Algorithmic Alpha in FOMC Prediction Markets

A SYSTEMATIC APPROACH TO FED COMMUNICATIONS

Andrew Cun | March 2026

Abstract. As prediction markets mature, they have birthed highly liquid, hyper-specific contracts centered around macroeconomic events. Among the most heavily traded are "Fed Speak" markets—contracts resolving on the exact phrases and word counts uttered by Federal Reserve Chair Jerome Powell during Federal Open Market Committee (FOMC) press conferences. Historically, trading these markets has relied on human intuition and manual execution. This paper outlines the architecture of a fully autonomous, algorithmic trading system designed to systematically capture alpha in FOMC prediction markets. The system synthesizes real-time macroeconomic data, historical transcript analysis, and live news sentiment. More importantly, it features a robust execution engine utilizing depth-adjusted pricing, correlation-aware phrase-family exposure limits, and continuous Bayesian updating with systematic loss-cutting during live speech ingestion.

Figure note. The embedded visuals are schematic, manuscript-aligned illustrations generated from the paper's conceptual framework rather than empirical backtest exhibits.

1. System Architecture: The Tripartite Model

The trading system is built on a tripartite architecture that strictly separates data processing, statistical prediction, and execution. This modularity ensures the system behaves as a disciplined quantitative pipeline rather than a monolithic oracle.

1. **The Sensory Engine:** Normalizes prediction market contracts, ingests live macroeconomic indicators, scrapes real-time financial news sentiment, and smooths volatile data inputs.
2. **The Statistical Engine:** Calculates dynamic probabilities via Mahalanobis distance regime matching, Negative Binomial count modeling, and tracks strict calibration metrics.
3. **The Risk & Execution Engine:** Enforces dynamic, correlation-aware portfolio constraints, evaluates live order books using depth-adjusted pricing, and executes orders via a multi-factor ranking algorithm.

2. The Data Pipeline: Signal Fidelity and Macro Robustness

Before any statistical model runs, the system addresses a critical operational moat: data quality and signal fidelity.

Market Canonicalization and Parsing

Prediction market titles across venues like Polymarket and Kalshi are notoriously messy, inconsistently formatted, and often ambiguous regarding resolution thresholds. The system employs a dedicated parsing engine that autonomously normalizes and canonicalizes raw market titles. It extracts precise numerical thresholds, resolves lexical inflection aliases via stemming and alias matching (for example, linking "rate cut" to "easing"), and systematically rejects ambiguous or malformed contracts. This eliminates a massive class of operational execution errors.

Anomaly Treatment and NFP Winsorization

Raw macroeconomic data is noisy and subject to massive exogenous shocks. For highly volatile series like Non-Farm Payrolls (NFP), the pipeline applies strict quantitative smoothing. Instead of relying solely on the absolute NFP surprise, the system calculates a three-month rolling mean blend. It applies strict statistical winsorization, capping extreme standard deviation outliers at 2.5 sigma and enforcing a hard absolute cap on NFP delta. This prevents a single anomalous macro print from aggressively skewing historical regime matching.

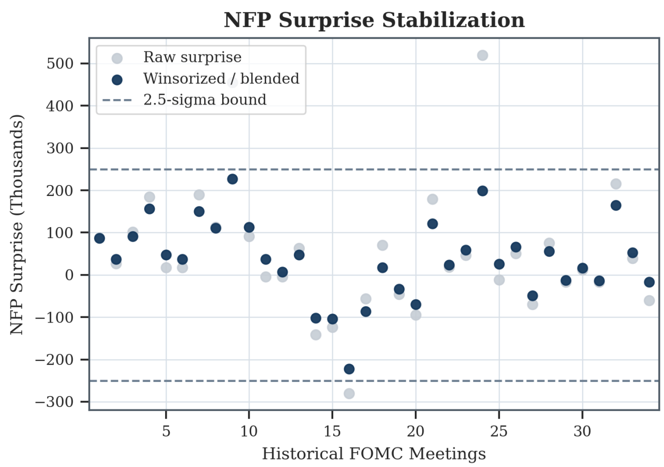

Figure 1: Illustrative NFP surprise stabilization. Light-gray points show raw historical surprises, while the navy series shows winsorized and blended values used to reduce regime-matching distortion from extreme macro shocks.

3. The Statistical Engine: Quantifying Fed Speak

The system translates the smoothed macroeconomic context into precise, tradable probabilities using historical distance matching and specialized distributional math.

Regime Similarity and the "Unprecedented Shock" Edge Case

The model uses Mahalanobis distance to match the current macroeconomic snapshot to historical FOMC eras. However, a critical edge case arises during unprecedented macro shocks where no close historical analog exists. While the distance algorithm will mathematically identify a "closest" match, the absolute distance may still be dangerously high. The system flags these low-confidence regime matches by calculating a thin effective sample size, which subsequently generates wide confidence intervals and triggers a severe execution haircut in the risk engine, effectively sidelining the bot until a new baseline is established.

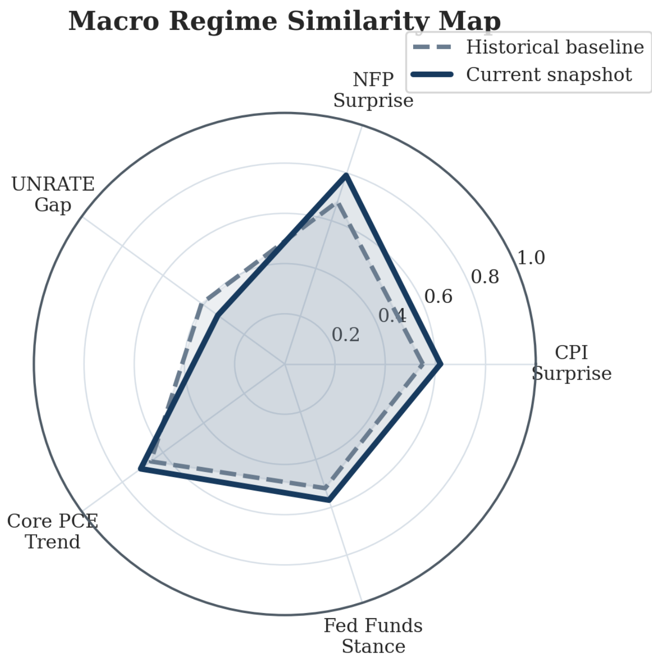

Figure 2: Illustrative macro regime similarity map. A current snapshot is compared against a historical baseline so the model can detect whether present conditions lie inside a familiar policy regime or represent a low-confidence, weak-analog shock state.

Negative Binomial Pricing for Count Markets

Because word frequencies exhibit overdispersion, the system models count contracts using a Negative Binomial distribution rather than a standard Poisson. By analyzing the mean mentions and variance in historically similar economic regimes, the model accurately prices the probability mass residing in the right tail of the distribution, identifying mispriced thresholds in the public market.

4. Phrase-Family Exposure and Correlation Risk

Before evaluating individual trades, the risk engine establishes a macro-portfolio view. A naive system might independently buy “Yes” contracts on “Inflation,” “Transitory,” and “CPI,” failing to realize it has taken a massively concentrated, highly correlated position.

To mitigate this, the system categorizes all tradable words into proprietary phrase families such as INFLATION, LABOR, GROWTH, and HOUSING. The risk engine enforces strict correlation-aware exposure limits. If the portfolio reaches its maximum capital allocation for the LABOR family, the system automatically restricts further exposure, forcing capital diversification across uncorrelated speech topics regardless of individual contract appeal.

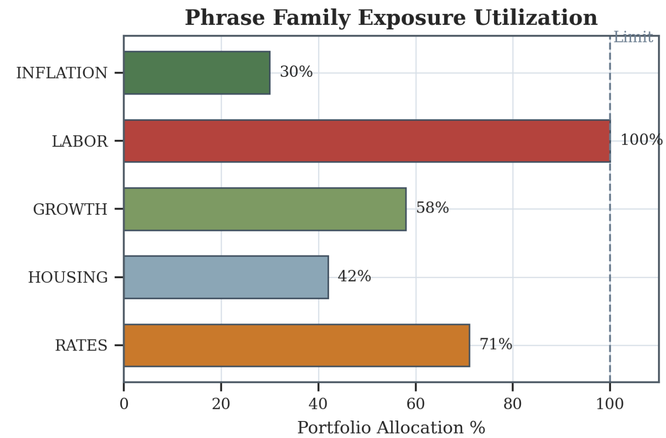

Figure 3: Illustrative phrase-family utilization map. Exposure is monitored at the family level so attractive single-contract trades cannot push the portfolio beyond aggregate thematic limits.

5. Multi-Factor Trade Ranking and Depth-Adjusted Execution

Candidate trades are passed through a proprietary multi-factor ranking algorithm that evaluates trades across multiple simultaneous dimensions.

Depth-Adjusted Expected Value (EV)

The model rejects the naive assumption that the best bid or ask represents the true cost of a trade. The execution engine reads the live order book and computes a depth-adjusted price. If the system intends to purchase 1,000 shares, it walks up the limit order book, calculating the volume-weighted average price required to fill the entire order. Expected value is then calculated against this depth-adjusted basis, ensuring that spread slippage, depth constraints, and taker fees do not compress the net EV below the hard-coded viability threshold.

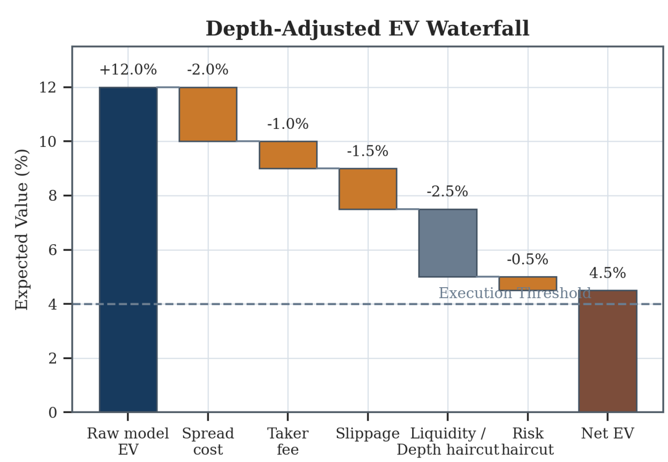

Figure 4: Illustrative depth-adjusted EV decomposition. A trade with strong raw model edge can fall below the execution threshold once spread, fees, slippage, liquidity depth, and risk haircuts are incorporated into the fill model.

The Scoring Matrix

The algorithm scores approved candidate trades based on:

- **Gross Edge & EV:** The divergence between the model's probability and the depth-adjusted market price.
- **Confidence Interval (CI) Scale:** Tighter confidence intervals receive higher execution priority.
- **Phrase Reliability & Liquidity:** Markets with deep books, tight spreads, and historically reliable tracking are heavily favored.

6. The Live Event Overlay: Dynamic In-Speech Adaptation

The system features a live tracking engine that operates dynamically during the press conference. As the event unfolds, the system ingests live chunks of the transcript, continuously updating its internal counters and Bayesian posteriors.

Trajectory Matching

The system does not merely track a running word count. It executes a historical trajectory match, calculating real-time similarity scores based on family overlap and phrase density against the arc of historical speech paths.

Tactical Adds and Emergency Exits

By projecting this trajectory, the system executes mid-speech adjustments:

- **Tactical Adds:** If at token 800 the live parsing engine detects an aggressive dovish tone and the posterior probability for a rate-cut phrase jumps from 52% to 68%, the system executes a tactical add, purchasing shares before the slower market reacts.
- **Systematic Loss Cutting:** Conversely, the system actively unwinds deteriorating positions. If the live trajectory deviates from the pre-speech prediction, it calculates a strict mathematical tradeoff between hold EV and exit EV. If exit EV crosses hold EV, the system executes an emergency position unwind, systematically capping downside variance before contract resolution.

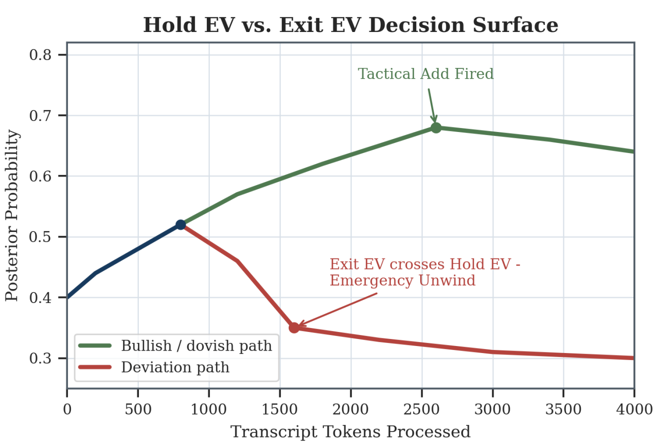

Figure 5: Illustrative in-speech decision surface. Once live transcript evidence diverges, the engine can either add to a strengthening theme or exit when realized trajectory causes exit EV to exceed hold EV.

7. Model Calibration and Paper Trading Infrastructure

A probabilistic model is only as useful as its calibration. A system predicting an event with 70% confidence must observe that event occurring exactly 70% of the time over a large sample.

Rigorous Calibration Tracking

The architecture includes a dedicated calibration module that actively tracks:

- **Brier Score:** Measuring the mean squared difference between predicted probabilities and actual binary outcomes.
- **Log-Loss (Cross-Entropy):** Heavily penalizing confident but incorrect predictions.
- **Expected Calibration Error (ECE):** Grouping predictions into decile buckets to measure reliability across the probability spectrum.

$$\text{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left|\text{acc}(B_m) - \text{conf}(B_m)\right|$$

By evaluating these metrics on a per-family basis, the system identifies systematic overconfidence in specific domains and autonomously recalibrates.

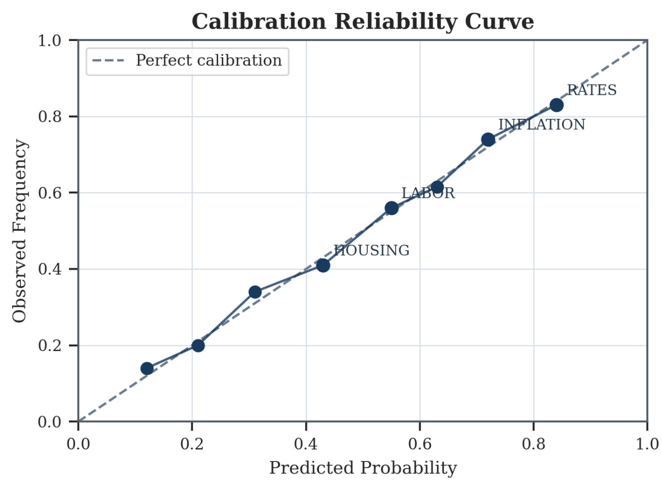

Figure 6: Illustrative reliability diagram. Phrase-family level calibration is monitored around the 45-degree line so the model can detect and correct persistent overconfidence or underconfidence.

Operational Discipline

To validate these metrics without risking live capital, the architecture features robust paper-trading and dry-run infrastructure, allowing the model to simulate execution, factor in historical slippage, and prove its calibrated edge in out-of-sample data prior to live deployment.

8. Limitations

While structurally robust, the model operates under inherent limitations:

- **Thin Historical Sample:** The FOMC meets only eight times per year. While the model maximizes this data via transcript tokenization, the absolute sample size of modern Fed press conferences remains statistically thin compared to high-frequency equity data.
- **Liquidity Constraints:** Prediction markets often lack the depth to support institutional-scale capital. The depth-adjusted pricing module effectively limits position sizing to what the current order book can support without massive slippage.
- **Latency Dependencies:** The live event overlay relies on high-speed API transcription and market-data polling. In a scenario with severe exchange-side lag, the system's mid-speech execution edge degrades.

Conclusion

The gap between discretionary macro trading and fully systematic execution in prediction markets is bridged by structural rigor. By combining robust data winsorization, lexical canonicalization moats, depth-adjusted liquidity analysis, and dynamic hold-versus-exit EV evaluations, the system effectively removes human bias and latency from the trading process. It stands as a production-grade architecture capable of systematically capturing alpha in the complex environment of Federal Reserve communications.

Figures in this document are schematic, manuscript-aligned illustrations generated from the user's requested conceptual specifications rather than empirical backtest exhibits.