The Nature of Volatility Across Financial Instruments*

Agustín Shehadi Candela
QUAFI Research
research@quafi.io

August 2026

Abstract

We study the distribution of returns for 4,334 instruments across six asset classes — US equities, exchange-traded funds, US and Argentine fixed income, Argentine equities, and Argentine depositary receipts over foreign shares — using 9.2 million daily observations spanning ten years. Three findings organise the paper. First, the estimated tail index lies between 2.3 and 3.3 for *every* class: the variance exists everywhere and the fourth moment nowhere, which renders the F-test, the χ^2 variance test and the Jarque–Bera test inadmissible on this data. We report them alongside their robust counterparts to show the difference. Second, tail heaviness and volatility are distinct and differently ordered attributes: the two fixed-income classes have the lowest volatilities and the heaviest relative tails, while the most volatile class in the sample has the lightest. Argentine hard-dollar sovereigns dominate US equity on mean, volatility and Sharpe ratio simultaneously while carrying the panel's lowest tail index — a case in which the mean–variance ranking and the tail ranking point in opposite directions. Third, aggregation thins the tails monotonically but does not deliver normality by one month, and the mean reversion that would justify horizon-dependent allocation is modest where it exists (six-month variance ratios of 0.71–0.77 for equity-like classes) and absent in Argentine sovereigns. Accounting for the observed distribution rather than a Gaussian match reduces the optimal US equity allocation by 23% for a log-utility investor, and the welfare cost of the Gaussian assumption rises with risk aversion. We separate inference at the pooled and instrument levels throughout, and show that the two can disagree sharply: the comparison of US equity against Argentine fixed income yields $p \approx 10^{-90}$ pooled and $p = 0.42$ across instruments. We argue the second is the defensible number, and draw the implications for how risk and horizon profiles should be constructed.

*QUAFI Working Paper Series, No. 2026-05. Preliminary; circulated to inform and to invite comment. The views are the author's own. *Nothing herein is investment advice, and nothing herein is a recommendation to buy or sell any security.* All data, code and intermediate results required to reproduce every number and figure are described in Section 2 and Appendix A. Any errors are mine.

Keywords: return distributions; heavy tails; tail index; volatility clustering; aggregational Gaussianity; variance ratios; risk aversion; investment horizon.

JEL: C12, C14, C58, G11, G12, G15.

Contents

1 Introduction

Every quantitative portfolio method rests on an assumption about the shape of the return distribution, and most rest on the Gaussian. The assumption is convenient rather than believed: it makes variance a sufficient description of risk, makes the Sharpe ratio a sufficient ranking, and makes the square-root-of-time rule exact. That it is false has been known since Mandelbrot (1963) and Fama (1965), and the stylised facts were codified by Cont (2001). What is less settled is how the departure varies *across* instruments, and what an analyst is entitled to conclude from it about the portfolio a particular investor should hold.

This paper takes a single panel — a broad cross-section of listed securities, 4,334 instruments and 9.2 million daily returns over ten years — and asks three questions of it.

What do the distributions look like, class by class? Section 4 reports moments, tail indices, and the sub-distributions within US equity by industry. The central finding is uniformity where one might expect variety: the tail index falls between 2 and 4 for every class, across three countries and four contractual forms. Variance exists; kurtosis does not. The finding is not new in isolation — estimates in this range are standard for equity indices — but its constancy across such different instruments is worth establishing, and it has an immediate methodological consequence, since the tests most commonly used to compare variances derive their null distributions from a Gaussian fourth moment.

How do the distributions differ, and how would one know? Section 4.6 runs the comparisons. The exercise is complicated by two forms of dependence that most applied work handles casually: volatility clusters through time, and instruments within a class co-move. We address the first with a stationary bootstrap (Politis and Romano, 1994) and the second by running every comparison at two levels — pooling returns, and taking one observation per instrument. The two disagree in an instructive way. With millions of pooled observations every test rejects, and the rejection reflects sample size rather than economics; at the instrument level, comparisons involving small classes often cannot resolve differences that the pooled test declares overwhelming.

What follows for investor profiles? Sections 5 and 6 take up the question the paper was written to answer. Horizon enters allocation only if returns depart from independence (Samuelson, 1969), so we test for that departure directly rather than assuming it, and find it modest in equities and absent in Argentine sovereigns. Risk aversion enters through the curvature of utility, so we compute optimal allocations and certainty equivalents under the empirical distributions and under Gaussian matches, and find that the tails cost a log-utility investor 23% of their equity allocation.

The paper's contribution is not a new estimator. It is a systematic, like-for-like distributional comparison across instrument types that are usually studied separately, carried out with inference that respects the dependence in the data, and carried through to the allocation question with the limitations stated rather than elided. Where the evidence does not support a conclusion — and on horizon it largely does not — we say so.

A note on corrections. Three errors in our own analysis were caught during this work, two of which would have reversed a conclusion. They are documented in Section 5 and Appendix A rather than quietly fixed, because in each case the erroneous result was the more publishable one, and because the means of detection — a result too uniform to be economic, a statistic determined by a single observation — generalise beyond this paper.

2 Data

2.1 Universe

The starting universe is a cross-section of 4,503 listed securities drawn from a commercial investable universe, each classified by asset class, industry, listing currency, and — for depositary receipts — the underlying security, its market, its currency, and the conversion ratio. Six classes are analysed:

- **US Equity** — common stock listed on US exchanges, screened for liquidity.
- **ETF** — exchange-traded funds across equity, sector, commodity and thematic mandates.
- **US Fixed Income** — a curated set of bond ETFs spanning the duration and credit spectrum. Bond *funds* rather than individual bonds, because daily marks on single corporate issues are sparse and stale-priced; the fund wrapper delivers a genuinely traded daily price.
- **Argentine Fixed Income** — hard-dollar sovereign bonds.
- **Argentine Equity** — common stock listed on BYMA.
- **CEDEAR** — Argentine depositary receipts over foreign equity.

2.2 Prices

Daily closing prices covering ten years to August 2026 are assembled from public market-data sources. Prices for exchange-listed instruments are adjusted for splits and dividends; Argentine sovereign prices are the local market's own closes, obtained from a public Argentine market-data interface.

Three construction decisions materially affect the statistics and are stated explicitly.

Argentine fixed income is restricted to USD-settled instruments. The Argentine sovereign curve trades in several legs: hard-dollar bonds settling in USD, peso bonds, and CER-linked (inflation-adjusted) peso bonds. Only the hard-dollar legs are analysed. Converting the peso legs to dollars would require applying the parallel exchange rate, which would make the resulting series a joint bet on credit and on the currency regime — and, given the volatility of that rate, overwhelmingly the latter. A distribution so constructed would describe Argentine

FX, not Argentine fixed income. The exclusion narrows the class to the most liquid sovereign complex and keeps the object of study interpretable.

Argentine equities are dollarized. These instruments quote in pesos, and a peso return series confounds the asset with inflation. Prices are therefore converted at the "contado con liquidación" rate, the market-implied exchange rate estimated each day as the cross-sectional median over liquid dual-listed pairs of

$$\mathrm{CCL}_t = \mathrm{median}_j \frac{P^{\mathrm{local}}_{j,t} \times (\text{shares per ADR})_j}{P^{\mathrm{ADR}}_{j,t}}. \tag{1}$$

The median across pairs, rather than any single pair, guards against a stale quote on one leg. Returns are scale-invariant to a constant, so an error in the share ratio of any one pair affects the level and not the distribution.

CEDEARs are reconstructed from their underlying, not fetched directly. A CEDEAR is a claim on a fixed fraction of a foreign share, priced as $P^{\mathrm{origin}}_{\mathrm{USD}}$/ratio. The ratio is a constant and therefore cancels in log returns: a CEDEAR over a US stock has, up to microstructure noise, *the same return distribution as the stock itself*. Treating the two as independent instruments would populate the sample with several hundred near-duplicates and would inflate every pooled test statistic. We therefore build each CEDEAR as

$$r^{\mathrm{CEDEAR}}_t \;=\; r^{\mathrm{origin}}_t \;+\; r^{\mathrm{FX}}_t, \tag{2}$$

which makes the wrapper's contribution — the currency overlay for non-USD-origin receipts — the object of measurement rather than an artefact. Section 4 reports what that overlay costs in variance.

2.3 Filters

An instrument enters the analysis if it has at least 250 return observations (approximately one trading year) and its price exceeds one currency unit. The first filter is a precision requirement: below roughly one year, tail-index and kurtosis estimates are too noisy to report. The second removes quantisation artefacts in sub-unit prices, where a one-tick move is a large percentage return and the measured volatility describes the tick size rather than the asset.

Both filters are applied to *estimation*, not to selection: no instrument is excluded on the basis of its returns, so the filters cannot induce a survivorship or performance bias in the cross-section. The panel is unbalanced by construction — instruments list and delist within the window — and every statistic is computed per instrument before aggregation, so unequal histories do not silently reweight the results.

2.4 Coverage

Table 1 reports the resulting panel.

Table 1: Panel coverage by asset class.

Asset class	Instruments	Median obs.
US Fixed Income	32	2,513
Argentine Fixed Income	12	1,191
ETF	1,105	2,510
US Equity	2,782	2,513
CEDEAR	371	2,513
Argentine Equity	32	1,380
All	4,334	9,153,393 obs.

Daily observations, 2016-08-16 to 2026-08-14. Instruments with fewer than 250 returns are excluded.

Remark 1 (Survivorship). The instrument list is a snapshot of the universe as constituted in August 2026, so securities that delisted before that date are absent. This biases the *level* of average returns upward, in common with most cross-sectional studies built on a current universe. It does not obviously bias the objects of interest here — the shape of the return distribution, the tail index, the persistence of volatility — except insofar as failed firms have heavier left tails than survivors, which would make the tail estimates reported below conservative. We flag this rather than correct it: correcting it would require a point-in-time universe that was not available for this study.

3 Methods

Daily returns violate the assumptions behind almost every standard test in three specific ways: the tails are heavy enough that high moments may not exist, volatility is serially dependent, and instruments within a class are cross-sectionally correlated. Each of the three breaks a different part of the usual inferential machinery. This section states each estimator, the assumption it needs, and the replacement used where that assumption fails.

3.1 Returns and the treatment of gaps

For instrument i with closing prices $P_{i,t}$ observed on trading dates $t_1 < t_2 < \cdots < t_{n_i}$, the log return is

$$r_{i,k} \;=\; \log P_{i,t_k} - \log P_{i,t_{k-1}}, \qquad k = 2, \ldots, n_i. \tag{3}$$

Returns are computed on *consecutive observed* closes. Missing dates are not forward-filled. The distinction is not cosmetic: filling a market holiday with the previous close inserts a return of exactly zero, and a panel spanning several exchange calendars would accumulate a spike of

artificial mass at the origin. Every statistic in this paper that is sensitive to the centre of the distribution — kurtosis above all — would be biased toward "peaked" by that choice. The cost of the convention is that a return spanning a gap covers more than one calendar day; we accept the mild time deformation in exchange for not manufacturing observations.

3.2 Moments and their existence

For a sample $r_1, \dots, r_n$ we report the mean, standard deviation, the standardised third and fourth cumulants

$$\hat{S} = \frac{\frac{1}{n}\sum (r_t - \bar{r})^3}{\hat{\sigma}^3}, \qquad \hat{K} = \frac{\frac{1}{n}\sum (r_t - \bar{r})^4}{\hat{\sigma}^4} - 3, \tag{4}$$

and the downside semideviation $\hat{\sigma}_- = \mathrm{sd}\{r_t : r_t < 0\}$, the dispersion measure that corresponds to a loss-averse investor, for whom upside and downside variation are not equivalent.

A warning attaches to $\hat{K}$ throughout. If the return distribution has a regularly varying tail with index α, then $\mathbb{E}|r|^p < \infty$ only for $p < \alpha$. Where $\alpha < 4$ — which, as Section 4 reports, covers most of the equities here — the population fourth moment does not exist, $\hat{K}$ has no target to estimate, and it diverges as n grows rather than converging. We report it because the literature does, and because it remains a valid *description of the observed sample*; we do not treat it as an estimate of a population quantity, and we do not test hypotheses about it.

3.3 Tail index

Let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ be the descending order statistics of $|r_t|$. The Hill estimator (Hill, 1975) of the tail index using the k largest is

$$\hat{\alpha}_k^{-1} \;=\; \frac{1}{k}\sum_{i=1}^{k} \big[\log X_{(i)} - \log X_{(k+1)}\big], \tag{5}$$

the reciprocal of the mean log-excess over the $(k{+}1)$-th order statistic. It is the maximum-likelihood estimator of the Pareto shape parameter on the exceedances and has the smallest asymptotic variance in its class.

Its entire practical difficulty is the choice of k: small k gives high variance, large k admits observations from the centre and biases $\hat{\alpha}$ upward. No automatic rule commands agreement. We therefore report Hill *plots* — $\hat{\alpha}_k$ traced against k — and read a tail index only where the curve is flat. A single number would conceal exactly the sensitivity that matters. As an independent cross-check we fit a Student-t by maximum likelihood; for a t distribution the tail index equals the degrees of freedom ν exactly, so agreement between $\hat{\nu}$ and a stable $\hat{\alpha}$ is evidence from a different part of the sample.

The thresholds that matter: $\alpha < 2$ implies infinite variance, $\alpha < 4$ infinite kurtosis.

3.4 Normality

The Jarque–Bera statistic (Jarque and Bera, 1980)

$$\mathrm{JB} = \frac{n}{6}\left(\hat{S}^2 + \tfrac{1}{4}\hat{K}^2\right) \xrightarrow{d} \chi^2_2 \tag{6}$$

is reported for continuity with the literature, with two caveats. Its null distribution presumes finite fourth moments, which is the assumption in doubt; and at n in the millions it rejects any departure however small. It functions here as a descriptive index of non-normality, not as a test. The Anderson–Darling statistic (Anderson and Darling, 1952)

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)\,[1 - F(x)]}\,\mathrm{d}F(x) \tag{7}$$

is more informative for our purpose because the weight $1/[F(1-F)]$ concentrates where the data is sparse — in the tails — unlike the Kolmogorov–Smirnov statistic, which is most sensitive near the median.

3.5 Comparing two classes

Location. Welch's test (Welch, 1947) of $H_0 : \mu_a = \mu_b$,

$$t = \frac{\bar{r}_a - \bar{r}_b}{\sqrt{s_a^2/n_a + s_b^2/n_b}}, \qquad \nu = \frac{\left(s_a^2/n_a + s_b^2/n_b\right)^2}{\dfrac{(s_a^2/n_a)^2}{n_a - 1} + \dfrac{(s_b^2/n_b)^2}{n_b - 1}}, \tag{8}$$

is used rather than the pooled-variance t-test because the variances differ across classes by an order of magnitude, which is the subject of the paper. The central limit theorem protects the test against non-normality at these sample sizes; nothing protects it against the dependence in the panel, which motivates the permutation version below.

Scale. The variance-ratio $F = s_a^2/s_b^2$ referred to $F_{n_a-1,\,n_b-1}$, and the one-sample χ^2 statistic $(n-1)s^2/\sigma_0^2 \sim \chi^2_{n-1}$, are both reported *in order to demonstrate their failure*. Their null distributions are derived under normality and are acutely sensitive to kurtosis; with the excess kurtosis observed here their true size is far above nominal, and a rejection carries no information about variances. The robust counterpart is Brown–Forsythe (Brown and Forsythe, 1974), a one-way ANOVA on the absolute deviations from the group *median*,

$$z_{ij} = |r_{ij} - \mathrm{med}_j(r)|, \tag{9}$$

which replaces Levene's mean-centring (Levene, 1960) and is markedly more robust under heavy tails. Brown–Forsythe is the scale test we quote.

The whole distribution. Two-sample Kolmogorov–Smirnov $D = \sup_x |F_a(x) - F_b(x)|$, the k-sample Anderson–Darling test of Scholz and Stephens (1987), and the Cramér–von Mises criterion are computed. Because every one of them rejects at our sample sizes, the reported quantity of substantive interest is the Wasserstein-1 distance

$$W_1(F_a, F_b) = \int_{-\infty}^{\infty} \left| F_a(x) - F_b(x) \right| \mathrm{d}x, \tag{10}$$

which is not a test but an *effect size* denominated in return units and therefore directly interpretable.

Permutation. Under H_0 the two samples are exchangeable, so the labels are uninformative. Recomputing the statistic over B random relabellings traces its exact null distribution with no distributional assumption at all, and the two-sided p-value is

$$\hat{p} = \frac{1 + \#\{b : |\theta^{(b)}| \geq |\hat{\theta}|\}}{1 + B}. \tag{11}$$

This is the appropriate tool when the fourth moment may not exist, and it is what the paper quotes for differences in means and in log-variances.

3.6 Dependence and multiplicity

Two features of the panel invalidate naive standard errors.

Serial dependence. Volatility clusters, so returns are not independent through time. Confidence intervals for time-series statistics use the stationary bootstrap of Politis and Romano (1994): blocks of Geometric$(1/b)$ length are resampled with replacement and wrapped circularly, with mean block length $b = 21$ trading days, comfortably beyond the horizon over which the autocorrelation of $|r_t|$ remains material. Random block lengths preserve stationarity of the resampled series, which the fixed-length moving-block bootstrap does not.

Cross-sectional dependence. Instruments within a class share factors. Pooling their returns and treating the result as one i.i.d. sample overstates the information content by roughly the number of instruments. We therefore run every comparison at two levels: *pooled*, stacking all daily returns of a class, and *instrument-level*, taking one summary statistic per instrument and comparing those distributions across classes with the Mann–Whitney U test. The instrument-level analysis has a sample size in the hundreds rather than the millions and does not pretend that co-moving assets are independent draws. Where the two levels disagree we follow the instrument level and say so.

Multiplicity. With $\binom{6}{2} = 15$ pairwise class comparisons and several tests each, we control the false discovery rate at $q = 0.05$ by the step-up procedure of Benjamini and Hochberg (1995): reject $H_{(i)}$ for all $i \leq \max\{i : p_{(i)} \leq iq/m\}$. FDR rather than family-wise control because the tests are strongly positively dependent, under which Bonferroni is needlessly conservative.

3.7 Horizon

Non-overlapping h-day returns are formed as $r_j^{(h)} = \sum_{k=(j-1)h+1}^{jh} r_k$. Non-overlapping windows are used deliberately: overlapping ones inflate the effective sample by a factor of h while inducing MA$(h-1)$ dependence, which makes standard errors for the aggregated moments appear far tighter than they are.

Under i.i.d. increments the excess kurtosis of the h-period return decays as

$$\kappa_h = \kappa_1 / h, \tag{12}$$

so the observed decay measured against (12) is a direct test of aggregational Gaussianity (Cont, 2001) and of the dependence that slows it.

The variance ratio of Lo and MacKinlay (1988),

$$\mathrm{VR}(q) = \frac{\mathrm{Var}(r_t^{(q)})}{q\,\mathrm{Var}(r_t)}, \tag{13}$$

equals one under a random walk, falls below one under mean reversion and rises above one under momentum. We use the heteroskedasticity-consistent statistic $z^*(q)$, since the homoskedastic version rejects the random walk merely because volatility clusters. This is the empirical hinge of the horizon question: under i.i.d. returns Samuelson's theorem (Samuelson, 1969) makes the optimal risky share independent of horizon, so any horizon-dependent advice must be earned by $\mathrm{VR}(q)$ departing from unity.

3.8 From distributions to preferences

For constant relative risk aversion $u(W) = W^{1-\gamma}/(1-\gamma)$, the annualised certainty equivalent of a return distribution is

$$\mathrm{CE}(\gamma) = \Big(\mathbb{E}\big[(1+R)^{1-\gamma}\big]\Big)^{\frac{1}{1-\gamma}\cdot 252} - 1, \qquad \gamma \neq 1, \tag{14}$$

with the logarithmic case $\mathrm{CE}(1) = \exp\{252\,\mathbb{E}[\log(1+R)]\} - 1$. Expectations are taken over the *empirical* distribution. Comparing this against the same functional evaluated on a Gaussian matched in mean and variance isolates the economic cost of assuming normality, and that cost is increasing in γ because a more risk-averse investor weights the left tail the Gaussian approximation fails to represent.

The optimal share w^* in a single risky asset is obtained by maximising $\mathbb{E}\,u\big((1+r_f)+w(R-r_f)\big)$ over a grid, again under the empirical law. The Gaussian benchmark is the textbook $w^*_{\mathrm{G}} = (\mu - r_f)/(\gamma\sigma^2)$; the gap between the two is the allocation error induced by ignoring higher moments.

4 What the distributions look like

4.1 First and second moments

Table 2 reports the central moments, computed per instrument and summarised by the median within each class.

Table 2: Descriptive moments by asset class.

	Mean (% p.a.)	Volatility (% p.a.)	Vol. IQR (% p.a.)	Semidev. (% p.a.)	Sharpe	Max DD (%)
US Fixed Income	-0.8	5.1	5.1	4.4	-0.16	-21.3
Argentine Fixed Income	13.5	38.7	1.8	29.5	0.35	-49.9
ETF	6.5	20.4	16.4	16.3	0.32	-39.0
US Equity	6.5	43.1	27.9	33.2	0.16	-69.6
CEDEAR	8.6	35.6	22.9	27.7	0.24	-62.0
Argentine Equity	-3.2	57.0	13.4	38.5	-0.07	-77.3

Median across instruments within each class; the interquartile range of volatility indicates dispersion *within* the class.

Two features deserve comment before anything else.

First, the ordering by volatility is not the ordering an intuitive risk ranking would produce. Argentine fixed income sits at 38.7% annualised, between the ETF class at 20.4% and US equity at 43.1%. A "bond" in this sample is not a low-volatility instrument; the label describes the contractual form, not the statistical behaviour. Meanwhile US fixed income — bond funds spanning the duration and credit spectrum — records a *negative* median mean return over the decade, an artefact of a sample that begins near the end of a forty-year decline in yields and ends after the 2022–2023 tightening.

Second, and more consequential for the paper's argument: Argentine fixed income dominates US equity on every mean–variance statistic simultaneously. Its median annualised return is 13.5% against 6.5%, its volatility is *lower* (38.7% against 43.1%), and its Sharpe ratio is more than double (0.35 against 0.16). An investor ranking assets on the mean–variance criterion alone would prefer the Argentine sovereign complex to the entire US stock market. Section 4.2 explains why that ranking is misleading, and Section 6 quantifies by how much.

Figure 1 shows the full distributions. The violin outlines make the point that a table of moments cannot: the classes differ far more in the extent of their tails than in the width of their central mass.

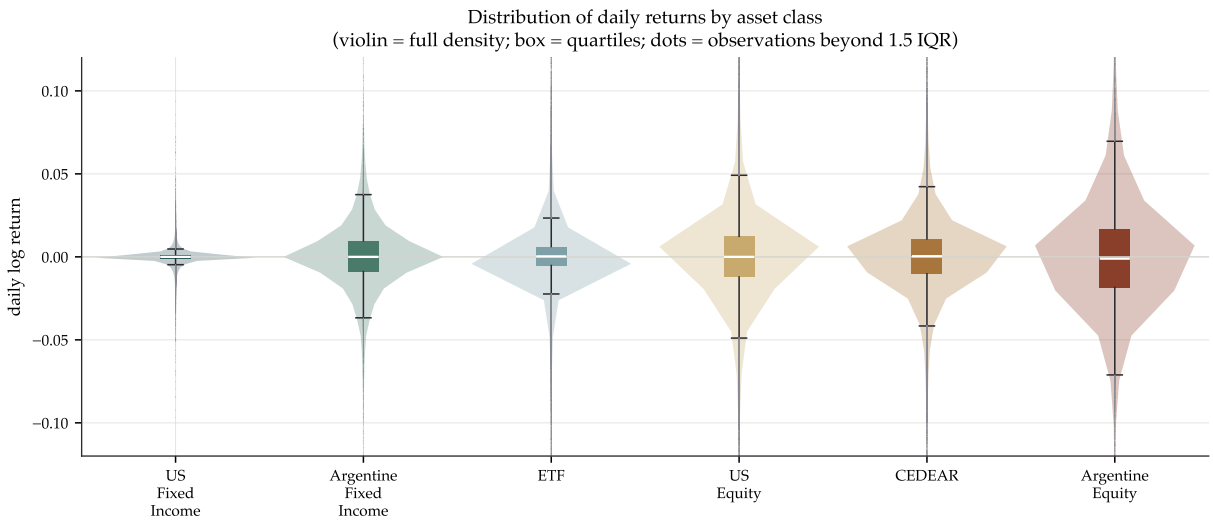

Figure 1: Distribution of daily log returns by asset class. Violins show the full kernel density; boxes the interquartile range; points every observation beyond 1.5 IQR. The whiskers, not the boxes, carry the information.

4.2 Shape: skewness, kurtosis, and the failure of the fourth moment

Table 3: Distributional shape and tail index.

Asset class	Excess kurtosis				Tail index		
	Skew	Pooled	95% CI	Median	$\hat{\alpha}$	$\hat{\alpha}$ med.	$\hat{\nu}_t$
US Fixed Income	-0.34	31.3	[22.3, 40.5]	25.3	3.25	2.90	1.93
Argentine Fixed Income	-1.82	60.3	[14.0, 134.2]	14.5	2.48	2.35	2.08
ETF	-4.29	275.8	[62.9, 1457.0]	10.1	3.15	3.06	1.43
US Equity	-2.33	138.3	[37.4, 137.3]	12.1	2.80	2.86	2.24
CEDEAR	-0.36	31.9	[23.5, 43.5]	9.0	3.65	3.10	2.28
Argentine Equity	-5.92	329.9	[23.2, 704.5]	8.1	2.92	3.29	3.54

Pooled stacks all daily returns of the class; *Median* is the median across instruments. Kurtosis intervals are stationary-bootstrap (mean block 21 days). $\hat{\alpha}$ is the Hill estimate at $k = \sqrt{n}$; $\hat{\nu}_t$ is the maximum-likelihood Student-t degrees of freedom, which equals the tail index for that family. Values below 4 imply an undefined population kurtosis — the column to its left then describes the sample only.

Table 3 contains the paper's central empirical result, and it is worth stating plainly: **every asset class in this sample has an estimated tail index between 2 and 4.** The Hill estimates run from 2.35 (Argentine fixed income) to 3.29 (Argentine equity); the maximum-likelihood Student-t degrees of freedom, estimated independently, agree in ordering and sit in the same range. The finding holds across three countries, four contractual forms and two currencies of denomination.

The consequence is not a technicality. $\alpha > 2$ means the variance exists, so the mean–variance apparatus has something to estimate. $\alpha < 4$ means the *fourth* moment does not, and therefore:

- the sample excess kurtosis reported in Table 3 has no population target and grows with the sample rather than converging — which is why the bootstrap intervals are so wide, and why the pooled and per-instrument medians disagree by an order of magnitude;
- the F-test for equality of variances and the χ^2 test for a single variance are inadmissible on this data, since both derive their null distributions from a Gaussian fourth moment;
- the Jarque–Bera statistic, whose asymptotic χ^2_2 null likewise presumes finite kurtosis, is not a valid test here either.

The last two points are demonstrated rather than asserted in Section 4.6.

Figure 2 shows the Hill estimates as functions of k rather than as points, and the picture is worth reading carefully because it is less tidy than the tabulated numbers suggest.

The curves do *not* exhibit a clean plateau. They are erratic for very small k, where few order statistics carry the estimate, and then decline monotonically as k grows — the textbook bias pattern, in which admitting observations from the centre of the distribution pulls $\hat{\alpha}$ toward the value implied by the bulk rather than the tail. There is no region where the estimate is stable enough for a reader to point at it and say the tail index is *that*.

What survives this is weaker than a point estimate but is the claim the paper needs. Across essentially the entire range of k — from a few dozen order statistics to the largest fraction plotted — every class remains inside the band $2 < \alpha < 4$. The conclusion is therefore not "the tail index of US equity is 2.80"; it is that no defensible choice of k places any class outside the range in which the variance exists and the kurtosis does not. That is the inference the rest of the paper uses, and it is robust to the choice that a single tabulated $\hat{\alpha}$ would have concealed.

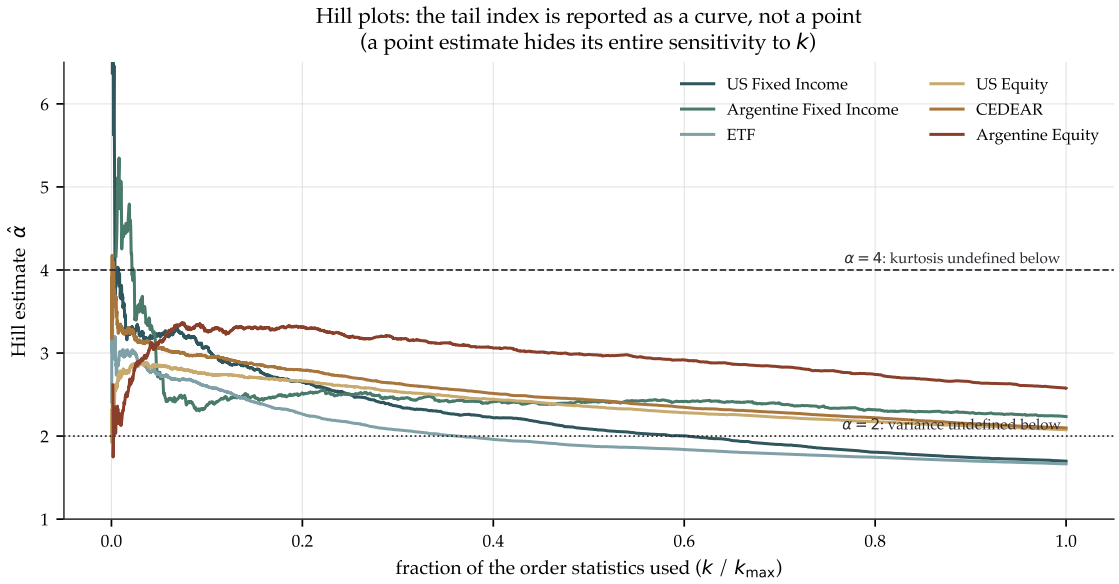

Figure 2: Hill plots. Reference lines mark the thresholds below which the variance ($\alpha = 2$) and the kurtosis ($\alpha = 4$) cease to exist. Every class lies in the intervening band.

Figure 3 makes the same point non-parametrically. Against a Normal the empirical quantiles depart from the diagonal in both tails and by a wide margin; against a Student-t fitted by maximum likelihood they track it into the extremes. The t family is not the only distribution consistent with these data, but it is a far better approximation than the Gaussian, and it is the minimum revision to the usual assumption that the evidence demands.

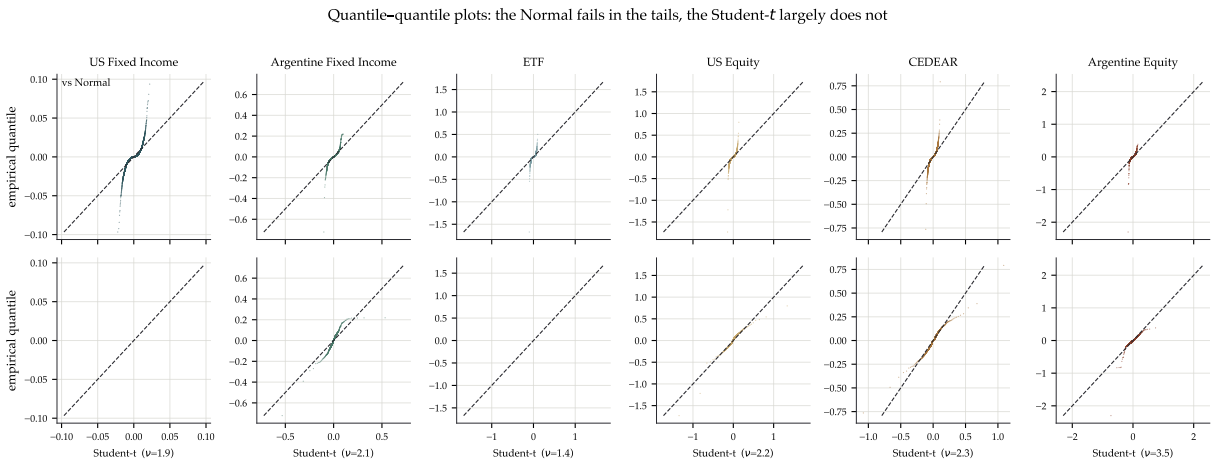

Figure 3: Quantile–quantile plots against the Normal (top) and a fitted Student-t (bottom).

4.3 Fixed income has the heaviest tails

The tail ordering inverts the volatility ordering. Measured by the tail ratio $(q_{99} - q_{01})/(q_{75} - q_{25})$ — a scale-free statistic that requires no moments, with a Gaussian reference value of 3.45 — the daily readings are: Argentine fixed income 6.88, US fixed income 5.99, US equity 5.51, CEDEAR 5.37, ETF 5.18, Argentine equity 5.06.

The two fixed-income classes have the *heaviest* relative tails and the lowest volatilities. Argentine equity, the most volatile class in the sample at 57% annualised, has the lightest relative tails.

This is a coherent economic picture rather than a curiosity. A bond spends most of its life delivering small, highly autocorrelated moves driven by the level of rates, punctuated by discrete repricings — a credit event, a restructuring, a policy shift. Its unconditional distribution is therefore narrow in the middle and long at the ends. An equity index absorbs a continuous flow of information and is volatile all the time; its extremes are large in absolute terms but not extreme *relative to its own everyday dispersion.* Volatility and tail risk are distinct attributes, and an instrument can rank low on one and high on the other.

4.4 Volatility clusters; returns do not

Figure 4 reproduces, on this panel, the oldest and most robust stylised fact in the literature (Mandelbrot, 1963; Cont, 2001): returns are close to serially uncorrelated, while their absolute

values are strongly and persistently autocorrelated. The first-order autocorrelation of $|r_t|$ is 0.30 for US fixed income, 0.24 for ETFs and Argentine fixed income, 0.19 for US equity and CEDEARs, and 0.15 for Argentine equity.

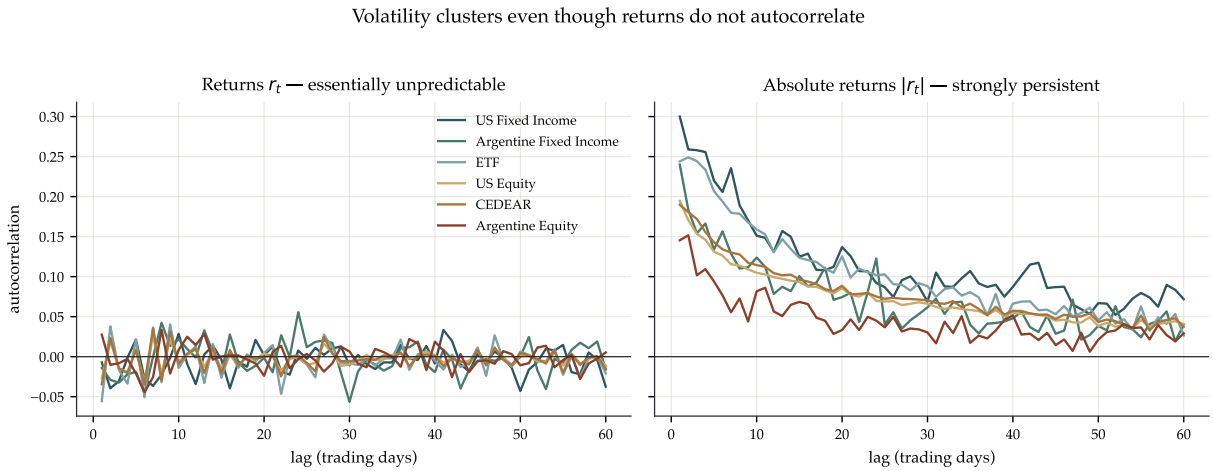

Figure 4: Median autocorrelation across instruments, by lag.

That persistence is what makes the stationary bootstrap necessary rather than optional: an i.i.d. resampling scheme would destroy exactly this structure and return intervals far narrower than the data supports.

4.5 Within US equity, industry is a volatility ordering

Table 4: Volatility by industry within US equity.

Industry	N	Volatility (% p.a.)	Mean (% p.a.)	Ex. kurt.	$\hat{\alpha}$
Biotechnology	129	75.5	0.7	77.6	2.34
Solar	8	72.2	3.6	9.5	3.46
Thermal Coal	2	71.1	7.0	9.6	2.74
Electrical Equipment & Parts	18	67.5	5.3	21.9	3.30
Uranium	8	67.3	21.3	4.5	4.38
Oil & Gas Drilling	8	65.6	-3.4	28.4	2.97
Computer Hardware	15	65.0	17.2	22.5	3.19
Silver	3	63.8	7.0	4.3	3.64
		...			
Financial Data & Stock Exchanges	10	27.8	11.9	12.5	3.57
Drug Manufacturers - General	15	27.5	5.5	38.1	3.07
Utilities - Regulated Gas	13	27.4	3.8	22.3	2.61
Insurance - Diversified	7	27.2	10.1	21.8	2.95
Utilities - Regulated Electric	35	27.1	3.1	94.1	2.70
REIT - Residential	14	27.0	2.5	21.5	3.13
Beverages - Wineries & Distilleries	2	26.2	-2.5	14.2	3.17
Confectioners	3	23.3	4.1	13.7	3.06

Eight most and eight least volatile industries within US equity, of 142 with sufficient coverage. One-way ANOVA of per-instrument volatility on industry: $F = 9.6$, $\eta^2 = 0.368$ — industry explains 36.8% of the cross-sectional variation in volatility.

Across 142 industries covering 2,467 instruments, a one-way ANOVA of per-instrument annualised volatility on industry gives $F = 9.6$ with $\eta^2 = 0.368$: **industry membership accounts for 36.8% of the cross-sectional dispersion in volatility.** The spread is wide — Biotechnology at 75.5% annualised against Confectioners at 23.3%, a factor of three.

Two cautions. The industry classification is a standard commercial taxonomy, and several categories contain only a handful of instruments, so the extremes of Table 4 are estimated on thin samples. And η^2 measures association, not mechanism: industries differ in leverage, in size, in the age of their constituent firms, and any of those could carry the effect. What the statistic establishes is that a sector label is materially informative about volatility — which is a claim about the usefulness of the label, not about industry as a cause.

Notably, the tail index does *not* order with volatility across industries. Biotechnology combines the highest volatility (75.5%) with one of the heaviest tails ($\hat{\alpha} = 2.34$), whereas Uranium is nearly as volatile (67.3%) with a much lighter tail ($\hat{\alpha} = 4.38$). Volatility and tail risk vary independently within an asset class as well as across them.

4.6 Formal comparisons

Table 5: Pairwise comparisons between asset classes.

Comparison	Location			Scale		Distribution	
	$\Delta\mu$ (pp p.a.)	d	p_{perm}	σ_a/σ_b	p_{perm}	W_1 ($\times 10^{-3}$)	p_{MW}
US Fixed Income vs. Argentine Fixed Income	-6.8	-0.024	0.048	0.21	0.000	0.0124	4.5e-07†
US Fixed Income vs. ETF	-3.7	-0.008	0.159	0.23	0.000	0.0088	5.1e-13†
US Fixed Income vs. US Equity	-1.4	-0.002	0.433	0.16	0.000	0.0170	2.5e-22†
US Fixed Income vs. CEDEAR	-7.0	-0.013	0.323	0.20	0.000	0.0139	1.4e-20†
US Fixed Income vs. Argentine Equity	2.1	0.004	0.917	0.14	0.000	0.0219	6.5e-12†
Argentine Fixed Income vs. ETF	3.1	0.005	0.359	1.11	0.887	0.0039	3.4e-04†
Argentine Fixed Income vs. US Equity	5.4	0.007	0.284	0.77	0.000	0.0045	4.2e-01
Argentine Fixed Income vs. CEDEAR	-0.1	-0.000	0.985	0.95	0.192	0.0018	3.6e-01
Argentine Fixed Income vs. Argentine Equity	8.9	0.010	0.433	0.67	0.000	0.0096	1.9e-05†
ETF vs. US Equity	2.3	0.003	0.040	0.70	0.000	0.0083	2.4e-193†
ETF vs. CEDEAR	-3.2	-0.005	0.189	0.86	0.000	0.0054	9.2e-49†
ETF vs. Argentine Equity	5.8	0.008	0.888	0.61	0.000	0.0133	1.6e-11†
US Equity vs. CEDEAR	-5.5	-0.007	0.011	1.23	0.000	0.0030	1.9e-12†
US Equity vs. Argentine Equity	3.5	0.004	0.307	0.87	0.000	0.0056	4.1e-04†
CEDEAR vs. Argentine Equity	9.0	0.012	0.109	0.71	0.000	0.0081	1.8e-07†

d is Cohen's d; W_1 the Wasserstein-1 distance between the pooled daily distributions. p_{perm} from 5,000 relabellings. p_{MW} is the Mann–Whitney test on *per-instrument* annualised volatilities — the comparison that does not treat co-moving assets as independent draws. † survives Benjamini–Hochberg at $q = 0.05$.

Table 5 reports all fifteen pairwise comparisons. The pattern across them supports three conclusions.

Pooled p-values are uninformative. With samples in the millions, every whole-distribution test rejects at any conventional level; the Anderson–Darling p-value is below its tabulated floor of 0.001 for every pair. This is a statement about sample size, not about economics. The Wasserstein distances in the same table are the quantities worth reading, because they are denominated in return units and do not grow with n.

The classical scale tests fail exactly as predicted. Take the comparison this paper was asked to examine: US equity against Argentine fixed income. The F-test returns $F = 0.599$ with $p = 0.000$ — an emphatic rejection of equal variances. Brown–Forsythe, the heavy-tail-robust counterpart, also rejects ($p \approx 10^{-90}$). But both are pooled tests, and Section 3 argued that pooling treats 2,782 co-moving equities as independent draws.

At the instrument level, the same comparison cannot be made. The Mann–Whitney test on per-instrument annualised volatilities — 12 Argentine bonds against 2,782 US equities — returns $p = 0.42$. The median volatilities are 38.7% and 43.1%, a difference of four percentage points that twelve instruments cannot resolve.

Remark 2 (On reading a null result). The correct conclusion is *not* that the two classes have the same volatility. It is that the Argentine hard-dollar sovereign complex contains twelve liquid

instruments, and twelve observations do not support a class-level inference. The pooled test's $p \approx 10^{-90}$ conveys a confidence the data does not contain. Where the two levels disagree, as here, the instrument-level result is the one that respects the dependence structure — and reporting only the pooled figure would be the more impressive and the less honest choice.

Figure 5 places every instrument in the panel on a single mean–volatility plane, which shows how much of the variation within each class would be lost by discussing class averages alone.

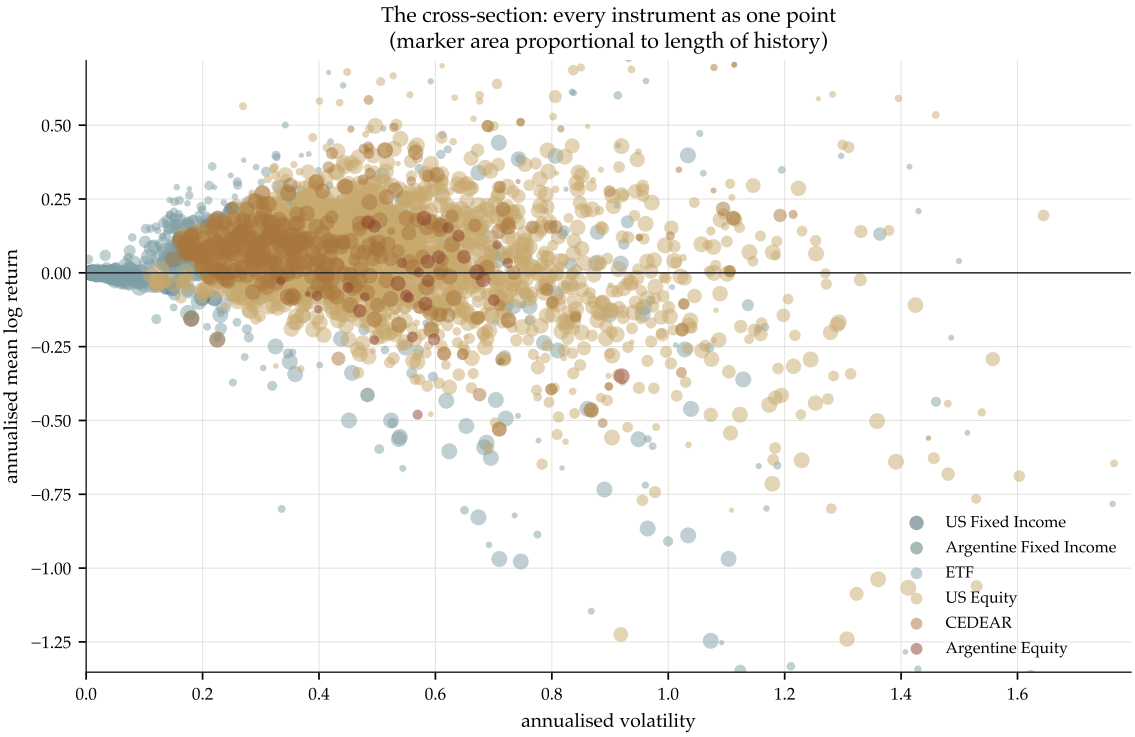

Figure 5: The cross-section. Each point is one instrument.

5 How the distributions change with horizon

Section 4 described daily returns. An investor does not hold an asset for a day, and the distribution of what they actually experience is the distribution of the *aggregate*. Whether that aggregate is a rescaled copy of the daily law or something qualitatively different is the question that decides whether investment horizon belongs in an allocation rule at all.

5.1 Aggregational Gaussianity is real and incomplete

Under i.i.d. increments the central limit theorem applies and the h-period return converges to normality at the rate (12). This is the stylised fact known as aggregational Gaussianity (Cont, 2001).

We measure it with the tail ratio rather than with kurtosis. The reason is the one established in Section 4.2: with $\alpha < 4$ the fourth moment does not exist, so sample kurtosis is not estimating anything. The practical consequence is severe at long horizons. A decade of daily data yields roughly 119 non-overlapping months, and the February–March 2020 block for the S&P 500 tracker sits at approximately -6.7 standard deviations; that single observation contributes about $6.7^4/119 \approx 17$ units of excess kurtosis on its own, out of a measured 19.8. A statistic that one observation can set is not a description of a distribution.

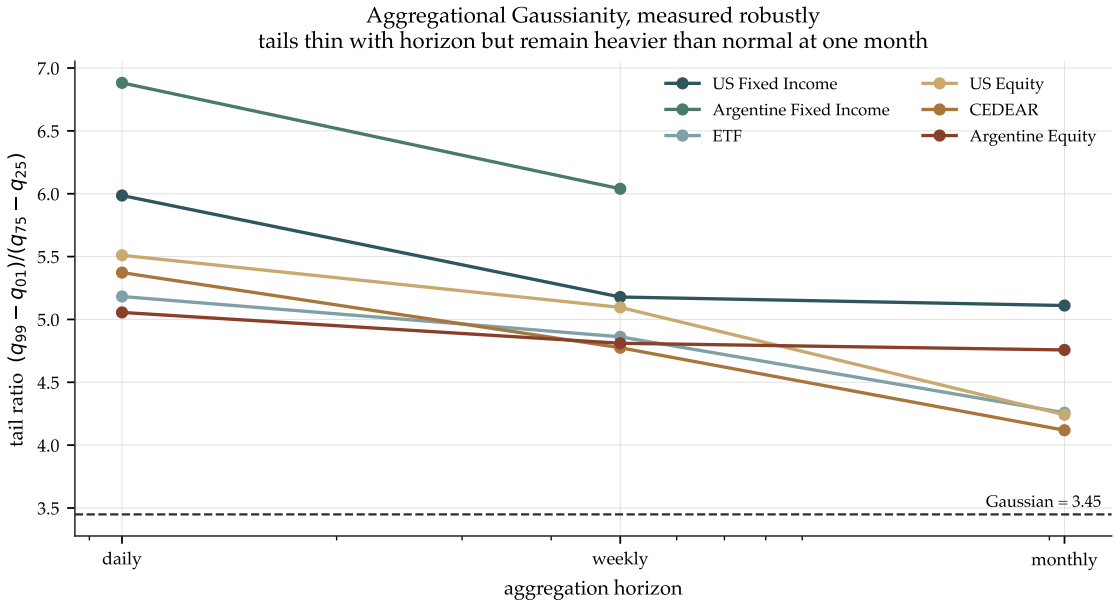

Figure 6: Tail ratio $(q_{99} - q_{01})/(q_{75} - q_{25})$ by aggregation horizon. The Gaussian value is 3.45. Quarterly and longer horizons are omitted: fewer than 60 non-overlapping blocks remain, and the estimate would not be identified.

Figure 6 shows the result. Tails thin monotonically with aggregation in every class — the central limit theorem is visible — but at monthly frequency the ratios remain between 4.1 and 5.1 against a Gaussian 3.45. Convergence is real, slow, and incomplete at the horizon at which most portfolios are reviewed.

The practical reading: a risk model calibrated on daily data and scaled to a month by $\sqrt{21}$ will understate the frequency of large monthly moves. It will understate it less than it understates large daily moves, but it will still understate it.

5.2 Does time diversify?

The question has a precise answer in theory. Samuelson (1969) showed that for an investor with constant relative risk aversion and i.i.d. returns, the optimal share in the risky asset is *independent of horizon*: a thirty-year investor and a one-year investor should hold the same portfolio. The widespread advice that the young should hold more equity is not implied by

risk aversion alone; it requires returns to depart from independence. Any horizon-dependent allocation rule must therefore be earned from the data, not assumed.

The variance ratio is the direct test.

Table 6: Lo–MacKinlay variance ratios, median across instruments.

Asset class	$q = 2$	$q = 5$	$q = 21$	$q = 63$	$q = 126$
US Fixed Income	0.99	0.91	0.86	0.88	0.86
Argentine Fixed Income	0.99	0.93	0.99	1.02	–
ETF	0.95	0.93	0.88	0.78	0.71
US Equity	0.97	0.96	0.91	0.83	0.77
CEDEAR	0.97	0.95	0.93	0.85	0.76
Argentine Equity	1.03	1.04	1.00	0.93	0.89

Median Lo–MacKinlay variance ratio across instruments. VR = 1 under a random walk; < 1 indicates mean reversion (time diversifies), > 1 momentum (it does not).

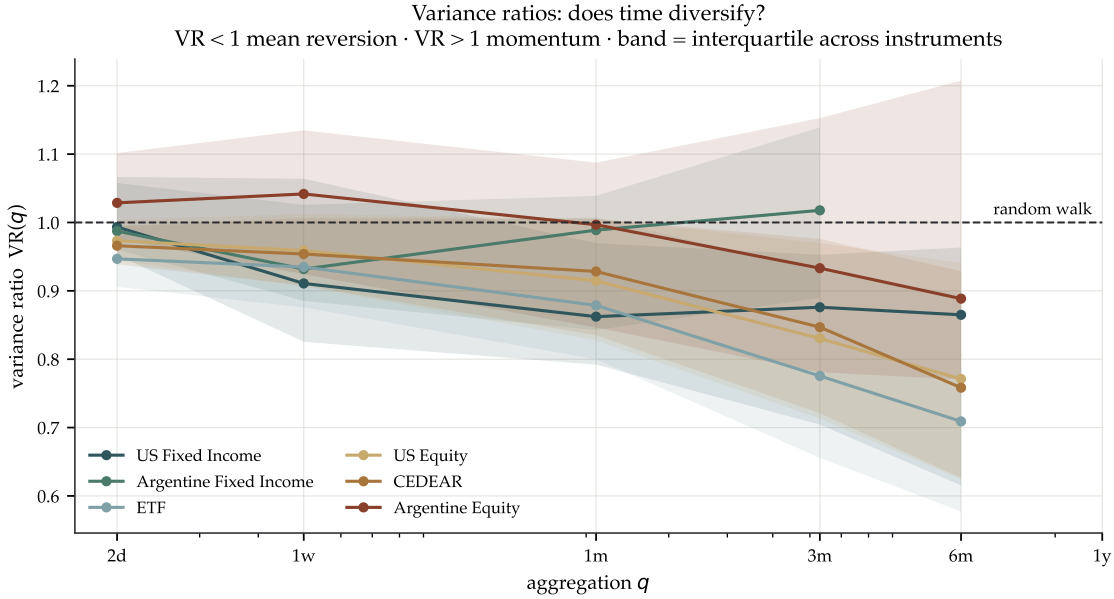

Figure 7: Variance ratios by aggregation. Bands show the interquartile range across instruments within each class.

The results split the panel in two.

Equity-like classes mean-revert, mildly. At a six-month aggregation, the median variance ratio is 0.709 for ETFs, 0.771 for US equity and 0.758 for CEDEARs. Between 81% and 85% of individual instruments in these classes have VR(126) < 1. Six-month variance is therefore roughly 71–77% of what the square-root-of-time rule predicts: holding the position longer does reduce annualised risk, by an amount that is economically meaningful but far short of the folklore.

Argentine fixed income does not. Its variance ratios are 0.987, 0.932, 0.989 and 1.018 at $q = 2, 5, 21, 63$ — indistinguishable from a random walk, with exactly half of instruments on either side of unity at the longer horizons. For this class Samuelson's benchmark holds, and horizon carries no information for allocation.

Remark 3 (A correction worth recording). An earlier implementation of the variance ratio divided by q twice, since the Lo–MacKinlay unbiasing constant already contains that factor. It returned $\mathrm{VR}(q) \approx 1/q$ for every instrument in every class — apparently overwhelming mean reversion everywhere, which would have provided strong support for horizon-dependent allocation. The error was detectable only because the result was too uniform: six asset classes across three countries agreeing to three decimal places is not an economic finding. The corrected estimator was validated against simulated series with known dynamics (i.i.d. random walk, AR(-0.3), AR($+0.3$)), reproducing the theoretical VR(2) of 1.00, 0.70 and 1.30 to within 0.03. We record this because the incorrect version was the convenient one.

5.3 What horizon does and does not license

Taking the two subsections together:

1. **Horizon is not a free lunch.** The mean reversion measured here is modest. A six-month variance ratio of 0.77 corresponds to annualised volatility roughly 12% lower than the i.i.d. scaling implies — not the elimination of equity risk that "stocks are safe in the long run" suggests.

2. **Horizon effects are asset-specific.** They are present in equities and equity funds, and absent in Argentine sovereigns. A single horizon adjustment applied uniformly across asset classes would be wrong for at least one of them.

3. **Longer horizons do not deliver normality.** Tails remain materially heavier than Gaussian at one month, so a long-horizon investor is not thereby released from tail risk — and the estimates here stop at one month precisely because the data cannot identify the quantity beyond it.

These conclusions are weaker than the practitioner literature generally assumes and are consistent with the more careful academic treatments. Pástor and Stambaugh (2012), accounting for parameter uncertainty, find that stocks may be *more* volatile in the long run from the perspective of an investor who does not know the true expected return — a channel this paper does not measure, since every statistic here is computed as though the sample moments were the population moments. Our estimates should be read as an upper bound on the benefit of a long horizon, not a lower one.

6 From distributions to investor profiles

The preceding sections establish what the return distributions are. This one asks what an analyst may legitimately infer from them about how a portfolio should be built for an investor of given risk aversion and horizon — and, equally, what cannot be inferred.

6.1 The cost of assuming normality

Table 7: Optimal allocation and the welfare cost of the Gaussian assumption.

	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7.5$	$\gamma = 10$
Panel A: optimal share w^* in the class index						
Argentine Fixed Income	0.23	0.12	0.08	0.05	0.03	0.02
Gaussian match	0.14	0.07	0.05	0.03	0.02	0.01
ETF	1.26	0.68	0.46	0.28	0.19	0.14
Gaussian match	1.38	0.69	0.46	0.28	0.19	0.14
US Equity	1.15	0.63	0.43	0.26	0.18	0.13
Gaussian match	1.50	0.75	0.50	0.30	0.20	0.15
CEDEAR	2.69	1.45	0.99	0.60	0.40	0.30
Gaussian match	2.96	1.51	1.01	0.61	0.40	0.30
Panel B: certainty-equivalent gap, Gaussian minus empirical (pp p.a.)						
US Fixed Income	-0.25	-0.25	-0.25	-0.26	-0.28	-0.29
Argentine Fixed Income	-1.06	-1.04	-0.93	-0.43	0.77	2.37
ETF	-0.30	-0.04	0.41	2.11	6.43	14.27
US Equity	0.51	1.38	3.04	9.99	27.93	50.03
CEDEAR	0.04	0.03	0.05	0.23	0.82	2.01
Argentine Equity	0.14	0.22	0.69	2.90	6.65	8.38

Non-overlapping monthly returns on the equal-weighted class index; both optima by identical grid search, so the difference isolates the higher moments. Classes with a negative sample mean take $w^* = 0$ at every γ and are omitted from Panel A. Panel B positive means normality *overstates* attainable welfare. Magnitudes at $\gamma \geq 7.5$ are sensitive to the single worst month of 98–122 and should be read as direction, not level.

Panel A of Table 7 reports the share of wealth a CRRA investor should hold in each class index, computed on non-overlapping monthly returns, alongside the same optimisation performed on a Gaussian sample matched in mean and variance. Because both optima come from an identical grid search, the difference between the rows isolates the contribution of the higher moments.

For US equity the effect is substantial: at $\gamma = 1$ the empirical optimum is $w^* = 1.15$ against 1.50 under the Gaussian match, so **accounting for the observed tails reduces the recommended equity allocation by 23%.** The effect persists across risk aversions (0.63 versus 0.75 at $\gamma = 2$). For ETFs and CEDEARs it is smaller, 6–9%, and for Argentine fixed income it runs the *other* way — the empirical optimum (0.23 at $\gamma = 1$) exceeds the Gaussian one (0.14), because that class's monthly index distribution is positively skewed over its short post-restructuring sample.

That reversal matters for the paper's argument. Heavy tails do not mechanically reduce

optimal allocations. What reduces them is a heavy *left* tail. Which tail is heavy is an empirical question, and it differs by asset class.

Panel B expresses the same information as welfare. The certainty-equivalent gap — what the Gaussian approximation claims an investor can achieve, minus what the empirical distribution actually delivers — is near zero at low risk aversion and rises steeply with γ: for US equity from 0.5 percentage points per year at $\gamma = 1$ to about 10 points at $\gamma = 5$. The direction is exactly what theory predicts, since a more risk-averse investor weights the left tail that the Gaussian misrepresents. The magnitudes at $\gamma \geq 7.5$ rest on the worst single month out of roughly 119 and should be read as direction rather than level.

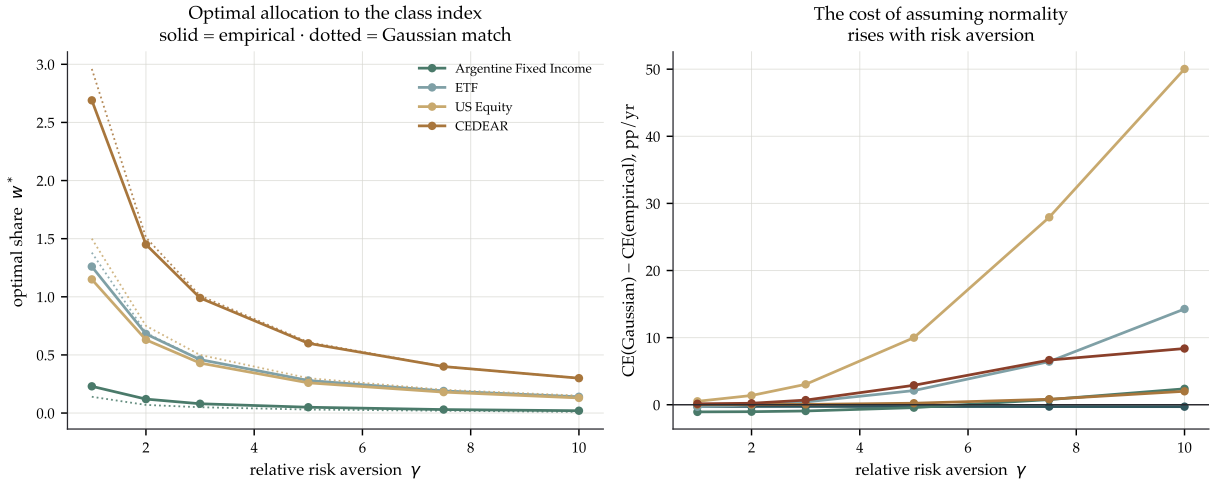

Figure 8: Optimal allocation and the welfare cost of normality.

6.2 Why the mean–variance ranking misleads

Section 4 noted that Argentine fixed income dominates US equity on mean, volatility and Sharpe ratio simultaneously. The distributional evidence explains why an analyst should not act on that ranking.

The Sharpe ratio is a functional of the first two moments only. It is a sufficient statistic for ranking assets when returns are Gaussian — and only then. Argentine fixed income has the lowest tail index in the panel ($\hat{\alpha} = 2.35$), the heaviest relative tails (daily tail ratio 6.88), and a pooled skewness of -1.82. Its superiority on the mean–variance criterion and its inferiority on tail risk are both real, and they point in opposite directions.

This is the practical core of the paper. **Any risk-profiling procedure that ranks assets by volatility or by Sharpe ratio alone will systematically misrepresent instruments whose risk is concentrated in the tail** — and, on this evidence, that description fits fixed income better than it fits equity.

6.3 What an analyst may infer, and what they may not

We can now state the inferential boundary precisely.

Supported by this evidence.

1. **Risk profiling requires a tail measure alongside a dispersion measure.** Volatility and tail heaviness order the classes differently and vary independently within them. A profile expressed only in volatility terms is incomplete in a way that matters most for the instruments a conservative investor is most likely to be offered.

2. **The relevant dispersion measure is asymmetric.** Median semideviation is well below total volatility in every class (Table 2), and the gap differs by class, so downside risk is not a fixed multiple of volatility. A preference specification that penalises downside and upside variation separately therefore captures something the symmetric formulation cannot.

3. **Horizon may enter the allocation for equities, weakly.** The variance-ratio evidence of Section 5 licenses a modest horizon tilt in equity-like classes and none in Argentine sovereigns.

4. **Diversification within an asset class is not optional.** The difference between pooled single-name returns and the class index is the difference between certainty-equivalent gaps of tens of percentage points and of low single digits. Concentration risk in a single name dwarfs every effect discussed in this paper.

Not supported by this evidence.

1. **Any claim that a long horizon makes equities safe.** It reduces annualised variance by roughly an eighth at six months and leaves the tails materially non-Gaussian.

2. **Calibrating γ from the data.** Nothing here identifies an investor's risk aversion; γ is varied parametrically. Inferring it requires either elicited preferences or observed choices under known beliefs, and the literature on doing so credibly (Rabin, 2000) is not encouraging about shortcuts.

3. **Class-level conclusions about Argentine fixed income.** Twelve instruments over 4.7 years cannot establish class-level differences. Its estimates are reported with that qualification throughout.

4. **Any forward-looking statement.** Every estimate here is unconditional and in-sample. The moments are treated as known, when in practice they are estimated with error — the channel by which Pástor and Stambaugh (2012) obtain the opposite sign on the horizon effect.

6.4 Implications for the construction of investor profiles

Three concrete design consequences follow, offered as engineering guidance rather than as advice to any investor.

First, a profile should be a pair, not a scalar: a dispersion tolerance and a tail tolerance. The two are not redundant, because they order the asset classes differently.

Second, the mapping from profile to asset class should not be expressed as "conservative ⇒ bonds". On this sample the two fixed-income classes carry the heaviest relative tails, and Argentine fixed income combines the best Sharpe ratio with the worst tail index. A conservative mandate implemented through a naive bond allocation may acquire precisely the risk it was meant to avoid.

Third, horizon should modulate the allocation only where the data supports it, and the modulation should be small. The variance-ratio evidence justifies a mild tilt in equities and none in Argentine sovereigns; a uniform horizon rule applied across classes would be wrong for at least one of them.

7 Conclusion

Return distributions differ across instrument classes in ways that a mean–variance summary does not capture, and the differences matter for how portfolios should be built.

The tail index lies between 2 and 4 for every class in this panel. Variance exists, so the mean–variance apparatus is not vacuous; kurtosis does not, so the standard tests for comparing variances are invalid and the robust alternatives are not optional refinements. Tail heaviness and volatility are distinct attributes that order the classes differently: the two fixed-income classes are the least volatile and the heaviest-tailed, and the most volatile class has the lightest tails. Within US equity, industry membership accounts for 37% of the cross-sectional variation in volatility, while tail index varies independently of it.

Aggregation thins the tails but does not normalise them: at one month every class remains well above the Gaussian benchmark. The mean reversion that alone could justify horizon-dependent allocation is present in equity-like classes and modest — six-month variance ratios around 0.71–0.77 — and absent in Argentine sovereigns. The common advice that a long horizon makes equities safe overstates what the data supports, and applying a single horizon rule across asset classes would be wrong for at least one of them.

Using the empirical distributions rather than Gaussian matches reduces the optimal equity allocation for a log-utility investor by 23%, and the welfare cost of assuming normality grows with risk aversion. But the effect is not mechanical: for a positively skewed class it runs the other way. What penalises an investor is a heavy *left* tail, and which tail is heavy is an empirical question.

Three limitations bound all of this. The universe is a current snapshot, so delisted instruments

are absent; this biases mean returns upward and probably makes the tail estimates conservative. Every moment is treated as known when it is in fact estimated, which is the channel through which Pástor and Stambaugh (2012) obtain the opposite sign on the horizon effect — our estimates are an upper bound on the benefit of a long horizon. And the Argentine fixed-income class contains twelve instruments over 4.7 years, which supports description but not class-level inference.

The natural extension is conditional rather than unconditional analysis: the tail index is time-varying, and an investor's exposure depends on the tail they face now rather than on its decade average. That requires the conditional machinery — GARCH-family models, realised measures, and the rough-volatility literature (Gatheral et al., 2018; Deep et al., 2026) — which we have deliberately kept out of a paper about unconditional shape.

A Reproduction

The study is reproducible from public data. The steps are:

1. **Panel.** Daily adjusted closes for the listed instruments are available from standard public market-data endpoints; the Argentine hard-dollar sovereigns from public Argentine market-data APIs. The window is ten years to August 2026.
2. **Currency treatment.** Argentine peso quotes are converted at the implied "contado con liquidación" rate, computed as the cross-sectional median over liquid dual-listed pairs as in Section 2. Depositary receipts are reconstructed from the underlying security and the relevant exchange rate divided by the conversion ratio, rather than taken from their local quotes.
3. **Returns.** Log returns on consecutive observed closes, no forward-filling. Instruments require at least 250 return observations and a price above one currency unit.
4. **Estimation.** All estimators and tests are those defined in Section 3; none is proprietary. Resampling uses a fixed seed, a stationary bootstrap with mean block length 21, and 5,000 permutations.

Three errors were found and corrected during the analysis. They are recorded here because in each case the incorrect result was the more publishable one, and because the means of detection generalise.

1. **The variance ratio was divided by q twice.** The Lo–MacKinlay unbiasing constant already contains that factor, so $\sigma_c^2(q)$ is a per-period variance. The error returned $\mathrm{VR}(q) \approx 1/q$ identically across all six classes, which would have been reported as overwhelming mean reversion and would have supported horizon-dependent allocation on fabricated grounds. It was detectable only because the result was too uniform to be economic. The corrected

estimator was validated against simulated series with known dynamics — i.i.d. random walk, AR(-0.3) and AR($+0.3$) — reproducing the theoretical VR(2) of 1.00, 0.70 and 1.30 to within 0.03.

2. **Kurtosis at long horizons was determined by one observation.** With $\alpha < 4$ the statistic has no population target, and with roughly 119 non-overlapping months a single -6.7σ block contributes most of the measured value. It was replaced by the quantile tail ratio, which requires no moments.

3. **Certainty equivalents were computed on pooled single-name returns.** That construction measures the fate of an investor holding one randomly chosen security, so individual issues losing most of their value dominated the expectation at high risk aversion. The analysis was recomputed on equal-weighted class indices at monthly frequency, which is the decision an investor actually faces.

References

T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212, 1952.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1): 289–300, 1995.

Morton B. Brown and Alan B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Akash Deep, Nicholas Appiah, and Svetlozar T. Rachev. Memory, roughness, and information persistence in financial markets: A structural approach to volatility forecasting. *arXiv preprint*, 2026. arXiv:2605.24285.

Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

Carlos M. Jarque and Anil K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980.

Howard Levene. Robust tests for equality of variances. In *Contributions to Probability and Statistics*, pages 278–292. Stanford University Press, 1960.

Andrew W. Lo and A. Craig MacKinlay. Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1):41–66, 1988.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4): 394–419, 1963.

Luboš Pástor and Robert F. Stambaugh. Are stocks really less volatile in the long run? *The Journal of Finance*, 67(2):431–478, 2012.

Dimitris N. Politis and Joseph P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.

Matthew Rabin. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292, 2000.

Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, 1969.

F. W. Scholz and M. A. Stephens. K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947.