

The Convex Origin of Fixed Costs¹

Online Appendix

Contents

A Proofs from Section 2 (The duality)	2
A.1 Proof of Proposition 1 (Cost function properties)	2
A.2 Proof of Corollary 1 (Shape of G determines curvature)	2
A.3 Proof of Corollary 2 (Canonical distributions)	2
A.4 Proof of Proposition 2 (Universality)	2
A.5 Topological genericity	3
A.6 Proof of Proposition 3 (Dispersion and convexity)	3
A.7 Proof of Proposition 4 (Entropy correspondence)	3
A.8 Detailed entropy discussion	4
A.9 Proof of Proposition 5 (Multi-action duality)	5
B Investment application	6
C Proofs from the pricing section	8
C.1 Proof of Proposition 7 (Stationary distribution)	8
C.2 Proof of Proposition 8 (Selection and the menu cost distribution)	9
C.3 Proof of Proposition 9 (Calvo limit)	10
C.4 Proof of Proposition 10 (Nonlinear Phillips curve)	10
C.5 Proof of Proposition 12 (Monotonicity in γ)	11
C.6 Proof of Proposition 13 (Frequency does not determine non-neutrality)	12
C.7 Proof of Corollary 5 ($\kappa = (1 + 2\gamma)\bar{\Lambda}$ for power laws)	12
C.8 Proof of Corollary 4 (Ordering of non-neutrality)	12
D Supplementary figures	13
E Comparison with Williams-Daly-Zachary	13

¹Hanbaek Lee, University of Cambridge, Email: hl610@cam.ac.uk

A Proofs from Section 2 (The duality)

A.1 Proof of Proposition 1 (Cost function properties)

Proof.

Differentiating $C(\varphi) = w \int_0^{G^{-1}(\varphi)} \xi dG(\xi)$: $C'(\varphi) = w G^{-1}(\varphi) \cdot g(G^{-1}(\varphi)) \cdot [g(G^{-1}(\varphi))]^{-1} = w G^{-1}(\varphi)$, using the inverse function theorem. Differentiating again: $C''(\varphi) = w \cdot [g(G^{-1}(\varphi))]^{-1} > 0$. ■

A.2 Proof of Corollary 1 (Shape of G determines curvature)

Proof.

By Proposition 1(iii), $C''(\varphi) = w/g(G^{-1}(\varphi))$. Since G^{-1} is increasing, g increasing implies $g(G^{-1}(\varphi))$ is increasing in φ , so $C''(\varphi) = w/g(G^{-1}(\varphi))$ is decreasing. The argument for g decreasing and g constant is analogous. ■

A.3 Proof of Corollary 2 (Canonical distributions)

Proof.

Apply $C(\varphi) = w \int_0^\varphi G^{-1}(u) du$ from Proposition 1(ii). (i) Uniform: $G^{-1}(\varphi) = \bar{\xi}\varphi$, so $C(\varphi) = w\bar{\xi} \int_0^\varphi u du = \frac{w\bar{\xi}}{2}\varphi^2$. (ii) Power law: $G^{-1}(\varphi) = \bar{\xi}\varphi^{1/\gamma}$, so $C(\varphi) = w\bar{\xi} \int_0^\varphi u^{1/\gamma} du = \frac{\gamma}{\gamma+1}w\bar{\xi}\varphi^{1+1/\gamma}$. (iii) Exponential: $G^{-1}(\varphi) = -\lambda^{-1}\ln(1-\varphi)$, so $C(\varphi) = -\frac{w}{\lambda} \int_0^\varphi \ln(1-u) du = \frac{w}{\lambda}[(1-\varphi)\ln(1-\varphi) + \varphi]$. (iv) Degenerate at $\bar{\xi}$: $G^{-1}(\varphi) = \bar{\xi}$ for all $\varphi \in (0, 1]$, giving $C(\varphi) = w\bar{\xi}\varphi$. ■

A.4 Proof of Proposition 2 (Universality)

Proof.

For $G_\epsilon = U[\bar{\xi} - \epsilon, \bar{\xi} + \epsilon]$, the quantile function is $G_\epsilon^{-1}(\varphi) = (\bar{\xi} - \epsilon) + 2\epsilon\varphi$ and the density is $g_\epsilon = 1/(2\epsilon)$. Applying $C(\varphi) = w \int_0^\varphi G^{-1}(t) dt$:

$$C_\epsilon(\varphi) = w \int_0^\varphi G_\epsilon^{-1}(t) dt = w \int_0^\varphi [(\bar{\xi} - \epsilon) + 2\epsilon t] dt = w(\bar{\xi} - \epsilon)\varphi + w\epsilon\varphi^2.$$

Strict convexity: $C''_\epsilon(\varphi) = 2w\epsilon > 0$ for all $\epsilon > 0$. The first-order condition $S = C'_\epsilon(\varphi) = w(\bar{\xi} - \epsilon) + 2w\epsilon\varphi$ gives $\varphi_\epsilon^*(S) = [S - w(\bar{\xi} - \epsilon)]/(2w\epsilon)$, which is interior whenever S falls in the interval $(w(\bar{\xi} - \epsilon), w(\bar{\xi} + \epsilon))$. As $\epsilon \rightarrow 0$, this interval collapses to the single point $w\bar{\xi}$, and the interior solution collapses to a step function: φ^* jumps discontinuously from 0 to 1 at $S = w\bar{\xi}$. ■

A.5 Topological genericity

The claim that lumpiness is non-generic can be made precise in the weak topology on the space of probability measures on $[0, \bar{\xi}]$. A sequence of measures μ_n converges weakly to μ if $\int f d\mu_n \rightarrow \int f d\mu$ for every bounded continuous f . The degenerate measure $\delta_{\bar{\xi}}$ (a point mass at $\bar{\xi}$) is the unique element of this space that induces a linear cost function and a binary threshold rule. Yet every weakly open neighbourhood of $\delta_{\bar{\xi}}$ contains measures with continuous positive densities, and hence strictly convex induced costs. To see this, let U be any weakly open neighbourhood of $\delta_{\bar{\xi}}$. By the Portmanteau theorem, U contains all measures μ satisfying $|\int f_i d(\mu - \delta_{\bar{\xi}})| < \epsilon$ for finitely many bounded continuous test functions f_1, \dots, f_k . The uniform perturbation $G_\epsilon = U[\bar{\xi} - \epsilon, \bar{\xi} + \epsilon]$ satisfies $\int f_i dG_\epsilon \rightarrow f_i(\bar{\xi})$ as $\epsilon \rightarrow 0$ by continuity, so $G_\epsilon \in U$ for ϵ small enough. Since G_ϵ has a continuous positive density, it induces a strictly convex C_ϵ by Proposition 2. Therefore $\delta_{\bar{\xi}}$ lies in the closure of the set of absolutely continuous measures: the degenerate (lumpy) case is a limit point of the convex (smooth) case, and every neighbourhood of it contains non-degenerate, strictly-convex instances.

A.6 Proof of Proposition 3 (Dispersion and convexity)

Proof.

Part (i): By Proposition 1(i), $C_i(1) = w E_{G_i}[\xi]$. Same mean gives $C_1(1) = C_2(1)$.

Part (ii): By Proposition 1(ii), $C_i(\varphi) = w \int_0^\varphi G_i^{-1}(u) du$. The standard quantile characterisation of mean-preserving spreads states that G_2 is a MPS of G_1 if and only if

$$\int_0^t G_2^{-1}(u) du \leq \int_0^t G_1^{-1}(u) du \quad \text{for all } t \in [0, 1],$$

with equality at $t = 1$. Therefore $C_1(\varphi) - C_2(\varphi) = w \int_0^\varphi [G_1^{-1}(u) - G_2^{-1}(u)] du \geq 0$ for all φ , with equality at $\varphi \in \{0, 1\}$.

Part (iii): Since $C'_i(\varphi) = w G_i^{-1}(\varphi)$ and both quantile functions are continuous and increasing with the same integral over $[0, 1]$, while G_2^{-1} is stretched relative to G_1^{-1} (lower at the bottom, higher at the top), the two marginal cost schedules cross exactly once. ■

A.7 Proof of Proposition 4 (Entropy correspondence)

Proof.

Part (i): The quantile function is $G^{-1}(\varphi) = -\lambda^{-1} \ln(1 - \varphi)$. Then

$$C(\varphi) = w \int_0^\varphi G^{-1}(t) dt = -\frac{w}{\lambda} \int_0^\varphi \ln(1 - t) dt = \frac{w}{\lambda} [(1 - \varphi) \ln(1 - \varphi) + \varphi].$$

Parts (ii)–(iii) follow from Corollary 2. ■

A.8 Detailed entropy discussion

Rational inattention and exponential fixed costs [Sims \(2003\)](#) introduces rational inattention, where agents face a Shannon mutual information cost of processing information about a continuous state. [Matějka and McKay \(2015\)](#) show that in the discrete-choice specialisation, this cost reduces to a Shannon entropy penalty over the choice probability vector, yielding logit choice probabilities. Proposition 4(i) shows that the cost function induced by exponentially-distributed fixed costs is exactly the Shannon entropy of the binary adjustment probability — the cost function in the binary discrete-choice rational inattention model. For binary choice, the two models are mathematically equivalent: they produce the same adjustment probabilities and the same payoffs. This equivalence does not extend to the full continuous-state RI framework of [Sims \(2003\)](#), where the mutual information cost has a richer structure involving Bayesian updating over a continuous state space.

Sparsity and uniform fixed costs [Gabaix \(2014\)](#) penalises attention allocation with a quadratic cost, yielding “sparse” decisions that underreact to stimuli. By Proposition 4(iii), the cost function induced by uniform fixed costs is exactly this quadratic penalty. The equivalence is at the level of the attention cost function for the binary decision; the full sparsity framework includes additional structure (endogenous attention allocation across multiple dimensions) not captured by the binary duality.

Scope of the correspondences The connection to rational inattention is worth making explicit. [Matějka and McKay \(2015\)](#) show that when an agent faces uncertainty about a state and processes information subject to a Shannon entropy cost, the optimal choice rule takes the multinomial logit form. In the multi-action extension, i.i.d. Gumbel-distributed fixed costs yield $C(\varphi) = \sigma \sum_j \varphi_j \ln \varphi_j$ (negative Shannon entropy), and the multinomial logit emerges as the optimal mixed strategy. Every agent who faces i.i.d. extreme-value adjustment costs and uses an optimal threshold rule is behaving exactly as if they were a rational inattention agent with Shannon entropy costs. The two models make identical predictions about adjustment frequencies and payoffs.

However, the economic mechanisms are distinct. In rational inattention, the agent is uncertain about the payoff-relevant state and incurs a cost of learning. In the present framework, the agent knows the surplus S and incurs a physical cost of acting. The reduced-form $C(\varphi)$ is the same; the welfare implications are not.

A.9 Proof of Proposition 5 (Multi-action duality)

Proof.

Part (i): General case. The agent observes cost draws (ξ_1, \dots, ξ_J) with $\xi_j \sim G_j$ independently, and selects the action with the highest net payoff:

$$\Pi^{FC}(\mathbf{S}) = E \left[\max_{j \in \{0, \dots, J\}} \{S_j - w\xi_j\} \right].$$

Partition the expectation by which action is optimal. For each $j \geq 1$, action j is chosen when $S_j - w\xi_j \geq S_k - w\xi_k$ for all $k \neq j$ and $S_j - w\xi_j \geq 0$. Under independence, the probability that action j is chosen can be written as a function $\varphi_j(\mathbf{S})$, and the expected cost paid conditional on choosing j is $E[w\xi_j \mid j \text{ chosen}]$.

Define $\varphi_0 = 1 - \sum_{j=1}^J \varphi_j$ (the probability of inaction). The expected payoff is:

$$\Pi^{FC}(\mathbf{S}) = \sum_{j=0}^J \varphi_j^* S_j - \sum_{j=1}^J E[w\xi_j \cdot \mathbf{1}_{j \text{ chosen}}].$$

By the Fenchel conjugate construction, $\Pi^{FC}(\mathbf{S}) = \sup_{\varphi \in \Delta^J} \{\sum_j \varphi_j S_j - C(\varphi)\}$ where $C(\varphi)$ is the convex conjugate of Π^{FC} . Under independence of the G_j , the cost function inherits a separable structure: the marginal cost of raising φ_j depends only on G_j , not on the other distributions.

Joint convexity of C follows from the fact that $\Pi^{FC}(\mathbf{S})$, as a pointwise maximum of affine functions of \mathbf{S} , is convex in \mathbf{S} . The Fenchel biconjugate theorem then implies that C , as the conjugate of Π^{FC} , is convex on Δ^J .

Part (ii): i.i.d. Gumbel specialisation. Let $\xi_j \stackrel{i.i.d.}{\sim} \text{Gumbel}(\mu, \sigma)$ for $j = 0, 1, \dots, J$, with CDF $G(\xi) = e^{-e^{-(\xi-\mu)/\sigma}}$. The expected maximum of $\{S_j - w\xi_j\}$ is a classical result (McFadden, 1974):

$$\Pi^{FC}(\mathbf{S}) = \sigma \ln \left(\sum_{j=0}^J e^{S_j/\sigma} \right) + \mu',$$

where $\mu' = w\mu - w\sigma\gamma_E$ and γ_E is the Euler–Mascheroni constant. This is the log-sum-exp function (up to constants), which is convex in \mathbf{S} .

The gradient gives the choice probabilities:

$$\varphi_j^* = \frac{\partial \Pi^{FC}}{\partial S_j} = \frac{e^{S_j/\sigma}}{\sum_{k=0}^J e^{S_k/\sigma}},$$

the multinomial logit. The Fenchel conjugate of the log-sum-exp is the negative

entropy:

$$C(\varphi) = \sup_{\mathbf{S}} \left\{ \sum_j \varphi_j S_j - \sigma \ln \left(\sum_j e^{S_j/\sigma} \right) \right\} - \mu' = \sigma \sum_{j=0}^J \varphi_j \ln \varphi_j,$$

which is non-negative on Δ^J (since $\ln \varphi_j \leq 0$) and jointly convex. \blacksquare

Remark 1. *The Gumbel distribution has unbounded support $(-\infty, \infty)$, so negative cost draws are possible. In the original fixed cost interpretation, this means that some realisations yield a “bonus” rather than a cost. The duality holds provided $E[\max_j \{S_j - w\xi_j\}]$ is finite, which is guaranteed by the double-exponential tails of the Gumbel. For distributions on $[0, \bar{\xi}]$ (bounded support), the convergence is automatic.*

B Investment application

Consider a firm with capital k choosing investment $I \geq 0$ subject to a random fixed cost $\xi \sim G$. If it adjusts, it pays $w\xi$ and obtains the continuation value at capital $k(1 - \delta) + I$. If not, it proceeds with depreciated capital $k(1 - \delta)$. The surplus from investing at the optimal scale is

$$S(k) = \max_{I \geq 0} \{-I + \beta(k(1 - \delta) + I)^\alpha\} - \beta(k(1 - \delta))^\alpha, \quad (1)$$

where I^* satisfies the first-order condition $1 = \alpha\beta(k(1 - \delta) + I^*)^{\alpha-1}$. By Theorem 1, the fixed cost formulation (Definition 1 in the main paper) is equivalent to the probability choice formulation (Definition 2), so the firm’s problem is exactly

$$V(k) = \max_{\varphi \in [0,1]} \{\varphi \cdot S(k) - C(\varphi)\} + \beta(k(1 - \delta))^\alpha, \quad (2)$$

with C given by the duality.

Proposition 1 (Conditions for smoothness of C).

$C \in C^2(0, 1)$ with $C'' > 0$ on the open unit interval — so that $\varphi^*(S)$ is a C^1 function of the surplus — if and only if:

- (i) G is absolutely continuous with density $g > 0$ on $(0, \bar{\xi})$.
- (ii) g is continuously differentiable ($g \in C^1$) on $(0, \bar{\xi})$.

Both conditions are automatically satisfied by any G with a C^1 , positive density. Smoothness fails when G has atoms (producing kinks in C'), when g has interior

zeros (producing points where C'' diverges), or when the support of G is disconnected (producing flat regions in C').

The degenerate case — G a point mass at $\bar{\xi}$ — violates condition (i) and produces the canonical lumpy investment model. C is linear, the only optimal choices are $\varphi \in \{0, 1\}$, and the firm either invests fully or not at all.

Proposition 2 (Investment response and the cost density).

In partial equilibrium, the sensitivity of the aggregate investment probability to a small productivity shock that raises the surplus by dS is

$$\frac{dE[\varphi^*(k)]}{dS} = E\left[\frac{g(G^{-1}(\varphi^*(k)))}{w}\right].$$

This is proportional to the expected density of cost draws at the current threshold, averaged across firms. Economies with higher g at the relevant threshold are more investment-elastic.

Proof.

By Theorem 1, the optimal investment probability is $\varphi^*(k) = G(S(k)/w)$. Differentiating with respect to the surplus:

$$\frac{d\varphi^*}{dS} = \frac{g(S/w)}{w} = \frac{g(G^{-1}(\varphi^*))}{w} = \frac{1}{C''(\varphi^*)},$$

where the last equality uses $C''(\varphi) = w/g(G^{-1}(\varphi))$ from Proposition 1(iii). Taking expectations across the cross-section of firms:

$$\frac{dE[\varphi^*(k)]}{dS} = E\left[\frac{g(G^{-1}(\varphi^*(k)))}{w}\right],$$

which is the expected density at the threshold, averaged over the distribution of capital stocks k . ■

Dispersion as a continuous index of lumpiness The curvature of the induced cost function provides a continuous metric of “how much lumpiness matters.” For the family $G = U[\mu - \epsilon, \mu + \epsilon]$, the curvature is $C'' = 2w\epsilon$, strictly increasing in the dispersion ϵ . As $\epsilon \rightarrow 0$, the model approaches the binary threshold rule. As ϵ grows, the model resembles a standard convex cost problem. There is no discrete boundary between “lumpy” and “smooth” economies.

Recovering the Khan-Thomas extensive margin under uniform G [Khan and Thomas \(2008\)](#) study a GE model in which firms face i.i.d. fixed costs drawn from a uniform distribution $G = U[0, \bar{\xi}]$. Their central finding is that lumpy micro-level investment has negligible aggregate consequences. The duality makes the mechanism transparent.

Under $G = U[0, \bar{\xi}]$, the duality (equation (2) in the main paper) yields

$$C(\varphi) = w \int_0^{\bar{\xi}\varphi} \xi \cdot \frac{1}{\bar{\xi}} d\xi = \frac{w\bar{\xi}}{2} \varphi^2.$$

The induced cost is exactly quadratic. Each firm’s investment problem therefore becomes

$$V(k) = \max_{\varphi \in [0,1]} \left\{ \varphi \cdot S(k) - \frac{w\bar{\xi}}{2} \varphi^2 \right\} + \beta(k(1 - \delta))^\alpha,$$

with interior first-order condition $\varphi^*(k) = S(k)/(w\bar{\xi})$. This is a smooth, concave problem — no kinks, no thresholds, no non-convexities. The firm’s investment probability responds linearly to the surplus. Aggregating across firms, the total investment rate is a smooth, linear function of the cross-sectional distribution of surpluses. There is nothing “lumpy” left at the aggregate level.

Uniform G is special only in producing *exactly* quadratic costs. More generally, any absolutely continuous G produces a strictly convex C (Proposition 1), which already smooths the extensive margin at the firm level.

The power law family $G(\xi) = (\xi/\bar{\xi})^\gamma$ generalises this. When $\gamma = 1$ (uniform), C is quadratic. When $\gamma > 1$ (concentrated G), C is more than quadratic — the cost of inducing high adjustment probabilities rises steeply, producing sharper investment spikes at the micro level. When $\gamma < 1$ (dispersed G), C is less than quadratic — adjustment probabilities respond more elastically to surplus changes. The parameter γ thus indexes the degree of lumpiness along a continuous spectrum, with the [Khan and Thomas \(2008\)](#) quadratic case as the midpoint.

C Proofs from the pricing section

C.1 Proof of Proposition 7 (Stationary distribution)

Proof.

The price gap evolves according to:

$$x_{t+1} = \begin{cases} z_{t+1} & \text{with probability } \Lambda(x_t), \\ x_t + z_{t+1} & \text{with probability } 1 - \Lambda(x_t), \end{cases}$$

where $z_{t+1} \sim \phi_z$ independently. The density of x_{t+1} conditional on the current cross-section f_t is:

$$f_{t+1}(x) = \int \Lambda(\tilde{x}) f_t(\tilde{x}) d\tilde{x} \cdot \phi_z(x) + \int (1 - \Lambda(\tilde{x})) \phi_z(x - \tilde{x}) f_t(\tilde{x}) d\tilde{x}.$$

The first term: adjusting firms (total mass $\bar{\Lambda}_t$) draw $z_{t+1} \sim \phi_z$ as their new gap. The second term: non-adjusting firms at \tilde{x} receive shock z_{t+1} and move to $\tilde{x} + z_{t+1}$; the density at x is the convolution of $(1 - \Lambda(\tilde{x})) f_t(\tilde{x})$ with ϕ_z .

Setting $f_{t+1} = f_t = f_{ss}$ yields the fixed-point equation of the main paper. Symmetry: if ϕ_z is symmetric around zero and $\Lambda(x) = \Lambda(-x)$ (which holds since $\Lambda(x) = G(Bx^2/w)$), then if f_t is symmetric, so is f_{t+1} . Since f_{ss} is the unique fixed point of this contraction, f_{ss} is symmetric around zero. \blacksquare

Remark 2. *The fixed-point equation corresponds, in discrete state space, to the eigenvector problem $\mathbf{f} = \mathbf{T}'\mathbf{f}$ where \mathbf{T} is the transition matrix encoding the adjustment and drift probabilities. This is the object computed numerically in the quantitative implementation.*

C.2 Proof of Proposition 8 (Selection and the menu cost distribution)

Proof.

Write the adjustment-weighted mean absolute gap as

$$\overline{|x|}^{adj} = \frac{E[\Lambda(x) \cdot |x|]}{E[\Lambda(x)]} = E[|x|] + \frac{\text{Cov}(\Lambda(x)/\bar{\Lambda}, |x|)}{1}.$$

The covariance is positive when Λ is increasing in $|x|$. Its magnitude depends on how sharply Λ responds to $|x|$.

(i) The elasticity of the hazard with respect to $|x|$ is

$$\frac{\partial \ln \Lambda(x)}{\partial \ln |x|} = \frac{g(Bx^2/w)}{G(Bx^2/w)} \cdot \frac{2Bx^2}{w},$$

which is twice the hazard rate of G evaluated at the firm's surplus. A higher hazard rate means Λ responds more sharply to $|x|$, increasing \mathcal{S} .

(ii) By Proposition 1(iii), $C'' = w/g(G^{-1}(\varphi))$. When g is small at the threshold, C'' is large and Λ responds steeply to changes in x^2 , concentrating adjustment among high-gap firms.

(iii) For constant Λ (Calvo), the covariance is zero and $\mathcal{S} = 0$. For a step function Λ (Golosov-Lucas), $\text{Cov}(\Lambda/\bar{\Lambda}, |x|)$ is maximal.

For the power law family, the hazard rate of G at ξ is $g(\xi)/G(\xi) = \gamma/\xi$, so the elasticity is 2γ — constant and strictly increasing in γ . ■

C.3 Proof of Proposition 9 (Calvo limit)

Proof.

As $\theta \rightarrow \infty$ with $\bar{\Lambda}_\theta = \lambda$ fixed, $Bx^2/(w\theta) \rightarrow 0$ for each x in the support of f_{ss} . Since G is differentiable at zero with $G(0) = 0$ and $g(0) > 0$, a first-order expansion gives $G(Bx^2/(w\theta)) \approx g(0) \cdot Bx^2/(w\theta)$. Note that $\Lambda_\theta(x) = G(Bx^2/(w\theta))$ and $\bar{\Lambda}_\theta = \int G(Bx^2/(w\theta))f_{ss}(x) dx$. For any compact set containing the support of f_{ss} , the argument $Bx^2/(w\theta)$ is uniformly small, so $G(Bx^2/(w\theta))/[g(0)Bx^2/(w\theta)] \rightarrow 1$ uniformly. The ratio $\Lambda_\theta(x)/\bar{\Lambda}_\theta \rightarrow Bx^2/E_{f_{ss}}[Bx^2]$, which is not constant. However, the *variation* in Λ_θ relative to its level vanishes: $\text{Var}_{f_{ss}}[\Lambda_\theta(x)]/\bar{\Lambda}_\theta^2 \rightarrow 0$, since both $\Lambda_\theta(x)$ and $\bar{\Lambda}_\theta$ are $O(1/\theta)$ while their difference is $O(1/\theta^2)$. For (i): since $\Lambda_\theta(x) = G(Bx^2/(w\theta)) \approx g(0)Bx^2/(w\theta)$ uniformly and $\bar{\Lambda}_\theta = \lambda$ by calibration, the ratio $\Lambda_\theta(x)/\lambda$ converges uniformly to 1 on the support of f_{ss} . For (ii): the variation in Λ_θ relative to its level vanishes, so the selection intensity $\mathcal{S}_\theta \rightarrow 0$, since the intensive margin is $O(\bar{\Lambda}_\theta/\theta)$, vanishing relative to the extensive margin $\bar{\Lambda}_\theta$ as $\theta \rightarrow \infty$. ■

C.4 Proof of Proposition 10 (Nonlinear Phillips curve)

Proof.

Assume that f_{ss} is continuously differentiable with $\int |x| f_{ss}(x) dx < \infty$ and that G is C^1 with bounded density g , so that differentiation under the integral sign is justified by the dominated convergence theorem. Write $f_t(x) = f_{ss}(x - \mu_t)$. Then

$$\pi_t = \int G\left(\frac{Bx^2}{w}\right) x \cdot f_{ss}(x - \mu_t) dx.$$

Differentiating with respect to μ_t using Leibniz's rule:

$$\frac{\partial \pi_t}{\partial \mu_t} = - \int G\left(\frac{Bx^2}{w}\right) x \cdot f'_{ss}(x - \mu_t) dx,$$

since $\frac{\partial}{\partial \mu_t} f_{ss}(x - \mu_t) = -f'_{ss}(x - \mu_t)$. Evaluating at $\mu_t = 0$:

$$\left. \frac{\partial \pi_t}{\partial \mu_t} \right|_{\mu_t=0} = - \int G\left(\frac{Bx^2}{w}\right) x \cdot f'_{ss}(x) dx.$$

Now integrate by parts with $u = G(Bx^2/w) \cdot x$ and $dv = f'_{ss}(x) dx$:

$$-\int u dv = -\left[G\left(\frac{Bx^2}{w}\right) x \cdot f_{ss}(x)\right]_{-\infty}^{\infty} + \int f_{ss}(x) \frac{d}{dx} \left[G\left(\frac{Bx^2}{w}\right) x\right] dx.$$

The boundary term vanishes because f_{ss} has tails decaying faster than any polynomial. Expanding the derivative:

$$\frac{d}{dx} \left[G\left(\frac{Bx^2}{w}\right) x\right] = g\left(\frac{Bx^2}{w}\right) \frac{2Bx}{w} \cdot x + G\left(\frac{Bx^2}{w}\right) = g\left(\frac{Bx^2}{w}\right) \frac{2Bx^2}{w} + G\left(\frac{Bx^2}{w}\right).$$

Substituting:

$$\kappa(G) = \left. \frac{\partial \pi_t}{\partial \mu_t} \right|_{\mu_t=0} = \int G\left(\frac{Bx^2}{w}\right) f_{ss}(x) dx + \int g\left(\frac{Bx^2}{w}\right) \frac{2Bx^2}{w} f_{ss}(x) dx.$$

The first integral is the extensive margin. The second integral is the intensive margin. Since $G(Bx^2/w)$ is even in x and f_{ss} is symmetric around zero, $\pi = 0$ at $\mu_t = 0$; the linearisation coefficient κ is well-defined and given by the formula above. ■

C.5 Proof of Proposition 12 (Monotonicity in γ)

Proof.

By Corollary 5 below, for the power law family $\kappa = (1 + 2\gamma)\bar{\Lambda}$ and the selection wedge equals $2\gamma\bar{\Lambda}$, both holding for any symmetric f_{ss} and any calibrated $\bar{\Lambda} > 0$.

(i) The selection intensity is $\mathcal{S} = (\kappa - \bar{\Lambda})/\bar{\Lambda} = 2\gamma$, strictly increasing in γ .

(ii) Monetary non-neutrality is proportional to $1 - \kappa = 1 - (1 + 2\gamma)\bar{\Lambda}$, strictly decreasing in γ for any $\bar{\Lambda} > 0$.

(iii) The Phillips curve slope is $\kappa = (1 + 2\gamma)\bar{\Lambda}$, strictly increasing in γ .

In each case, calibrating two power law economies to the same $\bar{\Lambda}$ (by adjusting $\bar{\xi}$) preserves γ as the sole determinant of κ , \mathcal{S} , and non-neutrality. ■

Remark 3. *Proposition 12 is stated within the power law family, which provides a clean one-parameter ordering at equal frequency. The underlying mechanism — concentrated G steepens the adjustment hazard, amplifying selection — is not family-specific. A general monotonicity result would require a stochastic-order condition on G that ensures higher $\kappa/\bar{\Lambda}$ whenever G is “more concentrated” in a suitable sense.*

C.6 Proof of Proposition 13 (Frequency does not determine non-neutrality)

Proof.

By the Phillips curve slope formula, $\bar{\Lambda}$ depends on $E[G(Bx^2/w)]$ while κ depends additionally on $E[g(Bx^2/w) \cdot 2Bx^2/w]$. Matching the first does not determine the second. For the power law $G(\xi) = (\xi/\bar{\xi})^\gamma$, direct computation gives $g(Bx^2/w) \cdot 2Bx^2/w = 2\gamma \Lambda(x)$, so $\kappa = (1 + 2\gamma)\bar{\Lambda}$. Two power laws with different γ , calibrated to the same $\bar{\Lambda}$ by adjusting $\bar{\xi}$, produce different κ . ■

C.7 Proof of Corollary 5 ($\kappa = (1 + 2\gamma)\bar{\Lambda}$ for power laws)

Proof.

For $G(\xi) = (\xi/\bar{\xi})^\gamma$, the density is $g(\xi) = (\gamma/\bar{\xi})(\xi/\bar{\xi})^{\gamma-1}$. Evaluate the intensive margin integrand at $\xi = Bx^2/w$:

$$\begin{aligned} g\left(\frac{Bx^2}{w}\right) \cdot \frac{2Bx^2}{w} &= \frac{\gamma}{\bar{\xi}} \left(\frac{Bx^2/w}{\bar{\xi}}\right)^{\gamma-1} \cdot \frac{2Bx^2}{w} \\ &= 2\gamma \cdot \frac{1}{\bar{\xi}} \left(\frac{Bx^2/w}{\bar{\xi}}\right)^{\gamma-1} \cdot \frac{Bx^2}{w} \\ &= 2\gamma \left(\frac{Bx^2}{w\bar{\xi}}\right)^\gamma = 2\gamma \Lambda(x). \end{aligned}$$

Substituting into the Phillips curve slope formula:

$$\begin{aligned} \kappa(G) &= \int G\left(\frac{Bx^2}{w}\right) f_{ss}(x) dx + \int 2\gamma \Lambda(x) f_{ss}(x) dx \\ &= \bar{\Lambda} + 2\gamma\bar{\Lambda} = (1 + 2\gamma)\bar{\Lambda}. \end{aligned}$$

The selection wedge is $\kappa - \bar{\Lambda} = 2\gamma\bar{\Lambda}$. The ratio $\kappa/\bar{\Lambda} = 1 + 2\gamma$ depends only on γ , not on f_{ss} , B , or w . ■

C.8 Proof of Corollary 4 (Ordering of non-neutrality)

Proof.

The real effect of a monetary shock dm is proportional to $(1 - \kappa) dm$, so non-neutrality is decreasing in κ .

(i) Calvo benchmark. When Λ is constant at λ , the intensive margin vanishes: $g(Bx^2/w) \cdot 2Bx^2/w = 0$ for all x in the Calvo limit, since a constant hazard implies

$g = 0$ almost everywhere. Therefore $\kappa = \bar{\Lambda} = \lambda$ and the selection wedge is zero. Non-neutrality is $1 - \lambda$, the maximum among models with the same frequency λ .

(ii) Absolutely continuous G with positive density. By Proposition 1(iii), $g > 0$ implies $C'' > 0$, so $\Lambda(x)$ is strictly increasing in $|x|$. The intensive margin $\int g(Bx^2/w)(2Bx^2/w)f_{ss}(x) dx > 0$. Therefore $\kappa > \bar{\Lambda}$ and the selection wedge is strictly positive.

(iii) Golosov-Lucas. When G is degenerate at $\bar{\xi}$, $\Lambda(x) = \mathbf{1}[Bx^2 \geq w\bar{\xi}]$. All adjusters have gaps at or beyond the threshold, so the intensive margin is maximised. Non-neutrality is minimal.

Monotonicity in γ : For the power law family, $\kappa = (1 + 2\gamma)\bar{\Lambda}$ by Corollary 5. At fixed $\bar{\Lambda}$, κ is strictly increasing in γ and non-neutrality $1 - \kappa$ is strictly decreasing. ■

D Supplementary figures

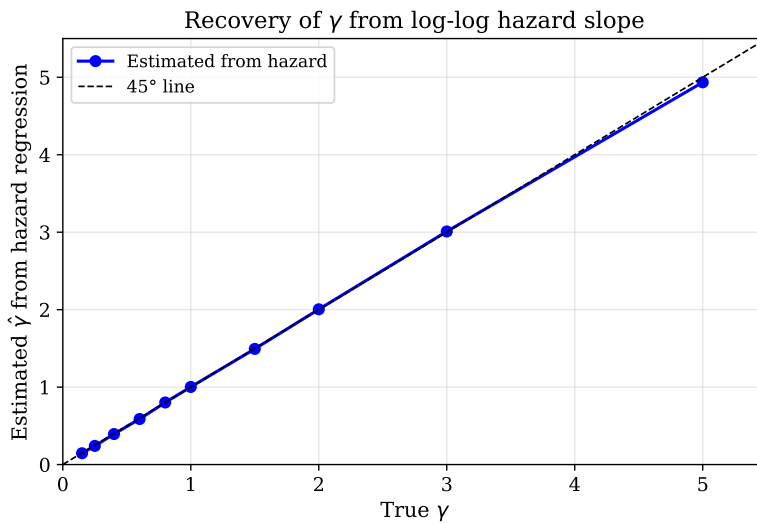


Figure 1: Recovery of γ from the log-log hazard regression. Each point is one simulated economy with 15,000 firms and 150 quarters. The regression recovers the true γ with mean absolute error 0.012. The identification requires only the conditional adjustment probability as a function of the price gap.

E Comparison with Williams-Daly-Zachary

Section 2.2 of the main paper discusses the relationship between Theorem 1 and the Williams-Daly-Zachary (WDZ) theorem (McFadden, 1978, 1981). This appendix pro-

Size distribution of price changes (het. σ_z , CV=1.0)

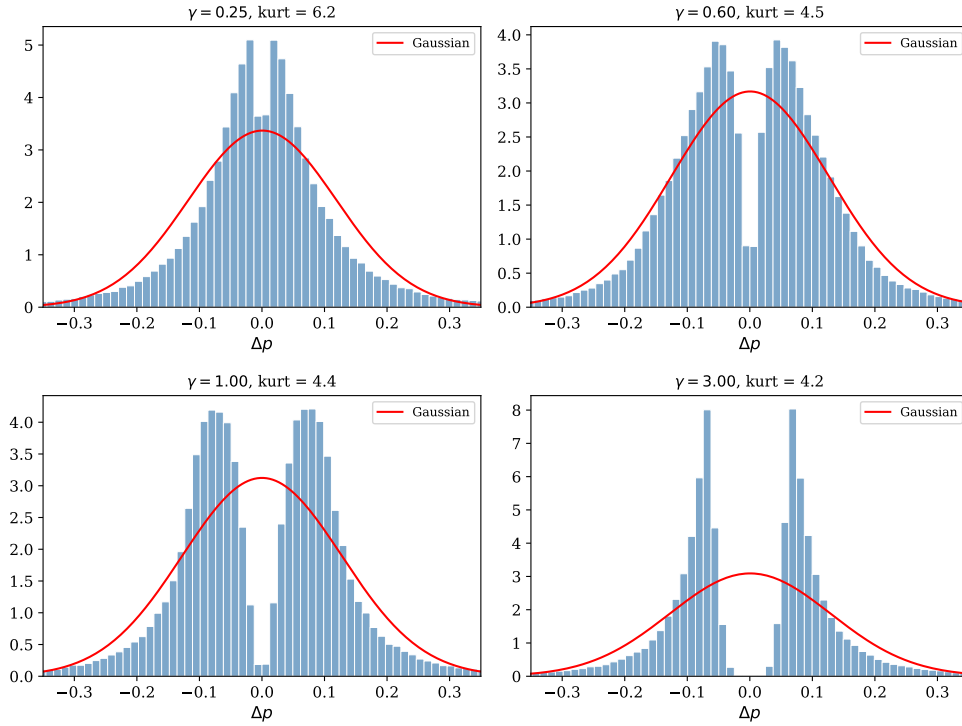


Figure 2: Distribution of price changes (heterogeneous σ_z , CV = 1). Red line: Gaussian with the same mean and variance. Low γ produces excess kurtosis through weak selection (small and large changes coexist). High γ concentrates changes through strong selection.

vides a systematic comparison. Both results share a common mathematical skeleton — an expected-maximum operator equals the Fenchel conjugate of a function determined by a distributional primitive — but differ in their economic content, scope, and implications. Table 1 summarises the comparison along six dimensions.

For binary choice ($J = 2$), the two results are formally equivalent under a sign change: a random benefit $S - w\xi$ versus a random cost $w\xi$ applied to a fixed benefit S are isomorphic. The binary conjugate identity is therefore not new. The contributions of this paper lie beyond the duality identity itself: the mapping $G \mapsto C$ and the distributional taxonomy (Propositions 3–5), the structural explanation for the [Khan and Thomas \(2008\)](#) result, and the continuous parametrisation of the pricing spectrum are entirely absent from WZD and the discrete choice literature.

Table 1: Williams-Daly-Zachary vs. Theorem 1

Dimension	WDZ	Theorem 1
Random element	Taste shocks ε_j on payoffs	Cost draws $w\xi$ on adjustment
Decision	Multi-alternative discrete choice	Binary: adjust or not
Surplus function	$E[\max_j\{u_j + \varepsilon_j\}]$	$\Pi^{FC}(S)$ (Definition 1)
Convexity result	Surplus convex in (u_1, \dots, u_J)	Payoff convex in S
Conjugate of	Generating function $\mathcal{G}(\varepsilon)$	$C(\varphi) = \int_0^\varphi G^{-1}(t) dt$
Choice probs	$p_j = \partial E[\max]/\partial u_j$	$\varphi^* = G(S)$
Primitive	Joint distribution of ε	Distribution G of costs
Coupling	Not constructed	Path-by-path equivalence
Economic context	Demand: consumer choice	Supply: adjustment frictions
Application	Multinomial logit (McFadden, 1974)	Calvo–GL spectrum

References

- Gabaix, Xavier. 2014. “A Sparsity-Based Model of Bounded Rationality.” *Quarterly Journal of Economics* 129 (4):1661–1710.
- Khan, Aubhik and Julia K. Thomas. 2008. “Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics.” *Econometrica* 76 (2):395–436.
- Matějka, Filip and Alisdair McKay. 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review* 105 (1):272–298.
- McFadden, Daniel. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Econometrics*, edited by Paul Zarembka. Academic Press, 105–142.
- . 1978. “Modelling the Choice of Residential Location.” In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist et al. North-Holland, 75–96.
- . 1981. “Econometric Models of Probabilistic Choice.” In *Structural Analysis of Discrete Data*, edited by Charles F. Manski and Daniel McFadden. MIT Press, 198–272.
- Sims, Christopher A. 2003. “Implications of Rational Inattention.” *Journal of Monetary Economics* 50 (3):665–690.