# The Measurement Problem

**Fat Tails, Estimation, and What Standard Finance Gets Wrong**

Jean-Marc Choufani

2026

# Table of contents


**1 Preface** 5

**Preface** 6
- 1.1 What this book argues . . . 6
- 1.2 Four central claims . . . 6
- 1.3 Who this is for . . . 7
- 1.4 On the author . . . 7
- 1.5 How to cite . . . 8
- 1.6 License . . . 8

**I Part I — The Taxonomy Problem** 9

**2 The Predictable Tail** 10
- 2.1 Abstract . . . 10
- 2.2 1. Introduction . . . 11
- 2.3 2. The Payoff Asymmetry Metric . . . 11
- 2.4 3. Main Results . . . 11
- 2.5 4. Why Standard Methods Miss It . . . 12
- 2.6 5. The Look-Ahead Contamination . . . 12
- 2.7 6. Conclusion . . . 13

**3 Not All Lottery Stocks Are Equal** 14
- 3.1 Abstract . . . 14
- 3.2 1. The Binary Event Problem . . . 15
- 3.3 2. Taxonomy Extension: FAT_STRUCTURAL and FAT_BINARY . . . 15
- 3.4 3. Empirical Results . . . 15
- 3.5 4. Sector Composition . . . 16
- 3.6 5. Implications . . . 16

**II Part II — The Estimation Problem** 17

**4 The Hill Estimator's Domain of Validity** 18
- 4.1 Abstract . . . 18
- 4.2 1. The Hill Estimator and Its Domain . . . 19
  - 4.2.1 1.1 The Estimator . . . 19
  - 4.2.2 1.2 Consistency Requires Regular Variation, Not Absolute Continuity 19
  - 4.2.3 1.3 The Mixture Process . . . 20
- 4.3 2. The Inconsistency Proposition . . . 20

4.4 3. Empirical Evidence . . . 21
4.4.1 3.1 Identification . . . 21
4.4.2 3.2 The 13.7pp Spread . . . 21
4.4.3 3.3 Binary Win Rates as a Diagnostic . . . 21
4.5 4. What This Does Not Claim . . . 21
4.6 5. Implications for the Taleb Taxonomy . . . 22

**5 Two Problems, One Incomplete Fix 23**
5.1 Abstract . . . 23
5.2 1. Two Problems . . . 24
5.2.1 1.1 Problem 1: Serial Correlation . . . 24
5.2.2 1.2 Problem 2: Directional Bias in the Estimand . . . 24
5.3 2. Formal Decomposition . . . 25
5.4 3. Empirical Magnitude . . . 25
5.5 4. Fat-Tail Amplification . . . 26
5.6 5. Differentiation from Prior Literature . . . 26
5.7 6. The Diagnostic . . . 27

**III Part III — The Inference Framework 28**

**6 The Nine-Gate Framework 29**
6.1 Overview . . . 29

**IV Part IV — Forthcoming 31**

**7 Ergodicity Deficit Across Tail Classes 32**
7.1 Motivation . . . 32
7.2 The Open Question . . . 33
7.3 Key prior work . . . 33

**8 Minimum Track Record Length Under Fat Tails 34**
8.1 Motivation . . . 34
8.2 The Claim . . . 35
8.3 Key prior work . . . 35

**9 Maximum Drawdown Under Power Laws 36**
9.1 Motivation . . . 36
9.2 The Claim . . . 36
9.3 Key prior work . . . 37

**10 Ontological vs Epistemic Extremes 38**
10.1 Motivation . . . 38
10.2 The Claim . . . 39
10.3 Key prior work . . . 39

**11 Against Optimization 40**
11.1 Motivation . . . 40

11.2 The Claim . . . . . 41
11.3 Connection to the Precautionary Principle . . . . . 41
11.4 Key prior work . . . . . 41

# 1 Preface

# Preface

This is a working monograph. It is incomplete by design — new chapters are added as the research develops. Every version is dated. Errors, when found, are corrected in place and noted in the session log.

## 1.1 What this book argues

Standard financial methodology is built on assumptions that break quietly under fat-tailed return distributions. The breakdowns are not exotic edge cases — they occur at tail exponents routinely observed in real equity data ( 2–3). Each chapter of this book isolates one such breakdown, states the mathematical condition under which it occurs, and quantifies its empirical consequence using a panel of 106,163 stock-month observations across 715 U.S. equities from 2009 to 2025.

The through-line is epistemological: **the method you choose determines what you can see, and the standard toolkit was not designed with fat tails in mind.** This does not mean the standard toolkit is always wrong — it means there is an identifiable class of problems for which it systematically misleads, and that class is larger than commonly acknowledged.

## 1.2 Four central claims

1. **The Hill estimator's domain of validity is narrower than its use implies.** Applied to equity returns driven by discrete binary events (regulatory decisions, clinical trial readouts), it is inconsistent for the structural tail exponent. The consequence is systematic misclassification within the Taleb (2025) taxonomy.
2. **Newey-West HAC standard errors address one of two distinct problems introduced by overlapping windows.** They correct serial correlation inflation. They

do not correct directional bias in the point estimate — a separate problem that is amplified, not attenuated, by fat-tailed return distributions.

3. **The ergodicity deficit, as formalized by Peters (2019), has not been calibrated to empirically estimated tail exponents.** The standard formula ( $^2/2$) is undefined for 2 and mis-specified for (2, 3). The empirical consequence for long-horizon compound growth has not been quantified across tail classes.

4. **The minimum track record required to distinguish skill from luck is a function of the tail exponent of the underlying return distribution.** At $\hat{} = 2.69$ — the median for U.S. equities in this sample — the required track record is substantially longer than Gaussian theory implies, with implications for how investment performance should be evaluated.

## 1.3 Who this is for

Anyone willing to think carefully about probability and measurement. The mathematics is precise but the motivation is accessible: these are questions about what we can and cannot know from the data we have, under the distributional conditions that actually obtain.

## 1.4 On the author

I am an orthodontist. I have no institutional affiliation in finance or mathematics. This work is guided entirely by curiosity and by a commitment to getting the mathematics right. Every substantive claim has been subjected to adversarial verification — an attempt to refute it before publishing it. Where the verification found errors, the errors are documented and the corrections noted.

Independent inquiry has a long tradition in quantitative research. What matters is whether the argument is correct, not whether the author holds the right credentials. Readers who find errors are encouraged to open an issue on the GitHub repository.

## 1.5 How to cite

> Choufani, J.-M. (2026). *The Measurement Problem: Fat Tails, Estimation, and What Standard Finance Gets Wrong* (working monograph). Retrieved from [URL].

Individual chapters have separate SSRN preprint identifiers where available; these are listed at the top of each chapter.

## 1.6 License



---

*First chapter written: 2026. This version: 2026.*

# Part I

# Part I — The Taxonomy Problem

# 2 The Predictable Tail

Lottery Characteristics Price Drawdown, Not Return

**SSRN:** abstract_id=6899979 **Status:** Live preprint

---

## 2.1 Abstract

We test whether payoff asymmetry (PA) — derived from the Hill estimator of individual equity return distributions — predicts forward equity returns when computed prospectively over 106,163 stock-month observations (715 U.S. equities, 2009–2025). Using distribution-free statistics appropriate for fat-tailed data, we find: (1) PA is a monotonic, symmetric, *inverted* predictor — low-PA names outperform with a 63.6% win rate and +4.90% 60-day median, while high-PA names underperform; (2) the same tail class (SUPER_FAT, 2) produces 70.4% win at the bottom of PA and 34.8% at the top — tail class alone is insufficient, direction matters; (3) Fama-MacBeth regressions fail to detect the signal (t = 1.37) because it is nonlinear and tail-concentrated; (4) common backtests using current PA values are look-ahead-contaminated, overstating the apparent edge by 35+ percentage points. The mechanism is the equity lottery premium: high-PA stocks exhibit positive-skew, negative-drift payoff structures; low-PA stocks exhibit positive drift with occasional large left-tail events.

---

## 2.2 1. Introduction

The practitioner intuition is intuitive: high payoff asymmetry — a return distribution with large upside relative to downside — should predict positive future performance. A stock whose historical gains have consistently exceeded its historical losses in magnitude sounds like a stock worth holding.

The data says the opposite.

This chapter documents the inversion and its mechanism. It also documents a methodological trap: computing payoff asymmetry from *current* data and testing it against *past* outcomes — a form of look-ahead contamination that inflates apparent predictive power by 35 percentage points and flips the sign of the conclusion.

The inversion is not a curiosity. It is a consequence of the equity lottery premium (Boyer et al. 2010): stocks with lottery-like payoff structures (large right tail, positive skew) attract speculative demand that raises prices and depresses subsequent returns. Payoff asymmetry, correctly computed prospectively, identifies this structure in real time.

## 2.3 2. The Payoff Asymmetry Metric

**Definition.** For a return series $\{r_t\}$ over a trailing $T$-day window:

$$PA = \frac{E[r \mid r > 0]}{E[|r| \mid r < 0]}$$

A stock with PA $> 1$ has average gains exceeding average losses in magnitude. A stock with PA $< 1$ has average losses exceeding average gains.

PA is computed from a 756-day (3-year) trailing window of daily returns, prospectively — using only data available before the evaluation date. The tail class (THIN, SUBEXP, FAT, SUPER_FAT) is derived from the Hill estimator applied to the same window.

## 2.4 3. Main Results

Table 2.1: Full-distribution PA results (N=16,948, 2009–2025, prospective, look-ahead-clean)

| PA Group | N | Win% | Median 60d Return |
|---|---|---|---|
| Bottom 2.5% | 423 | 66.7% | +6.14% |
| Bottom 10% | 1,694 | 63.6% | +4.90% |
| Baseline | — | 59.7% | — |
| Top 10% | 1,695 | 53.3% | +1.91% |
| Top 1% | 170 | 40.6% | −4.60% |

The bottom-10% 95% CI [61.2%, 65.9%] does not overlap with baseline [59.0%, 60.4%]. Mann-Whitney bottom 10% vs top 10%: p < 10 .

Within SUPER_FAT ( 2): bottom 10% PA → 70.4% win (+7.82% median); top 10% PA → 34.8% win (−8.76% median). The same tail class, opposite outcomes.

## 2.5 4. Why Standard Methods Miss It

On the same 112,152 observations (T = 202 monthly Fama-MacBeth periods):

- Spearman rank correlation (median-like, robust): PA return = **−0.045, p < 10**
- Linear Fama-MacBeth (conditional mean, OLS): PA coefficient t = +1.74 → +0.13 in the full model (effectively zero)

OLS fits the conditional mean. Fat tails inflate the mean at the high-PA extreme (top 1%: median −5.04% but mean +8.72%). Rank methods neutralize the moonshots and reveal the typical outcome. This is Taleb (2025) §3.4 Consequences 2 and 7 demonstrated simultaneously on real equity data.

## 2.6 5. The Look-Ahead Contamination

Computing PA from current data and testing against past outcomes creates an artifact: a stock that has already risen shows high PA and high realized returns simultaneously. This is circular.

| Method | Win Rate (top decile) |
|---|---|
| Contaminated (current PA, past outcomes) | 76–100% |
| Prospective (PA before evaluation date) | 40.6% |
| Contamination | ~35 percentage points |

The contaminated result is what most practitioners compute. The prospective result is what actually occurs.

## 2.7 6. Conclusion

Payoff asymmetry predicts forward equity returns — in the opposite direction from naive expectation. The mechanism is the equity lottery premium. Standard linear factor models cannot detect the signal because it is nonlinear and tail-concentrated. Common backtest implementations are contaminated by look-ahead bias that inflates apparent win rates by 35+ percentage points and reverses the conclusion.

The implication for the taxonomy developed in subsequent chapters: measuring the tail exponent alone is insufficient. The direction of payoff asymmetry within a tail class is the economically relevant variable.

---

*→ Chapter 2 explores why this direction varies systematically across identifiable subgroups within the SUPER_FAT class.*

# 3 Not All Lottery Stocks Are Equal

Source-of-Tail as a Missing Dimension in the Fat-Tail Taxonomy

**SSRN:** Pending review (submitted 2026-06-29) **Status:** Under editorial review

---

## 3.1 Abstract

The Taleb (2025) taxonomy classifies equities by tail exponent: THIN, SUBEXP, FAT, and SUPER_FAT ( 2). Among SUPER_FAT tickers (n = 78 in our 715-equity study universe), we identify a 13.7 percentage-point spread in payoff-weighted expected value between two economically distinct subpopulations that the taxonomy treats as equivalent: FAT_STRUCTURAL stocks (n = 70), whose extreme returns arise from a continuous power-law tail, achieve +7.74% expected value per signal with skill score +0.423; FAT_BINARY stocks (n = 8), whose extreme returns arise from the resolution of discrete binary events (FDA decisions, clinical trial readouts), achieve −5.96% with skill score −0.922. Binary win rates are approximately 49% for both groups — the Hill estimator generates the same classification for populations that are empirically indistinguishable by direction but have opposite economic outcomes.

Source-of-tail classification is not an extension of the existing taxonomy. It is a prerequisite for determining whether the Hill estimator's output is meaningful at all.

---

## 3.2 1. The Binary Event Problem

Not all extreme equity returns have the same origin. For most companies, the largest return realizations are draws from a continuous distribution — the structural right tail of the return process, driven by compounding of good outcomes. For a specific and identifiable class of companies, the largest returns are instead the resolution of discrete binary events of large magnitude: a Phase 3 clinical trial succeeds or fails; an FDA decision is approved or rejected.

These two generating mechanisms produce statistically similar tail exponents when measured by the Hill estimator, but they are economically and probabilistically distinct.

## 3.3 2. Taxonomy Extension: FAT_STRUCTURAL and FAT_BINARY

We define the mixture process:

$$X = (1 - B) \cdot Z_S + B \cdot J$$

where $B \sim \text{Bernoulli}(p)$, $Z_S$ is drawn from a continuous power-law distribution with tail exponent $\alpha_S$, and $J$ is a random variable representing the binary event payoff.

**FAT_STRUCTURAL:** The top-$k$ order statistics are dominated by draws from $Z_S$. The Hill estimator is consistent for $\alpha_S$.

**FAT_BINARY:** The top-$k$ order statistics contain a non-negligible fraction of $J$ realizations (see Chapter 3 for the formal inconsistency argument). The Hill estimator converges to a quantity depending on $(\alpha_S, p, J^+)$ that is not separately identified from order statistics alone.

## 3.4 3. Empirical Results

| Subgroup | n | EV per signal | Skill score | Binary win rate |
|---|---|---|---|---|
| FAT_STRUCTURAL | 70 | +7.74% | +0.423 | ~49% |
| FAT_BINARY | 8 | −5.96% | −0.922 | ~49% |

| Subgroup | n | EV per signal | Skill score | Binary win rate |
|---|---|---|---|---|
| **Spread** | | **13.7pp** | **1.345** | **~0** |

The binary win rates are statistically indistinguishable between groups. The Hill estimator cannot see the difference. The economic outcomes are opposite.

> **Caution**
>
> The FAT_BINARY result (n = 8) provides directional evidence only at this sample size. No magnitude claim should be made without confidence intervals. The structural result (n = 70) is on firmer statistical ground.

## 3.5 4. Sector Composition

The FAT_STRUCTURAL subgroup is dominated by Financials (25%), Energy (13%), Utilities (11%), and Real Estate (10%) — rate-sensitive, value-oriented names with left-tail risk and positive drift. The FAT_BINARY subgroup is dominated by Healthcare (biopharmaceutical names facing binary regulatory events).

This sector decomposition is the mechanism: the taxonomy conflates two economically opposite types of stock under a shared statistical label.

## 3.6 5. Implications

The consequence is not merely taxonomic. Risk models that apply uniform tail-class multipliers to all SUPER_FAT stocks are mis-specified. The appropriate multiplier for FAT_STRUCTURAL and FAT_BINARY stocks differs in sign, not just magnitude.

The formal mathematical argument for why the Hill estimator cannot distinguish these cases is developed in Chapter 3.

---

*→ Chapter 3 provides the mathematical foundation: why the Hill estimator is inconsistent for the structural tail exponent when the return process contains a binary-event component.*

# Part II

# Part II — The Estimation Problem

# 4 The Hill Estimator's Domain of Validity

Source-of-Tail Classification Reveals Systematic Misclassification in Power-Law Tail Analysis

**SSRN:** abstract_id=7022640 **Status:** Live preprint

---

## 4.1 Abstract

The Hill (1975) estimator is the canonical statistic for characterizing power-law tail behavior, and it underpins the tail class taxonomy formalized by Taleb (2025): THIN, SUBEXP, FAT, and SUPER_FAT. This estimator is consistent under a well-defined regularity condition: the survival function must satisfy

$$\lim_{u\to\infty} \frac{1-F(ux)}{1-F(u)} = x^{-\alpha} \quad \text{for all } x > 0$$

We argue that this condition is violated, in the practically relevant finite-sample sense, for an identifiable class of equities: those whose extreme return realizations arise predominantly from the resolution of discrete binary events of large magnitude. For such stocks — biopharmaceutical companies facing FDA decisions and clinical trial readouts are the canonical example — the top-$k$ order statistics used by the Hill estimator contain a non-negligible fraction of discrete jump observations that do not vanish as $n$ grows with $k = \lfloor 0.10n \rfloor$. The resulting $\hat{\alpha}$ is not a consistent estimator of the structural tail exponent $\alpha_S$; it converges to a quantity that depends on both $\alpha_S$ and the jump parameters $(p, J^+)$ in a way that is not separately identified from order statistics alone.

The consequence is systematic misclassification within the Taleb (2025) taxonomy. We demonstrate the empirical cost in a panel of 106,163 stock-month observations (715 U.S. equities, 2009–2025): a 13.7 percentage-point expected-value spread between FAT_STRUCTURAL and FAT_BINARY stocks that $\hat{\alpha}$ treats as identical. Source-of-tail classification is not an extension of the taxonomy — it is a prerequisite for determining whether the Hill estimator should be applied at all.

---

## 4.2 1. The Hill Estimator and Its Domain

### 4.2.1 1.1 The Estimator

Given an i.i.d. sample $X_1, \ldots, X_n$ with order statistics $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$, the Hill estimator with threshold parameter $k$ is:

$$\hat{\alpha} = \left( \frac{1}{k} \sum_{i=1}^{k} \ln X_{(i)} - \ln X_{(k+1)} \right)^{-1}$$

In our application, $k = \lfloor 0.10n \rfloor$ — the top 10% of return realizations.

### 4.2.2 1.2 Consistency Requires Regular Variation, Not Absolute Continuity

A key clarification: the regular variation condition

$$\lim_{u \to \infty} \frac{1 - F(ux)}{1 - F(u)} = x^{-\alpha}$$

does **not** require the distribution $F$ to be absolutely continuous. A discrete atom at a finite fixed point $J^+ < \infty$ does not violate regular variation asymptotically — for $u > J^+$, the atom contributes nothing to the tail ratio.

The problem is finite-sample, not asymptotic. The atom at $J^+$ contamination the top-$k$ order statistics at any fixed $k/n$ fraction, because the jump observations occupy stable positions in the ranking and do not vanish as $n \to \infty$.

### 4.2.3 1.3 The Mixture Process

Let the return process be:

$$X = (1 - B) \cdot Z_S + B \cdot J$$

where: - $B \sim \text{Bernoulli}(p)$, $p > 0$ - $Z_S \sim F_S$ with $F_S$ satisfying regular variation with exponent $\alpha_S$ - $J$ is a random variable with $P(J = J^+) = \pi$, $J^+ > 0$

## 4.3 2. The Inconsistency Proposition

**Proposition 4.1.** ***Proposition (Hill Estimator Inconsistency Under Mixture).*** *Under the mixture process above, with $k = \lfloor 0.10n \rfloor$ and $p > 0$:*

*(i) As $n \to \infty$, a positive fraction of the top-$k$ order statistics are $J$-component observations with probability approaching 1.*

*(ii) The Hill estimator $\widehat{\alpha}$ converges to a limit that depends on $(\alpha_S, p, J^+)$ rather than on $\alpha_S$ alone.*

*(iii) $\widehat{\alpha}$ is not a consistent estimator of $\alpha_S$; the estimand is a function of all three parameters and is not separately identified from order statistics alone.*

**Proof sketch.** The fraction of top-$k$ observations from the binary-jump component converges to a positive constant as $n \to \infty$ with $k = \lfloor 0.10n \rfloor$ fixed — it does not vanish. Each such observation contributes to the Hill sum a term $\ln J^+ - \ln X_{(k+1)}$ rather than the structural-tail term. The resulting limit is a function of $(\alpha_S, p, J^+)$. Since $p$ and $J^+$ are not separately identified from the order statistics $X_{(1)}, \ldots, X_{(k)}$ alone, consistency for $\alpha_S$ fails. □

**Differentiation from prior literature.** Cont and Tankov (2004) and Aït-Sahalia and Jacod (2009) address jump detection in continuous-time processes — separating continuous and jump components. Our claim is distinct: we do not argue that jumps can be detected; we argue that the Hill estimator applied to a mixture process produces an inconsistent estimator for the structural tail exponent, regardless of whether jumps are detected or not.

## 4.4 3. Empirical Evidence

### 4.4.1 3.1 Identification

FAT_BINARY stocks are identified by sector and catalyst type: - Biopharmaceutical companies with pending FDA decisions or Phase 3 trial readouts - Companies with outstanding regulatory binary events of large magnitude

In our 715-equity universe, n = 8 FAT_BINARY and n = 70 FAT_STRUCTURAL within the SUPER_FAT class (  2).

### 4.4.2 3.2 The 13.7pp Spread

| Classification | n | EV per signal | Hill class |
|---|---|---|---|
| FAT_STRUCTURAL | 70 | +7.74% | SUPER_FAT |
| FAT_BINARY | 8 | −5.96% | SUPER_FAT |

The Hill estimator assigns identical tail-class labels to both groups. Binary win rates are approximately 49% for both — the estimator is blind to the distinction.

### 4.4.3 3.3 Binary Win Rates as a Diagnostic

The binary win rate is the fraction of forward 60-day windows in which the return is positive. A binary win rate of ~49% for both FAT_STRUCTURAL and FAT_BINARY confirms that the estimator cannot distinguish the populations directionally.

## 4.5 4. What This Does Not Claim

This chapter does not claim that:

1. Every FAT_BINARY stock is a bad investment. The n = 8 result provides directional evidence only; individual FAT_BINARY names may have favorable characteristics.
2. The Hill estimator is wrong in general. It is consistent under regular variation; the critique is limited to the mixture case with a non-vanishing binary-event component.

3. Regular variation is violated asymptotically by binary events. The violation is finite-sample — practically relevant at n  750 daily observations (3 years), which is the estimation window used throughout.

## 4.6 5. Implications for the Taleb Taxonomy

Taleb (2025) (§18, p. 366) derives the Hill estimator and applies it to classify return distributions into THIN, SUBEXP, FAT, and SUPER_FAT. The taxonomy is correct for distributions satisfying regular variation with a continuous structural tail. The extension proposed here — adding a source-of-tail dimension — is not a critique of the taxonomy itself. It is a prerequisite for knowing when the estimator that builds the taxonomy is applicable.

---

*→ Chapter 4 addresses a separate methodological problem: what Newey-West standard errors do and do not fix when rolling-window statistics are evaluated against overlapping outcome windows.*

# 5 Two Problems, One Incomplete Fix

Newey-West Corrects for Overlap Autocorrelation But Not Overlap Bias, and Fat Tails Make the Uncorrected Problem Larger

**SSRN:** Pending review (submitted 2026-06-29) **Status:** Under editorial review

---

## 5.1 Abstract

The standard remedy for rolling-window statistics evaluated against overlapping outcome windows is to apply Newey-West HAC standard errors. This remedy is correct but incomplete. It addresses one of two distinct problems introduced by overlapping estimation and outcome windows: serial correlation inflation, which biases $t$-statistics upward. It does not address the second problem — directional bias in the point estimate, which biases the estimand itself in a predictable direction. These are separate problems requiring separate remedies.

We formalize this decomposition for the payoff asymmetry statistic $PA = E[r \mid r > 0]/E[|r| \mid r < 0]$, estimated over a trailing 756-day window of daily returns evaluated against 60-day forward outcomes (overlap fraction $\varphi = 60/756 \approx 7.9\%$). Using 106,163 stock-month observations across 715 U.S. equities from 2009 to 2025, we demonstrate that Problem 1 (serial correlation) is correctly addressed by Newey-West with $L = \lceil 60/21 \rceil = 3$ lags and effective sample size $n_{\text{eff}} \approx 28{,}240$; Problem 2 (directional bias) shifts the bottom-decile win rate by $-12.9$ percentage points and the top-decile win rate by $+3.6$ percentage points, compressing a prospective $+11.3$ pp spread to a $-5.2$ pp reversed spread — and is not corrected by any standard error adjustment.

Under fat-tailed return distributions (median $\hat{\alpha} = 2.69$, FAT class), Problem 2 is amplified beyond the $O(\varphi)$ first-order prediction by a factor we calibrate at approximately 1.6×. The mechanism is not that fat-tailed extremes are larger in absolute value than Gaussian extremes (they need not be at $\alpha = 2.69$), but that a single extreme contaminating observation contributes disproportionately to the ratio statistic relative to the typical observation.

Researchers who apply Newey-West and treat the overlap issue as resolved cannot determine from that result alone whether significance reflects genuine signal, Problem 1, or Problem 2.

---

## 5.2 1. Two Problems

### 5.2.1 1.1 Problem 1: Serial Correlation

When estimation windows overlap, successive observations of the statistic share data. For a 756-day trailing window evaluated monthly, consecutive monthly estimates share approximately 755/756 of their underlying data. This creates serial correlation in the test statistic that inflates the effective $t$-ratio.

**Problem 1 is well understood and correctly addressed by Newey-West.**

Hansen and Hodrick (1980) and Newey and West (1987) establish the correct HAC standard error for Problem 1. With overlap $h$ days and a monthly observation frequency (21 trading days), the appropriate lag order is $L = \lceil h/21 \rceil$. At $h = 60$ days: $L = 3$ lags. The effective sample size is $n_{\text{eff}} \approx n/(1 + 2 \times \rho_1/(1 - \rho_1))$ where $\rho_1 = 0.58$ is the empirically observed first-order autocorrelation. At our sample size: $n_{\text{eff}} \approx 28{,}240$.

### 5.2.2 1.2 Problem 2: Directional Bias in the Estimand

Overlap between the estimation window (used to compute PA) and the outcome window (used to evaluate the signal) introduces a directional bias in the point estimate itself. Observations in the overlap period contribute to both the PA computation and the outcome evaluation — creating a mechanical positive correlation when the outcome is positive and a mechanical negative correlation when the outcome is negative.

**Problem 2 is not addressed by Newey-West or any standard error adjustment.** Standard errors describe the precision of an estimate; they do not correct for systematic bias in what is being estimated.

## 5.3 2. Formal Decomposition

Let $PA_t$ denote the payoff asymmetry computed from the trailing 756-day window ending at $t$, and let $r_{t+1:t+h}$ denote the $h$-day forward return.

The contaminated estimand is:

$$\widetilde{PA}_t = \frac{\frac{n^+_{\text{clean}}\bar{r}^+_{\text{clean}} + n^+_{\text{overlap}}\bar{r}^+_{\text{overlap}}}{n^+_{\text{clean}} + n^+_{\text{overlap}}}}{\frac{n^-_{\text{clean}}|\bar{r}^-_{\text{clean}}| + n^-_{\text{overlap}}|\bar{r}^-_{\text{overlap}}|}{n^-_{\text{clean}} + n^-_{\text{overlap}}}}$$

where "overlap" observations are the $h$ days shared between the estimation and outcome windows. The prospective estimand excludes the overlap period entirely.

**Proposition 5.1.** ***Proposition 1 (Direction of Bias).*** *Conditional on a positive 60-day outcome ($r_{t+1:t+h} > 0$), the overlap observations are drawn from a truncated positive distribution. The contaminated $\widetilde{PA}_t$ exceeds the prospective $PA_t$ in expectation, creating upward bias in the numerator. The reverse holds for negative outcomes. The net effect biases $\widetilde{PA}_t$ toward positive values when measured contemporaneously with positive outcomes.*

## 5.4 3. Empirical Magnitude

| Metric | Prospective | Contaminated | Shift |
|---|---|---|---|
| Bottom-decile win rate | 63.9% | 51.0% | −12.9 pp |
| Top-decile win rate | 52.6% | 56.2% | +3.6 pp |
| Spread (bottom − top) | +11.3 pp | −5.2 pp | **−16.5 pp (sign reversal)** |

The contaminated measurement reverses the sign of the spread. Newey-West corrects the standard error on the contaminated estimate — it does not restore the prospective result.

## 5.5 4. Fat-Tail Amplification

Under Gaussian assumptions, Problem 2 is an $O(\varphi)$ effect — approximately proportional to the overlap fraction $\varphi = 7.9\%$.

Under fat-tailed distributions with $\hat{\alpha} = 2.69$, we observe a 1.6× amplification beyond the $O(\varphi)$ first-order prediction.

The mechanism: a single extreme return in the overlap period contributes to the PA ratio statistic with a weight proportional to its magnitude relative to the average. Under fat tails, this relative weight is larger than under Gaussian assumptions because:

1. Extreme observations are more extreme relative to the distributional mean
2. Conditioning on a positive 60-day outcome truncates extreme losses from the denominator by an amount that grows as $\alpha$ decreases

> **Note**
>
> The amplification mechanism is about *relative influence on the ratio statistic*, not about the absolute size of extreme returns. At $\alpha = 2.69$ and horizon $h = 60$: $60^{1/2.69} \approx 4.6 < 60^{0.5} \approx 7.75$ — fat-tailed extremes are not larger in absolute value than Gaussian extremes at this exponent. The amplification is a consequence of ratio statistic behavior, not absolute magnitude.

## 5.6 5. Differentiation from Prior Literature

Valkanov (2003) and Boudoukh et al. (2008) address the variance-vs-bias distinction in long-horizon OLS regressions with persistent regressors. This chapter extends to:

1. **Ratio statistics** (not OLS coefficients) — PA is a ratio of conditional means, not a regression coefficient, and exhibits different contamination behavior
2. **Fat-tailed return distributions** — first empirical calibration of the amplification factor (1.6×) at an empirically observed $\hat{\alpha}$
3. **A practical diagnostic** — the sign reversal between prospective and contaminated spreads as a signature of Problem 2 (Problem 1 alone cannot reverse a sign)

Hansen and Hodrick (1980) and Newey and West (1987) establish the correct treatment for Problem 1. This chapter does not challenge that treatment — it argues that Problem 1 correction is routinely applied while Problem 2 is left unaddressed.

## 5.7 6. The Diagnostic

A simple diagnostic distinguishes Problem 1 from Problem 2:

1. **Compute the statistic prospectively** (excluding the overlap period from the estimation window). If the $t$-statistic collapses, Problem 2 was driving apparent significance.
2. **Apply Newey-West to the prospective result** (addressing Problem 1). If significance survives, the signal is real.

A Newey-West $t > 2$ on a contaminated estimate does not survive this two-step filter if Problem 2 was active. The 12.9pp bottom-decile shift in our data illustrates what "Problem 2 active" looks like empirically.

---

*→ Chapter 5 develops the Nine-Gate Framework: a diagnostic protocol for distinguishing genuine signal from measurement artifacts across multiple simultaneous threats to validity.*

# Part III

# Part III — The Inference Framework

# 6 The Nine-Gate Framework

A Diagnostic Protocol for Fat-Tail Signal Validation

**Status:** Draft in progress (`paper5_gauntlet_framework.md`)

---

**Chapter in progress**

This chapter is under active development. The framework is designed and the gates are defined; the full write-up is forthcoming. The working draft is available on request.

## 6.1 Overview

The preceding four chapters document specific methodological failures: look-ahead contamination, estimator inconsistency under mixture processes, and incomplete bias correction. A practitioner encountering a new signal needs a systematic protocol for checking which of these failures — and others — may be active simultaneously.

The Nine-Gate Framework is that protocol. Each gate is a binary pass/fail test. A signal must pass all nine gates before it can be described as validated under fat-tail conditions. The gates address:

| Gate | Threat addressed |
|---|---|
| Gate 1: Point-in-time discipline | Look-ahead contamination (Chapter 4) |

| Gate | Threat addressed |
|---|---|
| Gate 2: Prospective vs contaminated divergence | Problem 2 bias (Chapter 4) |
| Gate 3: Source-of-tail pre-classification | Hill estimator domain of validity (Chapter 3) |
| Gate 4: Distribution-free test statistics | Mean inflation under fat tails (Chapter 1) |
| Gate 5: Multiple-testing correction | Deflated Sharpe Ratio / PBO |
| Gate 6: Effective sample size | Serial correlation via Newey-West (Chapter 4) |
| Gate 7: Regime robustness | Single-regime artifacts |
| Gate 8: Mechanism plausibility | Economic mechanism, not data mining |
| Gate 9: Out-of-sample accumulation | Live forward validation |

The framework operationalizes the epistemological argument running through this monograph: a result that cannot survive these nine gates is not a finding — it is a measurement artifact.

---

*Full chapter forthcoming. Subscribe to the Substack for updates, or watch the GitHub repository for new commits.*

# Part IV

# Part IV — Forthcoming

# 7 Ergodicity Deficit Across Tail Classes

When the Peters Formula Breaks Down

**Status:** Planned — research literature surveyed, writing not yet begun

---

> **Chapter forthcoming**
>
> Literature survey complete (see project repository). Writing begins after Chapters 1–5 are finalized.

## 7.1 Motivation

Peters (2019) establishes that expected utility maximization — the foundation of standard portfolio theory — implicitly assumes ergodicity: that the time-average growth rate of a compounding process equals its ensemble average. Under geometric Brownian motion, this assumption fails, and the ergodicity deficit equals $\sigma^2/2$.

The Peters formula, however, assumes Gaussian dynamics. It is undefined for $\alpha \leq 2$ (where variance is infinite) and has not been calibrated to an empirically estimated tail exponent $\hat{\alpha}$ across any equity cross-section.

## 7.2 The Open Question

How does the ergodicity deficit change as a function of $\alpha$ in the range $(2, 4)$? Is there a regime transition at $\alpha = 2$ where the standard formula breaks down? Can the deficit be estimated directly from Hill estimator output?

Our 715-stock cross-section with measured $\hat{\alpha}$ per ticker provides the instrument to answer these questions empirically for the first time.

## 7.3 Key prior work

- Peters (2019) — ergodicity economics framework
- Peters (2011) — optimal leverage from non-ergodicity
- Gabaix et al. (2003) — empirical equity tail exponent $\alpha \approx 3$
- Gap: no paper bridges Peters + Gabaix. **Novelty: HIGH.**

# 8 Minimum Track Record Length Under Fat Tails

How Long Must Performance History Be Before Skill Is Declared?

**Status:** Planned — research literature surveyed, writing not yet begun

---

ℹ Chapter forthcoming

## 8.1 Motivation

Bailey and López de Prado (2012) derive the Minimum Track Record Length (MinTRL) — the number of observations needed to reject, at a given confidence level, the null hypothesis that a manager's Sharpe ratio is zero. The derivation assumes finite fourth moment ($\kappa > 4$).

Kao et al. (2025) proved in 2025 that the Sharpe ratio converges at rate $n^{1-2/\kappa}$ when $\kappa \in (2, 4)$ — slower than the $\sqrt{T}$ Gaussian rate — toward a stable distribution with tail index $\kappa/2$.

The practical table that follows from these results does not exist: "At tail exponent $\hat{\alpha}$ and significance level $p$, the minimum track record is $N$ years."

## 8.2 The Claim

At $\hat{\alpha} = 2.69$ (the median in our equity sample), the minimum track record required to declare skill at 95% confidence is substantially longer than Gaussian theory implies. The precise factor is derivable analytically from the Kao et al. (2025) convergence rate and verifiable against our cross-sectional $\hat{\alpha}$ distribution.

The implication is philosophical as much as practical: for a large fraction of the equity universe, the sample sizes available to any practitioner are insufficient to distinguish skill from luck at conventional significance levels — not because the practitioner is unlucky, but because the distribution does not permit it.

## 8.3 Key prior work

- Bailey and López de Prado (2012) — Probabilistic Sharpe Ratio and MinTRL
- Bailey and López de Prado (2014) — Deflated Sharpe Ratio
- Kao et al. (2025) — convergence rate under fat tails (critical new anchor)
- Vlasiuk (2025) — Lévy-stable scaling of performance metrics
- Gap: no explicit $N(\hat{\alpha}, \text{significance})$ table. **Novelty: HIGH.**

# 9 Maximum Drawdown Under Power Laws

What the Gaussian Analytical Result Cannot Tell You

**Status:** Planned — research literature surveyed, writing not yet begun

---

Chapter forthcoming

## 9.1 Motivation

Magdon-Ismail and Atiya (2004) derive the analytical distribution of maximum drawdown under Gaussian (Brownian motion) dynamics. The result is elegant: expected drawdown scales as $\sqrt{T}$ at zero drift, $\log T$ at positive drift, linearly at negative drift.

For power-law return distributions, no analogous analytical result exists. Vlasiuk (2025) derives single-period drawdown scaling under Lévy-stable dynamics ($\propto \sigma T^{1/\alpha}$) but not the full multi-period distributional law.

## 9.2 The Claim

The distribution of running maximum drawdown as a function of $\alpha$ is not known. Our breach_35 and breach_50 empirical data — the fraction of stocks in each tail class that breach 35% and 50% peak-to-trough drawdown within a 12-month window — provides an empirical anchor for a simulation study calibrated to our observed $\hat{\alpha}$ cross-section.

The FAT_STRUCTURAL vs FAT_BINARY sub-classification (Chapter 3) predicts that drawdown distributions within the SUPER_FAT class should differ: FAT_BINARY stocks have discrete jump-driven drawdowns; FAT_STRUCTURAL stocks have continuous power-law drawdowns. The distributional shapes differ even when $\hat{\alpha}$ is identical.

## 9.3 Key prior work

- Magdon-Ismail and Atiya (2004) — Gaussian analytical result
- Filimonov and Sornette (2015) — empirical intraday drawdown power laws; dragon-king outliers
- Goldberg and Mahmoud (2017) — CED as coherent risk measure (finite variance)
- Vlasiuk (2025) — single-period scaling under stable dynamics
- Gap: full multi-period distributional law under $\alpha$-stable dynamics. **Novelty: HIGH.**

# 10 Ontological vs Epistemic Extremes

Why FAT_BINARY Events Are Neither Black Swans Nor Power Laws

**Status:** Planned — target journal: *Synthese*

---

Chapter forthcoming

## 10.1 Motivation

Taleb (2007) distinguishes Black Swans — high-impact, unpredictable events — from ordinary risks. The distinction is primarily epistemic: Black Swans are outside the observer's model, not necessarily outside the true distribution.

Colander (2010) argues further that Taleb's Black Swan is epistemological (limited information) rather than ontological (inherently indeterminate reality). Knight (1921) introduced the foundational distinction between measurable risk and unmeasurable uncertainty.

Neither framework has been formally operationalized using extreme value theory applied to the specific case of binary-event-driven fat tails.

## 10.2 The Claim

FAT_BINARY extreme returns occupy a distinct ontological category from both:

1. **Gaussian extremes** — which are epistemically uncertain (low probability, small magnitude, known distribution)
2. **Power-law extremes** — which are ontologically fat-tailed (heavy distributional tail, known generating process)

FAT_BINARY extremes are **structurally predictable** in their event space (the trial succeeds or fails; the FDA approves or rejects) but **epistemically uncertain** in direction. They are not black swans — the event is announced in advance, the magnitude is approximately known, the probability structure is binary. Yet they produce the same $\hat{\alpha}$ as genuine power-law tails.

This constitutes a third category that existing typologies do not accommodate: the **known-structure, unknown-direction extreme**. Formalizing this category using EVT and the mixture process of Chapter 3 is the contribution of this chapter.

## 10.3 Key prior work

- Knight (1921) — risk vs uncertainty
- Colander (2010) — epistemic vs ontological black swans
- Hansen and Sargent (2008) — robust control as formal Knightian uncertainty
- Taleb et al. (2014) — precautionary principle under fat tails
- Gap: no formal EVT operationalization of the ontological/epistemic distinction. **Novelty: HIGH.**

# 11 Against Optimization

Why Fat Tails Make the Optimal Strategy Epistemically Unavailable

**Status:** Planned — target journal: *Erkenntnis* or *Synthese*

---

Chapter forthcoming

## 11.1 Motivation

Portfolio optimization — maximize expected utility subject to distributional constraints — requires that the distribution of returns be estimable from available data. Under Gaussian assumptions, moment estimators converge at rate $\sqrt{T}$ and the optimization problem is well-conditioned for realistic sample sizes.

Under fat-tailed distributions, this assumption fails systematically. Taleb (2025) (Ch. 3–4) proves that sample estimators of mean and variance converge far slower than $\sqrt{T}$ under power-law tails, making standard statistical inference unreliable at realistic sample sizes. Michaud (1989) showed that even under Gaussian assumptions, sample estimation error dominates the optimization, producing portfolios that are optimized for measurement noise rather than true parameters.

Under fat tails with $\hat{\alpha} \in (2, 3)$, the Michaud problem is dramatically worse.

## 11.2 The Claim

The claim is epistemological: optimization is not merely imprecise under fat tails — it is epistemically unavailable. The data required to close the estimation gap exceeds any realistic investment horizon.

This can be stated precisely. Chapter 7 establishes the minimum track record required to detect a given information ratio at significance level $p$ under tail exponent $\hat{\alpha}$. The analogous result for optimization holds: the minimum sample size to identify the optimal portfolio within a given tolerance is a function of $\hat{\alpha}$, and at $\hat{\alpha} = 2.69$, that minimum exceeds the observable sample by a derivable factor.

The rational response to this result is robustness — the barbell strategy that Taleb advocates — rather than optimization. This chapter provides the mathematical justification for that preference as a theorem, not an intuition.

## 11.3 Connection to the Precautionary Principle

The Taleb et al. (2014) precautionary principle argues that for systems with fat-tailed harm distributions, expected-value frameworks fail. Combined with the unavailability of optimization established here, PP implies a specific portfolio structure: maximize resilience to estimation error (robustness) rather than expected utility of an estimated distribution (optimization). The two arguments — PP and estimation unavailability — converge on the same structural recommendation from different directions.

## 11.4 Key prior work

- Taleb (2025) — statistical consequences of fat tails; moment estimation failure
- Michaud (1989) — estimation error dominates optimization under Gaussian assumptions
- Taleb et al. (2014) — precautionary principle under systemic fat-tail risk
- Peters (2019) — ergodicity and the failure of ensemble optimization (Chapter 6 connection)
- Gap: mathematical proof that optimization is epistemically unavailable (not merely imprecise) at empirically calibrated $\hat{\alpha}$. **Novelty: MEDIUM-HIGH.**

Aït-Sahalia, Yacine, and Jean Jacod. 2009. "Testing for Jumps in a Discretely Observed Process." *Annals of Statistics* 37 (1): 184–222.

Bailey, David H., and Marcos López de Prado. 2012. "The Sharpe Ratio Efficient Frontier." *Journal of Risk* 15 (2).

Bailey, David H., and Marcos López de Prado. 2014. "The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality." *Journal of Portfolio Management* 40 (5): 94–107.

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw. 2008. "The Myth of Long-Run Diversification." *Review of Financial Studies* 21 (4): 1481–508.

Boyer, Brian, Todd Mitton, and Keith Vorkink. 2010. "Expected Idiosyncratic Skewness." *Review of Financial Studies* 23 (1): 169–202.

Colander, David. 2010. "Black Swans and Knight's Epistemological Uncertainty." *Journal of Post Keynesian Economics* 32 (4): 567–70.

Cont, Rama, and Peter Tankov. 2004. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC.

Filimonov, Vladimir, and Didier Sornette. 2015. "Power Law Scaling and 'Dragon-Kings' in Distributions of Intraday Financial Drawdowns." *Chaos, Solitons & Fractals* 74: 27–45.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley. 2003. "A Theory of Power-Law Distributions in Financial Market Fluctuations." *Nature* 423: 267–70.

Goldberg, Lisa R., and Ola Mahmoud. 2017. "Drawdown: From Practice to Theory and Back Again." *Mathematics and Financial Economics* 11: 275–97.

Hansen, Lars Peter, and Robert J. Hodrick. 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates." *Journal of Political Economy* 88 (5): 829–53.

Hansen, Lars Peter, and Thomas J. Sargent. 2008. *Robustness*. Princeton University Press.

Kao, Linda, Cheng-Few Lee, and John Lee. 2025. "Estimated Sharpe Ratio of Asset Returns with Fat Tails: Theory and Empirical Evidence." *Review of Quantitative Finance and Accounting*.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Houghton Mifflin.

Magdon-Ismail, Malik, and Amir F. Atiya. 2004. "On the Maximum Drawdown of a Brownian Motion." *Journal of Applied Probability* 41: 147–61.

Michaud, Richard O. 1989. "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analysts Journal* 45 (1): 31–42.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–8.

Peters, Ole. 2011. "Optimal Leverage from Non-Ergodicity." *Quantitative Finance* 11 (11): 1593–602.

Peters, Ole. 2019. "The Ergodicity Problem in Economics." *Nature Physics* 15: 1216–21. https://doi.org/10.1038/s41567-019-0732-0.

Taleb, Nassim Nicholas. 2007. *The Black Swan: The Impact of the Highly Improbable*. Random House.

Taleb, Nassim Nicholas. 2025. *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. 3rd ed.

Taleb, Nassim Nicholas, Rupert Read, Raphael Douady, Joseph Norman, and Yaneer Bar-Yam. 2014. *The Precautionary Principle (with Application to the Genetic Modification of Organisms)*.

Valkanov, Rossen. 2003. "Long-Horizon Regressions: Theoretical Results and Applications." *Journal of Financial Economics* 68 (2): 201–32.

Vlasiuk, Dmytro. 2025. *Lévy-Stable Scaling of Risk and Performance Functionals*.