RESEARCH NOTE RN-03 · FACTOR ATTRIBUTION, VALIDATION & RISK GOVERNANCE

Six-Factor Residual Attribution, Purged Walk-Forward Validation & *Volatility-Parity Risk Governance.*

Making money because the whole market or tech sector went up isn't stock-picking edge. Here is how we strip out GICS Sector and five style factors via Ridge regression, test scouts under 5-fold Purged Walk-Forward Cross-Validation and `Drop-1` Jackknife gates, and manage stop-loss path risk across market regimes.

AUTHOR: **STARLING QUANTITATIVE RESEARCH** PUBLISHED: **NOVEMBER 8, 2026** DESK: **STARLING QUANTITATIVE**

SECTION 01

Separating True Stock-Specific Alpha from Disguised Sector & Style Beta

If a strategy buys high-beta tech stocks during a bull market and beats the S&P 500 ($r_i - r_{\mathrm{SPY}}$), it hasn't necessarily found any stock-picking edge—it just took extra market, sector, or momentum risk that any institutional portfolio can already buy for almost zero fee. An alternative data scout is only worth funding if its predictions still make money *after* you strip out the sector it sits in and the common style factors driving the broader market.

That is why Starling Quantitative grades every scout, isolated sub-arm (arm_id), calibration weight, and promotion gate strictly on **6-dimension stock-level residual returns** ($R_{i,\mathrm{resid}}$) rather than raw benchmark-excess returns. Subjecting our historical research catalog to 5-fold Purged Walk-Forward evaluation across 3,607 Russell 3000 CIKs pruned over 75% of initial hypotheses: standalone technical price-momentum, pair-trading, and lagged policy screens turned negative after factor neutralization and transaction costs (for example, standalone technical momentum averaged -28.53 bps net 6-factor residual return across $N = 256$ Russell 3000 events, $\mathrm{PF} = 0.86$, and -66.37 bps in Fold-5 training) and were permanently retired; bearish forensic and product-recall sleeves were converted into $0 long-veto shields with zero single-stock inverse-ETF routing (while preserving target_ticker in Feed Tier A for institutional short books); and un-replayed feeds were quarantined to $0 shadow logging. What survived were structural primary-filing and microstructure mechanisms—such as Form 8-K Accelerated Share Repurchase (ASR) dealer-covering contracts, Form 4 insider clusters & net-retained option exercises,

sector-relative $T+3$ crash-stabilized zero-distress rebounds, single-stock follow-through-day compounder bids, and Form 10-K ASC-280 captive-supplier propagation.

SECTION 02

Six-Dimension Stock-Level Barra-Style Ridge Residualization

For each single-stock event $(i,\ t_{\text{entry}})$, daily returns are demeaned against the constituent's point-in-time **GICS Sector ETF** ($r^{\text{sec}}_{i,d} = r_{i,d} - r_{\text{ETF}(g(i)),\,d}$), and ℓ_2/Ridge factor loadings $\hat{\boldsymbol{\beta}}_i \in \mathbb{R}^5$ are estimated strictly inside the pre-entry window $[t_{\text{entry}} - 252,\ t_{\text{entry}} - 1]$ against five standardized OHLCV/liquidity style factors $\mathbf{f}_d = (\text{Mkt}_d,\ \text{Size}_d,\ \text{Mom}_d,\ \text{STR}_d,\ \text{Vol}_d)^\top$—with **zero pre-entry sample intercept subtracted from post-event out-of-sample returns** so pre-catalyst drift never contaminates post-event attribution:

$$R_{i,\text{resid}} = s_i \Big(R_{i,[t_{\text{entry}},t_{\text{exit}}]} - R_{\text{sec}(i),[t_{\text{entry}},t_{\text{exit}}]} - \hat{\boldsymbol{\beta}}_i^{\top} \mathbf{F}_{[t_{\text{entry}},t_{\text{exit}}]} \Big) - c_{\text{rt},i}$$

where $s_i \in \{-1, +1\}$ is the thesis sign, $\mathbf{F}_{[t_{\text{entry}},t_{\text{exit}}]}$ chains exact intraday entry ($\text{Open}(t_{\text{entry}}) \to \text{Close}(t_{\text{entry}})$) and multi-day holding ($\text{Close}(t_{\text{entry}}) \to \text{Exit}$) factor returns, and $c_{\text{rt},i}$ deducts round-trip half-spreads (plus a baseline $150\ \text{bps/yr}$ general-collateral short-borrow assumption on bearish evaluations). The five cross-sectional style dimensions span:

1. **Market** (Mkt): Broad market excess return exposure.
2. **Liquidity / Size** (Size): Cross-sectional small-minus-big quintile spread ranked by $-\text{ADDV}_{20}$ (inverse 20-day Average Daily Dollar Volume).
3. **12–1 Month Momentum** (Mom): Medium-term trend factor ($\text{Close}_{t-21} \,/\, \text{Close}_{t-252} - 1$; Jegadeesh & Titman, 1993).
4. **5-Day Short-Term Reversal** (STR): Trailing 5-trading-day (1-week) reversal factor ($-(\text{Close}_{t-1} \,/\, \text{Close}_{t-6} - 1)$; Lehmann, 1990).
5. **14-Day ATR Volatility** (Vol): Trailing 14-day Average True Range percentage ($\text{ATR}_{14} \,/\, \text{Close}_{t-1}$) realized volatility proxy.

Both scout Information Coefficients $\widehat{\text{IC}}_{\text{OOS},s} = \text{Corr}_{\text{rank}}(\tilde{z}_{s,i,t},\ R_{i,\text{resid}})$ and post-trade realized attribution are computed exclusively on the 6-dimension orthogonalized residual $R_{i,\text{resid}}$. Because this internal 6-dimension model uses point-in-time OHLCV and sector ETF proxies rather than commercial fundamental accounting factors (such as Book-to-Price Value, Earnings Yield, or Profitability), institutional counterparties evaluating Feed Tier A under mutual NDA are invited to residualize our point-in-time ScoutPitch stream directly against their own full commercial Barra (USE4 / USFAST) or Axioma risk models.

SECTION 03

Anchored 5-Fold Purged Walk-Forward CV & Deflated Sharpe Ratio

Standard k-fold cross-validation violates temporal ordering and leaks overlapping multi-day return windows across train/test boundaries. Following López de Prado (2018), Starling validates all backtest-eligible scouts across an anchored **5-Fold Purged Walk-Forward Cross-Validation** protocol over the survivorship-free Russell 3000 panel:

$$\mathcal{T}_{\text{train}}^{(m)} \cap \left[t_{\text{test_start}}^{(m)} - \Delta_{\text{embargo}},\ t_{\text{test_end}}^{(m)} + \Delta_{\text{embargo}} \right] = \varnothing, \qquad \Delta_{\text{embargo}} = 12 \text{ trading days}$$

Any training observation whose forward evaluation horizon overlaps the test fold—plus an explicit **12-trading-day embargo buffer** spanning our $H_{\max} = 10\text{-trading-day}$ canonical evaluation horizon plus the 2-trading-day Dual-Maturity settlement lag—is purged prior to fitting quantile tables, MAD scales, and shrinkage weights. Across the pooled 5-fold out-of-sample evaluation (2022–2026) on the 3,607-CIK Russell 3000 panel, walk-forward pruning cleanly separated structural catalysts from noise: while the unfiltered single-stock catalog ($N = 7,740$ events across 121 non-overlapping cohorts) averaged $+6.49$ bps net 6-factor residual return ($\text{PF} = 1.03$, $t_{\text{NW10}} = +0.65$), the walk-forward Admitted pool ($N = 794$ events across 94 non-overlapping 10-day cohorts) achieved $+70.10$ bps per event ($\text{PF} = 1.26$, event $t_{\text{NW10}} = +2.03$; tier-weighted non-overlapping cohort mean $+68.15$ bps, $\text{PF} = 1.88$), and the Graduated Core pool ($N = 218$ events across 54 non-overlapping cohorts) achieved $+144.57$ bps net 6-factor residual alpha per event ($\text{PF} = 1.54$, event $t_{\text{NW10}} = +2.05$; tier-weighted cohort mean $+114.43$ bps, $\text{PF} = 1.91$).

Two-Sided Tier-A Attribution vs. Long-Only Internal Execution: Note that the $N = 794$ Admitted ($+70.10$ bps) and $N = 218$ Graduated ($+144.57$ bps) pools evaluate two-sided signal quality across both bullish long catalysts and bearish short-direction disclosures (such as statutory Form NT late-filing notices, which are preserved as single-name short signals in Feed Tier A and evaluated at a baseline 150 bps/yr general-collateral borrow fee; institutional short books should stress-test distressed small-cap names against elevated hard-to-borrow specials fees). Isolating the **long-only internal executable subset** ($N = 576$ **bullish events across** 84 **non-overlapping cohorts, excluding all bearish veto sleeves**) yields $+68.84$ bps **net 6-factor residual return per event** ($\text{PF} = 1.27$, **date-clustered** $t_{\text{NW10}} = +1.64$, **event** $t_{\text{NW10}} = +1.74$).

To guard against selection bias across the historical research catalog, we freeze our cumulative trial denominator at $M = 45$ **evaluated configurations under an Earned-Seat Moratorium**, evaluating out-of-sample Sharpe ratios $\widehat{\text{SR}}$ via the **Bailey & López de Prado (2014) Deflated Sharpe Ratio** (DSR):

$$\mathrm{DSR} = \Phi\left(\frac{\left(\widehat{\mathrm{SR}} - \mathrm{SR}_0(M)\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\,\widehat{\mathrm{SR}} + \frac{\hat{\gamma}_4 - 1}{4}\,\widehat{\mathrm{SR}}^2}}\right)$$

Transparent Multiple-Testing Disclosure ($M = 45$ Across $T \in [54, 94]$ Sparse Event Cohorts): In a 4.8-year historical window comprising $T = 54$ to 94 non-overlapping 10-day event cohorts, penalizing openly for all $M = 45$ catalog trials sets an extreme-value maximum-Sharpe hurdle of $\mathrm{SR}_0(45) \approx 2.57$ annualized. Against that hurdle, the Graduated Core pool's realized $\widehat{\mathrm{SR}}_{\mathrm{ann}} = 1.16$ ($t_{\mathrm{NW10}} = +2.05$) evaluates to $\mathrm{DSR} = 0.0147$ (and $\mathrm{DSR} = 0.0002$ across all 94 Admitted cohorts at $\widehat{\mathrm{SR}}_{\mathrm{ann}} = 0.894$). Rather than gaming the denominator by pretending rejected scouts were never tested, we publish the full $M = 45$ penalty, enforce a hard New-Scout Moratorium so M cannot inflate further, and require live out-of-sample forward cohorts before scaling internal capital.

SECTION 04

How a Scout Earns Real Money: Dual-Maturity Gating & the Drop-1 Jackknife

One of the hardest operational questions in a live system is how quickly to promote a new scout out of shadow trading to spend real capital. If you simply check a scout's closed trades every night, you fall into a subtle trap: trades that hit a stop-loss on Day 2 close out right away, whereas trades that ride a full 10-day holding window take two full weeks to settle ("fast-stop / horizon-settlement maturity bias"). Conversely, if a scout gets lucky on one massive 40% buyout, its average return will look fantastic even if its other 19 trades lost money. We solve both problems with three mathematical rules:

1. **Dual-Maturity Cohort Gating:** A trade opened on date t_{entry} and closed on $t_{\mathrm{actual_exit}}$ is allowed to update a scout's promotion score at evaluation date t_{eval} if and only if *both* inequalities hold:

$$t_{\mathrm{actual_exit}} + 2\mathrm{d} \le t_{\mathrm{eval}} \qquad \text{and} \qquad t_{\mathrm{entry}} + H_{\max,s} + 2\mathrm{d} \le t_{\mathrm{eval}}$$

meaning the system waits until the entire entry cohort has reached its full scheduled horizon plus a 2-day settlement buffer before judging the scout.

2. **Drop-1 Leave-One-Out Jackknife Blend & Upsizing Cap:** One lucky outlier should never promote or double-size a weak signal. For every scout s, we compute its leave-one-out mean residual return $\hat{\mu}_{s,(-1^*)}$ after deleting its single largest winning trade, and define its outlier-robust mean $\hat{\mu}_{\mathrm{rob},s}$ via a 50/50 jackknife blend:

$$\hat{\mu}_{s,(-1^*)} = \min_{1 \le j \le N_s} \frac{1}{N_s - 1} \sum_{m \neq j} R^{\text{net}}_{s,m,\text{resid}}, \qquad \hat{\mu}_{\text{rob},s} = \min\left(\hat{\mu}_s, \;\; \frac{1}{2}\hat{\mu}_s + \frac{1}{2}\hat{\mu}_{s,(-1^*)} \right)$$

Both the scout's Bayesian posterior Sortino Information Ratio ($\text{IR}_{\text{post},s} > 0$) and its continuous Kelly multiplier are governed by $\hat{\mu}_{\text{rob},s}$, and Kelly upsizing is capped near baseline (blocking multi-unit scaling whenever $\hat{\mu}_{s,(-1^*)} \le 0$).

3. **Kish's Effective Sample Size (N_{eff}), Isolated Sub-Arm Quotas & Veto Counterfactual Attribution:** When observations carry exponential time-decay or conviction weights w_m, sample-size hurdles are evaluated using Kish's (1965) effective sample size $N_{\text{eff}} = \left(\sum_m w_m\right)^2 / \sum_m w_m^2$ rather than raw trade count N. Probationary sub-arms (arm_id = scout_name::sub_arm_regime) earn or lose promotion on isolated sub-arm ledgers under two-level (arm, family) slot quotas without contaminating parent core arms, while settled vetoed candidates in shadow_trades undergo continuous out-of-sample counterfactual attribution to verify that every active veto gate saves net residual alpha.

SECTION 05

Volatility-Parity Sizing, Crowding Guards, Volume-Clock Decay & Portable-Alpha Core

Unconstrained multi-asset Kelly sizing ($\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$) is notoriously sensitive to estimation error in expected returns on continuous equity distributions. For approved common-equity candidates clearing the $\eta_{\text{cost}} \cdot \text{RTCost}_{i,t}$ round-trip friction hurdle ($\eta_{\text{cost}} > 1$) via Barra-implied fair-midpoint limit collars (tracked alongside post-fill markout and unfilled-collar shadow TCA), Starling computes active weights via **Grinold active risk-aversion sizing** bounded simultaneously by **inverse-ATR constant-dollar stop-risk parity** and a hard single-issuer ceiling $w_{\max}$:

$$w_{i,t} = \min\left(\frac{\hat{\alpha}_{\text{rem},i,t} - \text{RTCost}_{i,t}}{\lambda_{\text{active}} \cdot \hat{\sigma}^2_{\text{resid},i,H}}, \;\; \frac{r_{\text{stop_max}}}{\text{eff_stop_pct}_{i,t}(\text{ATR}_{14})}, \;\; w_{\max} \right)$$

Crucially, as analyzed by **Kaminski & Lo (2014)** and **Lo & Remorov (2017)**, mechanical stop-loss and cash-switching policies on assets exhibiting return serial correlation, regime switching, and non-trivial transaction costs suffer severe whipsaw drag when switching thresholds trigger high turnover, whereas low-turnover regime-aware rules preserve capital across persistent macro breakdowns. In our 1,216-day out-of-sample Russell 3000 evaluation (2022–2026), a traditional 2-Gear SMA_200 trend-switching rule executed 38 full equity-to-cash regime switches and compounded at only $+38.45\%$—underperforming a constant 1.00x portable-alpha equity core ($+61.76\%$) by -23.31% due to switching friction and re-entry

lag after V-bottom reversals, directly corroborating the high-turnover whipsaw penalty modeled by Lo & Remorov (2017). Motivated by these path-dependence findings, Starling enforces four practical microstructure and portfolio rules:

▪ **Volume-Clock Subordinated Alpha Decay & Opportunity-Cost Replacement:** Rather than decaying strictly by calendar days, open position alpha decays over volume-clock subordinated elapsed time $d_{\text{eff}} = d_{\text{held}} \cdot \text{clip}\left(\sqrt{V_t / \text{ADV}_{20}},\, 0.75,\, 2.00\right)$ inside $[H_{\min},\, H_{\max}]$, exiting once remaining holding alpha drops below one-way exit friction; when active slots are full, a fresh graduated candidate replaces a mature non-loss incumbent only when net continuation advantage clears a strict positive basis-point hurdle.

▪ **Regime-Aligned Stop Boundaries, Volatility-Scaled Ratchets & Two-Sweep Confirmation:** Post-filing drift sleeves pair ATR_{14}-scaled stops and pre-exit stock-split reconciliation with volatility-scaled high-water-mark profit ratchets, whereas short-horizon liquidity rebounds reject pure GICS sector waterfalls, require zero SEC/FINRA distress filings and a multi-day ($T + 3$) crash-low stabilization gate before entry, and verify two consecutive daemon sweeps outside the stop boundary to avoid liquidating into transient single-print L1 vacuums.

▪ **24-Hour Post-Stop Quarantine, Ledoit-Wolf Clusters & Pairwise Crowding Guards:** Any issuer stopped out is quarantined from re-entry for 24 hours to eliminate whipsaw churn, while a rolling 60-day covariance kernel enforces hard Ledoit-Wolf GICS cluster caps, Percentage Contribution to Total Risk (PCTR) ceilings, and market-residualized pairwise factor crowding guards.

▪ **1.00x Cash-Funded Portable-Alpha Core (Zero Margin Debt & 10-Day Dwell + Contango Guard):** Over 1,216 out-of-sample trading days (2022–2026), maintaining a constant 1.00x cash-funded equity/Treasury core (managed with symmetric [-30d, +30d] wash-sale partner rings, partial whole-share Treasury trims, ex-dividend sawtooth guards, beta-aware funding waterfalls, and $0 settled margin debt) paired with our Purged-CV active overlay compounded at $+61.76\%$ ($+10.48\%$ CAGR, $\text{IR}_{\text{HAC}} = +0.807$, $\text{Max DD} = -25.47\%$ vs. SPY $+54.33\%$, $\text{Max DD} = -25.36\%$), avoiding the -23.31% whipsaw drag of moving-average cash switches ($+38.45\%$) while enforcing a 10-trading-day minimum dwell guard and VIX term-structure contango exemptions on volatility circuit breakers.

SECTION 06

Selected References & Technical FAQ

• Bailey, D. H., & López de Prado, M. (2014). The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.

• Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.

• Kaminski, K. M., & Lo, A. W. (2014). When do stop-loss rules stop losses? *Journal of Financial Markets*, 18, 234–254.

• Kish, L. (1965). *Survey Sampling*. John Wiley & Sons.

• Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, 105(1), 1–28.

• Lo, A. W., & Remorov, A. (2017). Stop-loss strategies with serial correlation, regime switching, and transactions costs. *Journal of Financial Markets*, 34, 1–15.

• López de Prado, M. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

How does Starling verify that its alternative signals are not simply disguised Market Beta or Momentum?

Every candidate signal is evaluated against stock-specific residual returns after demeaning by GICS Sector ETF and stripping out five standardized OHLCV/liquidity style factors—Market (Mkt), 20-Day Dollar-Volume Size (Size), 12–1 Month Momentum (Mom), 5-Day Short-Term Reversal (STR, Lehmann 1990), and 14-Day ATR Volatility (Vol)—via trailing 252-day Ridge regression. Signals whose raw excess returns are spanned by common style factors are permanently retired.

Why is a 12-trading-day embargo required in 5-fold Purged Walk-Forward Cross-Validation?

When evaluating multi-day holding horizons, adjacent training and testing observations share overlapping return intervals. Enforcing a 12-trading-day embargo buffer (López de Prado, 2018)—spanning our 10-trading-day maximum evaluation horizon plus a 2-trading-day Dual-Maturity settlement buffer—between training fold cutoffs and test fold starts prevents serial overlap from leaking test-period price shocks into training calibration.

What role do Drop-1 Jackknife tests and Dual-Maturity Cohort Gating play in live scout promotion?

Our 50/50 Drop-1 leave-one-out jackknife blend caps a scout's posterior mean residual return against its leave-one-out mean ($\hat{\mu}_{\mathrm{rob},s} = \min\left(\hat{\mu}_s, \frac{1}{2}\hat{\mu}_s + \frac{1}{2}\hat{\mu}_{s,(-1^*)}\right)$) and caps Kelly upsizing near baseline (blocking multi-unit scaling whenever $\hat{\mu}_{s,(-1^*)} \leq 0$), preventing single-outlier promotion. Dual-Maturity Cohort Gating evaluates trade cohorts only after every constituent position reaches its full scheduled horizon plus a 2-day settlement buffer, eliminating fast-stop / horizon-settlement maturity bias.

LIVE OUT-OF-SAMPLE ATTRIBUTION

© 2026 STARLING QUANTITATIVE · A BIWADA COMPANY

REGULATORY DISCLOSURE: Starling Quantitative is a proprietary quantitative research and internal capital allocation desk of Biwada LLC. Nothing in this monograph constitutes investment advice or an offer to solicit outside retail capital or manage third-party funds.