

A Measured Ladder of Pre-Announcement and In-Process Deal Probabilities, Reconstructed Entirely from Free EDGAR Filings

Maksim Aleksandrovich Andreev, Joseph Howden
Quark Quantitative Research

Abstract

We estimate a ladder of conditional probabilities along the path a company travels from being a plausible acquisition candidate to completing a transaction. The estimates draw only on sources the Securities and Exchange Commission publishes at no charge: the EDGAR full-text search index and the per-filer submissions record. A company that announces a review of strategic alternatives enters a definitive process within eighteen months at 31.71% ($n = 6,273$ episodes) against 3.44% on 2,500 seeded controls, a lift of 9.22 with bootstrap 95% interval [7.60, 11.65] and $z = 28.12$. The wording of that announcement carries no measurable information at the power available: across three occupied language classes the conversion rate runs 36.38%, 34.44% and 32.62%, a top-to-bottom gap of 3.76 points at $z = 1.57$ against a frozen bar of 3.0 and 5 points. A company that discloses in an 8-K that it has *received* an acquisition proposal converts within twenty-four months at 65.85% (81 of 123), Wilson 95% interval [57.11%, 73.64%], against 6.46% on 1,347 seeded control windows and $z = 8.14$ against the review rung. Once a transaction is public, a form-typed five-stage chain estimated on 2,970 decided episodes separates completion probabilities from 50.3% at the first deal-family filing to 77.6% at a definitive merger proxy, with time-in-stage quartiles attached. Among episodes reaching a definitive proxy or a tender offer, those still unresolved past the closers' 75th percentile of 113 days complete at 34.8% against 68.6% for reachers overall ($z = 16.35$, $n = 847$). Two filing-sequence features measured in a deal's first sixty days predict failure in their preregistered direction, and a third, the volume of additional soliciting material, runs strongly in the opposite direction to the one frozen before the data were seen. Every completion figure in this paper is a lower bound, because completion is detected through a delisting proxy that is tight for cash transactions and loose for stock-for-stock combinations whose targets' holders end up in a surviving listed entity. Every result is an in-sample measurement and none has been validated forward. The programme's own forward ledger records 136 resolutions against the thirty its decision rule requires, a count recovered largely by backfill from snapshot history, and the base rates that ledger supports remain unadopted pending settlement of the registered hypothesis on merger-arbitrage spreads.

1. Introduction

A merger-arbitrage decision decomposes into a small number of conditional probabilities. Before anything is announced, an analyst wants to know how much a public signal of corporate-control activity raises the chance that a definitive process follows. After a transaction is announced, the same analyst wants the probability that it completes, conditioned on how far along the process has travelled and on how long it has been sitting there. The quantities are elementary and the difficulty is entirely one of measurement, since each requires a population of episodes, a typed outcome, and a control group defined identically to the treated group.

This paper reports such measurements for three rungs of the resulting ladder, read against the hand-set screen band beneath them. The distinguishing feature of the exercise is its data. Each population is assembled from the EDGAR full-text search index, each outcome is read from the per-filer submissions endpoint, and each control is seeded from the Commission's own registrant file. No commercial transactions database is used at any point, and none was consulted in constructing any figure reported here. A curated transactions database would supply consideration per share as a typed field and would populate the pricing questions this programme has not been able to reach. It would not have supplied the

episode populations, which are derived from filings, and joining a purchased database to those episodes introduces a keying problem of its own.

Contribution. Three measured rungs of a corporate-control probability ladder, each with a preregistered bar, an interval, and a stated outcome proxy, computed entirely from free Commission endpoints. A stage-typed in-process chain with time-in-stage distributions and a confirmed overdue rider. Two preregistered failure predictors measured on 2,625 live-at-day-sixty episodes, together with a large effect that runs opposite the direction frozen before the data were seen and is reported as such rather than adopted. A settled null establishing that announcement wording does not condition conversion at the available power, whose instrument work independently confirmed that the review-rung result survives population cleaning.

Three limitations belong in the reader's hands before any number, and they are restated wherever the numbers appear rather than collected in a concluding section. First, completion throughout is detected by a delisting proxy, so every completion rate is a lower bound. That proxy is tight for cash transactions, whose targets reliably stop filing, and loose for stock-for-stock combinations, whose targets' holders receive shares in a surviving listed entity that the proxy scores as non-completion. Second, every result here is an in-sample measurement on historical filings and none has been validated forward. The forward ledger that would furnish out-of-sample confirmation records 136 resolutions against the thirty its rule requires, recovered largely by the backfill described in §10, and the rates it supports remain unadopted pending settlement of the registered hypothesis on merger-arbitrage spreads. Third, survivorship and population composition restrict what a measurement licenses rather than preventing it, and the distinction between a citable contrast and an uncitable level is drawn explicitly in §10.

2. Sources and the Free-Data Constraint

Two Commission endpoints carry the entire empirical content of this paper. The full-text search index returns filings matching exact phrases, with coverage beginning in 2001; the submissions endpoint returns,

per filer, the complete typed form and date history, paginated for filers with long records. Both are public, unmetered in the sense that matters here, and stable enough to support a resumable harvest at a courteous request rate. The harvests behind these units run at a 1.2-second throttle and cache to disk so that a rerun resumes rather than repeats.

The design decision that makes free data sufficient is the restriction of every outcome to typed form codes. Prose is used to define populations, since a strategic-review announcement or a disclosed proposal is a linguistic event, but no outcome anywhere in this paper is read from prose. A definitive process means the appearance in a filer's own submissions record of a form drawn from a frozen family, and a completion means the appearance of a delisting form. Form codes are unambiguous, machine-typed by the Commission, and immune to the extraction problems that have blocked this programme's attempts to recover per-share consideration from filing text.

Deal-family forms (frozen). 425, PREM14A, DEFM14A, SC TO-T, SC 13E3, SC 14D9, with S-4 additionally included in the strategic-review unit's outcome definition. Episodes are separated by gaps exceeding 365 days between family filings.

Completion proxy (frozen). Form 25 or Form 15-12B within 548 days of the last family filing of the episode. Episodes younger than 548 days are censored and excluded from probability estimates rather than counted as failures.

Control groups are drawn from the Commission's public registrant file, which lists company names and tickers against their filer identifiers. Each control receives a pseudo-episode date sampled from the empirical distribution of event dates and is then scored by the identical outcome detector. One caution attaches to that construction, and it was named by a sibling unit in this programme rather than by any of those reported here. A control pool seeded from all registrants contains funds, shells and inactive issuers; in the insider-precursor unit, only 31% of control windows showed any ownership-filing activity, against 79% of case windows. The direction that composition imparts to the contrasts reported below

has not been measured. A deal-family filing is a rarer event for a fund or a shell than for an operating company, which suggests the control base rates here are, if anything, lower than a matched pool would produce, and the corresponding lifts higher.

3. The Delisting Proxy and What a Completion Rate Means

Nothing in a filer’s submissions record states that a merger closed. What the record does contain is the sequence of registration-withdrawal forms that follow a completed acquisition of a listed target, and the programme reads completion from those. A target acquired for cash stops being a reporting company; a Form 25 or Form 15-12B appears; the proxy fires. A target acquired for stock in a combination that leaves a listed successor may never produce such a form, and the proxy records the episode as non-completion after its 548-day window elapses.

The consequence is arithmetically simple and pervasive. Every completion probability in §§8–9, together with the delisting-confirmed completion rate quoted in §7, is a probability of close-with-delisting, which bounds true completion from below. The conversion rates of §§5–7 are unaffected, since they measure entry into a definitive process and are read from deal-family filings rather than from delistings.

The bound is tight where consideration is cash and loose where it is stock. Because the same proxy is applied to every arm of every contrast, differences between arms are far better behaved than levels, and the reader who wants one number from this paper should take a difference. The failure-prediction unit of §9 makes the point explicit in its own settlement record, which states that its base rate of 56.19% pools genuine breaks with non-delisting closes, and that the finding is carried by the contrasts.

Binding caveat. Read every completion figure in this paper as $P(\text{close-with-delisting} \mid \text{conditioning event})$. It is a lower bound on true completion, tight for cash transactions and loose for stock-for-stock survivors. This sentence is repeated at each rung deliberately.

4. The Ladder, and Which Rungs Are Certified

The ladder as the register records it runs from a screen-level band through two pre-announcement rungs to the in-process chain. The lowest rung is a hand-set heuristic: the desk’s candidate screen attaches a band of 11 to 18% to names it surfaces on fundamental and filing grounds, and no registered unit supports that band. It is labelled as hand-set on the surface that displays it, and it is included here only so that the measured rungs above it are read against the right baseline.

RUNG	CONDITIONING EVENT	MEASURED P	INTERVAL	STATUS
0	Candidate screen	11–18%	none	Hand-set band, not certified
1	Strategic review announced	31.71% / 18mo	λ 95% [7.60, 11.65]	Confirmed, all three bars
2	Received proposal disclosed	65.85% / 24mo	Wilson [57.11%, 73.64%]	Confirmed measurement
3	In-process, stage-typed	50.3% → 77.6%	Wilson, per stage	Settled measurement, rider confirmed

Rungs 1 and 2 are certified in the sense the programme uses that word: a preregistration was frozen before any outcome statistic existed, instrument gates were passed or the unit was stopped, and the frozen primary bar cleared. Rung 3 is a measurement unit rather than a hypothesis test, and its one falsifiable rider cleared. None of the three measured rungs has been validated out of sample. The monotonicity of the ladder is itself a preregistered

result: the proposal rung’s primary bar was that its Wilson lower bound exceed the measured review rung of 0.317. The ordering of rungs 1 and 2 is therefore a tested claim rather than an arrangement of separately measured numbers.

5. Rung 1: A Review of Strategic Alternatives

The first measured rung asks what an announced review of strategic alternatives is worth. The population is the set of 8-K filings containing the fixed phrases “strategic alternatives” or “strategic review”, harvested from full-text search from 2010 forward, 17,258 filings in total. Those filings are deduplicated to one episode per filer per 365 days and restricted to episodes whose eighteen-month outcome window has fully closed, which yields 6,273 scored episodes. The outcome is any deal-family form, including S-4, in the filer’s own submissions record within 548 days. Controls are 2,500 registrants sampled under seed 20260810, excluding every filer in the event set, each assigned a pseudo-episode date drawn from the empirical event-date distribution and scored identically.

Three bars were frozen before any outcome statistic was computed: a two-proportion z of at least 3.0, a bootstrap 95% interval for the lift whose lower bound reaches at least 1.6, and stability in the sense that the lift exceeds 1.3 in both the pre-2018 and post-2018 halves. The measured event rate is 31.71% against a control rate of 3.44%, giving $z = 28.12$, a lift of 9.22 with a 10,000-resample bootstrap 95% interval of [7.60, 11.65], and half-sample lifts of 10.11 before 2018 and 8.34 after. All three bars cleared with room, and the preregistration’s own power note recorded in advance that the design was heavily powered and that an underpowered verdict was not a plausible outcome.

The unit’s preregistration recorded, before any measurement, that literature and practitioner estimates place the transaction rate following a strategic-review announcement at 30 to 50% within six to twenty-four months, and that the control base rate was expected near 4 to 8% per eighteen months. The measured 31.71% at eighteen months falls at the lower edge of the first band and the measured 3.44% at the lower edge of the second. We report this as the register does, as a replication of a stated expectation rather than as a citation, because the specific sources behind the band are not recorded in the platform’s own literature holdings.

Two riders bind the interpretation. The outcome definition detects involvement in a definitive process on either side, since an acquirer-side 425 or S-4 counts as readily as a target-side merger proxy. Strategic-review filers are presumptively targets, but the measured quantity remains either-side involvement, and controls receive the identical definition, which preserves the contrast as the object of interest. The completion caveat of §3 does not bind this rung directly, because its outcome is the appearance of a process rather than its completion. The consequence frozen at preregistration and wired on settlement is that the desk’s strategic-review flag became a probability modifier of the form $p = \min(p_0 \times \text{classifier lift} \times 9.22, 0.60)$. The cap binds at every point of the measured interval, so a flagged candidate reads at the cap.

6. Whether the Wording of the Announcement Matters

An obvious refinement of rung 1 is that not all strategic-review announcements say the same thing. A board that names a banker, or that states explicitly that a sale is among the alternatives, might plausibly be further along than one issuing a general statement. The programme froze four language classes in precedence order, sale-explicit above banker-named above activist-linked above vague, and set a primary bar requiring the top-to-bottom class gap to reach both $z = 3.0$ and five percentage points.

Conversion is flat across wording. On the classified population the banker-named class converts at 36.38% ($n = 492$), the sale-explicit class at 34.44% ($n = 540$), and the vague class at 32.62% ($n = 1,821$ after the frozen downward merge of the activist-linked class, which held 76 episodes against the occupancy floor of 200). The top-to-bottom gap is 3.76 points at $z = 1.57$, below both components of the bar. The preregistration’s power block put the minimum detectable effect between 4.4 and 6.2 points at realistic class occupancies against a plausible band of 8 to 20 points, so the null excludes a literature-scale effect and does not exclude an effect of three or four points.

The instrument work behind that verdict is the more consequential part of the unit. Four outcome-blind corrections were made before any conversion rate was read, each driven by a blind sample and each logged. The first hand validation scored 17 of 20 on class assignment but revealed that roughly half the corpus was not an announcement at all: expense line items of the form “Strategic alternatives, net 1,432”, non-GAAP and forward-looking boilerplate, an investor-meeting title, and acquirer-side deal news. An announcement-context eligibility gate cut the corpus from 6,273 episodes to 2,853, with document coverage of 98.9% on the retained set. The audit printer was then rebuilt to display matched evidence rather than the matched phrase, because three samples had been unscorable as printed. That change immediately exposed the activist class matching Exchange Act Rule 13d-5 boilerplate and, in one instance, a base64 blob. Banker recall was leaking passive constructions such as “is being advised by” and “selected X to serve as financial advisor”.

Residual banker recall remains imperfect and leaks into the vague class, which makes the classes more alike and biases the primary statistic toward zero, so the verdict reported here is conservative. A side finding of the cleaning carries more evidential weight than the primary statistic. On the cleaned announcement-only population the conversion rate runs approximately 33%, against the 31.71% measured on the raw corpus of rung 1, which establishes that the review-rung result survives population cleaning and is therefore not an artifact of phrase-match contamination.

7. Rung 2: Disclosure That a Proposal Has Been Received

The rung above a strategic review is a public statement that a specific approach exists. The operator question of record, frozen into the preregistration verbatim, was whether the pre-announcement probability can exceed 60% at all, since the review-rung formula caps at 0.60 by construction. The population is 8-K and 8-K/A filings from 1 January 2004 to 1 August 2024, the end date capped so that every episode carries a full twenty-four months of outcome runway. An eligible filing matches one of four frozen exact phrases, whose probe counts were recorded in the preregistration: “received a non-binding proposal” (151), “received an unsolicited proposal” (184), “non-binding proposal to acquire” (251), and “proposal to acquire all of the outstanding shares” (288).

The harvest returned 556 filings, deduplicated by accession and then by filer under a 365-day episode rule to 176 episodes. Fifty-three of those were removed by the pre-definitive exclusion described below, leaving 123 scored episodes, of which 81 produced a deal-family filing within twenty-four months. The conversion rate is therefore 65.85%, with a Wilson 95% interval of [57.11%, 73.64%] and $z = 8.14$ against the measured review rung of 0.317. The primary bar, frozen as a requirement that the Wilson lower bound exceed 0.317, cleared. Both parts of the operator’s question resolve affirmatively: the point estimate reaches 0.60 and the lower bound reaches 0.50.

QUANTITY	VALUE	N
Conversion to definitive process, 24 months	65.85%	81 / 123
Wilson 95% interval	[57.11%, 73.64%]	—
Seeded control windows	6.46%	1,347
Delisting-confirmed completion	60.98%	—
Days to first deal-family filing, p25 / p50 / p75	49 / 114 / 228	—
Episodes before 2015	71.01%	69
Episodes from 2015	59.26%	54
Unsolicited wording	63.46%	52
Non-binding wording	67.19%	64

The control comparison is the more informative one for a reader calibrating how much the disclosure is worth. Against 1,347 seeded control windows scored by the identical detector, the base rate is 6.46%, so the disclosure is associated with roughly a tenfold conversion rate. The delisting-confirmed completion of 60.98% is subject in full to the caveat of §3 and should be read as a lower bound on how many of these episodes reached a closed transaction rather than merely a definitive process.

7.1 The instrument record

This unit's measurement is only as good as its identification of who is disclosing what, and the first attempt failed that test. The frozen instrument gate required a blind twenty-sample read in which the filer is the target, disclosing receipt of a proposal about itself, in at least 85% of samples. The first read scored 10 of 20. The failures were interpretable: bidder-filed proposals, counterparty disclosures where the filer was describing someone else's approach, fiduciary-out boilerplate, and asset-only transactions. A document-level self-received filter was added before any outcome statistic was computed, and a fresh blind sample scored 19 of 20, or 95%.

The same reading exposed a second and more damaging contamination. Some matched 8-Ks quote a received proposal inside a background narrative filed at or after the definitive agreement, so the disclosure event postdates the process it appears to predict. A pre-definitive exclusion was frozen, removing any

episode whose filer had a deal-family filing in a window running from 180 days before to one day after the disclosure. That exclusion removed 53 of the 176 accepted episodes, or 30%, and every one of those episodes would have auto-scored as converted. Without the exclusion the measured conversion would have been inflated by a mechanism with no predictive content whatsoever.

Two residual biases are recorded and both run downward. Disclosures worded as declines may be under-captured by the frozen phrase set, and decline wording correlates with hostility, which converts at a lower rate; and approximately 5% residual identity contamination survives the self-received filter. The measured 65.85% is therefore conservative with respect to both.

The design planned for 300 to 600 episodes and scored 123, so the realised sample is well below the planning figure that underwrote the preregistered minimum detectable effect of 8.1 points. The primary bar nonetheless cleared at $z = 8.14$, which is an observed statistic rather than a power claim; the relevant caution is that the secondary splits reported above rest on subsamples of 52 to 69 episodes and should be read as descriptive. The pre-2015 rate of 71.01% against the post-2015 rate of 59.26% is the kind of difference this sample cannot adjudicate.

8. Rung 3: The In-Process Stage Chain

Once a transaction is public the question changes from whether a process will begin to whether the one under way will finish. The stage unit estimates a form-typed chain on the filer population assembled by the strategic-review outcome join, harvesting complete submissions histories per filer and cutting them into

episodes by the frozen 365-day gap rule. The result is 3,056 episodes, of which 86 are censored for insufficient runway and 2,970 are decided. Stages are defined by form code alone: the first family filing opens an episode, a PREM14A marks a preliminary proxy, a DEFM14A a definitive proxy, an SC TO-T a tender offer, and an SC 14D9 the target's response.

STAGE	N	CLOSED	P(CLOSE-WITH-DELISTING)	WILSON 95%	PRE-2018	POST-2018
First family filing	2,970	1,495	50.3%	[48.5%, 52.1%]	42.4%	63.0%
Preliminary proxy	905	660	72.9%	[69.9%, 75.7%]	66.2%	80.9%
Definitive proxy	1,243	965	77.6%	[75.2%, 79.9%]	72.0%	83.9%
Tender offer	575	292	50.8%	[46.7%, 54.9%]	44.3%	71.2%
Target response	338	258	76.3%	[71.5%, 80.6%]	72.5%	84.8%

The ordering is intelligible on its own terms. A definitive merger proxy is filed after a signed agreement and after the preliminary proxy has cleared staff review, and it carries the highest completion probability in the table. The tender-offer stage sits far lower than one might expect from the transaction type, and the pre-2018 and post-2018 split shows why the

level is untrustworthy. Tender completions rise from 44.3% to 71.2% across that split, a larger movement than any other stage exhibits, and one consistent with the delisting proxy behaving differently across eras and structures rather than with a change in deal outcomes of that magnitude. Every figure in the table is a lower bound in the sense of §3.

TRANSITION	P10	P50	P75	P90	N
First filing to definitive proxy	11	44	91	160	1,243
Definitive proxy to close	31	51	111	245	948
Tender to close	29	36	129	476	290
First filing to close	38	120	225	429	1,495

The duration distributions in days are the second deliverable, and they support the unit's one falsifiable rider. Among episodes that reach a definitive proxy or a tender offer, 1,759 in total, the completion rate is 68.6%. Restricting to the 847 that were still unresolved past 113 days, the 75th percentile of the closers' duration, the completion rate falls to 34.8%, a two-proportion z of 16.35 against a preregistered bar of 2.0. Elapsed time in stage is thus affirmative evidence about outcome rather than mere absence of news, and the desk's overdue trigger was converted from a hand-set fifteen-point penalty to this measured conditional on the day the unit settled. The immediate

effect on live positions was that seven of eleven open deals moved off a flat 91% class prior to stage-conditional values of 78% and 76%.

An engine defect was found after the selftest passed and before settlement: the censoring branch compared a string to a date, which the selftest could not catch because every planted episode in the fixture closed and the censoring branch therefore never executed. The register records this as the third instance of the same lesson, that a selftest must exercise every branch it is meant to protect. The defect was corrected before the reported statistics were computed.

9. Failure Predictors and the Inverted Soliciting-Material Result

The most informative finding in this paper is one the programme declined to use. Nine filing-sequence features were specified with their directions frozen before the data were seen, each measured over a deal's first sixty days and each tested against the delisting-proxy outcome on episodes still alive at day sixty. One of them was the count of DEFA14A filings, additional soliciting material, whose frozen direction was that a high volume signals vote trouble and therefore predicts failure. The measured effect is large and points the other way. Episodes in the top tercile of soliciting-material volume fail at 33.10% ($n = 725$) against 65.00% for the rest ($n = 1,900$), a two-proportion z of -14.73 and the largest effect anywhere in the study.

The reading that fits is that soliciting material is the machinery of a transaction being worked rather than the symptom of one in difficulty. A sponsor filing supplemental materials, correcting disclosure, and pressing for votes is a sponsor executing a process it expects to complete. Because the result is wrong-signed against the frozen direction, it is reported without being wired. The preregistration's settlement

rules admit into the desk's probability table only those features that clear the bar in the direction frozen for them, and a post-hoc sign flip is precisely the operation those rules exist to forbid. Promotion would require its own registration and a forward confirmation. That registration would also have to contend with a practical obstacle recorded at settlement, which is that the platform's event queue cannot currently observe DEFA14A filings at all, leaving the signal unobservable on the live surface as well as unwired.

We regard the inversion as the study's most credible result precisely because of its sign. A confirmatory finding on a feature whose direction was guessed correctly is compatible with a great many analytical errors that all push toward the guess; a finding that contradicts the analyst's stated prior, at a magnitude twice that of anything confirmed, is difficult to obtain from a leaking instrument or a hopeful specification. A second feature inverted as well, on smaller evidence: episodes with a family-filing gap exceeding thirty days fail at 49.33% against 58.21% for the rest, a z of -3.84 , so quiet deals complete more often than talkative ones, which is the same story told from the other side.

FEATURE	FROZEN DIRECTION	EXPOSED	UNEXPOSED	Z	DISPOSITION
Slow proxy path at day 60	higher failure	62.24% (1,671)	45.60% (954)	+8.27	Cleared here; withdrawn 11 Aug 2026 (see below)
Tender unresolved at day 60	higher failure	73.18% (343)	53.64% (2,282)	+6.80	Cleared, wired
Soliciting-material volume	higher failure	33.10% (725)	65.00% (1,900)	-14.73	Inverted, reported only
Family silence over 30 days	higher failure	49.33% (596)	58.21% (2,029)	-3.84	Inverted, reported only
Family amendments	higher failure	60.13% (454)	55.37% (2,171)	+1.86	Null
Activist ownership filing	higher failure	54.33% (727)	56.90% (1,898)	-1.19	Null
Insider-filing volume	two-sided	55.85% (965)	56.39% (1,660)	-0.26	Null
8-K volume	higher failure	56.11% (1,096)	56.25% (1,529)	-0.07	Null
Resolicitation	higher failure	50.00% (2)	56.20% (2,623)	-0.18	Too rare in window

Two features cleared the bar in their frozen direction. A proxy-path deal with no definitive merger proxy on file by day sixty fails at 62.24% against 45.60% for the rest, $z = 8.27$ on 1,671 exposed episodes; a tender offer still unresolved at day sixty fails at 73.18% against 53.64%, $z = 6.80$ on 343 exposed episodes. The family-wise control was 500 label permutations within announcement-year strata, whose maximum absolute z reached 3.77 at the 95th percentile and 2.59 at the median. The best real feature therefore exceeds the permutation ceiling by a factor of more than two, and the unit's rider, which requires at least two features clearing in the frozen direction with the best of them beating the permutation 95th percentile, is satisfied.

The population is 2,625 episodes alive at day sixty, drawn from the same 1,775 filers as the stage unit, with 86 censored episodes and 345 fast closes excluded by the frozen conditioning. The pooled failure rate is 56.19%, and the unit's settlement record states plainly that this figure pools genuine breaks with non-delisting closes, so the two cleared features should be read as conditionals on close-with-delisting

rather than on break in the ordinary sense. Read as a completion probability, the tender feature re-bases an unresolved tender offer to 26.8%, under a frozen precedence in which the overdue rider of §8 outranks the day-sixty features, which in turn outrank the stage base rate.

Withdrawal, 11 August 2026. The slow-proxy row above and the day-sixty re-base it supported are withdrawn. A deconfound unit, registered under its own identifier in the platform register and settled on that date, rebuilt the panel offline from the Commission's full-index files over 4,569 filer identifiers and 865,854 form rows, imported this unit's episode, feature and label code unchanged, and re-ran the two cleared features against a frozen four-arm ladder. The tender conditional survived every arm, with statistics between 11.16 and 12.25, and its wiring stands. The slow-proxy conditional did not: its primary arm passed at 3.24 against a bar of 3.0 while both remaining controls failed, at 2.32 and 2.54, and the frozen ladder required all three. The z of 8.27 reported above replicates at 3.12, below that arm's own family-wise permutation ceiling of 3.69, and the re-based rate of 37.8% does not replicate at all, the independent panel returning 48.6% and 53.5%. The mechanism is that the outcome window is anchored on each episode's own last family filing while the feature selects slow transactions, so feature and label share a cause. Under the consequence frozen before that unit ran, the slow-proxy conditional is a directional flag only, the figure of 37.8% may no longer be cited as a measured quantity, and short-side arming may cite only the tender conditional and the overdue rider of §8. This paper predates the deconfound; the correction is recorded here rather than in a revision of the text above so that what was claimed and what replaced it both remain legible. The companion staged-hazard paper reports the deconfound in full.

10. Survivorship, Contrasts, and Levels

Two distinct restrictions on interpretation operate in this programme and they are easy to conflate. The delisting proxy restricts levels on the deal side, as §3 sets out. Survivorship restricts levels on the equity side, and although no equity result appears in this paper, the convention the programme applies there is the one it applies here and is worth stating in full, because it is what licenses the tables above.

The programme's quality-dislocation results are computed on a panel of current index constituents, 3,298 episodes across 499 names. On that panel a stage-classification contrast returns +9.31% against +5.36% over a sixty-three-trading-day horizon on

arms of 2,051 and 1,247 episodes, a difference of 3.95 points with a Newey-West t of 5.18 against a permutation 95th percentile of 0.33 points. A time-conditioning contrast on the same panel returns +10.13% against +5.48% on arms of 1,656 and 1,642 episodes, a difference of 4.65 points at $t = 4.99$ against a permutation 95th percentile of 4.53 points, a margin of 0.12 points that the register discloses as thin. Because membership is current, the panel excludes names that left the index, which inflates its levels while leaving its contrasts unaffected. The programme states this wherever those numbers appear rather than once in a limitations paragraph.

The deal side inherits the identical discipline for a different reason. The stage ladder's 50.3% and 77.6%, the proposal rung's 65.85% and its delisting-confirmed 60.98%, and the failure study's 56.19% base are all lower bounds whose looseness varies with consideration type across the sample. A reader who takes the definitive-proxy stage as a completion probability for a particular cash transaction will understate it. A reader who takes the difference between the definitive-proxy stage and the first-filing stage, or between overdue and not overdue, is comparing quantities measured by the same proxy on the same population and is on much firmer ground. It is those differences that the paper offers as citable, and the levels are reported for completeness alone.

The forward record is the third restriction and the most consequential for anyone contemplating capital. Every result in this paper is in sample. The programme maintains a forward ledger intended to replace its class prior on definitive cash deals with a rate measured on its own resolutions, and that ledger records 136 resolutions against the thirty its rule requires. The count carries a provenance caveat. The desk archive holds 204 resolved deals, but for most of the ledger's life the routine that wrote those records discarded the entry spread, so usable spread-to-outcome pairs stood at zero until a backfill from the desk's snapshot history recovered entry records for 136 of them. The remaining 68 predate the snapshot history and cannot be recovered. The quorum is therefore met, and the measured base rates remain unadopted because the hypothesis registered over

them, which governs the merger-arbitrage spread, has not settled. No claim in this paper is supported by out-of-sample evidence, and the honest description of the ladder's status is that it is a set of well-instrumented historical measurements whose forward confirmation is pending that settlement.

11. Instrument Discipline

Two of the five units reported here, the proposal rung of §7 and the language unit of §6, changed their instruments before reading any outcome, and in each case the change was driven by a blind sample rather than by a result. Of the remaining three, the strategic-review and failure-prediction units record selftests passing with no instrument change, while the stage unit's single correction was the censoring defect of §8, found after its selftest passed. The procedure the two revised units followed is uniform enough to describe as a protocol. A hand-validation gate is frozen with the design, stating a sample size, a criterion, and an acceptance threshold. The sample is read blind to outcomes. If the gate fails, the instrument may be tightened and a fresh sample drawn, and the tightening is logged with its date; if it cannot be brought to threshold, the unit settles as measurement-limited without reading its statistics.

The proposal rung passed on its second attempt after the self-received filter took target identification from 10 of 20 to 19 of 20. The language unit made four such corrections and, in doing so, discovered that its population was half contaminated by documents that were not announcements. Two sibling units in the same family stopped instead of proceeding. A spread-calibration study found that 8 of 20 sampled rows carried junk tickers produced by a cover-page extraction bridge, and sealed its statistics unread; a post-termination study found that a classifier evaluated over a filer's entire history rather than over the terminated episode mislabelled 42 of 43 events. Those two failures are catalogued in a companion paper on the boundaries of free filing data. Their relevance here is that the same protocol which passed

the two rungs reported above stopped two neighbouring units, which is the behaviour a gate must exhibit to be worth anything.

Selftests precede every run and are required to demonstrate both detection and refusal: a planted effect must be found, a planted null must not clear, and a leakage tripwire must show zero exposure to information dated after the feature window. The stage unit's censoring defect, described in §8, is the standing counter-example and the reason the programme now requires that fixtures exercise every branch: a selftest whose protected branches never execute provides no evidence about the code those branches contain.

12. Reproducibility

Every figure in this paper is quoted either from the results file a unit wrote at settlement or from the register entry that records its verdict, and where the two are read together the results file governs, since the verdict is a prose summary of it. Each unit is implemented as a command-line program whose modes run in a fixed order. A harvest queries the Commission's endpoints and caches every response to disk; a selftest exercises the scoring code on planted fixtures and must pass before scoring is permitted; a scoring pass then reads only cached material and writes the results file. No network request is made during scoring, so a replicator who has completed a harvest can re-derive every statistic reported here without returning to the Commission.

The strategic-review rung of §5 is the one unit divided across three programs run in sequence. The harvester queries full-text search for 8-K filings matching the two frozen phrases and writes one record per matched filing. The outcome join then reads each filer's complete submissions history, types the outcome against the frozen deal-family form list, and writes one scored episode per event, carrying its outcome flag and the date of the filing that produced it. The scorer seeds the control pool, computes the two-proportion statistic, the bootstrap lift interval and the two half-sample lifts, and writes the results file the tables of §5 quote. Its selftest is a pair of planted panels: one with

a 20% event rate against 5% controls, which the scorer must clear, and one with 5% in both arms, which it must refuse.

The remaining four units are each a single program. The language study of §6 consumes the strategic-review event file together with the cached filing documents, applies the announcement-context eligibility gate, assigns the four frozen classes in precedence order, and writes per-class conversion rates with the top-to-bottom test statistic. Its four selftest checks cover class assignment under precedence, a shuffled-label null that must not clear, a leakage tripwire establishing that the classifier reads the document alone, and the tolerance of the join between an episode and its filing. The proposal rung of §7 harvests on its four frozen phrases, deduplicates to episodes, applies the self-received filter and the pre-definitive exclusion, and scores conversion against seeded control windows. Its five checks cover a planted converter, a planted non-converter, a leakage tripwire requiring zero filings dated on or before the disclosure, the deduplication rule at 100 and 400 days, and reproduction of the strategic-review control base rate on a 500-window draw.

The stage chain of §8 consumes complete submissions histories for the filer population, cuts them into episodes under the 365-day gap rule, and emits both the per-stage completion table and the duration percentiles. Its selftest plants a synthetic sequence set with known transitions, known durations and a known overdue effect, all of which the scorer must recover, together with a null plant whose overdue effect must fail the rider bar. The failure study of §9 consumes the same episode sequences, measures the nine frozen features over each deal's first sixty days, and writes the per-feature contrasts alongside the permutation ceiling. Its four checks are a planted soliciting-material effect that must clear, a random feature that must not, a leakage tripwire that plants a feature only in filings dated after day sixty and requires the window guard to refuse it, and an exact reproduction of the fixture's episode count and labels.

Every preregistration carries a power block stating the design's minimum detectable effect, a plausible effect size for the class, and an explicit statement of whether

the design can detect it, and each was frozen before its unit read any outcome statistic. Harvests are resumable and cache to disk, so a replicator repeating this work pays the Commission's rate limit once. The dominant cost is the wall-clock time of querying the submissions endpoint at a 1.2-second throttle across roughly 1,775 filers, and the computation itself is negligible by comparison. The exact names of every program, cache, results file and preregistration behind the five units are recorded in the platform repository and in each unit's register entry.

Calibration record. Each unit froze a single probability that its primary bar would pass, scored at settlement as a Brier contribution: strategic-review hazard 0.65 against a pass, proposal ladder 0.75 against a pass, stage chain 0.60 against a cleared rider, failure predictors 0.55 against a cleared rider, language conditioning 0.55 against a null. The corresponding Brier contributions are 0.1225, 0.0625, 0.16, 0.2025 and 0.3025. The programme's intent is that after roughly twenty such units the calibration curve replaces argument about confidence with arithmetic.

13. Discussion

The ladder is monotone where it has been tested and the increments are large. A candidate carrying no public signal sits at a hand-set band that no measurement supports. A strategic-review announcement raises the eighteen-month probability of entering a definitive process to 31.71%, roughly nine times the control rate, and the wording of that announcement adds nothing detectable at the power available. A disclosure that a proposal has been received roughly doubles that again, to 65.85%, with a lower bound above 0.50 and a control base rate of 6.46%. Once a process is public, the stage a transaction has reached and the time it has spent there both carry information, with completion probabilities running from 50.3% to 77.6% across the chain and an overdue condition cutting completion among advanced-stage episodes from 68.6% to 34.8%.

What is being measured, at every rung, is the observable consequence of processes conducted in private. A board announces a review when it has decided to test the market, and a target discloses a

received proposal when a legal obligation or a negotiating advantage requires it. Supplemental soliciting material, on the reading §9 supports, is filed in the course of working a vote the sponsor expects to win. The filing record does not observe the negotiation, and none of these measurements should be read as identifying a causal channel. They are conditional frequencies attached to publicly observable events, which is what a probability ladder needs to be useful and is also the limit of what filing data can supply.

The inverted soliciting-material result is the clearest evidence in this paper that the protocol is doing work. A study designed to find failure predictors returned its largest effect on a feature whose frozen direction was wrong, and the rules written before the data were seen required that the finding be reported without being adopted. The measurement is credible in a way a confirmed guess would not have been, since it survives the objection that the analyst found what the analyst expected. The refusal to adopt it preserves the property that makes the wired conditionals worth anything, which is that they were specified in advance of the evidence that supports them. Promotion of the inversion would require a fresh registration and a forward test, and until the platform's event queue can observe the filings in question, that test cannot begin.

Nothing in this paper licenses a change to a live decision rule beyond the changes already recorded at each unit's settlement, and none of those changes rests on out-of-sample evidence. The programme's forward ledger records 136 resolutions against its thirty-resolution quorum, recovered largely by the backfill described in §10, and the rates it supports remain unadopted pending settlement of the registered hypothesis on merger-arbitrage spreads. The ladder's value at present is that it replaces hand-set numbers with measured ones, states an interval on each, and names the proxy through which every completion figure must be read.

References

- [1] C.R. Harvey, Y. Liu & H. Zhu. ...and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016. The $t > 3.0$ convention adopted as the programme's default significance bar.
- [2] M.A. Andreev & J. Howden. Stochastic mechanical methods for quantitative portfolio management. *Quark Quantitative Research Working Paper*, 2026a.
- [3] M.A. Andreev & J. Howden. Ontological limits of historical validation for wavefunction-calibrated financial systems. *Quark Quantitative Research Working Paper*, 2026e.
- [4] M.A. Andreev & J. Howden. The reachable boundary of free filing data for event-driven research: a catalogue of measurement limits with named defect classes. *Quark Quantitative Research Working Paper*, 2026. Companion paper; catalogues the terms-extraction, delisting-and-price, and control-composition boundaries referenced in §11.
- [5] Preregistrations for the five units reported here, each frozen before its unit read any outcome statistic, held in the experiments directory of the platform repository under the register identifier of the unit it governs.
- [6] Register of hypotheses and verdicts, maintained in the platform's research engine, carrying for each of the five units its frozen hypothesis text, its terminal status, its verdict and the paths of the programs and results files behind it.

© 2026 Maksim Aleksandrovich Andreev & Joseph Howden.
All rights reserved.

Quark Quantitative Research Platform — Working Paper
Generated 2026-08-17 21:46 UTC

Every quantity reported here is recoverable at no cost from the Securities and Exchange Commission's full-text search index and per-filer submissions endpoint. Completion figures are delisting-

proxy lower bounds; all results are in-sample and none is forward-proven. Correspondence: M.A. Andreev, Quark Quantitative

Research.