The Convex Origin of Fixed Costs[†]

Hanbaek Lee[‡]

University of Cambridge

March 2026

(click here for the latest version)

Abstract

Are models with fixed and convex adjustment costs different in nature? This paper shows they are isomorphic. When fixed costs are heterogeneous — drawn from a distribution G — the optimal threshold rule exactly coincides with the one for the convex probability cost: same payoffs, same decisions, same sample paths. Each cost distribution G implies different convexity for menu costs, lumpy investment, rational inattention, discrete choice, and control costs, unifying these classes of models as special cases of one duality. Applied to pricing, a power law G with shape parameter γ — the degree of cost concentration — defines the Calvo–Golosov-Lucas spectrum continuously: non-neutrality decomposes into frequency and a selection wedge governed by γ alone. Micro price data yield $\hat{\gamma} \approx 0.4$: the economy sits between Calvo and Golosov-Lucas on the spectrum, with selection accounting for 44% of price flexibility.

Keywords: Fixed costs, convex adjustment costs, Fenchel duality, menu costs, discrete choice, rational inattention, Phillips curve.

JEL codes: D81, E31, E22, E32.

[†]I am grateful for helpful comments and discussions from Philip Ushchev and Mindy Z. Xiaolan. All errors are my own.

[‡]Email: hl610@cam.ac.uk

1 Introduction

Are models with fixed and convex adjustment costs different in nature? This paper shows they are isomorphic. The key insight is that when fixed costs are heterogeneous — drawn from a distribution G — the optimal adjustment rule under G is identical to the one from a convex probability cost $C(\varphi)$, where C is the Fenchel conjugate of G. The two formulations produce the same payoffs, the same adjustment probabilities, and the same realised outcomes. The reason is economic: agents with low cost draws adjust first, so the marginal non-adjuster faces a higher cost draw than the last. This increasing marginal cost is convexity.

The paper makes three contributions. First, it establishes the duality and its structural implications. The deterministic fixed cost — where every firm faces the same cost — is a knife-edge singularity; any heterogeneity, however small, restores strict convexity (Proposition 2). At the micro level, each firm's decision is always binary. At the aggregate level, there is no type distinction between "lumpy" and "smooth" — only a degree distinction, with the shape of G as the dial. Uniform G gives textbook quadratic costs (Khan and Thomas, 2008); concentrated G gives the Ss rule of Golosov and Lucas (2007); dispersed G approximates Calvo (1983). These are not different kinds of models but different shapes of one distribution. This also provides a partial-equilibrium foundation for the irrelevance result of Khan and Thomas (2008): with uniform fixed costs, each firm's problem is already convex — a representational artifact, not an economic force to be averaged out.

Second, the duality provides a unified economic interpretation for different classes of models with frictional adjustment: they all share the same mapping $G \mapsto C$. Different distributions G generate cost functions that coincide with well-known information-theoretic entropies (Proposition 4). Exponential G yields Shannon entropy — the cost function in the binary rational inattention model of Matějka and McKay (2015). Uniform G yields the quadratic attention penalty for Khan and Thomas (2008) and Gabaix (2014). Power law G yields Tsallis q-entropy (Tsallis, 1988). In the multi-action extension, i.i.d. Gumbel-distributed costs recover the multinomial logit of McFadden (1974). Five literatures — menu costs, lumpy investment, discrete choice, rational inattention, and control costs (Mattsson and Weibull, 2002; Hofbauer and Sandholm, 2002) — are manifestations of one duality under different distributional assumptions on G. The economic content and welfare implications of each literature may differ. But the shared structure implies that $C(\varphi^*)$ gives a direct welfare measure of probability adjustment in resource units, that any heterogeneity in fixed costs makes the individual problem smooth, and that the smooth first-order condition $S = C'(\varphi)$ replaces kink-based dynamic programming.

Third, the paper applies the duality to price setting, providing a unified framework that encompasses Calvo and Golosov-Lucas as polar cases of a single continuous spectrum. A power law family $G(\xi) = (\xi/\bar{\xi})^\gamma$ parametrises this spectrum through its exponent γ, the degree of cost concentration. When γ is low, menu costs are dispersed: firms reprice frequently but nearly at random, regardless of how far their price is from the optimum. When γ is high, menu costs are concentrated: only the most mispriced firms reprice, and each repricing event moves the price level substantially. The aggregate price response to a nominal shock therefore decomposes into two forces (Proposition 10): *frequency* — how many firms reprice — and *selection* — how well repricing is targeted at the firms that need it most. The parameter γ governs the balance between the two (Proposition 11). Alvarez, Le Bihan, and Lippi (2016) show that frequency and kurtosis of price changes — kurtosis being the observable footprint of selection — are sufficient statistics for non-neutrality. Their result characterises non-neutrality in terms of observable moments; the duality enables the reverse direction, from G to moments, since G determines the repricing hazard and hence both sufficient statistics. Estimating γ from micro price data yields $\hat{\gamma} \approx 0.4$: the economy sits between Calvo and Golosov-Lucas on the spectrum, with selection accounting for 44% of total price flexibility.

Although the pricing application is developed in greatest depth, the duality applies to any extensive-margin decision with heterogeneous costs — firm entry, search and matching, labour force participation, technology adoption, and credit default — as discussed in Section 5.

Related literature The analysis connects several literatures. In macroeconomic frictions: Caplin and Spulber (1987) on monetary neutrality under uniform price gaps; Gourio and Kashyap (2007), who show that the shape of the fixed cost distribution determines whether lumpy investment matters for aggregates in the Thomas (2002) model; Oskolkov and Lippi (2026) on investment misallocation. On pricing: Dotsey, King, and Wolman (1999), Golosov and Lucas (2007), Midrigan (2011), and Nakamura and Steinsson (2010); Alvarez, Le Bihan, and Lippi (2016) on sufficient statistics for non-neutrality; Vavra (2014) on the cross-sectional variance of price changes. In the present framework, the Vavra (2014) sufficient statistic is an endogenous function of G, so G governs both the variance and the structural friction generating it.

The analysis contributes to the generalised Ss literature of Caballero and Engel (1993, 1999), who introduce an "adjustment function" — the probability of adjustment as a function of the agent's state — as the primitive governing aggregate dynamics. The repricing hazard $\Lambda(x)$ in Definition 3 is the adjustment function for this class of models; its shape is not an assumption but a derived object, determined entirely by G. The duality thus micro-founds the adjustment function approach. Caballero and Engel (2007) show that the variance of idiosyncratic shocks relative to the inaction band governs whether state dependence amplifies or dampens aggregate responses; Proposition 12 makes an analogous monotonicity statement in terms of γ.

In discrete choice: the Williams-Daly-Zachary theorem (McFadden, 1974, 1978, 1981) shows that the social surplus function is the convex conjugate of a generating function determined by the shock distribution. As noted, Theorem 1 shares this algebraic skeleton for binary choice. The contributions here lie in the economic applications: the G-indexed taxonomy, the Khan-Thomas foundation, and the pricing spectrum are entirely absent from the discrete choice literature. In control costs and sparsity (Mattsson and Weibull, 2002; Hofbauer and Sandholm, 2002; Gabaix, 2014): the fixed cost model *is* a control cost model, with the penalty endogenously determined by G; the quadratic attention penalty is the uniform-G special case. In rational inattention (Sims, 2003; Matějka and McKay, 2015): Shannon entropy costs are equivalent to exponentially-distributed fixed costs for binary choice.

Roadmap Section 2 presents the duality theorem, characterises the induced cost function, establishes universality, develops the entropy correspondence, and applies the result to investment. Section 3 develops the pricing spectrum. Section 4 estimates γ from micro price data and quantifies the frequency-selection decomposition. Section 5 discusses further applications, interpretation, and scope. Section 6 concludes.

2 The duality

2.1 Primitives

Consider an agent who faces a binary decision: take an action ("adjust") or maintain the status quo ("do nothing"). Adjusting yields a gross surplus $S \geq 0$ but requires paying a fixed cost $w\xi$, where $w > 0$ is a wage or unit cost and ξ is a random draw from a distribution G on $[0, \bar{\xi}]$ with density $g > 0$. Not adjusting yields zero. The randomness in ξ captures heterogeneity in the cost of adjustment. This heterogeneity may be across agents (a cross-section of firms facing different menu costs) or across time (a single firm drawing a new cost each period, as in Dotsey, King, and Wolman 1999).

This setup is minimal. It strips away the details of particular applications to isolate the structure common to all of them. In the investment application, S is the surplus from investing at the optimal scale. In the pricing application of Section 3, S is the gain from closing the price gap. Because the setup is general, the duality theorem applies to any extensive-margin decision where the agent's cost of acting is uncertain.

Definition 1 (Fixed cost formulation)**.**
Given the payoff S, the following defines the fixed cost payoff $\Pi^{FC}(S)$ *and the* threshold rule ξ^**: The agent observes the realised cost ξ and adjusts if and only if $w\xi \leq S$:*

$$\Pi^{FC}(S) = \int_0^{\bar{\xi}} \max\{S - w\xi,\ 0\}\, dG(\xi) = \int_0^{\min\{S/w, \bar{\xi}\}} (S - w\xi)\, dG(\xi). \tag{1}$$

The agent adjusts for all cost draws below the threshold $\xi^* = S/w$ and does not adjust otherwise.

Definition 2 (Probability choice formulation)**.**
Given the payoff S, the following defines the probability choice payoff $\Pi^{PC}(S)$ *and the* optimal adjustment probability φ^**: The agent chooses $\varphi \in [0, 1]$, paying a convex cost $C(\varphi)$:*

$$\Pi^{PC}(S) = \max_{\varphi \in [0,1]} \big\{\varphi \cdot S - C(\varphi)\big\}, \tag{2}$$

where $C : [0, 1] \to \mathbb{R}_+$ with $C(0) = 0$, C increasing and convex. The agent trades off the expected benefit of adjusting ($\varphi \cdot S$) against the cost of choosing a high adjustment *probability ($C(\varphi)$).*

The duality theorem establishes that these two formulations are not merely analogous but identical. For each distribution G with a positive density, there exists a unique convex cost function C such that the two models yield the same payoff and the same optimal adjustment probability.

2.2 Main result

Theorem 1 (Convex duality of fixed costs)**.**
The fixed cost formulation (Definition 1) and the probability choice formulation (Definition 2) are equivalent under the cost function

$$C(\varphi) = w \int_0^{G^{-1}(\varphi)} \xi \, dG(\xi). \tag{3}$$

Specifically:

(i) ***Payoff equivalence.*** $\Pi^{FC}(S) = \Pi^{PC}(S)$ *for all* $S \geq 0$. *The two formulations are Fenchel conjugate duals:* $\Pi^{PC}(S)$ *is the conjugate of* C*, and* $C(\varphi)$ *is the conjugate of* Π^{FC}.

(ii) ***Policy equivalence.*** *The optimal adjustment probability is the same:* $\varphi^* = G(S/w)$. *Moreover, if the probability choice model randomises by drawing* $\xi \sim G$ *and adjusting whenever* $\xi \leq G^{-1}(\varphi^*)$*, the two models produce identical binary outcomes path by path.*

Proof.
(i) Payoff equivalence. Substituting the threshold $\xi^* = S/w$ (assuming $S/w \leq \bar{\xi}$; the corner is analogous) into (1):

$$\Pi^{FC}(S) = S \cdot G(S/w) - w \int_0^{S/w} \xi \, dG(\xi). \tag{4}$$

The first term is the expected surplus from adjusting: the probability of adjustment $G(S/w)$ times the surplus S. The second term is the expected cost paid by adjusters. Setting $\varphi = G(S/w)$ so that $S/w = G^{-1}(\varphi)$:

$$\Pi^{FC}(S) = \varphi \cdot S - w \int_0^{G^{-1}(\varphi)} \xi \, dG(\xi) = \varphi S - C(\varphi).$$

This is the objective of (2) evaluated at $\varphi^* = G(S/w)$. The conjugate relationship follows from the definition of the Fenchel conjugate: $\sup_\varphi\{\varphi S - C(\varphi)\} = \Pi^{PC}(S) = \Pi^{FC}(S)$.

(ii) Policy equivalence. The first-order condition of (2) is $S = C'(\varphi)$. Since $C'(\varphi) = wG^{-1}(\varphi)$ (Proposition 1 below), this gives $G^{-1}(\varphi^*) = S/w$, i.e., $\varphi^* = G(S/w)$ — the same threshold probability as in Definition 1. For the path-by-path equivalence: in the fixed cost model, the agent draws $\xi \sim G$ and adjusts whenever $\xi \leq S/w$, which occurs with probability φ^*. If the probability choice model uses the same draw and adjusts whenever $\xi \leq G^{-1}(\varphi^*)$, the two rules coincide and the realised adjustment indicators are identical. ■

The economic content is straightforward. In the fixed cost model, agents with low cost draws adjust and those with high draws do not. The expected payoff equals the adjustment probability times the surplus, minus the average cost paid by adjusters. In the probability choice model, the agent chooses an adjustment probability and pays a cost $C(\varphi)$. Proposition 1 below confirms that C is convex and characterises its shape. The theorem says these two descriptions yield exactly the same payoff and adjustment probability, provided C is constructed from G via (3).

The Fenchel conjugate structure places the two formulations on equal mathematical footing. The payoff function $\Pi^{PC}(S)$ and the cost function $C(\varphi)$ are conjugates of each other: the Fenchel conjugate of C yields Π^{PC}, and the conjugate of Π^{FC} yields C. The two express the same value function in dual variables — one in the surplus S, the other in the adjustment probability φ. Any result proved for one variable translates immediately to the other. The two are not "different theories" but different parametrisations of a single theory. When G is uniform, C is quadratic and the equivalence can be verified by direct calculation (see the investment application below).

Relation to discrete choice theory As noted in the introduction, the Williams-Daly-Zachary (WDZ) theorem (McFadden, 1978, 1981) establishes that the expected maximum utility under random taste shocks is the Fenchel conjugate of a generating function determined by the shock distribution. For binary choice, the WDZ identity and Theorem 1 share the same algebraic skeleton (Table G.1 in Online Appendix H). The binary conjugate identity is therefore not new; the contributions here lie in what it enables — the distribution-theoretic taxonomy, the structural foundation for Khan-Thomas, and the continuous pricing spectrum — all of which are absent from the discrete choice literature. The duality is the tool; the taxonomy and the pricing application are the results.

Two further distinctions are worth noting. First, the source of randomness differs: in WDZ, randomness enters payoffs via taste shocks ε_j; here, randomness enters costs via the draw $\xi \sim G$. For binary choice the two are isomorphic under a sign change. Second, the coupling is new. WDZ establishes equivalence of expected payoffs and choice probabilities; Theorem 1 further constructs a natural randomisation device under which the realised adjustment decisions coincide path by path — not merely in distribution.

Relation to control costs In the game theory literature, Mattsson and Weibull (2002) and Hofbauer and Sandholm (2002) study agents who choose mixed strategies subject to convex "control costs" that penalise departures from a default action. Deviating from habitual behaviour is costly, and the cost increases as the deviation becomes more extreme. The probability choice formulation (2) has exactly this structure: $C(\varphi)$ is the cost of choosing adjustment probability φ relative to the default of not adjusting ($\varphi = 0$). Theorem 1 therefore shows that the fixed cost model *is* a control cost model. The control cost is not an ad hoc penalty imposed by the modeller. It is the endogenous cost function determined by the distribution of frictions that the agent faces.

2.3 Properties of the induced cost function

The induced cost function $C(\varphi)$ inherits its shape entirely from the distribution G. The following proposition characterises its key properties and provides the economic interpretation that underlies all subsequent applications.

Proposition 1 (Cost function properties)**.**
The induced cost function (3) *satisfies:*

(i) $C(0) = 0$, $C(1) = w\, E_G[\xi]$.

(ii) $C'(\varphi) = w\, G^{-1}(\varphi)$*: marginal cost equals the cost of the marginal adjuster.*

(iii) $C''(\varphi) = w/g(G^{-1}(\varphi)) > 0$*: strictly convex whenever* $g > 0$.

(iv) C is C^2 if g is C^1 and positive.

Proof. See Online Appendix A. ∎

Part (ii) carries the core economic content. Agents are ranked by their cost draws: those with the lowest ξ adjust first. The marginal cost $C'(\varphi) = wG^{-1}(\varphi)$ is the cost of the agent just indifferent at the current threshold — the φ-th quantile of the cost distribution. Each increment in φ requires inducing a more expensive agent to adjust, which is the force that makes C convex. This is the economic origin of convexity: it is not assumed but arises endogenously from the ranking of heterogeneous agents by their costs.

Part (iii) reveals how the density g governs the strength of this convexity. Where g is large, many agents are bunched near the threshold. A small increase in the surplus tips many of them into adjustment, so the marginal cost rises gently. Where g is small, few agents are near the threshold. The same increase tips few additional agents, and the marginal cost rises sharply. In short: crowded margins mean gentle curvature; sparse margins mean sharp curvature.

Corollary 1 (Shape of G determines curvature)**.**

(i) g increasing $\Rightarrow$ C'' decreasing: there are relatively more agents at higher cost quantiles, so inducing additional adjustment becomes progressively easier. Adjustment is "cheap at the margin" for high φ.

(ii) g decreasing $\Rightarrow$ C'' increasing: agents thin out at higher cost quantiles, so inducing additional adjustment becomes progressively more expensive. Adjustment becomes disproportionately costly at high φ.

(iii) g constant (uniform) $\Rightarrow$ C'' constant: agents are equally spaced throughout the cost distribution, yielding the standard quadratic cost function with constant curvature.

Proof. See Online Appendix A. ∎

The following corollary applies the mapping (3) to four canonical distributions, recovering well-known cost function specifications as special cases of the single formula $C(\varphi) = w\int_0^{\varphi} G^{-1}(u)\,du$. Each distribution represents a different assumption about the cross-sectional heterogeneity of adjustment costs, and the induced cost function inherits its functional form from the shape of G.

Corollary 2 (Canonical distributions)**.**

(i) ***Uniform*** $G = U[0, \bar{\xi}]$*:* $C(\varphi) = \frac{w\bar{\xi}}{2}\varphi^2$*. The constant density produces constant curvature, so each unit increase in the adjustment probability costs the same marginal increment. This is the standard quadratic adjustment cost used throughout the macroeconomics and industrial organisation literatures.*

(ii) ***Power law*** $G(\xi) = (\xi/\bar{\xi})^\gamma$, $\gamma > 0$*:* $C(\varphi) = \frac{\gamma}{\gamma+1} w\bar{\xi}\,\varphi^{1+1/\gamma}$*. The cost function is iso-elastic with elasticity* $1 + 1/\gamma$*, and the shape parameter* γ *governs the curvature profile. For* $\gamma > 1$*, the density* $g(\xi) = \gamma\xi^{\gamma-1}/\bar{\xi}^\gamma$ *is increasing: most cost draws are high, so cheap adjusters are scarce and* C *is less curved than quadratic. For* $\gamma < 1$*, the density is decreasing: most cost draws are low, cheap adjusters are abundant, and* C *is more curved than quadratic. The uniform case is* $\gamma = 1$*. As* $\gamma \to \infty$*, the density concentrates near* $\bar{\xi}$ *and* C *approaches linearity (the degenerate limit). As* $\gamma \to 0$*, the density concentrates near zero and* C *approaches a step function at* $\varphi = 0$ *(nearly free adjustment). The power law family thus spans the full range from lumpiness to frictionless adjustment, indexed by a single parameter. This family plays the central role in the pricing application of Section 3.*

(iii) ***Exponential*** $G(\xi) = 1 - e^{-\lambda\xi}$ *on* $[0, \infty)$*:* $C'(\varphi) = -\frac{w}{\lambda}\ln(1-\varphi)$*, giving* $C(\varphi) = \frac{w}{\lambda}[(1-\varphi)\ln(1-\varphi) + \varphi]$*. The logarithmic marginal cost diverges as* $\varphi \to 1$*, severely penalising near-certain adjustment. This reflects the exponential distribution's unbounded support: to guarantee adjustment with probability close to one, the agent must be willing to pay arbitrarily high cost draws, which is infinitely expensive in expectation. As shown in Section 2.5, this cost function is proportional to the binary Shannon entropy, connecting the fixed cost model to the rational inattention literature (Sims, 2003; Matějka and McKay, 2015).*

(iv) ***Degenerate*** *at* $\bar{\xi}$*:* $C(\varphi) = w\bar{\xi}\varphi$*. The cost function is linear, with no strict convexity. The only optimal choices are* $\varphi = 0$ *(never adjust) or* $\varphi = 1$ *(always adjust), depending on whether* S *falls below or above* $w\bar{\xi}$*. This is the canonical*

binary threshold rule of the deterministic fixed cost model — the singular case analysed in Proposition 2 below.

Proof. See Online Appendix A. ■

These are not four different models — they are four special cases of one model, distinguished solely by the shape of G.

2.4 Universality of the convex representation

The degenerate case — a single, known fixed cost — is the only setting in which the problem is not strictly convex. This section shows that this case is a knife-edge singularity. Any perturbation, no matter how small, restores strict convexity. The convex model is the generic case; lumpiness is the singular exception.

Proposition 2 (Universality)**.**
Let $\bar{\xi} > 0$ be a deterministic fixed cost. For any $\epsilon > 0$, perturb the cost to $\xi \sim G_\epsilon \equiv U[\bar{\xi} - \epsilon,\ \bar{\xi} + \epsilon]$. The induced cost function is

$$C_\epsilon(\varphi) = w\big(\bar{\xi} - \epsilon\big)\varphi + w\epsilon\,\varphi^2, \tag{5}$$

which is strictly convex with $C''_\epsilon(\varphi) = 2w\epsilon > 0$ for every $\epsilon > 0$. As $\epsilon \to 0$:

(i) $C_\epsilon(\varphi) \to w\bar{\xi}\,\varphi$ pointwise: the linear (degenerate) cost.

(ii) $C''_\epsilon(\varphi) \to 0$: curvature vanishes.

(iii) The optimal $\varphi^*_\epsilon(S) = \min\left\{\frac{S - w(\bar{\xi}-\epsilon)}{2w\epsilon},\ 1\right\}$ converges to $\mathbf{1}[S \geq w\bar{\xi}]$: the binary threshold rule.

Thus the deterministic fixed cost is the singular limit of a family of strictly convex problems.

Proof. For $G_\epsilon = U[\bar{\xi} - \epsilon,\ \bar{\xi} + \epsilon]$, applying (3) gives $C_\epsilon(\varphi) = w(\bar{\xi} - \epsilon)\varphi + w\epsilon\varphi^2$, with $C''_\epsilon = 2w\epsilon > 0$. The full derivation is in Online Appendix A. ■

The result overturns the standard view that convex costs are a "smooth approximation" to the "true" lumpy friction. The lumpy model is the singular limit; the convex model is the generic case. The curvature $C''_\epsilon = 2w\epsilon$ measures the "softness" of the adjustment margin: even an infinitesimally small ϵ produces a strictly convex problem.

Epistemic interpretation and structural stability The deterministic fixed cost requires the agent to know the adjustment cost with infinite precision. In practice, this never holds. A firm repricing its goods faces uncertainty about staff time, managerial attention, and logistical disruption. A plant manager cannot know the exact installation cost until the project is underway. If there is *any* uncertainty — any ϵ-band of noise — the decision problem is already smooth and convex.

This is a statement about structural stability. The deterministic fixed cost is structurally unstable: any ϵ-perturbation changes the decision rule from a discontinuous step function to a smooth, increasing function of the surplus. The convex formulation is structurally stable: small changes in G produce small changes in C and in the optimal policy. Topological genericity is established in Online Appendix A: the set of distributions generating strictly convex C is open and dense in the space of distributions on $[0, \bar{\xi}]$.

Mean-preserving spreads and convexity Proposition 2 shows that any perturbation of a point mass restores strict convexity. The following result establishes the general principle: among distributions with the same mean, greater dispersion produces a lower — and more curved — cost function. The degree of convexity is monotone in the mean-preserving spread (MPS) order, the most natural partial order on distributions with a fixed mean.

Proposition 3 (Dispersion and convexity)**.**
Let G_1 and G_2 be distributions on $[0, \bar{\xi}]$ with the same mean, $E_{G_1}[\xi] = E_{G_2}[\xi]$, and let G_2 be a mean-preserving spread of G_1. Let C_1 and C_2 be the induced cost functions. Then:

(i) $C_1(1) = C_2(1) = w\,E[\xi]$*: the total cost of certain adjustment is identical.*

(ii) $C_2(\varphi) \leq C_1(\varphi)$ *for all* $\varphi \in [0, 1]$*: greater dispersion lowers the cost function in the interior.*

(iii) The marginal costs single-cross: there exists $\varphi_0 \in (0, 1)$ *such that* $C_2'(\varphi) \leq C_1'(\varphi)$ *for* $\varphi < \varphi_0$ *and* $C_2'(\varphi) \geq C_1'(\varphi)$ *for* $\varphi > \varphi_0$*.*

Consequently, greater dispersion of G lowers the cost of moderate adjustment, raises the cost of near-certain adjustment, and steepens the marginal cost schedule. The *degenerate distribution (a point mass) maximises C in the interior and minimises curvature (C is linear); any spread away from it lowers C and increases curvature.*

Proof. See Online Appendix A. ∎

The intuition is that greater dispersion makes the cheapest adjusters cheaper and the most expensive holdouts more expensive, steepening the marginal cost schedule while leaving the total cost $C(1) = wE[\xi]$ unchanged. Proposition 2 is the special case where $G_1 = \delta_{\bar{\xi}}$ and G_2 is any non-degenerate distribution with the same mean. In the pricing application, this linkage has direct aggregate consequences: greater dispersion weakens selection, flattens the Phillips curve, and increases monetary non-neutrality (Section 3.7).

2.5 Fixed costs as generalised entropies

The cost function $C(\varphi)$ measures how expensive it is to deviate from inaction. It is fully determined by G and inherits its shape from the quantile function G^{-1}. This section shows that canonical distributions G yield cost functions that coincide with well-known information-theoretic entropies, revealing that the fixed cost literature in macroeconomics, the discrete choice literature in industrial organisation, the rational inattention literature in information economics, and the sparse decision-making literature are all manifestations of the same duality viewed through different distributions G.

Proposition 4 (Entropy correspondence)**.**

(i) ***Exponential distribution****: $G(\xi) = 1 - e^{-\lambda\xi}$ on $[0, \infty)$. Then*

$$C'(\varphi) = -\frac{w}{\lambda}\ln(1-\varphi), \qquad C(\varphi) = \frac{w}{\lambda}\big[(1-\varphi)\ln(1-\varphi) + \varphi\big].$$

The adjustment probability under exponential costs is $\varphi^* = 1 - e^{-\lambda S/w}$, the CDF of the exponential distribution evaluated at S/w. Note that this is not *the logistic function $1/(1+e^{-S/\sigma})$ that arises in logit discrete choice; the two share a sigmoid shape but differ in functional form. The true logit correspondence arises in the multi-action extension (Section 2.7), where i.i.d. Gumbel-distributed costs across alternatives yield the multinomial logit.*

(ii) ***Power law distribution***: $G(\xi) = (\xi/\bar{\xi})^\gamma$ *on* $[0, \bar{\xi}]$. *Then*

$$C(\varphi) = \frac{\gamma}{\gamma+1} w\bar{\xi}\, \varphi^{1+1/\gamma}.$$

This is proportional to φ^q *with* $q = 1 + 1/\gamma$, *which coincides with the functional form of the Tsallis q-entropy (Tsallis, 1988), a one-parameter generalisation of Shannon entropy used in statistical mechanics. The parameter* γ *controls the index* q: $\gamma = 1$ *gives* $q = 2$ *(quadratic); as* $\gamma \to \infty$, $q \to 1$ *and the cost function approaches the Shannon form.*

(iii) ***Uniform distribution***: $G = U[0, \bar{\xi}]$. *Then* $C(\varphi) = \frac{w\bar{\xi}}{2}\varphi^2$, *a quadratic penalty. This is the simplest and most widely used convex cost specification, and it coincides with the* L^2 *penalty that appears in many regularisation frameworks.*

Proof. See Online Appendix A. ■

The correspondences have precise scope. Exponential G yields Shannon entropy — the cost function in binary rational inattention (Sims, 2003; Matějka and McKay, 2015); the equivalence is at the level of the binary cost function, not the full continuous-state RI framework. Uniform G yields the quadratic attention penalty of Gabaix (2014); again, the equivalence is at the binary decision level and does not capture the full multi-dimensional sparsity structure. In the multi-action extension (Section 2.7), i.i.d. Gumbel costs yield negative Shannon entropy and the multinomial logit. These are correspondences of functional form: the reduced-form $C(\varphi)$ is the same, but the economic mechanisms and welfare implications differ. The duality identifies the mathematical structure these models share; it does not assert that the underlying economics are the same. The detailed discussion is in Online Appendix A, and the welfare content of these correspondences is developed in Section 2.6.

The distribution G as a Rosetta Stone Table 1 summarises the mapping. Five literatures — menu costs, lumpy investment, discrete choice, rational inattention, and control costs — are unified by a single object: the distribution G. Each literature has independently developed models that are special cases of the duality, distinguished only by the distributional assumption on costs.

Table 1: The distribution G as a Rosetta Stone

Distribution G	**Cost** $C(\varphi)$	**Form**	**Entropy**	**Literature**
Exponential	$(1-\varphi)\ln(1-\varphi)+\varphi$	Log. MC	Shannon	Rational inattention
Power law (γ)	$\varphi^{1+1/\gamma}$	Iso-elastic	Tsallis q	Generalised discrete choice
Uniform ($\gamma=1$)	φ^2	Quadratic	L^2	Sparsity
Degenerate	$\bar{\xi}\,\varphi$ on $\{0,1\}$	Linear	—	Lumpy adjustment
Type I EV (multinomial)	$\sum \varphi_j \ln \varphi_j$	Neg. entropy	Shannon (mv)	Logit discrete choice

Notes: Each row specifies a distribution G of fixed costs. The second column gives the induced convex cost function $C(\varphi)$ via equation (3) (up to scale). The fourth column identifies the information-theoretic entropy to which C corresponds. The last column indicates the literature where this cost function appears. "Log. MC" = logarithmic marginal cost. "mv" = multivariate. The first four rows are binary choice; the last row is the multi-action extension (Proposition 5).

A robustness criterion The Shannon entropy (exponential G) is the unique entropy satisfying certain axioms (additivity, continuity, maximum entropy principle). If one believes these axioms should govern the cost of deviating from default behaviour, then exponential G is the "correct" fixed cost distribution, and the logit/rational inattention model is the uniquely justified reduced form. Departures from exponential G — such as the uniform distribution used in Khan and Thomas (2008) — correspond to departures from Shannon's axioms, and can be interpreted as reflecting specific features of the adjustment technology rather than information-processing constraints.

2.6 Welfare content of the duality

The welfare cost of adjustment at a given equilibrium is $C(\varphi^*)$. This identity has two uses: it gives macro researchers a smooth, closed-form welfare measure for fixed cost models, and it gives rational inattention and discrete choice researchers a physical-cost interpretation for their abstract penalty functions.

From fixed to convex: smooth computation and welfare analysis The probability choice formulation (Definition 2) replaces $\max\{S - w\xi,\ 0\}$ with $\max_\varphi\{\varphi S - C(\varphi)\}$, a concave objective with a smooth first-order condition $S = C'(\varphi)$. The value function is C^2 when g is C^1 (Proposition 1), and standard perturbation methods apply without special treatment of the threshold. For heterogeneous-agent models with adjustment frictions (Reiter, 2009; Auclert, Rognlie, and Straub, 2020), the convex formulation is directly pluggable into existing linearisation-based toolkits. The envelope theorem gives

$$\frac{\partial V}{\partial S} = \varphi^*(S), \qquad C'(\varphi^*) = wG^{-1}(\varphi^*).$$

The marginal cost of the friction is the cost of the marginal adjuster. If a policy reduces adjustment costs, the welfare gain is $\Delta W \approx \varphi^* \cdot \Delta S - \Delta C(\varphi^*)$, which can be evaluated in closed form for the power law family.

Beyond smoothness, the convex representation delivers closed-form solutions. The first-order condition $S = C'(\varphi) = wG^{-1}(\varphi)$ inverts to give the policy function $\varphi^* = G(S/w)$ directly — no value function iteration required. For the power law family, all aggregate objects follow in closed form: the repricing hazard $\Lambda(x) = (Bx^2/(w\bar{\xi}))^\gamma$, the Phillips curve slope $\kappa = (1+2\gamma)\bar{\Lambda}$, the selection wedge $2\gamma\bar{\Lambda}$, and the welfare cost $C(\varphi^*)$. The standard menu cost computation — solving a dynamic program with kinks, finding Ss bands numerically, then linearising — reduces to specifying G and evaluating known functions.

Most importantly, $C(\varphi^*)$ provides a closed-form summary of the total welfare cost of the friction at a given equilibrium:

$$\text{Welfare cost} = w\int_0^{G^{-1}(\varphi^*)} \xi \, dG(\xi) = C(\varphi^*). \tag{6}$$

Given the equilibrium adjustment probability, one does not need to know the full distribution of realised costs, the identity of each adjuster, or the cross-sectional distribution of gaps. However, φ^* itself depends on G through the entire equilibrium — the gap distribution, the hazard, the steady-state cross-section — so $C(\varphi^*)$ is a compact representation of welfare at a given equilibrium, not a shortcut to computing it without solving the model. For welfare comparisons across economies, one compares $C(\varphi^*_{\text{US}})$ and $C(\varphi^*_{\text{EA}})$, but each φ^* must be solved for in equilibrium.

From convex to fixed: welfare interpretation of information costs In the rational inattention literature, the Shannon entropy cost $\lambda \sum \varphi_j \ln \varphi_j$ is measured in nats or bits. The parameter λ is the shadow price of information capacity, calibrated to match choice probabilities, with no independent measurement strategy and no direct welfare interpretation in terms of resources consumed. The duality resolves this: under exponential G with rate λ, the entropy cost $C(\varphi^*)$ equals the expected physical cost — real dollars, real labour time, real managerial attention — paid by agents who adjust.

Every abstract penalty function in the RI and control cost literatures therefore acquires a concrete welfare interpretation in resource units. The parameter λ becomes the rate parameter of the exponential cost distribution, estimable from micro data on actual adjustment costs — converting a free parameter into a testable restriction. Welfare comparisons across cost specifications also become meaningful: two economies calibrated to the same φ^* but with different G's have different welfare costs via equation (6).

2.7 Beyond binary choice

The duality extends naturally beyond the binary adjustment decision. This section outlines extensions to multiple actions, state-dependent surpluses, and dynamic environments. The details are straightforward applications of the conjugate structure.

Multiple actions Suppose the agent chooses among $J+1$ actions $\{a_0, a_1, \ldots, a_J\}$, where a_0 is inaction. Each a_j for $j \geq 1$ yields surplus S_j at random cost $w\xi_j$ with $\xi_j \sim G_j$, independently across actions. The agent observes all J cost draws and selects the action with the highest net payoff. The duality extends to this multi-action setting:

Proposition 5 (Multi-action duality).
Let $J \geq 1$ and suppose $\xi_j \sim G_j$ independently for $j = 1, \ldots, J$, with each G_j absolutely continuous with positive density on $[0, \bar{\xi}_j]$. Define $S_0 = 0$, $\xi_0 = 0$. Then:

(i) The expected payoff $\Pi^{FC}(\mathbf{S}) = E\big[\max_{j\in\{0,\ldots,J\}}\{S_j - w\xi_j\}\big]$ equals

$$\Pi^{PC}(\mathbf{S}) = \max_{\boldsymbol{\varphi}\in\Delta^J} \sum_{j=0}^{J} \varphi_j S_j - C(\boldsymbol{\varphi}), \tag{7}$$

where $C : \Delta^J \to \mathbb{R}_+$ is jointly convex and determined by the vector of distributions $(G_1, \ldots, G_J)$.

(ii) When $\xi_j \overset{i.i.d.}{\sim}$ Gumbel(μ, σ) for all $j \geq 1$, the cost function is $C(\boldsymbol{\varphi}) = \sigma \sum_{j=0}^{J} \varphi_j \ln \varphi_j$

(negative Shannon entropy) and the optimal choice probabilities are

$$\varphi_j^* = \frac{e^{S_j/\sigma}}{\sum_{k=0}^{J} e^{S_k/\sigma}}, \tag{8}$$

the multinomial logit of McFadden (1974).

Proof.
Apply the Fenchel conjugate construction to the multi-alternative maximum and specialise to the Gumbel distribution. The detailed proof is given in Online Appendix A. ■

When the surplus depends on a state variable x (price gap, capital stock), the duality applies state by state: $\varphi^*(x) = G(S(x)/w)$ is the endogenous state-dependent hazard. In dynamic settings, the probability choice formulation (Definition 2) makes the Bellman equation smooth — standard existence, uniqueness, and concavity results apply — whereas the fixed cost formulation (Definition 1) has kinks at adjustment thresholds.

2.8 Application to investment

The duality has a direct implication for the lumpy investment literature. In the Thomas (2002) model, Gourio and Kashyap (2007) show that the shape of the fixed cost distribution is the key determinant of whether lumpy micro-level investment spikes matter for aggregates. The duality makes this precise: each G maps to a unique convex cost C, and the shape of G governs the curvature of C and hence the aggregate investment elasticity. Under uniform G, C is quadratic (Corollary 2), so each firm's investment problem is individually convex — providing a partial-equilibrium foundation for the irrelevance result of Khan and Thomas (2008). The GE mechanism (endogenous price adjustment absorbing investment spikes) is an additional sufficient condition; the PE convexity established here is a necessary one. More broadly, any G with a positive density delivers smoothness at the individual level. The dispersion of G provides a continuous index of lumpiness: concentrated G (high γ) produces sharp investment spikes; dispersed G (low γ) produces smooth adjustment. As dispersion shrinks to a point mass, the model recovers the binary threshold rule. Oskolkov and Lippi (2026) recover G from plant-level investment data, confirming that the shape of the cost distribution varies across sectors. Formal conditions for smoothness and the investment elasticity formula are in Online Appendix B.

3 The menu cost–Calvo spectrum

The deepest application of the duality is to price setting, where the distinction between time-dependent (Calvo) and state-dependent (menu cost) pricing is central to New Keynesian economics. The idea that these models sit on a spectrum is not new: Dotsey, King, and Wolman (1999) draw menu costs from a distribution, and Caballero and Engel (1993) parametrise the adjustment hazard directly. Without the duality, however, exploring this spectrum requires solving a dynamic program numerically for each candidate G, and the hazard shape must be assumed rather than derived. The duality changes this. It delivers the repricing hazard $\Lambda(x)$ in closed form, derives the hazard from G rather than assuming it, and yields an analytical frequency-selection decomposition indexed by a single parameter. The distribution G governs the repricing hazard, the selection effect, the Phillips curve slope, and the degree of monetary non-neutrality.

3.1 Setup

Consider a monopolistically competitive firm with current log price p and desired log price p^*. The *price gap* is $x \equiv p^* - p$. When the firm reprices (closing the gap to zero), it gains a surplus that depends on the distance from the optimum. Based on the standard second-order approximation of the profit function $S(x)$ around $p = p^*$, I assume

$$S(x) = Bx^2, \tag{9}$$

where $B > 0$ depends on the demand elasticity and the discount factor.[1] The quadratic form reflects the approximate symmetry of profits around the optimum. The loss from mispricing grows with the square of the price gap: a firm with a 2% gap loses roughly four times as much as one with a 1% gap.

The firm draws a menu cost $\xi \sim G$ each period and reprices if and only if the gain

[1] This is standard; see Golosov and Lucas (2007) and Alvarez, Le Bihan, and Lippi (2016). The quadratic loss approximation is $S(x) \approx \frac{\varepsilon}{2}x^2$ where ε is the demand elasticity, up to discounting.

from repricing exceeds the cost $w\xi$, where $w > 0$ is the wage (as in Section 2):

$$\text{Reprice if and only if } w\xi \leq Bx^2. \tag{10}$$

3.2 The repricing hazard

By Theorem 1, the firm's repricing decision under the fixed cost formulation (Definition 1) is equivalent to the probability choice formulation (Definition 2) with convex cost $C(\varphi)$. The first-order condition $Bx^2 = C'(\varphi) = wG^{-1}(\varphi)$ yields the optimal repricing probability as a function of the price gap:

Definition 3 (Repricing hazard)**.**
The repricing hazard — the probability that a firm with price gap x adjusts its price — is

$$\Lambda(x) \equiv G\left(\frac{Bx^2}{w}\right). \tag{11}$$

This is the central object of the pricing application. It is not a numerical output of a dynamic program; it is an explicit formula, given directly by the duality and the choice of G.[2] The repricing hazard $\Lambda(x)$ summarises everything about the firm's pricing behaviour that matters for aggregate dynamics. Conditional on (B, w) and the steady-state gap distribution f_{ss}, all aggregate pricing implications — inflation, the Phillips curve slope, the selection effect, monetary non-neutrality — flow through Λ. The distribution G is therefore the structural primitive from which aggregate pricing dynamics are derived.

Proposition 6 (Hazard properties)**.**
The repricing hazard $\Lambda(x)$ satisfies:

(i) $\Lambda(0) = G(0) = 0$*: a firm at its optimal price never reprices. The gain from doing so is zero, and every cost draw is positive.*

(ii) Λ *is symmetric in x and increasing in $|x|$: the repricing probability depends only on the magnitude of the gap, not its sign. The surplus Bx^2 depends only on x^2. Larger gaps produce larger surpluses, which exceed more cost draws.*

[2]The repricing hazard $\Lambda(x)$ is the "adjustment function" in the language of Caballero and Engel (1993, 1999). Those papers take the adjustment function as a primitive and characterise how its shape determines aggregate dynamics. The present contribution is to derive $\Lambda(x)$ from first principles as a function of the menu cost distribution G, establishing G as the deeper structural primitive.

(iii) $\Lambda'(x) = g(Bx^2/w) \cdot 2Bx/w$*: the sensitivity depends on the density* g *evaluated at the firm's surplus. Where* g *is high, many firms are near the threshold, and a small change in the gap tips many across it. Where* g *is low, few firms are near the threshold, and the same change has little effect.*

(iv) $\Lambda(x) \to 1$ *for* $|x| \geq \sqrt{w\bar{\xi}/B}$*: when the gap is sufficiently large, the surplus exceeds even the highest cost draw, and the firm reprices with certainty. This defines the "inaction band" beyond which all firms adjust.*

Together, these properties paint a simple picture of the repricing decision. A firm that is well priced ($x \approx 0$) has little to gain from repricing and almost never does. As its price drifts further from the optimum, the gain from repricing grows and exceeds more cost draws, so the repricing probability rises smoothly. Eventually, for a sufficiently large gap, the gain exceeds every possible cost draw, and the firm reprices with certainty. The shape of this transition — how quickly the probability rises from zero to one — is governed entirely by the density g of menu costs. A dense region of g near a given surplus level means many firms are close to indifferent there, producing a gentle slope in Λ. A sparse region produces a steep jump. This is the microeconomic mechanism through which G shapes aggregate price dynamics.

3.3 Stationary distribution of price gaps

The repricing hazard $\Lambda(x)$ governs individual pricing decisions. To close the model, one must also characterise the cross-sectional density of price gaps that arises in the stationary equilibrium. Let z denote the idiosyncratic desired-price shock with density ϕ_z, and suppose that in steady state inflation is zero. The price gap evolves as follows: a firm with gap x adjusts with probability $\Lambda(x)$, resetting its gap to a fresh draw $z' \sim \phi_z$; with the complementary probability $1 - \Lambda(x)$, it does not adjust and its gap drifts to $x + z'$.

Proposition 7 (Stationary distribution)**.**
Under the repricing hazard $\Lambda(x) = G(Bx^2/w)$ *and idiosyncratic shock density* ϕ_z*, the stationary density* f_{ss} *satisfies the fixed-point equation*

$$f_{ss}(x) = \underbrace{\bar{\Lambda} \cdot \phi_z(x)}_{\text{inflow from adjusters}} + \underbrace{\int (1 - \Lambda(\tilde{x})) \cdot \phi_z(x - \tilde{x}) \cdot f_{ss}(\tilde{x})\, d\tilde{x}}_{\text{inflow from non-adjusters}}, \tag{12}$$

where $\bar{\Lambda} = \int \Lambda(x) f_{ss}(x)\, dx$ is the aggregate adjustment frequency. In steady state with zero inflation, f_{ss} is symmetric around zero.

Proof. See Online Appendix C. ■

The equation describes a balance between two flows. The first term is the "reset" flow: each period, a fraction $\bar{\Lambda}$ of firms reprice. After repricing, their new gap is simply a fresh shock $z' \sim \phi_z$, drawn independently of their old gap. These firms are pulled back toward zero. The second term is the "diffusion" flow: firms that do not reprice keep their old gap and accumulate a new shock, drifting further from the optimum. The stationary density f_{ss} is the cross-section where these two flows balance.

The shape of G governs this balance. When G is concentrated (steep hazard), firms with large gaps reprice quickly, and the reset flow dominates. The cross-section is tightly concentrated near zero — most firms are close to their optimal price. When G is dispersed (flat hazard), firms with large gaps are almost as likely to keep their price as firms with small gaps, and the diffusion flow dominates. The cross-section is dispersed, with substantial mass at large gaps. The distribution G thus governs both the individual hazard and, through equation (12), the steady-state cross-section of price gaps. With the hazard and the stationary distribution in hand, the next question is how the composition of repricing firms — not just their number — shapes aggregate outcomes.

3.4 The selection effect

Under state-dependent pricing, firms with the largest price gaps are most likely to reprice, amplifying the aggregate price response relative to Calvo (Golosov and Lucas, 2007). The duality provides a precise characterisation of how G governs this selection. I first define a measure of selection intensity, then show how it depends on G.

Definition 4 (Selection intensity)**.**
Given a cross-sectional density $f(x)$ of price gaps, the average absolute adjusted price change *is defined as*

$$\overline{|x|}^{adj} := \frac{\int |x| \cdot \Lambda(x) \cdot f(x)\, dx}{\int \Lambda(x) \cdot f(x)\, dx} = \frac{E[\Lambda(x) \cdot |x|]}{E[\Lambda(x)]}. \tag{13}$$

This is the mean absolute price change among repricing firms. The selection intensity *is the proportional excess of the average adjuster's absolute price change over the population average:*

$$\mathcal{S} \equiv \frac{\overline{|x|}^{adj}}{E[|x|]} - 1. \tag{14}$$

When Λ is constant (Calvo), $\mathcal{S} = 0$: every firm is equally likely to reprice, so adjusters look just like the population. When Λ is a step function (Golosov-Lucas), $\mathcal{S}$ is maximal: only firms at the boundary of the inaction band reprice, so every adjuster has a large gap. In between, selection intensity is governed by how sharply Λ increases with $|x|$. The next proposition makes the link to G precise: since $\Lambda(x) = G(Bx^2/w)$, the shape of the cost distribution directly determines how strong selection is.

Proposition 8 (Selection and the menu cost distribution)**.**
Fix the cross-sectional density f. The selection intensity $\mathcal{S}$ is:

(i) Increasing in the responsiveness of Λ to $|x|$, i.e., in $\Lambda'(x)/\Lambda(x)$. The more sharply the hazard increases with the gap, the more adjusters are concentrated among high-gap firms.

(ii) Higher when g is smaller (agents sparse near the threshold), since a sparse density means Λ is more responsive to x^2. When g is large, Λ is nearly flat and selection is weak.

(iii) Zero in the Calvo limit (constant Λ) and maximal in the Golosov-Lucas limit (step Λ).

Proof.
Decompose the adjustment-weighted mean as $\overline{|x|}^{adj} = E[|x|] + \text{Cov}(\Lambda(x)/\bar{\Lambda},\ |x|)$. The covariance is positive when Λ is increasing in $|x|$, and its magnitude is governed by the elasticity $\partial \ln \Lambda / \partial \ln |x| = 2 \cdot g(Bx^2/w)/G(Bx^2/w) \cdot Bx^2/w$, which is twice the hazard rate of G at the firm's surplus. The full derivation is in Online Appendix C. ■

A concentrated G (high density near $\bar{\xi}$) means most firms face large menu costs. Only firms with very large gaps find it worthwhile to reprice, so the hazard is steep and selection is strong. A dispersed G (density spread across $[0, \bar{\xi}]$) means many firms face small costs. Even firms with small gaps reprice frequently, the hazard is flat, and selection is weak. The density g at the threshold is the key: it determines how many additional firms are tipped into adjustment by a small increase in the gap.

3.5 The Calvo–Golosov-Lucas spectrum

The results so far hold for any G with a positive density. To make the framework quantitatively useful, one needs a tractable parametric family that can span the range of pricing behaviour observed in the data. The two canonical models of price rigidity in the New Keynesian literature — Calvo and Golosov-Lucas — correspond to polar extreme shapes of the hazard $\Lambda(x)$. The goal of this subsection is to connect them through a single continuous family indexed by one parameter.

Calvo (1983) In the Calvo (1983) model, each firm reprices with a fixed probability $\lambda \in (0,1)$ regardless of its price gap. Since $\Lambda(x) = G(Bx^2/w)$ is always increasing in $|x|$, a strictly constant hazard cannot arise from any non-degenerate absolutely continuous G. The framework therefore does not literally nest Calvo; the following proposition makes the limiting sense precise.

Proposition 9 (Calvo limit)**.**
Let G be absolutely continuous with positive density on $[0, \bar{\xi}]$ and fix a target frequency $\lambda \in (0,1)$. Define a scaled family $G_\theta(\xi) = G(\xi/\theta)$ for $\theta > 0$, and let $\Lambda_\theta(x) = G_\theta(Bx^2/w) = G(Bx^2/(w\theta))$. For each θ, calibrate $\bar{\xi}$ so that the average frequency equals λ: $E_{f_{ss}}[\Lambda_\theta(x)] = \lambda$. As $\theta \to \infty$:

(i) The normalised hazard converges uniformly: $\sup_{x \in supp(f_{ss})} |\Lambda_\theta(x)/\lambda - 1| \to 0$.

(ii) Selection vanishes: $\mathcal{S}_\theta \to 0$.

That is, dispersing G flattens the hazard and eliminates selection. The implications for the Phillips curve slope and non-neutrality are derived in Sections 3.6 and 3.7.

Proof. See Online Appendix C. ■

Golosov and Lucas (2007): step-function hazard When G is degenerate at $\bar{\xi}$, $\Lambda(x) = \mathbf{1}[Bx^2 \geq w\bar{\xi}]$ — the canonical Ss rule with inaction band $\bar{x} = \sqrt{w\bar{\xi}/B}$. The step-function hazard generates maximal selection and dramatically reduces non-neutrality relative to Calvo.

The power law family The power law $G(\xi) = (\xi/\bar{\xi})^\gamma$, $\gamma > 0$, provides a tractable one-parameter family that spans the entire spectrum between Calvo and Golosov-Lucas. The parameter γ is the degree of cost concentration. When γ is low, the density g is decreasing: most firms draw small menu costs, adjustment is cheap, and the hazard rises gently with the gap — close to Calvo. When γ is high, the density concentrates near $\bar{\xi}$: most firms draw large menu costs, only firms with very large gaps find it worthwhile to reprice, and the hazard rises steeply — close to Golosov-Lucas. The induced hazard and cost function are:

$$\Lambda(x) = \left(\frac{Bx^2}{w\bar{\xi}}\right)^{\gamma}, \qquad C(\varphi) = \frac{\gamma}{\gamma+1} w\bar{\xi}\,\varphi^{1+1/\gamma}. \tag{15}$$

This family nests the uniform distribution ($\gamma = 1$) and extends to right-skewed cost distributions ($\gamma < 1$).

Example 1 (Power law menu costs)**.**
For $G(\xi) = (\xi/\bar{\xi})^{\gamma}$, the following relations hold:

(i) $G^{-1}(\varphi) = \bar{\xi}\varphi^{1/\gamma}$.

(ii) $C(\varphi) = \frac{\gamma}{\gamma+1} w\bar{\xi}\,\varphi^{(\gamma+1)/\gamma}$.

(iii) $C'(\varphi) = w\bar{\xi}\,\varphi^{1/\gamma}, \quad C''(\varphi) = \frac{w\bar{\xi}}{\gamma}\varphi^{(1-\gamma)/\gamma}$.

(iv) $\Lambda(x) = \left(\frac{Bx^2}{w\bar{\xi}}\right)^{\gamma}$.

The parameter γ governs the entire spectrum: $\gamma = 1$ gives uniform costs and quadratic C (Dotsey, King, and Wolman, 1999; Khan and Thomas, 2008); $\gamma \to \infty$ gives the step-function hazard of Golosov and Lucas (2007); $\gamma < 1$ gives right-skewed costs with a concave hazard and weak selection; $\gamma \to 0$ gives the "cheap adjustment" limit where nearly all firms reprice. Note that the $\gamma \to 0$ limit with fixed $\bar{\xi}$ yields near-certain adjustment, not a constant interior hazard. The Calvo approximation requires the frequency-calibrated limit of Proposition 9, which adjusts $\bar{\xi}$ with γ to hold $\bar{\Lambda}$ fixed.

3.6 Aggregate inflation and the Phillips curve

The repricing hazard $\Lambda(x)$ governs individual pricing behaviour. Because the duality gives $\Lambda(x) = G(Bx^2/w)$ in closed form, aggregate inflation can be written directly in terms of G — enabling analytical decompositions that would require numerical computation without the duality. Let $f_t(x)$ denote the cross-sectional density of price gaps at time t. Aggregate inflation is

$$\pi_t = E_{f_t}[\Lambda(x_t) \cdot x_t] = \int \Lambda(x)\, x\, f_t(x)\, dx. \tag{16}$$

Each firm contributes in proportion to $\Lambda(x) \cdot x$. Firms with large gaps and high repricing probabilities contribute the most; firms near the optimum contribute almost nothing.

To understand how aggregate inflation responds to nominal disturbances, consider a small aggregate shock μ_t that shifts the desired price of all firms by μ_t, moving the cross-sectional distribution of gaps from $f_{ss}(x)$ (the zero-inflation steady state) to $f_{ss}(x - \mu_t)$.

Proposition 10 (Nonlinear Phillips curve)**.**
Under the quadratic surplus approximation $S(x) = Bx^2$ and menu cost distribution G, aggregate inflation satisfies

$$\pi_t = E_t\left[G\left(\frac{Bx_t^2}{w}\right) \cdot x_t\right]. \tag{17}$$

For small aggregate disturbances around a zero-inflation steady state with gap distribution f_{ss}:

$$\pi_t \approx \kappa(G) \cdot \mu_t, \tag{18}$$

where the distribution-dependent slope *is*

$$\kappa(G) := \underbrace{\int G\left(\frac{Bx^2}{w}\right) f_{ss}(x)\, dx}_{\text{extensive margin (frequency)}} + \underbrace{\int g\left(\frac{Bx^2}{w}\right) \frac{2Bx^2}{w} f_{ss}(x)\, dx}_{\text{intensive margin (selection)}}. \tag{19}$$

Proof.
Write $f_t(x) = f_{ss}(x - \mu_t)$ and differentiate $\pi_t = \int G(Bx^2/w)\, x\, f_{ss}(x - \mu_t)\, dx$ with respect to μ_t at $\mu_t = 0$. Integrating by parts and expanding $\frac{d}{dx}[G(Bx^2/w) \cdot x]$ yields the two terms: the first from G itself (extensive margin) and the second from $g \cdot 2Bx^2/w$ (intensive margin). The full derivation is in Online Appendix C. ■

The extensive margin (first term) captures the average repricing frequency — the only margin under Calvo. The intensive margin (second term) captures selection:

a nominal shock tips firms near the threshold into adjustment, amplifying the price response. It is absent under Calvo and dominant under Golosov-Lucas.

The slope $\kappa(G)$ varies across the spectrum because G governs the relative strength of these two margins. The following corollary traces out the implications for the canonical cases.

Corollary 3 (How G governs the Phillips curve)**.**

(i) ***Calvo limit****: When G is highly dispersed, g is small across the region where f_{ss} has mass, and the intensive margin is negligible. The slope approaches $\kappa \approx \bar{\Lambda}$, approximating the textbook New Keynesian Phillips Curve.*

(ii) ***Golosov-Lucas****: When G is degenerate, $G(Bx^2/w) = \mathbf{1}[|x| \geq \bar{x}]$ and $g = \delta_{\bar{\xi}}$. The intensive margin dominates: inflation is driven by selection of firms at the threshold. The slope is very steep, implying low non-neutrality.*

(iii) ***Uniform****: When $G = U[0, \bar{\xi}]$, the slope simplifies to*

$$\kappa^U = \frac{3B}{w\bar{\xi}} Var_{ss}(x). \tag{20}$$

The slope depends on the steady-state dispersion of price gaps. Economies with more dispersed gaps have steeper Phillips curves, because the selection margin is stronger when more firms have large gaps near the threshold.

(iv) ***Power law*** *γ: The slope κ is strictly increasing in γ. Higher γ means more concentrated menu costs, a steeper hazard, stronger selection, and a steeper Phillips curve.*

3.7 Non-neutrality decomposition

The degree of monetary non-neutrality is inversely related to the price flexibility index κ. When κ is high, prices adjust quickly and real effects are small. When κ is low, prices are sticky and real effects are large. Because the duality expresses both κ and $\bar{\Lambda}$ as explicit functions of G and g, it yields a clean decomposition of κ into two forces. For the power law family, this decomposition is available in closed form.

Proposition 11 (Frequency vs. selection)**.**
The real effect of a monetary shock dm is proportional to $(1-\kappa)\,dm$. The flexibility index decomposes as

$$\kappa(G) = \underbrace{\bar{\Lambda}}_{frequency} + \underbrace{(\kappa - \bar{\Lambda})}_{selection\ wedge}, \tag{21}$$

where $\bar{\Lambda} = E_{f_{ss}}[\Lambda(x)]$ is the average adjustment frequency. Non-neutrality is governed by the selection wedge *$\kappa - \bar{\Lambda}$.*

The decomposition separates two conceptually distinct channels. Frequency ($\bar{\Lambda}$) measures how often firms reprice. The selection wedge ($\kappa - \bar{\Lambda}$) measures how well those repricings are targeted at the firms that need them most. Under Calvo, $\kappa = \bar{\Lambda} = \lambda$ and selection contributes nothing: repricing is random. For any absolutely continuous G, the selection wedge is strictly positive: firms with larger gaps are more likely to reprice, amplifying the aggregate price response beyond what frequency alone would produce.

This distinction has a striking implication. Two economies can share the same repricing frequency yet differ dramatically in non-neutrality. The difference is entirely attributable to the selection wedge, which depends on the shape of G.

Corollary 4 (Ordering of non-neutrality)**.**
Across models calibrated to the same average frequency $\bar{\Lambda}$:

(i) ***Calvo****: selection wedge $= 0$; maximum non-neutrality.*

(ii) ***Any absolutely continuous*** *G: positive selection wedge; intermediate non-neutrality.*

(iii) ***Golosov-Lucas****: maximum selection wedge; minimum non-neutrality.*

For the power law family, non-neutrality is strictly decreasing in γ. The duality embeds this well-known ordering (Golosov and Lucas, 2007) in a continuous framework parametrised by G.

The corollary establishes a qualitative ordering. Within the power law family, the ordering sharpens to a monotone relationship: γ alone determines selection, non-neutrality, and the Phillips curve slope.

Proposition 12 (Monotonicity in γ).
For the power law family $G(\xi) = (\xi/\bar{\xi})^\gamma$ with fixed $\bar{\xi}$ and B/w, calibrated to the same average adjustment frequency:

(i) Selection intensity $\mathcal{S}$ is strictly increasing in γ.

(ii) Monetary non-neutrality is strictly decreasing in γ.

(iii) The Phillips curve slope at $\mu = 0$ is strictly increasing in γ.

Thus γ indexes "how Calvo-like" the pricing environment is. Low γ (dispersed costs, weak selection, large non-neutrality) resembles Calvo. High γ (concentrated costs, strong selection, small non-neutrality) resembles Golosov-Lucas. The distribution G determines the economy's position on this spectrum.

Proof. See Online Appendix C. ■

A natural question is whether repricing frequency alone determines non-neutrality. The answer is no. Frequency is only half the story; what matters equally is how that repricing is allocated across firms. The following proposition makes this precise.

Proposition 13 (Frequency does not determine non-neutrality).
Fix f_{ss} and B/w. There exist absolutely continuous distributions G_1, G_2 with:

(i) The same average repricing frequency: $\bar{\Lambda}_1 = E_{f_{ss}}[\Lambda_1(x)] = E_{f_{ss}}[\Lambda_2(x)] = \bar{\Lambda}_2$.

(ii) Different flexibility indices: $\kappa(G_1) \neq \kappa(G_2)$, hence different selection wedges and different real effects of nominal shocks.

Proof. See Online Appendix C. ■

For the power law family, the decomposition takes a particularly clean form. The ratio of total flexibility to frequency is pinned down by γ alone — independent of the steady-state distribution, the surplus parameter, or the wage.

Corollary 5 (Selection-frequency ratio for power laws).
For $G(\xi) = (\xi/\bar{\xi})^\gamma$:

$$\kappa = (1 + 2\gamma)\,\bar{\Lambda}, \qquad \textit{selection wedge} = 2\gamma\,\bar{\Lambda}. \tag{22}$$

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

Table 2 illustrates. Six distributions calibrated to the same repricing frequency $\bar{\Lambda} = 0.10$ produce selection wedges ranging from 0 to 0.60. The implied cumulative output effects differ by a factor of three.

Table 2: Same repricing frequency, different selection wedges

Distribution G	$\bar{\Lambda}$	κ	$\kappa - \bar{\Lambda}$	$1 - \kappa$
Calvo	0.10	0.10	0	0.90
Power law ($\gamma = \frac{1}{2}$)	0.10	0.20	0.10	0.80
Exponential	0.10	0.27	0.17	0.73
Uniform ($\gamma = 1$)	0.10	0.30	0.20	0.70
Power law ($\gamma = 2$)	0.10	0.50	0.40	0.50
Power law ($\gamma = 3$)	0.10	0.70	0.60	0.30

Notes: Each row is calibrated to $\bar{\Lambda} = 0.10$. The last column, $1 - \kappa$, is proportional to the cumulative real output effect of a monetary shock. *Power law entries:* $\kappa = (1 + 2\gamma)\bar{\Lambda}$, which holds for any symmetric f_{ss}. *Exponential entry:* $G(\xi) = 1 - e^{-\lambda\xi}$ with $f_{ss} = N(0, \sigma_x^2)$. Setting $\rho = B\sigma_x^2/w$, the χ^2 moment generating function gives $\bar{\Lambda} = 1 - (1 + 2\lambda\rho)^{-1/2}$ and intensive margin $\lambda\rho(1 + 2\lambda\rho)^{-3/2}$. Calibrating $\lambda\rho$ to match $\bar{\Lambda} = 0.10$ gives $\kappa \approx 0.27$.

Figure 2 complements the table with the dynamic counterpart. Four power law economies — from near-Calvo ($\gamma = 0.20$) through uniform ($\gamma = 1$) to near-Golosov-Lucas ($\gamma = 5$) — are calibrated to the same quarterly repricing frequency. Each economy is hit with a one-time 1% monetary shock. The normalised output gap $y(t)/\delta$ is tracked as prices catch up to the new money supply.

The mechanism is selection: low γ means nearly random repricing and large persistent output effects; high γ means targeted repricing and small transient effects. The impulse responses are plotted in Figure 2 (Section 4.4).

3.8 Summary of the spectrum

The preceding subsections developed the pricing application in stages: from the repricing hazard, through the selection effect and the Calvo–Golosov-Lucas spectrum, to the Phillips curve and the non-neutrality decomposition. At every stage, a single object — the menu cost distribution G — determined the outcome. Table 3 collects these results into one view.

Reading the table from left to right traces the effect of concentrating the cost distribution. A dispersed G (left column) produces a nearly flat hazard: repricing

Table 3: The state-dependence spectrum

	Calvo	**Uniform G**	**Golosov-Lucas**
Distribution G	Dispersed	$U[0, \bar{\xi}]$	Point mass at $\bar{\xi}$
$C(\varphi)$	Highly curved	Quadratic	Linear on $\{0, 1\}$
$C''(\varphi)$	$\to \infty$	$w\bar{\xi}$	0 or ∞
Hazard $\Lambda(x)$	$\approx$ constant	$Bx^2/(w\bar{\xi})$	Step function
Selection	$\approx$ none	Moderate	Maximum
NKPC slope	$\approx$ constant	State-dependent	Very steep
Non-neutrality	Maximum	Intermediate	Minimum
Power law γ	$\to 0^+$	1	$\to \infty$

is close to random, selection is absent, the Phillips curve is nearly linear, and real effects of monetary shocks are large. This is the Calvo limit. A concentrated G (right column) produces a step-function hazard: only firms at the inaction boundary reprice, selection is maximal, the Phillips curve is very steep, and real effects are small. This is the Golosov-Lucas limit. Uniform G (middle column) sits halfway: quadratic costs, a linear hazard, and moderate selection.

The key message is that these are not different models with different structures. They are different shapes of one distribution, G, connected by the duality. The power law parameter γ indexes the position on this spectrum continuously. The empirical question — addressed in the next section — is where real economies sit.

4 Quantitative analysis

The theoretical results of Sections 2–3 establish that the distribution G is the structural primitive governing adjustment dynamics. A key payoff of the duality is that it makes G identifiable: the closed-form hazard $\Lambda(x) = G(Bx^2/w)$ implies that γ can be read off directly from the log-log slope of the repricing hazard, without solving the full model. This section asks: what does G look like in the data? The answer disciplines the theory and quantifies the frequency-selection decomposition.

The estimation proceeds in two steps. First, γ is identified from the shape of the repricing hazard — a strategy that requires only the conditional adjustment probability as a function of the price gap. Second, the model is simulated at the estimated γ and compared to the full set of micro-price moments reported in the literature.

4.1 Identification from the repricing hazard

The theoretical hazard for the power law family is $\Lambda(x) = (Bx^2/(w\bar{\xi}))^\gamma$. Taking logs:

$$\log \Lambda(x) = \gamma \log\left(\frac{Bx^2}{w}\right) - \gamma \log \bar{\xi} = 2\gamma \log |x| + \text{const.} \tag{23}$$

The slope of $\log \Lambda$ against $\log(x^2)$ is γ. This is a clean identification: the log-log regression uses only data from the interior of the hazard ($0 < \Lambda < 1$), where the power law structure is directly testable. The intercept identifies $\bar{\xi}$ given B/w; the slope identifies γ without knowledge of $\bar{\xi}$, B, or w separately.

A Monte Carlo exercise (Figure D.1 in the Online Appendix) confirms that the log-log regression recovers γ with mean absolute error 0.012 and $R^2 > 0.99$ for all values tested.

4.2 Calibration and moment matching

The baseline calibration sets $B = w = 1$ (normalisation) and $\sigma_z = 0.04$ (quarterly standard deviation of idiosyncratic desired-price shocks, consistent with Golosov and Lucas 2007 and Nakamura and Steinsson 2008). For each γ, the upper bound $\bar{\xi}$ is calibrated to match a quarterly repricing frequency of 24%, corresponding to approximately 8–9% monthly excluding temporary sales (Nakamura and Steinsson, 2008).

In the baseline model with homogeneous σ_z, kurtosis is everywhere below 3 (Table 4, left panel), a well-known limitation of single-product menu cost models. Following Midrigan (2011), the extended model draws per-product σ_z from a lognormal distribution with mean 0.04 and coefficient of variation 1.0. This single additional parameter generates kurtosis in the empirical range while preserving the identification of γ from the hazard shape.

4.3 Estimation results

Point estimate Matching the empirical kurtosis of approximately 5.0 in the heterogeneous-σ_z model yields

$$\hat{\gamma} \approx 0.4. \tag{24}$$

Table 4: Model moments: baseline and heterogeneous σ_z

	Baseline (homogeneous σ_z)					**Heterogeneous σ_z (CV = 1)**				
γ	Freq	$E[\|\Delta p\|]$	Kurt	Sm%	κ	Freq	$E[\|\Delta p\|]$	Kurt	Sm%	κ
0.15	0.24	7.2%	2.79	17%	0.31	0.24	7.9%	7.38	28%	0.31
0.25	0.24	7.5%	2.47	14%	0.36	0.24	8.8%	6.18	21%	0.36
0.40	0.24	7.8%	2.11	10%	0.43	0.24	9.7%	4.86	14%	0.44
0.60	0.24	8.1%	1.80	6%	0.53	0.24	10.3%	4.49	9%	0.53
0.80	0.24	8.3%	1.61	4%	0.63	0.24	10.6%	4.44	5%	0.63
1.00	0.24	8.4%	1.51	2%	0.72	0.24	10.8%	4.37	3%	0.73
1.50	0.24	8.6%	1.38	1%	0.96	0.24	11.0%	4.31	1%	0.97
2.00	0.24	8.7%	1.33	0%	1.20	0.24	11.1%	4.28	0%	1.21
Data targets		8–11%	4–6	25–40%			8–11%	4–6	25–40%	

Notes: All economies calibrated to quarterly repricing frequency $\bar{\Lambda} = 0.24$. "Freq" is the simulated frequency (confirming calibration). "$E[|\Delta p|]$" is the mean absolute (non-sale) price change. "Kurt" is the raw kurtosis. "Sm%" is the fraction of price changes smaller than 2.5% in absolute value. $\kappa = (1 + 2\gamma)\bar{\Lambda}$. Data targets from Nakamura and Steinsson (2008), Midrigan (2011), and Alvarez, Le Bihan, and Lippi (2016). 15,000 firms, 150 quarters per simulation.

This corresponds to a right-skewed cost distribution: the density $g(\xi) = \gamma\xi^{\gamma-1}/\bar{\xi}^{\gamma}$ is decreasing for $\gamma < 1$, meaning most firms face small menu costs while a few face large ones. The economy sits between Calvo and Golosov-Lucas on the spectrum, with meaningful selection.

Implied decomposition At $\hat{\gamma} = 0.4$, the flexibility index is

$$\kappa = (1 + 2 \times 0.4) \times 0.24 = 0.43.$$

The frequency component is $\bar{\Lambda} = 0.24$, and the selection wedge is $\kappa - \bar{\Lambda} = 0.19$. Selection accounts for 44% of total price flexibility:

$$\frac{\kappa - \bar{\Lambda}}{\kappa} = \frac{0.19}{0.43} = 0.44.$$

Non-neutrality, proportional to $1 - \kappa = 0.57$, is 57% of the Calvo benchmark — a substantial reduction due to selection, but far from the near-zero non-neutrality of Golosov-Lucas. Table 5 summarises.

Table 5: Frequency-selection decomposition at $\hat{\gamma} = 0.4$

	Value	**Share of κ**
Frequency ($\bar{\Lambda}$)	0.24	56%
Selection wedge ($\kappa - \bar{\Lambda}$)	0.19	44%
Flexibility (κ)	0.43	100%
Non-neutrality ($1 - \kappa$)	0.57	—

Notes: $\kappa = (1+2\hat{\gamma})\bar{\Lambda}$ with $\hat{\gamma} = 0.4$ and $\bar{\Lambda} = 0.24$. Non-neutrality as a fraction of Calvo: $(1-\kappa)/(1-\bar{\Lambda}) = 0.57/0.76 = 0.75$. Selection reduces non-neutrality by 25% relative to the Calvo benchmark at the same frequency.

Comparison with the literature The estimate $\hat{\gamma} \approx 0.4$ is consistent with several independent findings. Midrigan (2011) estimates a menu cost model with leptokurtic price changes and finds that matching the size distribution requires a multi-product structure that generates many small price changes alongside occasional large ones — precisely the pattern produced by a right-skewed G ($\gamma < 1$). Alvarez, Le Bihan, and Lippi (2016) report a kurtosis-frequency ratio of approximately 3.5, implying substantial but not maximal non-neutrality; the present estimate yields a CIR ratio of $\text{Kurt}/(6 \cdot \text{Freq}) \approx 4.9/(6 \times 0.24) = 3.4$, closely matching their sufficient statistic. Vavra (2014) documents that the cross-sectional variance of price changes comoves with inflation — a prediction that follows directly from the state-dependent hazard $\Lambda(x) = G(Bx^2/w)$ whenever G is non-degenerate.

4.4 Hazard functions and impulse responses

Figure 1 plots the simulated repricing hazard for four values of γ, along with the theoretical prediction (dashed). The hazard is concave in $|x|$ for $\gamma < 1$ (firms with moderate gaps adjust frequently, but the marginal increase for very large gaps is small) and convex for $\gamma > 1$ (adjustment is concentrated sharply at the threshold). The estimated $\hat{\gamma} \approx 0.4$ produces a gently increasing, concave hazard — qualitatively consistent with the smooth adjustment functions estimated by Caballero and Engel (1993).

Figure 2 shows the output impulse response to a one-time 1% monetary shock. All economies share the same repricing frequency; they differ only in γ. The Calvo benchmark (zero selection) decays exponentially at rate $\bar{\Lambda}$. State-dependent economies

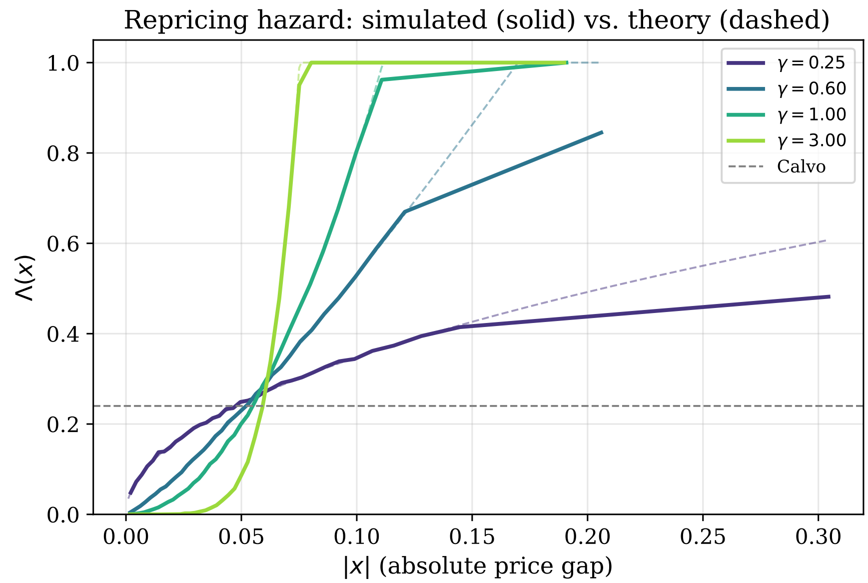

Figure 1: Repricing hazard $\Lambda(x)$ for selected γ values. Solid: simulated. Dashed: theoretical $\Lambda(x) = (Bx^2/(w\bar{\xi}))^\gamma$. All economies calibrated to $\bar{\Lambda} = 0.24$. The Calvo benchmark (constant hazard) is shown for reference.

converge faster because repricing is targeted at the most mispriced firms. The gap between Calvo and the state-dependent economy widens with γ: at $\hat{\gamma} = 0.4$, the output gap is approximately 75% of the Calvo path — selection matters, but frequency remains the dominant force.

Figure D.2 in the Online Appendix shows the distribution of price changes for several values of γ; the estimated $\hat{\gamma} = 0.4$ produces a moderately leptokurtic distribution with kurtosis around 5, matching the empirical evidence.

Relation to Alvarez, Le Bihan, and Lippi (2016) The ABLL sufficient statistic characterises non-neutrality in terms of the kurtosis and frequency of price changes. Within the power law family, both are endogenous functions of γ and $\bar{\xi}$. The present framework therefore provides the structural layer from which ABLL's observable diagnostics are derived: γ determines the kurtosis-frequency ratio, and hence the CIR, through the shape of the repricing hazard. The practical value of γ over the ABLL moments is that γ is a single structural parameter with a clear interpretation (the shape of the cost distribution), whereas the ABLL ratio combines two moments that can move independently in response to changes in the economic environment (e.g.,

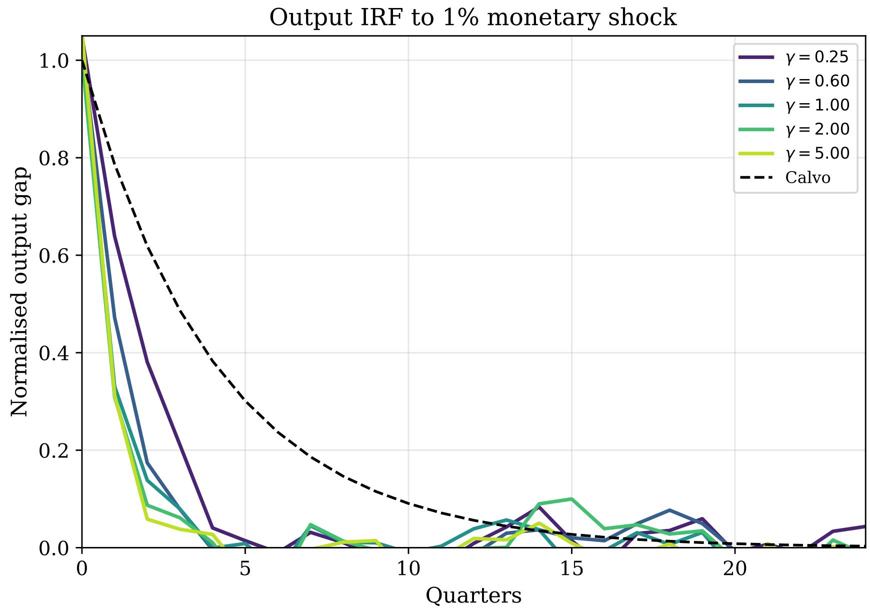

Figure 2: Output impulse response to a 1% monetary shock. All economies calibrated to $\bar{\Lambda} = 0.24$ quarterly. Higher γ implies stronger selection, faster convergence, and smaller cumulative output effects. The Calvo benchmark (dashed) has zero selection.

changes in trend inflation shift both frequency and kurtosis, but the mapping depends on G).

4.5 Evidence from micro price data

This section estimates γ from micro price data in the Dominick's Finer Foods scanner dataset (Midrigan, 2011; Chevalier, Kashyap, and Rossi, 2003). The dataset covers seven product categories (analgesics, cereals, cheeses, cookies, canned soup, laundry detergent, soft drinks) across approximately 100 stores and 400 weeks, yielding 42 million store-week-UPC observations. Temporary sales are removed using both the recorded sale indicator and a V-shaped filter that flags price drops reversed within three weeks.

Moments Table 6 reports the key moments. The kurtosis of non-sale price changes is 7.6, and 21% of changes are smaller than 2.5% in absolute value — both indicating substantial leptokurtosis and a large mass of small adjustments, consistent with weak selection.

Table 6: Dominick's price change moments (excl. sales)

Moment	Value
Frequency (weekly)	9.3%
Frequency (quarterly)	71.8%
Mean $\lvert\Delta p\rvert$	13.5%
Kurtosis	7.62
Fraction $\lvert\Delta p\rvert < 2.5\%$	21%
Fraction $\lvert\Delta p\rvert < 5\%$	42%
N (price changes, excl. sales)	1,656,025

Moment-based identification Matching the empirical kurtosis and fraction of small changes to the model's predictions (Table 4, heterogeneous-σ_z panel) yields $\hat{\gamma} \approx 0.15$ from kurtosis and $\hat{\gamma} \approx 0.25$ from the fraction of small changes, with a central estimate of $\hat{\gamma} \approx 0.2$. This places the Dominick's pricing environment even closer to the Calvo end of the spectrum than the calibration estimate of $\hat{\gamma} \approx 0.4$ from Section 4.3. The difference is consistent with the nature of the data: scanner data from a single grocery chain exhibits more product-level heterogeneity and higher repricing frequency than the broad CPI coverage underlying the Nakamura and Steinsson (2010) moments used in Section 4.2.

Duration dependence and cross-product heterogeneity Figure 3 shows the distribution of non-sale price changes: strongly leptokurtic, with a sharp peak near zero and heavy tails — qualitatively matching the model's prediction for low γ.

The duration-based repricing hazard is steeply decreasing — negative duration dependence documented by Nakamura and Steinsson (2008). Within the present framework, this is a signature of cross-product heterogeneity in G: the within-product hazard $\Lambda(x) = G(Bx^2/w)$ is increasing in $|x|$ for each product, but the pooled hazard across products with different G's can be decreasing in duration. Negative duration dependence in the aggregate is consistent with positive state dependence at the product level, provided menu costs are heterogeneous across products.

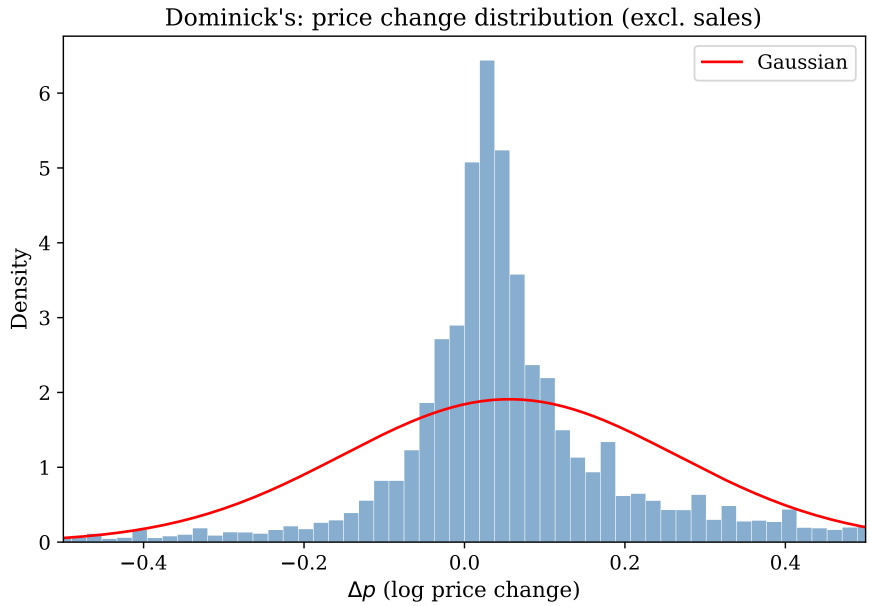

Figure 3: Distribution of non-sale price changes in Dominick's Finer Foods (7 categories, 42 million observations). The distribution is strongly leptokurtic (kurtosis = 7.6), with a sharp peak near zero and heavy tails relative to the Gaussian benchmark (red).

5 Discussion

Further applications Although the pricing application is developed in greatest depth (with investment in Online Appendix B), the duality applies to any extensive-margin decision with heterogeneous costs.

Heterogeneous entry costs micro-found smooth aggregate entry rates: the free-entry condition $S = C'(\varphi) = wG^{-1}(\varphi)$ equates the marginal entrant's cost to the marginal cost of raising the entry rate. Heterogeneous vacancy-creation costs micro-found the convex posting cost in Pissarides (2000), with C'' reflecting the shape of the cost distribution rather than a technological curvature.

Heterogeneous participation costs (childcare, commuting, health) generate a smooth labour supply without Rogerson (1988) lotteries: the participation elasticity is $g(G^{-1}(\varphi))/w$, the density of workers at the margin. The same logic extends to technology adoption, credit default (Bernanke, Gertler, and Gilchrist, 1999), and human capital investment. In each case, the shape of G governs the aggregate response.

Structural discipline, common language, identifiability The G framing provides structural discipline: quadratic costs require uniform G; iso-elastic costs require power law G. The functional form of C encodes a distributional assumption that G makes transparent and testable. It also provides a common language — the Calvo–Golosov-Lucas debate, selection in non-neutrality, and the Khan and Thomas (2008) irrelevance result are all statements about G — and an identifiable primitive, since G can be recovered from the adjustment hazard (Section 4.1), the size distribution of price changes, or investment spike frequencies. Oskolkov and Lippi (2026) recover G from Italian plant-level data; the duality implies their G simultaneously determines C, the investment hazard, and capital misallocation.

Generic smoothness of lumpy aggregates The generic smoothness of aggregate dynamics despite micro-level lumpiness resonates with a broader pattern. Auclert, Rognlie, and Straub (2020) document that individual households make discrete, lumpy consumption adjustments ("micro jumps"), yet aggregate impulse responses to monetary policy are smooth and hump-shaped ("macro humps"). The duality provides a structural explanation: whenever G has a positive density — the generic case by Proposition 2 — binary micro-level decisions aggregate into smooth macroeconomic dynamics. Cost heterogeneity is the structural mechanism.

Scope and limitations The equivalence is an identity of value functions and adjustment probabilities. It does not imply equivalence of welfare, interpretation, or causality. The identifiable object is G (equivalently C), not the "type" of friction. An economy where firms draw random menu costs (Dotsey, King, and Wolman, 1999) and one where firms choose repricing probabilities subject to a convex penalty make different predictions about how G responds to changes in the inflation environment. The duality is a statement about reduced-form outcomes given G; endogenising G is a natural next step.

6 Conclusion

The duality established in this paper changes how one should think about adjustment frictions. The distinction between fixed and convex costs is not a distinction between two kinds of economic friction. It is a distinction between two representations of the same friction, linked by Fenchel conjugacy. The fixed cost distribution G is the structural primitive that governs both.

Three implications follow. First, the lumpy–smooth divide dissolves. Any heterogeneity in fixed costs makes the individual decision problem smooth and convex. The deterministic fixed cost — the textbook case — is the singular exception, not the rule. This overturns the standard view that convex costs approximate the "true" lumpy friction: the convex model is the generic case. It also provides a partial-equilibrium foundation for the irrelevance result of Khan and Thomas (2008): with uniform fixed costs, each firm's problem is already convex, so aggregation is not the source of smoothness.

Second, the duality reveals that five seemingly unrelated literatures — menu costs, lumpy investment, discrete choice, rational inattention, and control costs — share the same mathematical structure under different assumptions on G. The welfare cost $C(\varphi^*)$ has a direct interpretation as the expected physical adjustment cost in resource units. The smooth first-order condition $S = C'(\varphi)$ replaces kink-based dynamic programming, accelerating computation.

Third, in the pricing application, the duality gives the repricing hazard in closed form, making G identifiable from data and enabling an analytical decomposition of non-neutrality into frequency and selection. This decomposition cannot be obtained without the duality: it requires the explicit functional form $\Lambda(x) = G(Bx^2/w)$.

Several directions remain open. Embedding the framework in general equilibrium would yield quantitative predictions about the welfare cost of inflation across the Calvo–Golosov-Lucas spectrum. Endogenising the fixed cost distribution — allowing it to respond to the inflation environment — would connect to the Lucas critique and to the time-varying non-neutrality documented by Vavra (2014). Extending the identification to multi-product and multi-sector settings would test whether γ varies across markets and what drives that variation.

References

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi. 2016. "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach." *American Economic Review* 106 (10):2817–2851.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model." Working Paper 26647, National Bureau of Economic Research.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, vol. 1. Elsevier, 1341–1393.

Caballero, Ricardo J. and Eduardo M. R. A. Engel. 1993. "Microeconomic Rigidities and Aggregate Price Dynamics." *European Economic Review* 37 (4):697–711.

———. 1999. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S, s) Approach." *Econometrica* 67 (4):783–826.

———. 2007. "Price Stickiness in Ss Models: New Interpretations of Old Results." *Journal of Monetary Economics* 54 (Supplement 1):100–121.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3):383–398.

Caplin, Andrew S. and Daniel F. Spulber. 1987. "Menu Costs and the Neutrality of Money." *Quarterly Journal of Economics* 102 (4):703–725.

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi. 2003. "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data." *American Economic Review* 93 (1):15–37.

Dotsey, Michael, Robert G. King, and Alexander L. Wolman. 1999. "State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output." *Quarterly Journal of Economics* 114 (2):655–690.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics* 129 (4):1661–1710.

Golosov, Mikhail and Robert E. Lucas, Jr. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy* 115 (2):171–199.

Gourio, François and Anil K. Kashyap. 2007. "Investment Spikes: New Facts and a General Equilibrium Exploration." *Journal of Monetary Economics* 54 (Supplement 1):1–22.

Hofbauer, Josef and William H. Sandholm. 2002. "On the Global Convergence of Stochastic Fictitious Play." *Econometrica* 70 (6):2265–2294.

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.

Mattsson, Lars-Göran and Jörgen W. Weibull. 2002. "Probabilistic Choice and Procedurally Bounded Rationality." *Games and Economic Behavior* 41 (1):61–78.

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.

Midrigan, Virgiliu. 2011. "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations." *Econometrica* 79 (4):1139–1180.

Nakamura, Emi and Jón Steinsson. 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics* 123 (4):1415–1464.

———. 2010. "Monetary Non-Neutrality in a Multisector Menu Cost Model." *Quarterly Journal of Economics* 125 (3):961–1013.

Oskolkov, Aleksei and Francesco Lippi. 2026. "Misallocation with Lumpy Investment." Discussion Paper 21212, Centre for Economic Policy Research.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory*. MIT Press, 2nd ed.

Reiter, Michael. 2009. "Solving Heterogeneous-Agent Models by Projection and Perturbation." *Journal of Economic Dynamics and Control* 53 (6):649–665.

Rogerson, Richard. 1988. "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics* 21 (1):3–16.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3):665–690.

Thomas, Julia K. 2002. "Is Lumpy Investment Relevant for the Business Cycle?" *Journal of Political Economy* 110 (3):508–534.

Tsallis, Constantino. 1988. "Possible Generalization of Boltzmann-Gibbs Statistics." *Journal of Statistical Physics* 52 (1–2):479–487.

Vavra, Joseph. 2014. "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation." *Quarterly Journal of Economics* 129 (1):215–258.