R}^{M \times M}$ is equal to $f(x_j \mid x_i, d)$ and the i th element of $e(d, P)$ is defined as $\mathbb{E}[\epsilon(d) \mid x = x_i, d^*(x_i, \epsilon) = d]$. Since $F(P) := \sum_{d \in \mathcal{D}} \text{diag}(P_d) F_d$ is a row-stochastic matrix, V_σ can be solved as

$$V_\sigma = Q(P) := (I_M - \rho F(P))^{-1} \left(\sum_{d \in \mathcal{D}} P_d \odot \{u_d + e(d, P)\} \right).$$

Finally, the CCPs are then determined via the fixed point relation

$$P_d(x) = \Psi(x, d; P) := \mathbb{P} \left(\bigcap_{d' \neq d} \left\{ \tilde{v}(x, d; P) + \epsilon(d) \geq \tilde{v}(x, d'; P) + \epsilon(d') \right\} \mid x \right),$$

where $\tilde{v}(x, d; P) = u(x, d) + \rho F_d(x)' Q(P)$ and $F_d(x_i)$ is the i th row of F_d .

Let θ_u denote the parameters that govern the flow utility $u(\cdot)$, and θ_f the parameters determining the evolution of the endogenous state. If the CCPs P are treated as a parameter representing the candidate optimal policy, the mapping $\Psi(x, d; \theta_u, \theta_f, P)$ specifies a likelihood for the observations $\{(d_{it}, x_{it})\}_{1 \leq t \leq T_i, i \in \mathbb{N}_+}$ of agent choices and endogenous states up to some finite time, as a parametric model with parameters θ_u, θ_f and P . We adopt two conventions that are commonly found in DDC specifications: that $u(x, d; \theta_u)$ is multiplicatively separable in the parameter θ_u , so that $u(x, d; \theta_u) = h(x, d)' g(\theta_u)$ for some function $g(\cdot)$; and that ϵ is a J -dimensional random vector whose coordinates follow independent standard type-I generalized extreme value distributions. The distributional assumption delivers the closed form expression for $e_x(d, P) = \gamma_e - \log(P_d(x))$, where γ_e is Euler's constant, so

$$\begin{aligned} Q(P; \theta_u, \theta_f) &= (I_M - \rho F(P; \theta_f))^{-1} \left(\sum_{d \in \mathcal{D}} \text{diag}(P_d) H_d \right) g(\theta_u) \\ &\quad + (I_M - \rho F(P; \theta_f))^{-1} \left(\sum_{d \in \mathcal{D}} P_d \odot (\gamma_e \cdot \iota_M - \log(P_d)) \right) \end{aligned}$$

where H_d is the matrix stacking the vectors $h(x_1, d)', \dots, h(x_M, d)'$ and ι_M is the M -dimensional vector of ones. Furthermore, since the difference of iid type-I generalized extreme value-distributed

random variables is the logistic distribution, the model-implied probabilities are given as

$$\mathbb{P}(d_{it} = d \mid x_{it}; \theta_u, \theta_f, P) = \frac{\exp(M_0(x_{it}, d; P, \theta_f) + M_1(x_{it}, d; P, \theta_f)g(\theta_u))}{1 + \sum_{d'=2}^J \exp(M_0(x_{it}, d'; P, \theta_f) + M_1(x_{it}, d'; P, \theta_f)g(\theta_u))}, \quad (1)$$

where we define

$$M_0(d; P, \theta_f) = \rho(F_d(\theta_f) - F_1(\theta_f))(I_M - \rho F(P; \theta_f))^{-1} \left(\gamma_e \cdot \iota_M - \sum_{d' \in \mathcal{D}} P_{d'} \odot \log(P_{d'}) \right)$$

$$M_1(d; P, \theta_f) = H_d - H_1 + \rho(F_d(\theta_f) - F_1(\theta_f))(I_M - \rho F(P; \theta_f))^{-1} \left(\sum_{d' \in \mathcal{D}} \text{diag}(P_{d'}) H_{d'} \right).$$

and $M_0(x, d; P, \theta_f)$ and $M_1(x, d; P, \theta_f)$ are obtained by interpolating among the rows of $M_0(d; P, \theta_f)$ and $M_1(d; P, \theta_f)$ (which correspond to the values $M_0(x, d; P, \theta_f)$ and $M_1(x, d; P, \theta_f)$ for x on the grid $\mathcal{X}_g$).

E Details on the Entry-Exit Model Simulation

E.1 Details on Baseline and Augmented Likelihood Derivations

E.1.1 Likelihood of the Enlarged Model

Let $F_1(\theta_f)$ and F_0 denote the $(M+1) \times (M+1)$ transition matrices for the observed states given $d = 1$ and $d = 0$, respectively. In the present model, we have

$$F_0 = \begin{pmatrix} \iota_{M+1} & O_{(M+1) \times M} \end{pmatrix}, \quad F_1(\theta_f) = \begin{pmatrix} 0 & F_{01}(\theta_f)' \\ O_{M \times 1} & F_{11}(\theta_f) \end{pmatrix}$$

where ι_{M+1} is the $M+1$ -dimensional vector of ones, $F_{01}(\theta_f)$ is an M -dimensional vector representing a discretized version of the distribution $\text{LogNormal}[b_e, \sigma_u^2]$, and $F_{11}(\theta_f)$ is the Tauchen transition matrix for the process $\log k_{t+1} \sim \mathcal{N}[b_0 + b_1 \log k_t, \sigma_u^2]$. Explicitly, letting $\{k_1, \dots, k_M\}$ be a fixed grid over capital,

$$[F_{01}(\theta_f)](j) = \begin{cases} \Phi\left(\frac{\log k_j - b_e}{\sigma_u}\right) & \text{if } j = 1 \\ \Phi\left(\frac{\log k_{j+1} - b_e}{\sigma_u}\right) - \Phi\left(\frac{\log k_j - b_e}{\sigma_u}\right) & \text{if } 1 \leq j < M \\ 1 - \Phi\left(\frac{\log k_j - b_e}{\sigma_u}\right) & \text{if } j = M \end{cases}$$

$$[F_{11}(\theta_f)](i, j) = \begin{cases} \Phi\left(\frac{\log k_j - b_0 - b_1 \log k_i}{\sigma_u}\right) & \text{if } j = 1 \\ \Phi\left(\frac{\log k_{j+1} - b_0 - b_1 \log k_i}{\sigma_u}\right) - \Phi\left(\frac{\log k_j - b_0 - b_1 \log k_i}{\sigma_u}\right) & \text{if } 1 \leq j < M \\ 1 - \Phi\left(\frac{\log k_j - b_0 - b_1 \log k_i}{\sigma_u}\right) & \text{if } j = M \end{cases}$$

for each $1 \leq i, j \leq M$.

Given a sample $\{(d_{it}, k_{it})\}_{0 \leq t \leq T, i \in \mathbb{N}_+}$ of entry/exit choices and capital stock that are independent across firms, the log-likelihood is given by

$$\begin{aligned} \ell(\theta, P; \{(d_{it}, k_{it})\}_{0 \leq t \leq T}) &= \sum_{t=1}^T \{\alpha_d(x_{it}; \theta, P) d_{it} - A_d(\alpha_d(x_{it}; \theta, P))\} \\ &\quad + \sum_{t=1}^T d_{i,t-1} \{f(k_{it}) + \alpha_f(x_{i,t-1}; \theta_f)' S_f(k_{it}) - A_f(\alpha_f(x_{i,t-1}; \theta_f))\} \end{aligned}$$

with initial values $d_{i,-1} = 0$ to indicate that every firm starts as a potential entrant. Here, $\alpha_d(x; \theta, P)$ is the element of the vector $\alpha_d(\theta, P)$ corresponding to the value of the state x , which is defined as

$$\alpha_d(\theta, P) = M_0(\theta_f, P) + M_1(\theta_f, P)\theta_u \in \mathbb{R}^{(M+1) \times 1}$$

$$M_0(\theta_f, P) = \rho(F_1(\theta_f) - F_0) (I_{M+1} - \rho F(P, \theta_f))^{-1} \left(\gamma_e \cdot \iota_{M+1} - \sum_{d' \in \{0,1\}} P_{d'} \odot \log P_{d'} \right)$$

$$M_1(\theta_f, P) = H + \rho(F_1(\theta_f) - F_0) (I_{M+1} - \rho F(P, \theta_f))^{-1} \text{diag}(P_1)H$$

$$F(P, \theta_f) = \text{diag}(P_0)F_0 + \text{diag}(P_1)F_1(\theta_f), \quad H = \begin{pmatrix} 0 & -1 \\ k_1 & 0 \\ \vdots & \vdots \\ k_M & 0 \end{pmatrix}, \quad A_d(\alpha) = \log(1 + \exp(\alpha))$$

for Euler's constant γ_e . Meanwhile,

$$f(k) = -\frac{1}{2} \log(2\pi) - \log(k),$$

$$\alpha_f(k, s; \theta_f) = \begin{pmatrix} \frac{1}{\sigma_u^2} [s(b_0 + b_1 \log(k)) + (1-s)b_e] \\ -\frac{1}{2\sigma_u^2} \end{pmatrix},$$

$$S_f(k) = \begin{pmatrix} \log(k) \\ \log(k)^2 \end{pmatrix}, \quad A_f(\alpha) = -\frac{\alpha_1^2}{4\alpha_2} - \frac{1}{2} \log(-2\alpha_2)$$

follows from the known exponential family representation of the log-normal distribution.

E.1.2 The Augmented Likelihood

Let P^*, θ_f^* be first-stage estimators for the CCPs P and state transition parameters θ_f , and $\{d_{it}, k_{it}\}_{0 \leq t \leq T, i \in \mathbb{N}_+}$ be a panel used to estimate the flow utility parameters in the second stage that is independent of the sample used to construct P^* and θ_f^* . Then, the augmented principle reveals that the log-likelihood of the augmented model will be given as

$$\begin{aligned} \ell^{\mathcal{A}}(\theta_u, \delta; \{d_{it}, k_{it}\}_{0 \leq t \leq T}) &= \sum_{t=1}^T \left\{ \alpha_d^{\mathcal{A}}(x_{it}; \beta, \delta) d_{it} - A_d(\alpha_d^{\mathcal{A}}(x_{it}; \beta, \delta)) \right\} \\ &\quad + \sum_{t=1}^T d_{i,t-1} \left\{ f(k_{it}) + \alpha_f^{\mathcal{A}}(x_{i,t-1}; \delta)' S_f(k_{it}) - A_f(\alpha_f^{\mathcal{A}}(x_{i,t-1}; \delta)) \right\} \end{aligned}$$

where $\beta = \theta_u$ and

$$\begin{aligned} \alpha_d^{\mathcal{A}}(x; \beta, \delta) &= \alpha_d(x; \beta, \theta_f^*, P^*) + \frac{\partial \alpha_d(x; \beta, \theta_f^*, P^*)}{\partial \theta_f'} \delta \\ &= \left[M_0(x; \theta_f^*, P^*) + \sum_{j=1}^{k_{\theta_f}} \frac{\partial M_0(x; \theta_f^*, P^*)}{\partial \theta_{f,j}} \delta_j \right] + \left[M_1(x; \theta_f^*, P^*) + \sum_{j=1}^{k_{\theta_f}} \frac{\partial M_1(x; \theta_f^*, P^*)}{\partial \theta_{f,j}} \delta_j \right] \beta \\ \alpha_f^{\mathcal{A}}(x; \delta) &= \alpha_f(x; \theta_f^*) + \frac{\partial \alpha_f(x; \theta_f^*)}{\partial \theta_f'} \delta. \end{aligned}$$

The derivative of the natural parameter of capital can easily be obtained as

$$\frac{\partial \alpha_f(k, s; \theta_f)}{\partial \theta_f'} = \begin{pmatrix} \frac{s}{\sigma_u^2} & \frac{s \log(k)}{\sigma_u^2} & \frac{1-s}{\sigma_u^2} & -\frac{1}{\sigma_u^4} [s(b_0 + b_1 \log(k)) + (1-s)b_e] \\ 0 & 0 & 0 & \frac{1}{2\sigma_u^4} \end{pmatrix}.$$

The derivative of the natural parameter of the choice probabilities involves the derivatives of the coefficients $M_0(\theta_f, P)$ and $M_1(\theta_f, P)$ with respect to θ_f , which are given as

$$\begin{aligned} \frac{\partial M_0(\theta_f, P)}{\partial \theta_{f,j}} &= \left\{ \rho \frac{\partial F_1(\theta_f)}{\partial \theta_{f,j}} (I_{M+1} - \rho F(P, \theta_f))^{-1} \right. \\ &\quad \left. + \rho^2 (F_1(\theta_f) - F_0) (I_{M+1} - \rho F(P, \theta_f))^{-1} \text{diag}(P_1) \frac{\partial F_1(\theta_f)}{\partial \theta_{f,j}} (I_{M+1} - \rho F(P, \theta_f))^{-1} \right\} \end{aligned}$$

$$\times \left(\gamma_e \cdot I_{M+1} - \sum_{d' \in \{0,1\}} P_{d'} \odot \log P_{d'} \right)$$

and

$$\begin{aligned} \frac{\partial M_1(\theta_f, P)}{\partial \theta_{f,j}} &= \left\{ \rho \frac{\partial F_1(\theta_f)}{\partial \theta_{f,j}} (I_{M+1} - \rho F(P, \theta_f))^{-1} \right. \\ &\quad \left. + \rho^2 (F_1(\theta_f) - F_0) (I_{M+1} - \rho F(P, \theta_f))^{-1} \text{diag}(P_1) \frac{\partial F_1(\theta_f)}{\partial \theta_{f,j}} (I_{M+1} - \rho F(P, \theta_f))^{-1} \right\} \\ &\quad \times \text{diag}(P_1) H \end{aligned}$$

for each $1 \leq j \leq k_{\theta_f}$. The derivative of the transition matrix $F_1(\theta_f)$ involves differentiating the elements of $F_{01}(\theta_f)$ and $F_{11}(\theta_f)$: for each $1 \leq i, j \leq M$, we have

$$\mu_i = b_0 + b_1 \log k_i, \quad z_{ij} = \frac{\log k_j - \mu_i}{\sigma_u}.$$

Then,

$$\begin{aligned} \frac{\partial [F_{11}(\theta_f)](i, j)}{\partial b_0} &= \begin{cases} -\frac{1}{\sigma_u} \phi(z_{ij}), & j = 1, \\ -\frac{1}{\sigma_u} \{ \phi(z_{i,j+1}) - \phi(z_{ij}) \}, & 1 < j < M, \\ \frac{1}{\sigma_u} \phi(z_{ij}), & j = M, \end{cases} \\ \frac{\partial [F_{11}(\theta_f)](i, j)}{\partial b_1} &= \log k_i \frac{\partial [F_{11}(\theta_f)](i, j)}{\partial b_0} \\ \frac{\partial [F_{11}(\theta_f)](i, j)}{\partial \sigma_u^2} &= \begin{cases} -\frac{\log k_j - \mu_i}{2\sigma_u^3} \phi(z_{ij}), & j = 1, \\ -\frac{1}{2\sigma_u^3} \{ (\log k_{j+1} - \mu_i) \phi(z_{i,j+1}) - (\log k_j - \mu_i) \phi(z_{ij}) \}, & 1 < j < M, \\ \frac{\log k_j - \mu_i}{2\sigma_u^3} \phi(z_{ij}), & j = M \end{cases} \end{aligned}$$

Now define

$$\mu = b_e, \quad z_j = \frac{\log k_j - \mu}{\sigma_u}.$$

Then,

$$\frac{\partial[F_{01}(\theta_f)](j)}{\partial b_e} = \begin{cases} -\frac{1}{\sigma_u} \phi(z_j) & j = 1, \\ -\frac{1}{\sigma_u} \{\phi(z_{j+1}) - \phi(z_j)\}, & 1 < j < M, \\ \frac{1}{\sigma_u} \phi(z_j), & j = M, \end{cases}$$

$$\frac{\partial[F_{01}(\theta_f)](j)}{\partial \sigma_u^2} = \begin{cases} -\frac{\log k_j - \mu}{2\sigma_u^3} \phi(z_j), & j = 1, \\ -\frac{1}{2\sigma_u^3} \{(\log k_{j+1} - \mu)\phi(z_{j+1}) - (\log k_j - \mu)\phi(z_j)\}, & 1 < j < M, \\ \frac{\log k_j - \mu}{2\sigma_u^3} \phi(z_j), & j = M \end{cases}$$

E.2 Prior Selection

Suppose that we use a sample $\{d_{it}, k_{it}\}_{0 \leq t \leq T, 1 \leq i \leq n_1}$ to form the first-stage estimates of θ_f^* based only on the log-normal state transition. If θ_f^* is found as the maximum likelihood estimator of the n_1 -sample likelihood

$$\ell_{n_1}^S(\theta_f) = \sum_{i=1}^{n_1} \sum_{t=1}^T d_{i,t-1} \{f(k_{it}) + \alpha_f(x_{i,t-1}; \theta_f)' S_f(k_{it}) - A_f(\alpha_f(x_{i,t-1}; \theta_f))\}.$$

Using the fact that, for each firm i , the sequence of score increments of the score function forms a martingale difference sequence with respect to the filtration $\mathcal{F}_i$ generated by $\{d_{it}, k_{it}, s_{it}\}_{0 \leq t \leq T}$, we can see that the per-firm Fisher information admits a closed form solution as

$$\begin{aligned} \mathbb{I}(\theta_f) &= \sum_{t=1}^T \mathbb{E} \left[d_{i,t-1} \left(\frac{\partial \alpha_f(x_{i,t-1}; \theta_f)}{\partial \theta_f'} \right)' \left\{ S_f(k_{it}) - \mathbb{E}[S_f(k_{it}) | \mathcal{F}_{i,t-1}] \right\} \right]^2 \\ &= \sum_{t=1}^T \mathbb{E} \left[d_{i,t-1} \left(\frac{\partial \alpha_f(x_{i,t-1}; \theta_f)}{\partial \theta_f'} \right)' \text{Var}[S_f(k_{it}) | \mathcal{F}_{i,t-1}] \left(\frac{\partial \alpha_f(x_{i,t-1}; \theta_f)}{\partial \theta_f'} \right) \right] \\ &= \sum_{t=1}^T \mathbb{P}(d_{i,t-1} = 1, s_{i,t-1} = 0) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sigma_u^2 & 0 \\ 0 & 0 & 0 & 1/(2\sigma_u^4) \end{pmatrix} \end{aligned}$$

$$+ \sum_{t=1}^T \mathbb{P}(d_{i,t-1} = 1, s_{i,t-1} = 1) \begin{pmatrix} 1/\sigma_u^2 & m_{1t}/\sigma_u^2 & 0 & 0 \\ m_{1t}/\sigma_u^2 & m_{2t}/\sigma_u^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/(2\sigma_u^4) \end{pmatrix}$$

for

$$m_{1t} = \mathbb{E} \left[\log k_{i,t-1} \mid d_{i,t-1} = s_{i,t-1} = 1 \right], \quad m_{2t} = \mathbb{E} \left[(\log k_{i,t-1})^2 \mid d_{i,t-1} = s_{i,t-1} = 1 \right].$$

The finite-sample analogue $\hat{\mathbb{I}}_{n_1}(\theta_f)$ of this quantity can be computed by replacing the probabilities and conditional expectations with the analogous finite-sample quantities. Then, a natural estimator of the sampling variance of θ_f^* is

$$\hat{V}^* = \frac{1}{n_1} \hat{\mathbb{I}}_{n_1}(\theta_f)^{-1}.$$

We let the prior covariance for the auxiliary parameters δ be given as $a_{n_1} \hat{V}^*$ for $a_n = \log(n)$.

E.3 Monte Carlo Design

Table E.1: **Monte Carlo Design: Dynamic Discrete Choice Model.** Structural calibration and data-generating process.

Parameter or setting	Symbol	Value
<i>Panel A: Structural parameters</i>		
Operating-profit coefficient	α	0.18
Entry cost	δ_p	0.80
Incumbent productivity intercept	b_0	0.00
Incumbent productivity persistence	b_1	0.78
Entrant productivity mean	b_e	0.50
Productivity innovation variance	σ_u^2	0.42
Discount factor	ρ	0.98
<i>Panel B: Data generation</i>		
Number of firms	N	1,000
Time periods per firm	T	10
Number of productivity grid points	M	25
Monte Carlo replications per design	R	500
Local misspecification magnitudes	h	$\{-4, -3, \dots, 3, 4\}$
First-stage sample size	N_1	1,000 firms
First-stage panel length	T_1	10 periods
First-stage random seed		64

Notes: The parameters of interest are $\theta_u = (\alpha, \delta_p)'$, while the nuisance parameters are $\theta_f = (b_0, b_1, b_e, \sigma_u^2)'$. The data are generated at the fixed true parameter vector (θ_u, θ_f) for every value of h . The nuisance calibration used for estimation is $\theta_f^*(h) = \theta_f + hd/\sqrt{N}$, where d is the worst-case first-stage direction described in Table E.2. The numerical productivity grid and the reference conditional choice probabilities are held fixed across misspecification designs.

Table E.2: **Monte Carlo Design: Dynamic Discrete Choice Model.** Estimation methods, priors, and numerical implementation.

Setting	Specification
<i>Panel A: Estimation methods</i>	
Baseline model	$\theta_u \in \mathbb{R}^2$; nuisance parameters fixed
Parsimoniously augmented model	$\theta_u \in \mathbb{R}^2$, auxiliary $\delta \in \mathbb{R}^2$
Joint model	$\theta_u \in \mathbb{R}^2$, nuisance deviation in $\mathbb{R}^4$
Reference CCPs	Fixed at $P^* = P_{eq}$
Productivity grid	Fixed across h and model evaluations
<i>Panel B: Priors and local perturbations</i>	
Prior for each structural parameter	$\mathcal{N}(0.5, 0.25^2)$, truncated below at zero
First-stage covariance	$\widehat{I}_{ff}^{-1} \log(N_1)/N_1$
Joint-model nuisance prior	Four-dimensional Gaussian, truncated at the 90% Mahalanobis radius
Parsimonious nuisance prior	Two-dimensional induced Gaussian, truncated at the 90% Mahalanobis radius
Worst-case direction	Computed from the joint Fisher information and first-stage information
Parsimonious loading	$B_0 = \widehat{I}_{ff}^{-1} \widehat{I}_{fu}$
<i>Panel C: Posterior computation</i>	
Sampling algorithm	NUTS, target acceptance 0.8
Chains	4
Requested iterations per chain	1,000
Initial iterations discarded	500
Nominal credible interval level	95%
Credible interval construction	Posterior quantiles 0.025 and 0.975

Notes: The first-stage information matrix is estimated from a fixed, independently simulated panel. The worst-case perturbation direction is computed once and held fixed across the Monte Carlo experiments. The parsimonious loading matrix and its induced prior covariance are recomputed for each Monte Carlo replication using the baseline posterior medians and the corresponding sample information matrix. The joint model recomputes the Hotz–Miller continuation-value objects at each proposed nuisance parameter vector on the fixed productivity grid. The baseline and augmented models use precomputed continuation-value objects.

E.4 Sampler Diagnostics

Table E.3: HMC Sampling Diagnostics: Dynamic Discrete Choice Model

Model	Parameter	ESS	Tail ESS	$\widehat{R}$
Baseline	α	1120.9	1116.9	0.9995
	δ_p	1156.2	1211.5	1.0071
Parsimonious Augmentation	α	767.7	1178.5	1.0021
	δ_p	1081.9	1071.1	1.0046
	z_1	745.2	1164.5	1.0030
	z_2	868.7	1374.5	1.0042
Full Estimation	α	2174.3	1558.7	1.0013
	δ_p	2600.5	1649.3	1.0012
	z_1	1905.6	1423.1	1.0008
	z_2	2150.3	1493.3	1.0017
	z_3	2483.0	1621.2	1.0029
	z_4	2768.6	1393.5	1.0024

Notes: Diagnostics are computed from four HMC chains, each containing 500 retained draws after 500 warmup iterations. ESS and $\widehat{R}$ are obtained using `MCMCchains.ess_rhat`. Tail ESS is computed as the minimum of the estimated effective sample sizes for the indicator processes associated with the 2.5th and 97.5th posterior quantiles, summed across the four chains. The auxiliary coordinates in the parsimonious augmentation are rotated nuisance directions and are not directly comparable to the original nuisance coordinates in the full model.

F Details on Krusell-Smith Model Simulation

F.1 Numerical Differentiation for KS Model

The GE Jacobian for output depends directly on the nuisance parameters η (through, for instance, computation of steady state production), on the steady state interest rate implied by η (through steady state capital), and the sequence space Jacobians $\mathcal{J}(\eta)$:

$$G(\eta) = G_f[\eta, r^*(\eta), \mathcal{J}(\eta)].$$

The sequence space Jacobians $\mathcal{J}(\eta)$ in turn are functions directly of the nuisance parameters η , and depend on η indirectly through the steady state interest rate $r^*(\eta)$, the steady state policy vector $y_{ss}(\eta)$ and the steady state distribution vector $D_{ss}(\eta)$. Furthermore, the latter two quantities also depend directly on η and indirectly on η through $r^*(\eta)$. These dependencies are made clear through

the expression

$$\mathcal{J}(\eta) = \mathcal{J}_f [\eta, r^*(\eta), y_{ss}(\eta, r^*(\eta)), D_{ss}(\eta, r^*(\eta))] .$$

The numerical two-sided derivative of $G(\eta)$ with respect to η_i at η^* can be computed as

$$\frac{G(\eta^* + h \cdot e_i) - G(\eta^* - h \cdot e_i)}{2h}$$

where e_i is the i th standard basis vector of $\mathbb{R}^{k_\eta}$ and $h > 0$ is some small positive constant (we put $h = 1e - 5$) in practice. We can compute $G(\eta^* \pm h \cdot e_i)$ as soon as we obtain an expression for

$$r^*(\eta^* \pm h \cdot e_i),$$

since then computing $G(\eta^* \pm h \cdot e_i)$ is simply a matter of plugging in the values $\eta^* \pm h \cdot e_i$ and $r^*(\eta^* \pm h \cdot e_i)$ into the functions $\mathcal{G}_f(\cdot)$ and $\mathcal{J}_f(\cdot)$ that we have already constructed.

Conceptually, it should be possible to compute $r^*(\eta^* \pm h \cdot e_i)$ directly through the bisection loop; but this would then involve solving for the steady state $2k_\eta$ times, which is computationally burdensome. A computationally simpler approach is to employ a first order approximation of $r^*(\eta^* \pm h \cdot e_i)$ around η^* :

$$r^*(\eta^* \pm h \cdot e_i) \approx r^*(\eta^*) + \frac{\partial r^*(\eta^*)}{\partial \eta_i} h.$$

Given any η , the steady state interest rate $r^*(\eta)$ is defined implicitly via the fixed point relation

$$K^s(\eta, r^*(\eta)) - K^d(\eta, r^*(\eta)) = 0$$

where $K^s(\eta, r)$ denotes aggregate savings and

$$K^d(\eta, r) = \left(\frac{\alpha}{r + \delta} \right)^{\frac{1}{1-\alpha}} L^*(\eta)$$

is capital demand given interest rate r and the nuisance parameters $\eta = \{\alpha, \beta, \delta, \gamma, \rho_n, \sigma_n^2, m_n\}$.

Note further that aggregate savings are given as the inner product

$$K^s(\eta, r) = D_{ss}(\eta, r)' y_{ss}(\eta, r).$$

As such, by the implicit function theorem,

$$\frac{\partial r^*(\eta^*)}{\partial \eta_i} = - \frac{\left(\frac{\partial D_{ss}(\eta^*, r^*(\eta^*))}{\partial \eta_i} \right)' y_{ss}(\eta^*, r^*(\eta^*)) + D_{ss}(\eta^*, r^*(\eta^*))' \left(\frac{\partial y_{ss}(\eta^*, r^*(\eta^*))}{\partial \eta_i} \right) - \frac{\partial K^d(\eta^*, r^*(\eta^*))}{\partial \eta_i}}{\left(\frac{\partial D_{ss}(\eta^*, r^*(\eta^*))}{\partial r} \right)' y_{ss}(\eta^*, r^*(\eta^*)) + D_{ss}(\eta^*, r^*(\eta^*))' \left(\frac{\partial y_{ss}(\eta^*, r^*(\eta^*))}{\partial r} \right) - \frac{\partial K^d(\eta^*, r^*(\eta^*))}{\partial r}}.$$

We compute the derivatives of $D_{ss}(\eta^*, r^*(\eta^*))$ and $y_{ss}(\eta^*, r^*(\eta^*))$ via two-sided numerical differentiation. Generically, given two input values $\bar{\eta}_+$ and $\bar{\eta}_-$ around $(\eta^*, r^*(\eta^*))$, we denote

$$dD_{ss} = D_{ss}(\bar{\eta}_+) - D_{ss}(\bar{\eta}_-)$$

$$dy_{ss} = y_{ss}(\bar{\eta}_+) - y_{ss}(\bar{\eta}_-).$$

dy_{ss} we can compute by computing $y_{ss}(\bar{\eta}_+)$ and $y_{ss}(\bar{\eta}_-)$ separately using the household EGM algorithm. Computation of dD_{ss} is more complicated, since $D_{ss}(\eta, r)$ is defined as the stationary distribution of the transition matrix $\Lambda_{ss}(\eta, r)$:

$$D_{ss}(\eta, r) = \Lambda_{ss}(\eta, r) D_{ss}(\eta, r).$$

It follows that

$$dD_{ss} \approx (d\Lambda_{ss}) D_{ss}(\eta^*, r^*(\eta^*)) + \Lambda_{ss}(\eta^*, r^*(\eta^*)) dD_{ss}$$

where

$$d\Lambda_{ss} = \Lambda_{ss}(\bar{\eta}_+) - \Lambda_{ss}(\bar{\eta}_-)$$

can be computed by passing $y_{ss}(\bar{\eta}_+)$ and $y_{ss}(\bar{\eta}_-)$ through the histogram method. Now dD_{ss} solves the equation

$$[I_{NJ} - \Lambda_{ss}(\eta^*, r^*(\eta^*))] dD_{ss} = (d\Lambda_{ss}) D_{ss}(\eta^*, r^*(\eta^*)).$$

Since $\Lambda_{ss}(\eta^*, r^*(\eta^*))$ has exactly one eigenvalue equal to 1 (being a column stochastic matrix), $I_{NJ} - \Lambda_{ss}(\eta^*, r^*(\eta^*))$ has rank $NJ - 1$ and this equation cannot be inverted immediately. Instead, we make use of the additional condition $\iota' dD_{ss} = 0$, where ι is the NJ -dimensional vector of ones; this equation requires simply that the total mass of households remains unchanged across different values of the nuisance parameter. Now we can solve for dD_{ss} by replacing the first row of $I_{NJ} - \Lambda_{ss}(\eta^*, r^*(\eta^*))$ with ι' , the first element of $(d\Lambda_{ss}) D_{ss}(\eta^*, r^*(\eta^*))$ by 0 and inverting the resulting linear equation.

This first order approximation approach is much less computationally intensive compared to computing $r^*(\eta^* \pm h \cdot e_i)$ through the bisection loop, since it involves running only a few backward + forward recursions to obtain the derivatives of the policy rule and steady state distributions. The gain in terms of computational time scales with the number of fixed point relations that characterize the steady state values of the model.

To summarize, the numerical differentiation procedure for the GE Jacobian consists of three nested computations. Fix $1 \leq i \leq k_\eta$, $h > 0$ and denote by r^* the steady state interest rate under η^* .

1) Derivative of Interest Rate

We compute $\frac{\partial r^*(\eta^*)}{\partial \eta_i}$ using the implicit function theorem:

$$\frac{\partial r^*(\eta^*)}{\partial \eta_i} \approx - \frac{d_i D'_{ss} y_{ss} + D'_{ss}(d_i y_{ss}) - d_i K^d}{d_r D'_{ss} y_{ss} + D'_{ss}(d_r y_{ss}) - d_r K^d}$$

where

$$d_i D_{ss} = D_{ss}(\eta^* + h \cdot e_i, r_{ss}) - D_{ss}(\eta^* - h \cdot e_i, r_{ss})$$

$$d_r D_{ss} = D_{ss}(\eta^*, r_{ss} + h) - D_{ss}(\eta^*, r_{ss} - h)$$

and likewise for the other quantities. We define

$$\bar{\eta}_\pm = \left(\eta^* \pm h \cdot e_i, r_{ss} \pm \frac{\partial r^*(\eta^*)}{\partial \eta_i} h \right).$$

2) Deviations of Steady State Quantities

We compute the perturbed steady state savings policy function and steady state distribution

$$y_{ss}^\pm = y_{ss}(\bar{\eta}_\pm), \quad D_{ss}^\pm = D_{ss}(\bar{\eta}_{pm}).$$

3) Numerical Derivative of GE Jacobians

We compute the perturbed Jacobians $G^\pm$ using $(\bar{\eta}_{pm}, y_{ss}^\pm, D_{ss}^\pm)$ as inputs. The numerical derivative of $G(\eta)$ at η^* with respect to η_i is then given as

$$\frac{\partial G(\eta^*)}{\partial \eta_i} \approx \frac{G^+ - G^-}{2h}.$$

F.2 Monte Carlo Design

Table F.1: **Monte Carlo Design: Krusell–Smith Model.** Structural calibration and numerical solution.

Parameter or setting	Symbol	Value
<i>Panel A: Structural calibration</i>		
Capital share	α^*	0.33
Depreciation rate	δ_d	0.025
Risk aversion	γ	1
Idiosyncratic productivity persistence	ρ_n	0.966
Idiosyncratic productivity grid scale	Δ_n	0.50
Idiosyncratic innovation variance	σ_n^2	0.016711
Steady-state interest rate target	r_{ss}	0.01
Steady-state output target	Y_{ss}	1
Household discount factor	β_d	Endogenously calibrated
Log-productivity mean	m_n	Endogenously calibrated
<i>Panel B: Numerical solution</i>		
Idiosyncratic productivity states	$n + 1$	7
Household asset-grid points	M	500
Distribution histogram points	N_{hist}	5,000
Minimum asset holdings	$a_{\min}$	0
Maximum asset holdings	$a_{\max}$	100
Equilibrium convergence tolerance		10^{-8}
Household convergence tolerance		10^{-8}
Stationary-distribution tolerance		10^{-8}
Equilibrium interest-rate update weight		0.05
Finite-difference step size	ε	10^{-5}
Sequence-space Jacobian horizon	T_{IRF}	300

Notes: Idiosyncratic labor productivity is approximated by a seven-state Rouwenhorst Markov chain. The innovation variance is calibrated as $\sigma_n^2 = \Delta_n^2(1 - \rho_n^2)$. The household discount factor is calibrated to reproduce the steady-state interest-rate target, while the productivity mean is calibrated to match the labor-supply normalization. The household problem is solved using the endogenous grid method, and aggregate dynamics are computed using the sequence-space Jacobian method. Numerical derivatives are obtained using centered finite differences.

Table F.2: **Monte Carlo Design: Krusell–Smith Model.** Local misspecification, priors, and posterior computation.

Setting	Specification
<i>Panel A: Data-generating process</i>	
Aggregate productivity persistence	$\rho_{z,0} = 0.70$
Aggregate innovation variance	$\sigma_{z,0}^2 = 0.10$
Local misspecification	$\alpha_{0,T} = \alpha^* + h/\sqrt{T}$
Reference capital share	$\alpha^* = 0.33$
Misspecification magnitudes	$h \in \{0.3, 0.5\}$
Sample sizes	$T \in \{500, 1000\}$
Monte Carlo replications	100 per design
Data-generating distribution	Stationary Gaussian time series
<i>Panel B: Estimation methods and priors</i>	
Parameters of interest	(ρ_z, σ_z^2)
Baseline nuisance calibration	$\alpha = \alpha^*$
Augmented nuisance dimension	1
Prior for ρ_z	Beta(2.625, 2.625)
Prior for σ_z^2	IG(4.56, 0.5696)
Untruncated auxiliary prior	$\mathcal{N}(0, \Sigma_\delta)$
Auxiliary prior variance	$\Sigma_\delta = 0.15^2/(\partial x/\partial \alpha)^2$
Auxiliary prior truncation	$\delta^2/\Sigma_\delta \leq \chi_{1,0.90}^2$
<i>Panel C: Posterior computation</i>	
Sampling algorithm	Laplace-informed adaptive RWMH
Mode initialization	Nelder–Mead optimization
Proposal covariance	Regularized inverse negative Hessian
Hessian calculation	Forward-mode automatic differ- entiation
MCMC iterations	10,000
Burn-in iterations	5,000
Target acceptance probability	0.25
BLAS threads	1
Parameter credible interval level	95%
IRF interval construction	Pointwise 95% posterior quan- tiles
IRF variables	Output, capital, consumption

Notes: For each Monte Carlo replication, a Gaussian time series is generated using the equilibrium mean and covariance implied by the locally perturbed capital share. The baseline model fixes the capital share at α^* , whereas the augmented model estimates a scalar auxiliary parameter representing its local deviation. The auxiliary prior is a centered Gaussian truncated to its 90% Mahalanobis region. The MCMC proposal is initialized using a local Gaussian approximation at the posterior mode, with scalar proposal-scale adaptation during burn-in. Parameter coverage is evaluated using equal-tailed credible intervals. Impulse-response coverage is computed pointwise at each reported horizon.

F.3 Sampler Diagnostics

Below we present effective sample sizes for the RWMH sampler used to estimate the baseline and augmented models, as well as trace plots for the sampled parameters.

Table F.3: MCMC Sampling Diagnostics: Krusell–Smith Model

Model	Parameter	ESS	Tail ESS	Acceptance Rate
Baseline	ρ_z	601.8	977.5	0.2488
	σ_z^2	485.0	599.1	0.2488
Augmented	ρ_z	471.9	642.1	0.2488
	σ_z^2	387.1	403.7	0.2488
	δ_1	425.9	493.2	0.2488

Notes: The table reports sampling diagnostics for the baseline and augmented random-walk Metropolis–Hastings samplers for a representative Monte Carlo replication. Each sampler uses 10,000 iterations, with the first 5,000 discarded as burn-in, as in the Monte Carlo design detailed in the preceding section. ESS denotes the effective sample size computed from the autocorrelation of retained posterior draws. Tail ESS is the minimum of the effective sample sizes for the indicator sequences associated with the 2.5th and 97.5th posterior quantiles. Both ESS measures use an initial-positive-sequence autocorrelation estimator. Acceptance rates are computed over post-burn-in iterations.

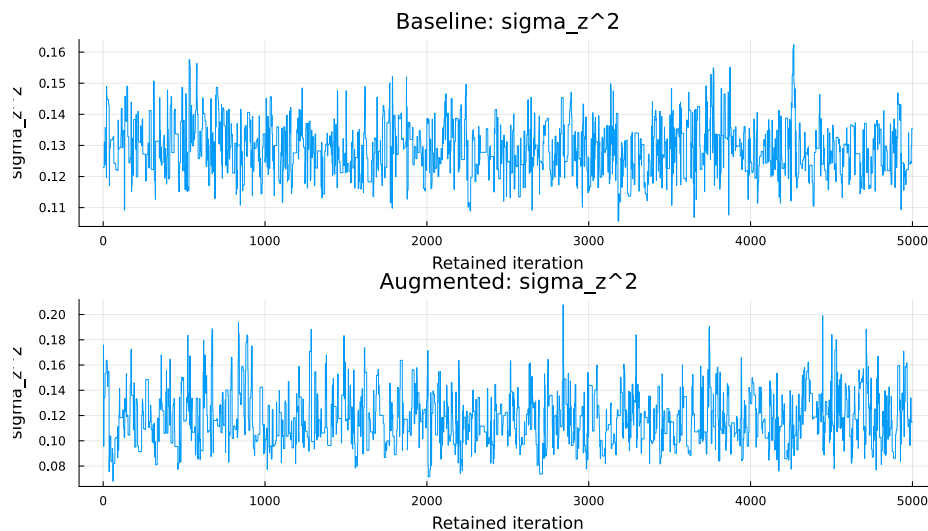


Figure F.1: MCMC trace plots for σ_z^2 in the Krusell–Smith model.

Notes: Baseline (upper panel) and augmented (lower panel) specifications; 5,000 retained draws after 5,000 burn-in iterations.

F.4 Robustness Checks for Misspecification Magnitude

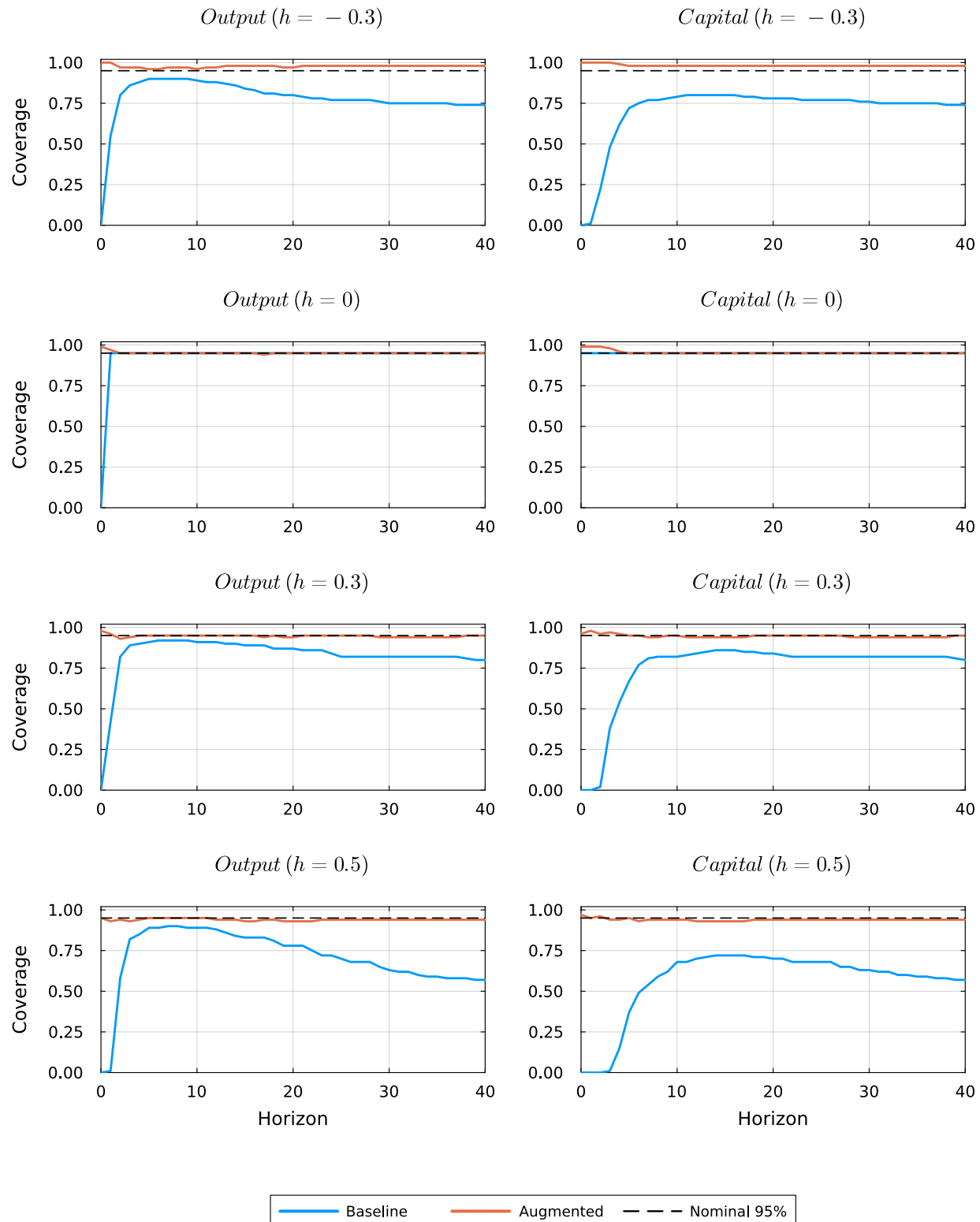
Here we present results for the baseline and augmented model for an expanded set of local misspecification magnitudes. In the main text, we focused on local misspecification of the form $\alpha_0 = \alpha^* + h/\sqrt{T}$ where $h = \{0.3, 0.5\}$. Below are the shock parameter and impulse response coverage results for $h \in \{-0.3, 0, 0.3, 0.5\}$. The same qualitative results hold for $h = -0.3$, demonstrating that the robustness of the augmented model and the failure of the baseline model does not depend on the direction of the local misspecification. Meanwhile, for $h = 0$ the models exhibit similar coverage, although the augmented model reports substantially wider uncertainty for σ_z^2 .

Table F.4: **Monte Carlo study under local misspecification.** We report the coverage of nominal 95% credible intervals, the mean absolute error of the posterior mode, and the average posterior standard deviation for the shock parameters ρ_z and σ_z^2 .

h	T	Model	Coverage		Mean Absolute Error		Posterior SD	
			ρ_z	σ_z^2	ρ_z	σ_z^2	ρ_z	σ_z^2
-0.3	500	Baseline	0.95	0.31	0.0260	0.0147	0.0319	0.0054
		Augmented	0.96	1.00	0.0270	0.0111	0.0324	0.0162
	1000	Baseline	0.98	0.30	0.0175	0.0113	0.0227	0.0040
		Augmented	0.97	0.95	0.0179	0.0093	0.0229	0.0121
0	500	Baseline	0.95	0.95	0.0260	0.0054	0.0320	0.0063
		Augmented	0.95	0.98	0.0270	0.0117	0.0323	0.0168
	1000	Baseline	0.98	0.94	0.0174	0.0040	0.0227	0.0044
		Augmented	0.97	0.94	0.0179	0.0095	0.0229	0.0123
0.3	500	Baseline	0.96	0.22	0.0256	0.0192	0.0317	0.0075
		Augmented	0.95	0.96	0.0269	0.0128	0.0321	0.0175
	1000	Baseline	0.96	0.22	0.0171	0.0127	0.0227	0.0050
		Augmented	0.98	0.94	0.0179	0.0098	0.0229	0.0126
0.5	500	Baseline	0.95	0.01	0.0255	0.0347	0.0313	0.0085
		Augmented	0.96	0.95	0.0268	0.0140	0.0323	0.0179
	1000	Baseline	0.95	0.00	0.0172	0.0227	0.0226	0.0055
		Augmented	0.97	0.94	0.0179	0.0102	0.0229	0.0128

Notes: Coverage is computed from nominal 95% posterior credible intervals over 100 Monte Carlo replications. Mean absolute error is the Monte Carlo average of $|\hat{\theta}_{\text{mode}} - \theta_0|$. Posterior standard deviations are averaged across Monte Carlo replications. The nuisance parameter is the capital share α , with local misspecification $\alpha_0 = \alpha^* + h/\sqrt{T}$ and $\alpha^* = 0.33$. The case $h = 0$ corresponds to correct nuisance calibration.

Figure F.2: **Monte Carlo study for impulse response coverage.** Below we display the coverage of pointwise credible bands for the impulse response of output and capital to TFP shocks up to 40 horizons. The impulse responses are evaluated for the baseline and augmented models. The nuisance parameter is the capital share α , with local misspecification $\alpha_0 = \alpha^* + h/\sqrt{T}$ and $\alpha^* = 0.33$.



References

- Aguirregabiria, Victor and Pedro Mira (2002). “Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models”. In: *Econometrica* 70.4, p. 1519.
- Kleijn, B. J. K. and A. W. van der Vaart (2012). “The Bernstein-Von-Mises theorem under misspecification”. In: *Electronic Journal of Statistics* 6, pp. 354–381.
- Rust, John (1987). “Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher”. In: *Econometrica: Journal of the Econometric Society*, pp. 999–1033.