Break-Even Volatility for Leveraged ETFs: A Higher-Moment Framework and a Volatility-Gated Switching Rule

Arpit Pareek, Suleiman Rahimzai, Arya Dev, Derek Horstmeyer, Zihao Li
Department of Finance, Costello College of Business, George Mason University

September 1, 2026

Abstract

Leveraged exchange-traded funds amplify benchmark exposure, but their volatility loss grows with the square of the leverage ratio while the compensating risk premium grows only linearly, implying a break-even volatility (σ^*) at which the two exactly offset. We extend this framework in three directions. First, we generalize the Gaussian break-even condition to a quartic admitting any combination of skewness and kurtosis through the Pearson family, so the threshold no longer assumes normally distributed returns. Second, we show the annualized threshold is invariant to reset frequency under independent returns, so daily, weekly, monthly, and quarterly resets share one break-even volatility, and characterize how serial correlation qualifies this. Third, we build a switching strategy holding leverage only while a GARCH forecast of conditional volatility sits below the threshold, tested against implied-volatility and trend-following alternatives. The threshold proves governed almost entirely by the spread between expected equity return and financing cost; higher moments enter with the signs theory requires, but too small to change any switching decision. This predicts, before any test, that the CBOE SKEW index cannot inform the leverage choice, and our results confirm it. In totality, the results detail the feasibility of an optimal switching leverage product based on higher moments, reset periods, and financing costs.

Keywords: leveraged ETFs, volatility decay, break-even volatility, Pearson moments, GARCH, Gaussian, dynamic leverage, look-ahead bias.

1 Introduction

The leveraged exchange-traded funds (LETFs) are designed to provide a fixed multiple of the daily return of a reference index. So, a 3× ETF aims to deliver three times the reference's percent return on any given trading day. However, investors tend to misunderstand the daily target return for a long-horizon performance guarantee. Given that exposures get recalculated over time, an LETF is unlikely to outperform the benchmark multiple of times over one month, one year, or possibly ten years. The performance over many periods is determined by the series of returns through which the benchmark achieves its final value. Higher exposure to the market brings additional return proportional to the leverage multiple L, but the variance cost to the expectation of the log-return growth is proportional to L^2. Financing costs, fund management fees, and the cost of leverage itself reduce the expected excess returns. Thus, volatile enough markets could lead to under-performance of the leveraged fund despite positive gains from the reference index. Therefore, the issue isn't whether leverage increases returns, but rather the specific market regime in which marginal returns generated by leverage will be big enough to offset the excess compounding and trading costs.

Historical experience shows that this distinction makes a big difference. The S&P delivered an annualized return of 8.33% over the 1996-2024 time-line used in our analysis. The annualized return of a daily-reset 3× leveraged position, which includes effects of financing and fund charges, was 6.90% , and the maximum draw-down of the position reached −98.8%. At 4× leverage, the annualized return was just 0.69% , while maximum draw-down was close to −99.9%. Hence, permanent leverage in an unfavorable volatility environment may under-perform the asset to which it provides exposure. The particular path of returns determine the multi-period performances that the index follows, not just where it finishes because of path dependence [Cheng and Madhavan, 2009, Avellaneda and Zhang, 2010, Jarrow, 2010]. The source of this gap is volatility decay, which is negative on average: exposure payoff scales linearly with the leverage ratio L , but variance cost of the expected logarithmic growth scales with L^2 . There are additional drags from financing costs and fund expenses. The numbers above illustrate the point directly, without resorting to any modeling. Static leverage does not simply fail to match its multiple; it underperformed the underlying index outright once the opposing effects of volatility are accounted for. The question this raises is not whether leveraged products are usable, but under what conditions holding one is justified — which is the question this paper answers. For geometric Brownian motion, the expectation of the continuously compounded growth rate of an $L\times$ investment is

$$g(L) \;=\; L\mu \;-\; (L-1)r_f \;-\; \phi \;-\; \tfrac{1}{2}L^2\sigma^2, \tag{1}$$

where μ is the expected index return, r_f is the financing cost, ϕ is the management fee and σ is the annualized volatility.

Instead of solving for the optimal leverage for the given volatility level — the Kelly-Merton question [Kelly, 1956, Merton, 1969], whose answer is non-purchasable continuum — we turn the question around and ask which volatility makes a given, invest-able leverage break even against the index. This is known as the break-even volatility (σ^*).

Under Gaussian returns it has the closed form:

$$\sigma^* = \sqrt{\frac{2\left[(L-1)(\mu - r_f) - \phi\right]}{L^2 - 1}}. \tag{2}$$

Below σ^* the levered fund beats the index; above σ^*, the fund fails to beat the index. When calibrated to the S&P 500, σ^* lies within the empirical support of realized volatility, defining its role as a regime signal.

The research framework introduced in this paper makes use of the following four pieces of related literature. Avellaneda and Zhang [2010] proves that the terminal value of a leveraged fund under continuous rebalancing depends on the integrated variance of the path of realized volatility. The findings encourage the use of volatility as the main state variable. Li [2026] applies the break-even problem to the central moment space to give a Pearson-based representation that gives rise to the fourth-order model developed here. Return persistence is modeled along the lines of Hsieh et al. [2025]. A first-order auto-regressive process is used to approximate the framework in the present study. Finally, the conditional-volatility component builds on the ARCH and GARCH frameworks of Engle [1982] and Bollerslev [1986].

The Gaussian break-even threshold condition is extended to a fourth-order relation that accommodates skewness and kurtosis using the Pearson central moments representation. While the higher order moments affect the break-even relation with the theoretically predicted sign, the practical impact of their inclusion in the daily rebalancing break-even threshold is negligible. Under calibrated S&P 500 moments, accounting for skewness and kurtosis would shift σ^* by only 0.15 percentage point at 3× leverage. The higher moments therefore enter the threshold with the sign theory requires but at a magnitude far below the decision band that governs the rule, and it is that smallness — established from the polynomial rather than discovered in the data — which makes the SKEW result of Section 5.6 a structural statement rather than an empirical one.

In the independent case, the annualized break-even threshold is independent of the reset frequency when moments of the return distribution are properly defined for the respective horizon. Serial correlation impacts variance accumulate in multi-period intervals and, hence, changes the relationship between the measured daily volatility and volatility associated with weekly and monthly resetting. Rather than imposing a momentum parameter, we estimate the auto-regressive coefficient from historical data. It is negative — $\hat{\varphi} = -0.0841$, with a heteroskedasticity-consistent standard error of 0.0238 and $t = -3.54$ — indicating short-horizon mean reversion rather than momentum in daily S&P 500 returns.

Furthermore, the comparative statics result shows that $\mu - r_f$ makes up the primary determinant of σ^* , with equity risk-premium spread dominating in terms of quantitative importance. For a leverage ratio for $3\times$, a one percentage point increase in the financing rate leads to a decrease in the breakeven threshold of approximately 1.51 percentage points, which is an order of magnitude greater than the total higher moment effect in the calibrated return distribution. This demonstrates the concept that keeping nominal expected equity returns fixed and varying financing costs mechanically yields a time-varying equity risk premium. The empirical model thus considers both the fixed-nominal-return and the fixed-equity-risk-premium specifications.

2 Literature Review

The mechanics of how leveraged and inverse ETFs really work was first modeled by Cheng and Madhavan [2009]. That is, under a strategy of maintaining fixed leverage, the fund must be able to rebalance every day, which means that the fund buys on days when prices rise and sell when prices fall. As a result, Avellaneda and Zhang [2010] arrive at a result that is coherent with the former paper. Under constant rebalancing, the value of the fund after time T is $X_T/X_0 \approx (S_T/S_0)^L \exp\big(-\frac{1}{2}L(L-1)\int_0^T \sigma_t^2\,dt\big)$, corrected only by a coefficient dependent on the realized variance process. No other features of the variance path are significant. Using the volatility threshold as a mechanism for control makes more sense than an arbitrary modeling assumption, as the integrated variance is the only state variable considered. The empirical size of the losses incurred over the holding period was researched by Lu et al. [2012], Trainor and Carroll [2013] and SLCG [2009], and Jarrow [2010] obtained similar results using a different approach. More recently, Trainor [2026] revisits the approximation itself, showing that the relative expected-return estimates issuers publish deviate materially from the statistically correct values at a given index return and standard deviation — evidence that even the characterisation of decay, let alone a rule for acting on it, is not settled. Our contribution is to calculate the exact volatility at which it is offset and extend the calculation beyond diffusion.

The question of how much leverage is optimal was being asked before these funds existed. Kelly [1956] and Merton [1969] derived the growth-optimal fraction $(\mu - r_f)/\sigma^2$ under log utility, and MacLean et al. [2011] reviewed it along with its susceptibility to parameter uncertainty. The above studies ask about the optimal L when σ is known. Our paper uses the opposite approach, asking instead what σ is required for a known tradable L to break even. We are not aware of prior work that poses the problem in this direction; Li [2026] provides the Pearson central-moment form of the quartic that we extend in Section 3.3. That formulation suits our purpose, because the decision facing an investor is discrete — hold the $3\times$ fund or the $1\times$ fund — rather than the continuum of exposures the growth-optimal literature solves for.

Beyond the second moment, for a higher-order extension of the break-even analysis, it must be assumed that the mean and variance remain constant while skewness and kurtosis change.

The fact that higher moments representing risk are rewarded is illustrated: Kraus and Litzenberger [1976] and Harvey and Siddique [2000] demonstrate that co-skewness receives a premium. The leveraged-ETF literature has not carried that treatment into the break-even problem: the closest study, Hsieh et al. [2025], models return dynamics in detail but assumes Gaussian innovations throughout, so skewness and kurtosis play no role in it. The statistical tools consist of the Pearson [1895] family of distributions, indexed precisely by its first four moments, which is why the manifolds of Section 4 provide well posed comparative statics instead of the confounding of multiple moments changing simultaneously. Our conclusion that the impact of the higher-moment adjustment is small in case of daily rebalancing — 0.15 percentage points of σ^* at $L = 3$ — as far as we know, has not been previously quantified in this context, and it is what makes our SKEW null a structural rather than empirical statement.

Serial Dependence enters through the Rebalancing Horizon. The variance ratio statistic, which was introduced by Lo and MacKinlay [1988] for summarizing the impact of autocorrelation on multi-period variance, is the one used here. It is Poterba and Summers [1988] who noted the horizon dependence in the sign of the same. Momentum enters the process of price formation explicitly in the model developed by Hsieh et al. [2025] and it is their version of the same, which we have used in section 3.4; we estimate its coefficient from our own sample rather than assuming a value. Our result is negative in the sense that momentum affects nothing else but the units of measurement of volatility.

The implementation of σ^* requires a volatility forecast. The conditional-variance filter provided by Engle [1982] and Bollerslev [1986] can be estimated via rolling maximum likelihood and it is precisely its resilience that makes the threshold strategy work prior to realizing full realized variance. The alternative of the asymmetric version of the GJR-GARCH model in [Glosten et al., 1993] specification, where negative returns increase variance more than positive ones, motivated as it was by the empirical in-sample success, degraded out-of-sample performance in our earlier specification. Section 5.2 explains why we withdraw that quantitative comparison rather than restate it: it predates the signal-timing correction of Section 5.7 and has not been re-run under corrected timing. The forward-looking measure of volatility would naturally be an implied volatility measure but Carr and Wu [2009] and Bollerslev et al. [2009] show that the difference between the expectation of the variance under the risk-neutral and physical measures is positive, persistent and informative. As such, assuming VIX as a physical measure introduces a known upward bias which, for the purpose of a de-levering strategy, is precisely the wrong one.

An extension to the third moment is the CBOE SKEW index, its content has been examined by Bevilacqua and Tunaru [2021] and Elyasiani et al. [2021] with generally modest predictive results, and whose tail-risk-premium interpretation is developed in Bollerslev and Todorov [2015]. Our null result is consistent with that literature but lands on a different mechanism: not that SKEW fails to predict, but that the quantity it would inform is insensitive to it at the horizon in question.

The gap this paper fills is the connection between these ideas. The leveraged-ETF literature characterizes decay but stops short of a decision rule. The growth-optimal literature produces a rule, but for a continuum of leverage that is not investable. The higher-moment and volatility-forecasting literature builds up to a break-even threshold but stops short of one that can be evaluated as a model. We derive σ^* under non-Gaussian returns and autocorrelation, establish which of its inputs actually govern it, estimate it in real time from a rolling conditional-variance forecast, and evaluate the resulting strategy against the benchmark an investor would actually hold. The implied-volatility and higher-moment signals we also test turn out to add nothing once signal timing is corrected, which is itself one of the paper's results.

3 Theoretical Framework

This section develops the break-even volatility σ^* in four stages. Section 3.1 fixes notation. Section 3.2 derives the threshold under geometric Brownian motion, which is exact but assumes Gaussian returns. Section 3.3 extends it to a quartic polynomial admitting arbitrary skewness and kurtosis. Section 3.4 treats the rebalancing horizon and first-order return momentum, and establishes that the annualised threshold is exactly invariant to both. Section 3.5 then asks which of the model's inputs actually move σ^*, and finds an answer that reorders the priorities of everything downstream. Section 3.6 specifies the Monte Carlo procedure used to check the closed form where the expansion is least trustworthy.

3.1 Notation

S_t denotes the reference index and X_t the net asset value of the leveraged fund. The fund targets constant leverage L and resets its exposure every W trading days, with $N = 252$ trading days per year. We work throughout with the moments of the *daily* simple return of the index: mean, daily volatility v, skewness s, and excess kurtosis κ, the last normalised so that a Gaussian law gives $\kappa = 0$. The annualised volatility is $\sigma = v\sqrt{N}$. The annualised expected return of the index is μ, the financing rate is r_f, and ϕ is the annual expense ratio of the leveraged fund. Momentum enters through the first-order autoregressive coefficient φ, and $C_f(\varphi, W)$ denotes the resulting variance-ratio correction defined in Section 3.4.

Remark 1 (Notation collisions). Three collisions arise when this paper is read alongside its sources and must be kept in mind. First, Li [2026] writes W_t for the leveraged fund's wealth process; we reserve W for the rebalancing period and write X_t for fund value. Second, s denotes skewness throughout this paper, never a standard deviation; volatility is always v (daily) or σ (annualised). Third, κ is *excess* kurtosis, whereas Li [2026] reports raw kurtosis k, so $\kappa = k - 3$; the calibration $\kappa = 16.818$ corresponds to a raw kurtosis of 19.818. Both calibrated moments are rolling-window medians rather than full-sample statistics; see Section 3.3.

3.2 The geometric Brownian motion baseline

Let the index follow $dS_t/S_t = \mu\, dt + \sigma\, dW_t$. A fund that continuously rebalances to maintain exposure L must borrow $(L-1)$ units of capital per unit of equity and pays an expense ratio, so its value evolves as

$$\frac{dX_t}{X_t} \;=\; L\,\frac{dS_t}{S_t} \;-\; (L-1)\, r_f\, dt \;-\; \phi\, dt, \tag{3}$$

and by Itô's lemma its expected continuously compounded growth rate is

$$g(L) \;=\; L\mu - (L-1)r_f - \phi - \tfrac{1}{2}L^2\sigma^2. \tag{4}$$

Equation (4) contains the entire economic argument of this paper and repays a slow reading. The term $L\mu$ is the reward for exposure and is linear in leverage. The term $-(L-1)r_f - \phi$ is the deterministic cost of obtaining that exposure and is likewise close to linear. The term $-\frac{1}{2}L^2\sigma^2$ is volatility decay, and it is *quadratic*. Its origin is the wedge between arithmetic and geometric averaging: a sequence of returns that averages zero compounds to less than zero, by approximately $\sigma^2/2$ per unit time, and levering the position by L multiplies the squared increments by L^2. Leverage is therefore rewarded in proportion to L and penalised in proportion to L^2. Some volatility level must exist at which the two exactly offset, and locating it is the object of this paper.

Setting $g(L) = g(1)$ in (4), the ϕ term of the unlevered benchmark nets out and the condition becomes

$$(L-1)(\mu - r_f) - \phi \;=\; \tfrac{1}{2}\big(L^2 - 1\big)\sigma^2,$$

which rearranges to the break-even volatility

$$\boxed{\sigma^* \;=\; \sqrt{\frac{2\big[(L-1)(\mu - r_f) - \phi\big]}{L^2 - 1}}} \tag{5}$$

Below σ^* the $L\times$ fund out-compounds the index; above it, it does not. When financing and fees are set to zero, (5) collapses to $\sigma^* = \sqrt{2\mu/(L+1)}$, the form given by Li [2026]. Evaluated at $\mu = 8\%$ with $r_f = \phi = 0$, it yields 23.09%, 20.00% and 17.89% for $L = 2, 3, 4$; these are the values against which our numerical solver is verified.

Two structural features of (5) deserve emphasis because they govern the comparative statics of Section 3.5. First, σ^* depends on the square root of the risk-premium spread $(\mu - r_f)$, so the threshold is far less sensitive to that input than the input's own estimation uncertainty might suggest — but it depends on the spread, not on μ alone, a distinction that turns out to matter a great deal. Second, the fee enters divided by $(L-1)$ once the expression is written per unit of incremental leverage, so expense ratios bite hardest at low leverage, where there is least incremental risk premium to absorb them.

Remark 2. Equation (5) conditions on constant σ. Under stochastic volatility, Avellaneda and Zhang [2010] show that a continuously rebalanced fund satisfies $X_T/X_0 \approx (S_T/S_0)^L \exp\bigl(-\frac{1}{2}L(L-1)\int_0^T \sigma_t^2\,dt\bigr)$, so the fund's terminal value depends on the volatility path only through integrated variance. The correct object to compare against σ^* is therefore the expected average variance over the holding period rather than any spot reading, which is the theoretical justification for the GARCH-based forecast of Section 5.2 in preference to a trailing realised window.

3.3 The Pearson quartic: admitting skewness and kurtosis

Equity index returns are not Gaussian. They are negatively skewed, because markets fall faster than they rise, and leptokurtic, because extreme moves occur far more often than a normal law permits. Throughout this paper we calibrate to $s = -0.514$ and excess kurtosis $\kappa = 16.818$. These are the medians of the rolling 1,260-day estimates of the daily S&P 500 return moments over 1990–2024, not the full-sample statistics, which are -0.181 and 10.48. We calibrate to the rolling medians deliberately: the switching rule of Section 5.4 evaluates the break-even polynomial on rolling moments, so the rolling distribution is the one the rule actually faces, and its median is the representative draw from that distribution. The choice is also the conservative one — the rolling medians are more adverse than the full-sample values on both moments, so calibrating to them overstates rather than understates the higher-moment correction whose smallness is the finding of this section. Both features are adverse for a levered position and both should therefore reduce σ^*.

Following Li [2026], we retain the first four moments in the expansion of expected log-wealth, which gives the growth rate of the $L\times$ fund as

$$g_L \;\approx\; L\,m \;-\; \frac{L^2 v^2}{2} \;+\; \frac{L^3 s\, v^3}{6} \;-\; \frac{L^4 \kappa\, v^4}{24}, \tag{6}$$

where m is the daily mean return. Imposing the break-even condition $g_L = g_1$, every coefficient appears as a difference $L^j - 1$; collecting terms and restoring the financing and fee drag yields the central result of this section.

Proposition 1 (Break-even quartic)**.** *The daily break-even volatility v^* is the smallest positive real root of*

$$\bigl(L^4-1\bigr)\kappa\, v^4 \;-\; 4\bigl(L^3-1\bigr)s\, v^3 \;+\; 12\bigl(L^2-1\bigr)v^2 \;-\; \frac{24\bigl[(L-1)(\mu-r_f)-\phi\bigr]}{N} \;=\; 0, \tag{7}$$

and the annualised threshold is $\sigma^* = v^*\sqrt{N}$. Setting $s = \kappa = 0$ recovers (5) *exactly.*

Two features of (7) are worth drawing out. The cubic term carries $-4(L^3-1)s$; for the empirically relevant case $s < 0$ this term is positive, adds to the left-hand side, and therefore forces v^* down. Negative skew tightens the threshold, because a levered position converts the index's asymmetry into a compounding asymmetry — a given percentage loss requires a larger percentage gain to recover, and leverage widens that gap. The quartic term carries $+(L^4-1)\kappa$ with $\kappa > 0$ for any realistic return distribution, so fat tails tighten the threshold unconditionally, regardless of which tail is heavier.

The polynomial admits no closed-form solution and is solved numerically. We extract all four roots, discard those with imaginary part exceeding 10^{-9} in absolute value and those with real part at or below 10^{-6}, and take the minimum of the survivors. The economically meaningful root is the smallest positive one: at $v = 0$ the polynomial is negative (assuming a positive net risk premium) and increasing, so the first crossing marks the volatility at which leverage ceases to pay. Larger roots are artefacts of truncating the expansion at fourth order. When $\kappa \approx 0$ the quartic degenerates and we solve the quadratic $v^* = \sqrt{-D/C}$ directly.

How much do the higher moments actually matter?

Having built the apparatus, intellectual honesty requires reporting how much work it does. Table 1 compares σ^* computed with $s = \kappa = 0$ against σ^* computed at the full S&P 500 calibration, at the long-run drift $\mu = 12.3\%$ and daily rebalancing.

	$L=2$	$L=3$	$L=4$
σ^*, Gaussian ($s=\kappa=0$)	28.64%	24.80%	22.18%
σ^*, calibrated ($s=-0.514$, $\kappa=16.818$)	28.50%	24.65%	22.02%
Shift attributable to higher moments	-0.13 pp	-0.15 pp	-0.16 pp

Table 1: Effect of skewness and kurtosis on the break-even threshold at $\mu = 12.3\%$, $W = 1$, with $r_f = \phi = 0$ so that the moment terms are isolated. The signs are as theory predicts and the magnitudes are small.

The higher moments move the threshold by roughly *fifteen hundredths of a percentage point.* The signs are correct and the direction is theoretically required, but at daily rebalancing the cubic and quartic terms of (7) are numerically negligible against the quadratic term. The reason is a mismatch of scale rather than of order: the third and fourth cumulants are the same order in the time step as the variance, so they do not vanish asymptotically, but at realistic daily parameters their coefficients are roughly two orders of magnitude smaller.

This is not a negative result about the framework so much as a precise statement of where the framework earns its keep. It has three consequences that we carry forward. First, it explains the SKEW null result of Section 5.6 *structurally* rather than empirically: the usual explanation, that physical skewness is too stable to serve as a dynamic signal, is true but secondary. The stronger statement is that no movement in skewness within any plausible range can shift σ^* by enough to change a switching decision, because the measured effect (a ± 2 standard deviation shock to skewness moves σ^* by less than 0.10 percentage points at every leverage level) is two orders of magnitude below the hysteresis band of 3 percentage points that governs the rule. The null is a property of the polynomial, not an accident of the sample. Second, it predicts where the higher moments *would* bind: the kurtosis coefficient grows as L^4 against L^2 for the variance term, so the correction grows with leverage, and it grows further at longer rebalancing horizons. Both are directions this paper explores rather than assumes. Third, it establishes the scale against which Section 3.5 measures everything else.

Remark 3 (On the expansion coefficients)*.* The coefficients 1/6 and 1/24 in (6) are those of Li [2026] and are used throughout this paper and in all figures. A direct Taylor expansion of $\ln(1+LR) = LR - \frac{1}{2}(LR)^2 + \frac{1}{3}(LR)^3 - \frac{1}{4}(LR)^4$ instead produces 1/3 and 1/4. We have verified by simulation (2×10^7 draws matched to the S&P 500 calibration) that at daily rebalancing and $L = 3$ the two conventions give σ^* of 20.91% and 20.78% against an exact simulated value of 20.69%. The discrepancy of roughly 0.2 percentage points is smaller than the effect of a 15 basis point move in the financing rate (Section 3.5) and is immaterial at $W = 1$. It is not immaterial at $W = 21$ or at $L = 4$, and we flag it as a limitation of extrapolating the closed form to those settings rather than as a correction applied here.

3.4 Rebalancing horizon, momentum, and W-invariance

A fund that resets every W days is exposed to the moments of the W-day return rather than the daily return. Under independent daily increments the moments aggregate according to

$$m_W = m\,\frac{W}{N}, \qquad s_W = \frac{s}{\sqrt{W}}, \qquad \kappa_W = \frac{\kappa}{W}, \tag{8}$$

which are the standard cumulant scaling laws: the third and fourth cumulants are additive over independent increments, and dividing by the appropriate power of the standard deviation — itself growing as $\sqrt{W}$ — produces the decay rates shown. These are the same relations that make monthly returns look more Gaussian than daily ones.

Substituting (8) into (7) and re-annualising produces a result that is easy to state and, on first encounter, surprising.

Proposition 2 (W-invariance)**.** *Under independent daily returns, the annualised break-even threshold is exactly independent of the rebalancing period W, for all four moments:*

$$\sigma_W^* \;=\; \sigma_1^* \qquad \textit{for every } W.$$

Sketch. Write the quartic in terms of the annualised σ_W using $v_W = \sigma_W \sqrt{W/N}$. Each term acquires a factor of W/N and no more: the quartic term contributes $(\kappa/W)\cdot(W/N)^2 = \kappa W/N^2$, the cubic contributes $(s/\sqrt{W}) \cdot (W/N)^{3/2} = sW/N^{3/2}$, the quadratic contributes W/N, and the constant term is already proportional to W/N. Dividing through by W/N eliminates W entirely. □

Numerically the invariance holds to machine precision. At $\mu = 12.3\%$, $s = -0.514$, $\kappa = 16.818$ and $\varphi = 0$, the thresholds at $W \in \{1, 5, 21\}$ agree to within 3×10^{-16} for every leverage level: 28.50% at $L = 2$, 24.65% at $L = 3$, and 22.02% at $L = 4$. The three manifold surfaces of Section 4 are not merely similar under $\varphi = 0$; they are identical.

First-order momentum

Serial dependence breaks the independence assumption underlying (8). Following Hsieh et al. [2025] we model the daily return as first-order autoregressive with coefficient φ, under which the variance of a W-day sum is inflated by the variance ratio

$$C_f(\varphi, W) \;=\; 1 + \frac{2}{W}\sum_{j=1}^{W-1}(W-j)\,\varphi^j \;=\; 1 + \frac{2\varphi}{1-\varphi} - \frac{2\varphi\,(1-\varphi^W)}{W(1-\varphi)^2}, \tag{9}$$

the Lo and MacKinlay [1988] statistic, with $C_f(\varphi, 1) = 1$ and $C_f(\varphi, \infty) = (1+\varphi)/(1-\varphi)$.

We estimate φ from the sample rather than assuming it. Over 1996–2024 the first-order autocorrelation of daily S&P 500 returns is $\hat{\varphi} = -0.0841$, with a heteroskedasticity-consistent standard error of 0.0238 and $t = -3.54$. The sign is negative: daily returns partly *reverse* rather than trend. Consequently $C_f < 1$ and multi-period variance accumulates more slowly than independence would imply, taking the values 1.000, 0.874 and 0.852 at $W = 1, 5, 21$. Monthly returns are $\sqrt{0.852} = 0.92$ times as volatile as an independent-increments model would predict from the same daily variance.

The economically important point is what C_f does and does not affect. It does *not* enter the break-even condition, which depends only on the moments of the W-day return; a threshold expressed in terms of the volatility actually measured at the rebalancing horizon remains σ_W^* from Proposition 2, momentum or no momentum. What C_f changes is the relationship between that quantity and the daily volatility a GARCH filter reports. Since $\sigma_W^2 = \sigma_1^2\, C_f$, the equivalent threshold on *daily-measured* volatility is

$$\sigma_{1|W}^* \;=\; \frac{\sigma_W^*}{\sqrt{C_f(\varphi, W)}}. \tag{10}$$

At $\mu = 12.3\%$, $L = 3$ and the estimated $\hat{\varphi} = -0.0841$ this gives 24.65% at $W = 1$, 26.38% at $W = 5$, and 26.71% at $W = 21$. Because returns mean-revert, a fund that resets less often accumulates less variance over the block and therefore tolerates *more* volatility, not less. Momentum moves the signal, not the bar; but if the signal and the bar are quoted in different units, the practitioner must convert one of them, and (10) is that conversion.

Remark 4 (The sign of φ, not its magnitude, is what matters). Earlier drafts of this paper assumed $\varphi = 0.30$, a positive coefficient implying momentum. That assumption reverses the conclusion above. At $\varphi = +0.30$ the variance ratio takes the values 1.000, 1.613 and 1.799 at $W = 1, 5, 21$, and (10) gives 24.65%, 19.41% and 18.38%: the threshold *falls* with the rebalancing horizon rather than rising, and the ladder inverts. We report this counterfactual once, and nowhere else in the paper, because it isolates what the reset-horizon result actually rests on. It is not the size of the autocorrelation but its sign, and the sign is an empirical question we answer with an estimate rather than an assumption.

Remark 5 (Correction to the implementation). An earlier version of the manifold code applied an additional adjustment to the quadratic coefficient, $C \mapsto 12(L^2-1) - 12(L-1)(C_f-1)/C_f$, alongside the conversion (10). That adjustment does not follow from the derivation above, and its effect runs opposite to its intent: evaluated at the $\varphi = 0.30$ that draft assumed, it *raises* σ^*, giving 20.38% rather than 19.41% at $W = 5$ and 19.47% rather than 18.38% at $W = 21$, thereby making the rule less conservative at exactly the horizons where the assumed momentum argued for more caution. The results reported here use (10) alone.

Remark 6 (On the stability of $\hat{\varphi}$). The full-sample estimate conceals substantial instability, which we report because the qualitative conclusion of this subsection depends on a sign. Judged against heteroskedasticity-consistent standard errors throughout, $\hat{\varphi}$ is $+0.018$ in the 1990s ($t = 0.55$), -0.088 in the 2000s ($t = -2.60$), -0.047 in the 2010s ($t = -1.47$) and -0.182 in the 2020s ($t = -2.19$). Mean reversion is therefore statistically detectable only in the two decades containing a major crisis and is indistinguishable from zero in the two calm ones. A rolling five-year window ranges from -0.252 to $+0.076$ and is positive in 14% of windows. No single value is defensible as a structural constant, which is why we report $\hat{\varphi}$ alongside the $\varphi = 0$ baseline of Proposition 2 rather than treating it as a fixed parameter. We note also that the naive i.i.d. standard error $1/\sqrt{n}$ would report $t = -7.85$ on the full sample and would call the 2010s significant; it is not appropriate for a series with this much volatility clustering and we do not use it.

3.5 What actually moves the threshold

Sections 3.3 and 3.4 establish that skewness, kurtosis, the rebalancing horizon and momentum each shift σ^* in the theoretically expected direction. This section asks the quantitative question those results leave open: among all the model's inputs, which ones actually determine the threshold? The answer is not the one the structure of the model suggests.

Differentiating (5) gives the two elasticities that matter:

$$\frac{\partial \sigma^*}{\partial r_f} = \frac{-1}{(L+1)\,\sigma^*}, \qquad \frac{\partial \sigma^*}{\partial\,\phi} = \frac{-1}{(L^2-1)\,\sigma^*}. \tag{11}$$

Evaluated at $L = 3$ and a threshold near 16.5%, a one percentage point rise in the financing rate lowers σ^* by 1.57 percentage points. The same perturbation at $L = 2$ and $L = 4$ gives 1.91 and 1.38 percentage points. A ten basis point increase in the expense ratio lowers σ^* by 7.5 basis points at $L = 3$. These are exact perturbations of the quartic; the linearised elasticities of (11) give 1.51, 1.82 and 1.33, the difference being the curvature of the square root over a perturbation this large.

Set that against the higher-moment correction. Table 2 makes the comparison directly, with every row evaluated at a single common baseline so the magnitudes are strictly comparable. That baseline matters: Table 1 isolates the moment terms at $\mu = 12.3\%$ with $r_f = \phi = 0$, where $\sigma^* = 24.65\%$ and the moment correction is 0.15 percentage points; at the lower threshold prevailing once financing and fees are charged the same correction is 0.06 percentage points, because the moment terms scale with powers of σ^* itself. Both figures are correct for their own configuration, and the comparison in Table 2 uses the latter.

Perturbation	Shift in σ^* at $L = 3$
Financing rate +1 pp	-1.57 pp
Expense ratio +10 bp	-0.075 pp
Skewness and kurtosis, Gaussian $\to$ full calibration	-0.06 pp
Excess kurtosis 16.8 $\to$ 100	-0.06 pp
Excess kurtosis 16.8 $\to$ 400	-0.27 pp

Table 2: Sensitivity of the break-even threshold to its inputs, every row evaluated at the same baseline: $L = 3$, $\mu - r_f = 5.96\%$, $\phi = 95$ basis points, giving $\sigma^* = 16.50\%$. A single percentage point of financing cost moves the threshold twenty-six times as far as the entire higher-moment correction. Excess kurtosis would have to reach roughly 3,000 — a distribution unlike any observed equity index — to match one point of financing cost.

The break-even threshold is, to first order, a function of the risk-premium spread $(\mu - r_f)$ and of essentially nothing else. This has an immediate implication for how μ is specified. If μ is held fixed in nominal terms while r_f floats with the policy rate, then the implied equity risk premium is not a modelling assumption at all — it is a residual that collapses as rates rise. Under a fixed $\mu = 8\%$, the spread falls from 8.0% when rates are at zero to 1.6% when the three-month bill yields 6.4%, and σ^* correspondingly traverses the range in Table 3.

	3-month bill	σ^*, fixed nominal $\mu = 8\%$	σ^*, fixed ERP = 5.5%
2000	5.80%	8.14%	15.85%
2003	1.00%	17.50%	15.85%
2006	4.80%	10.78%	15.85%
2011	0.05%	18.81%	15.85%
2021	0.04%	18.82%	15.85%
2024	5.20%	9.81%	15.85%

Table 3: Break-even threshold at $L = 3$ under two specifications of the expected return, using approximate annual averages of the three-month Treasury bill plus the 40 basis point financing spread. Under a fixed nominal drift the threshold varies by nearly eleven percentage points across the sample; under a fixed risk premium it is constant.

The economics of the rate effect are real: higher financing costs genuinely do make leverage less attractive, and a threshold that tightens when money is expensive is behaving correctly.

What is not obviously correct is the magnitude, which under a fixed nominal μ is driven by an implied risk premium that no one has specified and that varies by a factor of five across the sample. Nominal expected equity returns co-move with interest rates; the premium is considerably more stable than the nominal return it sits inside.

This matters for interpreting the empirical results of Sections 5 and 6, because the policy-rate path over 1996–2024 is correlated with the regimes the strategy must call. Rates were high through 1996–2000 and again in 2006–2007, giving a threshold near 8–11% that a GARCH volatility reading of 15–20% clears easily, forcing the strategy defensive ahead of both major drawdowns; and rates were at zero from 2009 to 2021, giving a threshold near 19% that keeps the strategy levered through the bull market. Those are the correct calls. But under a fixed nominal μ a material share of them is attributable to the policy rate rather than to the volatility signal, and a reader who plots σ^* against the bill yield will see the relationship immediately.

We therefore report the empirical results under both specifications: the fixed nominal $\mu = 8\%$ used in the original implementation, and a constant-premium specification $\mu_t = r_{f,t} + \text{ERP}$ with ERP fixed at 5.5%, which holds the threshold constant and isolates the contribution of the volatility signal. Table 4 reports the comparison.

	CAGR	Vol	Sharpe	MaxDD
1× buy-and-hold	8.33%	19.2%	0.378	−56.8%
2×, fixed nominal $\mu = 8\%$	10.97%	24.8%	0.439	−61.5%
2×, constant ERP = 5.5%	9.44%	26.1%	0.377	−66.0%
3×, fixed nominal $\mu = 8\%$	12.77%	29.8%	0.466	−56.8%
3×, constant ERP = 5.5%	11.74%	32.3%	0.427	−66.9%

Table 4: The two specifications of expected return, at daily decision frequency with the defensive position set to 1×, under the GARCH-only final specification.

The improvement narrows under the constant-premium specification but does not vanish: at 3× the Sharpe ratio falls from 0.466 to 0.427, still clearly above the index's 0.378, while at 2× it falls from 0.439 to 0.377 — to the index's level, with the residual gap of 0.0009 far too small to read as an advantage in either direction. A material share of the fixed-nominal result is therefore attributable to the policy-rate path rather than to the volatility signal, and across the two leverage levels of practical interest the share is larger at lower leverage. We regard this as the more informative of the two possible outcomes: the strategy is part volatility timing and part rate timing, and reporting the decomposition is more useful than claiming it is purely the former.

The monotonicity does not extend to 4×. Table 8 reports the constant-premium specification *outperforming* the fixed-nominal one at that leverage, 0.466 against 0.423, reversing the pattern at 2× and 3×. We flag this rather than explain it. The two specifications do not hold the same positions on the same days — time in leverage under the fixed-nominal gate falls monotonically in L, from 0.499 at $2\times$ to 0.400 at $4\times$ — so the comparison at each leverage level is between different day sets, and at $4\times$ the fixed-nominal gate is out of the market for three days in five. Attributing the reversal to the rate signal rather than to that selection difference would require holding the selected days fixed across specifications, which we have not done. We note also that $4\times$ lies outside the range any issuer currently offers on a broad equity index, so nothing in the paper's recommendations turns on this row.

3.6 Closed form and Monte Carlo

Proposition 1 gives σ^* as the root of a polynomial, which is cheap enough to evaluate on a dense grid and is what makes the manifolds of Section 4 feasible at 60×60 resolution across three leverage levels and three horizons. It is also an approximation, obtained by truncating an expansion at fourth order, and its accuracy degrades in ways that are predictable but not self-diagnosing. The natural check is to solve the break-even condition a second way, by simulation. We set out that procedure here and the conditions under which it would be expected to disagree with the closed form.

Such a procedure would target the exact condition $\mathbb{E}[\ln(1 + LR_W)] = \mathbb{E}[\ln(1 + R_W)] + \delta$ directly, where R_W is the W-day index return and δ the financing and fee drag. Returns are generated from the Pearson diffusion of Li [2026],

$$dY_t \;=\; -\theta\,(Y_t - \mu)\,dt \;+\; \sqrt{2\theta\,(aY_t^2 + bY_t + c)}\;dW_t, \tag{12}$$

whose parameters are fixed by the first four moments through $A = 10k - 12s^2 - 18$, $a = (2k - 3s^2 - 6)/A$, $b = s\sqrt{v}\,(k+3)/A$ and $c = v(4k - 3s^2)/A$, with $k = \kappa + 3$ the raw kurtosis. The Pearson family is the appropriate choice because it is indexed exactly by its first four moments, which lets skewness and kurtosis be varied independently while mean and variance are held fixed — a property that a Student-t or skew-normal specification does not possess, and without which the manifold surfaces would confound the moment being swept with the ones that move alongside it. For each candidate σ one would simulate paths over a fixed horizon, compound the $L\times$ and $1\times$ positions with resets every W days, and locate σ^* by bisection on the difference in mean log growth.

We expect close agreement at $W = 1$ and divergence at $W = 21$, for three reasons of increasing severity. The truncation error of the fourth-order expansion is $O\big((LR_W)^5\big)$: at 20% annualised volatility a two-standard-deviation daily move at $L = 3$ gives $|LR_W| \approx 0.076$ and an error near 5×10^{-7}, while the same move over a month gives $|LR_W| \approx 0.35$ and an error near 1.4×10^{-3}, which cumulated over twelve resets is of the order of a hundred basis points of annual log growth. The error is signed: for losses every term of the series is negative, so truncation understates the loss and the closed form is biased toward leverage. More fundamentally, the exact integrand $\ln(1 + LR_W)$ diverges as $R_W \to -1/L$, a -33.3% move at $L = 3$; a polynomial has no singularity there and reports a finite value. Over a day such a move lies outside the historical record, but over a month it does not. Finally, under conditional heteroskedasticity realised variance over a holding period is far more dispersed than an unconditional model implies, and since Avellaneda and Zhang [2010] place integrated variance inside an exponential, that dispersion is not a second-order detail; conditioning on a single unconditional triple (σ, s, κ), as the closed form must, averages it away.

These three arguments predict where the closed form should be trusted and where it should not, and they do so before any simulation is run: the truncation error is $O((LR_W)^5)$ and therefore negligible at daily rebalancing and material at monthly; it is signed, so the closed form is biased toward leverage; and the exact integrand has a singularity at $R_W = -1/L$ that no polynomial can represent. Every empirical result reported in this paper is at daily reset, where the expansion is accurate to well under a basis point of the effects being measured, and we therefore use the closed form throughout. Verifying the predicted divergence at $W = 21$ and $L = 4$ by simulation, and establishing which method should govern a product specification at those settings, is left to future work. We flag it here rather than omitting it because the direction of the bias is known: a reader extrapolating our thresholds to a monthly-reset product should treat them as an upper bound on tolerable volatility rather than a point estimate.

4 Manifold Results

The closed form of Proposition 1 is cheap enough to evaluate on a dense grid, which allows the comparative statics of Section 3.5 to be displayed as surfaces rather than asserted as derivatives. We sweep σ^* over the annual drift $\mu \in [3\%, 20\%]$ against each of four second axes in turn — skewness, excess kurtosis, the financing rate, and the autocorrelation coefficient — at three reset horizons and at $L = 2$ and $L = 3$. All surfaces use the rolling-median calibration $s = -0.514$, $\kappa = 16.818$ of Section 3.3 and the 95 basis point expense ratio, and each panel marks the S&P 500 calibration point.

Figures 1 and 2 present the full sheet at $L = 3$ and $L = 2$. Each carries a translucent plane at the index's own realised volatility of 19.2%: where the surface sits above that plane, the leveraged fund out-compounds the index at typical market volatility, and where it sits below, it does not.

Four features are visible, and each corresponds to a result established analytically above.

The moment surfaces are nearly flat along their own axis. Sweeping skewness across $[-1.5, +0.5]$ and excess kurtosis across $[0, 20]$ moves σ^* by less than a quarter of a percentage point at any drift level, while the same surfaces are steeply sloped in μ. This is Table 1 rendered geometrically: the higher moments enter the quartic with the correct sign but with coefficients two orders of magnitude below the variance term.

The financing surface collapses to zero. Along the front edge, where $(L-1)(\mu - r_f) \leq \phi$,

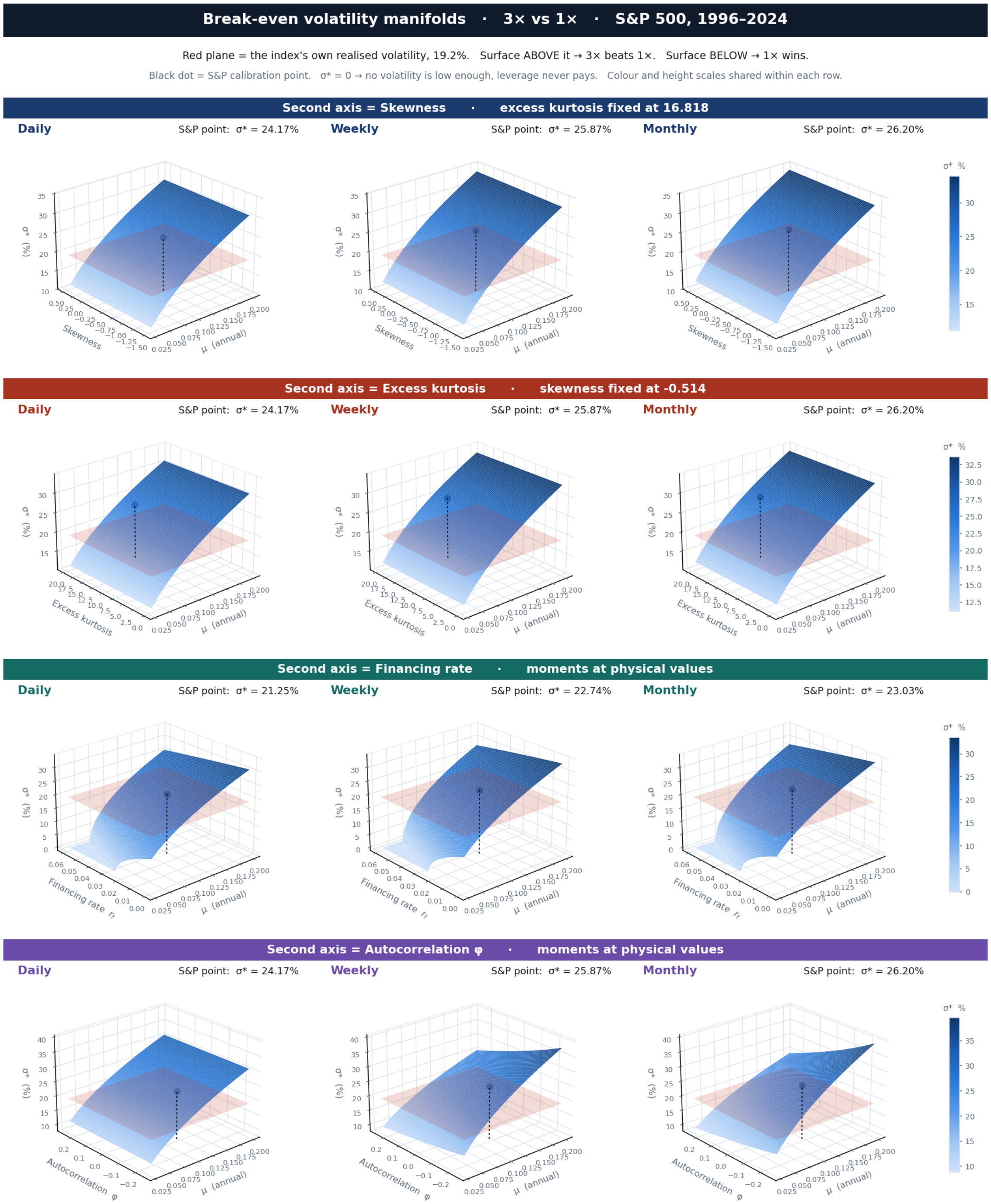

Figure 1: Break-even volatility manifolds at $L = 3$. Rows sweep skewness, excess kurtosis, the financing rate and the autocorrelation coefficient φ against the annual drift μ; columns are daily, weekly and monthly reset. The red plane marks the index's realised volatility of 19.2% and the black dot the S&P 500 calibration point. Colour and height scales are shared within each row, so panels are comparable across reset horizons.

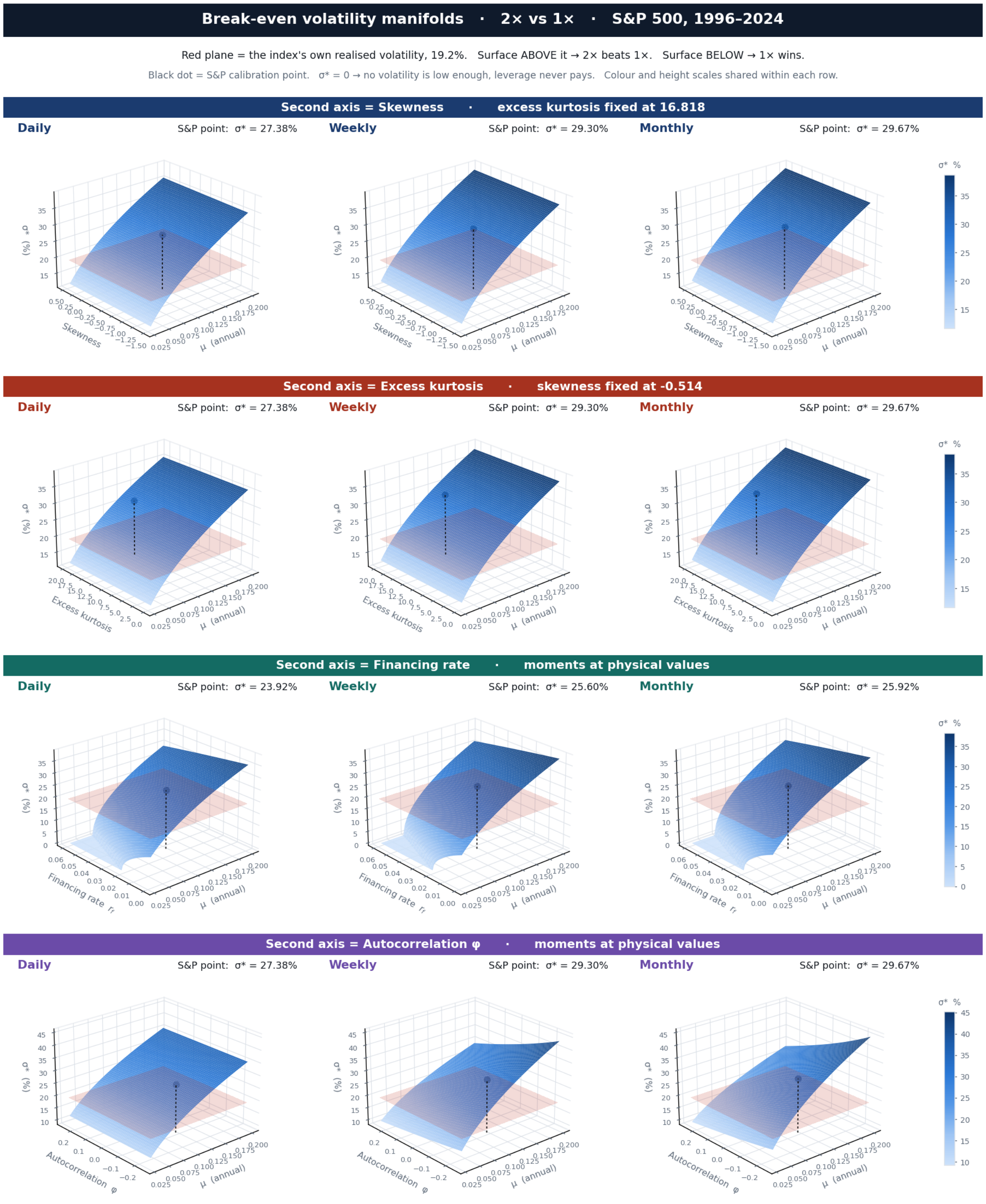

Figure 2: The same sheet at $L = 2$. Every surface sits higher than its $L = 3$ counterpart in Figure 1 — at the S&P calibration the daily threshold is 27.38% against 24.17% — which is the comparative static of Section 3.5 made visible: the threshold falls with leverage, so a more levered product requires a calmer market to be worth holding. The qualitative shape of every row is unchanged, confirming that the features below are properties of the polynomial rather than of one leverage level.

the quartic has no positive root and $\sigma^* \to 0$: no volatility is low enough for leverage to pay, because the risk premium no longer covers the cost of obtaining it. That boundary is the only place in the entire parameter space where the threshold vanishes, and it is reached by moving the financing rate alone.

The autocorrelation surface is flat at daily reset and tilts at longer horizons. At $W = 1$ the surface is exactly constant along the φ axis, because $C_f(\varphi, 1) = 1$ for every φ: with nothing to aggregate, autocorrelation cannot enter. At $W = 5$ and $W = 21$ the surface tilts, and the direction of the tilt follows the sign of φ. This is Proposition 2 together with its qualification, in a single panel: the reset horizon acts on the threshold only through serial dependence.

The three horizons are otherwise identical. In the skewness, kurtosis and financing rows, the daily, weekly and monthly panels differ only by the level shift that C_f produces. Under $\varphi = 0$ they are not merely similar but coincide to machine precision.

We also stress skewness by ± 2 standard deviations of its rolling estimate, holding the other three moments fixed. The threshold moves by -0.10 and $+0.09$ percentage points at $L = 2$, by -0.10 and $+0.10$ at $L = 3$, and by -0.10 and $+0.10$ at $L = 4$. This is the counterfactual behind the SKEW null of Section 5.6: the null holds not because skewness fails to move, but because σ^* barely responds when it does. A two-standard-deviation shock — a crash regime by the standard of the sample — shifts the threshold by less than a tenth of a point against a decision band of three.

5 Empirical Model

5.1 Data

We use daily closing levels of the S&P 500 (`^GSPC`), the CBOE VIX and SKEW indices, and the 13-week Treasury bill yield (`^IRX`), from January 1990 to December 2024 (8,711 trading days after aligning the four series). The strategy is evaluated from January 1996, allowing 1,510 days of warm-up for the rolling moment window (1,260), the momentum window (200) and a buffer — 28.6 years of out-of-sample evaluation.

Leveraged returns are constructed from the index rather than taken from live funds, since no 3× S&P product existed before 2008 and none exists at 4×. A daily $L\times$ position earns $L\, r_{\text{spx},t} - (L-1)\, r_{f,t}/252 - \phi/252$, with the expense ratio set at 95 basis points and financing at the T-bill plus a 40 basis point spread. A natural validation, which we leave to future work, is to compare the constructed series against UPRO or SPXL from 2009 onward.

5.2 GARCH(1,1) and rolling estimation

The conditional variance follows $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ with Gaussian innovations, estimated by maximum likelihood via Nelder–Mead on returns pre-scaled by 100 for numerical stability, subject to $\omega > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 0.999$. Parameters are re-estimated every 252 trading days on a rolling 2,500-day window, with the first fit at day 1,000; the variance recursion runs forward between refits. Every forecast $\hat{\sigma}_t$ therefore uses only returns through $t-1$, making it a genuine one-step-ahead prediction usable at t. Across the 32 refits the fitted parameters range over $\omega \in [0.0002,\ 0.0505]$, $\alpha \in [0.0184,\ 0.2049]$ and $\beta \in [0.7440,\ 0.9804]$, with persistence $\alpha+\beta \in [0.9459,\ 0.9990]$ and a mean of 0.9866. The process is therefore close to integrated throughout, as is standard for daily equity returns, but never estimated at the constraint boundary.

We also tested the asymmetric extension of Glosten et al. [1993], which allows negative returns to raise future variance more than positive ones. Despite strong in-sample motivation — the leverage effect is well documented — it degraded out-of-sample performance in our earlier specification. We withdraw the quantitative comparison rather than restate it, because it predates the signal-timing correction of Section 5.7 and we have not re-run it under corrected timing. The qualitative reading we retain is the familiar one: a richer conditional variance specification evaluated out of sample rather than in tends to fit noise rather than dynamics.

5.3 The variance risk premium correction

The VIX is forward-looking but risk-neutral, and Carr and Wu [2009] and Bollerslev et al. [2009] establish that the gap between risk-neutral and physical expected variance is large and persistently positive. In our sample raw VIX averages 19.48% against a variance-risk-premium-corrected 16.43% — a systematic bias of roughly three percentage points, which for a de-levering rule errs in exactly the wrong direction. We estimate the premium as a rolling 252-day mean of $\mathrm{VIX}_\tau - \hat{\sigma}_\tau^{\mathrm{GARCH}}$ and subtract it, and separately against subsequently realised 21-day volatility, the textbook definition, lagged 22 days so no future information enters. The two give mean corrected levels of 16.43% and 15.35%.

The correction is sound and the bias it removes is real. It does not, however, make the resulting signal useful, as Section 5.5 shows.

5.4 The switching rule

Let $\hat{\sigma}_t$ be the GARCH forecast, σ_t^* the threshold from Proposition 1 evaluated on rolling physical moments, and δ_{band} a hysteresis band. The rule is

$$\mathrm{position}_t = \begin{cases} L\times & \hat{\sigma}_t < \sigma_t^* - \delta_{\mathrm{band}} \\ 1\times & \hat{\sigma}_t > \sigma_t^* + \delta_{\mathrm{band}} \\ \mathrm{position}_{t-1} & \text{otherwise.} \end{cases} \tag{13}$$

Every input is observable at the close of $t-1$. The defensive position is $1\times$, not cash: switching to cash confers a market-timing option no leveraged product actually offers and flatters drawdown accordingly. Section 6 reports both. Position changes incur 5 basis points.

The band is set at 3 percentage points, a value fixed before the analysis in Section 5.5. Table 6 reports the full sweep.

5.5 Signal ablation

Strategy	CAGR	Vol	Sharpe	MaxDD	Info. ratio
1× buy-and-hold	8.33%	19.2%	0.378	−56.8%	—
2× buy-and-hold	9.15%	38.3%	0.353	−91.2%	0.329
2× momentum only	10.19%	28.3%	0.393	−73.5%	0.323
2× GARCH only	10.71%	23.7%	0.439	−58.7%	0.391
2× GARCH, no momentum	**10.97%**	24.8%	**0.439**	−61.5%	**0.402**
2× GARCH+VIX	10.14%	23.0%	0.423	−57.3%	0.337
2× GARCH+VIX, no mom.	10.80%	23.6%	0.443	−56.8%	0.404
2× GARCH+VIX, fwd. VRP	9.64%	23.0%	0.403	−57.6%	0.274
2× GARCH+SKEW blend	10.63%	23.7%	0.436	−58.7%	0.383
3× buy-and-hold	6.90%	57.5%	0.362	−98.8%	0.353
3× momentum only	11.60%	39.1%	0.412	−85.3%	0.367
3× GARCH only	11.46%	28.3%	0.434	−58.8%	0.342
3× GARCH, no momentum	**12.77%**	29.8%	**0.466**	**−56.8%**	**0.412**
3× GARCH+VIX	10.53%	27.1%	0.410	−59.8%	0.285
3× GARCH+VIX, no mom.	11.17%	27.9%	0.427	−56.8%	0.325
3× GARCH+VIX, fwd. VRP	11.26%	27.3%	0.433	−58.1%	0.333
3× GARCH+SKEW blend	11.46%	28.3%	0.434	−58.8%	0.342

Table 5: Signal ablation, 1996–2024, net of fees, financing and switching costs, with all signals correctly lagged. Information ratios are measured against 1× buy-and-hold; multiplying by $\sqrt{28.6}$ gives a t-statistic, so 0.412 corresponds to $t = 2.20$.

Three readings of Table 5 matter.

The volatility gate works, and it is the gate that works. A momentum-only rule — the 200-day moving average, a well known standalone strategy — reaches an information ratio of 0.367 at 3× but with a −85.3% drawdown. The GARCH gate reaches 0.412 with a −56.8% drawdown. Whatever the rule is doing, it is not trend-following in disguise.

Option-implied volatility subtracts. At 3×, adding the VRP-corrected VIX to the GARCH gate lowers the information ratio from 0.412 to 0.325; the textbook forward-VRP definition recovers part of that loss, at 0.333, but neither reaches the GARCH-only gate. Split by sample half, the GARCH–VIX rule's Sharpe gain over passive is +0.001 in 1996–2010 and −0.005 in 2011–2024. This is not a weak signal; it is no signal. Given that VIX is a 30-day risk-neutral expectation and the decision is a one-day conditional one, the natural reading is that a properly specified conditional variance model already contains what the option market knows about the near horizon, plus less noise.

The trend overlay subtracts. Removing it raises the 3× information ratio from 0.342 to 0.412. It was never derived from the theory, and the theory is right that it does not belong.

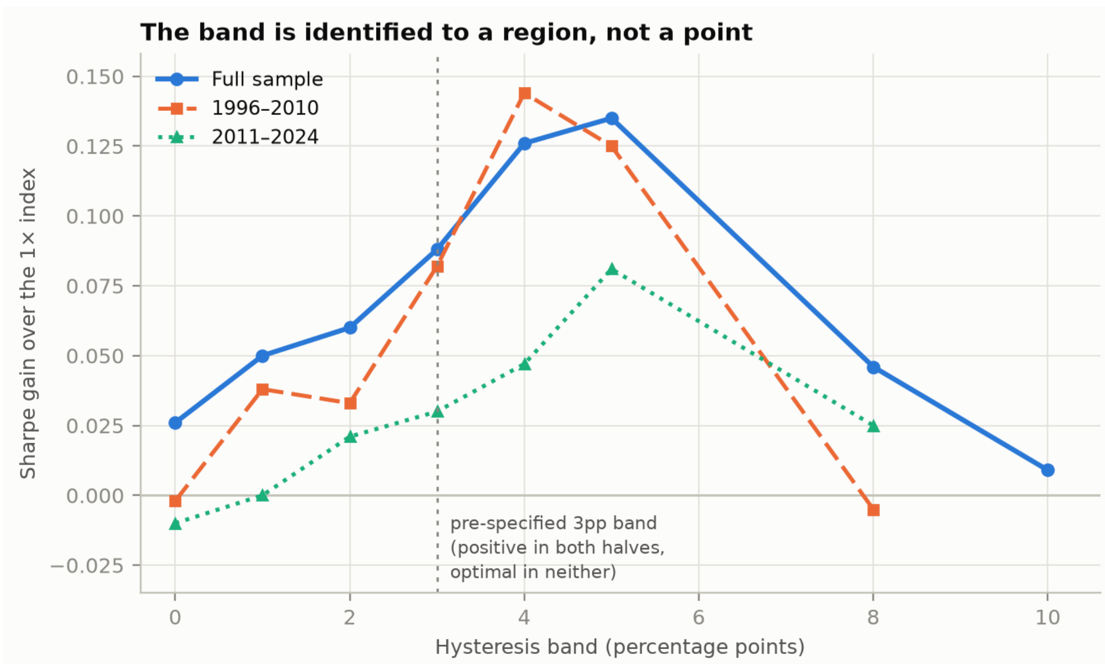

Figure 3: The hysteresis band sweep of Table 6, plotted as Sharpe gain over the 1× index so that the full sample and the two halves share one scale. The argmax moves between halves — 4pp in 1996–2010, 5pp in 2011–2024 — while the shape does not: no improvement below 3pp, a maximum at 4–5pp, and decay by 8pp in all three series. It is the region that replicates, not the point, which is why we report the pre-specified 3pp band rather than a fitted optimum.

Band	CAGR	Sharpe	Info. ratio (t)	1996–2010	2011–2024
0 pp	10.82%	0.404	0.306 (1.64)	−0.002	−0.010
1 pp	11.59%	0.428	0.346 (1.85)	+0.038	−0.000
2 pp	11.85%	0.438	0.361 (1.93)	+0.033	+0.021
3 pp	**12.77%**	**0.466**	**0.412 (2.20)**	**+0.082**	**+0.030**
4 pp	14.09%	0.504	0.482 (2.58)	+0.144	+0.047
5 pp	14.48%	0.513	0.500 (2.67)	+0.125	+0.081
8 pp	11.20%	0.424	0.326 (1.74)	−0.005	+0.025
10 pp	9.14%	0.387	0.174 (0.93)	—	—

Table 6: Hysteresis band sweep, 3× GARCH-only gate. The last two columns give the Sharpe gain over the index within each half of the sample.

The sweep in Table 6 and Figure 3 has an interior maximum on the full sample, but it is *not* identified to a point. Re-running it separately on each half places the optimum at 4pp in 1996–2010 and at 5pp in 2011–2024; no single value wins both. What replicates is the *shape* — essentially nothing below 3pp, a maximum at 4–5pp, and decay by 8pp — which establishes that hysteresis performs real work rather than absorbing noise. We report the pre-specified 3pp throughout because it is positive in both halves and optimal in neither, which is the conservative choice and the one that avoids selecting a parameter on the sample used to evaluate it. That choice costs 0.047 of Sharpe against the full-sample maximum.

5.6 The SKEW null result

The CBOE SKEW index is risk-neutral, and substituting it directly into a polynomial calibrated to physical moments inflates σ^* roughly threefold. Blending in only its deviation from its long-run mean removes the distortion, and leaves an average threshold adjustment of -0.02 percentage points on the 16% of days when SKEW is most elevated. In Table 5 the SKEW rows reproduce the GARCH-only rows to every reported digit.

The usual explanation — that physical skewness is too stable to serve as a dynamic signal — is true but secondary. The stronger statement follows from Section 3.3: the cubic and quartic terms of the break-even polynomial move σ^* by roughly 0.15 percentage points at realistic calibration, two orders of magnitude below the 3-point decision band. *No* movement in skewness within any plausible range can change a switching decision. The null is a property of the polynomial, not an accident of the sample, and Section 4 confirms it survives a ± 2 standard deviation shock to skewness.

5.7 Look-ahead audit

An earlier version of this study compared the closing VIX of day t against the threshold governing the position that earns day t's return. VIX at the close of t already reflects that day's move, so the rule was de-levering into declines it had observed. The GARCH forecast, the momentum filter and the rolling moments were correctly lagged; only the implied-volatility level was not.

3× specification	Contaminated	Corrected
Band = 3pp, momentum on	0.537	0.410
Band = 0pp, momentum off	0.838	0.351

Table 7: Sharpe ratios before and after correcting a one-day look-ahead in the implied-volatility gate.

Table 7 gives the magnitude. Beyond inflating levels, the error manufactured two false conclusions — that the momentum filter and the hysteresis band should both be removed — because both changes raise switching frequency and so use the peek more often. Under corrected timing the band sweep reverses direction entirely and the momentum comparison narrows. We report this at length because the diagnostic generalises: any signal that improves monotonically with trading frequency, in a strategy whose economic rationale is regime detection rather than high-frequency prediction, should be checked for timing before it is believed.

5.8 Robustness and limitations

We report six limitations. Each is quantified, because a limitation without a magnitude is a disclaimer rather than a finding.

The edge is concentrated in the first half of the sample. Splitting at 2010, the 3× daily-gated rule improves on the index by 0.082 of Sharpe over 1996–2010 and by 0.030 over 2011–2024. At 3× with weekly reset the split is starker: +0.121 and −0.002. The strategy is defensive by construction — it earns its advantage by declining to hold leverage through high-volatility regimes — so a fourteen-year period containing no sustained crisis is precisely where it should add least. This is not a robustness failure so much as a statement of what the product is, but it means full-sample statistics overstate what an investor should expect in a calm decade.

The one configuration that does *not* decay is 2× with weekly reset, which improves on the index by 0.076 and 0.078 of Sharpe in the two halves. That stability, rather than its full-sample level, is why we treat it as the preferred specification in Section 6.

The hysteresis band is identified to a region, not a point. As Section 5.4 reports, the split-half optima differ — 4pp in the first half, 5pp in the second — while the shape of the sweep replicates in both. The band is therefore identified to a region and not to a value, and the pre-specified 3pp we carry throughout is optimal in neither half. The limitation is that any band we could report is chosen from a range the data does not resolve; the mitigation is that we chose ours before running the sweep.

The autocorrelation coefficient is not stable. Remark 6 gives the decade-by-decade estimates: $\hat{\varphi}$ is statistically distinguishable from zero only in the two decades containing a major crisis, and a rolling five-year window is positive in 14% of cases. The limitation this creates is specific. Where only the sign of φ matters — the direction in which the reset horizon shifts the daily-comparable threshold — a sample drawn from the 1990s alone would reverse our conclusion. We therefore report $\varphi = 0.30$ as a labelled scenario alongside the estimate rather than discarding it, and treat the direction of the reset-horizon result as conditional on a sign estimated over a full cycle rather than as a structural constant.

The empirical reset-frequency result of Section 6 does *not* inherit this limitation: it is a backtest of realised fund constructions, not an extrapolation from an estimated parameter.

Reset-calendar sensitivity. A fund that resets every W days is one of W distinct funds, depending on which day the calendar starts. We re-ran every one: five weekly calendars, twenty-one monthly, twenty-one quarterly. At 2×, weekly reset beat daily on four of the five calendars (Sharpe 0.438 to 0.502, against 0.439 for daily); at 3× it beat daily on all five, but by a margin of only 0.002 on the full sample, which we treat as noise. The result is not an artefact of one alignment. Equally, the wipeouts are not: a 3× quarterly-reset fund was destroyed on *all* twenty-one calendars, and a 4× monthly fund on eighteen of twenty-one.

The unguarded fund construction. Our multi-period-reset funds hold notional exposure and borrowed principal fixed between resets, with no intra-period intervention. A fund constructed this way fails permanently when the index falls $100/L$ percent inside a single block: -50% at $L = 2$, -33% at $L = 3$, -25% at $L = 4$. Live products carry intraday rebalancing triggers and circuit breakers precisely to prevent this, so our construction should be read as establishing why such guardrails are necessary at longer reset horizons rather than as a forecast of what a real product would do. The design implication — that the maximum safe reset horizon shortens as leverage rises, and that weekly is the only frequency safe across 2× through 4× — survives either reading.

The weekly advantage is specific to 2×. At 2×, weekly reset improves Sharpe over daily by 0.036 (0.476 against 0.439). At 3× the same comparison gives 0.002 (0.468 against 0.466), which is within noise. The reset-frequency result should therefore be stated as a property of moderate leverage rather than of leveraged funds generally: at higher multiples the un-rebalanced tail risk described below offsets the threshold benefit almost exactly, and beyond weekly it dominates.

Scope. All results are for the S&P 500 over 1996–2024, a single index over a single sample containing two major crises. Leveraged returns are constructed from the index rather than taken from live funds, since no 3× S&P product existed before 2008 and none exists at 4×. The financing spread is assumed at 40 basis points over the T-bill; Section 3.5 shows the threshold is more sensitive to this assumption than to any other input in the model, which makes it the single most valuable parameter to replace with a practitioner's actual cost of funds.

6 Optimal Product Design

6.1 Switching to $1\times$, not to cash

The defensive stance drives the strategy. Shifting to cash provides an implied optionality on market timing: the portfolio sells out of equities that decision, not any feature of the leverage ratio itself. None of the levered strategies do this, and no one wanting to do it would require σ^* in order to achieve it.

Shifting to $1\times$ leads to a maximum draw-down of $-\mathbf{56.78}\%$ — the same, down to the basis-point, as that of the index.The leverage rule does not reduce draw-downs. Instead, it takes advantage of $\mathbf{12.77}\%$ per year where the index earns 8.33% while avoiding increasing draw-down.

6.2 Backtest results

Table 8 reports the full scorecard, and Figure 4 shows what it looks like as a path.

	CAGR	Vol	Sharpe	Sortino	MaxDD	IR	Turn.	Time lev.
$1\times$ buy-and-hold	8.33%	19.2%	0.378	0.531	−56.8%	—	—	—
$2\times$ buy-and-hold	9.15%	38.3%	0.353	0.496	−91.2%	0.329	—	—
$3\times$ buy-and-hold	6.90%	57.5%	0.362	0.508	−98.8%	0.353	—	—
$4\times$ buy-and-hold	0.69%	76.6%	0.366	0.514	−99.9%	0.362	—	—
$2\times$ gated, fixed nominal	10.97%	24.8%	0.439	0.612	−61.5%	0.402	43	0.499
$2\times$ gated, constant ERP	9.44%	26.1%	0.377	0.521	−66.0%	0.254	67	0.633
$3\times$ gated, fixed nominal	12.77%	29.8%	0.466	0.644	−56.8%	0.412	49	0.443
$3\times$ gated, constant ERP	11.74%	32.3%	0.427	0.583	−66.9%	0.355	65	0.560
$4\times$ gated, fixed nominal	11.80%	34.5%	0.423	0.577	−56.8%	0.330	43	0.400
$4\times$ gated, constant ERP	13.65%	35.6%	0.466	0.644	−62.6%	0.402	65	0.462

Table 8: Final scorecard under the GARCH-only gate, band = 3pp, defensive position $1\times$, daily decision frequency, 1996–2024. IR is the information ratio against the $1\times$ index. Turnover is the total number of position changes over the sample; time in leverage is the fraction of trading days spent at $L\times$ rather than $1\times$. The Sharpe and Sortino ratios of the $3\times$ fixed-nominal and $4\times$ constant-ERP rows coincide at three decimal places but are distinct: 0.4661 against 0.4663, and 0.6440 against 0.6445. The agreement is a coincidence of rounding and carries no interpretation.

Two findings carry the table. First, static leverage is risky: $3\times$ buy-and-hold returned 6.90% against the index's 8.33% with a −98.8% maximum drawdown, and $4\times$ returned 0.69% with −99.9%. Second, the gated rule is robust through the sample split. Splitting at 2010, the $3\times$ daily-gated rule improves on the index by 0.082 of Sharpe over 1996–2010 and by 0.030 over 2011–2024, and $2\times$ with weekly reset improves by 0.076 and 0.078. We report both halves

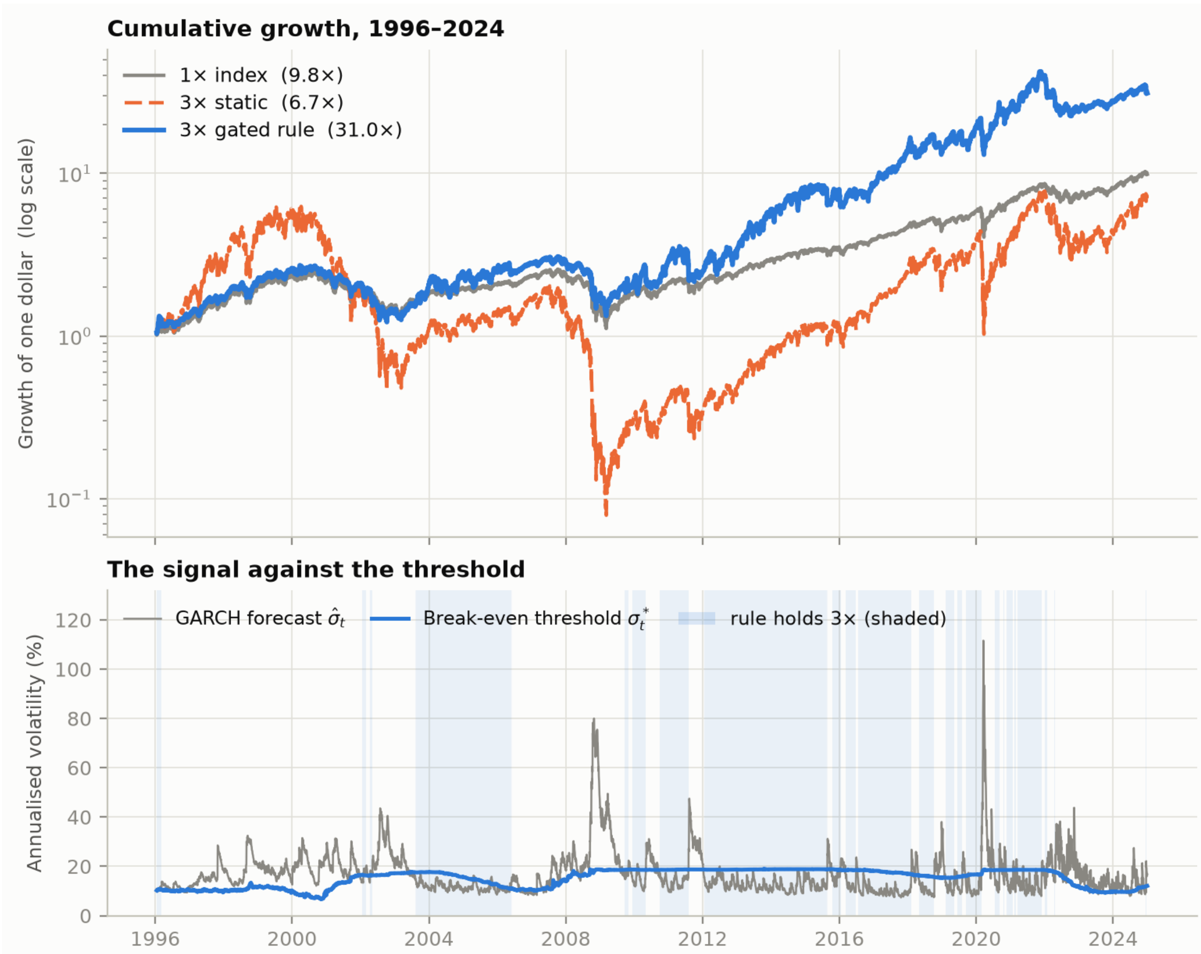

Figure 4: Top: cumulative growth of one dollar at 3×, log scale, net of financing, fees and switching costs. Static 3× ends below the index it levers — 6.7 against 9.8 — having lost more than 90% of its value twice; the gated rule ends at 31.0. Bottom: the GARCH forecast $\hat{\sigma}_t$ against the break-even threshold σ_t^*, with shading on the days the rule holds 3×. The rule is defensive through 2000–2003 and 2008–2009 and levered through the calm stretches, which is the entire mechanism: the threshold is crossed repeatedly rather than approached once, and it is the crossings that do the work.

because a strategy that works only after 2010 is a strategy that has been fitted to one regime, and the reader is entitled to see which of ours survive that test and which do not.

6.3 Rebalancing horizon

Two objects are distinguished by the W notation and need to be distinguished. The *fund's* reset cycle is the factor σ^* depends on, as Proposition 2 reveals that the annualized threshold does not depend on it when returns are independent. The fact that observed returns exhibit mean reversion ($\hat{\varphi} = -0.0841$) means that a fund whose reset happens more rarely will have a smaller variance over several periods than a daily-reset one, hence it will have a *higher* break-even threshold. At $L = 3$, and charging the 95 basis point expense ratio, the daily threshold is 24.2% against a monthly-equivalent threshold of 26.2% using the estimated autoregressive coefficient, or 27.8% using the measured variance ratio $VR(21) = 0.756$ directly. Under either estimator the direction is the same. Such an infrequently reset leveraged product would stand more volatility than those that are now being used — which is a design implication, since no such products are available now.

However, the *strategy's* frequency of decisions is a separate question, and the answer is not the same. Re-evaluating the gate every 1, 5 and 21 trading days gives Sharpe ratios of 0.439, 0.476 and 0.418 at $L = 2$, and 0.466, 0.468 and 0.436 at $L = 3$. Weekly decisions are the best at both leverage levels and monthly the worst. At $L = 3$ the spread across the three is 0.032 of Sharpe, small enough that the choice can reasonably be made on turnover; at $L = 2$ it is 0.058, large enough that weekly evaluation is preferable on performance grounds alone. Weekly evaluation also cuts the number of position changes — by roughly a fifth at $L = 2$ (43 to 35) and a third at $L = 3$ (49 to 35) — and monthly evaluation roughly halves them (26 and 22 respectively), though at a cost in Sharpe at both leverage levels.

6.4 A recommendation framework

The results above reduce to a rule an investor or an issuer can apply without re-estimating anything. Three inputs are required: the leverage under consideration, the prevailing cost of funds, and a forecast of volatility over the intended holding period.

Three properties of Table 9 carry the practical content. The threshold falls with leverage, so a 4× product requires a calmer market than a 2× product to be worth holding — the opposite of the intuition that higher leverage is a more aggressive bet on the same view. It falls steeply with the financing rate, by roughly 1.5 percentage points per point of cost at $L = 3$, so an issuer's own funding advantage translates directly into a wider usable volatility range. And every entry sits within the empirical support of realised S&P 500 volatility, which averaged 19.2% over our sample: the threshold is not a corner case but a line the market crosses repeatedly, which is what makes it usable as a regime signal rather than a theoretical curiosity.

Financing rate r_f	$L = 2$	$L = 3$	$L = 4$
1%	20.02%	17.99%	16.27%
2%	18.30%	16.56%	15.01%
3%	16.39%	14.99%	13.63%
4%	14.23%	13.24%	12.10%
5%	11.67%	11.21%	10.33%

Table 9: Break-even volatility σ^* at $\mu = 8\%$ with a 95 basis point expense ratio and the physical moment calibration. Hold the leveraged position while forecast volatility sits below the entry; hold the index above it.

Two cautions attach to using the table directly. The forecast should be of average volatility over the holding period rather than a spot reading, for the reason given in Section 3.2: the fund's terminal value depends on integrated variance. And a rule applied without hysteresis switches roughly seven times a year rather than under two, which Table 6 shows costs more than it saves.

7 Conclusion

Leverage is compensated according to L and punished according to L^2; a volatility at which the two exactly offset. We derive that break-even volatility under geometric Brownian motion, extend it to a quartic admitting arbitrary skewness and kurtosis, and show that the annualized threshold is invariant to the rebalancing horizon when returns are independent. Comparative statics indicate that it depends primarily on the risk-premium spread: a single percentage point of financing cost moves the threshold by 1.5 percentage points at $L = 3$, while the entire higher-moment apparatus moves it by 0.15.

In practice, however, static leverage doesn't work even according to its own lights — 3× buy-and-hold lost ground relative to the index over 1996–2024, with a −98.8% draw-down. A strategy of buying only when the GARCH forecast is lower than the threshold yields 12.77% versus 8.33% at identical maximum draw-down, and an information ratio of 0.412, while making less than two trades a year. The effect is modest, and we report it without embellishment.

Three potential aids fall short. Implied volatility adds nothing above a conditional variance forecast at this horizon. The CBOE SKEW index adds nothing either, because the break-even polynomial's response to skewness is two orders of magnitude below the decision band — a null our theory predicts rather than merely confirms. A trend overlay subtracts. Finally, we report a one-day look-ahead we found in our own implied-volatility gate, which inflated the Sharpe ratio of the headline specification by 0.127 and of the most affected specification by 0.487, and produced two false conclusions along the way.

The practical implications of our results on product design are limited but, we believe, nonetheless interesting. A leveraged product does not require an options data feed to gate itself in a sensible manner; what it requires are a conditional variance model, a correctly specified risk premium, and correct signal timing. The first is standard; the second and third are where most of the sensitivity lies, and the third is where we went wrong.

Three extensions follow directly. First, single-stock leveraged ETFs — the fastest-growing segment of the category — carry considerably higher volatility and fatter tails than an index, which is precisely the setting in which the higher-order moments stop being negligible and the quartic earns its place. Second, the framework transfers to international indices such as the FTSE 100, the Nikkei 225 and the DAX. That transfer is helped by our own null result: because the GARCH-only rule performs at least as well as any specification using implied volatility, it can be applied wherever a local implied-volatility series is unavailable or too illiquid to trust, which is the common case outside the largest markets. Third, the invariance result implies a product that does not yet exist — a weekly- or monthly-reset fund, which tolerates more volatility than its daily-reset counterpart precisely because returns mean-revert.

References

Avellaneda, M. and Zhang, S. (2010). Path-dependence of leveraged ETF returns. *SIAM Journal on Financial Mathematics*, 1(1), 586–603.

Bevilacqua, M. and Tunaru, R. (2021). The SKEW index: extracting what has been left. *Journal of Financial Stability*, 53, 100816.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492.

Bollerslev, T. and Todorov, V. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1), 113–134.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3), 1311–1341.

Hsieh, C.-H., Chang, J.-R. and Chen, H. H. (2025). Compounding effects in leveraged ETFs: beyond the volatility drag paradigm. arXiv:2504.20116 [q-fin.PM]. Preprint, under review.

Cheng, M. and Madhavan, A. (2009). The dynamics of leveraged and inverse exchange-traded funds. *Journal of Investment Management*, 7(4), 43–62.

Elyasiani, E., Gambarelli, L., and Muzzioli, S. (2021). Skew index: descriptive analysis, predictive power, and short-term forecast. *North American Journal of Economics and Finance*, 55, 101334.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5), 1779–1801.

Lu, L., Wang, J., and Zhang, G. (2012). Long term performance of leveraged ETFs. *Financial Services Review*, 21, 63–80.

Trainor, W. J., Jr. (2026). Beyond volatility decay: correcting relative expected return estimates for leveraged exchange traded funds. *Journal of Risk and Financial Management*, 19(1), 20. DOI:10.3390/jrfm19010020.

Trainor, W. J., Jr. and Carroll, M. G. (2013). Forecasting holding periods for leveraged ETFs using decay thresholds: theory and applications. *Journal of Financial Studies and Research*, 2013, Article ID 715425, 13 pages. DOI:10.5171/2013.715425.

Harvey, C. R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, 55(3), 1263–1295.

Jarrow, R. A. (2010). Understanding the risk of leveraged ETFs. *Finance Research Letters*, 7(3), 135–139.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4), 917–926.

Kraus, A. and Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *Journal of Finance*, 31(4), 1085–1100.

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

MacLean, L. C., Thorp, E. O., and Ziemba, W. T., eds. (2011). *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, 51(3), 247–257.

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A*, 186, 343–414.

Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, 22(1), 27–59.

Securities Litigation and Consulting Group (2009). *Leveraged ETFs, Holding Periods and Investment Shortfalls*. Working paper.

Li, Z. (2026). Leverage break-even point model with central moments. Unpublished manuscript, George Mason University. Available on request.