

LAMINAR: Measuring an Order-Flow Coherence Label Across Twelve Futures Instruments

Bee Rosa Davis
Davis Geometric
`bee_davis@alumni.brown.edu`
ORCID: 0009-0009-8034-4308

10 September 2026

Working paper · DOI: 10.5281/zenodo.22681795

Abstract

We evaluate LAMINAR, a session-relative order-flow classifier, to measure the incidence, perturbation sensitivity and context dependence of its coherence label. A fixed configuration is evaluated on a common numerical price scale across 15 futures tapes representing 12 instruments, 36,533,928 trades and 194 trading-day sessions. Its LAMINAR label appears in every included session, with pooled incidence from 6.35% to 9.67% among non-initializing reporting observations.

Incidence alone does not identify market-specific organization. A control grid of 70 executions covers seven source sessions in four instruments. Surrogates that independently resample session marginals produce mean incidence at least as high as the real inputs in three of those instruments. The tested S&P session shows a larger response to price-path permutation: incidence falls from 7.94% to 4.92%–5.56% across two seeds.

Selection within inserted episodes provides a different measurement. Grafted real segments have higher mean inside than outside incidence on gold and the S&P, but lower mean inside incidence on yen and the 2-year Treasury future. Synthetic episodes yield a different pattern. These results distinguish frequency of selection, sensitivity to the chosen perturbation, and retention of a label in a changed context. A separate controlled replay also shows why price representation and withheld classifications must be recorded explicitly. The findings concern the measurement and its controls; predictive returns, execution performance and the classifier’s proprietary construction are outside the study’s scope.

1 Introduction

Order flow is both a sequence of transactions and an input to price formation. Market microstructure research studies how its composition and ordering relate to prices and liquidity [2, 3]. A classifier can summarize that activity by selecting observations that meet an operational criterion. Interpreting those selections requires more than counting them: the same frequency could arise from organized market activity, from the mechanics of a relative threshold, or from their interaction.

This paper separates three empirical questions. *Incidence* asks how frequently the criterion is met. *Sensitivity* asks which controlled changes in the input alter its output. *Portability* asks whether selected segments tend to be selected again when their surrounding context changes. These quantities need not agree. A detector may select at a similar rate on real and surrogate inputs while assigning its selections to different places.

In classical fluid mechanics, *laminar* flow is orderly motion in layers: neighboring layers slide smoothly past one another without the irregular cross-stream mixing characteristic of turbulence. In Reynolds’ pipe experiments, a filament of dyed water remained distinct in the ordered regime and dispersed into the surrounding water as the flow became turbulent [1, p. 942]. The distinction concerns the organization of motion; neighboring layers can still move at different speeds.

We call the selected financial state *laminar* because the classifier seeks a corresponding form of collective organization in order flow. In the configuration studied here, the label is assigned when a spectral measure of trade-transition connectivity and a measure of temporal persistence in a derived flow signal are jointly elevated relative to their recent session histories. The name connects two aspects of the criterion: how transactions are connected and how the resulting flow signal evolves through time. The physical analogy motivates the criterion; these financial measurements define the selected state.

We investigate that state through a cross-instrument sample, a controlled representation experiment, a surrogate ladder, and synthetic and grafted episodes. The contribution is a measurement account: which conclusions the observed frequencies support, and which depend on the perturbation or reference context used to interpret them.

2 Measurement and observation units

The classifier consumes trades in arrival order, using price, quantity, aggressor side and prevailing quote prices. The replay uses a spectral representation of trade transitions with rolling, session-relative calibration. Each session starts with a fresh classifier; its history continues across reporting-bin boundaries. The trade-transition configuration uses fixed bid and ask sizes, so these measurements do not incorporate cross-instrument variation in actual displayed depth.

Every instrument is supplied on the same numerical price convention. If δ is the smallest nonzero absolute price change observed in its tape, the replay supplies the classifier with the scale factor $0.25/\delta$ for prices and quotes. Quantity, aggressor side and timestamps retain their input values. This aligns one observed price step with the reference increment; it does not equalize event rates, volatility, trade-size distributions or market economics. An observed minimum step is also distinct from an independently verified exchange tick-size schedule.

2.1 Sampling and denominators

The reporting grid contains non-overlapping 120 s bins. Each populated bin records the classifier state at its last observed trade. That state need not have held throughout the bin. Empty bins contribute no labels. The final pending bin is flushed and stored separately in each session receipt.

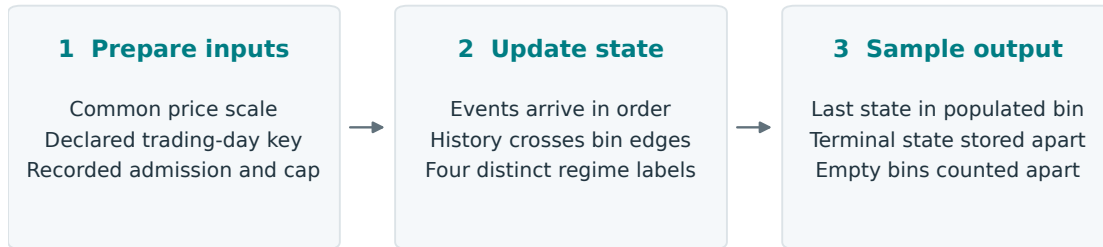
We retain four labels: LAMINAR, TURBULENT, TRANSITIONAL and INITIALIZING. TRANSITIONAL withholds a regime assertion; it is not a synonym for TURBULENT. Let L , U , R and I denote their respective counts. We report

$$p = \frac{L}{L + U + R}, \quad p_A = \frac{L}{L + U}, \quad a = \frac{L + U}{L + U + R}.$$

Thus $p = a p_A$: a tape's incidence combines how often the detector asserts a regime with how often an asserted regime is laminar. Undefined ratios are left undefined. Tape rates pool counts across sessions rather than average session percentages. Session coverage counts sessions containing at least one LAMINAR observation.

What one observation represents

SCHEMATIC / Event-driven state, sampled on a two-minute reporting grid



Two denominators answer different questions



L: laminar U: turbulent R: transitional I: initializing

Figure 1: The measurement contract. The diagram is schematic. Event-history memory and reporting cadence are distinct. Full regime labels preserve the difference between a turbulent assertion and a transitional reading. Neither reported share estimates continuous elapsed-time occupancy.

The cross-instrument table excludes separately stored terminal bins from its primary rates. The controls include them. The real control executions reproduce the corresponding session counts exactly after adding those terminal labels, providing a direct check that the two reporting conventions reconcile.

3 Sample, admission and representation

The sample uses licensed CME Globex order-by-order data, obtained from Databento. Continuous-contract pulls are admitted using exchange status intervals, a leading instrument within each interval, and valid reconstructed books. Eligible runs shorter than 500 trades are excluded before classification. The replay orders admitted runs by their first trade and requires one lead instrument within each session. A 700,000-trade cap applies per session; the included sample contains 2 cap flags.

For these daylight-saving-time dates, the declared trading-day key is the UTC date of the status-interval start plus two hours. This aligns the intended 22:00 UTC opening with the following date. The implementation refuses interval starts outside its specified daylight-saving range. Session receipts record interval bounds, lead instrument, first and last used trade, admitted counts and cap status. This is a declared partition of the admitted tape, rather than an assumption that every session contains a complete exchange day.

Tape	Trades	N	p (%)	p_A (%)	Sessions	Period
RTY	421,982	3,370	9.67	16.47	5/5	Jul
6E	133,221	3,231	8.82	14.91	5/5	Jul
YM	277,026	3,299	8.58	15.17	5/5	Jul
CL	3,392,888	12,964	8.07	14.00	19/19	Mar–Apr
ZT	128,267	2,976	7.90	11.60	5/5	Jul
NQ	1,795,518	3,385	7.68	13.18	5/5	Jul
GC	331,552	3,377	7.67	13.45	5/5	Jul
6J	95,882	3,221	7.48	12.31	5/5	Jul
ES	10,298,409	13,303	7.35	12.50	20/20	Mar–Apr
ZB	144,380	2,904	7.30	9.23	5/5	Jul
ZF	263,750	3,060	7.06	8.90	5/5	Jul
ES-apr	8,887,782	13,720	7.01	11.46	20/20	Apr–May
ES-jul	2,088,165	3,385	7.00	11.62	5/5	Jul
ZN-summer	5,647,076	42,144	6.73	9.20	65/65	Apr–Jul
ZN	2,628,030	13,118	6.35	8.55	20/20	Mar–Apr

Table 1: The measured sample. $N = L + U + R$ excludes initializing and separately recorded terminal bins. $p = L/N$; $p_A = L/(L + U)$ conditions on an asserted regime. Sessions reports those with at least one LAMINAR observation over all included sessions. Periods are in 2026.

The sample contains 127,457 primary non-initializing observations: 9,208 laminar, 75,433 turbulent and 42,816 transitional, with 397 initializing observations recorded separately. There are 194 terminal records and 2,463 skipped grid bins between observed bins. Skipped-bin counts can include gaps between admitted runs and are not a census of all missing exchange activity. Including terminal records changes a tape’s pooled rate by at most 0.027 percentage points in this sample.

3.1 A controlled price-representation experiment

A separate yen replay holds the events, configuration, fixed replay-group partition and reporting convention constant while varying the numerical price representation. It covers 95,882 trades in five groups and 3,221 non-initializing observations. Applying the reference scale externally and supplying the same scale through the classifier’s configuration produce identical saved regime counts in every group.

Price representation	LAMINAR	TURBULENT	TRANSITIONAL	p (%)
Native numerical units	0	0	3,221	0.00
External reference scaling	241	1,717	1,263	7.48
Configured reference scaling	241	1,717	1,263	7.48

Table 2: Yen representation control with fixed events and segmentation. Each condition also has 18 initializing observations. Terminal bins are excluded. The experiment’s fixed group partition is retained for the paired comparison; these counts are not added to the main sample.

The native-unit zero here is entirely transitional: the detector makes no laminar or turbulent assertion. Treating it as evidence of uniformly turbulent market behavior would change the meaning of the recorded output. The experiment establishes sensitivity to representation and count-level agreement between two implementations of the reference scaling. It does not establish invariance under every change of units or equality of every event-level trajectory.

3.2 Incidence across instruments

Every included session contains at least one LAMINAR observation. Tape rates span 6.35%–9.67%; rates conditional on an asserted regime span 8.55%–16.47%. Figure 2 places those two tape-level quantities beside the session observations. The distributions show that a pooled rate is not a narrow interval within which every session must fall.

The asset labels identify the sampled instruments, not exchangeable draws from their entire asset classes. Some instrument comparisons also involve different calendar periods. Incidence can therefore be reported for this sample without attributing its differences to liquidity, depth, event density or an inherent asset-class property. Those explanations require their own matched measurements.

Incidence and variation across the measured tapes

Points: session rates | Diamonds: pooled incidence | Circles: asserted-only incidence

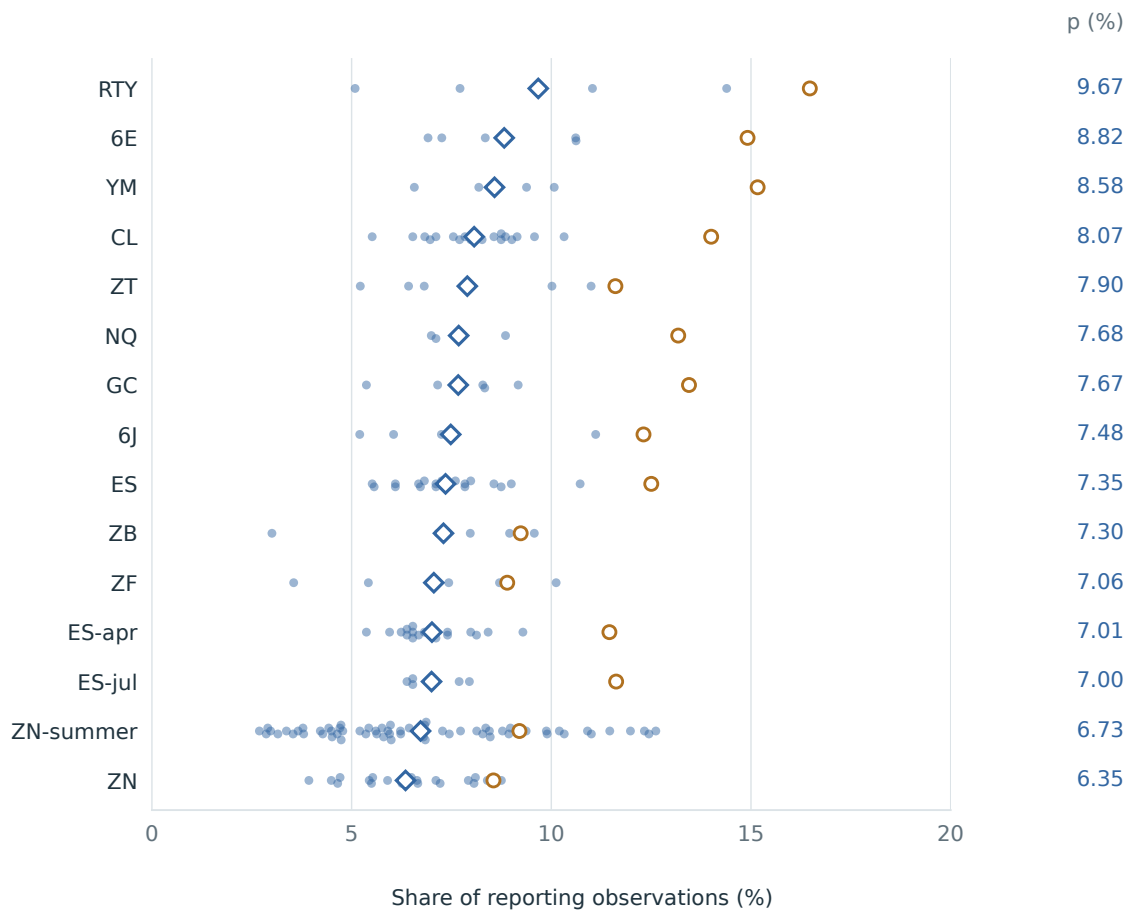


Figure 2: Observed session variation and two tape-level denominators. Blue points are individual session rates p ; blue diamonds pool their counts. Amber circles show p_A , the rate among asserted regimes only. Small vertical offsets separate nearby points without changing their measured rates. The right column gives pooled p . These distributions are descriptive observations, not confidence intervals.

Conditioning on assertion raises the numerical rate because it removes transitional observations from the denominator. The size of that change differs between tapes. Reporting p and p_A together makes this contribution visible rather than treating every non-initializing reading as an asserted market regime.

4 Surrogate and inserted-episode controls

The control grid comprises 70 executions: one real-input execution and six control types for each of four instruments. Yen, 2-year Treasury and gold each use the sessions keyed 6 and 7 July 2026 with three seeds per control type. The S&P uses 6 July with two seeds per control type. There are seven distinct source sessions across four instruments, not 70 independent market samples. The three instruments with three seeds contribute 19 executions each; the S&P contributes 13.

Signs. Aggressor signs are permuted within a session, with prices, absolute trade sizes, quotes and timestamps held in their slots. This changes the assignment and ordering of flow direction without changing the price path.

Path. Observed price increments are permuted and accumulated from the first price. Quotes are recentered using each event’s half-spread; sizes, signs and timestamps remain in their slots. The transformation preserves the increment multiset while changing its order and its pairing with other inputs.

Independent resampling. Price increments, signs, sizes and half-spreads are sampled independently from the source session’s empirical marginals. Event timestamps are retained. This is a conditional surrogate: it still inherits the session’s empirical distributions, event-time schedule and finite numerical lattice. Its prices are accumulated increments, not independent price levels. We use it to measure selection under a specified loss of dependence, not as a universal model of randomness.

Blocks. Ten-minute source blocks are reordered, their internal event sequences retained, and prices rebased to join the blocks. The implementation starts each block one millisecond after its predecessor. This compresses gaps and changes alignment with the reporting grid, so it can also change the number of reporting observations. A block contrast jointly measures sensitivity to reordering, timing and sampling; it does not isolate a pure clock effect.

Synthetic episodes. Organized price mixing, drift and one-sided flow are inserted into an independently resampled background, targeting roughly a tenth of the events. Episode length is derived from mean session event rate with a minimum event count. The construction supplies a known intervention; it does not establish that this particular pattern exhausts the economic meaning of coherence.

Grafts. Segments from bins selected on the real source session are rebased and transplanted into a resampled background, subject to an event budget. Their price and signed-size sequences are retained, but recipient timestamps and surrounding history change. This is a conditional portability probe based on previously selected segments, rather than an independently labelled ground-truth classification task.

5 Selection frequency under perturbation

The independently resampled controls produce mean incidence of 10.33%, 8.97%, 8.47% and 6.98% for yen, 2-year Treasury, gold and S&P, respectively. Real-input rates on those same source sessions are 9.36%, 7.67%, 8.19% and 7.94%. Mean surrogate incidence is therefore higher in three of four tested instruments. A positive incidence reading alone does not distinguish these real inputs from the tested surrogate process.

Observed incidence alongside surrogate controls

Dashed line: real tape | Points and segment: seed results | Diamond: seed mean

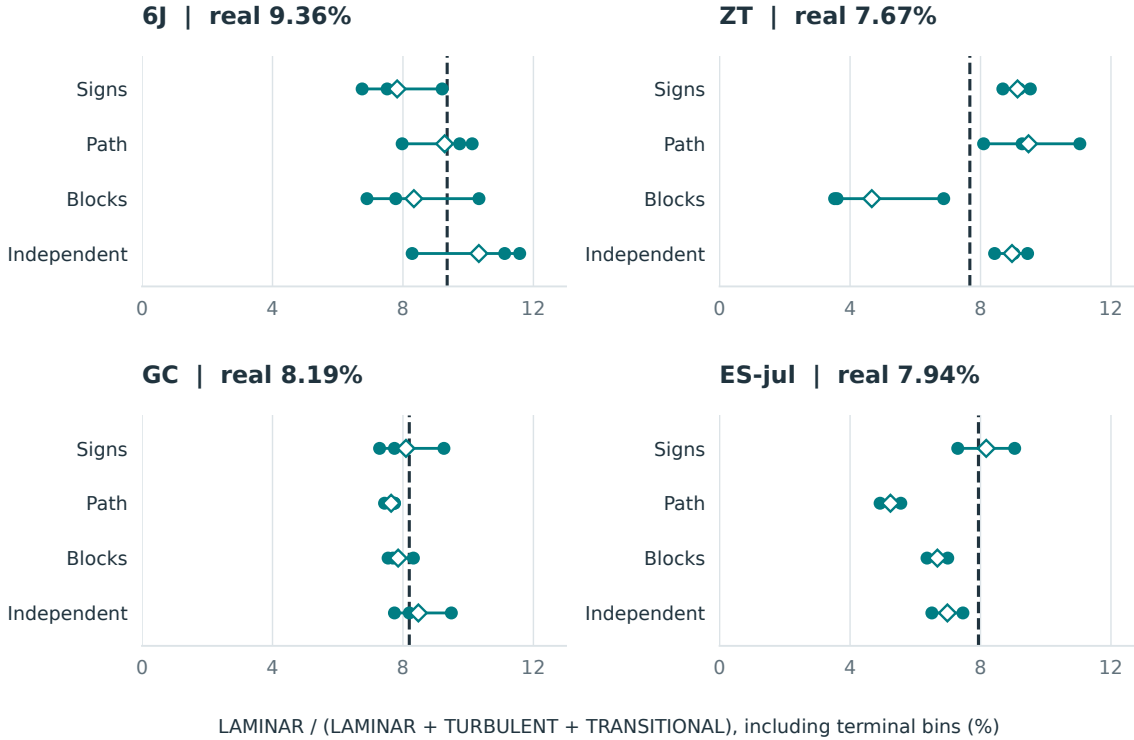


Figure 3: The surrogate ladder on the seven control sessions. Each point is one seed's pooled rate over the instrument's included control sessions; segments span the observed seed range and diamonds mark its mean. Dashed lines show the matched real-input rate. There are three seeds per surrogate for 6J, ZT and GC and two for ES. All panels use the same horizontal scale. Seed ranges are not confidence intervals.

The largest price-path contrast occurs in the tested S&P session: its 7.94% real rate exceeds the 4.92% and 5.56% path-permuted rates, a mean difference of 2.70 percentage points. Gold also exceeds all three of its path-permuted rates, with a smaller mean difference of 0.56 points. Thus the result is not uniform insensitivity to input order. The magnitude and direction of the response depend on the instrument and transformation.

5.1 What the null comparison establishes

A session-relative criterion can select a nonzero fraction of observations from a surrogate as well as from a market sequence. Relative thresholds and dependence among the underlying statistics both affect that fraction. The control grid measures the resulting incidence under explicit input models; it does not establish a mathematically fixed selection quota.

The 2-year Treasury block rate averages 4.67%, below its 7.67% real rate. The block implementation also modifies timing and grid alignment, so this difference cannot by itself establish that event density causes the change. Likewise, changing signs or increments changes their relationships with other inputs. Each surrogate is a specified intervention with a limited interpretation, not a selective removal of every possible form of structure.

6 Concentration and portability of selections

For inserted episodes, a reporting bin is designated “inside” when at least half its events belong to the inserted material. Its label remains the last observed state in that bin. Inside and outside rates each exclude initializing observations and include terminal bins. We compute the difference within each session–seed pair. Table 3 reports equal-weight means of the recorded, rounded pair rates and their ratio; it does not pool all event counts or average individual ratios.

Market	Control	Inside (%)	Outside (%)	Ratio	$\Delta > 1$ pp
6J	Synthetic	15.10	8.98	1.68	4/6
6J	Grafted	6.18	7.55	0.82	2/6
ZT	Synthetic	11.55	8.08	1.43	3/6
ZT	Grafted	2.72	10.45	0.26	0/6
GC	Synthetic	16.77	8.65	1.94	5/6
GC	Grafted	13.90	7.78	1.79	4/6
ES-jul	Synthetic	3.25	8.15	0.40	0/2
ES-jul	Grafted	12.15	8.25	1.47	2/2

Table 3: Equal-weight means over session–seed pairs of the recorded inside/outside rates. Ratio divides those two means. The final column counts paired differences exceeding one percentage point; it is a descriptive cutoff, not a significance test. Six pairs comprise two sessions and three seeds; the two ES pairs comprise one session and two seeds.

The one-percentage-point cutoff in the final column is a descriptive summary. Counting every strictly positive difference instead gives five of six yen synthetic pairs and four of six 2-year synthetic pairs. Showing the continuous differences in Figure 4 avoids making the interpretation depend on that cutoff.

Where selections fall within each control session

Every point is one session-seed pair; diamonds show their mean difference

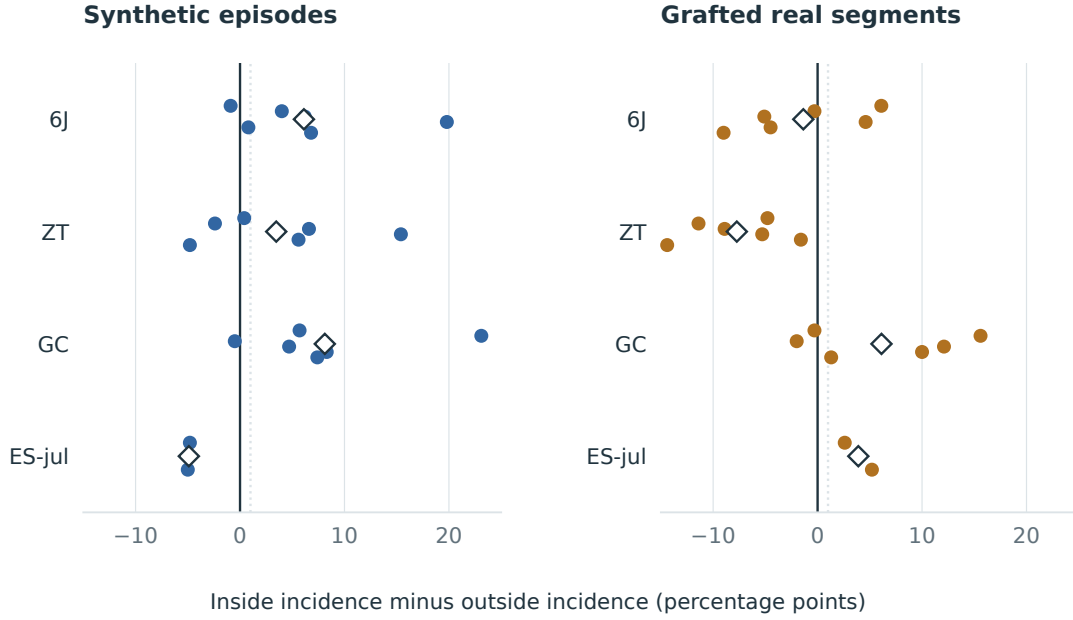


Figure 4: Within-session concentration of selections. Positive values mean higher LAMINAR incidence inside inserted material than outside it. Every session-seed pair is plotted, with small vertical offsets for legibility; diamonds show mean paired differences. The solid line marks zero and the dotted line marks one percentage point. Six points represent two source sessions across three seeds; the two ES points represent one source session across two seeds.

Gold’s synthetic episodes show mean inside incidence approximately 1.94 times outside incidence, with five of six differences exceeding one percentage point. Its grafted segments show a ratio of 1.79, with four of six differences above that cutoff. In the S&P session, both grafted pairs exceed the cutoff and their ratio of mean rates is 1.47, while neither synthetic pair has higher inside incidence. A detector’s response to one constructed pattern therefore does not determine its response to selected market segments.

Yen grafts have a ratio of mean rates of 0.82, despite two positive pairs. The 2-year graft ratio is 0.26; all six pairs have lower inside than outside rates and three have zero observed inside incidence. Under these graft conditions, selected source segments do not reliably retain elevated selection in the recipient context. This is evidence about that transplantation experiment, not proof that the source market lacks organization.

The contrasting synthetic and graft results matter as much as the successful cases. Portability is conditional on segment selection, rebasing, recipient timing, episode length and the classifier’s surrounding reference history. Repeated seeds vary those constructions; they do not create additional independent source sessions.

7 Temporal support and the memory scale

Figure 5 places the available ES, CL and ZN session rates on one calendar axis. Its purpose is to make the temporal support of a tape mean explicit. Horizontal segments summarize observed periods; they neither fit a trend nor interpolate across unmeasured dates. A rate computed from one period does not establish continuous persistence over the months surrounding it.

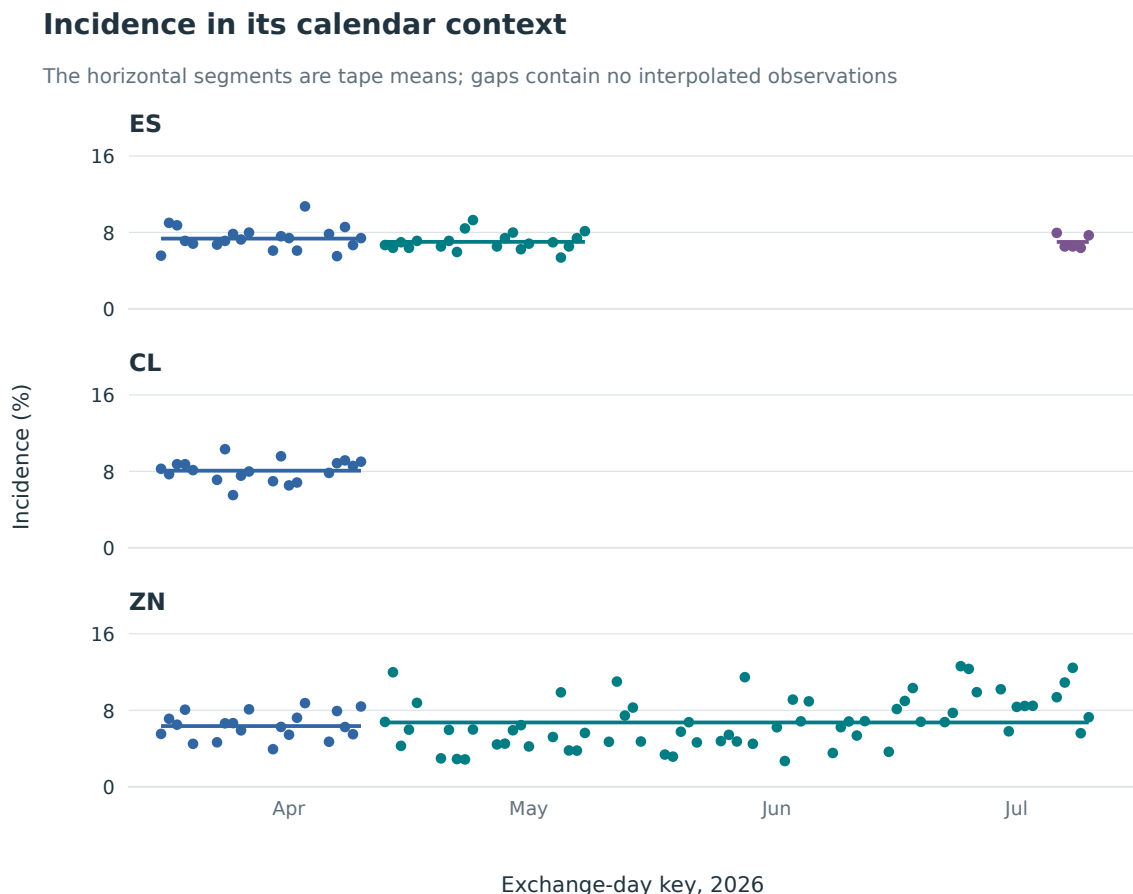


Figure 5: Calendar support for three instruments in the measured sample. Points are individual session rates and horizontal segments are pooled tape rates. All panels share their calendar and incidence scales. Multiple periods, where present, have distinct colors. Only sessions represented in the evidence receipts are plotted.

The classifier's rolling history advances in events; the reporting grid advances in clock time. Equal event counts can span different elapsed durations across markets or within a day. Event density is consequently a plausible moderator of the control results, but the grid does not isolate it. A direct test would vary event timing or history length while controlling the event sequence and reporting convention. The present evidence supports this as a testable explanation, not an established dense-versus-thin market law.

8 Interpretation and conclusion

The measured incidence is a property of a declared procedure applied to a specified tape. Full labels and explicit denominators separate positive selection from withholding an assertion. The common price convention and controlled representation experiment make numerical scale part of the measurement specification. Together, these choices make a cross-instrument frequency comparison interpretable without treating its labels as direct observations of an intrinsic market regime.

The control results distinguish occurrence from selectivity. Independently resampled inputs sustain similar incidence, while path permutation and grafting produce instrument-specific responses. Gold and the tested S&P session show concentration within grafted segments; yen and the 2-year future show lower mean inside than outside incidence under the same type of intervention. These findings warrant conditional statements about the tested sequences and controls, rather than a universal fixed quota or a causal claim about density.

The controls cover seven source sessions and use a small number of seeds. Their contrasts are descriptive; no population confidence intervals or significance claims are assigned to the seed ranges. Wider matched-session coverage, event-time controls and an unused temporal holdout would test how these patterns generalize. Directional outcomes and executable trading value require separate measurements on the selected observations. Neither successful grafts nor null occupancy establish those outcomes.

Data availability

Evidence package. The completed session receipts, the 70 control executions and the three representation-control receipts, with source hashes, exact-count checks, the generated tables and figures, the manuscript source and the audit scripts, are deposited as a dataset on Zenodo: doi:10.5281/zenodo.22681795. The real control counts reconcile to the main session receipts after terminal-bin inclusion. Tables and figures are generated from these saved artifacts; no new classifier replay is performed by the manuscript builder.

Derived counts, current version. The per-session window counts for every tape, as served by the project site: helicity.io/evidence/reynolds_laminar.json. The deposited package is the frozen copy of record.

Source data. Licensed CME Globex market-by-order events, Databento dataset GLBX.MDP3 (databento.com/datasets/GLBX.MDP3). The licence does not permit redistribution of the events; the derived receipts above are sufficient to regenerate every table and figure.

This paper. Working-paper record: doi:10.5281/zenodo.22681795.

Disclosure

The author developed the classifier evaluated here, which Davis Geometric offers commercially under the name LAMINAR. United States provisional patent applications covering the method are pending. The study reports measurements and controls; none of its findings depends on the status of those applications.

References

- [1] Reynolds, O. (1883). An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Philosophical Transactions of the Royal Society of London*, 174, 935–982.
- [2] Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335.
- [3] Easley, D., López de Prado, M. M., & O'Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies*, 25(5), 1457–1493.