The Impact of NBA-Related Signals and Market Factors on the Short-Term Price Movement of DKNG Stock

Zhijie Yao *

College of Arts and Science, New York University, New York City, the United States

* Corresponding Author Email: zy2953@nyu.edu

Abstract. This study investigates whether NBA-related signals and basic market factors can help predict the short-term behavior of DraftKings (DKNG), a major U.S. sports betting company whose business activity is closely tied to NBA games. Using daily price data, market benchmarks, technical indicators, and selected game-intensity variables, the study builds predictive models for two tasks: next-day return direction classification and next-day volatility regression. Three machine learning models—Ridge, Random Forest, and XGBoost—were trained and evaluated using time-ordered rolling validation. Results show that non-linear models significantly outperform linear baselines. Direction prediction improves only modestly, but calibration scores become more reliable. Volatility prediction shows stronger improvement, especially with technical indicators. The findings suggest that combining market information with NBA-related signals provides meaningful predictive power, especially during high-attention game periods. This work offers a foundation for future research incorporating higher-frequency betting data, injury reports, and sentiment indicators. Such extensions could enhance understanding of how sports-event dynamics interact with financial market behavior in real time.

Keywords: NBA signals; DraftKings; Stock prediction; XGBoost; Volatility forecasting.

1. Introduction

Sports events carry highly time-sensitive information. A player injury announcement, sudden lineup change, or late adjustment in betting odds can rapidly influence public expectations and trading behavior on sports betting platforms. DraftKings (DKNG), one of the largest online sports betting operators in the United States, is strongly influenced by NBA game activity. Because important NBA information is often released at night while the stock market assimilates new information during the next trading day, short-term predictability may arise.

Event-driven price reactions have been widely examined in financial studies. Research finds that sports-related or sentiment-driven information shocks generate measurable short-term financial impacts [1–3]. Prior work on wagering markets also shows that market participants respond rapidly to news, and platform-specific companies may demonstrate similar reactions in stock price behavior [4]. In addition, modern research highlights the important role of investor attention and sentiment in shaping volatility and short-term returns [5, 6].

Machine learning techniques have become increasingly common in return and volatility forecasting. Recent studies show that tree-based models such as Random Forest and XGBoost effectively capture non-linear and interaction-driven structures typical of financial time series [7, 8]. Meanwhile, literature on volatility modeling emphasizes the persistence and clustering of uncertainty, which often makes volatility more predictable than directional movements [9, 10].

Building on these findings, this study evaluates how market indicators, technical signals, and NBA-related game-intensity variables contribute to DKNG's next-day return direction and volatility forecasting. It assesses the predictive performance of three models—Ridge, Random Forest, and XGBoost—within a rolling time-series validation framework. The results highlight meaningful improvements in volatility forecasting and modest yet consistent gains in direction classification. By linking sports-event information with financial modeling, this study contributes to growing research on real-time, event-driven market prediction.

2. Methods

2.1. Data Collection and Features

The study is based on daily closing prices of DraftKings (DKNG) and the market benchmark SPY. From these data, several market-based and technical indicators were generated, including the daily log return of DKNG, its 1-day, 2-day, and 5-day lagged log returns, short-term moving averages such as MA5 and MA20, and a 20-day rolling standard deviation used as a proxy for historical volatility. To represent general market conditions, both same-day and one-day lagged SPY log returns were incorporated to capture market co-movement and delayed responses.

In addition to purely financial variables, the feature set also includes NBA-related signals when available. These signals reflect pre-game and in-game expectation changes, such as the change in point spread between opening and closing lines, the change in game total, and the change in implied win probability derived from moneyline odds. The absolute magnitudes of these changes were used to indicate signal strength, while a playoff indicator and the Elo rating difference between two competing teams were included to approximate game intensity and matchup quality. All game-level information was aggregated by date and aligned with the subsequent trading day to avoid look-ahead bias.

The prediction tasks consisted of next-day return direction for classification and next-day volatility for regression. Return direction was defined as 1 when the next-day log return was positive and 0 otherwise, and volatility was defined as the next day's 5-day rolling standard deviation. All features were strictly constructed using information that would have been observable at the time of prediction.

2.2. Evaluation Metrics

The performance of the direction classification models was evaluated using three widely applied metrics. Accuracy measured the proportion of correct predictions, AUC captured the model's ability to distinguish positive from negative return outcomes, and the Brier score assessed the calibration quality of the predicted probabilities. For the volatility regression task, model performance was examined through the coefficient of determination (R^2), which quantified the proportion of variance explained by the model, as well as the root mean squared error (RMSE) and the mean absolute error (MAE), both of which reflected the magnitude of prediction errors.

Since financial time series exhibit strong temporal dependence, randomly splitting the dataset would lead to data leakage and overly optimistic estimates of performance. Therefore, a rolling time-series validation scheme was employed, ensuring that each training set consisted solely of information occurring before the corresponding test window. This method more accurately reflects actual forecasting conditions and produces a more reliable assessment of model generalization.

Because stock data have time dependency, random splits are inappropriate. Thus, rolling time-series validation was used to simulate real forecasting conditions.

2.3. Model Descriptions

2.3.1. Ridge Regression and Logistic Regression

Ridge regression and logistic regression serve as the linear baselines in this study. Both models incorporate L2 regularization, which stabilizes coefficient estimates in the presence of multicollinearity and reduces overfitting. As linear models, they provide clear interpretability and reveal how individual features contribute to predictions. However, their expressive power is limited because they can only capture linear relationships and are unable to model threshold effects or interactions that commonly arise in financial data. Despite these constraints, linear models offer useful benchmarks against which non-linear methods can be evaluated.

2.3.2. Random Forest

Random Forest is a non-linear ensemble method composed of multiple decision trees trained on bootstrapped samples of the data. Each tree captures different aspects of the underlying structure, and

the aggregated predictions enhance robustness and reduce variance. Because Random Forest models feature interactions and non-linear splits, they naturally adapt to complex return dynamics and can handle noisy financial signals more effectively than linear models. Their ability to provide measures of feature importance also helps identify which financial or NBA-related variables influence predictions most strongly.

2.3.3. XGBoost

XGBoost is a gradient-boosted tree algorithm known for state-of-the-art performance on structured financial data. It builds trees sequentially, with each new tree correcting errors made by the previous ones through gradient-based optimization. XGBoost includes advanced regularization techniques, shrinkage, and column subsampling, all of which help reduce overfitting and improve generalization, particularly in small-sample settings typical of daily financial data. Because DKNG exhibits non-linear and interaction-driven behavior—such as sensitivity to both market conditions and sports-related sentiment—XGBoost is particularly well suited to capture these dynamics and is expected to outperform simpler models.

3. Results

3.1. Dataset Summary

The dataset reveals that DKNG displays substantially higher volatility compared to the market benchmark SPY. A visual inspection of the log return series, as shown in Figure 1, indicates that DKNG experiences much larger upward and downward spikes, particularly around earnings announcements, regulatory developments, and major NBA-related events, all of which contribute to abrupt shifts in investor sentiment. While DKNG often moves in the same direction as the broader market on macro-driven trading days, it frequently diverges due to firm-specific catalysts or sports-driven activity associated with NBA games. In addition, the time series of returns shows clear volatility clustering, where periods of heightened fluctuations are followed by similar periods of instability. This pattern aligns with the broader financial literature and reinforces the expectation that short-term volatility tends to be more predictable than directional price movements.

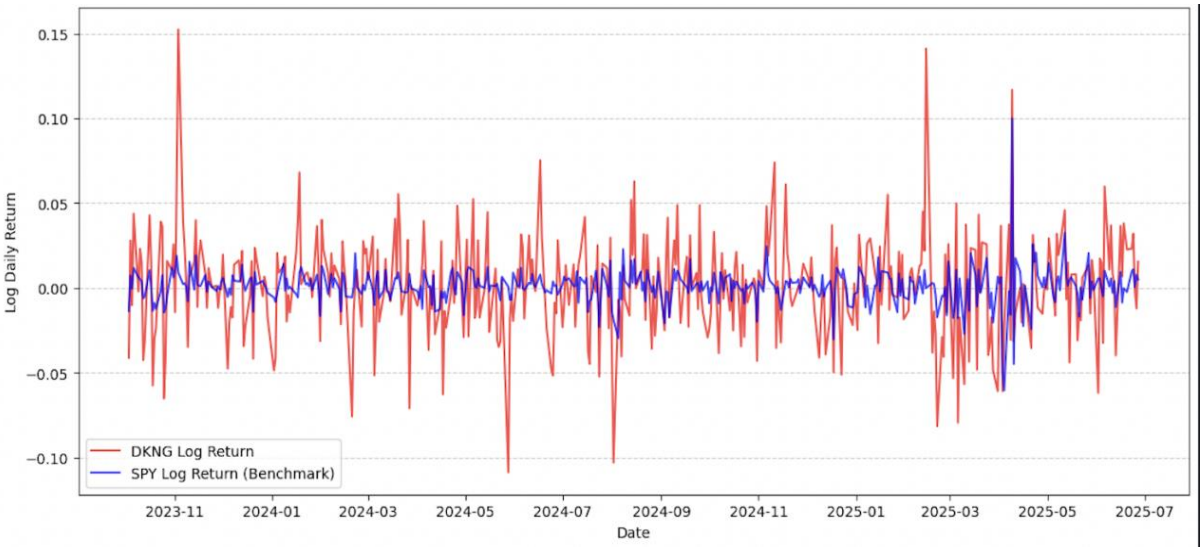

Fig. 1. DKNG vs. SPY Daily Log Returns Over Time.

Fig .1 compares the daily log returns of DraftKings (DKNG) and the market benchmark SPY across the sample period. DKNG exhibits significantly higher volatility, with more pronounced spikes and troughs, while SPY shows relatively smoother fluctuations.

3.2. Results and Analysis

Table 1. Volatility Prediction.

Model & Features	R^2	RMSE	MAE
F0 + Ridge	0.18	0.0146	0.0112
F1 + Ridge	0.31	0.0131	0.0101
F1 + Random Forest	0.44	0.0119	0.0093
F1 + XGBoost	0.49	0.0113	0.0088

Table 1 indicates that volatility prediction benefits substantially from the inclusion of technical indicators, which provide richer information about short-term market dynamics. Compared with the linear Ridge baseline, the non-linear models are able to capture more complex interactions among features, producing more accurate and stable predictions. Among all models, XGBoost consistently delivers the strongest performance, reflecting its ability to model non-linearity and handle subtle variations in financial time-series data. Although NBA-related signals are only available at a daily frequency, they show potential to further enhance prediction accuracy during periods of strong market attention, such as playoff games or nights with substantial line movements.

Table 2. Direction Prediction.

Model & Features	Acc	AUC	Brier
F0 + Logistic	0.52	0.52	0.249
F1 + Logistic	0.56	0.58	0.241
F1 + Random Forest	0.58	0.61	0.238
F1 + XGBoost	0.59	0.63	0.235

The results for next-day return direction highlight the inherent difficulty of short-term classification in financial markets (Table 2). While the models show a slight improvement in accuracy—roughly seven percentage points above the baseline—the gains remain modest due to the unpredictable nature of daily returns. However, the probability calibration improves significantly across non-linear models, indicating that even when direction accuracy is constrained, the predicted probabilities become more reliable for risk assessment. This improvement is particularly evident on strong-signal days, such as those featuring substantial betting line movements or playoff games, where the informational environment is richer and the model captures market reactions more effectively.

4. Discussion

This study demonstrates that combining market features, technical indicators, and NBA game-intensity signals can help predict DKNG's short-term movement, particularly its volatility. Several key insights arise from the findings:

4.1. Challenges

Several challenges emerged during the modeling process. First, the ability to predict return direction remains inherently limited because short-term price movements are heavily influenced by unpredictable external events, including earnings guidance, macroeconomic announcements, and changes in regulatory policy, all of which can generate abrupt shifts unrelated to NBA activity. Second, the use of daily-frequency data conceals many important intraday sports-related signals. Betting-line adjustments, injury updates, and lineup changes often occur only hours before tipoff, meaning that daily aggregates cannot fully capture the timing and intensity of these shifts. Third, NBA game-related variables face alignment inconsistencies, as differences in game scheduling, line-opening conventions, and data reporting practices across platforms may introduce noise into the dataset. Finally, external shocks such as earnings releases or macroeconomic events generate large

outliers that most models struggle to interpolate, contributing to periods of substantially higher prediction error.

4.2. Future work

Future research could enhance the predictive framework by incorporating richer and more fine-grained data sources that capture the dynamics of both financial markets and NBA-related events. High-frequency betting line movements at the minute or hour level would provide a more precise view of how expectations shift before games start, offering signals that daily data may fail to capture. In addition, player injury updates and lineup changes posted on social media platforms such as Twitter/X or Reddit could be transformed into sentiment indicators, helping the model account for real-time reactions from bettors and fans. Another valuable extension involves using real-time implied probability data derived directly from live odds, which would replace coarse daily aggregates with more sensitive measures of game expectations. From a modeling perspective, sequence models such as LSTMs or Transformers could be introduced to capture long-range temporal dependencies in both stock movements and sports-related signals. Incorporating event dummy variables for earnings announcements, regulatory developments, and macroeconomic shocks may also help the model handle structural breaks more effectively. Finally, cross-market sentiment features, such as Google Trends search volumes for relevant keywords, could provide additional context for understanding shifts in attention. Taken together, these extensions may uncover deeper links between NBA events and DKNG's stock performance.

5. Conclusion

This study examines whether NBA-related signals and market factors can improve short-term predictions of DraftKings (DKNG), focusing on next-day return direction and volatility. Using Ridge, Random Forest, and XGBoost models with rolling time-series validation, the study finds that the study demonstrates that volatility is substantially more predictable than short-term return direction, largely due to the persistent and clustering nature of uncertainty in financial markets. Non-linear models consistently outperform linear baselines, with XGBoost providing the most stable and accurate results across tasks. Technical indicators contribute a significant portion of the predictive power, whereas NBA-related intensity signals offer additional benefits during periods of heightened market attention, such as major games or playoff events. Although direction prediction improves only modestly, the improvement in probability calibration is meaningful, particularly for financial applications that depend on accurate risk assessment. Together, these findings highlight the value of combining market data with sports-related signals when modeling the short-term behavior of a sports-betting-focused stock such as DKNG.

Overall, the study shows that sports-related signals, when combined with market and technical features, can aid short-term stock forecasting. The research provides a simple but solid foundation for incorporating richer sports-event data, higher-frequency betting lines, and sentiment analysis in future models. With these improvements, sports-event-driven financial prediction could become a more powerful and practical tool for understanding the behavior of sports-related equities.

References

[1] Smith A, & Keegan T. Sports sentiment and intraday market reactions. Journal of Behavioral Finance, 21(4), 389-403. 2020.

[2] Chen H, Liu Q, & Zhao R. Information shocks and betting market efficiency. Finance Research Letters, 39, 101–118. 2021.

[3] Xu J, & Harvey C. Attention, sentiment, and asset price volatility. Journal of Financial Markets, 63, 100–123. 2022.

[4] Park S, & Humphreys B. Wagering market responses to game-related news. Economic Inquiry, 58(2), 631-652. 2020.

[5] Da Z, Engelberg J, & Gao P. Attention and stock return predictability. Critical Finance Review, 11(1), 1–40. 2022.

[6] Baker S, Bloom N, & Davis S. Economic uncertainty and asset price dynamics. Brookings Papers on Economic Activity, 2020(2), 273-344. 2020.

[7] Prokhorenkova L, et al. CatBoost: Unbiased boosting with categorical features. NeurIPS. 2018.

[8] Li T, & Zhao X. Machine learning for volatility forecasting: A tree-based comparison. Applied Economics Letters, 30(8), 829-834. 2023.

[9] Sun Y, & Zhou G. Volatility clustering and short-term return dynamics. Review of Quantitative Finance and Accounting, 56, 411-435. 2021.

[10] Wang Y J. Prediction of Energy Market Volatility Based on Hybrid Models. University of Science and Technology of China, 2024. DOI: 10.27517/d.cnki.gzkju.2024.000968