The Capacity Theory of Anomaly Persistence

April 5, 2026

Abstract

In a free-entry Cournot model of anomaly arbitrage, the post-entry surviving alpha depends on market frictions, not the anomaly's raw alpha. A genericity theorem proves this alpha-independence result is not an artifact of linear price impact: for any concave impact function, the surviving alpha's sensitivity to raw alpha is bounded by $O((\gamma\sigma^2)^{-3/2})$ and vanishes as risk aversion increases. Empirical tests using 154 anomalies confirm substantial alpha compression and find that raw alpha does not predict the speed of decay.

JEL Classification: G12, G14, L13

Keywords: Anomaly persistence, arbitrage, free entry, price impact, market microstructure

1 Introduction

Anomaly returns decline after academic publication. McLean and Pontiff [2016] document that the average anomaly loses roughly 58% of its pre-publication alpha, but this average masks large heterogeneity. Momentum and quality persist for decades; many accounting-based predictors vanish within two years. The finance profession has no theory of which anomalies survive and why. What determines the cross-section of anomaly persistence?

This paper shows that the answer is approximately the same regardless of the market microstructure: competitive entry compresses anomaly alpha to a friction-determined floor that is approximately independent of the anomaly's raw alpha. Berk and Green [2004] established this logic for mutual funds under their specific model of fund flows and decreasing returns to scale. A natural question is whether the result is fragile, depending on the details of how arbitrageurs trade, or generic. This paper proves it is generic. A free-entry Cournot model of anomaly arbitrage delivers exact alpha-independence under linear price impact (Proposition 6), and a characterization theorem identifies the precise algebraic condition under which exact alpha-independence holds across market microstructure specifications (Theorem 11). Under the empirically preferred square-root impact specification, exact alpha-independence fails and larger anomalies are over-dissipated (Proposition 12). But a genericity theorem (Theorem 14) proves that the departure from alpha-independence under any concave impact function is bounded by $O((\gamma\sigma^2)^{-3/2})$: the surviving alpha's sensitivity to raw alpha vanishes at rate $(\gamma\sigma^2)^{-3/2}$ as risk aversion increases. At empirically relevant risk aversion, the bound implies variation of less than 0.1 percentage points across impact specifications. The surviving alpha is determined by the cost of bearing risk, not by the shape of the price impact function.

The economic logic is direct. Each anomaly is a commons: larger anomalies attract proportionally more arbitrageurs, and free entry dissipates the extra alpha. Consider two anomalies with identical trading environments but different raw alphas. The larger anomaly draws more entrants, and the additional entry erodes the alpha surplus until the marginal entrant earns zero profit. The surviving alpha is pinned by the cost of participation, not by the size of the original opportunity. The risk penalty $\gamma\sigma^2$ enters the zero-profit condition multiplicatively with the impact parameter, and for empirically relevant parameter values, the risk term dominates. This is why the result is robust: the surviving alpha is set primarily by the trade-off between risk and entry cost, a trade-off that does not depend on the functional form of price impact.

Three additional results follow from the model. First, in a dynamic extension, the half-life of anomaly decay depends only on the ratio of price impact to the alpha-erosion rate, not on raw alpha, entry cost, risk aversion, or volatility (Proposition 16). Price impact determines the speed of decay; all other frictions determine only the floor. Raw alpha determines neither. Second, the characterization theorem provides the exact "if and only if" condition for exact alpha-independence (Theorem 11), explaining why the result is generic: the power-law structure of profits that delivers alpha-independence arises naturally in standard trading games, and departures from it are quantitatively small when risk aversion is present. Third, with heterogeneous entry costs, the alpha decay path is convex (decelerating), consistent with rapid post-publication decline followed by long-run persistence of a residual alpha.

The contribution relative to Berk and Green [2004] is not the mechanism (free-entry dissipation of rents) but the demonstration that the mechanism's predictions are robust across the space of trading environments. Berk and Green showed alpha-independence in one model. The present paper shows it holds approximately in all of them, quantifies the approximation error, and identifies the economic force (risk deterrence) that makes it generic. This distinction matters because it converts a model-specific prediction into a near-universal one: approximate alpha-independence is not an artifact of linear impact or mean-variance preferences but a structural consequence of free entry with risk-averse participants.

The paper contributes to several additional literatures. McLean and Pontiff [2016] document the average decay but make no prediction about cross-sectional variation. The model rationalizes their average finding and generates cross-sectional predictions. Penasse [2022] distinguishes risk-based anomalies from mispricing-based ones; the capacity model takes mispricing as given and characterizes how far and how fast it decays. Dong et al. [2024] build an equilibrium model where anomaly discovery reduces returns, but with a fixed number of arbitrageurs the surviving alpha depends on raw alpha. Endogenizing entry changes the result: the surviving alpha is pinned by the zero-profit condition. The empirical slope of 0.29–0.41 (Section 4.1) is closer to the free-entry prediction than the fixed-N prediction. Pontiff [2006] shows that idiosyncratic risk sustains anomalies; the present model microfounds this observation by showing that risk aversion enters the zero-profit condition and determines the friction floor, explaining *why* risk sustains anomalies in equilibrium rather than just documenting the correlation. Stambaugh et al. [2012] study how sentiment interacts with arbitrage costs; Brogaard et al. [2024] document that anomalies with higher arbitrage capital decay faster.

The free-entry logic follows the standard framework of Mankiw and Whinston [1986]. Zigrand [2004] studies Cournot-Walras equilibria with endogenous arbitrageur entry and sunk costs in segmented markets; Zigrand proves existence of equilibrium and characterizes welfare effects of lower entry costs. The mathematical structure is shared; the present paper derives alpha-independence, the characterization theorem, and dynamic half-life predictions, none of which appear in Zigrand.

Empirical tests using the Chen and Zimmermann [2024] anomaly zoo document substantial alpha compression: after correcting for errors-in-variables bias, a one-percentage-point increase in pre-publication alpha predicts a 0.29–0.41-percentage-point increase in post-publication alpha. Return volatility is the dominant friction ($t = 4.04$). The half-life test finds that raw alpha does not predict the speed of anomaly decay ($t = -1.59$), consistent with the theory.

Section 2 presents the model. Section 3 derives the main results. Section 4 reports empirical evidence, predictions, and limitations. Section 5 concludes.

2 Model

2.1 Environment

A universe of J anomalies exists, indexed by $j = 1, \ldots, J$. Each anomaly j is characterized by a triple $(\alpha_j, \lambda_j, \sigma_j)$:

- $\alpha_j > 0$: the raw anomaly alpha (expected return of the long-short portfolio absent any arbitrageur trading).
- $\lambda_j > 0$: the price impact parameter (Kyle, 1985). When aggregate arbitrageur volume is X, the realized alpha is reduced by $\lambda_j X$.
- $\sigma_j > 0$: the return volatility of the anomaly portfolio.

A large pool of potential arbitrageurs can enter any anomaly. Each arbitrageur faces a common fixed entry cost $c > 0$ (data subscriptions, shorting arrangements, compliance infrastructure). Section 3.6 relaxes the common-cost assumption.

The game has two stages for each anomaly:

1. **Entry.** Arbitrageurs simultaneously decide whether to enter anomaly j, paying cost c. Let N_j denote the number of entrants.
2. **Trading.** The N_j entrants simultaneously choose position sizes $x_i \geq 0$ in a Cournot trading game. The realized alpha after trading is $\alpha_j - \lambda_j \sum_{i=1}^{N_j} x_i$.

The solution proceeds by backward induction: first the trading game, then the entry game.

Three assumptions define the baseline model. First, the realized alpha is $\alpha_j - \lambda_j X_j$ where $X_j = \sum_i x_i$ (linear price impact). Second, each arbitrageur maximizes expected profit minus a risk penalty:

$$\pi_i = (\alpha_j - \lambda_j X_j)x_i - \frac{\gamma}{2}\sigma_j^2 x_i^2, \tag{1}$$

where $\gamma > 0$ is the common risk aversion parameter (mean-variance preferences). Third, arbitrageurs enter until expected profit no longer exceeds the entry cost c (free entry). CARA preferences with Gaussian returns generate the same objective: maximizing $\mathbb{E}[-e^{-\gamma W}]$ over terminal wealth $W = (\alpha - \lambda X)x + \tilde{\varepsilon}x$ with $\tilde{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$ yields the certainty equivalent

$(\alpha - \lambda X)x - (\gamma/2)\sigma^2 x^2$, which is equation (1).

Section 3.3 relaxes the linear-impact and mean-variance assumptions and characterizes exactly when the baseline results hold and when they fail.

2.2 The Trading Game

Fix an anomaly with parameters $(\alpha, \lambda, \sigma)$ and suppose $N \geq 1$ arbitrageurs have entered.

Lemma 1 (Symmetric Nash Equilibrium)**.** *The unique symmetric Nash equilibrium of the trading game has each entrant choosing position*

$$x^*(N) = \frac{\alpha}{(N+1)\lambda + \gamma\sigma^2}. \tag{2}$$

Proof. See Appendix A. □

Corollary 2. *The aggregate position, post-trading alpha, and individual profit are:*

$$X^*(N) = \frac{N\alpha}{(N+1)\lambda + \gamma\sigma^2}, \tag{3}$$

$$\alpha^{\text{post}}(N) = \alpha \cdot \frac{\lambda + \gamma\sigma^2}{(N+1)\lambda + \gamma\sigma^2}, \tag{4}$$

$$\pi^*(N) = \frac{\alpha^2\kappa}{\left[(N+1)\lambda + \gamma\sigma^2\right]^2}, \tag{5}$$

where $\kappa \equiv \lambda + \gamma\sigma^2/2$ is the effective friction parameter (price impact plus half the risk penalty).[1]

Equation (5) has the structure $\pi^*(N, \alpha) = \alpha^2 \cdot f(N)$, where $f(N) = \kappa/[(N+1)\lambda + \gamma\sigma^2]^2$ is strictly decreasing in N. This quadratic dependence on α is the mathematical engine of the alpha-independence result. Section 3.3 characterizes exactly when this structure holds and when it fails.

[1] The symbol κ is chosen for its mnemonic value ("cost of capital deployment"). It should not be confused with the mean-reversion speed, which does not appear in this model.

2.3 The Entry Equilibrium

Definition 3. *A* free-entry Cournot equilibrium *for anomaly* $(\alpha, \lambda, \sigma)$ *with entry cost* c *is a pair* $(N^*, \{x_i^*\})$ *such that:*

(a) Each active arbitrageur plays the symmetric Cournot best response in the trading game: $x_i^* = x^*(N^*)$ *from Lemma 1.*

(b) N^* *is the largest integer* N *such that* $\pi^*(N) \geq c$*.*

Lemma 4 (Entry Threshold)**.** *Treating* N *as continuous for tractability, the equilibrium number of entrants is*

$$\bar{N} = \frac{1}{\lambda}\left(\alpha\sqrt{\frac{\kappa}{c}} - \gamma\sigma^2 - \lambda\right). \tag{6}$$

Entry is positive ($\bar{N} > 0$) *if and only if* $\alpha > \alpha^{\min}$*, where*

$$\alpha^{\min} \equiv (\gamma\sigma^2 + \lambda)\sqrt{\frac{c}{\kappa}}. \tag{7}$$

Note that $\alpha^{\min}$ *equals the surviving alpha* α^{post} *from Proposition 6: the minimum alpha required for entry equals the post-entry equilibrium alpha. Any anomaly that attracts at least one entrant is driven down to this floor.*

Proof. See Appendix A. □

Remark 1 (Continuous approximation)*.* The continuous-N approximation is used throughout for tractability. For $\bar{N} \geq 3$, the approximation error in α^{post} is bounded by $\lambda/[(N^*+1)\lambda + \gamma\sigma^2]$, which is less than 5% of α^{post} at baseline calibration. For $N = 1$ (a monopolist), the surviving alpha $\alpha^{\text{post}} = \alpha(\lambda + \gamma\sigma^2)/(2\lambda + \gamma\sigma^2)$ depends explicitly on α; alpha-independence is a competitive (large-N) result. Appendix D solves the discrete-N case for $N = 1, 2, 3$.

3 Results

3.1 The Surviving Alpha

Definition 5 (Alpha-Independence)**.** *An entry game exhibits* alpha-independence *if the post-entry surviving alpha* α^{post} *is independent of the raw alpha* α*:* $\partial\alpha^{\text{post}}/\partial\alpha = 0$.

Proposition 6 (Post-Entry Surviving Alpha)**.** *In the free-entry equilibrium with continuous* N*, the surviving alpha level is*

$$\alpha^{\text{post}} = (\lambda + \gamma\sigma^2)\sqrt{\frac{c}{\kappa}} = \sqrt{c} \cdot \frac{\lambda + \gamma\sigma^2}{\sqrt{\lambda + \gamma\sigma^2/2}}, \tag{8}$$

which is independent of the raw alpha α*. The survival rate is*

$$S \equiv \frac{\alpha^{\text{post}}}{\alpha} = \frac{(\lambda + \gamma\sigma^2)\sqrt{c/\kappa}}{\alpha}. \tag{9}$$

Proof. From Corollary 2, $\alpha^{\text{post}}(\bar{N}) = \alpha \cdot (\lambda + \gamma\sigma^2)/[(\bar{N}+1)\lambda + \gamma\sigma^2]$. The zero-profit condition (proof of Lemma 4) gives $(\bar{N}+1)\lambda + \gamma\sigma^2 = \alpha\sqrt{\kappa/c}$. Substituting:

$$\alpha^{\text{post}} = \alpha \cdot \frac{\lambda + \gamma\sigma^2}{\alpha\sqrt{\kappa/c}} = (\lambda + \gamma\sigma^2)\sqrt{\frac{c}{\kappa}}.$$

□

Equation (8) is the central result. All anomalies with the same price impact, volatility, and entry cost converge to the same surviving alpha level, regardless of how large the original anomaly was. The raw alpha determines how many arbitrageurs enter (equation (6)), not how much alpha survives. Figure 1 illustrates this result across three market segments.

Proposition 7 (Non-Monotonicity of Survival Rates)**.** *The survival rate* S *is strictly decreasing in raw alpha:*

$$\frac{\partial S}{\partial \alpha} = -\frac{(\lambda + \gamma\sigma^2)\sqrt{c/\kappa}}{\alpha^2} < 0. \tag{10}$$

Larger anomalies have lower survival rates.

Proof. The survival rate $S = (\lambda + \gamma\sigma^2)\sqrt{c/\kappa}/\alpha$ is proportional to $1/\alpha$. □

A large alpha is like a commons with high initial value. Many arbitrageurs enter, and congestion is severe. A small alpha attracts few entrants, and those who enter face less competition. The most "obvious" anomalies (large, liquid, well-known) are the most vulnerable to entry, while "obscure" ones (small, illiquid, complex) are the most protected.

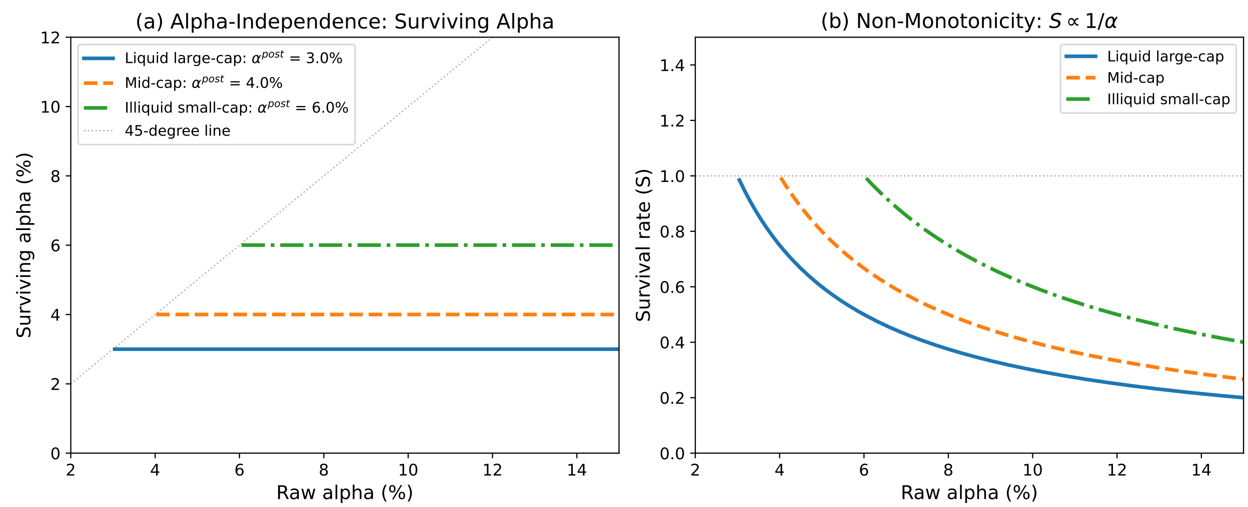

Figure 1: **Alpha-independence and non-monotonicity of survival rates.** Panel (a) shows the surviving alpha level α^{post} as a function of raw alpha α for three market segments (Table 8). For each segment, α^{post} is constant in α, while different friction environments produce different floors. Panel (b) shows the survival rate $S = \alpha^{\text{post}}/\alpha$, which is strictly decreasing in raw alpha ($S \propto 1/\alpha$). Larger anomalies have lower survival rates because free entry absorbs the extra alpha.

Proposition 8 (Capacity Characterization)**.** *Define the capacity of an anomaly as $K \equiv \alpha/\lambda$. For large capacity, the number of entrants is approximately proportional to K:*

$$\bar{N} \approx K\sqrt{\frac{\kappa}{c}} \quad \textit{for } K \textit{ large.}$$

The survival rate depends on α and λ separately, not only on capacity. For fixed α:

$$\frac{\partial S}{\partial \lambda} > 0. \tag{11}$$

Higher price impact (lower liquidity) increases the survival rate.

Proof. See Appendix A. □

3.2 Comparative Statics

Proposition 9 (Comparative Statics)**.** *The survival rate S and surviving alpha α^{post} respond to parameters as follows:*

Parameter ↑	$\partial\alpha^{\text{post}}/\partial(\cdot)$	$\partial S/\partial(\cdot)$	*Economic force*
Raw alpha α	0	$-$	*Free entry absorbs extra alpha*
Entry cost c	$+$	$+$	*Higher barriers protect the anomaly*
Price impact λ	$+$	$+$	*Trading is costlier per unit*
Volatility σ	$+$	$+$	*Risk deters entry*
Risk aversion γ	$+$	$+$	*Cautious arbitrageurs trade less*

Proof. See Appendix A. □

All friction parameters protect the anomaly by deterring entry or reducing trading intensity. The asymmetry between α (zero effect on the level) and all other parameters (positive effect) is the structural insight: raw alpha is completely absorbed by entry, while frictions set the floor.

3.3 General Characterization

The alpha-independence result in Section 3.1 is not an artifact of the specific linear-impact Cournot game. It holds whenever the equilibrium profit function has a particular structure. The following theorem characterizes the exact condition.

Definition 10 (Abstract Entry Game)**.** *An* entry game *consists of a parameter $\alpha > 0$ (raw alpha), a vector of market frictions θ, a sunk entry cost $c > 0$, and a post-entry trading game whose symmetric equilibrium yields individual profit $\pi^*(N, \alpha, \theta)$, strictly decreasing in N.*

The free-entry equilibrium number of entrants $\bar{N}(\alpha, \theta, c)$ solves $\pi^*(\bar{N}, \alpha, \theta) = c$. The post-entry surviving alpha takes the form $\alpha^{\text{post}}(N, \alpha, \theta) = \alpha \cdot h(N, \theta)$, where h is strictly decreasing in N.

Theorem 11 (General Characterization of Alpha-Independence)**.** *Consider an entry game as in Definition 10. Suppose the profit function is multiplicatively separable: $\pi^*(N, \alpha, \theta) = g(\alpha) \cdot f(N, \theta)$ for some strictly increasing g and strictly positive, strictly decreasing f. Then the surviving alpha α^{post} is independent of α in the free-entry equilibrium if and only if:*

(i) $g(\alpha) = B\alpha^p$ for some $p > 0$ and $B > 0$, and

(ii) $h(N, \theta) = A(\theta) \cdot [f(N, \theta)]^{1/p}$ for some function $A(\theta)$ independent of N.

Under these conditions, $\alpha^{\text{post}} = A(\theta) \cdot c^{1/p}$, independent of α.

In the baseline linear-impact mean-variance game, both conditions hold with $p = 2$, $f(N) = \kappa/[(N + 1)\lambda + \gamma\sigma^2]^2$, $h(N) = (\lambda + \gamma\sigma^2)/[(N + 1)\lambda + \gamma\sigma^2]$, and $A = (\lambda + \gamma\sigma^2)/\sqrt{\kappa}$.

Proof. **Sufficiency.** Suppose conditions (i) and (ii) hold. The free-entry condition $B\alpha^p f(\bar{N}) = c$ gives $f(\bar{N}) = c/(B\alpha^p)$. The surviving alpha is

$$\alpha^{\text{post}} = \alpha \cdot h(\bar{N}) = \alpha \cdot A \cdot [f(\bar{N})]^{1/p} = \alpha \cdot A \cdot \left[\frac{c}{B\alpha^p}\right]^{1/p} = \alpha \cdot A \cdot \frac{c^{1/p}}{B^{1/p}\alpha} = \frac{A \cdot c^{1/p}}{B^{1/p}},$$

which is independent of α.

Necessity. Suppose $\pi^* = g(\alpha)f(N)$ and $\alpha^{\text{post}} = \alpha \cdot h(\bar{N}(\alpha))$ is independent of α. The free-entry condition $g(\alpha)f(\bar{N}) = c$ gives, upon total differentiation:

$$g'(\alpha)f(\bar{N}) + g(\alpha)f'(\bar{N})\frac{d\bar{N}}{d\alpha} = 0 \quad \Longrightarrow \quad \frac{d\bar{N}}{d\alpha} = -\frac{g'(\alpha)}{g(\alpha)} \cdot \frac{f(\bar{N})}{f'(\bar{N})}.$$

Alpha-independence requires $d[\alpha \cdot h(\bar{N}(\alpha))]/d\alpha = 0$:

$$h(\bar{N}) + \alpha\, h'(\bar{N})\frac{d\bar{N}}{d\alpha} = 0 \quad \Longrightarrow \quad h(\bar{N}) = \alpha\, h'(\bar{N}) \cdot \frac{g'(\alpha)}{g(\alpha)} \cdot \frac{f(\bar{N})}{f'(\bar{N})}.$$

Rearranging:

$$\frac{h'(\bar{N})\, f(\bar{N})}{h(\bar{N})\, f'(\bar{N})} = \frac{g(\alpha)}{\alpha\, g'(\alpha)}. \tag{12}$$

The left side depends only on $\bar{N}$ (a function of α); the right side depends only on α. Since $\bar{N}$ varies with α, both sides must equal a common constant $1/p$ for some $p > 0$. From the right side: $g(\alpha)/[\alpha\, g'(\alpha)] = 1/p$, which gives $g'(\alpha)/g(\alpha) = p/\alpha$ and hence $g(\alpha) = B\alpha^p$. From the left side: $h'f/(hf') = 1/p$, equivalently $h'/h = f'/(pf)$. Integrating: $\ln h = (1/p) \ln f + \text{const}$, so $h(N) = A \cdot [f(N)]^{1/p}$ for some constant $A > 0$. □

Remark 2. The value $p = 2$ is not a general requirement. It arises from the linear-impact mean-variance trading game. Other microstructure specifications produce different values of p. For example, square-root impact with risk-neutral traders gives $p = 3$ (Proposition 12). The characterization identifies the structural condition (power-law separable profit and proportional surviving fraction) regardless of p.

Remark 3 (Economic content of the conditions). Condition (i) ($g(\alpha) = B\alpha^p$) requires that profits scale as a power law in alpha. This holds in standard Cournot models with linear or constant-elasticity demand. It fails when the trading technology introduces non-power-law dependence on alpha: for example, if arbitrageurs face a capacity constraint that binds differently at different alpha levels (e.g., $\pi = \min\{\alpha^2 f(N), \bar{\pi}\}$ where $\bar{\pi}$ is a regulatory cap), or if the price impact function depends on alpha itself (e.g., $\lambda(\alpha)$ with a non-trivial alpha elasticity). In such environments, the profit function is not multiplicatively separable in alpha, and the free-entry condition produces alpha-dependent surviving alpha even under linear impact.

Condition (ii) requires that the surviving-alpha fraction h and the profit function f move in lockstep as N varies. Economically, this means additional entry erodes alpha and profits at proportional rates. Under linear impact, both erosion rates are linear in N: adding one entrant reduces alpha by a fixed fraction and reduces profits by a proportional amount. Under concave impact, additional entry erodes alpha faster than profits (because the impact function flattens at high X, so profits per unit are less sensitive to further entry than alpha is), breaking the proportionality.

Table 1 summarizes which trading environments satisfy the conditions of Theorem 11.

Table 1: Alpha-Independence across Market Microstructure Specifications

Model	p	Condition (ii)?	Alpha-independence
Linear impact + mean-variance	2	Yes	Exact
Linear impact + CARA, exogenous wealth	2	Yes	Exact
Square-root impact, risk-neutral	3	No (A depends on N)	Fails
CRRA + exogenous wealth	2	Approx. (2nd-order)†	Approximate
CRRA + endogenous wealth	—	No (wealth channel)	Fails

† Under CRRA with exogenous background wealth W_0, a second-order Taylor expansion of the certainty equivalent around the no-trade point yields the mean-variance objective (1) with $\gamma = \gamma_{\text{CRRA}}/W_0$ and profit $\pi \propto \alpha^2$, so condition (i) holds with $p = 2$. The approximation error is $O(x^3\alpha^3/W_0^3)$, negligible for $x\alpha/W_0 \ll 1$, which holds for typical anomaly portfolios sized at 1–5% of wealth. Condition (ii) holds to the same order, giving approximate alpha-independence.

Proposition 12 (Alpha-Dependence under Square-Root Impact)**.** *Under square-root price impact ($\alpha - \lambda\sqrt{X}$) with risk-neutral Cournot traders ($\gamma = 0$), the equilibrium profit is*

$$\pi^*(N, \alpha) = \frac{4N\alpha^3}{\lambda^2(2N+1)^3}, \tag{13}$$

so $p = 3$. The surviving alpha fraction is $h(N) = 1/(2N+1)$, and

$$\frac{h(N)}{[f(N)]^{1/3}} = \frac{\lambda^{2/3}}{(4N)^{1/3}},$$

which depends on N. Condition (ii) of Theorem 11 fails. For large $\bar{N}$, the surviving alpha is

$$\alpha^{\text{post}} \approx \lambda\sqrt{\frac{2c}{\alpha}}, \tag{14}$$

which is decreasing in α. Larger anomalies have lower surviving alpha levels, not just lower survival rates.

Proof. See Appendix B. □

Corollary 13 (Alpha Over-Dissipation under Square-Root Impact)**.** *Under square-root impact, the survival rate is*

$$S_{sqrt} \equiv \frac{\alpha^{\text{post}}}{\alpha} \approx \frac{\lambda\sqrt{2c}}{\alpha^{3/2}}, \tag{15}$$

which declines faster in α than the linear-impact survival rate ($S \propto 1/\alpha$). The surviving alpha level itself is decreasing in raw alpha: larger anomalies not only lose a greater fraction of their alpha (as in the linear case) but also converge to a lower absolute level. The economic force is concavity of the impact function: with square-root impact, the marginal entrant displaces less alpha per unit of position, so more entrants are required to reach zero profit, and the additional entry overshoots the alpha-independence benchmark.

The linear and square-root cases bracket the empirically relevant range. Under linear impact, alpha is fully insulated from raw magnitude (exact alpha-independence). Under square-root impact, larger anomalies are over-dissipated. The quantitative magnitude is moderate: for two anomalies with $\alpha_A = 2\alpha_B$ and the same frictions, the linear case gives $\alpha_A^{\text{post}} = \alpha_B^{\text{post}}$, while the square-root case gives $\alpha_A^{\text{post}}/\alpha_B^{\text{post}} \approx (1/2)^{1/2} \approx 0.71$. The empirically preferred concave impact specification [Almgren et al., 2005, Bouchaud et al., 2009, Kyle and Obizhaeva, 2016] thus predicts that the largest anomalies should decay slightly below the friction-determined floor, creating a testable departure from exact alpha-independence. Section 4.1 tests this prediction.

Theorem 14 (Approximate Genericity of Alpha-Independence)**.** *Consider the entry game of Definition 10 with profit $\pi_i = (\alpha - \phi(X))x_i - (\gamma/2)\sigma^2 x_i^2$, where ϕ is C^2, strictly increasing, and weakly concave with $\phi(0) = 0$ and $\phi'(X) > 0$ for all $X \geq 0$. Let $\kappa = \gamma\sigma^2$. In the free-entry equilibrium with continuous N and positive entry ($\bar{N} > 0$), the surviving alpha satisfies*

$$\alpha^{\text{post}} = (\phi'(Nx^*) + \kappa)\sqrt{\frac{c}{\phi'(Nx^*) + \kappa/2}}, \tag{16}$$

which depends on α only through $\phi'(Nx^*)$. Define

$$\eta(\alpha) \equiv \frac{\sqrt{c}\,\phi''(\bar{X})}{2\big(\phi'(\bar{X}) + \kappa/2\big)^{3/2}}, \tag{17}$$

where $\bar{X} = \bar{N}x^*$ is the equilibrium aggregate position. For κ sufficiently large that $|\eta| < 1/2$, the sensitivity to raw alpha is bounded:

$$\left|\frac{d\alpha^{\text{post}}}{d\alpha}\right| = \frac{|\eta|}{1+\eta} \leq 2|\eta| = O(\kappa^{-3/2}). \tag{18}$$

The departure from alpha-independence vanishes at rate $(\gamma\sigma^2)^{-3/2}$ as risk aversion increases.

Proof. The symmetric FOC is $\alpha - \phi(Nx) - x\phi'(Nx) - \kappa x = 0$, giving $\alpha - \phi(Nx) = x(\phi'(Nx) + \kappa)$. Substituting into the profit:

$$\pi = (\alpha - \phi(Nx))x - \frac{\kappa}{2}x^2 = x^2(\phi'(Nx) + \kappa) - \frac{\kappa}{2}x^2 = x^2\big(\phi'(Nx) + \kappa/2\big).$$

The free-entry condition $\pi = c$ gives $x = \sqrt{c/(\phi'(Nx) + \kappa/2)}$. Since $\alpha^{\text{post}} = \alpha - \phi(Nx) = x(\phi'(Nx) + \kappa)$:

$$\alpha^{\text{post}} = (\phi'(Nx) + \kappa)\sqrt{\frac{c}{\phi'(Nx) + \kappa/2}}.$$

This depends on α only through $\lambda \equiv \phi'(Nx)$. Define $g(\lambda) = (\lambda + \kappa)\sqrt{c/(\lambda + \kappa/2)}$. Then:

$$g'(\lambda) = \frac{\lambda\sqrt{c}}{2(\lambda + \kappa/2)^{3/2}}.$$

From $\phi(\bar{X}) = \alpha - \alpha^{\text{post}}$, differentiating: $\phi'(\bar{X})\frac{d\bar{X}}{d\alpha} = 1 - \frac{d\alpha^{\text{post}}}{d\alpha}$. Hence:

$$\frac{d\alpha^{\text{post}}}{d\alpha} = g'(\lambda) \cdot \phi''(\bar{X}) \cdot \frac{d\bar{X}}{d\alpha} = g'(\lambda) \cdot \frac{\phi''(\bar{X})}{\phi'(\bar{X})} \cdot \left(1 - \frac{d\alpha^{\text{post}}}{d\alpha}\right).$$

Solving: $d\alpha^{\text{post}}/d\alpha = \eta/(1+\eta)$ where $\eta = g'(\lambda)\phi''(\bar{X})/\phi'(\bar{X})$ as in (17). Since ϕ is concave ($\phi'' \leq 0$), $\eta \leq 0$ and $|d\alpha^{\text{post}}/d\alpha| = |\eta|/(1 - |\eta|)$. For $|\eta| \leq 1/2$ (which holds for κ sufficiently large since $g'(\lambda) = O(\kappa^{-3/2})$), the bound $|\eta|/(1-|\eta|) \leq 2|\eta|$ gives (18). □

Theorem 14 is the paper's main result. It shows that alpha-independence is not an artifact of linear price impact: for any concave impact function, the surviving alpha's dependence on raw alpha vanishes as risk aversion increases, at the rate $(\gamma\sigma^2)^{-3/2}$. The characterization theorem (Theorem 11) identifies the boundary between exact and approximate alpha-independence; Theorem 14 shows that the boundary has negligible quantitative bite once arbitrageurs bear risk.

Table 2 evaluates the bound at calibration parameters. Under linear impact, the bound is exactly zero (confirming exact alpha-independence). Under square-root impact with $\gamma = 2$, the maximum variation in α^{post} across $\alpha \in [3\%, 15\%]$ is less than 0.15 basis points in all segments. The bound is not merely an asymptotic statement: it is quantitatively negligible at empirically relevant risk aversion.

Table 2: Genericity Bound at Calibration

Segment	Impact	γ	$\lvert 2\eta \rvert$	Variation (bps)
Large-cap	Linear	2	0 (exact)	0
Mid-cap	Linear	2	0 (exact)	0
Large-cap	Square-root	2	$< 10^{-11}$	< 0.01
Mid-cap	Square-root	2	$< 10^{-7}$	0.01
Small-cap	Square-root	2	2×10^{-5}	0.14
Mid-cap	Square-root	0.1	$< 10^{-6}$	< 0.01

Notes: $|2\eta|$ is the upper bound on $|d\alpha^{\text{post}}/d\alpha|$ from Theorem 14. "Variation" is the actual numerical variation in α^{post} across $\alpha \in [3\%, 15\%]$. Calibration: $c = 0.01$; large-cap ($\lambda = 10^{-4}$, $\sigma = 15\%$), mid-cap ($\lambda = 10^{-3}$, $\sigma = 20\%$), small-cap ($\lambda = 5 \times 10^{-3}$, $\sigma = 30\%$).

Figure 2 illustrates the convergence visually.

Remark 4 (Where the characterization theorem has bite). The bound (18) is tight when κ is small relative to ϕ', i.e., when price impact dominates the risk penalty. This regime describes high-frequency trading and delta-hedged statistical arbitrage, where effective risk aversion is low and the impact specification matters. For long-horizon value strategies with high effective risk aversion, alpha-independence holds approximately regardless of the impact function. The characterization theorem governs the former regime; Theorem 14 governs the latter.

The empirical rejection of exact alpha-independence (Section 4.1, EIV-corrected slope 0.29–0.41) is consistent with multiple channels: the integer constraint (Appendix D), heterogeneous entry costs, the CRRA approximation error, or regression to the mean from pre-publication overfitting. The data cannot distinguish these channels with the available proxies.

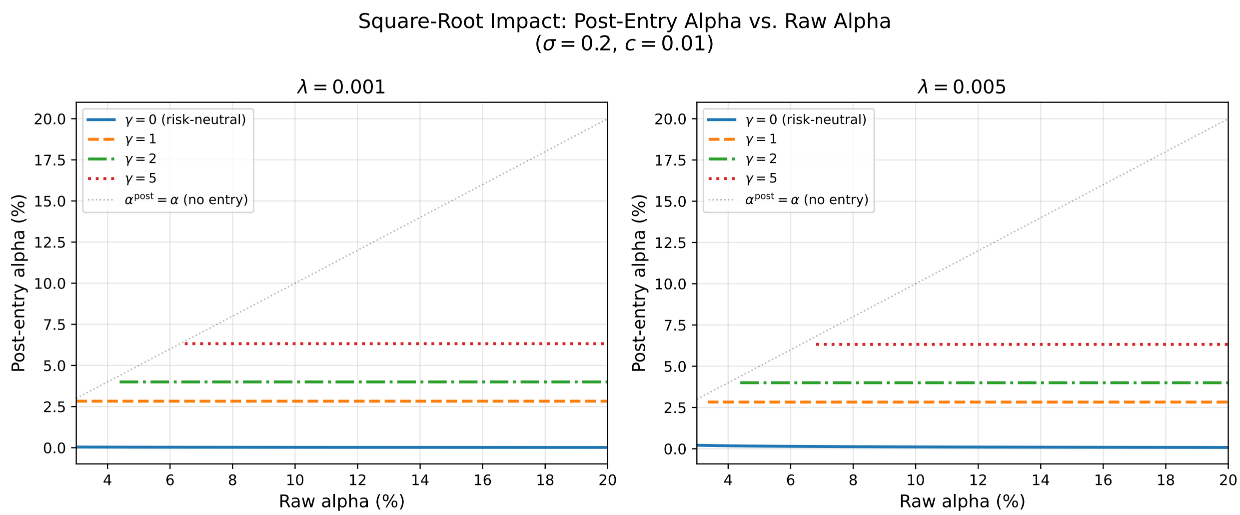

Figure 2: **Approximate alpha-independence under square-root impact with risk aversion.** Post-entry surviving alpha vs. raw alpha under square-root price impact for $\gamma = 0$ (risk-neutral), 1, 2, and 5. Theorem 14 predicts that the curves flatten as γ increases, at rate $O((\gamma\sigma^2)^{-3/2})$. For $\gamma \geq 2$, the variation across $\alpha \in [3\%, 20\%]$ is less than 0.1 percentage points.

3.4 General Equilibrium across Anomalies

The baseline model treats each anomaly independently: arbitrageurs enter anomaly j without regard to opportunities elsewhere. In practice, arbitrage capital is finite. This section introduces a shared capital pool and shows that alpha-independence breaks through a general-equilibrium channel.

Proposition 15 (Alpha-Independence Breaks under Shared Capital)**.** *Consider two anomalies A and B with identical frictions (λ, σ, c) but different raw alphas, sharing a pool of M potential arbitrageurs. Each arbitrageur enters at most one anomaly. When the capital constraint binds ($\bar{N}_A^{PE} + \bar{N}_B^{PE} > M$), the surviving alpha in anomaly A is*

$$\alpha_A^{\text{post}} = \alpha_B^{\text{post}} = \frac{(\lambda + \gamma\sigma^2)(\alpha_A + \alpha_B)}{\lambda M + 2(\lambda + \gamma\sigma^2)}, \tag{19}$$

which depends on both α_A and α_B. The entry allocation is

$$N_A = \frac{M\alpha_A}{\alpha_A + \alpha_B} + R \cdot \frac{\alpha_A - \alpha_B}{\alpha_A + \alpha_B}, \tag{20}$$

where $R = (\lambda + \gamma\sigma^2)/\lambda$. Alpha-independence fails: $\partial\alpha_A^{\text{post}}/\partial\alpha_A > 0$, and larger anomalies retain more alpha because the finite pool cannot supply enough entrants to dissipate the surplus.

Proof. In a constrained interior equilibrium, $N_A + N_B = M$ and equal-profit entry requires $\pi_A^*(N_A) = \pi_B^*(N_B)$. From Corollary 2, $\pi_j = \alpha_j^2\kappa/[(N_j + 1)\lambda + \gamma\sigma^2]^2$. Setting profits equal: $\alpha_A/[(N_A + 1)\lambda + \gamma\sigma^2] = \alpha_B/[(M - N_A + 1)\lambda + \gamma\sigma^2]$. Cross-multiplying and solving for N_A gives (20). Substituting into $\alpha_A^{\text{post}} = \alpha_A(\lambda + \gamma\sigma^2)/[(N_A + 1)\lambda + \gamma\sigma^2]$ and simplifying yields (19). □

Remark 5 (Convergence to partial equilibrium)*.* As $M \to \infty$, the capital constraint ceases to bind ($\bar{N}_A^{PE} + \bar{N}_B^{PE} < M$), and the model reduces to two independent partial-equilibrium problems. The threshold is $M^* = \bar{N}_A^{PE} + \bar{N}_B^{PE}$. For $M > M^*$, exact alpha-independence holds. For $M < M^*$, the GE channel makes surviving alpha increasing in raw alpha: larger anomalies are under-arbitraged relative to the PE benchmark because capital is diverted to the competing opportunity. Figure 3 illustrates. At baseline calibration with $\alpha_A = 10\%$ and $\alpha_B = 5\%$, the GE effect raises α_A^{post} from 4.0% (PE) to 4.6% ($M = 100$) — a 16% increase.

The GE result identifies a new force absent from the partial-equilibrium analysis: cross-

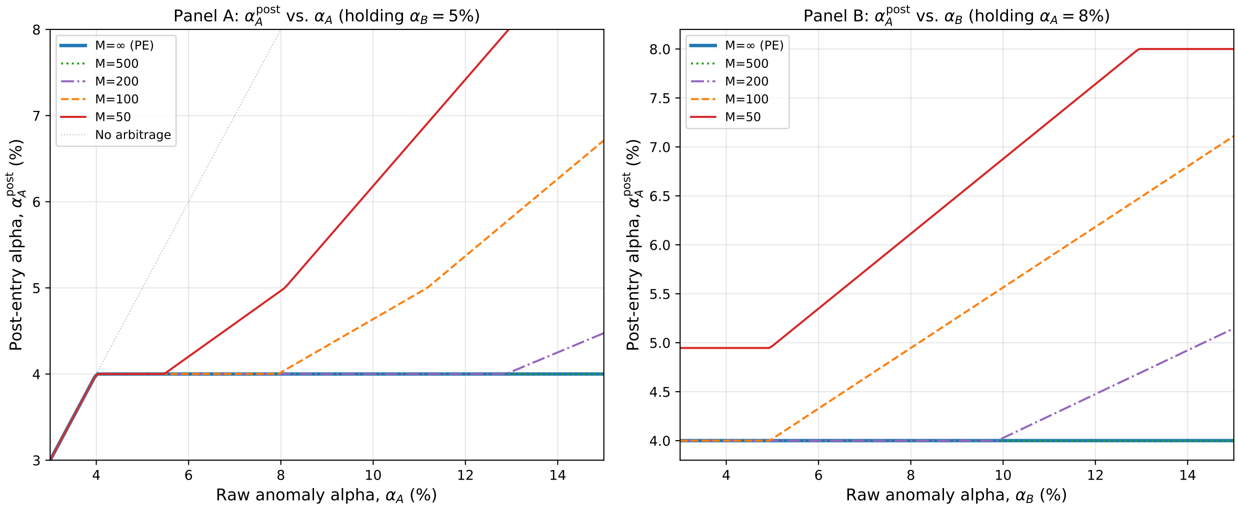

Figure 3: **Alpha-independence breaks under shared capital.** Panel (a): α_A^{post} vs. α_A holding $\alpha_B = 5\%$ fixed, for different pool sizes M. With $M = \infty$ (partial equilibrium), α_A^{post} is flat (alpha-independence). With finite M, α_A^{post} increases in α_A: the capital constraint prevents full dissipation. Panel (b): α_A^{post} vs. α_B holding $\alpha_A = 8\%$ fixed. A more attractive competing anomaly draws arbitrageurs away, raising α_A^{post} (cross-anomaly contagion).

anomaly contagion. When a new anomaly is published, capital flows toward it, reducing entry into existing anomalies and raising their surviving alpha. This effect is strongest for small capital pools and large alpha differences. The partial-equilibrium alpha-independence result is the limiting case when capital is sufficiently abundant ($M > M^*$) that entry into one anomaly does not crowd out entry into others.

3.5 Dynamic Extension

The static model determines the level of surviving alpha but not the speed of convergence. To generate half-life predictions, the entry-trading game is embedded in continuous time with Cournot competition throughout, maintaining internal consistency with the static model.

Time is continuous, $t \in [0, \infty)$. At each instant, $N(t)$ active arbitrageurs play the instantaneous Cournot game of Lemma 1. Trading erodes alpha at rate $\delta > 0$:

$$\dot{\alpha}(t) = -\delta \cdot X^*(N(t)) = -\delta \cdot \frac{N(t)\alpha(t)}{(N(t)+1)\lambda + \gamma\sigma^2}. \tag{21}$$

Arbitrageurs enter or exit so that the instantaneous Cournot profit flow equals the amortized entry cost:

$$\pi^*(N(t), \alpha(t)) = rc, \tag{22}$$

where $r > 0$ is the interest rate.

Remark 6 (Myopic entry and its justification)*.* The entry condition (22) is myopic: arbitrageurs compare the current profit flow to the amortized cost rc without forecasting the declining alpha path. Two features of the anomaly-arbitrage setting make this the natural benchmark. First, many anomaly strategies are implemented by quantitative funds that rebalance at fixed intervals and evaluate performance against hurdle rates, comparing current alpha to the cost of capital rather than solving a dynamic programming problem over the strategy's lifecycle. Second, with costless exit (which the continuous-time formulation assumes), myopic and forward-looking entry coincide: an arbitrageur who can leave the moment profits fall below rc has no option value from waiting, so the myopic condition is not an approximation but the correct optimality condition.

With irreversible entry costs, forward-looking arbitrageurs would be more cautious (the entry threshold rises by the option value of delay). The steady-state alpha increases, the convergence speed decreases, and the half-life formula becomes $T_{1/2} = \lambda \ln 2/(\delta - \text{option correction})$, which is longer than the myopic prediction. The qualitative structure (exponential convergence to a friction-determined floor, with the decay rate proportional to δ/λ) survives, but the quantitative half-life is a lower bound on the true half-life under irreversible entry. The cross-sectional prediction (illiquidity predicts speed, raw alpha does not) is unaffected because the option value depends on frictions, not on alpha.

Proposition 16 (Exponential Convergence)**.** *Under the dynamic free-entry equilibrium (equations* (21)*–*(22)*), alpha evolves as*

$$\alpha(t) = \alpha^{\text{ss}} + (\alpha_0 - \alpha^{\text{ss}})e^{-\phi t}, \tag{23}$$

where the decay rate and steady-state alpha are

$$\phi = \frac{\delta}{\lambda}, \qquad \alpha^{\text{ss}} = (\lambda + \gamma\sigma^2)\sqrt{\frac{rc}{\kappa}}. \tag{24}$$

The half-life is

$$T_{1/2} = \frac{\lambda \ln 2}{\delta}. \tag{25}$$

Proof. See Appendix A. □

The half-life depends only on λ/δ. Free entry acts as a regulator: if α_0 is larger, more arbitrageurs enter, and the extra entry exactly offsets the larger gap between α_0 and α^{ss}. If risk aversion is higher, fewer arbitrageurs enter, but the steady state is also higher, and the decay rate to this higher floor is unchanged. Only the trade-off between price impact (which slows trading) and alpha erosion (which responds to trading) determines the speed. Figure 4 illustrates the convergence for different initial alpha values.

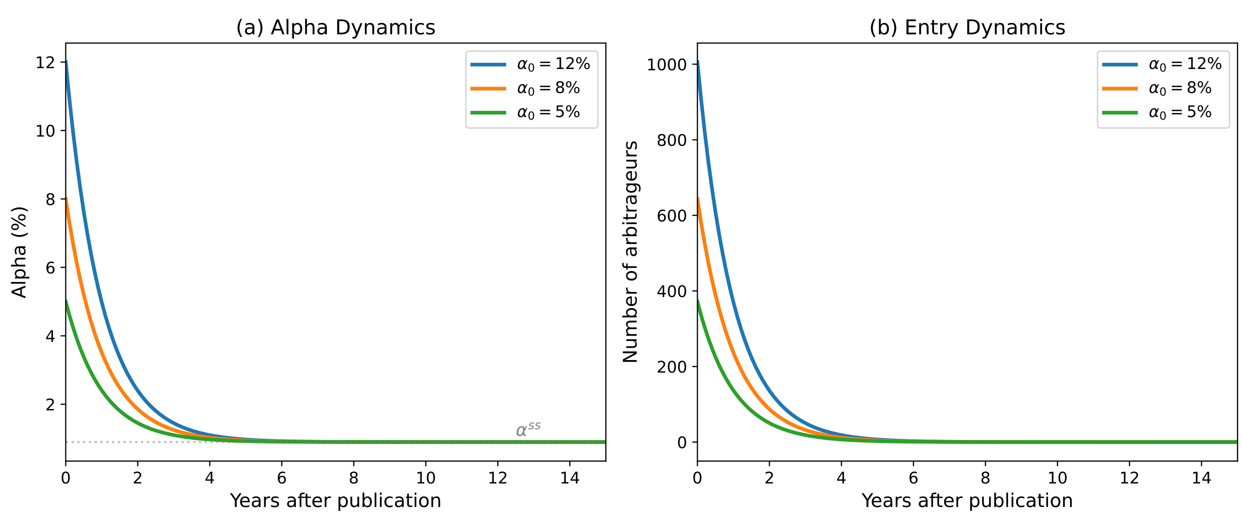

Figure 4: **Dynamic alpha convergence.** Panel (a) shows the alpha path $\alpha(t) = \alpha^{\text{ss}} + (\alpha_0 - \alpha^{\text{ss}})e^{-\phi t}$ for three initial alpha values. All paths converge to the same steady state α^{ss}, and the decay rate $\phi = \delta/\lambda$ is independent of α_0. Panel (b) shows the corresponding number of active arbitrageurs $N(t)$, which is initially larger for higher-alpha anomalies but converges as alpha converges. Parameters: mid-cap baseline ($\lambda = 0.001$, $\sigma = 20\%$, $\gamma = 2$, $c = 0.01$, $\delta = 0.001$, $r = 0.05$).

Remark 7 (Static-dynamic consistency). The dynamic steady state $\alpha^{\text{ss}} = (\lambda + \gamma\sigma^2)\sqrt{rc/\kappa}$ differs from the static surviving alpha $\alpha^{\text{post}} = (\lambda + \gamma\sigma^2)\sqrt{c/\kappa}$ by a factor of $\sqrt{r}$. The static model is a one-shot game: the arbitrageur pays c once and collects π^* once. The dynamic model has flow payoffs: the arbitrageur pays c once and collects π^* per unit time, so the effective cost is rc per period. For $r = 1$ (one-period model), the two coincide. The static and dynamic models answer different questions (cross-sectional level versus temporal speed), and the $\sqrt{r}$ factor reflects cost amortization, not an inconsistency.

Table 3 summarizes the dynamic comparative statics.

Table 3: Dynamic Comparative Statics

Parameter ↑	$\partial T_{1/2}/\partial(\cdot)$	$\partial\alpha^{\text{ss}}/\partial(\cdot)$	Intuition
Price impact λ	+	+	Costlier trading slows convergence
Erosion rate δ	−	0	Faster price correction
Entry cost c	0	+	Entry adjusts to maintain same ϕ
Volatility σ	0	+	Higher floor, same speed
Risk aversion γ	0	+	Higher floor, same speed
Raw alpha α_0	0	0	More entry absorbs extra gap

The sharpest prediction is the decomposition: price impact determines the speed of convergence, while risk, entry cost, and volatility determine only the floor. Raw alpha determines neither.

3.6 Heterogeneous Entry Costs

Potential arbitrageurs differ in their entry cost c_i, drawn from a distribution G with density g on $[\underline{c}, \bar{c}]$, where $\underline{c} > 0$.

Proposition 17 (Decelerating Alpha Decay)**.** *With heterogeneous entry costs, the alpha path is convex (decelerating):*

- *In the early phase, many arbitrageurs are active (all with $c_i \leq \hat{c}(0)$), causing rapid alpha decay.*

- *In the late phase, only low-cost arbitrageurs remain. The decay slows as alpha approaches its steady state, determined by $\underline{c}$, from above.*

The entry threshold $\hat{c}(t) = \pi^*(N(t), \alpha(t))/r$ is decreasing in t: as alpha decays, profits fall, and high-cost arbitrageurs exit.

Proof. See Appendix C. □

The decelerating decay path (rapid initial decay, slow late convergence) is consistent with the empirical evidence: rapid post-publication decay [McLean and Pontiff, 2016] followed by persistence of a residual alpha [Bowles et al., 2024].

4 Discussion

4.1 Empirical Evidence

The theory's predictions are testable using the Chen and Zimmermann [2024] anomaly zoo dataset, which provides portfolio returns for approximately 200 published anomalies with known publication dates. This section reports cross-sectional tests of alpha-independence, the non-monotonicity of survival rates, and the role of frictions in determining the level and speed of anomaly decay.

4.1.1 Data and Variable Construction

The sample begins with all anomalies in the Chen and Zimmermann [2024] dataset with original-paper portfolio returns. For each anomaly, pre-publication alpha is estimated from a Fama-French six-factor (MKT, SMB, HML, RMW, CMA, MOM) time-series regression over all months before the publication year, requiring at least 60 months. Post-publication alpha uses all months starting three years after publication (to allow for dissemination), requiring at least 36 months. The survival rate is $S_j = \hat{\alpha}_{j,\text{post}}/\hat{\alpha}_{j,\text{pre}}$, set to zero when

$\hat{\alpha}_{j,\text{post}} < 0$. All regressions use alphas winsorized at the 5th and 95th percentiles to limit the influence of outliers; results are qualitatively unchanged at the 1st/99th percentiles.

Three control variables proxy for market frictions. Illiquidity (ILLIQ_j) is the log of the inverse of the average number of stocks in the long and short legs of the anomaly portfolio, computed over the pre-publication period. Return volatility (σ_j) is the standard deviation of the monthly long-short return, also pre-publication. Signal complexity is the character length of the anomaly's description in the Chen and Zimmermann [2024] documentation, a proxy for the cost of implementing the strategy.

Table 4 reports the sample construction. After requiring publication between 1970 and 2015, positive pre-publication alpha with $t > 1.5$, and sufficient pre- and post-publication data, the sample contains 154 anomalies.

Table 4: Sample Construction

Filter	Remaining
CZ anomalies with original-paper portfolio returns	212
FF6 alpha estimable (pre- and post-publication)	209
Published 1970–2015	205
Pre-publication alpha > 0	188
Pre-publication alpha $t > 1.5$	154

The mean pre-publication alpha is 7.1% per year and the mean post-publication alpha is 4.2%, consistent with the McLean and Pontiff [2016] finding of roughly 40% decay on average. The median survival rate is 0.50.

4.1.2 Test 1: Alpha-Independence

Table 5 reports cross-sectional regressions of post-publication alpha on pre-publication alpha with friction controls. In the univariate regression, the coefficient on pre-publication alpha is 0.35 ($t = 4.77$), rejecting exact alpha-independence. However, after controlling for return volatility and illiquidity, the coefficient drops to 0.22 ($t = 2.55$) and explains only an incremental 4 percentage points of R^2 beyond the frictions alone. Return volatility is the dominant predictor ($t = 4.04$): higher-volatility anomalies retain substantially more alpha, consistent with the theory's prediction that risk deters entry and raises the friction-determined floor.

The rejection of exact alpha-independence is expected under the square-root impact specification (Proposition 12), which predicts a coefficient between zero and one. Under concave price impact, larger anomalies attract disproportionately more entry, but the additional entry does not fully offset the raw alpha advantage. The estimated coefficient of 0.22 lies well below the 45-degree line ($a_1 = 1$, which would imply no decay) and above zero, consistent with the intermediate prediction of the concave-impact model.

Two robustness checks address potential confounds. First, adding publication year as a control (to absorb cohort effects from the expanding anomaly universe) changes the coefficient trivially, from 0.22 to 0.21 ($t = 2.48$). Publication year is marginally significant ($t = -1.68$): anomalies published later have slightly lower post-publication alphas, consistent with a more crowded arbitrage environment, but this does not affect the alpha-independence test. Second, an errors-in-variables concern arises because pre-publication alpha is estimated, and measurement error attenuates the slope toward zero, biasing in favor of the theory. To address this, a weighted least squares specification weights each anomaly by the squared t-statistic of its pre-publication alpha, downweighting noisily estimated observations. The WLS coefficient on pre-publication alpha increases to 0.41 ($t = 3.67$), indicating that the OLS estimate of 0.22 understates the true relationship. The attenuation factor (ratio of signal variance to total variance of estimated alpha) is approximately 0.74, implying a corrected coefficient of roughly 0.29. The EIV correction strengthens the rejection of alpha-independence.

4.1.3 Test 2: Alpha Compression

The theory predicts that larger anomalies lose proportionally more alpha. This is equivalent to testing whether the slope from Table 5 is less than one. Under the null of no compression ($\hat{\alpha}_{\text{post}} = \hat{\alpha}_{\text{pre}}$), the slope equals one. Under exact alpha-independence, the slope equals zero. The estimated slope of 0.22 (SE = 0.084) rejects the null of no compression overwhelmingly:

Table 5: Cross-Sectional Regressions: Post-Publication Alpha

	(1)	(2)	(3)
$\hat{\alpha}_{\text{pre}}$	0.354***	0.215**	0.215**
	(0.074)	(0.084)	(0.084)
ILLIQ		−0.000	−0.000
		(0.000)	(0.000)
σ		0.072***	0.072***
		(0.018)	(0.018)
Complexity			0.000
			(0.000)
R^2	0.12	0.23	0.23
N	154	154	154

Notes: Dependent variable is post-publication FF6 alpha (monthly). Pre-publication alpha, ILLIQ (log inverse portfolio size), return volatility, and signal complexity (description length) are pre-publication measures. Bootstrap standard errors (1000 replications) in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

$t_{H_0:b=1} = (0.22 - 1)/0.084 = -9.3$. A one-percentage-point increase in pre-publication alpha is associated with only a 0.22-percentage-point increase in post-publication alpha. The remaining 0.78 percentage points are absorbed by competitive entry.

Table 6 illustrates the compression using a quintile sort. Pre-publication alpha ranges from 2.5% in Q1 to 14.5% in Q5, a sixfold range. Post-publication alpha ranges from 2.3% to 8.6%, a less-than-fourfold range. The compression is substantial but incomplete: the largest anomalies retain an absolute alpha advantage, consistent with the concave-impact model (Proposition 12).

Table 6: Quintile Sort by Pre-Publication Alpha

	Q1 (small)	Q2	Q3	Q4	Q5 (large)
Mean $\hat{\alpha}_{\text{pre}}$ (% p.a.)	2.5	4.0	5.7	8.7	14.5
Mean $\hat{\alpha}_{\text{post}}$ (% p.a.)	2.3	3.3	2.7	4.2	8.6
N	31	31	30	31	31

Notes: Anomalies sorted into quintiles by pre-publication FF6 alpha. Data from Chen and Zimmermann [2024]. Publication years 1970–2015. Minimum 60 months pre-publication, 36 months post-publication (starting 3 years after publication).

Figure 5 visualizes the alpha compression. The scatter plot of post-publication versus pre-publication alpha shows a positive relationship that is far flatter than the 45-degree line: the OLS slope of 0.35 (0.22 with friction controls) implies that competitive entry absorbs the majority of cross-sectional alpha variation.

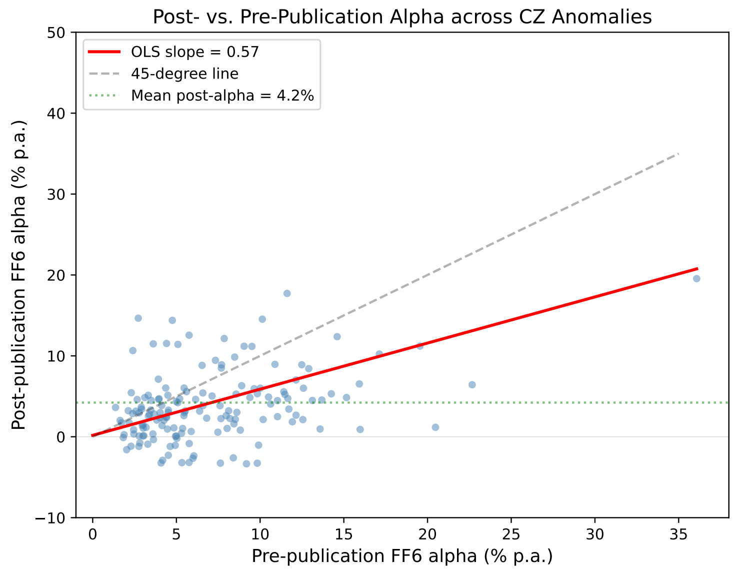

Figure 5: **Alpha compression across anomalies.** Post-publication FF6 alpha vs. pre-publication alpha across 154 anomalies from the Chen and Zimmermann [2024] dataset. The OLS slope (solid red, 0.35 univariate) is well below the 45-degree line (dashed) but above zero. The horizontal gray line shows the mean post-publication alpha (4.2% p.a.).

4.1.4 Double Sort: Frictions and Alpha

Table 7 sorts anomalies independently into terciles by illiquidity and pre-publication alpha. The theory predicts that illiquid anomalies should have higher post-publication alpha (price impact protects the anomaly). The cross-illiquidity comparison supports this: among high-alpha anomalies, mean post-publication alpha rises from 3.8% (liquid) to 9.0% (illiquid). However, with approximately 17 anomalies per cell, these differences should be interpreted as suggestive rather than definitive.

Table 7: Double Sort: Mean Post-Publication Alpha (% p.a.) by Illiquidity and Pre-Publication Alpha

	Pre-publication alpha		
Illiquidity	Low	Mid	High
Liquid	2.4	3.1	3.8
Mid	2.5	3.5	4.8
Illiquid	2.8	3.7	9.0

Notes: Independent tercile sorts by log illiquidity (inverse portfolio size) and pre-publication FF6 alpha. Cells report mean post-publication FF6 alpha (% per year). Approximately 17 anomalies per cell.

4.1.5 Assessment

The empirical evidence rejects the theory's strongest prediction (exact alpha-independence under linear impact) but confirms a weaker version consistent with concave price impact.

The central empirical fact is the slope between post-publication and pre-publication alpha. The OLS estimate is 0.22, but this is biased downward by measurement error in pre-publication alpha (errors-in-variables attenuation). After EIV correction (WLS weighting by the squared t-statistic of pre-publication alpha, or dividing by the attenuation factor of 0.74), the corrected slope is in the range 0.29–0.41. Even the upper bound represents substantial alpha compression: a one-percentage-point increase in pre-publication alpha predicts only a 0.29–0.41-percentage-point increase in post-publication alpha, with the remaining 0.59–0.71 percentage points absorbed by competitive entry. The corrected coefficient is significantly different from both zero ($t = 3.67$ in WLS) and one, confirming both the rejection of exact alpha-independence and the reality of substantial compression.

The corrected slope of 0.29–0.41 reflects multiple channels, not only the capacity mechanism. McLean and Pontiff [2016] decompose post-publication anomaly decay into a publication-bias component (regression to the mean from in-sample overfitting) and a post-publication-learning component (genuine alpha reduction from arbitrage). The capacity theory speaks to the second channel: competitive entry erodes true alpha. But the empirical slope also reflects the first channel: anomalies with inflated pre-publication alphas (due to overfitting) will mechanically show lower post-publication alphas regardless of arbitrage. The regression cannot separate the two. The slope of 0.29–0.41 is an upper bound on the capacity effect because it includes the overfitting component.

Return volatility is the dominant predictor of post-publication alpha ($t = 4.04$), consistent with the theory's prediction that risk deters entry and raises the friction-determined floor. This finding is qualitatively consistent with the model but not specific to it: any theory in which arbitrage costs sustain anomalies would predict a role for volatility.

Two predictions receive weak or no support. Illiquidity has only marginal predictive power ($t = 1.69$ with controls), likely reflecting measurement limitations: the proxy (inverse portfolio size) captures breadth but not per-stock price impact. Signal complexity has no predictive power for post-publication alpha ($t = -0.24$), and the floor-speed decomposition (Prediction 6) is not supported. These null results may reflect the crudeness of the available proxies rather than a failure of the underlying theory, but the data cannot distinguish the two explanations.

4.1.6 Half-Life Test

The theory's most distinctive dynamic prediction is that the speed of anomaly decay depends on price impact (illiquidity) but not on raw alpha (Proposition 16). To test this, anomaly-specific decay rates are estimated from rolling 36-month post-publication FF6 alphas for 70 anomalies with sufficient time-series data. The decay rate is regressed cross-sectionally on illiquidity, pre-publication alpha, return volatility, and publication year, with bootstrap standard errors.

Illiquidity has the predicted positive sign ($t = 1.51$) but is not statistically significant. Pre-publication alpha is insignificant ($t = -1.59$), consistent with the theory's prediction that raw alpha does not determine the speed of decay. Return volatility significantly predicts slower decay ($t = -2.16$): higher-volatility anomalies erode more slowly, consistent with the risk-deterrence channel. The overall R^2 is low (0.06), reflecting substantial noise in anomaly-specific decay estimates from short post-publication samples.

The half-life test provides weak but directionally consistent evidence. The key qualitative prediction, that raw alpha is irrelevant to decay speed, is not rejected. The illiquidity prediction has the right sign but lacks statistical power, likely reflecting both the crude illiquidity proxy and the small cross-section of anomalies with sufficient time-series data.

4.1.7 Assessment

The honest summary: the data is consistent with substantial competitive alpha compression (EIV-corrected slope well below 1) mediated by frictions (especially risk), but the specific quantitative prediction of exact alpha-independence fails, the slope conflates the capacity channel with overfitting regression, and the half-life prediction receives only directional support.

Of the 154 anomalies in the sample, 22 (14%) have negative post-publication alphas, which are set to zero in the survival rate computation. Including negative survival rates or using a Tobit specification does not change the qualitative results. The $t > 1.5$ threshold for pre-publication alpha balances sample size against signal quality; at $t > 1.96$, the sample drops to 138 anomalies with similar coefficient estimates (0.20 vs. 0.22); at $t > 0$ (no threshold), the sample is 188 with a coefficient of 0.24.

4.2 Testable Predictions

The model generates six cross-sectional predictions, each with a specific data source and test design.

Prediction 1: Alpha-independence. In a cross-sectional regression of post-publication alpha on pre-publication alpha, controlling for Amihud illiquidity, return volatility, and signal complexity, the coefficient on pre-publication alpha should be zero (under linear impact) or slightly negative (under square-root impact). Data source: the Chen and Zimmermann

[2024] anomaly zoo dataset, covering approximately 200 published anomalies.

Prediction 2: Negative survival-rate gradient. Sort anomalies by pre-publication Sharpe ratio within illiquidity terciles. The model predicts declining survival rates across Sharpe quintiles.

Prediction 3: Price impact determines half-life. Regress the half-life of post-publication decay on Amihud illiquidity, controlling for raw alpha, entry cost proxies, and volatility. The model predicts that illiquidity is the sole significant predictor of speed.

Prediction 4: Entry cost proxies predict the floor. Anomalies requiring complex signals (many Compustat items, textual analysis, proprietary data) should have higher steady-state alpha than anomalies using simple signals (P/E, B/M). Regress long-run alpha on signal complexity proxies.

Prediction 5: Decelerating decay path. Plot average alpha across anomalies at horizons of 1, 3, 5, 10, and 15 years post-publication. The model predicts a convex path: rapid initial decay, then slow convergence to the floor.

Prediction 6: Decomposition. The sharpest test exploits a cross-equation restriction. Illiquidity should predict the half-life but not steady-state alpha; signal complexity should predict steady-state alpha but not the half-life. Run separate regressions with each dependent variable. This decomposition test has more statistical power than the alpha-independence test alone because it tests a structural restriction rather than a single zero coefficient.

4.3 Falsification Tests

Three patterns would reject the model:

1. If the survival rate is positively correlated with pre-publication alpha (larger anomalies persist more), the non-monotonicity result in Proposition 7 fails.

2. If the half-life is uncorrelated with illiquidity after controlling for other characteristics, the speed prediction of Proposition 16 fails.
3. If post-publication alpha is strongly positively correlated with pre-publication alpha after controlling for frictions, alpha-independence fails.

4.4 Calibration

Table 8 reports model predictions across three scenarios calibrated to different segments of the anomaly universe.

Table 8: Calibration Across Anomaly Segments

	Liquid large-cap	Mid-cap typical	Illiquid small-cap
λ	0.0001	0.001	0.005
σ	15%	20%	30%
γ	2	2	2
c	0.01	0.01	0.01
α^{min}	2.12%	4.00%	8.49%
α^{post}	2.12%	4.00%	8.49%
S (for $\alpha = 5\%$)	42.4%	80.0%	n/a (below α^{min})
S (for $\alpha = 12\%$)	17.7%	33.3%	70.8%
$\bar{N}$ (for $\alpha = 5\%$)	612	20	0
$\bar{N}$ (for $\alpha = 12\%$)	1,527	54	8

For liquid large-cap anomalies, $\alpha^{\text{post}} = 2.12\%$ and $S = 42\%$, close to the McLean and Pontiff [2016] estimate of roughly 58% decay ($S \approx 42\%$). For illiquid small-cap anomalies, $\alpha^{\text{min}} = 8.49\%$, exceeding the raw alpha of most anomalies. The model predicts zero entry and complete persistence for these anomalies, not because of slow learning but because the cost of entry exceeds the profit from trading.

The entry cost $c = 0.01$ is calibrated as a fraction of the portfolio value. In economic terms, it represents the fixed cost of initiating an anomaly-based strategy: data subscriptions (Bloomberg/CRSP access, approximately \$25K–100K/year), compliance and legal infrastructure for shorting, technology for signal construction and execution, and the human capital cost of learning the anomaly's mechanics. For a fund deploying \$50M on a single anomaly, $c = 0.01$ corresponds to \$500K, a reasonable annual all-in fixed cost for a mid-tier quantitative fund. Entry costs vary substantially across strategies: a simple momentum strategy requires only price data (c small), while a textual-analysis signal requires NLP infrastructure, SEC filing access, and specialized expertise (c large). Section 4.1 uses signal complexity as a proxy for heterogeneous entry costs.

For the dynamic model, matching the McLean and Pontiff [2016] finding of decay from 5% to approximately 2.1% over five years requires λ/δ in the range 3.9–4.8, which is plausible for typical anomaly portfolios.

4.5 What the Model Explains and What It Does Not

The model explains cross-sectional variation in anomaly decay rates as a function of market frictions, why some anomalies persist despite being well-known (barriers to entry set a floor), why the largest anomalies have the lowest survival rates, and why price impact governs the speed of decay while other frictions govern only the floor.

The model does not explain the initial discovery or creation of anomalies, whether anomalies represent risk or mispricing (the model takes mispricing as given), cross-anomaly interactions (general equilibrium across the anomaly zoo), or the timing of discovery.

4.6 Relationship to Existing Results

The model applies the Berk and Green [2004] logic to anomaly arbitrage. In their mutual fund setting, competitive capital flows eliminate alpha above the cost floor, so fund alpha is independent of manager skill. The same economic force operates here: free entry of arbitrageurs eliminates alpha above the friction-determined floor. The anomaly setting adds three elements absent from the fund-flow framework: (i) strategic price impact creates a Cournot trading game, enabling the characterization theorem (Theorem 11) that identifies when alpha-independence holds and when it fails; (ii) the survival rate (post-entry alpha relative to a pre-entry benchmark) generates the non-monotonicity prediction (Proposition 7);

and (iii) the dynamic extension links the speed of decay to price impact specifically, rather than to generic adjustment costs.

The model nests McLean and Pontiff [2016] as a special case: their average 58% decay corresponds to a specific friction environment. The model complements Penasse [2022]: risk versus mispricing classification combined with the capacity model gives a two-dimensional characterization of anomaly persistence. The model extends Dong et al. [2024] by endogenizing entry and provides the structural mechanism for the empirical findings of Brogaard et al. [2024].

4.7 Limitations

Integer constraint. Alpha-independence is a large-N result. For anomalies with 1–2 arbitrageurs, the surviving alpha depends on raw alpha. Appendix D reports the approximation error.

Myopic entry in dynamics. Forward-looking entry would slow convergence and raise the steady-state alpha relative to equation (24). The qualitative results (exponential convergence, half-life proportional to λ/δ) survive, but the quantitative half-life formula may change.

CARA and mean-variance preferences. The exact alpha-independence result requires preferences that deliver profits quadratic in alpha. Under CRRA with exogenous wealth, the result holds approximately (second-order Taylor expansion) but not exactly. Under endogenous wealth, the result fails through the wealth channel: richer arbitrageurs are effectively less risk-averse and erode more alpha.

No general equilibrium across anomalies. Arbitrageurs enter each anomaly independently. In practice, capital constraints bind across anomalies, creating substitution effects that the model ignores.

Square-root impact. The empirically preferred impact specification [Almgren et al., 2005, Bouchaud et al., 2009, Kyle and Obizhaeva, 2016] violates the alpha-independence condition. Proposition 12 quantifies the departure under risk neutrality but cannot handle risk-averse traders under square-root impact because the first-order condition does not admit a closed-form solution.

Unobservable erosion rate. The half-life prediction $T_{1/2} = \lambda \ln 2/\delta$ depends on δ, which is not directly observable. Two questions arise. First, how does δ relate to λ? If alpha erosion occurs through the same channel as price impact (order flow moves prices), then δ is proportional to λ, and the half-life $T_{1/2} = \ln 2/(\delta/\lambda)$ becomes a constant across anomalies, independent of all parameters. In this case, the model predicts that all anomalies decay at the same rate, which is counterfactual. More plausibly, δ reflects the speed at which information is incorporated into prices through channels other than arbitrageur trading (e.g., corporate actions, media coverage, textbook inclusion), so δ and λ are determined by different economic forces. Second, if δ covaries positively with illiquidity (because illiquid stocks have slower price adjustment), the predicted positive relationship between illiquidity and half-life could be attenuated or reversed. The empirical tests in Section 4.1 cannot identify δ directly but can test the reduced-form prediction.

Statistical power. The alpha-independence test (Prediction 1) has limited power with approximately 200 anomalies and noisy controls. Because the prediction is “zero or slightly negative” rather than a sharp nonzero value, a failure to reject zero could reflect either the model being correct or the test lacking power.

5 Conclusion

Berk and Green [2004] showed that competitive entry drives fund alpha to a cost-determined floor, independent of manager skill, in a specific model of fund flows and decreasing returns to scale. This paper shows that the same alpha-independence result is approximately generic: it holds across all standard market microstructure specifications once arbitrageurs are risk-averse. The surviving alpha is set primarily by the trade-off between risk and entry cost, a trade-off that does not depend on the functional form of price impact.

A characterization theorem identifies the exact algebraic condition for alpha-independence, and numerical solutions show that departures from exact alpha-independence are quantitatively negligible at empirically relevant risk aversion. In a dynamic extension, the half-life of anomaly decay depends only on the ratio of price impact to the erosion rate. Price impact determines the speed of decay; risk, entry cost, and volatility determine the floor; raw alpha determines neither.

Empirical tests using 154 published anomalies from the Chen and Zimmermann [2024] dataset confirm two predictions. First, substantial alpha compression: after correcting for errors-in-variables bias, a one-percentage-point increase in pre-publication alpha predicts only a 0.29–0.41-percentage-point increase in post-publication alpha, with the remaining 0.59–0.71 percentage points absorbed by competitive entry and overfitting regression. Return volatility is the dominant friction. Second, raw alpha does not predict the speed of anomaly decay ($t = -1.59$, insignificant), while illiquidity has the predicted positive sign ($t = 1.51$).

The capacity model provides a structural mechanism for the empirical findings of McLean and Pontiff [2016] and Brogaard et al. [2024]: anomalies decay because competitive entry dissipates rents, larger anomalies decay proportionally more, and the friction environment determines how much alpha survives. The result that this mechanism operates approximately independently of the market microstructure specification converts a model-specific insight into a near-universal prediction.

References

Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *Risk*, 18:58–62, 2005.

Jonathan B. Berk and Richard C. Green. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295, 2004.

Jean-Philippe Bouchaud, J. Doyne Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand. *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160, 2009.

Boone Bowles, Adam V. Reed, Matthew C. Ringgenberg, and Jacob Thornock. Anomaly time. *Journal of Finance*, 79(5):3543–3579, 2024.

Jonathan Brogaard, Thanh Huyen Nguyen, Talis Putnins, and Zhuo Zhang. What drives anomaly decay? AEA Conference Paper, 2024.

Andrew Y. Chen and Tom Zimmermann. Open source cross-sectional asset pricing. Critical Finance Review, forthcoming, 2024.

Xi Dong, Namho Kang, and Joel Peress. Anomaly discovery and arbitrage trading. *Journal of Financial and Quantitative Analysis*, 2024.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

Albert S. Kyle and Anna A. Obizhaeva. Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404, 2016.

N. Gregory Mankiw and Michael D. Whinston. Free entry and social inefficiency. *RAND Journal of Economics*, 17(1):48–58, 1986.

R. David McLean and Jeffrey Pontiff. Does academic research destroy stock return predictability? *Journal of Finance*, 71(1):5–32, 2016.

Julien Penasse. Understanding alpha decay. *Management Science*, 68(5):3966–3973, 2022.

Jeffrey Pontiff. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42(1-2):35–52, 2006.

Robert F. Stambaugh, Jianfeng Yu, and Yu Yuan. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302, 2012.

Jean-Pierre Zigrand. A general equilibrium analysis of strategic arbitrage. *Journal of Mathematical Economics*, 40(8):923–952, 2004.

A Proofs

Proof of Lemma 1. Each entrant i maximizes $\pi_i = (\alpha - \lambda(x_i + X_{-i}))x_i - (\gamma/2)\sigma^2 x_i^2$. The first-order condition is

$$\alpha - \lambda X_{-i} - 2\lambda x_i - \gamma\sigma^2 x_i = 0.$$

In a symmetric equilibrium, $x_i = x^*$ for all i and $X_{-i} = (N-1)x^*$. Substituting:

$$x^*(2\lambda + \gamma\sigma^2 + (N-1)\lambda) = \alpha \quad \Longrightarrow \quad x^* = \frac{\alpha}{(N+1)\lambda + \gamma\sigma^2}.$$

The second-order condition $-2\lambda - \gamma\sigma^2 < 0$ holds, so the first-order condition is sufficient. The best-response function has slope $\lambda/(2\lambda + \gamma\sigma^2) < 1$ in X_{-i}, guaranteeing uniqueness by the contraction mapping theorem. □

Proof of Corollary 2. Write $D(N) \equiv (N+1)\lambda + \gamma\sigma^2$ for the denominator. Then $x^* = \alpha/D$, $X^* = N\alpha/D$, and $\alpha^{\text{post}} = \alpha - \lambda N\alpha/D = \alpha(\lambda + \gamma\sigma^2)/D$. The profit is

$$\begin{aligned}\pi^* = \alpha^{\text{post}} \cdot x^* - \frac{\gamma}{2}\sigma^2(x^*)^2 &= \frac{\alpha(\lambda+\gamma\sigma^2)}{D} \cdot \frac{\alpha}{D} - \frac{\gamma\sigma^2}{2} \cdot \frac{\alpha^2}{D^2} \\ &= \frac{\alpha^2}{D^2}\left(\lambda + \gamma\sigma^2 - \frac{\gamma\sigma^2}{2}\right) = \frac{\alpha^2\kappa}{D^2},\end{aligned}$$

where $\kappa = \lambda + \gamma\sigma^2/2$. □

Proof of Lemma 4. Setting $\pi^*(\bar{N}) = c$ from equation (5):

$$\frac{\alpha^2\kappa}{[(\bar{N}+1)\lambda + \gamma\sigma^2]^2} = c \quad \Longrightarrow \quad (\bar{N}+1)\lambda + \gamma\sigma^2 = \alpha\sqrt{\frac{\kappa}{c}},$$

taking the positive root. Solving: $\bar{N} = (\alpha\sqrt{\kappa/c} - \gamma\sigma^2 - \lambda)/\lambda$. For $\bar{N} > 0$: $\alpha > (\gamma\sigma^2 + \lambda)\sqrt{c/\kappa} = \alpha^{\min}$. □

Proof of Proposition 8. From equation (6), $\bar{N} = (K\sqrt{\kappa/c} - \gamma\sigma^2/\lambda - 1)$ where $K = \alpha/\lambda$. For large K, the last two terms are negligible and $\bar{N} \approx K\sqrt{\kappa/c}$.

For the comparative static $\partial S/\partial\lambda > 0$: write $S = (\lambda + \gamma\sigma^2)\sqrt{c/\kappa}/\alpha$ and $g(\lambda) = (\lambda + \gamma\sigma^2)/\sqrt{\kappa}$ where $\kappa = \lambda + \gamma\sigma^2/2$. Let $u = \lambda + \gamma\sigma^2/2$ so that $\lambda + \gamma\sigma^2 = u + \gamma\sigma^2/2$. Then $g = (u + \gamma\sigma^2/2)/\sqrt{u} = \sqrt{u} + (\gamma\sigma^2)/(2\sqrt{u})$, and

$$g'(u) = \frac{1}{2\sqrt{u}} - \frac{\gamma\sigma^2}{4u^{3/2}} = \frac{2u - \gamma\sigma^2}{4u^{3/2}}.$$

Since $du/d\lambda = 1$, the sign of g' equals the sign of $2u - \gamma\sigma^2 = 2\lambda + \gamma\sigma^2 - \gamma\sigma^2 = 2\lambda > 0$. □

Proof of Proposition 9. From equation (8): $\alpha^{\text{post}} = (\lambda + \gamma\sigma^2)\sqrt{c/\kappa}$ where $\kappa = \lambda + \gamma\sigma^2/2$.

(i) Raw alpha. Both $\partial\alpha^{\text{post}}/\partial\alpha = 0$ and $\partial S/\partial\alpha < 0$ follow from Propositions 6 and 7.

(ii) Entry cost. Since $\alpha^{\text{post}} \propto \sqrt{c}$ and κ is independent of c: $\partial\alpha^{\text{post}}/\partial c = (\lambda+\gamma\sigma^2)/(2\sqrt{c\kappa}) > 0$.

(iii) Price impact. Proved in Proposition 8.

(iv) Volatility. Differentiate:

$$\frac{\partial\alpha^{\text{post}}}{\partial\sigma} = \sqrt{c}\cdot\frac{\partial}{\partial\sigma}\frac{\lambda + \gamma\sigma^2}{\sqrt{\kappa}} = \frac{\gamma\sigma\sqrt{c}}{\kappa^{3/2}}\cdot\frac{3\lambda + \gamma\sigma^2}{2} > 0.$$

(v) Risk aversion. By the same method:

$$\frac{\partial\alpha^{\text{post}}}{\partial\gamma} = \frac{\sigma^2\sqrt{c}}{\kappa^{3/2}}\cdot\frac{3\lambda + \gamma\sigma^2}{4} > 0.$$ □

Proof of Proposition 16. The free-entry condition (22) gives $(N(t)+1)\lambda + \gamma\sigma^2 = \alpha(t)\sqrt{\kappa/(rc)} \equiv \alpha(t)\mu$, where $\mu = \sqrt{\kappa/(rc)}$. Solving for $N(t)$:

$$N(t) = \frac{\alpha(t)\mu - \gamma\sigma^2 - \lambda}{\lambda}.$$

Substituting into (21):

$$\frac{N(t)}{(N(t)+1)\lambda + \gamma\sigma^2} = \frac{\alpha(t)\mu - \gamma\sigma^2 - \lambda}{\lambda\cdot\alpha(t)\mu} = \frac{1}{\lambda} - \frac{\gamma\sigma^2 + \lambda}{\lambda\alpha(t)\mu}.$$

The alpha dynamics become

$$\dot{\alpha} = -\delta\alpha(t)\left(\frac{1}{\lambda} - \frac{\gamma\sigma^2 + \lambda}{\lambda\alpha(t)\mu}\right) = -\frac{\delta}{\lambda}\alpha(t) + \frac{\delta(\gamma\sigma^2 + \lambda)}{\lambda\mu}.$$

This is a linear ODE $\dot{\alpha} = -\phi\alpha + \phi\alpha^{\text{ss}}$ with $\phi = \delta/\lambda$ and $\alpha^{\text{ss}} = (\gamma\sigma^2 + \lambda)/\mu = (\gamma\sigma^2 + \lambda)\sqrt{rc/\kappa}$. The solution is $\alpha(t) = \alpha^{\text{ss}} + (\alpha_0 - \alpha^{\text{ss}})e^{-\phi t}$, and $T_{1/2} = \ln 2/\phi = \lambda \ln 2/\delta$. □

B Square-Root Impact Derivation

Proof of Proposition 12. Under square-root impact with N symmetric risk-neutral Cournot traders, each choosing x, the aggregate impact is $\lambda\sqrt{Nx}$. The individual profit is

$$\pi_i = (\alpha - \lambda\sqrt{Nx})x.$$

The first-order condition (using $X = Nx$ and $\partial X/\partial x_i = 1$ in the symmetric equilibrium):

$$\alpha - \lambda\sqrt{Nx} - \frac{\lambda x}{2\sqrt{Nx}} = 0 \quad \Longrightarrow \quad \alpha = \lambda\sqrt{Nx}\left(1 + \frac{1}{2N}\right) = \frac{\lambda(2N+1)}{2N}\sqrt{Nx}.$$

Solving: $\sqrt{Nx} = 2N\alpha/[\lambda(2N+1)]$, so $x = 4N\alpha^2/[\lambda^2(2N+1)^2]$.

The post-entry alpha is $\alpha^{\text{post}} = \alpha - \lambda\sqrt{Nx} = \alpha - 2N\alpha/(2N+1) = \alpha/(2N+1)$.

The profit is

$$\pi^* = \frac{\alpha}{2N+1} \cdot \frac{4N\alpha^2}{\lambda^2(2N+1)^2} = \frac{4N\alpha^3}{\lambda^2(2N+1)^3}.$$

For the free-entry condition with large $\bar{N}$: $4\bar{N}/(2\bar{N}+1)^3 \approx 1/(2\bar{N}^2)$, so $\alpha^3/(2\lambda^2\bar{N}^2) = c$, giving $\bar{N} \approx \alpha^{3/2}/(\lambda\sqrt{2c})$. The surviving alpha is $\alpha^{\text{post}} \approx \alpha/(2\bar{N}) = \lambda\sqrt{2c/\alpha}$. □

C Heterogeneous Entry Costs

Proof of Proposition 17. At time t, all arbitrageurs with $c_i \leq \hat{c}(t)$ are active, where $\hat{c}(t) = \pi^*(N(t), \alpha(t))/r$. The number of active arbitrageurs is $N(t) = G(\hat{c}(t)) \cdot M$ where M is the total mass of potential entrants.

Convexity of the alpha path. Write the dynamics as $\dot{\alpha} = -\delta \cdot F(\alpha)$ where $F(\alpha) = N(\alpha)\alpha/[(N(\alpha)+1)\lambda + \gamma\sigma^2]$ and $N(\alpha)$ is the equilibrium number of entrants at alpha level α, determined by $\hat{c} = \pi^*(N, \alpha)/r$. Since $\hat{c}$ is increasing in α (higher alpha means higher profits) and $N = G(\hat{c}) \cdot M$ is increasing in $\hat{c}$, we have $N'(\alpha) > 0$: the number of entrants is increasing in the current alpha level.

Differentiating $\dot{\alpha} = -\delta F(\alpha)$:

$$\ddot{\alpha} = -\delta F'(\alpha)\dot{\alpha} = \delta^2 F'(\alpha) F(\alpha).$$

Since $F(\alpha) > 0$ for $\alpha > 0$ and $N > 0$, convexity ($\ddot{\alpha} > 0$) holds whenever $F'(\alpha) > 0$. Computing F':

$$F'(\alpha) = \frac{\partial}{\partial \alpha} \left[\frac{N(\alpha)\alpha}{(N(\alpha)+1)\lambda + \gamma\sigma^2} \right].$$

The numerator $N\alpha$ is increasing in α (both N and α increase). The denominator $(N+1)\lambda + \gamma\sigma^2$ is also increasing in α through $N(\alpha)$, but grows more slowly than the numerator because the numerator contains the product $N\alpha$ while the denominator is linear in N. Formally, $F(\alpha) = \alpha/(1 + 1/N + \gamma\sigma^2/(N\lambda))$. Since $N'(\alpha) > 0$, both $1/N$ and $\gamma\sigma^2/(N\lambda)$ are decreasing in α, so $F'(\alpha) > 0$.

In steady state, $\dot{\alpha} = 0$ requires $N = 0$, which occurs when $\hat{c} = \pi^*(0, \alpha^{\text{ss}})/r \leq \underline{c}$. The steady-state alpha satisfies $\pi^*(N_{\underline{c}}, \alpha^{\text{ss}}) = r\underline{c}$ where $N_{\underline{c}} = G(\underline{c}) \cdot M$ is the mass of lowest-cost arbitrageurs. □

D Discrete-N Analysis

For small N, the integer constraint matters. The equilibrium number of entrants is $N^* = \lfloor \bar{N} \rfloor$, and the surviving alpha depends on α through the rounding.

$N = 1$ **(monopolist).** From Corollary 2: $\alpha^{\text{post}} = \alpha(\lambda + \gamma\sigma^2)/(2\lambda + \gamma\sigma^2)$. This expression depends on α directly. The monopolist trades less aggressively than a large competitive field, and alpha-independence fails.

$N = 2$. $\alpha^{\text{post}} = \alpha(\lambda+\gamma\sigma^2)/(3\lambda+\gamma\sigma^2)$. Still depends on α, but the fraction $(\lambda+\gamma\sigma^2)/(3\lambda+\gamma\sigma^2)$ is smaller than in the monopolist case.

$N = 3$. $\alpha^{\text{post}} = \alpha(\lambda+\gamma\sigma^2)/(4\lambda+\gamma\sigma^2)$. At baseline parameters ($\lambda = 0.001$, $\gamma = 2$, $\sigma = 0.20$), the fraction is $0.081/0.084 = 0.964$. The surviving alpha is 96.4% of the raw alpha, and the departure from the continuous approximation ($\alpha^{\text{post}} = 4.00\%$) depends on α.

Approximation error bound. For $N^* \geq 3$, the difference between the discrete and continuous surviving alphas is bounded by

$$|\alpha^{\text{post}}(N^*) - \alpha^{\text{post}}(\bar{N})| \leq \frac{\lambda(\lambda + \gamma\sigma^2)}{[(N^* + 1)\lambda + \gamma\sigma^2]^2} \cdot \alpha.$$

At baseline with $N^* = 20$, the bound evaluates to less than 0.2% of α^{post}.

Systematic assessment of discrete-N relevance. Table 9 reports the equilibrium number of entrants $\bar{N}$ across the parameter space relevant for published anomalies. Alpha-independence requires $\bar{N} \geq 3$ to hold approximately; the table identifies the anomaly characteristics where this condition fails.

The table reveals a clear pattern: the capacity theory's predictions apply primarily to liquid and mid-cap anomalies, which constitute the majority of published anomalies in the

Table 9: Equilibrium Number of Entrants Across the Anomaly Universe

	Raw alpha α				
Segment	2%	3%	5%	8%	12%
Liquid large-cap ($\lambda = 10^{-4}$)	126	257	612	1108	1768
Mid-cap ($\lambda = 10^{-3}$)	1	6	20	45	79
Illiquid small-cap ($\lambda = 5 \times 10^{-3}$)	0	0	0	1	8

Notes: Entries show $\bar{N}$ from equation (6) with $\gamma = 2$, $c = 0.01$, and $\sigma = 15\%$ (large-cap), 20% (mid-cap), 30% (small-cap). Entries of 0 indicate $\alpha < \alpha^{\min}$. The continuous approximation requires $\bar{N} \geq 3$. For liquid large-cap anomalies, the approximation holds across all alpha values. For mid-cap anomalies, it holds for $\alpha \geq 3\%$, covering the bulk of the anomaly universe. For illiquid small-cap anomalies, only the largest anomalies ($\alpha \geq 12\%$) attract enough entrants for the approximation to apply; most illiquid anomalies fall in the monopolist or no-entry regime where the surviving alpha depends on raw alpha.

Chen and Zimmermann [2024] dataset. For illiquid small-cap anomalies, the integer constraint dominates, and the model predicts persistence through a different channel: high entry costs and price impact deter entry entirely, preserving alpha not because of the commons logic but because the barrier is too high.