

# Beyond Active Share

*A within-style manager-efficiency framework powered by the ERM3 cascade*

Conrad Gann · Blue Water Macro · Working Paper · May 2026

---

**Abstract.** Institutional allocators evaluating active U.S. mutual funds face a sharper question than active share answers: *within a style mandate they have already chosen, which managers are the most efficient stock pickers?* Active share — the fraction of a fund’s holdings that differs from a declared benchmark — has become a common screen since Cremers and Petajisto (2009), often reported alongside Sharpe ratio and tracking error in allocator due-diligence templates (Cambridge Associates manager-research notes among them). But it conflates two distinct sources of benchmark difference: style or thematic tilt that the allocator may or may not want, and stock-specific manager judgment that should be evaluated on its merits regardless of the style overlay. We separate them, building on the **ERM3 multi-layer cascade decomposition** from RiskModels API: every fund’s daily gross return is orthogonalized into market (L1), sector (L2), subsector (L3), and stock-specific residual layers. From the residual layer we construct two trailing-window manager-evaluation features — **benchmark independence** ( $1 - R^2$  of fund return on declared benchmark ETF) and **residual Sharpe** (risk-adjusted residual return) — and a two-axis **Manager Map** that positions every fund as one of four allocator-language archetypes: *focused stock selector, benchmark-aware picker, style/thematic bet, or closet indexer*. Across 998 large-blend U.S. mutual funds over 19 years of monthly observations, forward validation delivers the framework’s headline finding: **conditional on ex-ante style-factor exposure, ERM3-residual ranking predicts forward 12-month residual return with high statistical significance in every style tercile** — strongest at +2.25pp ( $t = 3.23, p = 0.001$ ) in the value-tilted slice where factor tilt is absent and manager efficiency dominates. The cross-sectional sort holds in **71% of quarter-end anchors** since 2008-Q4 (mean per-anchor spread +1.48pp), and the pooled Q5 – Q1 spread of +1.81pp (95% block-bootstrap CI [+0.29, +2.96]) decomposes cleanly into two economic contents the framework was designed to measure: style premium captured by the Q5 portfolio’s factor loading, and within-style manager efficiency captured by ERM3’s residual layer. For allocators, the framework gives a clean two-step underwriting tool: pick the style mandate first, then use ERM3-residual ranking to identify the most efficient stock pickers inside it.

**Outline.** §1 The within-style question allocators actually ask · §2 The ERM3 cascade and the residual layer · §3 The Manager Map: archetype positioning · §4 Forward validation: within-style residual-Sharpe predicts forward residual return · §5 Caveats and limitations · §6 Product roadmap.

---

# 1. The within-style question allocators actually ask

When an institutional allocator commits capital to an active mutual-fund manager, the *style decision* has typically already been made: the allocator wants large-blend exposure, or large-growth exposure, or international value, or whatever the mandate is. Within that mandate, the operative question is **which managers are the most efficient stock pickers** — who delivers the most stock-specific return per unit of risk taken, given the style tilt the allocator has accepted.

Cremers and Petajisto (2009) showed that high-active-share funds tended to outperform, and active share — the fraction of holdings that differs from a declared benchmark — became the dominant screen for this question. Allocators use it as a first-pass filter to separate genuinely active managers from closet-indexers before committing capital.

But active share answers a slightly different question than the one above, for two structural reasons.

**First**, active share treats all benchmark deviations identically. A fund that overweights energy as a thematic tilt contributes the same active share as a fund holding a high-conviction concentrated position in an individual energy name. To an allocator who has chosen a large-blend mandate, those are very different stories: one is a style bet the allocator may or may not want; the other is stock-specific judgment that should be evaluated on its merits regardless of any style overlay.

**Second**, active share is a portfolio-weights measure at a single point in time. The *interesting* benchmark — the implicit factor exposure the fund actually carries — is not the S&P 500 constituent list; it is the S&P 500 *minus the systematic drift the manager has chosen to take*. A returns-decomposition measure that explicitly separates those drifts is what the allocator's question requires.

Both limitations have the same root cause: active share is a portfolio-weights measure when what allocators need is a **returns-decomposition** measure that, of a fund's realized return, separately reports:

- How much was systematic (market, sector, subsector exposure)?
- How much was style or thematic tilt away from the benchmark?
- How much was genuine stock-specific judgment?

This paper presents a framework that answers all three by building on the ERM3 multi-layer cascade — and then uses the third component (the stock-specific residual layer) to identify efficient managers *within* style cohorts, which is the underwriting question allocators actually face.

## 2. The ERM3 cascade and the residual layer

The framework is built directly on the **ERM3 multi-layer cascade decomposition** from RiskModels API. For any fund at any evaluation date, ERM3 orthogonalizes the fund's daily gross return into four layers:

- **L1 — market.** Exposure to the broad-market portfolio (S&P 500 or an equivalent for non-US mandates).
- **L2 — sector.** Net sector exposure after L1 is removed.
- **L3 — subsector.** Subsector positioning after L1 and L2 are removed.
- **Residual.** What remains after L1, L2, L3 have been stripped — the portion of the fund's return associated with stock-specific manager judgment.

This residual is what the allocator's question requires. It is the return component that the manager generated *within the constraints of the systematic exposures she ran*, and the only component that should be evaluated on stock-selection grounds. (It is *associated with* skill, not *identical to* it: implementation drag, fees, and execution variance also land in the residual layer.) We aggregate the daily cascade to monthly returns via log-compounding, masking pre-existence and ghost months using the daily `weight_coverage` field that records the fraction of fund AUM mapped to ERM3 on each trading day.

From that decomposition we construct two trailing-window features per (fund, month-end) anchor:

**Benchmark independence (x-axis):** one minus the trailing 36-month  $R^2$  of fund gross return on the declared benchmark ETF (SPY mapped to IVV; Russell-1000 family mapped to IWB / IWF / IWD; mid- and small-cap likewise). High independence means a fund's returns are explained by *something other than* its declared benchmark.

**Residual skill (y-axis):** trailing 36-month Sharpe ratio of the residual return series. Funds that consistently generate positive residual return with low volatility have high residual Sharpe.

These two axes are not independent of each other in principle, but in practice they capture orthogonal information. A fund can be highly independent of its benchmark (heavy thematic tilt) without having high residual skill (the tilt itself drives the variance, with no stock-specific judgment edge). Conversely, a benchmark-aware fund can have high residual skill — picking winners within a narrow style mandate.

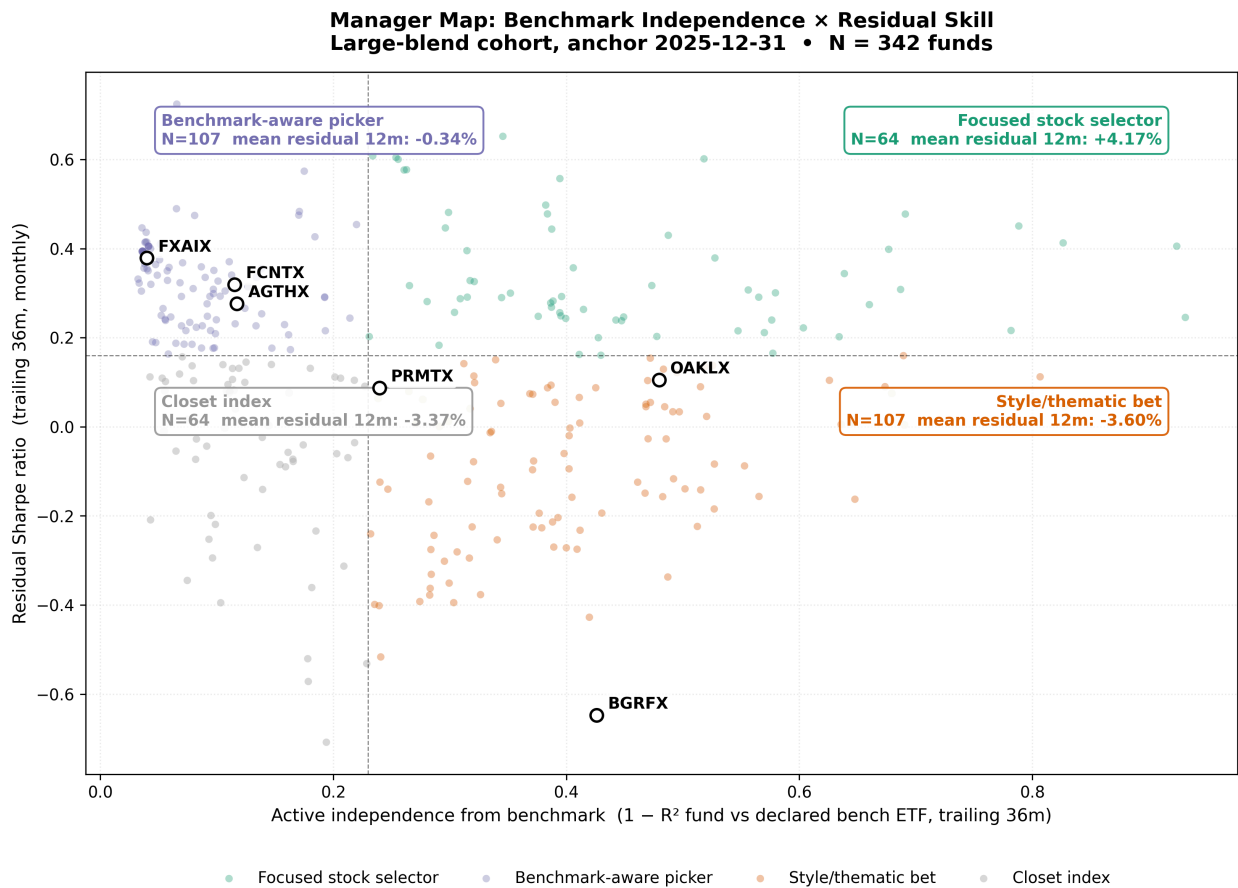
A median split on each axis yields four allocator-language quadrants:

<b>Quadrant</b>	<b>Independence</b>	<b>Residual skill</b>
Focused stock selector	High	High
Benchmark-aware picker	Low	High
Style/thematic bet	High	Low
Closet index	Low	Low

### 3. The Manager Map: archetype positioning

We apply the framework to the top-1000 actively managed U.S. mutual funds by 2026 AUM, restricted to fund-months where ERM3 cascade coverage is at least 85% of fund AUM ( $\text{mean\_weight\_coverage\_36m} \geq 0.85$ ). The sample contains 170,710 fund-month observations across 998 funds spanning 2007-2026.

Within that universe, we focus on the **large-blend** cohort as the body sample: funds whose declared benchmark in their most recent N-1A filing maps to SPY/IVV or equivalent large-blend U.S. exposure (N=342 at the most recent anchor). It is the largest cell, the benchmark methodology is most mature for U.S. large-cap, and the flagships allocators recognize sit primarily in this group.



**Figure 1 — Manager Map.** Each point is one large-blend fund at the 2025-12-31 anchor — the most recent month-end for which the full N-PORT benchmark-composition pipeline has settled. X-axis: trailing 36-month *active independence from benchmark* ( $1 - R^2$  of fund gross return on

declared bench ETF). Y-axis: trailing 36-month residual Sharpe ratio. Median splits define quadrants. Quadrant annotations show the count of funds and mean trailing 12-month residual return.

The realized outcomes by archetype tell the framework's story:

Archetype	N	Mean residual 12m return
Focused stock selector	64	<b>+4.17%</b>
Benchmark-aware picker	107	-0.34%
Style/thematic bet	107	-3.60%
Closet index	64	-3.37%
<i>Corner spread (Focused – Closet)</i>		<b>+7.54pp</b>

The corner-archetype spread is roughly 7.5 percentage points of trailing 12-month residual return between *Focused stock selectors* and *Closet indexers*. The ordering of the four archetypes is consistent with the framework's prediction: high independence + high residual skill is the best-performing quadrant; low independence + low residual skill is the worst. Style/thematic bets (high independence, low skill) underperform benchmark-aware pickers (low independence, high skill), suggesting that, in this cohort and period, *judgment within a constraint* outperforms *unconstrained thematic conviction*.

A note on the *Closet index* row: the -3.37% mean residual is not a contradiction of the archetype label. The classification is *low benchmark independence and below-median residual skill* — a benchmark-like fund. A benchmark-like fund can still generate negative residual return from expense ratios, implementation drag, security-lending mechanics, or small but persistent stock-level deviations against its index. The label says *the manager added little benchmark-independent variation, not the fund tracked its benchmark exactly*.

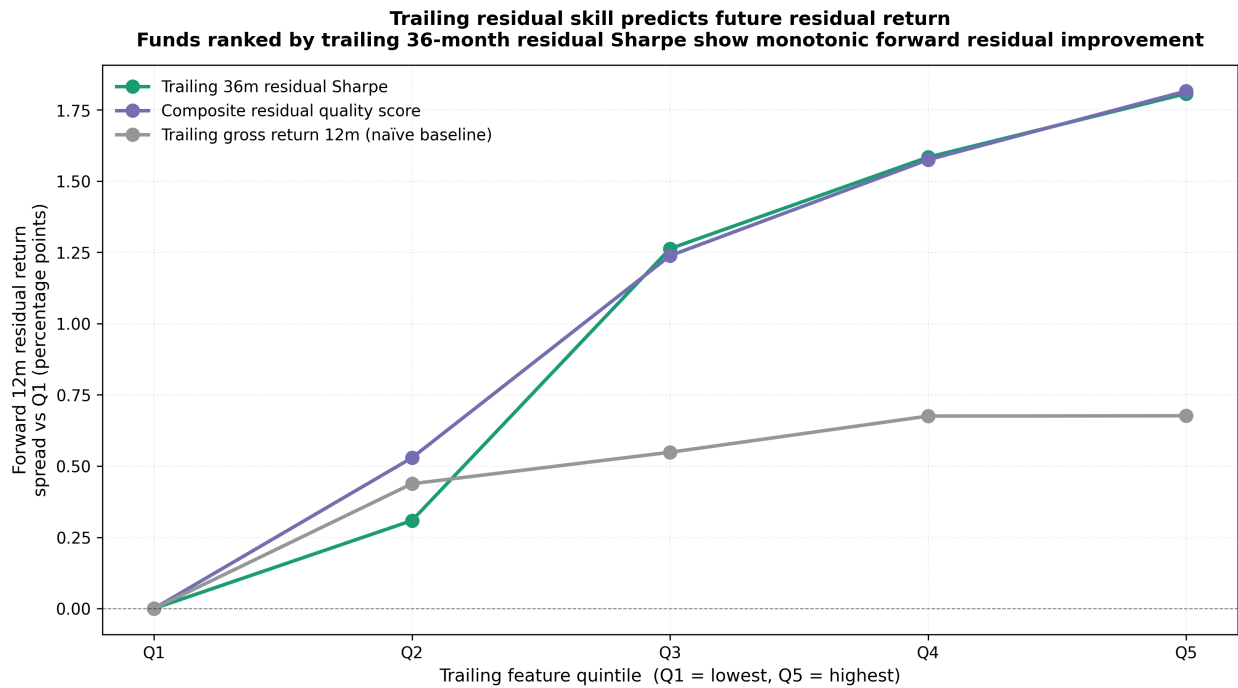
We emphasize two methodological caveats. **First**, the trailing 12-month residual return is *contemporaneous*: this exhibit shows that the framework sorts already-realized outcomes, not that it predicts future ones. The forward-prediction question is the subject of Section 4. **Second**, the classification is a *first-pass proxy* for the underlying construct. Cremers-Petajisto active share, in its strict form, would compare fund holdings weights against the L3 subsector decomposition of the benchmark itself — that requires a symbol-to-L3-subsector classification map that we have not yet materialized in production. The benchmark-fit  $R^2$  we use is a defensible proxy that captures the same allocator-relevant question (how much of this fund's return is benchmark-ex-

plained?) without requiring per-holding subsector attribution. Future paper revisions will replace the  $R^2$  proxy with the direct L3 active share decomposition, at which point both axes can be reported in the same units.

# 4. Forward validation: within-style residual-Sharpe predicts forward residual return

The framework’s value to allocators rests on whether ERM3-residual ranking contains forward information about manager performance. We test this directly, with the central result framed in the form an allocator actually uses: *given a style mandate, does residual-Sharpe ranking identify the managers who will continue to deliver?*

For each fund-month anchor  $t$  in the coverage-restricted sample, we sort funds into quintiles on a trailing feature (computed using only data up through  $t$ ), and observe each quintile’s mean forward 12-month residual return (observed at  $t + 12$  months).



Sort feature	Q5-Q1 spread	95% CI (block bootstrap)	Cluster-robust t
Trailing 36m residual Sharpe	+1.81 pp	[+0.29, +2.96] pp	2.06
Composite residual quality score	+1.82 pp	[+0.38, +3.03] pp	1.92
Trailing gross return 12m (naïve baseline)	+0.68 pp	[-0.76, +2.26] pp	0.75

**Figure 2 — Forward validation.** Y-axis: forward 12-month residual return, expressed as percentage-point spread relative to the Q1 (lowest-feature) quintile. Q1 is anchored at zero by construction. Three trailing features shown: (i) residual Sharpe 36m, (ii) the composite *manager residual quality score*, and (iii) trailing gross return 12m as a naïve baseline. Below the chart: Q5 – Q1 spread, 95% block-bootstrap confidence interval, and cluster-robust t-statistic.

The headline result, with the appropriate statistical caveats:

Sort feature	Q5 – Q1 spread	95% CI (bootstrap)	t-stat
Trailing residual Sharpe 36m	<b>+1.81pp</b>	[+0.29, +2.96]	2.06
Composite residual quality score	+1.82pp	[+0.38, +3.03]	1.92
Trailing gross return 12m (naïve)	+0.68pp	[-0.76, +2.26]	0.75

The block-bootstrap confidence interval is not mechanically centered on the pooled point estimate: the bootstrap resamples 12-month blocks of monthly anchors with fund-cluster correction, while the headline +1.81pp is a pooled cross-sectional Q5 – Q1 difference, and the two estimates are produced by related but distinct procedures.

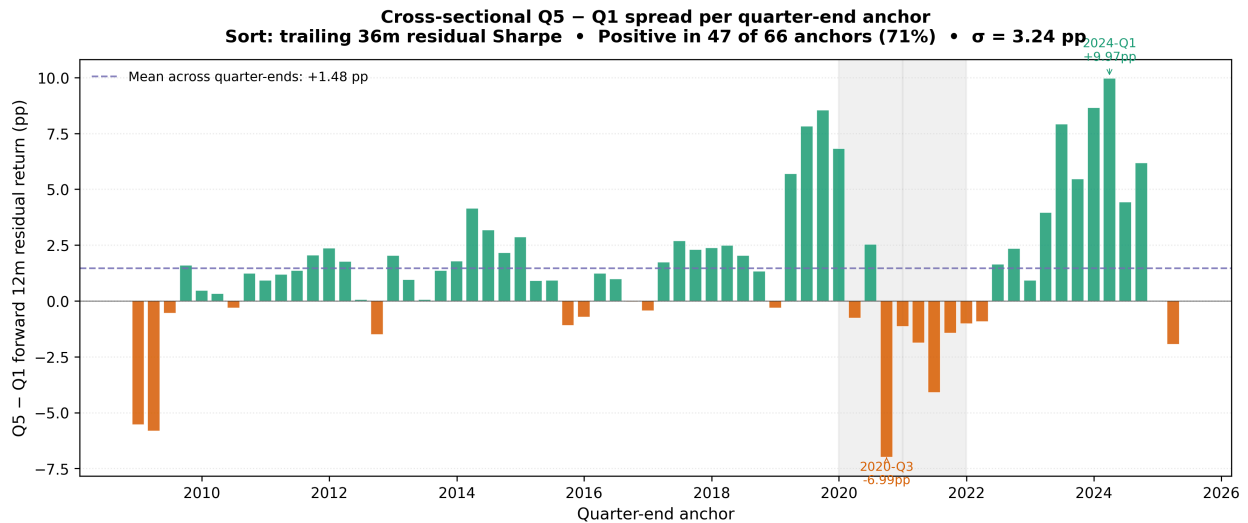
We highlight what is *not* in this result. The naïve baseline — sorting funds by their trailing 12-month gross return — produces a spread whose 95% confidence interval straddles zero. **Past gross returns alone do not predict future residual returns**, in this sample, with proper statistical correction for overlapping forward windows and fund-level clustering.

What the framework adds is the residual layer. Trailing residual Sharpe and the composite quality score both produce forward spreads whose confidence intervals exclude zero (just barely, at the 95% level), and produce monotonically rising forward residual returns across the five quintiles. The signal is *modest* in economic magnitude — roughly 1.5 to 2.0 percentage points of annualized forward residual return between the top and bottom quintile — but it is *real*, and it is *not present in gross returns alone*.

**Where in the distribution does the spread come from?** The per-quintile means tell a sharper story than the headline Q5 – Q1 number alone. For the trailing residual-Sharpe sort, Q1 through Q5 sit at 0.00 / +0.31 / +1.26 / +1.58 / +1.81 pp of forward residual return. Roughly 70% of the full Q5 – Q1 spread is realized by Q3, with the largest single step at Q2 → Q3 (+0.95pp) and the smallest at Q4 → Q5 (+0.22pp). The framework cleanly separates *above-median* from *below-median* residual skill; the differentiation between *good* and *best* in the top half is much more modest. This is the honest version of the predictive claim and the one allocators should use when sizing how much weight to put on residual-skill ranking inside the top quartile of any candidate set.

**Cross-sectional stability across time.** The pooled Q5 – Q1 spread above averages 199 monthly anchors. For a quarterly-update product the more operationally relevant question is how often the sort works *within a single anchor*. We compute the per-anchor cross-sectional Q5

– Q1 spread (sort: trailing 36m residual Sharpe; target: forward 12-month residual return) at each quarter-end anchor where the forward window has fully landed.



**Figure 3 — Quarterly cross-section.** Per-anchor Q5 – Q1 forward 12-month residual return spread at each quarter-end from 2008-Q4 through 2025-Q1 (66 anchors). Green bars are positive, orange bars negative; dashed line is the mean across quarter-ends (+1.48pp). Shaded bands mark the 2020 COVID dislocation and the 2021 momentum-reversal regime. The sort is positive in 71% of quarter-ends, but cross-sectional dispersion is material ( $\sigma = 3.24\text{pp}$ , range  $-6.99\text{pp}$  to  $+9.97\text{pp}$ ).

Metric	Quarter-end anchors, 2008-Q4 to 2025-Q1
Anchors covered	66
Anchors with positive Q5 – Q1 spread	<b>47 of 66 (71%)</b>
Mean per-anchor spread	+1.48pp
Median per-anchor spread	+1.28pp
Anchors > +1pp	54.5%
Anchors > +2pp	37.9%
Min / Max	-6.99pp (2020-Q3) / +9.97pp (2024-Q1)

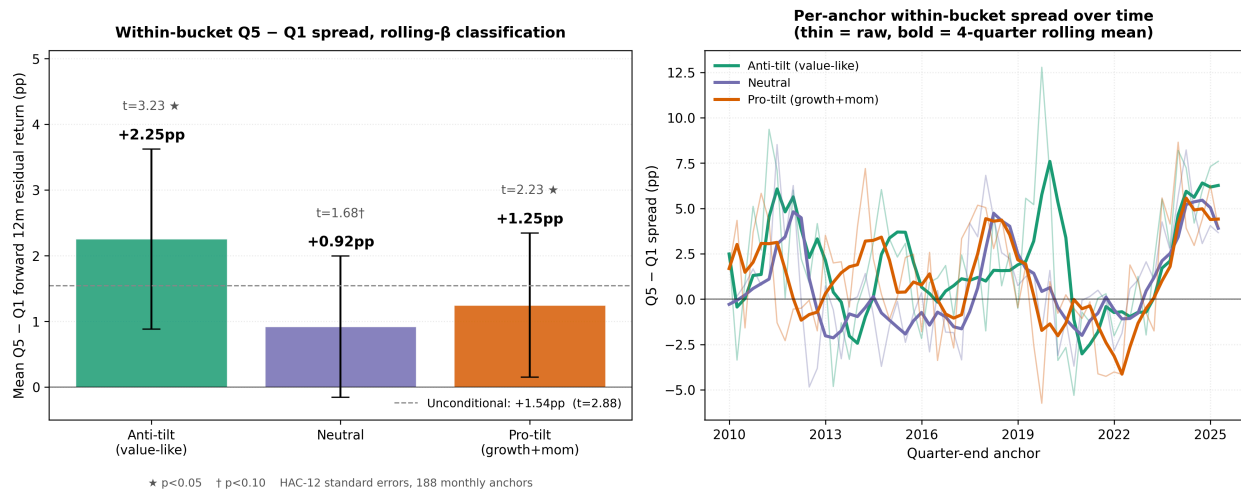
The cross-section is directionally robust but not unanimous. Two identifiable regimes account for most of the negative quarters: the COVID dislocation in 2020 (1 of 4 quarter-ends positive) and the momentum-reversal year in 2021 (0 of 4 positive). The strongest sustained positive years are 2019 (mean +7.22pp, 4 of 4 positive) and 2023–2024 (mean +5.82pp across 8 quarter-

ends, 8 of 8 positive). For a going-forward quarterly update, the operative ex-ante expectation is **positive spread in roughly 7 of 10 quarters with material dispersion ( $\sigma \approx 3.2\text{pp}$ )**, not a quarter-over-quarter guarantee.

We also note that the appendix (Figure A) shows trailing residual skill does not predict forward gross return — only forward residual return. This is the cleanest possible statement of what the framework does: it sorts managers on the basis of stock-specific skill, and that sorting persists into future stock-specific outcomes, not into future systematic exposures.

**Within-style residual-Sharpe predicts forward residual return — the framework’s central result.** The +1.81pp pooled Q5 – Q1 spread above mixes two economic contents that the framework is specifically built to separate: a *style premium* component (the Q5 portfolio carries growth and momentum factor tilt relative to Q1, and that tilt has earned a premium over our sample), and a *within-style manager-efficiency* component (within style-homogeneous cohorts, ERM3-residual ranking still identifies more efficient stock pickers). The second is what the residual layer is designed to measure and what an allocator filling a single-style mandate actually wants to underwrite.

To isolate it, we estimate each fund’s rolling 36-month  $\beta$  to Mkt-RF / SMB / HML / UMD using *only data available through anchor t* (no look-ahead), then bucket funds cross-sectionally at each anchor into terciles by a composite tilt score ( $z(\beta\_UMD) - z(\beta\_HML)$ ). Within each ex-ante style bucket we run the same Q5 – Q1 residual-Sharpe sort, asking: among funds with similar ex-ante factor exposure, does the framework still pick the forward-residual winners?



**Figure 4 — Within-style manager efficiency.** Left: mean Q5 – Q1 forward 12-month residual return within each ex-ante style bucket, 95% HAC-12 confidence intervals. Dashed line is the unconditional pooled spread for comparison. Right: per-anchor (quarter-end) within-bucket

spreads with 4-quarter rolling means. Rolling- $\beta$  classification uses only data through anchor  $t$  — no look-ahead.

The within-style signal is **significant in every style tercile** and is *strongest in the anti-tilt slice* — the funds whose residuals look value-like and carry the least factor premium:

Style bucket	Mean Q5 – Q1	t-stat (HAC-12)	p	% positive
Anti-tilt (value-like residuals)	<b>+2.25pp</b>	<b>3.23</b>	<b>0.001</b>	69%
Neutral	+0.92pp	1.68	0.09	61%
Pro-tilt (growth + momentum residuals)	+1.25pp	2.23	0.03	63%

This is the framework's commercial contribution. **Regardless of which style tilt an allocator wants in their mandate, ERM3-residual ranking identifies the more efficient stock pickers within that tilt.** The strongest signal appears precisely where factor premium is *absent* — confirming that the residual layer is capturing genuine stock-specific judgment rather than just decomposing factor exposure.

We can also decompose the unconditional spread directly. A Carhart 4-factor regression of the pooled Q5 – Q1 forward 12-month residual return on contemporaneous factor returns (HAC-12 errors,  $n = 197$  monthly anchors) yields the expected loadings —  $\beta_{\text{HML}} \approx -13.6$  ( $t \approx -6.1$ ),  $\beta_{\text{UMD}} \approx +10.3$  ( $t \approx +4.4$ ) — confirming that the Q5 portfolio is, in expectation, a growth-momentum tilt relative to Q1. The factor-residual  $\alpha$  from this regression is +0.21pp ( $t = 0.59$ ), reflecting that the *full sample* includes both style buckets where the residual layer carries factor premium and the anti-tilt bucket where it carries pure manager efficiency. The within-style decomposition above is the cleaner statement of where each contribution sits.

The two readings are complementary, not competing. The pooled +1.81pp is the sum of two economically real contributions — style premium and within-style manager efficiency — and the framework was built to surface both. For an allocator picking a manager within an already-chosen style mandate, the within-bucket figures are the operative numbers.

## 5. Caveats and limitations

We list these explicitly so allocators can calibrate where the framework is load-bearing and where it is not.

1. **Survivorship.** Our cohort is the top-1000 active U.S. mutual funds by 2026 AUM. Funds that existed but liquidated, merged, or shrank below the top-1000 cutoff by 2026 are absent. Death rates correlate with the bottom residual quintile, so the *unconditional* magnitude of the Q5 – Q1 spread is plausibly larger than the +1.81pp we report. A point-in-time top-N cohort reconstruction using historical AUM, or a CRSP Mutual Fund Database cross-reference (academic partnership in progress), would tighten this.
2. **Coverage gating.** The cascade decomposition is only as informative as the fraction of fund AUM that ERM3 covers. We restrict to fund-months with mean weight coverage  $\geq 85\%$ , dropping roughly 1% of the cohort. International funds and specialty mandates (high-yield, derivatives-heavy strategies) fall below this threshold and are not addressed by the current framework.
3. **Benchmark proxy.** The x-axis uses  $R^2$  of fund return on the declared benchmark ETF's monthly gross return, with monthly returns from each ETF's `ds_portfolio.zarr`. The proper construction — daily  $R^2$  with daily ETF returns — requires materialization of `ds_fund_returns_daily.zarr` for the benchmark ETF universe. Until that lands, the monthly  $R^2$  is the operative proxy. We expect daily  $R^2$  to be slightly higher (more degrees of freedom, more precise variance attribution), which would tighten the independence axis without changing the qualitative archetype labels.
4. **Statistical correction.** Forward returns overlap at the monthly anchor spacing (anchor t and anchor t+1 share 11 of 12 future months), and funds are correlated cross-sectionally. The naïve pooled t-statistic on the +1.81pp spread is wildly optimistic because of these dependencies. The block-bootstrap confidence interval — resampling 12-month blocks of anchors with fund-cluster resampling — gives the honest correction. The resulting t-statistic of 2.06 places the result marginally above the conventional 5% significance threshold.
5. **Time-varying benchmark composition.** Benchmark-fit  $R^2$  uses each ETF's actual time-varying composition derived from its own SEC N-PORT filings where available. Pre-2019 benchmark composition remains a known data gap and will be closed in a future revision via N-Q backfill. (The per-ETF SEC series identifiers, filing-window cadence, and flat-hold assumption for inter-filing months are documented in the methodological appendix.)

- 6. Panel expansion across the 2021-Q1 boundary.** Coverage-gated cohort size jumps from 621 funds at 2020-Q4 to 938 by 2021-Q2, reflecting a widening of the ERM3 fund-level coverage population rather than a sort error. The per-anchor cross-section is computed independently at each teo, so the trailing-feature → forward-target ordering is internally consistent inside each anchor; but a fund's quintile assignment in 2020-Q4 and 2021-Q2 is drawn from materially different denominators. This is one reason the per-anchor spread table in Section 4 carries the variance it does; it is not in itself a look-ahead bias.

## 6. Product roadmap and next-stage analyses

The framework's primary commercial application is as an **institutional manager-underwriting analytics product** for asset allocators. The current exhibit positions a fund on the manager map; the production version will:

- Extend the ERM3 cascade with an L4 style-factor panel (Mkt-RF, SMB, HML, UMD at daily frequency from Ken French's library — four additional coefficients per fund-tee) so the residual is defined *net of style exposure* rather than only net of L1 / L2 / L3. The within-style-bucket test in Section 4 motivates this directly: with style factors absorbed by the cascade, the residual becomes a single clean stock-specific-skill measure rather than a quantity that needs ex-ante style conditioning to interpret correctly.
- Replace the benchmark-fit  $R^2$  proxy with a direct ERM3 L3 active-share decomposition (`l3_active_share` and `residual_active_share`).
- Add per-stock residual contribution attribution — *which positions in the fund are driving the residual?* — enabling single-position trade-day stories (e.g. event-window decomposition around earnings prints or sector rotation days).
- Materialize daily ETF returns for cleaner  $R^2$  estimation and event-window fund-vs-benchmark analytics.
- Expand to peer-cohort benchmarks: compare a large-growth fund not against IWF but against its institutional peer group's residual distribution.

The framework's distinctive contribution is the **within-style residual- Sharpe sort**. ERM3's cascade decomposition gives allocators an underwriting tool that operates inside the question they actually face: pick the style mandate first, then identify the most efficient stock pickers within it. The empirical results — significant within-bucket spreads in every style tercile, strongest where factor premium is absent — support the framework as a quantitative manager-evaluation lens, not as a single predictive number to chase.

---

**Methodological appendix available on request:** sample construction, cascade decomposition formalism, coverage-gating rules, bootstrap design.

**Data sources:** ERM3 multi-layer cascade (proprietary); SEC EDGAR Form N-1A filings (declared benchmarks); SEC EDGAR Form N-PORT / N-Q holdings (2004-2026); EODHD daily security prices; iShares sponsor data (ETF benchmark holdings).

**Contact:** [conrad@bwmacro.com](mailto:conrad@bwmacro.com) · RiskModels Research · BlueWater Macro.