The Index Effect Before the Announcement: Selection-List Front-Running, Residual Tracking Costs, and the Attenuating Premium in the STOXX Europe 600

A quantitative study of STOXX Europe 600 quarterly index reviews, 2014–2025

v1.2 — 4 July 2026 · all numbers from audited artifacts; see the provenance manifest

Contents

Abstract

We dissect the index effect in Europe's most-tracked benchmark, the STOXX Europe 600, using 1,219 constituent changes over 2014–2025 on a survivorship-free daily panel, a five-window decomposition of the review timeline, and inference that respects the batching of events into 49 quarterly review cycles. Because STOXX selects constituents by published rule from a published monthly ranking, the setting isolates how markets price a predictable index change. The answer is: before it is announced. Scheduled additions earn a median +410 basis points and deletions −476 between the selection-list publication and the announcement — the only windows that survive both a peer-placebo null and review-cycle bootstrap — and the run-up concentrates in marginal candidates whose inclusion the rule had not yet resolved. After the announcement little remains: the apparent give-back and post-effective reversal fail cluster-honest inference (12 of 31 naively significant cells die under clustering), closing-auction pressure is detectable only on the deletion side and only marginally (median +45 basis points, interval touching zero), neither side shows a statistically identified post-effective reversal, execution cost has no cross-sectional structure at the close, and is not forecastable out of sample from strictly point-in-time features; only a flow-times-arbitrage-risk interaction shows structure, and even its precision leans on a single review

cycle. The implied hidden cost to trackers is 2.8 to 5.1 basis points per year — roughly four to eight times below Petajisto's S&P 500 benchmark. Per unit of passive demand, the addition-side premium is directionally attenuating but not statistically confirmed, and straddles a disclosed data seam, while deletion-side repricing persists. The European index effect is alive at selection, dead at execution, and cheap for the funds that track it.

---

1. Introduction

For four decades the index effect was the cleanest anomaly in equity markets: a stock added to a major benchmark rose on no news about its cash flows, a deleted stock fell, and the pattern was strong enough to name a cottage industry of rebalancing arbitrage. The most consequential recent finding in this literature is that, in its original habitat, the effect is dying — Greenwood and Sammon document S&P 500 announcement returns collapsing toward zero as index-change arbitrage matured into a competitive business. Yet almost everything we know quantitatively about the effect, its costs, and its decay comes from US indices governed by committee discretion. Europe's most-tracked benchmark has never received the same autopsy at the same evidentiary standard, and it differs from the S&P 500 in exactly the dimension the theory cares about: information.

The STOXX Europe 600 selects constituents by published rule from a published ranking. A monthly selection list discloses every candidate's position weeks before anything is announced; a buffered threshold — entrants must rank 550 or better, incumbents survive until 750 — makes most changes mechanically predictable; and the calendar itself is fixed by rule, with the selection list preceding the announcement by about 21 trading days and the announcement preceding implementation by about 15. Where the S&P framework asks how markets absorb a *surprise*, the STOXX framework asks how they price a *schedule*. That makes the European case the natural laboratory for the questions passive investing actually poses at scale: how much of the effect moves ahead of the announcement when the announcement is anticipatable, what residual cost still reaches the funds that must trade mechanically, and whether the effect decays here as it did in the US.

This paper answers those questions on 1,219 constituent events over 2014–2025, built on a 34-million-row daily panel with survivorship-free coverage of deleted names, an authoritative review-timeline calendar, and event cohorts that separate ordinary quarterly changes from the M&A-, spin-off-, and distress-driven changes that contaminate naive samples. Four findings organize everything that follows. First, the index effect is alive but has moved upstream of the announcement: scheduled additions carry a median +410 basis points and deletions −476 between the selection-list date and the announcement — the only windows in the entire study that decisively beat both a peer-placebo null and review-cycle-clustered confidence bands — and the run-up concentrates almost entirely in *marginal* candidates whose fate the published rule had not yet settled — a concentration Section 4.1 is careful not to over-read, because selection into the addition sample conditions on exactly that late rank movement; the selection-resistant evidence comes from the deletion side, where names the rule had already condemned still fall roughly 280 basis points before anything is announced. Second, everything after the announcement is small or fragile: the post-announcement "give-back" fails placebo and bootstrap scrutiny, closing-auction pressure is detectable only on the deletion side and only marginally, and neither side shows a statistically identified post-effective reversal — the near-complete unwind documented for the US has no European counterpart in our windows. Third, the residual cost to index trackers is bounded between 2.8 and 5.1 basis points per year — roughly four to eight times below Petajisto's S&P 500 benchmark — and the execution layer behind it is cross-sectionally unstructured, out-of-sample unforecastable, and economically conditional only on the interaction of passive flow with hedging difficulty. Fourth, the addition-side premium per unit of passive demand is directionally attenuating across our sample, concentrated where the effect itself lives, but twelve years of data cannot statistically certify the decline — and the deletion side shows no attenuation at all.

Three methodological commitments distinguish the exercise, and we flag them because they change conclusions, not just error bars. Inference respects the batch structure of index reviews throughout: 827 scheduled events supply only 49 independent quarterly clusters, and twelve of the thirty-one window-level results a naive sign test would certify as significant die under cluster-honest inference — a discipline whose absence in older event studies we suspect explains part of the literature's larger reported effects. Every claim is stratified across a disclosed three-era data provenance rather than pooled over a silently stitched panel, which matters most when Section 7 must distinguish economic attenuation from a measurement seam. And the paper maintains a strict separation between what is *predictable* (membership changes, from public rules) and what is *exploitable* (residual returns after the market has priced that predictability) — a distinction the out-of-sample null of Section 5 makes concrete.

The paper speaks to four literatures at once: the classical index-effect canon (Shleifer's demand curves; Petajisto's cost accounting; Patel and Welch's window structure and peer placebo, which we adopt as primary inference), the arbitrage-limits tradition (Wurgler and Zhuravskaya's arbitrage risk, which survives here only as a flow interaction), the disappearing-effect evidence (Greenwood and Sammon, whose multiplier framework we adapt to a buffered single-index setting where their migration channel is structurally zero), and the anomaly-decay literature (McLean and Pontiff, whose out-of-sample discipline our predictability layer operationalizes). Section 2 describes the data and its seams; Section 3 fixes methodology and the inference doctrine; Sections 4 through 7 present results by question; Section 8 collects robustness; Section 9 concludes with what an index-fund desk, an index provider, and a would-be arbitrageur should each take away.

---

2. Data and sample construction

2.1 A three-era price and return spine

The empirical backbone of this study is a unified daily security-level panel covering 34,366,585 observations on 27,650 securities between 2 January 2012 and 1 April 2026 . The panel is deliberately wider than the STOXX Europe 600 itself: it retains every security that appears in STOXX's European production universe, so that deleted constituents, declined candidates, and peer stocks remain observable after they leave the index. No survivorship filter is applied at any stage; a security that delists retains its full price history up to its final trading day, and event-study windows that end early because a name stops trading are flagged rather than silently completed.

The spine stitches three data eras, and we disclose the seams rather than present the panel as a single homogeneous source. From January 2012 to 11 October 2017 (the *api* era, 1,825,677 rows on 1,400 securities), prices, shares, and free-float factors are reconstructed from LSEG/Refinitiv data because STOXX's own archive does not reach that far back at daily security level. From 12

October 2017 to 13 June 2023 (the *legacy* era, 17,854,503 rows on 19,456 securities), the panel is built from STOXX's legacy component-level archive. From 14 June 2023 onward (the *ab* era, 14,686,405 rows on 24,716 securities), it consumes STOXX's native daily production feed with no reconstruction of any kind. Each observation carries its era tag, and the tag agrees with the calendar boundaries on every row.

Five properties of the seams matter for interpretation, and we state them here once so that every later section can refer back. First, for non-euro names in the api era, euro price series use LSEG-mid foreign-exchange rates because the STOXX FX archive does not reach earlier than October 2017; LSEG-mid and STOXX-fix rates differ by roughly 0.3% on a given day, a mean-reverting gap with a median euro-level deviation near 0.5%, and the difference affects only the api era. Second, api-era shares outstanding and free-float factors are reconstructed and carry an uncertainty band of roughly ±10–20% against native STOXX free floats; the legacy and ab eras use native STOXX floats. Third, industry classification switches taxonomies at the 2023 seam — ICB-2008 supersector codes before, ICB-2019 codes after — so no analysis in this paper pools sector groupings across that boundary. Fourth, corporate-action coverage is era-scoped by construction: a dedicated 25-event spin-off register covers the api era, while STOXX's own corporate-action feed covers the legacy and ab eras, and LSEG-sourced information is never used to adjust prices inside the STOXX eras. Fifth, deletion coverage in the api era is slightly incomplete: six removed lines (all rights issues or temporary listings routed to the excluded "other" cohort) and one scheduled deletion without a usable price window are absent, whereas retention from October 2017 onward is complete. Events whose estimation or event window straddles a seam are flagged with a dedicated indicator and retained.

Because the era boundaries almost coincide with the calendar thirds of the sample, the period stratification used throughout this paper (Early 2014–2017, Middle 2018–2021, Late 2022–present) is inevitably entangled with reconstruction methodology: Early-period estimates rest almost entirely on api-era data, Middle-period estimates entirely on the legacy archive, and Late-period estimates roughly two-thirds on the native feed. Wherever a time trend carries economic weight — most prominently in the attenuation analysis of Section 7 — we treat the native ab era as the methodologically cleanest stratum and say explicitly when a temporal contrast straddles a seam.

2.2 Events, cohorts, and the review timeline

The event sample contains 1,219 index-change events on 721 securities between 2014 and early 2026: 609 additions and 610 deletions, of which 843 take effect on one of the 49 quarterly review dates and 376 occur mid-cycle. All four timeline anchors of a STOXX review — the interim selection-list publication T_{sl}, the announcement T_{ann}, the underlying-data cut T_{uda}, and the effective date T_{eff} — come from a single authoritative calendar table assembled from STOXX's review-composition files; no date is ever parsed from a file name or reconstructed from a rule of thumb. For the typical cycle the selection list precedes the announcement by about 21 trading days and the announcement precedes the effective date by about 15 .

Events are partitioned into seven cohorts by cause. The scheduled cohort (827 events: 421 additions, 406 deletions) contains ordinary quarterly-review changes and is the headline sample throughout the paper. The remaining cohorts — fast-track entries (173), M&A-driven deletions (134), spin-offs (33), forced removals such as bankruptcies (8), mid-cycle demotions (8), and a residual “other” class (36) that is excluded from analysis — capture changes whose price dynamics are dominated by the triggering corporate event rather than by index membership itself . Sixteen review-date changes that were mechanically stapled to M&A or redomiciliation outcomes are classified into their causal cohorts rather than left in the scheduled sample. No signed level statistic in this paper pools additions and deletions: the index effect is signed, and pooling would average opposite-signed responses toward zero. Two deliberately sign-safe constructions are the only exceptions, each disclosed where it appears — the reversal-ratio slope of Section 4.3, which is also reported per side, and the absolute-value premium blend the Petajisto cost identity of Section 6 requires.

Each event also carries a contamination flag: if a capital-structure corporate action (split, rights issue, spin-off distribution, or similar) falls within five trading days of either the announcement or the effective date, the event is marked contaminated. Headline statistics are reported on the full sample and on the scrubbed (`ex_contam`) subsample in parallel, and Section 8 shows the headline results are insensitive to the scrub.

For scheduled events we additionally observe each candidate’s position on the pre-review selection list. STOXX’s published buffer rule adds a non-constituent ranked 550 or better and retains an incumbent until it falls below rank 750; we therefore label each scheduled event by its selection-list marginality (*core* versus *buffer*) and by whether the buffer rule alone, applied to the last pre-announcement selection list, already predicted the change. Because that list is published about 21 trading days — roughly a calendar month — before the announcement and ranks continue to move afterwards, these labels are point-in-time lower bounds on predictability: a significant core-versus-buffer contrast is informative, while a null would not be.

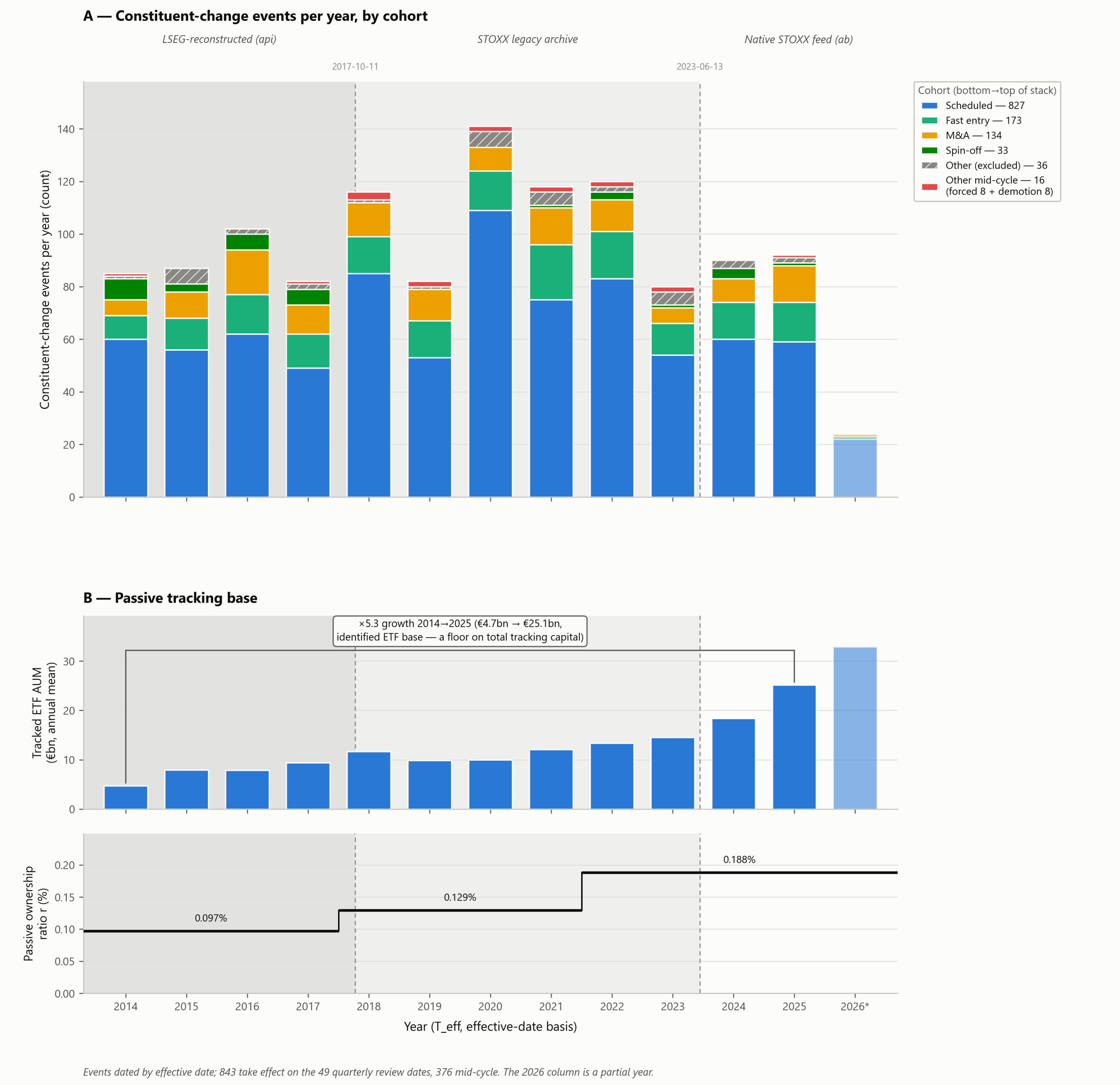

Figure 7. Sample architecture. Panel A: the 1,219 constituent-change events, 2014–2026, by effective year and causal cohort; scheduled quarterly-review changes (827) are the headline sample, mid-cycle cohorts are analyzed separately, and the residual "other" class (36) is excluded. Shaded bands mark the three data-provenance eras of the price panel — LSEG-reconstructed (through 2017-10-11), STOXX legacy archive (to 2023-06-13), and native STOXX production feed thereafter — whose seams every temporal claim in the paper discloses. Panel B: annual-mean assets of the nine identified physically replicating STOXX Europe 600 ETFs (a floor on index-tracking capital) and the implied passive ownership ratio per period. The 2026 column is a partial year.

2.3 Ancillary panels

Four ancillary panels feed specific questions. A passive-ownership panel aggregates the assets under management of the nine physically replicating UCITS ETFs identified as tracking the STOXX Europe 600, at monthly frequency from Datastream, with annual means used throughout; tracked assets grow from €4.7 billion in 2014 to €25.1 billion in 2025. This panel is ETF-only by construction: index funds, segregated mandates, and synthetic replication are not captured, so it is a lower bound on true index-tracking capital — a property whose consequences we carry explicitly into the flow analysis of Section 5, the cost estimates of Section 6, and the multiplier levels of Section 7. A volume panel of 1,685,024 security-days covers 717 of the 721 event securities from LSEG, and average daily value traded is always computed as raw share volume times the raw euro close, a product invariant to share splits. Volume covers the primary listing's lit line only; in Europe's fragmented market total tradable volume is materially higher, so all days-of-ADV figures in this paper are conservative upper bounds on the true demand-to-liquidity ratio . A book-equity panel (33,345 fiscal observations on 2,596 of 2,607 candidate-pool securities) supplies book-to-market inputs with a six-month reporting lag; because the vendor supplies latest-restated rather than first-print values, these inputs are approximately rather than strictly point-in-time, a limitation we disclose wherever book equity enters . Finally, two arbitrage-risk measures are computed per event, described in Section 3.5.

2.4 Benchmarks and return conventions

All abnormal returns in the paper are computed against the STOXX Europe 600 Net Return index (SXXR) using each stock's dividend-reinvested euro return series, so that numerator and benchmark treat dividends identically. The price-return pair (SXXP with unadjusted price series) is used as a robustness benchmark; a 24-cell parallel check of the headline window results found no sign flip under the swap, and the peer-placebo quantiles concord in sign on every headline window (Section 8). The gross-return variant (SXXGR) is excluded by design: it assumes full withholding-tax recovery, which no investable tracker achieves, and would bias abnormal returns downward mechanically. One reconstruction note applies to the native era: STOXX's daily production feed resets its security-level return index at each ex-date rather than compounding it, so the ab-era net-return series was rebuilt into an accumulating total-return series using STOXX's own adjustment factors before any analysis; the rebuild adds the expected $≈$1.7% per year of dividend accrual and all downstream artifacts were regenerated from it. In the LSEG-reconstructed era the net-dividend withholding rate per market is backfilled as a constant at its earliest observable (2018–2020) modal value — a second-order approximation, on the order of a basis point per year, disclosed here.

---

3. Methodology

3.1 Market model

Expected returns follow the single-factor market model estimated per event and benchmark over the window $[T_{ann} - 120, T_{ann} - 21]$ in trading days, with a minimum of 60 valid observations:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \varepsilon_{i,t}, \qquad \widehat{AR}_{i,t} = R_{i,t} - \hat{\beta}_i R_{m,t}.$$

The fitted intercept is deliberately discarded when forming abnormal returns. Index candidates are selected in part on past performance, so $\hat{\alpha}_i$ embeds exactly the run-up the event study is trying to measure; subtracting it would mechanically depress measured abnormal returns (Patel and Welch, 2016). The one place this choice bites is the small pre-selection-list window, where cumulated abnormal returns overlap the estimation window itself and therefore contain a mechanical $\hat{\alpha}$ residue; every table flags that window's overlap share, and no headline claim rests on it.

Return-quality filters are applied before estimation: deletion-day technical prints, stale closes, carried-forward quotes, and their next-day catch-up returns are removed from both estimation and event windows, since a spurious 0% or −100% print inside the estimation window would bias $\hat{\beta}_i$ and the residual volatility on which the parametric tests depend.

3.2 Event windows and terminology

The review timeline is tiled into five non-overlapping windows. Because one of the artifact-level window names collides with the economic term "front-running," the paper fixes its vocabulary here and uses it consistently; the mapping to machine-readable artifacts is part of the replication package.

Paper term	Window	Artifact name
Pre-selection window	$[T_{sl} - 5,\ T_{sl} - 1]$	`pre_leak` (α-overlap flagged)
Pre-announcement run-up	$[T_{sl},\ T_{ann} - 1]$	`sl_to_ann`
Announcement build-up	$[T_{ann},\ T_{eff} - 2]$	`front_run`
MOC window	$[T_{eff} - 1,\ T_{eff}]$	`moc`
Post-effective reversal window	$[T_{eff} + 1,\ T_{eff} + 42]$	`reversal`

Two composite windows recur: the announcement-to-effective window $[T_{ann}, T_{eff}]$ (the sum of the build-up and MOC windows), and the Petajisto-convention net-effect window $[T_{ann} - 10, T_{eff}]$, which anchors the turnover-cost and multiplier analyses because it is the only window whose sign matches the direction a tracking fund actually trades into.

3.3 Inference under review-date clustering

STOXX changes constituents in batches: the 827 scheduled events fall on just 49 quarterly effective dates, roughly seventeen names per cycle, and same-cycle abnormal returns are cross-sectionally correlated through common flows and common market conditions. Treating events as independent therefore overstates precision, and this is not a technicality: of the 31 scheduled SXXR window-cells in our main tables that a naive sign test would call significant at the 5% level, 12 fail once inference respects the 49-cycle structure . The paper's significance language is built entirely on three cluster-honest instruments. Point estimates are medians (abnormal-return distributions around index events are heavy-tailed, and the 2020 COVID cycle and sanction-driven delistings would otherwise dominate means); each cell carries a 95% confidence interval for the median from a block bootstrap that resamples whole review cycles; each scheduled cell carries the quantile of its observed median within a 5,000-replication peer placebo distribution, constructed by re-running the identical event pipeline on size-matched non-constituent peers assigned the same event calendar (Patel and Welch, 2016); and cluster-robust t-statistics on cycle-aggregated means are reported alongside. Classical BMP and Patell statistics appear in the tables for comparability with the older literature but are never the basis of a claim, because both assume the cross-sectional independence that review-date batching violates. Across the roughly 1,700 test statistics the full table set contains, we apply no family-wise correction; instead, every window-level CAAR claim promoted to the text must clear the placebo and bootstrap gates simultaneously, a stricter requirement than a false-discovery adjustment on nominal p-values, and the one headline-adjacent quantity that fails those gates — the post-announcement give-back — is reported as fragile throughout. Regression coefficients, for which neither instrument exists, are held to a different and weaker standard — review-cycle-clustered standard errors plus a specification battery, including leave-one-cycle-out — and are labeled as such wherever they appear.

For the mid-cycle cohorts the effective date is event-specific, so the cycle block degenerates to the event and their bootstrap intervals carry no clustering correction; tables expose the cycle-to-event ratio so the two dependence structures cannot be confused, and single-cycle cells carry no interval at all.

3.4 Stratification

Every statistic is computed within cohort and side — scheduled additions, scheduled deletions, and each mid-cycle cohort separately — and then stratified three further ways: by period (Early 2014–2017, Middle 2018–2021, Late 2022– — a partition motivated by the time-heterogeneity evidence of Bennett, Stulz, and Wang, 2020), which Section 2.1 already qualified as entangled with the data-era seams; by data era directly; and by the contamination scrub. The scheduled cohort on the SXXR benchmark with full-sample scrub is the headline cell everywhere, with the ex-contaminated companion reported in parallel.

3.5 Arbitrage risk

Following Wurgler and Zhuravskaya (2002), a stock's arbitrage risk is the variance of the residual left after hedging it with its closest substitutes under a zero-net-investment constraint. The primary measure A_1 hedges with the index itself, $A_1 = \text{Var}(R_i - R_{SXXR})$ over $[T_{ann}-252, T_{ann}-22]$ — the constraint $\sum x = 1$ forces the unit hedge, which is Wurgler and Zhuravskaya's own specification, not a simplification — and is available for all cohorts. The preferred multi-peer variant A_2 hedges with the three nearest non-constituent peers in the same supersector, matched by Mahalanobis distance on size, book-to-market, and momentum, with the hedge weights solving the constrained minimum-variance problem in closed form; it exists for 740 of 827 scheduled events. The two measures correlate at 0.98, so they are never entered in the same regression: A_1 is the primary regressor and A_2 a scheduled-only robustness column.

---

4. The anatomy of the effect (Q1, Q2, Q4)

4.1 The effect lives before the announcement

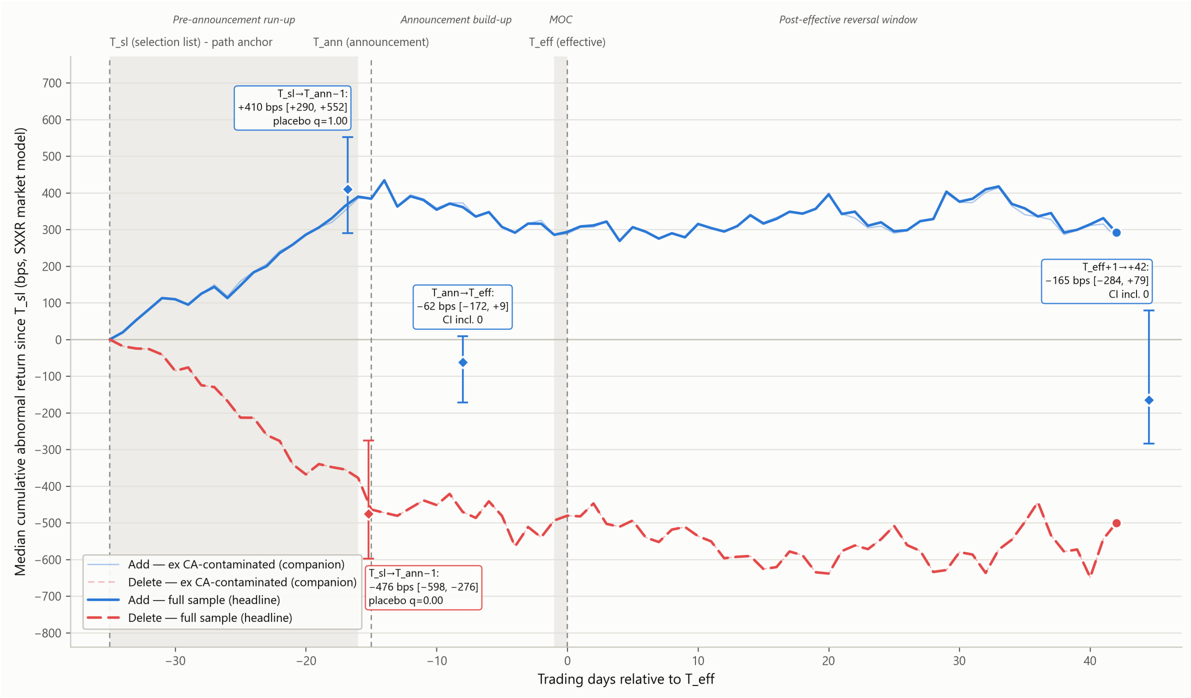

Figure 1. Median cumulative abnormal return (SXXR market model, basis points) around scheduled STOXX Europe 600 review events, 2014–2025, in trading days relative to the effective date; each event's path is anchored at zero on its selection-list date (T_{sl}). Shaded bands mark the event windows; bracketed values are the tested window-level median CAARs with review-cycle block-bootstrap 95% CIs and peer-placebo quantiles from the published CAAR tables. The curve is a median trajectory: the median of cumulative paths does not equal the median of per-event window sums, so effect sizes are read from the bracketed window statistics, never from differences of the plotted line (the line shows +384/−464 between T_{sl} and T_{ann}). Events with a short announcement-to-effective gap retain at most two estimation-window-adjacent days at the anchor. n per day across the plotted domain: 345–369 (additions) and 357–379 (deletions) on the full-sample lines, marginally lower on the decontaminated companions; the coverage clip never binds.

Figure 1 summarizes the paper in a single picture. Anchored at zero on each event's selection-list date, the median scheduled addition climbs roughly +384 basis points by the announcement and then gives part of it back; the median scheduled deletion falls roughly −464 basis points into the announcement and does not recover over the following two months. Formally, the pre-announcement run-up window $[T_{sl}, T_{ann} - 1]$ carries a median cumulative abnormal return of **+410 basis points** for additions and **−476 basis points** for deletions (naive sign tests $p = 2.4 \times 10^{-17}$ and 1.9×10^{-14}, descriptive under review-date clustering — the certification is the clustered stack that follows). These are the only window-level results in the paper that survive every layer of the inference stack simultaneously: the review-cycle block bootstrap places the addition median between +290 and +552 basis points and the deletion median between −598 and −276, both far from zero, and against the 5,000-replication peer placebo the observed medians sit at the 1.00 and 0.00 quantiles respectively — no size-matched peer basket assigned the same event calendar produces anything comparable.

Two words of interpretive caution accompany the headline, because the honest reading is subtler than "arbitrageurs front-run the announcement." First, the peer placebo is matched on size but not on selection: stocks enter the addition sample precisely because their ranks improved into the review cut-off, so part of the measured run-up is the tail of the very drift that caused selection, and no size-matched null can remove that component. Second, the decomposition by selection-list marginality in Figure 3 shows where the run-up lives, but it cannot by itself say why: additions whose entry the published rule already predicted (the *core* group) show essentially no net effect over the Petajisto window (−11 basis points), while *buffer* additions — names whose fate was still open at the list date — carry +389 basis points with an interval excluding zero. That pattern is exactly what genuine anticipation of a resolving membership change would produce, and it is also exactly what pure selection with no index effect at all would produce — a buffer name enters the sample only by climbing through the threshold inside this very window, while core names face no such conditioning — so the split localizes the run-up without discriminating between the two readings. The design does contain one selection-resistant witness, and we weight it accordingly: deletions the rule had already sealed at the list date face the weakest conditioning available (staying in the sample requires only *not recovering*), yet they still fall −282 basis points before the announcement, below 99.8% of peer-placebo draws whose own drift is nearer −50 — directionally supported rather than certified, as the 56-event bootstrap interval is wide. The decisive within-sample test does not exist in this calendar: across all 49 cycles the announcement never trails the composition cut-off (it is the next trading day after the month-end cut-off in 33 cycles and precedes the formal cut-off date in 16), so there is no determined-but-unannounced window in which selection is structurally zero — itself a manifestation of the always-on sunshine design. A rank-path-matched placebo — peers whose ranks improved comparably without crossing the threshold — is the feasible future design, and until it is run we present the addition-side run-up as the pre-announcement repricing of impending membership change, an unresolved blend of selection and anticipation, with the deletion side supplying the selection-resistant floor. The deletion side carries the same structure in a subtler form: rule-predicted demotions and late-resolving ones show similar point declines (−282 versus −293 basis points), but only the late-resolving majority is statistically identified — the 56 demotions the list had already sealed are too few and too dispersed to pin down, so on that side marginality separates precision rather than magnitude. To keep the two schemes distinct: on the marginality split, core additions sit at −11 and buffer additions at +389; on the rule-prediction split, predicted deletions sit at −282 and late-resolving deletions at −293 — and on the addition side the two splits coincide exactly (Figure 3 shows both). Because the marginality labels are point-in-time lower bounds (Section 2.2), the true concentration in marginal names is, if anything, understated.

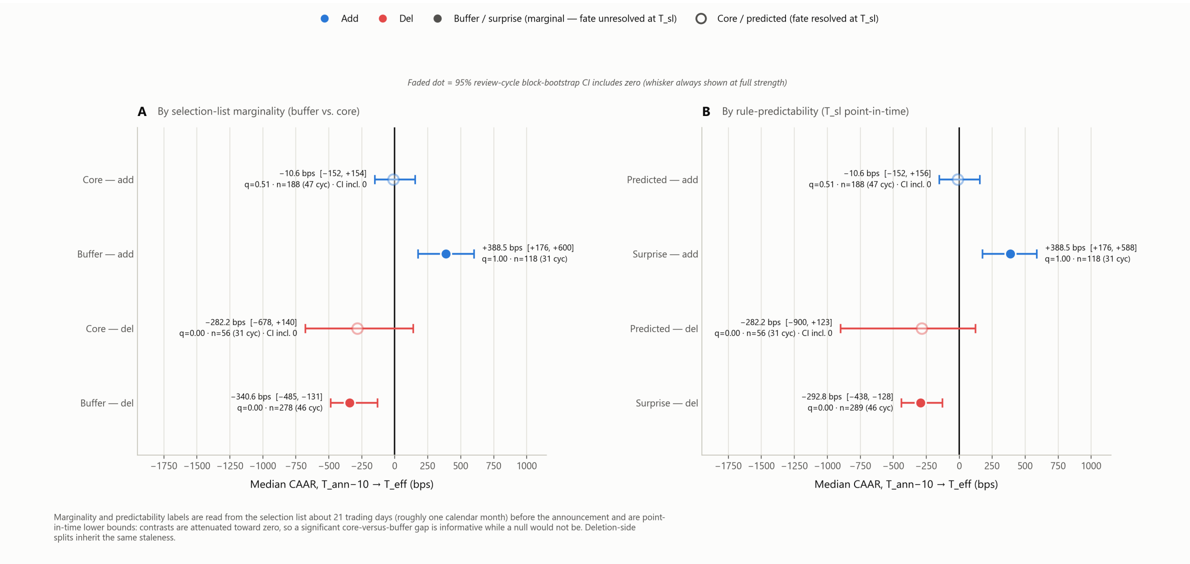

Figure 3. Median cumulative abnormal return over the net-effect window $[T_{ann} - 10, T_{eff}]$ *for scheduled STOXX Europe 600 changes, 2014–2025, SXXR market model, full sample, split by selection-list marginality (Panel A) and by whether the published buffer rule applied to the point-in-time list already predicted the change (Panel B). Whiskers are review-cycle block-bootstrap 95% intervals;* q *is the peer-placebo quantile. The effect concentrates in marginal (buffer) additions; rule-predicted additions carry no net effect. Deletion-side groups show similar point declines with only the late-resolving majority statistically identified. Labels are point-in-time lower bounds read about 21 trading days before the announcement, so the true concentration is understated.*

4.2 After the announcement: a fragile give-back

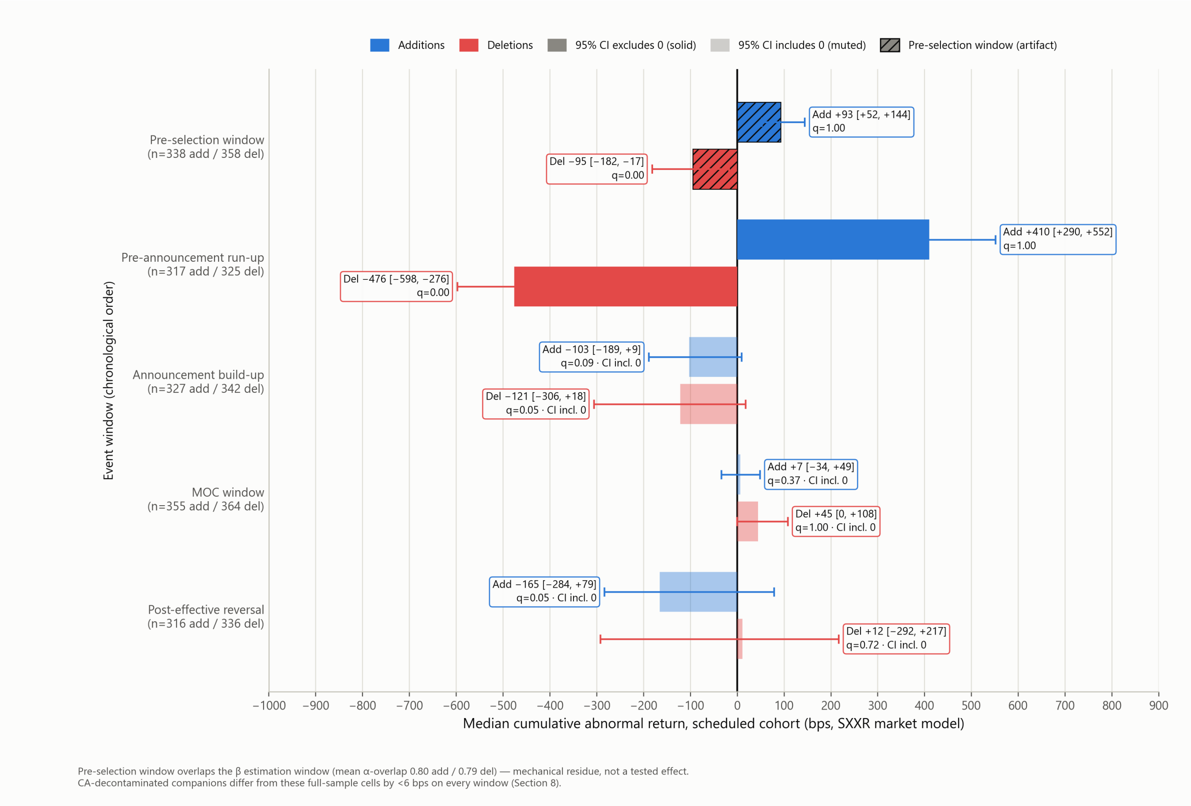

Figure 2. Median cumulative abnormal returns by event window for scheduled STOXX Europe 600 changes, 2014–2025, SXXR market model, full sample; error bars are review-cycle block-bootstrap 95% intervals and q is the peer-placebo quantile of the observed median (scheduled cohort). Only the pre-announcement run-up windows carry intervals excluding zero on both sides; the post-announcement build-up, MOC, and reversal windows are statistically fragile, and the pre-selection window largely reflects estimation-window overlap rather than a tradable effect. Decontaminated companions differ by less than six basis points on every window.

Over the announcement-to-effective window $[T_{ann}, T_{eff}]$ the median scheduled addition *loses* 62 basis points. A naive reading would call this significant — the sign test gives $p = 0.028$ and the cycle-clustered t is −2.19 ($p = 0.034$) — but it is exactly the kind of cell that motivates the paper's inference doctrine: its block-bootstrap interval, [−172, +9], includes zero, and its placebo quantile of 0.35 means size-matched peers drifted similarly over the same weeks. We therefore describe the post-announcement period as a *give-back* — the sign is consistently negative across specifications — while refusing to certify it as a distinct priced phenomenon. The deletion-side counterpart (−55 basis points) is not significant on any test. Within the composite window, the build-up segment $[T_{ann}, T_{eff} - 2]$ contributes a median −103 basis points for additions and likewise fails the placebo (quantile 0.09); we flag this deliberately, because the artifact-level name of that window could invite confusion with front-running, and the data say the opposite: by the time the change is public, the profitable part of the trade is over.

4.3 The effective day and what follows

The closing-auction window shows the classic asymmetry of the literature in attenuated European form — but by this paper's own certification standard it is a marginal effect, and we present it as one. The deletion-side median is +45 basis points with a bootstrap interval, [−0.2, +108], that touches zero: it clears the placebo gate decisively (quantile 1.00 — no peer basket bounces like this) but not the bootstrap gate, so under the Section 3.3 doctrine it is directionally supported rather than certified. The mean-based companion is stronger (+73 basis points, cycle-clustered $p = 0.029$) — prices bounce as the forced selling clears — while additions show nothing on either statistic (median +7, mean +22, $p = 0.38$). The asymmetry itself echoes Petajisto's finding that deletion-side price pressure exceeds addition-side pressure, though no commensurable US figure exists for this window: his deletion estimate is a full announcement-to-effective, selection-laden decline, whereas ours is a two-day closing rebound after the forced selling clears.

Post-effective dynamics are where the European evidence departs most clearly from the US narrative of transient pressure. The regression of reversal-window abnormal returns on MOC-window abnormal returns yields a pooled slope of $\lambda_1 = -0.72$, but its cluster-robust interval, [−1.77, +0.33], spans both zero (no reversal) and −1 (full reversal), and the pooled coefficient is dominated by the deletion side (per-side estimates: additions −0.21, deletions −1.00) ; the estimate also shares its $\widehat{\beta}$ with the MOC regressor, a quantified mechanical bias of about −0.05, and a second, unmodeled bias runs through β itself: it is estimated before the announcement but applied through the post-effective window, and inclusion raises a stock's comovement with its index (Barberis, Shleifer, and Wurgler, 2005), so post-effective abnormal returns retain some residual market exposure — second-order for the medians, but a reason beyond power to read the reversal estimates gently. We therefore make no claim of correspondence with the near-complete reversal Patel and Welch report for the S&P 500. What the data do support is the absence of any statistically identified reversal: the addition-side reversal window carries a point estimate of −165 basis points with an interval spanning both zero and a full unwind, and the deletion-side estimate is +12 basis points — essentially nothing — inside an even wider interval. Nothing here resembles the systematic post-effective unwind the US literature documents, and Figure 1's paths read the same way descriptively: neither median recovers toward its pre-event level over the following two months. On the deletion side in particular, the repricing shows no identified tendency to unwind.

4.4 Stage decomposition and the sunshine test

Summing the five windows for the 487 scheduled events that are complete across all of them (230 additions, 257 deletions — the mid-cycle cohorts fail the completeness floor because their reversal windows are truncated by delisting), mean stage contributions decompose the net move. For additions, the pre-announcement run-up contributes +524 basis points against a net total of +338 — a stage share of 1.55, with the post-announcement stages giving back the excess . Two accounting caveats travel with any share number. The denominators are means, not the medians used everywhere else (shares must add up; medians do not), so stage *levels* here run above their median counterparts. And the net total includes the small pre-selection window whose content is mostly the mechanical $\widehat{\alpha}$ residue of Section 3.1; excluding it from the denominator raises the run-up share to roughly 2.5. Either way the qualitative statement is unchanged and matches Figure 1: more than the entire net index effect accrues before the announcement.

STOXX's four-anchor timeline offers, in principle, a test of Admati–Pfleiderer sunshine trading:

does a longer publicly known lead time reduce the post-effective reversal? In practice the test is underpowered *by construction*, and we report it as such rather than as a null result. Because STOXX sets its calendar by rule, the two lead times barely vary — the selection-to-announcement lead has a standard deviation of 2.7 trading days around its 21-day median, and the announcement-to-effective lead 2.6 around 15 — so the identifying variation that index-committee discretion supplies in the US setting simply does not exist here. The estimated lead-time coefficients are individually and jointly insignificant (addition-side $\delta_1 = +0.018$, $\delta_2 = -0.045$, joint $p = 0.31$), and the honest conclusion is that Europe's rule-based calendar is *always-on sunshine*: the mechanism is plausibly one reason the measured effect is small and front-loaded, but the sample contains no variation with which to prove it.

5. Cross-section and predictability (Q3, Q5)

If the effective-day price impact were a compensated risk or a capacity constraint, it should be explicable in the cross-section and forecastable ahead of time. This section shows it is neither, and that the one structural regularity that does survive is conditional, not unconditional. The two subsections tell a single arbitrage-risk story, and we present them together precisely so that no reader mistakes them for a contradiction. One measurement caveat applies to the whole section: the cross-sectional dependent variable here, the slippage target of the predictability layer, and the reversal-ratio regressor of Section 4.3 all read the same two-day MOC-window abnormal return, so the execution-layer results are one noisy object examined from three angles, not three independent confirmations.

5.1 No unconditional cross-sectional structure at the close

For the clean two-day MOC window, we regress each scheduled event's abnormal return on the candidate determinants of Section 3.5 and the standard controls — arbitrage risk, idiosyncratic volatility, relative size, log market capitalization, pre-announcement run-up, selection-list marginality, and turnover — with regressors standardized, additions and deletions estimated separately, and inference clustered by review cycle. The pre-registered Fama–MacBeth estimator turns out to be infeasible in this setting, and we say so rather than substitute silently: with only five to ten same-side scheduled events per quarterly cycle, the per-cycle cross-sectional regression is estimable in at most one of the 49 cycles — far short of any usable minimum — so the estimator is undefined and the pooled cycle-clustered regression — its designated fallback — is the headline .

The result is a null with teeth. Arbitrage risk, the theoretically central regressor, attracts a t-statistic of −0.51 on the addition side and +1.42 on the deletion side; no regressor in the main specification clears conventional significance on the MOC window, and the Chow test cannot reject coefficient equality across sides ($p = 0.70$). Two sample caveats bound the claim. The regression runs on scheduled events only, and the deletion-side sample is further truncated to names that kept trading through the effective date, so it describes orderly demotions rather than distressed exits. Country and sector fixed effects are omitted — the former for thin cells, the latter because the industry taxonomy breaks at the 2023 seam (Section 2.1) — so the null is about characteristics, conditional on neither geography nor sector.

One coefficient elsewhere in the table deserves preemptive disarming, because it is the kind of number that invites over-reading. On the secondary announcement-to-effective window, addition-side arbitrage risk enters at $t = -3.76$. That window is the give-back of Section 4.2, and the coefficient says only that harder-to-hedge additions give back more of their run-up — raw idiosyncratic volatility explains the same cells slightly better ($t = -4.11$), as befits a mechanical retracement. It is not evidence that arbitrage risk raises the effective-day impact, and it never appears in this paper as such.

5.2 A conditional structure that does survive — and a convexity that does not

Wurgler and Zhuravskaya documented an *unconditional* arbitrage-risk effect on the event-day jump of their 1976–1989 S&P 500 sample; we do not recover that level effect on the modern two-day MOC window — coherent with this paper's central finding that the effect has moved into the pre-announcement run-up — but their mechanism equally implies an interaction, and that is what survives here: scarce arbitrage capital should steepen the *slope* of price impact in flow. Measuring the passive demand shock as the tracked assets each event must absorb relative to the stock's average daily value traded (F/ADV), and execution slippage as the signed MOC-window abnormal return (positive = costly for the tracker on both sides), the interaction term is the one structural survivor in the entire execution layer: the flow-times-arbitrage-risk cross-partial enters at $p = 0.0015$ under cycle clustering and holds across the winsorized, raw, scrubbed, and year-fixed-effect specifications. A leave-one-cycle-out sweep adds the sharpest qualifier: the coefficient's magnitude is stable (22 to 29 whichever single cycle is dropped, and excluding the 2020Q1 COVID cycle leaves $p = 0.003$), but its precision is concentrated — excluding the single 2015Q4 cycle leaves the magnitude at 26 with $p = 0.20$, while excluding any other cycle keeps $p \leq 0.016$ — so we read the interaction as magnitude-stable structural evidence whose statistical certification leans on one joint-tail-rich cycle, not as a certified effect. Expressed in interpretable units, one standard deviation of arbitrage risk raises slippage by about 2.9 percentage points per day-of-ADV of passive demand (0.029 in return terms) . The magnitude lives in the joint tail: it is expressed per day of average volume — itself an upper bound on true days-of-liquidity, since ADV counts the primary listing's lit volume only in a fragmented market — and the median event's demand shock is only about a third of such a day, so the typical high-arbitrage-risk differential is nearer one percentage point, reaching multiple-percentage-point territory only where large flow meets a hard-to-hedge name. We report the effect only in per-standard-deviation form; the raw coefficient is a cross-partial in awkward natural units, and the accompanying level terms are centering artifacts with no standalone interpretation. Read jointly with Section 5.1: there is no unconditional arbitrage-risk premium at the close, but when large flow meets a hard-to-hedge name, the flow bites harder. The cross-section never estimated that cross-partial, so the two findings are complementary, not contradictory.

The classical convexity prediction fares worse. The quadratic flow term is at best suggestive in a pooled regression ($p = 0.11$) and collapses entirely once year fixed effects absorb the secular growth of passive assets ($p = 0.91$ thereafter), and a non-parametric LOWESS fit of slippage in flow is not even monotone. What looked like convexity was the AUM time trend wearing a quadratic costume. Deletion-side estimates are reported but carry no weight in either direction: the deletion MOC window is contaminated by the incipient reversal of Section 4.3, and a deleted name's collapsing ADV is itself an outcome of the demotion, so both sides of the deletion ratio are endogenous. Figure 6 shows the addition-side scatter, the non-monotone fit, and the interaction visually.

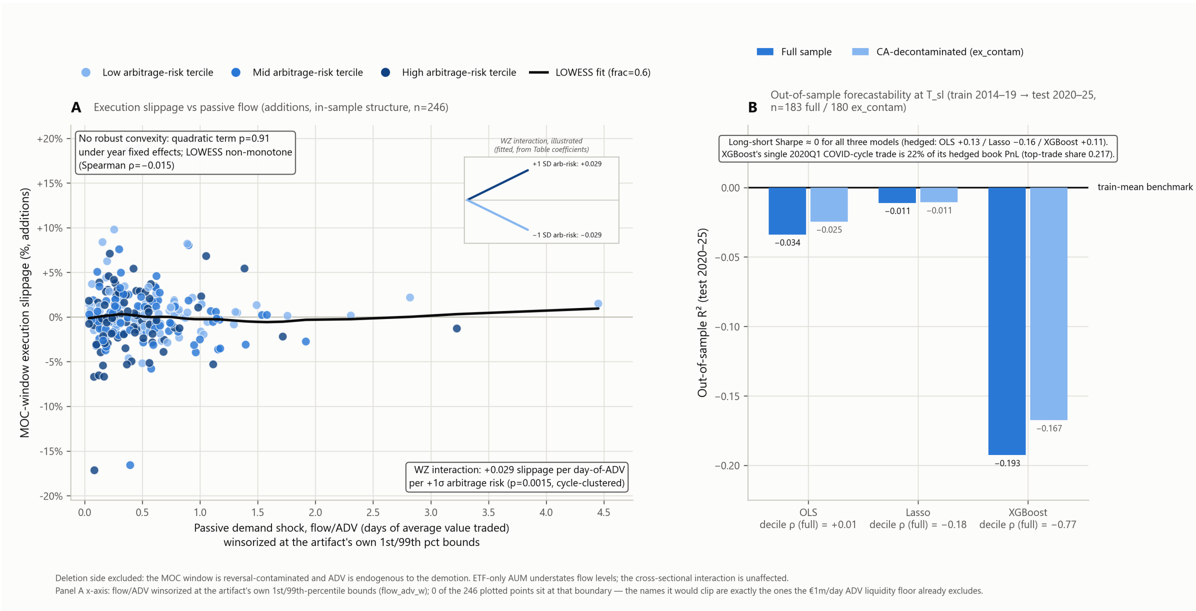

Figure 6. Panel A: MOC-window execution slippage against the passive demand shock in days of average value traded, scheduled additions, 2014–2025, with a LOWESS fit and points shaded by arbitrage-risk tercile; the inset illustrates the fitted interaction as the model-implied slope differential at ±1σ arbitrage risk — drawn from the estimated coefficient, because a raw subgroup fit cannot recover a controlled cross-partial. The quadratic flow term is not robust ($p = 0.91$ under year fixed effects) while the flow-times-arbitrage-risk interaction is (+0.029 slippage per day-of-ADV per one standard deviation of arbitrage risk, $p = 0.0015$, review-cycle-clustered). Panel B: out-of-sample R^2 of point-in-time slippage forecasts trained on 2014–2019 and tested on 2020–2025, against the training-mean benchmark; every model is negative under both the full and decontaminated samples, decile rank correlations are zero or negative, and the long-short implementation earns approximately nothing once a single COVID-cycle trade (22% of the gradient-boosted model's hedged profit) is recognized as a tail event.

5.3 Out of sample, there is nothing to harvest

The predictability layer asks the practitioner's question directly: standing at the selection-list date with strictly point-in-time features — every input rebuilt at a T_{sl} cutoff, scalers and winsorization fit on training data only, hyperparameters chosen on a validation slice inside the training years — can one forecast which events will be expensive? Training on 2014–2019 and testing on 2020–2025, the answer is no, uniformly. Out-of-sample R^2 against the training-mean benchmark is negative for every model on the addition side (OLS −0.03, Lasso −0.01, gradient boosting −0.19), decile rank correlations between predicted and realized slippage are zero or negative, and a long-short implementation — short the predicted-costly additions, long the predicted-costly deletions, per-cycle, equal-weighted, hedged — earns a Sharpe ratio indistinguishable from zero once a single COVID-cycle deletion trade — 22% of the gradient-boosted model's hedged profit — is recognized as the tail event it is . The in-sample fit (addition-side R^2 of +0.07) against the out-of-sample failure is the textbook signature of a memorized rather than generalizing relationship. Ex-contamination results are indistinguishable from the full sample. The null is also consistent with clean point-in-time discipline: any residual look-ahead leakage in our feature construction could only have inflated out-of-sample fit, so a negative R^2 is what an honest feature set should produce when there is nothing to find. The absence of out-of-sample forecastability at the one horizon where a forecast would be actionable is consistent with the McLean–Pontiff pattern of post-discovery decay, though our single train–test split identifies the absence, not the decay path itself.

5.4 What this means for practitioners

The three results compose into one practical sentence: the economically meaningful object in the STOXX rebalancing cycle is the *list*, not the *trade*. The membership changes themselves are substantially anticipatable from a published rule applied to a published ranking — that is where the +410/−476 of Section 4.1 lives — while the residual execution cost at the effective date is small, cross-sectionally unstructured, unforecastable out of sample, and materially conditional only on the interaction of flow with hedging difficulty. An execution desk should therefore spend its effort on flow-aware scheduling around marginal names rather than on stock-picking models of expected slippage; Section 6 prices what remains.

6. What passive investors actually pay (Q6)

The rebalancing premium is not merely an anomaly; for anyone tracking the index it is a hidden cost. A tracker mechanically buys additions after they have run up and sells deletions after they have fallen, so the premium of Section 4 is paid by the fund and captured by whoever positioned earlier. Petajisto's accounting identity turns the premium into an annual performance drag: with p the average absolute premium a fund trades into, a and d the one-sided annual addition and deletion turnover, and s the share of the tracked universe with no index-neutral substitute, the annual cost lies between a no-reversal bound $p\,[s(a-d)+d]/(1+p)$ and a full-reversal bound $p\,(a+d)/(1+p)$.

Our inputs are deliberately matched to the objects a fund actually trades. The premium p is the event-count-weighted average absolute median cumulative abnormal return over the $[T_{ann} - 10, T_{eff}]$ window — the Petajisto convention, and the only window whose sign matches the direction of the forced trade — computed on scheduled, decontaminated events against SXXR; it averages 255 basis points across 2014–2025. We stress that this p is a blend of both sides by construction and is not comparable to any single addition or deletion CAAR quoted elsewhere in the paper. Turnover comes from STOXX's own review files restricted to scheduled review dates so that the turnover universe matches the premium universe: additions average 1.20% and deletions 0.84% of index weight per year, one-sided — roughly a 2% annual one-sided rebalancing churn, far below S&P 500 levels.

The resulting cost estimate for a STOXX Europe 600 tracker over 2014–2025 is **between 2.8 and 5.1 basis points per year**. We quote the range rather than a point deliberately. The upper figure is the full-reversal bound, which has the practical virtue of not depending on the uncalibrated substitution share s; the lower figure fixes s at Petajisto's US value of 0.70 for want of a European calibration. Our own reversal evidence (Section 4.3) cannot reject either polar assumption — the post-effective windows carry confidence intervals spanning both no-reversal and full-reversal — so the honest deliverable is the bounded interval, with the s-free upper bound as the quotable single number. One tension deserves stating rather than hiding: our own Section 4.3 finds no identified reversal, and a literal no-reversal reading pushes toward the lower bound. But a premium that never reverses is not thereby free — if it reflects a permanent demand-curve shift, the tracker still pays it on the way in — and distinguishing permanent repricing from transient pressure is precisely what the reversal windows cannot do here, which is why the range, not either endpoint, is the deliverable. In euro terms, applied year by year to the tracked assets we can identify, the full-reversal bound averages €6.5 million per year across the period (the per-year product, not 5.1 basis points times the €12.1 billion mean base — cost and assets co-move, and the distinction is worth a footnote in any client deck). The same annual-mean asset series reappears as the demand denominator of Section 7, so the euro cost here and the multiplier decline there are two readings of one series, not independent corroboration. Because the asset base covers only the nine physical ETFs of Section 2.3, the euro figure is a floor on the aggregate cost borne by all index-tracking capital; the basis-point figure is the transferable number, valid for any SXXP-tracking book regardless of size.

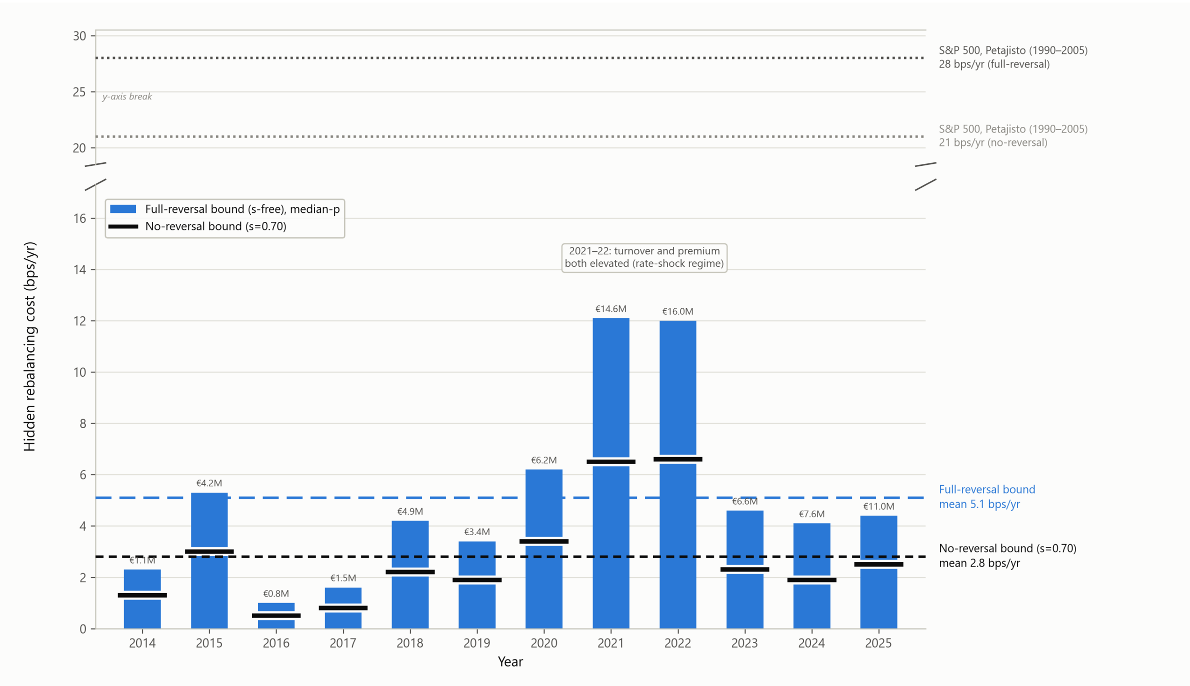

Figure 4. Annual hidden rebalancing cost for a STOXX Europe 600 tracker under the Petajisto (2011) accounting, 2014–2025. Bars show the s-free full-reversal bound with the median-premium input; markers show the no-reversal bound at Petajisto's US substitution share $s = 0.70$ *(uncalibrated for Europe), so each year reads as a bounded range. Euro labels apply the year's cost to that year's identified passive-tracking ETF assets and are floors on the aggregate cost across all index-tracking capital; the basis-point figures are asset-invariant. The premium input is the event-count-weighted absolute median CAAR over* $[T_{ann} - 10, T_{eff}]$ *on scheduled, decontaminated events; turnover is restricted to scheduled review dates so premium and turnover describe the same universe. US reference lines are Petajisto's 1990–2005 S&P 500 estimates; part of the US–Europe gap reflects that earlier, higher-premium era rather than geography.*

Set against the US benchmark, the European figure is strikingly small. Petajisto's S&P 500 estimates for 1990–2005 are 21 to 28 basis points per year; bound-matched, the STOXX cost is roughly 4.4–5.5 times lower on the full-reversal comparison and about 7.5 times lower on the no-reversal comparison. Two forces produce the gap, and only one of them is geography. The STOXX 600's buffered, rule-based review generates far less turnover than the S&P 500's committee process did, and its premium per unit of turnover is smaller. But part of the gap is era: the US numbers date from the high-premium 1990s, and Section 7 shows the premium itself has been shrinking. A fund selecting European large-cap exposure today pays single-digit basis points for the privilege of transparent, rule-based index construction — a number small enough that, for most mandates, it is dominated by fee differences, yet large enough to reward execution desks that internalize Section 5's flow-conditionality on the handful of marginal names each quarter.

7. Is the European index effect disappearing? (Q7)

The US literature's most consequential recent finding is that the S&P 500 index effect has all but vanished. Greenwood and Sammon organize that evidence around a price multiplier M — the abnormal return per unit of passive demand shock — and document an order-of-magnitude collapse across decades. This section runs the European edition, with two structural adaptations and, thanks to the uncertainty machinery of Section 3.3, error bars that the descriptive US-style comparison lacks.

The multiplier is $M = \overline{CAAR}/\bar{D}$, where the numerator is the median net-effect CAAR over the Petajisto window and the denominator is the passive demand shock as a fraction of float. For a float-weighted index the demand shock collapses to the passive ownership ratio — tracked assets over total index float capitalization — identically for every event in a period: because weights are proportional to float, the flow a tracker must trade in any name is the same fraction of that name's float, about 0.10% of float in the Early period rising to 0.19% in the Late period . This algebraic fact has a useful corollary: the Greenwood–Sammon *composition-via-demand* channel — smaller multipliers because marginal events carry smaller per-name demand — is structurally zero here, just as their migration channel is (the STOXX 600 has no co-tracked disjoint neighbor index to net flows against). Any composition effect must run through multiplier heterogeneity across event classes, which is exactly what the buffer split delivers. Because our asset base is ETF-only (Section 2.3), the denominator is understated and every multiplier level in this section is an upper bound; comparisons with US levels are meaningless, and we confine all statements to *trends and contrasts within our own panel*.

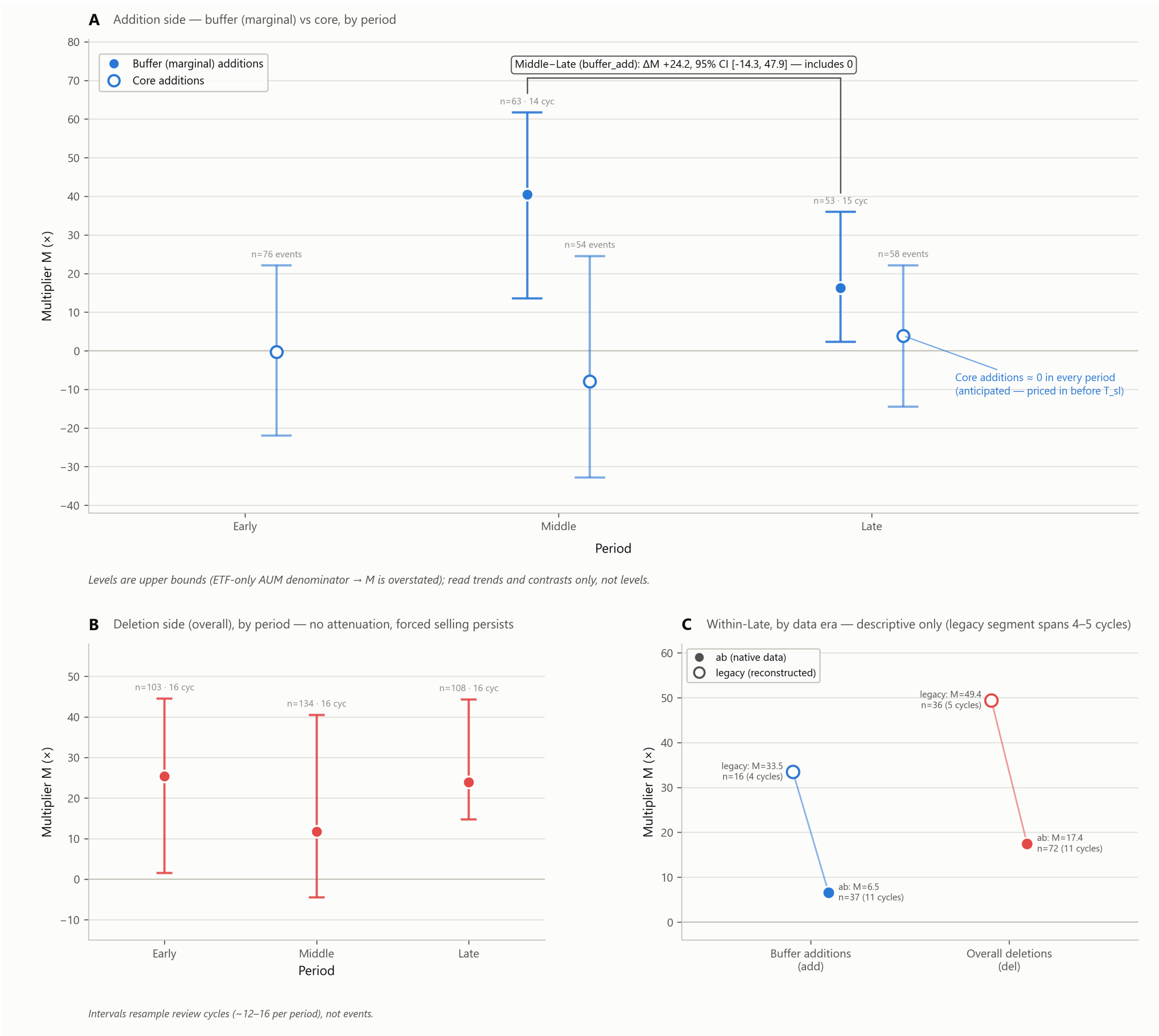

Figure 5. Greenwood–Sammon price multipliers $M = \overline{CAAR}/\bar{D}$ for scheduled STOXX Europe 600 events by period (Early 2014–2017, Middle 2018–2021, Late 2022–), net-effect window $[T_{ann} - 10, T_{eff}]$, median CAAR, full sample. Whiskers are review-cycle block-bootstrap 95% intervals mapped through the fixed period demand ratio; the demand denominator is ETF-only, so all levels are upper bounds and only trends and within-panel contrasts are interpretable. Panel A: the marginal (buffer) addition multiplier is well separated from zero in both periods, but the Middle-to-Late decline itself carries a confidence interval that includes zero — attenuation is directional, not confirmed. Panel B: deletion multipliers do not attenuate. Panel C: within the Late period, multipliers are lower in the native-data segment on both sides; the legacy segment spans only four to five review cycles, so this split is descriptive, and genuine attenuation, the 2022 rate-shock regime, and the reconstruction-to-native measurement change are observationally equivalent at this boundary.

The point estimates trace an attenuation pattern concentrated exactly where Section 4.1 located the effect. For buffer additions — the marginal names that carry the run-up — the multiplier falls from 40.4 in the Middle period to 16.3 in the Late period, while core additions sit indistinguishably near zero throughout; overall deletion multipliers, by contrast, are as high at the end of the sample as at the beginning (25.4 Early, 23.9 Late). The new confidence machinery then disciplines the narrative in both directions. Each period's buffer-addition multiplier is individually well separated from zero (Middle [13.6, 61.7], Late [2.3, 36.0]): the marginal-addition effect is real in both halves of the sample, so nothing here has "disappeared." But the *decline itself* — a Middle-minus-Late difference of +24.2 — carries a bootstrap interval of [−14.3, +47.9] that includes zero. Twelve years and 49 cycles are simply not enough data to certify a trend in a heavy-tailed ratio; we note that the US result needed four decades. Both statements also inherit one scope limit: the intervals propagate CAAR sampling error only, with the demand denominator held at its measured value, so denominator measurement error would widen them further — reinforcing, not weakening, the non-rejection. Mechanically, decomposing the multiplier's log change shows about 59% of the decline comes from a smaller CAAR numerator and 41% from the growing passive denominator — the premium per event shrank *and* the assets absorbing it grew, and only the first component is "the effect getting weaker" in the economic sense . For the same reason, this section and Section 6 are not independent corroboration of anything: the same asset series that raises euro costs there lowers multipliers here.

An identification caveat specific to our data must be stated plainly. The Middle period is measured entirely on the legacy archive and the Late period roughly two-thirds on the native feed (Section 2.1), so the Middle-to-Late contrast straddles a data seam. Splitting the Late period at the seam is only descriptively possible — the legacy segment spans four to five review cycles depending on the cell, too few for trustworthy resampling — and what it shows is that multipliers are markedly lower in the native-feed segment on *both* sides of the book (overall deletions: 49.4 in Late-legacy versus 17.5 in Late-native). Three readings of that step — genuine post-2022 attenuation, the 2022 rate-shock regime inflating the legacy-side segment, and a residual measurement difference between reconstruction and native data — are observationally equivalent at this seam, and we present all three rather than adjudicate on insufficient evidence.

The composite answer to this section's title question is therefore *attenuating, not disappearing — and only on one side*. The addition-side premium per unit of passive demand is directionally lower late in the sample, concentrated in marginal names, but the decline is not statistically confirmed and is partly denominator growth. The deletion side shows no attenuation at all: forced selling still reprices demoted names, with no identified reversal (Section 4.3). The European market appears to have competed away the *predictable, buy-side* component first — consistent with a liquidity-provision industry that grew into the anticipatable part of the calendar — while the uncomfortable, inventory-consuming business of absorbing deleted names remains priced. At most, the addition side shows the directional *beginning* of the attenuation Greenwood and Sammon documented for the US — unconfirmed here, and absent on the deletion side — and Europe started from a far smaller premium in the first place (the bound-matched cost comparison of Section 6), not from anything resembling the US multiplier levels our data cannot speak to.

8. Robustness

The inference stack of Section 3.3 already embeds the two most important robustness layers — the peer placebo and the review-cycle bootstrap — into every table. This section reports the remaining checks and, just as importantly, is explicit about which claims each check covers.

<table>
<tr><th>Claim</th><th>Sign / cluster</th><th>Placebo</th><th>Bootstrap CI</th><th>CA scrub</th><th>FF3</th><th>β-window</th></tr>
<tr><td>Run-up, additions (+410)</td><td>✓</td><td>✓ (q=1.00)</td><td>✓ (excl 0)</td><td>✓</td><td>◐indirect</td><td>—</td></tr>
<tr><td>Run-up, deletions (−476)</td><td>✓</td><td>✓ (q=0.00)</td><td>✓ (excl 0)</td><td>✓</td><td>◐indirect</td><td>—</td></tr>
<tr><td>Give-back, additions (−62)</td><td>✓ naive</td><td>✗ (q=0.35)</td><td>✗ (incl 0)</td><td>✓ stable</td><td>✓</td><td>✓</td></tr>
<tr><td>MOC, deletions (median +45; mean +73)</td><td>✓ mean</td><td>✓ (q$≈1.0)|7(touches0)|3|—|—||Reversal, additions(−165)|3naive|7(q = 0.047→$)</td><td>✗ (incl 0)</td><td>✓ stable</td><td>—</td><td>—</td></tr>
</table>

✓ = passes; ✗ = fails (the claim is reported as fragile); ◐= partial. "Stable" in the scrub column means the value moves immaterially, not that a fragile claim becomes robust.

Factor-model adjustment. Because no daily European three-factor series exists at our sample's resolution, we built one — market (SXXR net return), size, and value factors from canonical June-formation double sorts on the same candidate-pool universe that defines our peer baskets, with value-weighted simple returns, total-market-equity value sorts, and negative-book-equity names excluded. Re-running the event study with three-factor expected returns leaves the effect intact: the net-effect window shows additions at +198 basis points and deletions at −258 (sign tests $p \approx 0.001$), against +154 and −293 under the single-factor model on the identical events and identical window dates. The deletion effect attenuates and the addition effect strengthens under the factor model — the signature of size and value loadings absorbing part of the raw deletion move — though we have not decomposed how much each factor contributes. A second limit is structural: the factor portfolios are formed on the same candidate-pool universe that contains the event stocks themselves, so on the deletion side genuine size and value loadings cannot be separated from mechanical absorption of the event names' own returns into the factors. One phrasing subtlety belongs in print: the factor series has no single-name gaps, so factor-adjusted CARs average about 1.5 more daily observations per event inside the wide windows; cell membership and window dates are identical, per-event daily counts are not, and the difference is immaterial to medians.

Corporate-action scrub. Comparing full and decontaminated samples cell by cell, on medians *and* on means (a median-only check would be blind to exactly the tail events the scrub targets), the scheduled headline cells move by about one percent of their value or less . Of 210 comparable cells, 63 are scrub-sensitive — overwhelmingly small-n sub-strata and mid-cycle cohorts where a single spin-off or distressed name dominates — and four cells consist entirely of contaminated events and are scoped out of any claim.

Benchmark swap. Re-running the headline window cells against the price-return pair leaves every sign in place — a 24-cell parallel check across the Q1/Q2 windows records zero flips — and the peer-placebo quantiles concord in sign on every headline window under either benchmark.

Estimation-window sensitivity. Re-estimating every β on 90- and 150-day windows and rebuilding the headline CAAR on the common event set moves the paired per-event difference by less than half a basis point, with bootstrap intervals on the paired difference containing zero in all cells ; the canonical 250-day variant and the propagation of alternative windows into the decomposition and cost tables were not run, a gap we disclose rather than paper over.

Coverage honesty. The matrix's first row shows a deliberate asymmetry: the pre-announcement run-up — the paper's central claim — is certified by the placebo, bootstrap, and scrub layers, while the factor-model and β-window checks were executed on the announcement-anchored windows and touch it only indirectly (the factor-adjusted net-effect window, which contains most of the run-up, is consistent in sign and magnitude but not cluster-significant on its own). Given the run-up's size and its pre-announcement timing — β misspecification and factor exposure accrue over weeks, not the days that separate these windows — we judge the residual risk second-order, and flag it as the natural first target for any referee replication.

9. Conclusion

In the STOXX Europe 600, the index effect has not disappeared; it has relocated. The economically large and statistically unambiguous price action — a median +410 basis points for additions and −476 for deletions — occurs between the publication of the selection list and the announcement, concentrates in the marginal candidates whose inclusion the published rule had not yet settled, and leaves the post-announcement landscape nearly empty: a give-back that fails placebo and bootstrap scrutiny, closing pressure that is modest and deletion-sided, no identified post-effective reversal, no cross-sectional structure at the close, and no out-of-sample forecastability of execution cost. What remains for the funds that must trade the changes is a hidden cost of 2.8 to 5.1 basis points per year — several times below the historical US benchmark — of which the addition-side component per unit of passive demand appears to be shrinking, while the deletion-side component persists undiminished.

Each of our three audiences reads this differently. For an index-tracking desk, the effect's move upstream means benchmark-relative risk lives in the weeks before the announcement, not at the effective date; the residual cost is small, unforecastable name by name, and worth attacking only through flow-aware scheduling of the handful of marginal, hard-to-hedge names each quarter. For an index provider, the European evidence is a quiet vindication of rule-based transparency: an always-on sunshine regime is associated with a small, front-loaded, and apparently shrinking transfer from trackers to arbitrageurs — several times below what committee-era US indices imposed. For a would-be arbitrageur, the message is sobering: the predictable part of the calendar is already priced by the time it is announced, the residual is noise at the horizons that matter, and the surviving edge — anticipating *rank resolution* among marginal candidates before the list crystallizes — is exactly the part this study shows to be competitive already.

The limitations are the flip side of the paper's honesty and bound what the numbers can carry. Our passive-asset base captures physically replicating ETFs only, so euro cost figures are floors and multiplier levels are ceilings. Book-equity inputs are approximately rather than strictly point-in-time. The price panel stitches three data eras whose seams we disclose but cannot fully neutralize, and the sharpest temporal contrast in the paper — the Late-period multiplier step — sits on one of them, where genuine attenuation, the 2022 regime, and measurement change are observationally equivalent. Selection-list marginality labels are point-in-time lower bounds on predictability. And twelve years of quarterly reviews yield forty-nine independent clusters — enough to certify the run-up beyond reasonable doubt, and not enough to certify its decay.

The natural extensions map onto those limits: a calibrated European substitution share to tighten the cost interval's lower bound, intraday closing-auction data to decompose the MOC window that daily bars can only bracket, a passive-ownership panel beyond ETFs to convert multiplier trends into levels, and, in time, the additional cycles that will tell whether the European index effect is following its American ancestor into irrelevance — or has already settled at the small, rule-disciplined equilibrium its architecture was designed to produce.

References

Admati, A. R., and P. Pfleiderer (1991). "Sunshine Trading and Financial Market Equilibrium." *Review of Financial Studies* 4(3), 443–481.

Barberis, N., A. Shleifer, and J. Wurgler (2005). "Comovement." *Journal of Financial Economics* 75(2), 283–317.

Bennett, B., R. M. Stulz, and Z. Wang (2020). "Does Joining the S&P 500 Index Hurt Firms?" ECGI Finance Working Paper 690/2020.

Boehmer, E., J. Musumeci, and A. B. Poulsen (1991). "Event-Study Methodology under Conditions of Event-Induced Variance." *Journal of Financial Economics* 30(2), 253–272.

Greenwood, R., and M. Sammon (2023). "The Disappearing Index Effect." Harvard Business School Working Paper 23-025, revised November 2023.

McLean, R. D., and J. Pontiff (2016). "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* 71(1), 5–32.

Patel, N., and I. Welch (2016). "Extended Stock Returns in Response to S&P 500 Index Changes." Working paper, UCLA Anderson (SSRN 2638660).

Patell, J. M. (1976). "Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests." *Journal of Accounting Research* 14(2), 246–276.

Petajisto, A. (2011). "The Index Premium and Its Hidden Cost for Index Funds." *Journal of Empirical Finance* 18(2), 271–288.

Shleifer, A. (1986). "Do Demand Curves for Stocks Slope Down?" *Journal of Finance* 41(3), 579–590.

STOXX Ltd. *STOXX Index Methodology Guide* (portfolio-based indices), §7 (selection lists, buffer rules, and review timeline), edition current to the sample period.

Wurgler, J., and E. Zhuravskaya (2002). "Does Arbitrage Flatten Demand Curves for Stocks?" *Journal of Business* 75(4), 583–608.