Online Resource 2

Supplementary Appendices and Predeclared Message-Level Validation Protocol

Shear-Thickening Markets: State-Dependent Liquidity Frictions and Endogenous Tail Risk

Review of Quantitative Finance and Accounting

Anonymous peer-review version

Author-identifying metadata are omitted from this review file in accordance with the journal's double-blind policy. Author details and declarations are supplied separately through the submission interface and title-page file.

Contents

Appendix A. Historical data, measurement, and extended results

A.1 Data source and exact analysis vintage

The historical analysis uses the daily Fama/French U.S. market factor constructed from CRSP records. The total market return is (Mkt-RF + RF)/100. The archived input is the November 2024 data vintage, locally named F-F_Research_Data_Factors_daily_202411.csv. Its SHA-256 hash is 8a42ddaa2bfd9fcb71a6b4daac8c3aebd98ebcb681cfd683ce74c193340e91d9. Complete calendar years 1927-2023 yield 25,499 daily observations. The source file is public but is not redistributed; Online Resource 1 provides the exact hash, transformation code, and all non-proprietary derived outputs.

A.2 Measurement definitions

Let r_t denote the daily simple market return and l_t = log(1 + r_t) the daily log return. The manuscript reports four complementary measurements rather than treating volatility as one scalar object.

Horizon-specific volatility. Within each trailing ten-calendar-year window, nonoverlapping h-day log returns are formed for h in {1, 5, 20}. Their sample standard deviation is annualized by sqrt(252/h).

Calendar-year volatility. Calendar-year log returns are formed within each trailing ten-year window; their standard deviation measures long-horizon return dispersion.

Stress volatility. The daily series of twenty-one-session rolling standard deviations of log returns is annualized by sqrt(252). The median describes typical conditions and the 99th percentile describes severe rolling stress.

Ex ante standardized returns. Each daily log return is divided by the lagged sixty-three-session standard deviation, using information available before day t. This permits a tail-frequency comparison that is not mechanically driven by contemporaneous volatility.

Tail concentration. Days are ranked by absolute simple return, while their contribution to variation is measured by squared log return. The top-q share is the cumulative squared variation generated by the most extreme q percent of days.

Variance ratios. VR(h) = Var(sum_{j=1}^h l_{t+j}) / [h Var(l_t)] using nonoverlapping blocks. Values below one imply that a portion of daily movement cancels over the aggregation horizon.

A.3 Postwar horizon gradient

Horizon	1955-1964	2014-2023	Endpoint change	Log trend / decade	HAC p-value
1 day	10.60%	18.25%	+72.1%	+12.7%	2.54e-11
5 days	11.79%	17.50%	+48.4%	+6.4%	0.00151
20 days	12.08%	16.47%	+36.3%	+3.6%	0.0742
Calendar year	15.30%	15.68%	+2.5%	+1.6%	0.604

The log-linear trend uses overlapping ten-year windows and Newey-West covariance estimates with nine lags. The overlapping-window inference is descriptive; the endpoint comparison and horizon ordering are more important than a literal constant trend. The daily increase is pronounced, the five-day increase is smaller, the twenty-day trend is marginal, and the annual-return trend is statistically uninformative.

A.4 Period comparisons: typical versus stress volatility

Period	Days	Pooled vol.	Median annual RV	Median 21-day vol.	99th pct. 21-day vol.
1927-1949	6,793	21.5%	16.1%	13.1%	62.5%
1950-1974	6,350	11.1%	9.9%	8.4%	29.7%
1975-1999	6,319	13.9%	12.1%	10.9%	32.5%
2000-2023	6,037	19.9%	15.6%	14.4%	67.8%
1984-2003	5,048	16.4%	13.9%	11.9%	40.6%
2004-2023	5,033	19.4%	13.9%	13.1%	73.5%

Period	Days |r|>2.5%	Variance share from those days	Top 1% variance share	|z|>4	z<-4	Max drawdown
1927-1949	6.42%	64.1%	29.2%	0.72%	0.57%	83.9%
1950-1974	0.80%	19.4%	21.8%	0.28%	0.19%	48.4%
1975-1999	1.09%	27.3%	26.7%	0.30%	0.25%	33.2%
2000-2023	4.84%	50.5%	25.4%	0.45%	0.40%	54.7%

Period	Days |r|>2.5%	Variance share from those days	Top 1% variance share	|z|>4	z<-4	Max drawdown
1984-2003	2.34%	37.0%	26.6%	0.36%	0.34%	49.6%
2004-2023	4.43%	52.6%	28.8%	0.48%	0.42%	54.7%

The 1984-2003 and 2004-2023 comparison illustrates the central distinction. Median annual realized volatility is nearly unchanged, while the 99th percentile of rolling twenty-one-day volatility rises from 40.6 to 73.5 percent. The frequency of absolute daily moves above 2.5 percent rises from 2.34 to 4.43 percent, and those days account for 37.0 versus 52.6 percent of squared variation.

A.5 Tail concentration and variance ratios

Most extreme share of days	Number of days	Share of squared variation
0.1%	26	9.5%
0.25%	64	15.7%
0.5%	128	22.3%
1%	255	30.3%
2.5%	638	44.1%
5%	1,275	56.7%
10%	2,550	70.3%

Period	VR(5)	VR(20)	VR(63)
1955-1964	1.235	1.292	1.614
2014-2023	0.919	0.812	1.034
1984-2003	1.115	1.067	0.979
2004-2023	0.769	0.893	0.630

Between 1955-1964 and 2014-2023, VR(5) falls from 1.235 to 0.919 and VR(20) from 1.292 to 0.812. This is consistent with a larger fraction of daily movement cancelling at short multi-day horizons. It is not a structural identification result: non-synchronous trading, composition changes, and changing autocorrelation can also affect variance ratios.

A.6 Robustness limits and interpretation discipline

The historical findings should be read as stylized facts. The analysis does not identify algorithmic trading as the cause, distinguish public-news jumps from order-book overshoot, or hold constant the composition of the listed market. A later Fama/French data vintage may revise historical returns. The manuscript therefore uses the historical section to motivate a mechanism and predeclared microstructure test, not to claim causal proof of the policy premise.

Appendix B. Analytical proof sketches

This appendix supplies compact derivations for the propositions in Section 5. They establish logical conditions and comparative statics; they are not a complete general-equilibrium welfare theorem.

B.1 Proposition 1: temporal compression under convex impact

$$I(x) = k \, \mathrm{sign}(x) \, |x|^{\beta}, \quad \beta \geq 1$$

Suppose total directional demand Q is divided into N equal increments and depth D replenishes between increments. The absolute impact of each increment is $k|Q/(ND)|^{\beta}$. Summing over increments gives

$$N k \, |Q/(ND)|^{\beta} = k \, |Q/D|^{\beta} N^{(1-\beta)}.$$

The derivative with respect to N is non-positive for $\beta \geq 1$ and strictly negative for $\beta > 1$. Hence concentrating the same total flow into fewer separated increments weakly raises cumulative transient impact. At a fixed sampling frequency, the sum of squared returns also rises when a given terminal adjustment is concentrated into fewer observations.

B.2 Proposition 2: speed-depth mismatch

Let a denote the arrival rate of aggressive demand and g the replenishment rate of durable depth. Peak stress is $X = A/D$. A reduction in reaction latency raises peak X whenever the proportional increase in accumulated flow over the stress interval exceeds the proportional increase in available depth. Endogenous withdrawal makes the condition easier to satisfy because D falls as recent price changes, flow toxicity, or common-direction movement rise. The proposition is conditional: faster information is not destabilizing while durable replenishment keeps pace.

B.3 Proposition 3: state-contingent marginal damage

Let G(X) be expected external damage from temporary displacement, cascades, inventory losses, and propagation, with G increasing and convex above a normal-state region. A marginal charge proportional to G'(X) concentrates distortion where marginal damage is high. For a fixed expected-damage reduction, a constant fee necessarily charges some low-X trades with low marginal external cost and may undercharge high-X trades. When normal trading has positive surplus, the state-contingent schedule attains the target with weakly lower average deadweight loss under standard single-crossing demand responses.

B.4 Proposition 4: stabilizing condition

$$\mathrm{Var}(\Delta p) \approx \mathrm{Var}(Q) / D^2.$$

Taking a local log derivative with respect to fee intensity τ gives

$$d \log \mathrm{Var}(\Delta p)/d\tau = d \log \mathrm{Var}(Q)/d\tau - 2 \, d \log D/d\tau.$$

Both right-side derivatives can be negative. Volatility falls only when the proportional reduction in destabilizing flow variance exceeds approximately twice the adverse proportional reduction in durable depth. A durability-weighted rebate improves this condition by making the depth response less negative or positive. This is an empirical restriction, not an assumption.

B.5 Proposition 5: controller stability

$$x_{t+1} = a x_t + b - e \, \tau(x_t).$$

Linearizing around an equilibrium x_0 yields local slope $a - e \, \tau_x(x_0)$. The scalar controller is locally stable when $|a - e \, \tau_x(x_0)| < 1$. A fee slope that is too small fails to offset persistence; a slope that is too large overshoots. In the joint flow-depth system, the analogous condition is that the spectral radius of the Jacobian be below one. Smoothing, caps, hysteresis, and rebates widen but do not guarantee the stable region.

Appendix C. Numerical experiment: complete specification and diagnostics

C.1 Purpose and state equations

The experiment asks whether state dependence can matter in a system with feedback flow, convex impact, finite depth replenishment, and a direct fee cost to liquidity. Variables are normalized and the experiment is not calibrated to any particular exchange or time interval. All policy regimes face identical random shocks.

q_d(t) = rho_q q_(t-1) + theta u_t + phi r_(t-1) - omega z_(t-1) + sigma_q epsilon_q(t)

X_t = |q_d(t)| / D_t

tau_t = min[tau_max, alpha max(X_t-c,0)^p],   q_t = q_d(t) exp(-chi tau_t)

z_t = (1-delta_z)z_(t-1) + lambda sign(q_t) |q_t/D_t|^beta

r_t = u_t + z_t - z_(t-1)

D_(t+1) = clip[D_t + gamma(Dbar-D_t) - delta_q|q_t| - delta_r|r_t| - eta tau_t + xi rho tau_t|q_t| + sigma_D epsilon_D(t)]

The flat fee is selected by grid search to match the adaptive regime's expected charge E[tau_t |q_t|] within one percent. The adaptive rule is active only above c. The rebate term is present only in the adaptive-plus-rebate regime.

C.2 Baseline parameterization

Parameter	Value	Interpretation
T	4200	Intervals per simulated path
burn	700	Burn-in intervals
n_paths	240	Independent paths
seed	8417	Random seed
rho_q	0.42	Persistence of aggressive flow
theta_news	18	Flow response to fundamental news
phi_mom	2.6	Short-horizon feedback strength
omega_dev	2	Stabilizing response to transitory mispricing
sigma_q	0.05	Flow-shock standard deviation
q_cap	1.65	Absolute flow cap
sigma_news	0.0032	Gaussian news standard deviation
jump_prob	0.002	Rare-jump probability per interval
jump_scale	0.013	Rare-jump scale
jump_df	3.5	Student-t jump degrees of freedom
impact_lambda	0.009	Impact scale
impact_beta	1.75	Impact convexity
impact_cap	0.08	Absolute impact cap
dev_decay	0.34	Transitory displacement decay
Dbar	1	Normal durable depth
depth_recovery	0.09	Depth replenishment rate
flow_depletion	0.075	Depth loss per unit flow
vol_withdrawal	1.35	Depth withdrawal with absolute return
fee_depth_cost	0.02	Direct participation cost of the fee
depth_noise	0.009	Depth-shock standard deviation
Dmin	0.18	Minimum depth
Dmax	1.55	Maximum depth
chi	2.2	Flow elasticity to normalized friction
c	0.16	Activation threshold
p	2	Fee convexity
alpha	5	Fee slope/scale
tau_max	0.28	Fee-intensity cap
rebate_share	0.75	Revenue share eligible for rebate
rebate_efficiency	0.5	Conversion of rebates into durable depth
equal_revenue_flat_fee	0.005	Matched constant fee intensity

C.3 Simulation protocol

1. Generate one common library of Gaussian news, rare Student-t jumps, flow shocks, and depth shocks from the fixed seed.

2. Run the no-intervention, adaptive, and adaptive-plus-rebate regimes against the identical library.

3. Find the constant flat fee whose expected charge matches the adaptive regime within one percent; rerun it against the same library.
4. Retain 840,000 post-burn-in observations across 240 paths.
5. Compute absolute-return expected shortfall at 99.5 and 99.9 percent, root-mean-square transitory pricing error, depth-shortfall probabilities, ordinary-state flow reduction, and fee incidence.
6. Run a separate common-random-number grid over adaptive slope alpha and direct fee-induced depth cost eta to map the failure region.
7. Run a deterministic impulse response with one negative fundamental shock and no random noise.

C.4 Baseline policy metrics

Policy	99.9% tail ES	RMS pricing error	P(depth<0.8)	Ordinary flow reduction	Fee active
No intervention	2.616%	5.25 bp	0.140%	0.000%	0.0%
Equal-revenue flat fee	2.609%	5.10 bp	0.136%	1.094%	100.0%
Adaptive fee	2.427%	3.77 bp	0.055%	0.000%	10.0%
Adaptive fee + rebate	2.427%	3.75 bp	0.012%	0.000%	9.9%

Relative to no intervention, the adaptive fee reduces 99.9 percent absolute-return expected shortfall by 7.23 percent and RMS transitory pricing error by 28.25 percent. The equal-charge flat fee reduces the same metrics by 0.30 and 2.92 percent while reducing ordinary-state flow by 1.09 percent because it is always active. The rebate changes the tail metric only marginally in the baseline but reduces the probability of depth below 80 percent of normal by 91.33 percent relative to no intervention.

C.5 Deterministic shock response

Policy	Minimum transitory error	Minimum durable depth	Maximum fee intensity
No intervention	-24.1 bp	0.915	0.000
Equal-revenue flat fee	-23.5 bp	0.915	0.005
Adaptive fee + rebate	-9.2 bp	0.956	0.280

C.6 Stability map and explicit failure region

Direct depth cost eta	Adaptive slope alpha	Change in 99.5% tail ES	Mean depth	P(depth<0.8)
0	5	-4.2%	0.904	0.011%
0.4	20	-4.9%	0.881	5.949%
0.8	20	+0.4%	0.781	30.774%
1.2	10	+1.3%	0.781	28.532%
1.6	20	+20.6%	0.469	76.174%
2.4	80	+31.6%	0.272	97.182%

Negative expected-shortfall changes denote improvement. When the direct participation cost is low, steeper adaptive response improves the tail metric across the displayed range. When liquidity is highly fee-sensitive, the same response thins depth enough to reverse the benefit. At eta=2.4 and alpha=80, the 99.5 percent tail expected shortfall is 31.6 percent worse than no intervention. The experiment therefore reports a genuine failure region rather than assuming the policy works.

C.7 What the experiment does and does not establish

The model establishes internal coherence and comparative mechanisms under explicit assumptions. It does not estimate real demand elasticity, welfare weights, cross-venue substitution, adverse-selection responses, or an optimal fee schedule. The parameter values should not be interpreted as basis-point recommendations. The second-stage empirical study must estimate durability and elasticities before any pilot can be designed.

Appendix D. Predeclared message-level validation protocol

D.1 Primary empirical question

Conditional on the same displayed depth, aggressive flow, spread, prevailing volatility, volume, and conventional order-book imbalance, does a book with a higher ex ante probability of cancellation before execution exhibit larger and more transitory price dislocations? The theory is supported only if survival-adjusted durable depth adds out-of-sample information beyond standard book variables and the incremental effect is stronger for subsequent reversal than for permanent price discovery.

D.2 Required data and institutional scope

The required data must preserve individual order lifecycles: submission, modification, partial cancellation, deletion, partial execution, full execution, side, price, quantity, exchange timestamp, sequence number, and persistent order identifier. Limited-depth snapshots are insufficient when orders can enter the retained price range after being submitted outside it, because true age and survival history are then missing.

The preferred design uses one order-driven venue as the discovery sample, a later period and additional securities as a locked confirmation sample, and a second exchange architecture for external replication. Consolidated best-quote and marketwide controls are required because local orders can cancel in response to superior away-market liquidity. Auctions, halts, corrections, and crossed markets are excluded from the primary continuous-trading specification and analyzed separately.

D.3 Sample stratification and pre-registration

Before opening the confirmation sample, freeze the security universe, date ranges, horizons, price bands, event exclusions, predictor set, outcome definitions, fixed effects, matching variables, and primary evaluation metrics. The discovery universe should span large-, mid-, and small-cap securities; high and low volume; large and small relative tick size; ordinary stocks and heavily traded exchange-traded funds; calm, ordinary, and stressed market regimes.

D.4 Order-level competing-risk model

$$d_{jt}(H) = 1 - \Pr(\text{order } j \text{ cancels unexecuted before } t+H \mid F_t).$$

For each resting order and horizon H, estimate competing outcomes: cancellation/deletion before execution, execution, or continued survival. Partial cancellation is handled at the share level. An order that executes is successful liquidity and is not discounted. The primary model should be interpretable, such as a discrete-time multinomial hazard or cause-specific competing-risk model. Gradient boosting or neural survival models are robustness checks rather than the sole specification.

Features available at t may include order age, remaining size, price distance, queue position, volume ahead, spread, multi-level imbalance, recent aggressive flow, recent cancellations and additions, partial-execution history, midpoint direction, volatility, time of day, stock liquidity, and marketwide state. Predictions must be generated strictly out of sample by rolling prior-year estimation or repeated cross-fitting. Generated-regressor uncertainty should be propagated with stock-day blocks or repeated cross-fit folds.

D.5 Durable depth and directional stress

$$D_{dur}^{s}(t,H,b) = \text{sum over orders } j \text{ in side } s \text{ and band } b \text{ of } q_{jt}\, w(\text{distance}_{jt})\, d_{jt}(H).$$

$$X_{sell}(t) = \text{aggressive sell flow / durable bid depth}; \quad X_{buy}(t) = \text{aggressive buy flow / durable ask depth}.$$

Primary specifications use fixed tick or basis-point bands and report distance-unweighted robustness. The main benchmark replaces durable depth with raw displayed depth. Other benchmarks include top-of-book and multi-level depth, queue imbalance, order-flow imbalance, book slope, cancellation intensity, order age, spread, volume, and recent volatility.

D.6 Predeclared hypotheses

H1 Survival discount. Displayed depth overstates short-horizon absorption capacity, and the discount becomes larger during high volatility, synchronized movement, and intense same-side cancellations.

H2 Incremental impact. Conditional on the same displayed depth and aggressive flow, lower durable depth predicts larger immediate directional price impact beyond conventional controls.

H3 Convexity. Impact and depth withdrawal rise disproportionately once aggressive flow approaches durable depth; a spline in X outperforms a purely linear specification.

H4 Tail prediction. Flow divided by durable depth predicts rare short-horizon moves out of sample better than displayed-depth ratios, raw order-flow imbalance, spread, volatility, and volume.

H5 Transitory displacement. The incremental predictive power is stronger for peak displacement that subsequently reverses than for the permanent price component after news controls.

H6 Systemic state dependence. The relation strengthens when many securities move in the same direction, cross-asset correlation rises, or linked futures and ETFs transmit common flow.

D.7 Outcomes and primary models

Tail outcomes should be a family rather than one arbitrary threshold: stock-specific 99.9th-percentile midpoint returns; moves above four or five lagged local standard deviations; unusually large impact per unit executed volume; rapid depth collapse; volatility-auction or limit events; and simultaneous jumps across linked instruments. Rare-event evaluation emphasizes precision-recall, log loss, Brier score, calibration in the upper risk bins, and incremental likelihood.

$$\textit{Peak impact} = \max \text{ over } 0<u<=H1 \text{ of } |m_{(t+u)}-m_t|.$$

$$\textit{Permanent impact} = |m_{(t+H2)}-m_t|, \quad H2 > H1.$$

$$\textit{Transitory impact} = \textit{Peak impact} - \textit{Permanent impact}.$$

The primary panel model relates future directional midpoint change or transitory impact to a flexible function of X, with security, date, and intraday-bin fixed effects and controls listed above. Standard errors are clustered at a level that captures both security-day dependence and common market intervals. Quantile or tail regressions and rare-event classifiers supplement the mean model.

D.8 Matched-event and scheduled-news designs

For communication and identification, match observations within the same security and time of day on displayed depth, aggressive flow, spread, recent volatility, volume, market return, and conventional imbalance. Compare high- and low-durability books. The concrete question is whether two apparently equally deep books facing equal sell flow experience different peak and reversal profiles because one book is more likely to cancel.

Scheduled macroeconomic announcements provide a second setting. Control for announcement magnitude and sign, measure the initial response, later persistent response, overshoot, spread, cancellation, and replenishment. Durable-depth stress should predict the excess initial movement beyond the later consensus response, not merely the direction of the news.

D.9 Locked confirmation and external replication

All transformations and primary tests are frozen before opening the locked sample. A substantial discovery requires: economically material impact conditional on identical displayed state; stronger effects in the upper X tail; incremental out-of-sample performance; disproportionate explanatory power for reversal; and replication in a later period, additional securities, and a second venue. A coefficient that is statistically significant only because of billions of observations is insufficient.

D.10 Falsification criteria

1. Survival-adjusted depth adds no out-of-sample information beyond standard order-book variables.
2. The relation is linear and no stronger in high-stress states.
3. Durable-depth stress predicts permanent information-driven movement as strongly as temporary reversal.
4. Opposite-side durable depth has the same directional effect as same-side capacity.
5. Results disappear outside one venue, one year, or a narrow group of highly liquid securities.
6. Predictive improvement requires future cancellation information rather than ex ante probabilities.
7. Estimated demand and depth elasticities imply that any plausible surcharge violates the stabilizing condition in Proposition 4.

D.11 Data engineering and audit trail

The order-book engine should validate feed sequence continuity, nonnegative order sizes, cancellation quantities not exceeding outstanding size, execution links to live orders, reconstructed best quotes against vendor snapshots, and reconciliation of venue trades with official totals. Features should be computed in a streaming process rather than storing a full-depth snapshot after every event. The final archive should contain code, field definitions, exclusions, checksums, frozen configuration files, and synthetic or vendor-permitted test fixtures, without redistributing licensed raw data.

D.12 Interpretation of a successful second-stage result

A successful result would not by itself prove that the proposed fee is optimal. It would establish an empirically measurable externality: marginal aggressive flow has disproportionately large transitory impact when displayed liquidity is statistically unlikely to remain executable. The estimated external-cost curve and behavioral elasticities could then discipline a structural or agent-based counterfactual and a controlled market-design pilot.

Appendix E. Reproducibility archive and file map

Online Resource 1 is a ZIP archive containing portable Python scripts, fixed seeds, derived public outputs, the exact source-file hash, and all manuscript figures. The raw Fama/French source data are not included.

Path	Content
README.md	Environment, public input instructions, exact hash, and reproduction commands
DATA_VINTAGE_AND_HASH.txt	Source vintage, filename, hash, analysis interval, and observation count
requirements.txt	Python dependencies
run_all.py	One-command orchestration
src/historical_analysis.py	Historical volatility, tail concentration, variance ratios, and HAC trends
src/simulate_model.py	Common-random-number policy experiment, impulse response, and stability grid
src/make_figures.py	Regeneration of all eight figures
outputs/history/*.csv	Derived historical tables
outputs/simulation/*.csv	Derived numerical tables and diagnostics
figures/*.png	600-dpi manuscript figures

E.1 Reproduction commands

python run_all.py --fama-french-file /path/to/F-F_Research_Data_Factors_daily_202411.csv

The scripts regenerate all outputs from the supplied public input and fixed random seed. On the tested environment, the numerical CSVs reproduce exactly. Later revisions to the Fama/French source history may create small differences unless the archived November 2024 vintage is used.

E.2 Online Resource captions

Online Resource 1. Reproducibility archive containing Python code, fixed random seeds, data-vintage and checksum documentation, non-proprietary derived tables, and all manuscript figures.

Online Resource 2. Supplementary appendices containing detailed historical methods and results, analytical proof sketches, the complete numerical specification and diagnostics, and the predeclared message-level validation protocol.