Prediction Market Efficiency: Analyzing Kalshi Contract Dynamics

Emma Nagy

December 2024

Abstract

This study examines the efficiency and predictability of prediction markets by analyzing three Kalshi contracts spanning cryptocurrency ($\geq$ 129k), equity indices (S&P 500 $\geq$ 6999), and monetary policy (Fed December 2025 no-hike). Through correlation analysis, lead-lag dynamics, and machine learning forecasting across seven models, we document that Kalshi contracts efficiently track underlying assets with behavior patterns reflecting structural differences across asset classes. Bitcoin contracts lag spot prices by one day due to liquidity differentials (correlation 0.61 at lag +1), S&P 500 contracts adjust synchronously (peak correlation 0.49 at lag 0), and Fed contracts respond to discrete information events (peak correlation 0.32 at lag 0). While machine learning models achieve 64–83% in-sample accuracy, performance is unstable and deteriorates out of sample, consistent with limited predictability. Comparison with CME FedWatch reveals structural consistency in monetary policy expectations despite inverse contract designs (correlation -0.63). Results demonstrate that prediction markets function as efficient information aggregation mechanisms without persistent arbitrage opportunities.

1 Introduction

Prediction markets provide a structured mechanism for aggregating dispersed information about future events by allowing participants to trade contracts whose payoffs depend on event outcomes. Unlike traditional forecasting methods, prediction markets harness the wisdom of crowds and create financial incentives for accurate information revelation. This study evaluates the efficiency and predictability of prediction markets by examining three Kalshi contracts across distinct asset classes.

Kalshi operates as a CFTC-regulated prediction market exchange where users trade binary options on real-world events. Contracts settle based on objective outcomes, creating a direct link between market prices and probability assessments. This analysis addresses three fundamental questions: How closely do prediction market contracts track their underlying assets? Do contracts anticipate or react to movements in reference markets? Can machine learning models identify exploitable patterns in contract price dynamics?

We analyze three contracts representing different market structures and information environments:

Bitcoin $\geq$ \$129,000: Binary contract on whether Bitcoin closes above \$129,000, tracking a 24/7 highly liquid cryptocurrency market with continuous price discovery.

S&P 500 $\geq$ 6999: Binary contract on year-end S&P 500 level, tied to a major equity index with standard trading hours and established market infrastructure.

Federal Reserve December 2025 (No Hike): Binary contract on the Fed maintaining unchanged rates, reflecting event-driven monetary policy expectations influenced by discrete information releases.

These contracts span continuous markets (cryptocurrency, equities) and event-driven expectations (monetary policy), enabling assessment of how prediction market efficiency varies across fundamentally different information structures. The analysis combines correlation metrics, cross-correlation lead-lag analysis, comparison with alternative probability measures (CME FedWatch), and machine learning forecasting to provide a comprehensive evaluation of market dynamics.

2 Data and Methodology

2.1 Data Sources

Contract price data were obtained from the Kalshi API, providing daily closing prices for each prediction market contract. Underlying asset data include Bitcoin spot prices and S&P 500 index values from Yahoo Finance, and CME FedWatch probabilities derived from federal funds futures for monetary policy expectations.

The sample period spans contract inception through December 2024, with observation counts varying by contract based on launch dates. Bitcoin and S&P 500 contracts provide sufficient data for time series analysis, while the Fed contract reflects a shorter horizon given its specific event focus.

2.2 Analytical Framework

Correlation Analysis: Two measures characterize the contract-underlying relationship. Level correlation assesses long-run co-movement between contract prices and underlying asset values, capturing whether contracts track broad trends. Return correlation measures day-to-day synchronization, revealing whether contracts react to short-term fluctuations or primarily respond to sustained movements.

Lead-Lag Dynamics: Cross-correlation analysis examines returns at lags from -5 to +5 days. Negative lags indicate contracts leading underlying markets (predictive power), while positive lags suggest reactive behavior. Lag 0 represents synchronous adjustment. This approach identifies whether prediction markets anticipate price movements or incorporate information after it appears in reference markets.

Machine Learning Forecasting: Seven models were trained to predict next-day contract direction: Logistic Regression, Ridge Classifier, LASSO, Random Forest, XGBoost, LightGBM, and Neural Networks. Features include lagged returns (1-5 days), rolling volatility (3, 5, 10-day windows), and volume-based metrics. Training employed 70/30 time-series splits preserving temporal ordering. Performance metrics include accuracy and ROC-AUC scores.

Platform Comparison: The Fed contract analysis includes comparison with CME FedWatch probabilities. While Kalshi provides binary "no change" probabilities, FedWatch distributes probability across all potential rate outcomes. This comparison tests whether different market structures produce consistent probability assessments for identical events.

3   Results

3.1   Market-Underlying Asset Relationships

Contract-underlying relationships reveal distinct patterns across asset classes, reflecting differences in market structure, information flow, and trading environments.

Bitcoin ≥ \$129K: The Bitcoin contract exhibits strong long-run alignment (level correlation 0.95) but weak day-to-day co-movement (return correlation 0.17). This divergence reflects different trading environments. Bitcoin operates as a continuous 24/7 global market with high liquidity and immediate reaction to information. The Kalshi contract, trading in a thinner market with standard hours, adjusts when participants revise threshold-crossing probabilities rather than tracking every Bitcoin price fluctuation. The contract accurately captures Bitcoin's directional trend without mirroring short-term volatility.

S&P 500 ≥ 6999: The equity contract shows moderate trend alignment (level correlation 0.80) but minimal daily synchronization (return correlation 0.05). Despite both the index and contract rising over the sample period, the contract adjusts in discrete steps. This pattern stems from the binary contract structure: small index movements often fail to meaningfully alter year-end threshold-crossing probability. The contract remains stable during minor fluctuations and moves when fundamental outlook shifts, producing smooth underlying trends but jumpier contract prices.

Fed December 2025 (No Hike): The monetary policy contract displays weak full-period alignment (level correlation 0.35) but relatively stronger day-to-day connection (0.32) compared to other contracts. This inverted pattern reflects how policy expectations evolve. Extended periods show minimal change, punctuated by sharp revisions following data releases or Fed communications. Rather than tracking gradual trends, the contract responds to discrete information events, producing weak long-run correlation but stronger short-term co-movement during active periods.

These patterns demonstrate that Kalshi contracts follow broad directional movements more reliably than daily fluctuations. Observed differences reflect underlying asset characteristics—continuous price discovery for Bitcoin and equities, event-driven expectations for monetary policy—rather than market inefficiency.

3.2   Lead-Lag Dynamics

Cross-correlation analysis reveals whether prediction markets anticipate or react to movements in underlying markets, providing insight into information incorporation timing.

Bitcoin Contract (Figure 1): Analysis shows a clear peak at lag +1 with correlation approximately 0.61. The Kalshi contract typically adjusts one day after Bitcoin spot price movements. Given liquidity differences between a global 24/7 cryptocurrency market and a prediction market with limited depth, this one-day lag is economically rational. Kalshi participants update positions once underlying movements appear persistent rather than reacting to every price fluctuation. This pattern indicates reactive rather than predictive behavior, consistent with efficient incorporation of confirmed information.
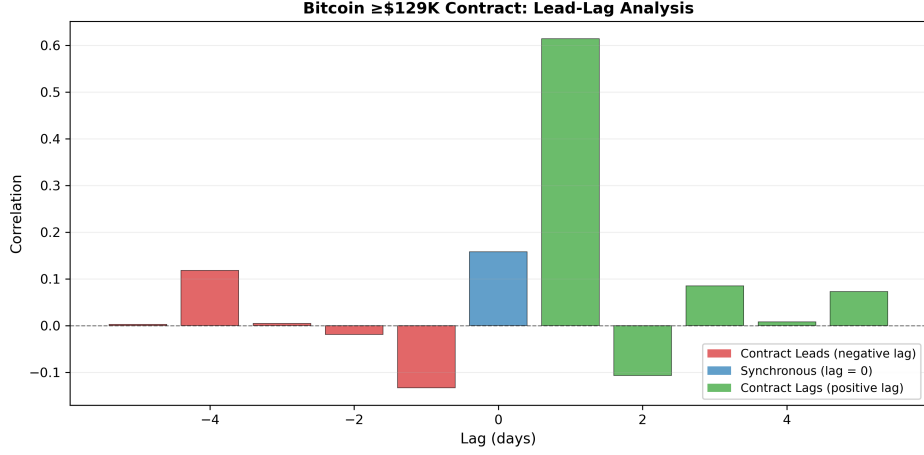
Figure 1: Lead-lag correlations for Bitcoin $\geq$ \$129K contract. Peak at lag +1 indicates contract adjusts one day after underlying Bitcoin price movements, reflecting liquidity-driven delay rather than forecasting ability.

S&P 500 Contract (Figure 2): The strongest correlation occurs at lag 0 (approximately 0.49), indicating synchronous adjustment. When the equity index moves, the Kalshi contract updates on the same day. Negligible correlations at adjacent lags confirm efficient real-time information incorporation without meaningful predictive power. This pattern suggests prediction market participants process equity market movements promptly, adjusting probabilities as new information emerges rather than anticipating future changes.
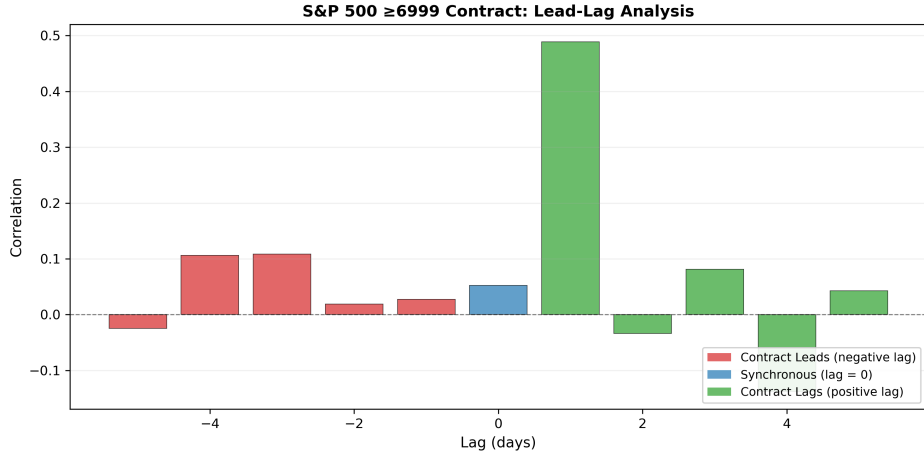

Figure 2: Lead-lag correlations for S&P 500 $\geq$ 6999 contract. Peak at lag 0 demonstrates synchronous adjustment, consistent with efficient information incorporation and limited forecasting power.

Fed December 2025 Contract (Figure 3): The Fed contract peaks at lag 0 with correlation near 0.32. Unlike equity contracts, several adjacent lags show smaller but noticeable correlations, producing a diffuse pattern. This reflects monetary policy's event-driven nature: expectations shift sharply around scheduled announcements (CPI releases, FOMC statements) and drift between events. The diffuse correlation pattern indicates that policy expectations adjust around informa-

tion clusters rather than continuously, consistent with the scheduled nature of monetary policy communication.
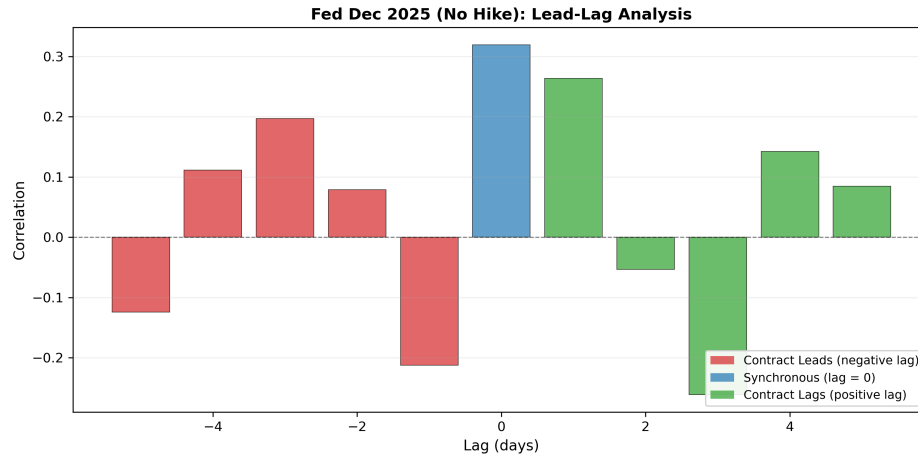

Figure 3: Lead-lag correlations for Fed December 2025 (No Hike) contract. Peak at lag 0 with diffuse adjacent correlations reflects event-driven policy expectations that shift around scheduled announcements rather than continuously.

Summary: No contract exhibits consistent predictive power. Bitcoin contracts lag underlying markets by one day (liquidity-driven delay), S&P 500 contracts adjust synchronously (efficient incorporation), and Fed contracts respond to discrete events (scheduled information releases). These patterns highlight asset class differences—crypto shows liquidity constraints, equities incorporate quickly, policy expectations shift discretely—without revealing persistent arbitrage opportunities or systematic predictability.

3.3 FOMC Case Study: Kalshi vs. CME FedWatch

To evaluate whether different market structures produce consistent probability assessments, we compare Kalshi's binary "no change" contract with CME FedWatch probabilities derived from federal funds futures. While both reflect Federal Reserve meeting expectations, their structures differ fundamentally:

Kalshi assigns a single probability to "no change" in the federal funds rate. FedWatch distributes probability across all potential rate outcomes based on futures pricing. These structural differences create apparent divergence that requires careful interpretation.

During the sample period, markets generally expected rate cuts. FedWatch assigned high probability to lower-rate outcomes while Kalshi's "no change" probability remained low. As rate-cut expectations subsided, FedWatch probabilities declined and Kalshi's probability rose, producing an inverse relationship.

This generates strong negative correlation at levels (-0.63) and changes (-0.35). Crucially, this negative correlation does not indicate disagreement between markets. Instead, it reflects opposite contract designs expressing identical information. When FedWatch shows high probability of cuts, Kalshi shows low probability of no change—logically consistent assessments viewed through different lenses.

Neither market demonstrates consistent timing advantage over the other. Both adjust around the

same information events (economic data releases, Fed communications) without one systematically leading. Once payoff structures are accounted for, the two markets provide coherent, mutually consistent views of shifting policy expectations.

This comparison illustrates the importance of understanding contract mechanics when interpreting prediction market data. Apparent disagreements may reflect structural differences rather than information discrepancies, and markets using different instruments can provide complementary perspectives on identical events.

3.4   Machine Learning Analysis

To test whether historical price patterns contain exploitable structure, seven machine learning models were trained to forecast next-day contract direction. Models ranged from linear classifiers (Logistic Regression, Ridge, LASSO) to ensemble methods (Random Forest, XGBoost, LightGBM) and neural networks. Input features included lagged returns (1-5 days), rolling volatility measures (3, 5, 10-day windows), and volume-based metrics.

Performance Results: Predictive performance proved weak across all markets. Bitcoin contracts achieved the highest accuracy (82.6%), but this result relies on only 23 test observations and exhibits high instability. Small sample size makes this metric unreliable—random chance could produce similar results. S&P 500 accuracy peaked at 64.3%, marginally above the 50% baseline for binary classification. The Fed contract's minimal test set rendered its metrics essentially meaningless for evaluation purposes.

ROC-AUC scores varied widely across models and contracts, frequently indicating overfitting. Models that performed well on training data showed substantial degradation on test sets, suggesting they captured noise rather than signal. This pattern appears consistently across asset classes, indicating the problem stems from fundamental unpredictability rather than inadequate model specification.
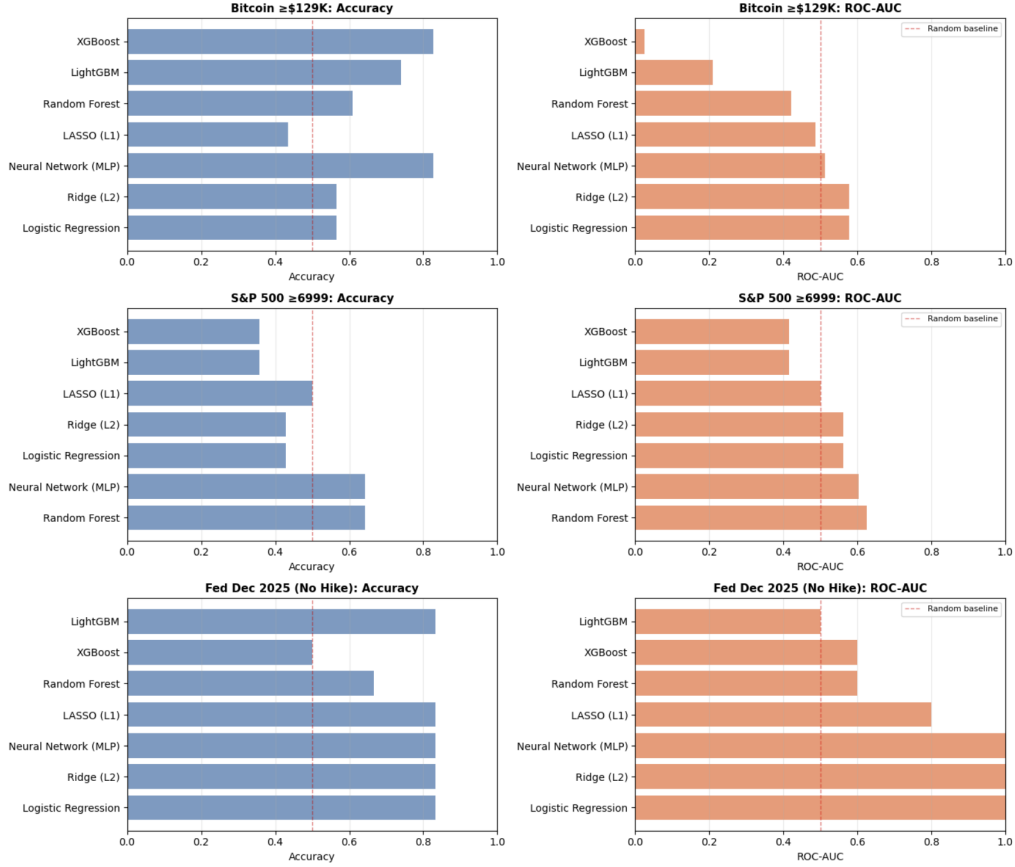
Figure 4: Machine learning model performance comparison across contracts. Wide variance in accuracy and ROC-AUC scores, particularly for Bitcoin (small test set) and Fed contracts (event-driven), indicates weak predictive power and substantial overfitting. S&P 500 results marginally exceed baseline but lack robustness for practical application.

Feature Importance: Certain features consistently ranked as important across models. Lagged returns (particularly 1-day lag) and short-horizon volatility (3-day rolling standard deviation) appeared most relevant. However, statistical importance did not translate into forecasting power. While these features correlate with price movements historically, they fail to provide reliable forward-looking signals.

This disconnect between feature importance and predictive performance is characteristic of efficient markets. Historical patterns that appear meaningful in-sample reflect how markets processed past information, not how they will respond to future events. Once information is incorporated into prices, residual patterns are too weak and unstable to support systematic forecasting.

Implications: The weak machine learning performance supports market efficiency. If simple historical features (returns, volatility, volume) sufficed for prediction, arbitrageurs would exploit these patterns until they disappeared. The observed lack of exploitability suggests Kalshi markets incorporate available information rapidly, leaving insufficient predictable structure for mechanical trading strategies.

For practitioners, results indicate that technical analysis and machine learning approaches offer limited value for prediction market trading. Patterns that appear robust in backtesting prove frag-

ile in forward testing. For researchers, findings suggest that improving forecast accuracy requires incorporating richer information sources—news sentiment, order flow, macroeconomic release timing—beyond simple price-based features.

4 Discussion

4.1 Market Efficiency Across Asset Classes

Evidence consistently supports efficient information incorporation across all three contracts. While efficiency manifests differently depending on asset class characteristics, no contract exhibits persistent exploitable patterns or systematic biases.

Bitcoin contracts demonstrate liquidity-driven dynamics. The one-day lag reflects practical constraints of a thinner prediction market responding to a highly liquid 24/7 underlying asset, rather than informational inefficiency. Participants rationally wait for confirmation of price movements before updating positions, avoiding overreaction to noise.

Equity contracts show near-instantaneous adjustment. Synchronous correlation at lag 0 indicates prediction market participants process S&P 500 movements in real time, updating probabilities as new information emerges. The absence of predictive power at negative lags confirms that participants do not systematically anticipate index movements.

Monetary policy contracts reflect event-driven information structure. Discrete jumps around scheduled announcements are economically sensible given that policy expectations change primarily when the Fed communicates or economic data surprises. Between events, expectations remain stable, producing the observed correlation pattern.

These asset-specific patterns demonstrate that efficiency does not require identical behavior across markets. Instead, efficient markets incorporate information appropriately given their structural constraints. The differences observed reflect how information flows in cryptocurrency, equity, and policy markets rather than indicating inefficiency.

4.2 Platform Consistency and Contract Design

The Kalshi-FedWatch comparison illustrates how contract structure affects probability representation without indicating market disagreement. Strong negative correlation stems from opposite payoff designs—Kalshi measures "no change" probability while FedWatch distributes probability across outcomes. When rate cuts appear likely, both markets agree: FedWatch shows high cut probability, Kalshi shows low "no change" probability.

This finding has practical implications. Researchers and practitioners must understand contract mechanics before interpreting cross-market comparisons. Apparent divergences may reflect measurement differences rather than information disagreements. Multiple markets can provide complementary perspectives on identical events without contradicting each other.

The consistency between platforms also validates prediction markets as information aggregation mechanisms. Despite different structures and participant bases, Kalshi and CME futures-based probabilities converge on similar assessments. Neither systematically leads the other, suggesting both incorporate available information efficiently.

4.3 Predictability and Trading Implications

Machine learning results reinforce that prediction markets, like traditional financial markets, do not offer easy profits through pattern recognition. Historical price data, even when processed through sophisticated models, provides limited forecasting power. This has several implications:

For traders: Technical strategies and machine learning approaches appear unpromising for prediction market speculation. Markets incorporate information sufficiently quickly that simple historical features cannot support systematic profitable trading. Participants likely need fundamental analysis—understanding event drivers, information sources, probability modeling—rather than relying on price patterns.

For Kalshi: Weak predictability indicates healthy market functioning. Prices adjust appropriately to new information without exhibiting persistent biases or exploitable inefficiencies. This validates the platform as a credible aggregation mechanism rather than a predictable system vulnerable to systematic exploitation.

For researchers: Improving prediction market forecasts likely requires incorporating richer data. News sentiment analysis, order flow metrics, macroeconomic release calendars, and social media signals may provide incremental information beyond price histories. However, even these sources face the fundamental challenge that markets incorporate publicly available information rapidly.

5 Limitations and Future Research

Several limitations warrant consideration when interpreting these results. First, sample sizes vary substantially across contracts, with particularly small test sets for Bitcoin and Fed contracts limiting statistical power. Results showing high accuracy or ROC-AUC scores for these contracts should be interpreted cautiously given potential instability.

Second, the analysis focuses on daily data, potentially missing higher-frequency patterns. Intraday analysis could reveal whether contracts lead or lag on shorter timescales, though practical trading costs and liquidity constraints likely limit the exploitability of any patterns discovered.

Third, machine learning models employ relatively simple features (returns, volatility, volume). More sophisticated features—derived from news sentiment, social media activity, or macroeconomic calendars—might improve performance, though efficiency hypothesis suggests public information should already be incorporated.

Fourth, the study examines only three contracts across specific time periods. Results may not generalize to all Kalshi markets or other prediction market platforms. Different contract types (sports, politics, weather) or different time periods (crisis vs. calm) could exhibit different dynamics.

Future research could address these limitations through several approaches. Expanding the sample to include more contracts and longer time series would improve statistical power and enable formal tests of market efficiency. Incorporating alternative data sources (news, social media, order flow) could test whether richer information improves forecasting. Comparing Kalshi with other prediction market platforms (Polymarket, PredictIt) would assess whether results reflect platform-specific features or general prediction market characteristics.

Event study methodology could examine how contracts respond to specific information releases, identifying whether reaction times vary by announcement type or market conditions. This would provide more granular evidence on information incorporation speed. Finally, examining contract

pricing around settlement dates could test for behavioral biases or calendar effects.

6   Conclusion

This study provides comprehensive evidence that Kalshi prediction markets function as efficient information aggregation mechanisms across cryptocurrency, equity, and monetary policy contracts. While efficiency manifests differently depending on asset class characteristics—liquidity-driven lags for Bitcoin, synchronous adjustment for equities, event-driven responses for policy—no contract exhibits persistent exploitable patterns or systematic predictability.

Lead-lag analysis reveals that contracts respond to rather than anticipate underlying market movements. Bitcoin contracts lag by one day due to liquidity differentials, S&P 500 contracts adjust synchronously, and Fed contracts react around discrete information events. Machine learning models achieve weak predictive performance despite testing seven algorithms with optimized features, confirming limited exploitable structure in historical price patterns.

Comparison with CME FedWatch demonstrates that apparent cross-platform divergence reflects contract design rather than information disagreement. Different probability measures can provide consistent assessments when structural differences are properly accounted for, validating prediction markets as credible aggregation mechanisms.

These findings have practical implications for multiple constituencies. Traders should recognize that prediction markets, like traditional financial markets, incorporate information efficiently and resist exploitation through simple pattern recognition. Kalshi can view these results as evidence of healthy market functioning without systematic biases. Researchers seeking to improve prediction market forecasts should focus on incorporating richer information sources beyond historical prices, while recognizing fundamental limits to predictability in efficient markets.

Most fundamentally, the analysis demonstrates that prediction markets aggregate dispersed information effectively across diverse asset classes and market structures. They provide a window into evolving market beliefs rather than a predictable system offering systematic trading opportunities. This dual function—information aggregation without exploitability—represents the hallmark of well-functioning markets and validates prediction markets as tools for understanding how participants assess uncertain future events.

References

Note: This report is based on analysis of Kalshi prediction market data. Interactive analysis and detailed methodology available in accompanying Jupyter notebook.

Kalshi Market Data. Available at: https://kalshi.com

CME FedWatch Tool. Available at: https://www.cmegroup.com/markets/interest-rates/cme-fedwatch-tool.html

---

For questions or collaboration opportunities, please contact via portfolio website or LinkedIn.