Distributional Jump Bridges for Disclosure Response Forecasts

Aryan Gupta
aryan.cs.app@gmail.com

May 2026

Abstract

Financial disclosure models often combine event text with market history and predict every market-response target through one shared path. This paper proposes Distributional Jump Bridges (DJB), a neural event architecture that first estimates a no-event response distribution from pre-disclosure market state, then uses SEC filing text to transport the event response through bounded shifts in target means and predictive log-variances. A return-conservative variant (RC-DJB) adds a return-mean constraint, providing a controlled test of whether disclosure text should directly move near-term abnormal-return point forecasts. In a leakage-controlled public-data study with 7,236 SEC 8-K events, 7,236 matched controls, and 2,463 held-out rows from 98 large U.S. firms, DJB achieves the lowest multi-target MSE, 0.0552 with 95% ticker-clustered interval [0.0396, 0.0773], after aligning firm and market return windows. Its abnormal-return Rank IC has the largest point estimate, 0.0330, although the ticker-clustered interval includes zero [-0.0042, 0.0710]. Paired bootstrap comparisons show DJB improves MSE over concat fusion by 0.0114 [0.0092, 0.0135] and over RC-DJB by 0.0010 [0.0005, 0.0015]. Leave-one-ticker-out Rank IC remains positive for every held-out ticker deletion, with minimum 0.0263. RC-DJB bridge interventions raise event-row MSE by 0.0183, 0.0037, and 0.0025 when the bridge is removed, text is zeroed, or text is shuffled. The results support a response-transport modeling principle: in this experiment, disclosure text contributes more clearly to event risk, liquidity, and uncertainty than to cluster-significant return ranking.

1 Introduction

Public disclosures are timed interventions in the market information set [Fama, 1970, Beaver, 1968, Patell and Wolfson, 1984]. A filing can change volatility, liquidity, and uncertainty even when its effect on cross-sectional abnormal-return ranking is weak or unstable [Bamber, 1986, Kothari, 2001]. Neural sequence and multimodal forecasters often fuse histories, event information, or text before predicting market targets [Vaswani et al., 2017, Lim et al., 2021, Zhang et al., 2025]. This design can ask the same text representation to explain risk response, volume response, uncertainty, and return ranking. In financial data, that coupling can be problematic: return labels are noisy, text is high dimensional, and small leakage, data snooping, or backtest overfitting can be mistaken for predictive structure [Campbell et al., 1997, Gu et al., 2020, White, 2000, Bailey et al., 2016, López de Prado, 2018].

This paper studies a stricter architectural question: should disclosure text enter an event model as an unconstrained predictor of every outcome, or as a distributional transport operator layered on top of ordinary market dynamics? Distributional Jump Bridges (DJB) first predict a no-event Gaussian response distribution from pre-event price, volume, and market-relative features. They then use disclosure text to generate bounded shifts in target means and predictive log-variances. The return-conservative variant (RC-DJB) adds the sharper restriction $\Delta\mu_{return} = 0$, so text may move volatility, volume, and uncertainty but cannot directly overwrite the abnormal-return point forecast.

The contributions are:

1. A distributional bridge architecture for event-driven financial response modeling.

2. A leakage-controlled SEC 8-K experiment with matched no-event controls and public post-filing market outcomes.

3. An evaluation protocol that aligns firm and market return windows and reports both ticker-clustered and leave-one-ticker stability audits for Rank IC.

4. Evidence that DJB improves response-regression error over fusion and bottleneck baselines, with RC-DJB intervention tests isolating what the return-mean constraint changes.

2 Related Work

Classical event studies estimate abnormal returns around information releases [Fama et al., 1969, Brown and Warner, 1985, MacKinlay, 1997]. Disclosure and earnings-announcement studies document price, volatility, and trading-volume responses after public information releases [Beaver, 1968, Patell and Wolfson, 1984, Bamber, 1986, Boehmer et al., 1991, Kothari and Warner, 2007]. DJB keeps this information boundary but replaces scalar event-window estimation with a neural distributional response model; RC-DJB is the return-conservative constrained variant. The abnormal-return construction follows the market-model and factor-model tradition [Sharpe, 1964, Fama and French, 1993, Carhart, 1997, Kothari and Warner, 2007]; this study uses a SPY-adjusted public-data variant rather than CRSP/factor residuals.

The architecture is related to recurrent and state-space sequence models, including LSTMs and gated recurrent units [Hochreiter and Schmidhuber, 1997, Cho et al., 2014], structured state-space sequence models [Gu et al., 2022], and selective state-space models [Gu and Dao, 2023, Wang et al., 2025]. It also draws on probabilistic forecasting and time-series foundation model work [Dawid, 1984, Gneiting and Raftery, 2007, Salinas et al., 2020, Lim et al., 2021, Nie et al., 2023, Ansari et al., 2024, Goswami et al., 2024].

Financial text contains information relevant to prices and fundamentals [Tetlock, 2007, Tetlock et al., 2008, Kogan et al., 2009, Loughran and McDonald, 2011, Cohen et al., 2020]. Recent financial NLP systems and benchmarks study domain language models, multimodal financial forecasting, SEC filing reasoning, and leakage in financial agents [Jegadeesh and Wu, 2013, Araci, 2019, Wu et al., 2023, Yang et al., 2023, Zhang et al., 2025, Li et al., 2025]. RC-DJB is also adjacent to counterfactual prediction, although this paper does not estimate structural causal effects [Rubin, 1974]. DJB differs by routing event text through an explicit residual distributional transport module; RC-DJB further imposes a target-specific restriction in which disclosure text may transport response distribution components while the abnormal-return mean remains no-event anchored.

3 Method

Throughout, DJB denotes the unconstrained distributional bridge, and RC-DJB denotes the return-conservative ablation. Both models share the same no-event dynamics and disclosure bridge. DJB allows text to shift all three target means; RC-DJB keeps the abnormal-return mean fixed and uses text only for volatility, volume, and log-variance transport.

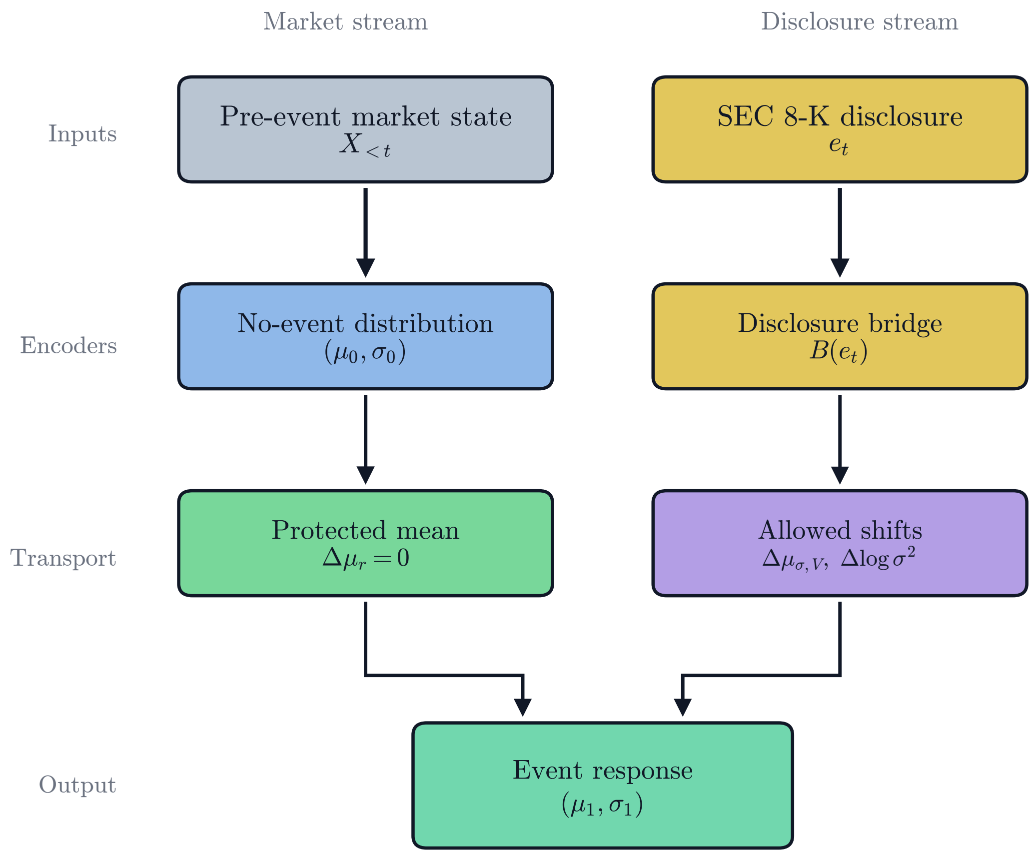

Figure 1: Distributional Jump Bridge architecture. The model first estimates a no-event response distribution from market history; disclosure text then generates bounded distributional shifts. The return-conservative variant enforces $\Delta\mu_r = 0$.

Inputs and targets. Each sample i has label start date τ_i, a 40-trading-day pre-event tensor $X_i \in \mathbb{R}^{40\times 8}$, disclosure embedding e_i, metadata $m_i \in \mathbb{R}^6$, and target vector $y_i \in \mathbb{R}^3$. The targets are five-trading-day SPY-adjusted return, realized-volatility jump, and volume jump. Text-aware variants use chunk-pooled BGE-small disclosure embeddings, following the dense sentence and passage embedding literature [Reimers and Gurevych, 2019, Karpukhin et al., 2020, Xiao et al., 2023].

Event-study labels. Let $P_{i,t}$, $V_{i,t}$, and $r_{i,t}$ denote adjusted close, volume, and daily close return for firm i, and let r_t^M denote the SPY return. The label window $\mathcal{W}_i = \{\tau_i, \ldots, \eta_i\}$ starts on the disclosure date for before-close traded-day filings and on the next traded date otherwise; η_i is the fifth tradable session in the window. Let $\mathcal{R}_i = \mathcal{W}_i \setminus \{\tau_i\}$ denote the close-to-close return dates strictly after the label-start close. The pre-window $\mathcal{P}_i$ is the 20 traded sessions strictly before τ_i. These choices follow the event-study convention that features must precede the public event time while outcomes are measured only after the event window begins [Brown and Warner, 1985, MacKinlay, 1997, Patell and Wolfson, 1984, Boehmer et al., 1991]. Realized-volatility differences are reported as a simple daily-data analogue of the realized-variation literature [Andersen et al., 2003]. The implemented targets are

$$y_{i,r} = \frac{P_{i,\eta_i}}{P_{i,\tau_i}} - 1 - \left(\prod_{t\in\mathcal{R}_i}(1 + r_t^M) - 1\right), \tag{1}$$

$$y_{i,\sigma} = \mathrm{sd}_{t\in\mathcal{W}_i}(r_{i,t}) - \mathrm{sd}_{t\in\mathcal{P}_i}(r_{i,t}), \tag{2}$$

$$y_{i,V} = \frac{|\mathcal{P}_i|}{|\mathcal{W}_i|}\frac{\sum_{t\in\mathcal{W}_i} V_{i,t}}{\sum_{t\in\mathcal{P}_i} V_{i,t}} - 1. \tag{3}$$

The three targets are deliberately not identical event-study estimands. The return target is a conservative forward-drift label anchored at the label-start close; for a filing accepted before the market close, it therefore does not try to identify the intraday filing-day price jump from daily bars. A previous-close denominator would capture more immediate repricing, but with daily data it also mixes pre-filing intraday movement into the disclosure response. Volatility and volume are different response-distribution targets, so the filing-day session is included whenever the filing is public before the close. This convention tests whether disclosure text helps transport risk/liquidity response while leaving cross-sectional return ranking to the no-event state. Rows are admissible only when their maximum feature date is strictly before the label start, $\max\{t : t \in X_i\} < \tau_i$.

No-event distribution. A recurrent encoder maps pre-event market history to a latent state [Cho et al., 2014]:

$$h_i = \mathrm{GRU}_\theta(X_i) + W_m m_i. \tag{4}$$

A bounded gate produces the no-event state $\tilde{s}_i$, from which the model predicts a diagonal Gaussian distribution, a standard proper-scoring-rule choice for probabilistic forecasts [Gneiting and Raftery, 2007, Salinas et al., 2020]:

$$\tilde{s}_i = h_i \odot (1 + 0.1\tanh(g_\theta(h_i, m_i))), \tag{5}$$

$$[\mu_{0,i}, \ell_{0,i}] = f_\theta(\tilde{s}_i, m_i), \tag{6}$$

where $\ell_{0,i}$ is log-variance.

Distributional bridge. Disclosure text generates a bounded latent state jump and distributional transport:

$$b_i = B_\theta(e_i), \tag{7}$$

$$j_i = \alpha\tanh(u_\theta(b_i)) \odot \sigma(v_\theta(b_i)), \tag{8}$$

$$[\Delta\mu_i, \Delta\ell_i] = T_\theta(\mathrm{LN}(\tilde{s}_i + j_i), b_i, m_i). \tag{9}$$

Here LN denotes layer normalization [Ba et al., 2016]. The return-conservative architectural constraint is $\Delta\mu_{i,return} = 0$. For RC-DJB,

$$\mu_{1,i} = \mu_{0,i} + [0, \Delta\mu_{i,volatility}, \Delta\mu_{i,volume}], \qquad \ell_{1,i} = \mathrm{clip}(\ell_{0,i} + \Delta\ell_i). \tag{10}$$

DJB removes the zero-return-mean restriction and uses the full $\mu_{1,i} = \mu_{0,i} + \Delta\mu_i$. Both variants predict $p_\theta(y_i) = \mathcal{N}(\mu_{1,i}, \mathrm{diag}(\exp(\ell_{1,i})))$.

Objective. For target dimension k, the Gaussian negative log likelihood used for probabilistic training and evaluation is a proper likelihood-based scoring objective [Gneiting and Raftery, 2007]:

$$\mathcal{L}_{NLL} = \frac{1}{2NK} \sum_{i=1}^{N} \sum_{k=1}^{K} \left[\exp(-\ell_{1,ik})(y_{ik} - \mu_{1,ik})^2 + \ell_{1,ik} \right], \qquad K = 3, \quad \ell_{1,ik} \in [-7, 3]. \tag{11}$$

The pairwise rank regularizer is computed only over real-disclosure pairs in a minibatch, following standard learning-to-rank objectives [Burges et al., 2005, Cao et al., 2007]:

$$\mathcal{L}_{rank} = \frac{1}{|Q|} \sum_{(i,j)\in Q} \log \left(1 + \exp \left[-\operatorname{sign}(y_{i,r} - y_{j,r}) \frac{\mu_{1,i,r} - \mu_{1,j,r}}{T} \right] \right), \tag{12}$$

where Q contains event pairs with unequal abnormal-return labels and $T = 0.02$. The supervised term is the weighted Smooth-L1 loss after optional target standardization [Huber, 1964]:

$$\mathcal{L}_{Huber} = \frac{1}{NK} \sum_{i=1}^{N} \sum_{k=1}^{K} w_k \, \mathrm{SmoothL1}_{\beta} \left(\frac{\mu_{1,ik} - \bar{y}_k}{s_k}, \frac{y_{ik} - \bar{y}_k}{s_k} \right). \tag{13}$$

For bottleneck models with posterior parameters (μ_i^z, ℓ_i^z), the KL penalty follows variational and information-bottleneck regularization [Kingma and Welling, 2014, Tishby et al., 2000],

$$\mathcal{L}_{KL} = -\frac{1}{2} \operatorname{mean}_{i,k} \left(1 + \ell_{ik}^z - (\mu_{ik}^z)^2 - \exp(\ell_{ik}^z) \right). \tag{14}$$

The control and consistency penalties are

$$\mathcal{L}_{control} = \frac{\sum_i \mathbb{1}\{a_i = 0\} \|\Delta\mu_i\|_2^2}{\sum_i \mathbb{1}\{a_i = 0\}}, \tag{15}$$

$$\mathcal{L}_{consistency} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{Var}_i(\Delta\mu_{i,k}), \tag{16}$$

where $a_i = 1$ for real disclosures and $a_i = 0$ for matched controls. The full training loss combines target-standardized Huber regression, Gaussian likelihood, control suppression, bottleneck regularization, bridge sparsity, residual consistency, and rank learning:

$$\mathcal{L} = \lambda_y \mathcal{L}_{Huber} + \lambda_n \mathcal{L}_{NLL} + \lambda_c \mathcal{L}_{control} + \lambda_{KL} \mathcal{L}_{KL} + \lambda_s \|z_i\|_1 + \lambda_v \mathcal{L}_{consistency} + \lambda_r \mathcal{L}_{rank}. \tag{17}$$

Here z_i is the stored disclosure-bridge latent used for sparsity diagnostics. The main run uses $w_y = (3.0, 1.0, 0.25)$, $\lambda_n = 0.10$, $\lambda_c = 0.35$, $\lambda_{KL} = 0.0$, $\lambda_s = 0.001$, $\lambda_v = 0.05$, and $\lambda_r = 0.15$.

4 Data and Leakage

The experiment uses public SEC EDGAR APIs [U.S. Securities and Exchange Commission, 2026]. The sample covers 98 usable large U.S. firms, SEC 8-K filings from 2020–2025, and public daily prices. Tickers are mapped to CIKs through the SEC company-tickers file. Filings are read from SEC submissions JSON, including archived submissions, using `acceptanceDateTime` and the primary EDGAR document URL. The parser stores source URL, accession number, accepted timestamp, available timestamp, extracted text, and text hash. Prior filing-text studies motivate using public SEC text as a market-response substrate [Loughran and McDonald, 2011, Jegadeesh and Wu, 2013, Cohen et al., 2020].

Daily prices are downloaded from public price endpoints, with SPY used as the market proxy for abnormal returns. The trading calendar is inferred from observed price rows. Event label windows start on the same observed trading day for filings accepted before 16:00 ET; filings accepted after 16:00 ET or on non-trading days start on the next observed trading day. Features use only prices strictly before the label start date. Labels use future returns only as targets. One deterministic same-ticker no-event control is constructed for each filing, excluding dates within a 10-calendar-day blackout window around known events. The no-lookahead rule and blackout controls are motivated by standard event-study design and known financial backtesting leakage failure modes [MacKinlay, 1997, White, 2000, Hansen, 2005, Bailey et al., 2016, López de Prado, 2018, Li et al., 2025]. Figure 2 visualizes the chronological sample and the feature-to-label guard margin used by every row.

Table 1: Main public-data sample.

Artifact	Count
Usable companies	98
SEC 8-K events	7,236
Matched no-event controls	7,236
Feature rows	14,472
Public price rows	159,093
Held-out rows	2,463

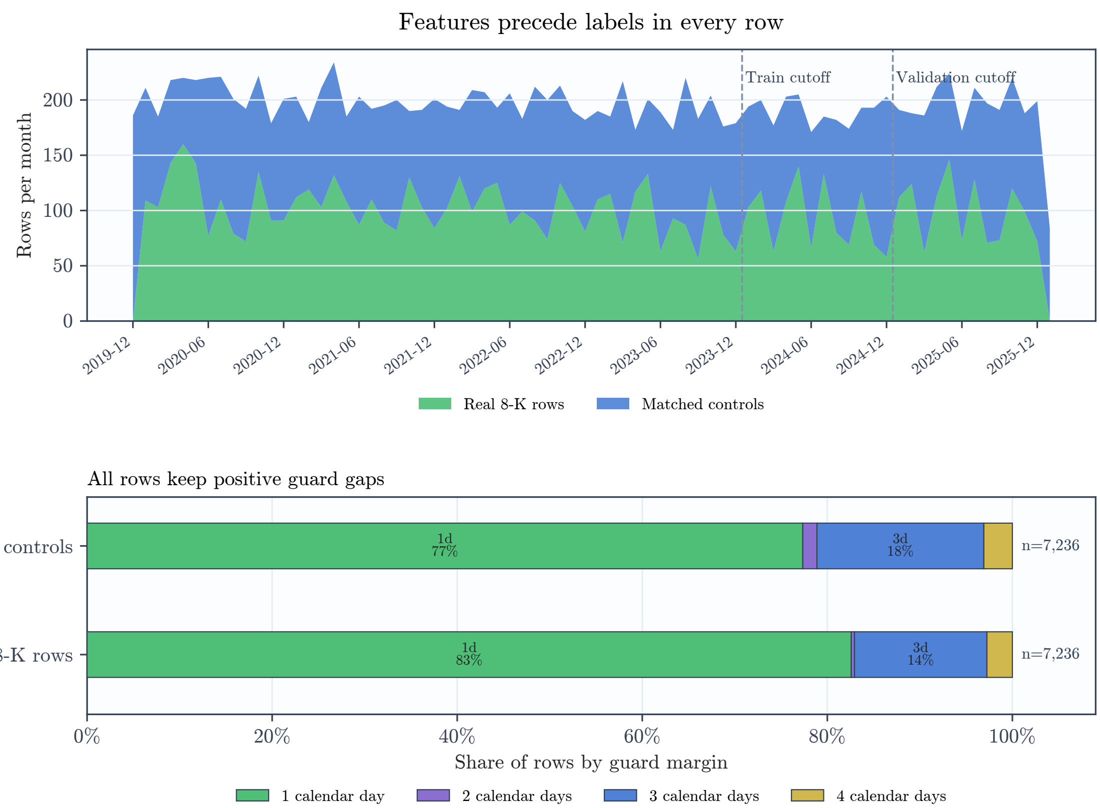

Figure 2: Chronological data substrate and feature-label leakage audit. The upper panel shows monthly real-event and matched-control rows with chronological split boundaries; the lower panel shows the share of rows by positive feature-label guard margin and the zero-violation leakage audit.

5 Evaluation Setup

Rows with $\tau_i \leq$ 2023-12-31 are training rows, rows from 2024 are validation rows, and later rows are held out. This yields 9,729 training rows, 2,280 validation rows, and 2,463 held-out rows across 98 held-out tickers and 13 held-out label-start months. Checkpoints are selected by the declared validation criterion for each run and evaluated once on the held-out split. All models use fixed seed 29. The chronological split follows time-series evaluation practice, where future observations must not influence model selection or feature construction [Bergmeir and Benítez, 2012, Hyndman and Athanasopoulos, 2021, López de Prado, 2018].

Baselines include no-event dynamics, text-only MLP, concat fusion, a stochastic event-bottleneck baseline, a low-rank disclosure-operator baseline, an event-jump state-space baseline with BGE embeddings, and an unconstrained Distributional Jump Bridge (DJB). The No-event baseline is trained on the same chronological event/control rows and event-window labels as the other models; it removes disclosure text rather than changing the training substrate. Confidence intervals use percentile bootstrap resampling; ticker-clustered intervals resample tickers with replacement [Efron, 1979, Davison and Hinkley, 1997, Cameron et al., 2008]. Paired comparisons use row-paired bootstrap resampling. Rank IC is Spearman correlation between predicted and realized abnormal return [Spearman, 1904, Grinold and Kahn, 2000].

Evaluation statistics. For model m, the reported multi-target error is

$$\mathrm{MSE}(m) = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_{ik} - \hat{y}_{ik}^{(m)}\right)^2. \tag{18}$$

Rank IC is the Spearman correlation between predicted and realized abnormal return:

$$\mathrm{IC}(m) = \rho_P\left(\mathrm{rank}(\hat{y}_{1:N,r}^{(m)}), \mathrm{rank}(y_{1:N,r})\right). \tag{19}$$

For a statistic S, bootstrap replicates $S^{*(1)}, \dots, S^{*(B)}$ form percentile intervals

$$\mathrm{CI}_{0.95}(S) = \left[Q_{0.025}\{S^{*(b)}\}_{b=1}^{B}, Q_{0.975}\{S^{*(b)}\}_{b=1}^{B}\right], \tag{20}$$

The model-level intervals in Table 2 use $B = 1000$. Ticker-clustered intervals sample tickers with replacement and include all held-out rows for each sampled ticker. Rank IC uses integer ranks of predicted and realized abnormal returns. Paired MSE comparisons against DJB use $B = 2000$ row-paired bootstrap replicates and

$$\Delta_{\mathrm{MSE}}(m) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{K}\|y_i - \hat{y}_i^{(m)}\|_2^2 - \frac{1}{K}\|y_i - \hat{y}_i^{DJB}\|_2^2\right], \tag{21}$$

with rows resampled as matched pairs. Paired rank comparisons use the same joined rows but report $\mathrm{IC}(DJB) - \mathrm{IC}(m)$.

Probabilistic diagnostics. For nominal level $1-\alpha$, the diagonal Gaussian predictive interval for target k is

$$I_{ik}(1-\alpha) = \left[\mu_{1,ik} - z_{1-\alpha/2}\exp(\ell_{1,ik}/2), \mu_{1,ik} + z_{1-\alpha/2}\exp(\ell_{1,ik}/2)\right], \tag{22}$$

and empirical coverage is

$$\widehat{C}(1-\alpha) = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{1}\{y_{ik} \in I_{ik}(1-\alpha)\}. \tag{23}$$

Mean predictive standard deviation is $\frac{1}{NK}\sum_{i,k}\exp(\ell_{1,ik}/2)$. Interval coverage is reported because calibrated predictive uncertainty is distinct from point forecast error [Winkler, 1972, Gneiting and Raftery, 2007].

Table 2: Held-out results after firm/market return-window alignment. MSE is over abnormal return, volatility jump, and volume jump. Intervals are 95% ticker-clustered bootstrap intervals.

Model	MSE	MSE CI	Rank IC	Rank IC CI
No-event	0.0695	[0.0517, 0.0933]	-0.0393	[-0.0886, 0.0064]
Concat fusion	0.0666	[0.0499, 0.0886]	0.0007	[-0.0422, 0.0389]
Event bottleneck	0.0696	[0.0524, 0.0923]	-0.0295	[-0.0752, 0.0160]
Disclosure operator	0.0698	[0.0526, 0.0926]	-0.0043	[-0.0487, 0.0422]
Event-jump SSM	0.0566	[0.0406, 0.0788]	0.0254	[-0.0107, 0.0633]
DJB	**0.0552**	[0.0396, 0.0773]	**0.0330**	[-0.0042, 0.0710]
RC-DJB	0.0562	[0.0404, 0.0783]	0.0154	[-0.0288, 0.0559]

Table 3: Paired comparisons against DJB. Positive MSE values mean DJB has lower MSE. Positive rank values mean DJB has higher rank IC.

Baseline	MSE Diff.	95% CI	Rank IC Diff.	95% CI
No-event	0.01424	[0.01145, 0.01765]	0.0722	[0.0204, 0.1293]
Concat fusion	0.01140	[0.00919, 0.01353]	0.0323	[-0.0214, 0.0875]
Event bottleneck	0.01443	[0.01182, 0.01727]	0.0625	[0.0144, 0.1151]
Disclosure operator	0.01459	[0.01200, 0.01735]	0.0372	[-0.0257, 0.0993]
Event-jump SSM	0.00136	[0.00085, 0.00188]	0.0076	[-0.0192, 0.0370]
RC-DJB	0.00097	[0.00052, 0.00146]	0.0176	[-0.0174, 0.0557]

Table 4: Rank-temperature sensitivity. LOTO is the minimum leave-one-ticker-out Rank IC over the 98 held-out tickers.

Model	Temperature	MSE	Rank IC	Rank IC CI	LOTO min
DJB	$T = 0.02$	0.0552	0.0330	[-0.0042, 0.0710]	0.0263
DJB	$T = 0.10$	0.0559	0.0152	[-0.0239, 0.0537]	0.0086
RC-DJB	$T = 0.02$	0.0562	0.0154	[-0.0288, 0.0559]	0.0095
RC-DJB	$T = 0.10$	0.0562	0.0201	[-0.0207, 0.0594]	0.0146

Bridge audits. The target-specific response-transport statistic for group $c \in \{0, 1\}$ and target k is

$$T_{c,k} = \text{mean}_{i:a_i=c} \left|\Delta\mu_{i,k}\right|, \tag{24}$$

where $c = 1$ denotes real 8-K rows and $c = 0$ denotes controls. The aggregate transport reported below is $\bar{T}_c = K^{-1} \sum_k T_{c,k}$. Response-channel AUC uses $\frac{1}{2}(|\Delta\mu_{i,\sigma}| + |\Delta\mu_{i,V}|)$ to classify real disclosures against matched controls. Latent bridge AUC uses the score $s_i = \text{mean}_k |z_{i,k}|$, with average ranks used for tied scores [Hanley and McNeil, 1982].

Table 5: Distributional bridge diagnostics on the held-out split. Log-variance Rank IC audits test whether return uncertainty shifts carry signed return-ranking information.

Diagnostic	DJB	RC-DJB
Gaussian NLL	-2.2815	-2.2961
Observed 80% interval coverage	0.8948	0.9012
Observed 95% interval coverage	0.9636	0.9632
Mean predictive standard deviation	0.1351	0.1390
Mean response transport, real 8-Ks	0.0235	0.0832
Mean response transport, controls	0.0075	0.0684
Return-mean transport, real 8-Ks	0.0053	0.0000
Return-mean transport, controls	0.0050	0.0000
Volatility transport, real 8-Ks	0.0160	0.1977
Volatility transport, controls	0.0116	0.1932
Volume transport, real 8-Ks	0.0493	0.0521
Volume transport, controls	0.0060	0.0121
Response-channel transport AUC	0.9269	0.8393
Volatility-channel transport AUC	0.9948	0.9813
Volume-channel transport AUC	0.8801	0.7860
Latent bridge AUC, events vs. controls	0.5297	0.6427
Return log-var delta signed Rank IC	-0.0046	0.0274

The per-channel audit clarifies the RC-DJB transport profile. Its return-mean transport is exactly zero by construction, while the response bridge remains active in volatility and volume. The volatility channel shows high transport for both real disclosures and controls, whereas the volume channel separates real events more clearly. The aggregate control transport therefore reflects a specific response-channel behavior, not a violation of the return-mean constraint.

Interventions. The no-bridge intervention sets the bridge state jump, bridge hidden state, mean delta, and log-variance delta to zero, so $\mu_1 = \mu_0$ and $\ell_1 = \ell_0$. The zero-text intervention replaces e_i with the zero vector before disclosure routing. The shuffled-text intervention permutes disclosure embeddings across held-out rows. For RC-DJB, $\Delta\mu_{return} = 0$ in every variant, so abnormal-return point predictions and rank IC are invariant to these bridge interventions. These ablations test a learned model component rather than identify causal market effects [Rubin, 1974, Imbens and Rubin, 2015].

Table 6: RC-DJB intervention tests. Rank IC is unchanged by bridge interventions because the return-mean constraint fixes the abnormal-return point prediction to the no-event mean. The final column reports latent bridge AUC, not a direct text-discrimination score.

Variant	All MSE	Event MSE	Event NLL	Rank IC	Latent AUC
Full RC-DJB	**0.0562**	**0.0852**	**-2.2189**	0.0154	0.6427
No bridge	0.0713	0.1035	4.8117	0.0154	0.5000
Zero text	0.0580	0.0889	-2.2119	0.0154	0.0067
Shuffled text	0.0574	0.0877	-2.2160	0.0154	0.3554

The no-bridge NLL degradation indicates that the base distribution alone is poorly calibrated for event rows; the bridge is therefore important for event-response uncertainty as well as point error. The zero-text latent AUC should be interpreted narrowly. With a constant zero text input, the latent score reflects bridge biases and metadata interactions rather than disclosure-content separation, which explains the inverted event/control ordering.

Table 7: Paired event-row MSE stress tests. Positive values mean full RC-DJB has lower event-row MSE than the intervention.

Intervention	Event MSE increase	95% paired CI
No bridge	0.0183	[0.0165, 0.0203]
Zero text	0.0037	[0.0025, 0.0051]
Shuffled text	0.0025	[0.0014, 0.0036]

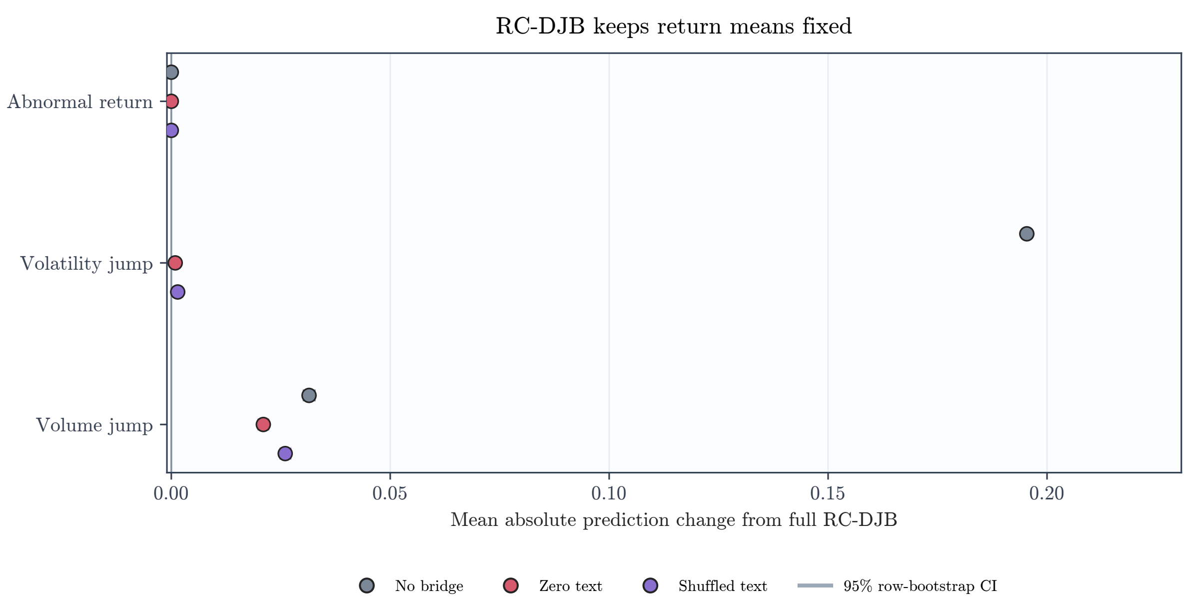

Figure 3: Return-mean constraint audit. Points show mean absolute point-prediction change from full RC-DJB under bridge interventions, with 95% row-bootstrap intervals. The abnormal-return mean prediction remains exactly invariant while response targets move; Table 5 separately audits return log-variance transport.

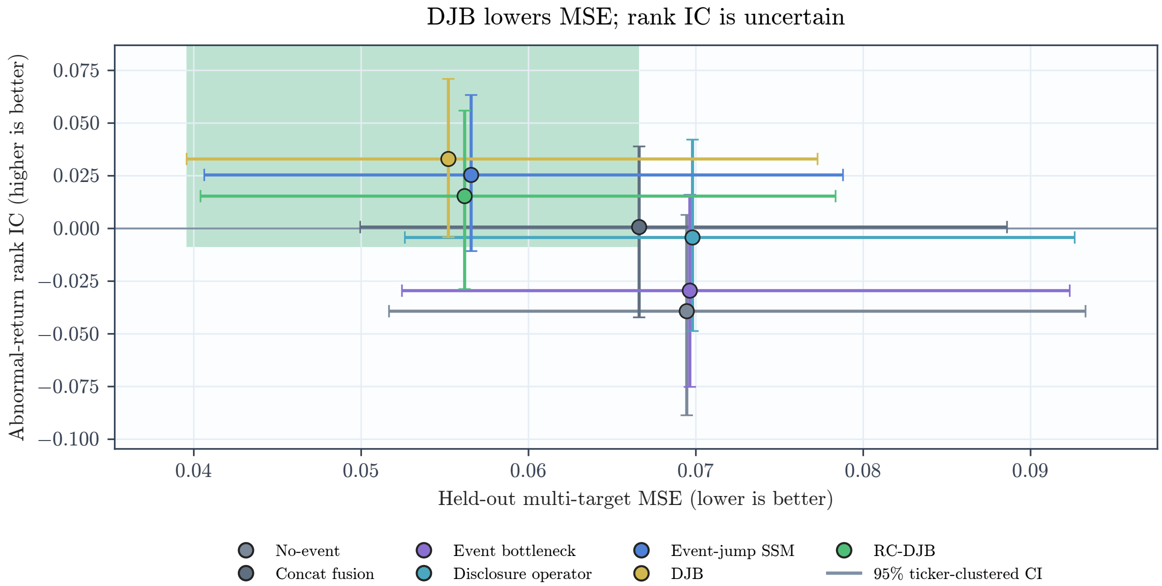

Figure 4: Response/ranking tradeoff after firm/market return-window alignment. DJB has the lowest held-out MSE and largest point Rank IC, while ticker-clustered Rank IC intervals remain close to zero and require stability auditing.

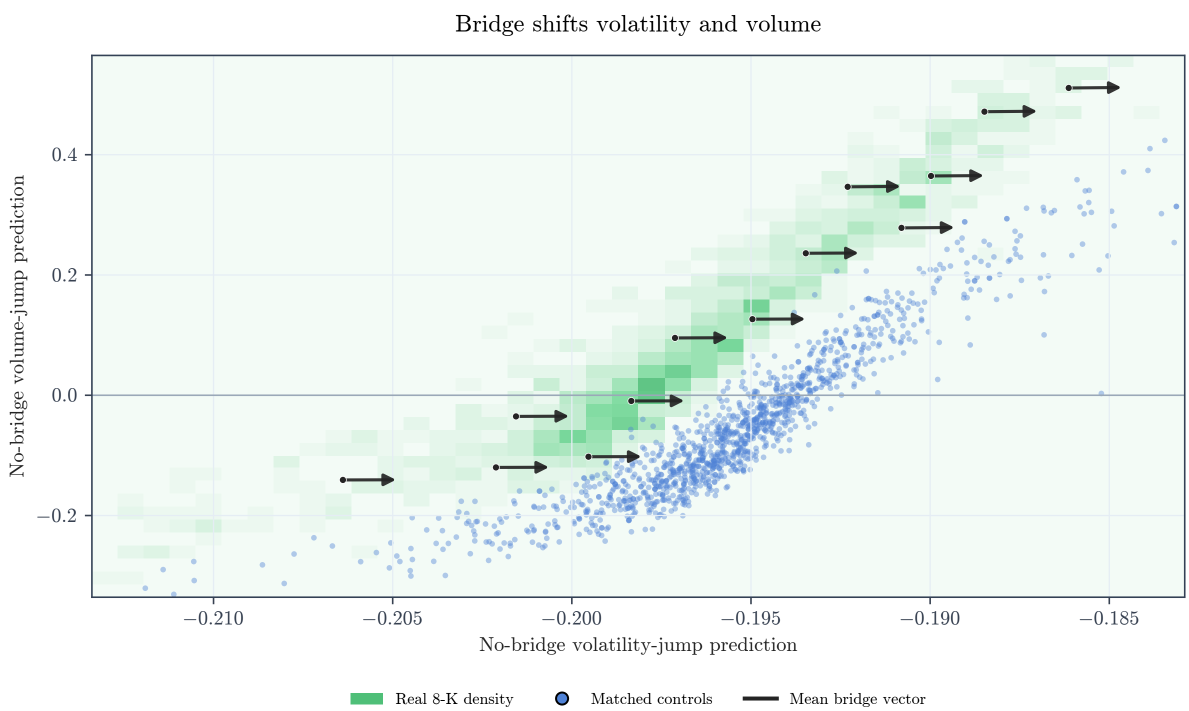

Figure 5: Held-out response-state transport induced by the disclosure bridge. The background shows real 8-K density in the no-bridge volatility/volume response plane, blue points show matched controls, and arrows show binned mean shifts in volatility-jump and volume-jump predictions. The figure shows disclosure text acting through response-distribution transport rather than a single unconstrained fusion head.

6 Results

6.1 DJB lowers response error

After aligning the firm and market return windows, DJB obtains the lowest multi-target MSE (0.0552) and the largest point Rank IC (0.0330). Paired comparisons show lower MSE than every baseline, including concat fusion by 0.0114 [0.0092, 0.0135], the event-jump state-space baseline by 0.0014 [0.0009, 0.0019], and RC-DJB by 0.0010 [0.0005, 0.0015]. The gain is not solely a property of the training substrate; it is associated with the learned residual distributional bridge.

The standalone No-event model has negative held-out rank IC, but this does not imply that pre-event market state is intrinsically anti-predictive. That baseline must explain disclosure-window returns, volatility, and volume using only ordinary market history and metadata. DJB trains its no-event state jointly with a disclosure bridge, control penalty, likelihood term, and rank objective. The result does not imply that an isolated no-event forecaster is sufficient; it shows that no-event dynamics become useful when event text is routed through residual response-distribution transport.

6.2 Return ranking remains uncertain

DJB's ticker-clustered Rank IC interval is [-0.0042, 0.0710], so the rank result is not cluster-significant under ticker resampling. The leave-one-ticker-out audit is still informative: deleting any single held-out ticker leaves DJB Rank IC positive, with minimum 0.0263 after removing KLAC and maximum 0.0381 after removing COST. The supplementary decile profile in Figure A.2 shows the same pattern visually: the rank profile contains signal, but it is not a monotone large-effect curve.

The tighter rank temperature $T = 0.02$ is not the sole source of the rank result. DJB with $T = 0.10$ has lower Rank IC (0.0152), but RC-DJB moves from 0.0154 at $T = 0.02$ to 0.0201 at $T = 0.10$. In both cases the ticker-clustered intervals cross zero. The rank regularizer helps shape the ordering objective, but it does not make the return-ranking evidence cluster-significant.

6.3 RC-DJB isolates bridge transport

RC-DJB does not have the lowest response error after firm/market return-window alignment. It is a controlled ablation that asks what is lost when text is prevented from directly moving abnormal-return means. Its point Rank IC is lower than DJB's (0.0154 vs. 0.0330), while its Gaussian NLL is slightly better (-2.2961 vs. -2.2815). This tradeoff clarifies the role of the constraint: it isolates risk, liquidity, and uncertainty transport from return-mean transport, but the unconstrained bridge is the stronger point forecaster in this sample.

Removing the RC-DJB bridge raises event-row MSE by 0.0183 [0.0165, 0.0203]; zeroing or shuffling text raises event-row MSE by 0.0037 [0.0025, 0.0051] and 0.0025 [0.0014, 0.0036]. These changes occur even though abnormal-return point predictions are unchanged by construction, which means the bridge is acting through volatility, volume, and uncertainty transport. The return log-variance delta has signed Rank IC 0.0274 in RC-DJB. The intervention therefore supports a return-mean constraint, not complete return-distribution isolation.

6.4 Transport varies across disclosures

The RC-DJB control penalty does not eliminate all response-channel movement on matched controls. Table 5 shows that return-mean transport is zero, but the volatility channel has similar magnitude on events and controls. The constraint therefore answers a target-specific routing question, while the per-channel transport audit shows where the bridge remains active.

The bridge-activation ladder in Figure A.3 shows that bridge transport is not a uniform offset applied to all disclosures. Rows with larger learned transport are primarily rows where the volume-response channel moves more strongly, and the highest-transport deciles also have larger realized volume responses. This is consistent with event-study evidence that disclosure response depends on the event, firm, and market state [Beaver, 1968, Patell and Wolfson, 1984, Kothari and Warner, 2007].

7 Reproducibility

The public code repository provides the configurations, data-ingestion scripts, model implementations, evaluation code, tables, and figure-generation scripts used in this study: github.com/aryan-cs/distributional-jump-bridges. Large generated artifacts such as processed metadata, checkpoints, cached predictions, and downloaded market files are not tracked in Git; they are reproducible from the documented pipeline. Automated tests cover timestamp gating, after-hours label starts, feature leakage rejection, control construction, deterministic tensor generation, finite model losses, probabilistic loss integration, and evaluation leakage guards. Labels are computed only from future price windows and are never included in model inputs.

8 Limitations

The study is scoped to a leakage-controlled architecture test. The 98-firm universe is a dense large-cap disclosure-response panel, not an all-equity trading universe; scaling the same protocol to broader S&P 500, Russell 3000, small-cap, and low-liquidity samples is an important external-validity test. The price source is public daily data rather than CRSP or intraday trades and quotes, so the return label is designed as a conservative forward-drift target rather than an immediate jump estimator. A previous-close or intraday label would answer a complementary question about first-session repricing. Abnormal returns are SPY-adjusted rather than factor-model residuals [Fama and French, 1993, Carhart, 1997]. BGE-small is a general-purpose encoder with chunk mean pooling; domain-specific or long-document financial encoders such as FinBERT, BloombergGPT-style models, FinGPT, or SEC-specialized long-context encoders could change absolute response error [Araci, 2019, Wu et al., 2023, Yang et al., 2023]. Because all text-aware models in the main experiment share the same frozen text representation, the reported comparison isolates the routing architecture rather than encoder scale. Rank IC is positive in point estimate and stable under leave-one-ticker deletion, but its ticker-clustered interval crosses zero; net performance after spreads, commissions, market impact, borrow constraints, capacity, and turnover requires a separate portfolio-construction study [Grinold and Kahn, 2000, Novy-Marx and Velikov, 2016, Frazzini et al., 2018, López de Prado, 2018]. The return-mean constraint is architectural and predictive, not structural causal identification [Rubin, 1974, Imbens and Rubin, 2015]; it blocks direct movement of the return mean but still allows return log-variance transport.

9 Responsible Use

This paper studies event representations using public filings and public daily prices. The results do not constitute trading recommendations. The replication package is intended to support leakage auditing, independent reproduction, and falsification of the empirical claims.

10 Conclusion

DJB shows that the route by which disclosure text enters a financial event model matters. In this study, disclosure text contributes most clearly when it acts as a residual distributional transport operator layered on top of no-event market dynamics. In this public-data experiment, the unconstrained DJB has the lowest response error, while RC-DJB clarifies what changes when a return-mean constraint is imposed. The resulting principle is architectural and testable: model disclosures as distributional jumps, audit return-ranking claims under ticker-clustered uncertainty, and separate return-mean transport from risk, liquidity, and uncertainty transport.

A Supplementary Diagnostics

The following figures provide secondary robustness and heterogeneity views without changing the paper's main empirical claims.

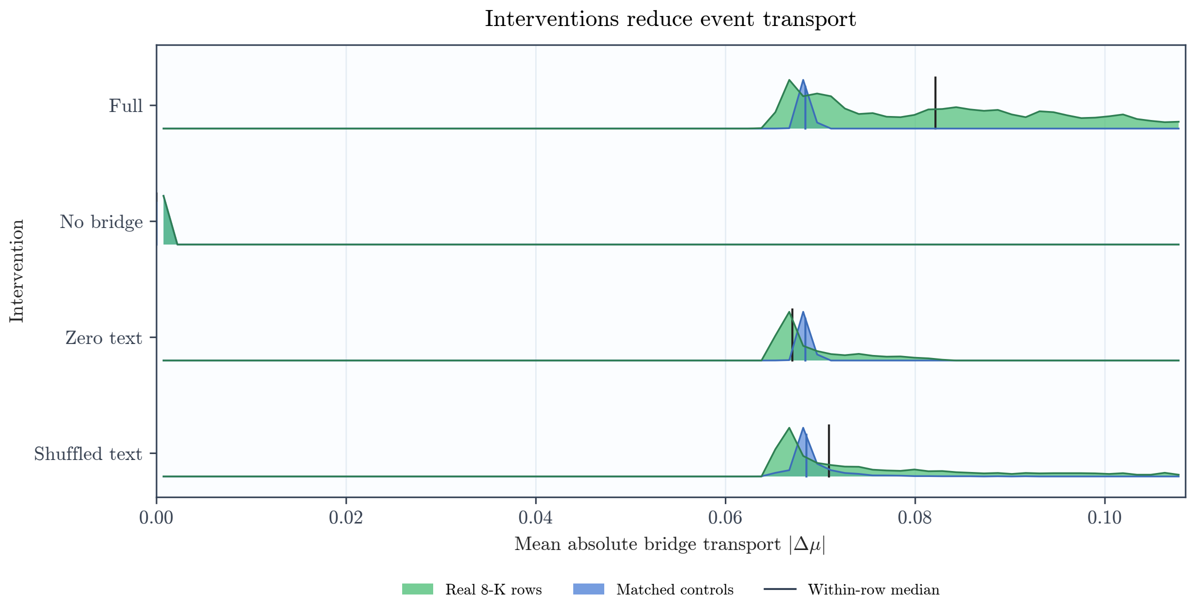

Figure A.1: Response-transport fingerprints under bridge interventions. Ridge densities compare mean absolute response transport for real disclosure rows and matched controls; vertical marks show within-row medians.

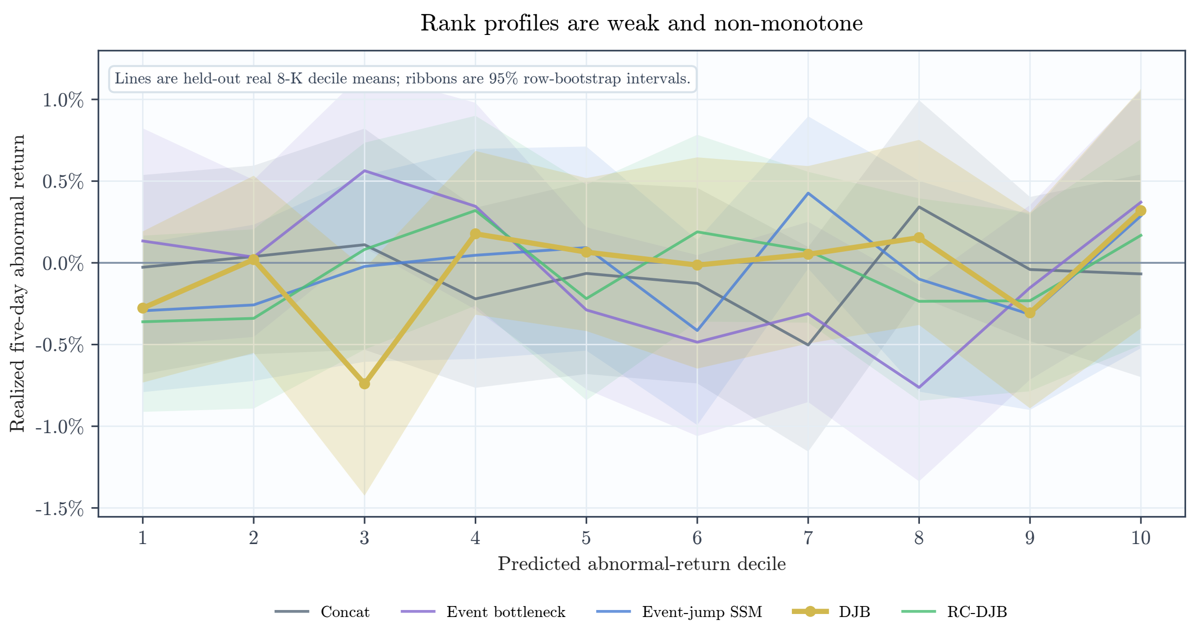

Figure A.2: Predicted abnormal-return decile profiles on held-out real 8-K rows. Lines show realized five-day abnormal-return means; ribbons are 95% row-bootstrap confidence intervals computed within each predicted decile. The non-monotone profile is consistent with a positive but statistically uncertain ranking signal rather than a large trading-style effect.

Figure A.3: Bridge-activation ladder for held-out real 8-K rows. Rows are sorted by RC-DJB response transport and grouped into deciles. The upper panel shows that variation in the bridge is carried mostly by the volume-response channel, while volatility transport is nearly constant; the lower panel reports realized post-filing response magnitudes in sample-standard-deviation units with 95% row-bootstrap intervals.

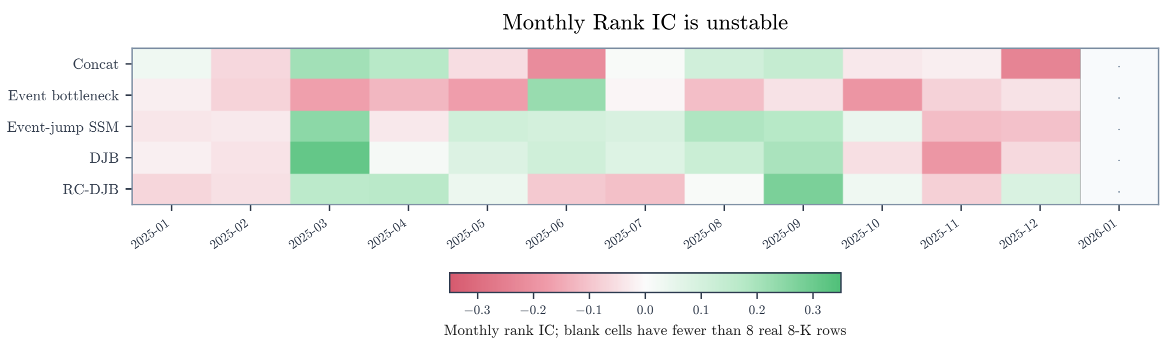

Figure A.4: Monthly event-rank IC over held-out real 8-K rows. The model-level rank result is not uniform across months, motivating ticker-clustered intervals and leave-one-ticker-out stability checks rather than a single unqualified rank statistic.

References

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2): 579–625, 2003. doi:10.1111/1468-0262.00418.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=gerNCVqqtR.

Dogu Araci. FinBERT: Financial sentiment analysis with pre-trained language models, 2019. URL https://arxiv.org/abs/1908.10063.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016. URL https://arxiv.org/abs/1607.06450.

David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. The probability of backtest overfitting. *Journal of Computational Finance*, 20(4):39–69, 2016. doi:10.21314/JCF.2016.322.

Linda Smith Bamber. The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research*, 24(1): 40–56, 1986. doi:10.2307/2490803.

William H. Beaver. The information content of annual earnings announcements. *Journal of Accounting Research*, 6:67–92, 1968. doi:10.2307/2490070.

Christoph Bergmeir and José M. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012. doi:10.1016/j.ins.2011.12.028.

Ekkehart Boehmer, Jim Musumeci, and Annette B. Poulsen. Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2):253–272, 1991. doi:10.1016/0304-405X(91)90032-F.

Stephen J. Brown and Jerold B. Warner. Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(1):3–31, 1985. doi:10.1016/0304-405X(85)90042-X.

Christopher J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory N. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 89–96, 2005. doi:10.1145/1102351.1102363.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008. doi:10.1162/rest.90.3.414.

John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, pages 129–136, 2007. doi:10.1145/1273496.1273513.

Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997. doi:10.1111/j.1540-6261.1997.tb03808.x.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, 2014. doi:10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Lauren Cohen, Christopher Malloy, and Quoc Nguyen. Lazy prices. *The Journal of Finance*, 75(3):1371–1415, 2020. doi:10.1111/jofi.12885.

Anthony C. Davison and David V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997. doi:10.1017/CBO9780511802843.

A. Philip Dawid. Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A*, 147(2):278–292, 1984.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. doi:10.1214/aos/1176344552.

Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970. doi:10.1111/j.1540-6261.1970.tb00518.x.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. doi:10.1016/0304-405X(93)90023-5.

Eugene F. Fama, Lawrence Fisher, Michael C. Jensen, and Richard Roll. The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21, 1969. doi:10.2307/2525569.

Andrea Frazzini, Ronen Israel, and Tobias J. Moskowitz. Trading costs. *Available at SSRN 3229719*, 2018. doi:10.2139/ssrn.3229719. URL https://ssrn.com/abstract=3229719.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi:10.1198/016214506000001437.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models, 2024. URL https://arxiv.org/abs/2402.03885. Accepted at ICML 2024.

Richard C. Grinold and Ronald N. Kahn. *Active Portfolio Management*. McGraw-Hill, 2 edition, 2000.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces, 2023. URL https://arxiv.org/abs/2312.00752.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi:10.1093/rfs/hhaa009.

James A. Hanley and Barbara J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143 (1):29–36, 1982. doi:10.1148/radiology.143.1.7063747.

Peter Reinhard Hansen. A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380, 2005. doi:10.1198/073500105000000063.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi:10.1214/aoms/1177703732.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 3 edition, 2021. URL https://otexts.com/fpp3/.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015. doi:10.1017/CBO9781139025751.

Narasimhan Jegadeesh and Di Wu. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729, 2013. doi:10.1016/j.jfineco.2013.08.018.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 6769–6781, Online, 2020. Association for Computational Linguistics. doi:10.18653/v1/2020.emnlp-main.550. URL https://aclanthology.org/2020.emnlp-main.550/.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. doi:10.48550/arXiv.1312.6114. URL https://arxiv.org/abs/1312.6114.

Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. Predicting risk from financial reports with regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 272–280, Boulder, Colorado, 2009. Association for Computational Linguistics. URL https://aclanthology.org/N09-1031/.

S. P. Kothari. Capital markets research in accounting. *Journal of Accounting and Economics*, 31(1–3):105–231, 2001. doi:10.1016/S0165-4101(01)00030-1.

S. P. Kothari and Jerold B. Warner. Econometrics of event studies. In B. Espen Eckbo, editor, *Handbook of Corporate Finance: Empirical Corporate Finance*, volume 1, pages 3–36. Elsevier, 2007. doi:10.1016/B978-0-444-53265-7.50015-9.

Xiangyu Li, Yawen Zeng, Xiaofen Xing, Jin Xu, and Xiangmin Xu. Profit mirage: Revisiting information leakage in LLM-based financial agents, 2025. URL https://arxiv.org/abs/2510.07920.

Bryan Lim, Sercan Ö. Arik, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021. doi:10.1016/j.ijforecast.2021.03.012.

Marcos López de Prado. *Advances in Financial Machine Learning*. Wiley, 2018.

Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65, 2011. doi:10.1111/j.1540-6261.2010.01625.x.

A. Craig MacKinlay. Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39, 1997.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Robert Novy-Marx and Mihail Velikov. A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147, 2016. doi:10.1093/rfs/hhv063.

James M. Patell and Mark A. Wolfson. The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics*, 13(2):223–252, 1984. doi:10.1016/0304-405X(84)90024-2.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3982–3992, Hong Kong, China, 2019. Association for Computational Linguistics. doi:10.18653/v1/D19-1410. URL https://aclanthology.org/D19-1410/.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5): 688–701, 1974.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020. doi:10.1016/j.ijforecast.2019.07.001.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964. doi:10.1111/j.1540-6261.1964.tb02865.x.

Charles Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904. doi:10.2307/1412159.

Paul C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007. doi:10.1111/j.1540-6261.2007.01232.x.

Paul C. Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467, 2008. doi:10.1111/j.1540-6261.2008.01362.x.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method, 2000. URL https://arxiv.org/abs/physics/0004057.

U.S. Securities and Exchange Commission. EDGAR application programming interfaces (APIs). https://www.sec.gov/edgar/sec-api-documentation, 2026. Accessed May 3, 2026.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Peihao Wang, Ruisi Cai, Yuehao Wang, Jiajun Zhu, Pragya Srivastava, Zhangyang Wang, and Pan Li. Understanding and mitigating bottlenecks of state space models through the lens of recency and over-smoothing. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=pymXpl4qvi.

Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000. doi:10.1111/1468-0262.00152.

Robert L. Winkler. A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, 67(337):187–191, 1972. doi:10.1080/01621459.1972.10481224.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. BloombergGPT: A large language model for finance, 2023. URL https://arxiv.org/abs/2303.17564.

Shitao Xiao, Zheng Liu, Peitian Zhang, Niklas Muennighoff, Defu Lian, and Jian-Yun Nie. C-Pack: Packed resources for General Chinese Embeddings, 2023. URL https://arxiv.org/abs/2309.07597.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. FinGPT: Open-source financial large language models, 2023. URL https://arxiv.org/abs/2306.06031.

Yang Zhang, Wenbo Yang, Jun Wang, Qiang Ma, and Jie Xiong. CAMEF: Causal-augmented multi-modality event-driven financial forecasting by integrating time series patterns and salient macroeconomic announcements. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3867–3878, 2025. doi:10.1145/3711896.3736872. URL https://arxiv.org/abs/2502.04592. Accepted at SIGKDD 2025.