Vol. 19 (2026): Proceedings of 2026 International Conference on Financial Technology, Mathematical Statistics, and Business Analytics (ICFMB 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/icfmb.202619290295

# FinTech-based Joint Pricing of Options and Stocks: Data Quality, Mathematical Modeling, and Visualization Framework

**Zhengxuan Qu***

*School of University College Dublin, Institute of Business and Management, Dublin, Ireland*

**Corresponding author: Zhengxuan Qu*

## Abstract

With the rapid development of FinTech technologies such as big data analytics, cloud computing, and algorithmic modeling, joint pricing of options and stocks faces new opportunities and challenges. Traditional option pricing methods often underperform in high-dimensional, noisy, and asynchronous market data. This paper proposes a FinTech-based joint pricing framework, encompassing data quality control, robust modeling methods, and result visualization strategies. We analyze common data collection problems, such as missing values, inconsistent formats, time asynchrony, and high-frequency noise, and propose corresponding data cleaning, feature engineering, and modeling processes, including classic parametric models, machine learning models, and hybrid models. Finally, this paper discusses visualization methods, such as time series plots, implied volatility surfaces, and residual analysis, to support model validation and risk management. This framework provides academic researchers and investment practitioners with a systematic and reproducible joint pricing process.

### Keywords

FinTech, joint pricing, options, stocks, data cleaning

## 1. Introduction

FinTech, through core technologies such as big data analytics, cloud computing infrastructure, and algorithmic modeling, is profoundly reshaping asset pricing in capital markets. These technologies can process massive amounts of market data, support complex stochastic modeling, and achieve near real-time analysis, thereby improving the accuracy of derivative valuations and enhancing our understanding of real-world market behavior [1-3].

Options, as derivatives closely related to the value of underlying stocks, are particularly sensitive to market dynamics. Traditional pricing models, such as the Black-Scholes model, rely on simplifying assumptions such as constant volatility, frictionless markets, and continuous trading [4]. However, empirical studies show that these assumptions often do not hold true. Research using high-frequency data has revealed that markets exhibit volatility fluctuations, price jumps, trading frictions, data asynchrony, and microstructural noise, all of which significantly impact asset and option pricing [5].

To address these real-world market characteristics, recent research has proposed pricing models that combine realized volatility, stochastic volatility, and price jumps, while also considering microstructural noise in high-frequency trading environments. For example, the Realized GARCH model utilizes high-frequency market data to improve the accuracy of volatility prediction and option pricing [6]; the BARVJ model, by incorporating overnight information and price jumps, reduces option pricing errors by approximately 20% compared to traditional models. Furthermore, stochastic volatility models employing fractional Brownian motion or double exponential jumps can better capture market characteristics such as implied volatility smiles, long-term memory effects, and rough volatility [3, 5]. These studies demonstrate that FinTech-driven data analytics and algorithmic modeling play a crucial role in overcoming the limitations of traditional option pricing methods, making asset valuations more realistic and robust under complex market conditions.

## 2. Literature Review

### 2.1 Classical and Stochastic Models

The Black–Scholes model (1973) provides a closed-form solution for European option pricing, if the logarithmic return of the underlying asset follows a normal distribution and volatility is constant. For path-dependent or early exercising options, a binomial tree model can be used for numerical approximation [7].

Due to the characteristics of jumps, heavy tails, and volatility smiles in real markets, scholars have proposed jump-diffusion models [8]and two-factor stochastic volatility jump models (2FSVJ) to describe market behavior and provide pricing formulas more realistically. While these models improve realism, they increase mathematical complexity, model calibration difficulty, and computational burden, especially with high-dimensional or high-frequency data [7].

### 2.2 Machine Learning and Data-Driven Approaches

With increased computing power and data availability, machine learning (ML) and deep learning (DL) methods are widely used in option pricing. ML algorithms do not require strict assumptions about the stochastic processes of the underlying asset and can learn complex nonlinear relationships between features and option prices or implied volatility [9, 10].

Research shows that advanced network architectures (such as Highway Networks and Transformers) outperform simple feedforward networks in pricing and implied volatility prediction and can serve as efficient alternatives to traditional analytical models [11]. Some studies have incorporated physical or financial constraints into neural networks to further improve the accuracy of volatility surface prediction [10].

### 2.3 Hybrid Models and Reviews

Reviews emphasize the importance of machine learning (ML), neural networks, and hybrid models in option pricing. In hybrid models, parametric models (such as Black-Scholes or Heston) are combined with nonparametric ML methods to capture residuals or theoretical assumption biases. Meanwhile, deep learning is used to solve the pricing mechanism of high-dimensional BSDEs, outperforming traditional models in pricing complex derivatives [12, 13].

The literature shows that option pricing is shifting from purely theoretical models to data-driven, hybrid, and ML-enhanced frameworks. However, systematic research on the complete process (data quality, preprocessing, modeling, and visualization) is still lacking, which is the motivation for this study [14].

## 3. Methodology

### 3.1 Data Preprocessing and Quality Control

In the joint pricing of options and stocks, data is typically merged from multiple sources, which often leads to several data quality issues. Missing values commonly occur for illiquid option contracts, such as deep out-of-the-money options with unavailable quotes; if left untreated, they may bias estimation results or reduce sample size and are therefore handled through deletion or interpolation. Heterogeneous data formats arise because variables such as interest rates, dividends, and implied volatility are reported using different conventions across data providers; failing to standardize these formats may introduce systematic pricing errors, making numerical transformation and unification necessary. Time asynchrony between high-frequency stock prices and sparsely observed option quotes can generate spurious noise if ignored and is typically addressed through timestamp matching or resampling. In addition, high-frequency noise, including bid–ask spreads and short-term price fluctuations, may distort observed prices and impair model stability, particularly for machine learning models; thus, mid-quote prices or smoothing techniques are often applied. Finally, cleaned data is normalized and transformed into relevant features-such as historical volatility, time to maturity, and moneyness-to ensure reliable and consistent inputs for both traditional pricing models and machine learning approaches [15, 16].

### 3.2 Modeling Strategy

Based on the research objectives and data characteristics, this paper adopts a comparative modeling strategy that incorporates classical parametric models, machine learning models, and hybrid approaches. Classical option pricing models, including the Black–Scholes model, stochastic volatility models, and jump-diffusion models, are first employed due to their strong theoretical foundations and economic interpretability. Despite its simplifying assumptions, the Black–Scholes model remains a widely used benchmark in empirical studies, as it provides a transparent reference point against which pricing biases and model improvements can be clearly evaluated.

Machine learning and deep learning models, such as feedforward neural networks, high-speed networks, and physics-informed neural networks, are then considered for their ability to capture highly nonlinear relationships between input features and option prices or implied volatility without imposing restrictive parametric assumptions. However, while these models often exhibit strong in-sample fitting and predictive performance, their limited interpretability and potential overfitting raise concerns regarding generalization, especially in sparse or noisy option markets.

To address the limitations of both approaches, this study further considers hybrid models that combine classical pricing outputs with machine learning-based residual corrections or BSDE-based formulations. In the context of joint pricing of stocks and options, such hybrid frameworks preserve theoretical consistency across assets while allowing flexible adjustments for empirical deviations, thereby achieving a balance between economic interpretability and predictive accuracy [17].

### 3.3 Diagnosis and Visualization

Diagnostic and visualization tools mainly include time series graphs, used to compare actual and predicted prices, observe co-movement, volatility clustering, and potential market shifts; implied volatility surfaces, used to show the relationship between strike price and expiration date and implied volatility, thereby identifying smiles or skewness; residual analysis, using scatter plots, histograms, or heatmaps to detect model bias and heteroscedasticity; and feature importance analysis (for machine learning or mixture models), used to identify key variables affecting predictions, such as volatility, out-of-the-money value, expiration date, and trading volume. These visualization methods can effectively assist in result interpretation, model validation, and risk assessment [6, 18].

## 4. Conceptual Empirical Design

Assuming the use of the S&P 500 index and its option chain data, this study considers features including underlying asset price, strike price, time to expiration, historical volatility, implied volatility, bid-ask spread, trading volume, dividends, interest rates, out-of-the-money status, and volatility surface parameters. The models employed include the classic 2FSVJ model, feedforward networks or high-speed networks, and hybrid models (classical model output with ML residual correction). Evaluation metrics include root mean square error (RMSE), mean absolute error (MAE), out-of-sample predictive power, computational speed, and model stability. The hypothetical results show that hybrid models and machine learning models outperform classic models in predictive accuracy, especially in out-of-the-money options or periods of volatility aggregation.

Furthermore, volatility surface and residual analysis can reveal error-prone option ranges, providing a reference for risk management [10, 19].

## 5. Discussion

The main advantage of this framework lies in its ability to improve data consistency and robustness. By preprocessing and synchronizing multi-source, high-frequency, and heterogeneous data, it makes the input information more reliable, thereby enhancing the overall stability of the model. Simultaneously, the framework offers modeling flexibility and adaptability, capable of combining classical parametric models, machine learning/deep learning models, and hybrid models to select the most suitable pricing method based on different option types and market conditions. Furthermore, the framework exhibits good scalability and reproducibility, facilitating replication experiments and comparative analyses across different markets, asset classes, or time periods. Interpretability is also a significant advantage; through residual analysis, feature importance assessment, and volatility surface visualization, it can assist in risk management and strategy decision-making, providing actionable reference information for traders and risk managers.

However, this framework also faces several challenges in practical applications. Data limitations are a key issue, including missing or illiquid option contracts, asynchronous trading data, and high-frequency market noise. Data imputation or smoothing may also introduce systematic biases. Model generalization ability is also a bottleneck; machine learning and deep learning models may perform well on training data, but their predictive ability declines under extreme market conditions, posing a risk of overfitting. Furthermore, the framework needs to strike a trade-off between interpretability and performance: more complex models may improve prediction accuracy but reduce interpretability. Computational cost is also a significant factor; training and predicting high-dimensional data and complex models requires substantial computational resources, limiting its application in real-time trading or high-frequency strategies. Overall, the framework shows significant potential in improving option pricing accuracy, capturing complex market behavior, and supporting risk management, but trade-offs and optimizations are still needed regarding data quality, model selection, and computational resources [20].

## 6. Conclusions and Future Research

This study constructs and validates a financial technology-based option-equity co-pricing framework, systematically integrating data quality management, feature engineering, modeling techniques, and visualization analysis to provide asset pricing research with a process that combines theoretical depth and operational feasibility. The study emphasizes the fundamental role of high-quality data cleaning and time synchronization in co-pricing, as stock and option data often originate from different sources, exhibiting issues such as frequency mismatch, missing values, and noise. Inadequate handling of these issues directly impacts model stability and predictive performance. Furthermore, this paper compares the performance of traditional models with machine learning and deep learning models. Results show that ML and hybrid models significantly outperform classic methods such as Black-Scholes when dealing with complex phenomena such as nonlinear structures, volatility clustering, and market jumps. Python-based visualization diagnostics help researchers understand price dynamics, error structures, and risk exposure from multiple dimensions, contributing to the scientific rigor of quantitative strategies and risk control systems.

Future research directions remain broad. Firstly, large-scale empirical testing can be conducted by introducing richer cross-asset real-world market data (such as commodities, foreign exchange, and interest rate options) to improve model universality. Secondly, stress tests should be conducted under different market conditions to assess the model’s robustness during crises or in highly volatile environments. Thirdly, the research can be expanded to more complex derivatives, such as pricing problems involving path dependence, early exercise, or multiple factors. Fourthly, it is recommended to integrate interpretable machine learning methods, such as SHAP and LIME, to enhance the model’s transparency and compliance in real-world investment decisions. Finally, this research framework can be further developed into a real-time pricing and risk management system, providing technical support for quantitative trading, market making, and portfolio management, thereby promoting the deep application of fintech in asset pricing.

## References

[1] Fang, Z. and Han, J.-Y. Realized GARCH Model in Volatility Forecasting and Option Pricing. Computational Economics. 2025, 66(5), pp. 3637-3657. https://doi.org/10.1007/s10614-024-10826-8.

[2] Chong, C. H. and Todorov, V. Asymptotic expansions for high-frequency option data. The Annals of Applied Probability. 2025, 35(3), pp. 1941-1979. https://doi.org/10.1214/25-AAP2163.

[3] Zhang, S., Yong, H. and Xiao, H. Option Pricing with Fractional Stochastic Volatilities and Jumps. Fractal and Fractional. 2023, 7(9), p. 680. https://doi.org/10.3390/fractalfract7090680.

[4] Black, F. and Scholes, M. The Pricing of Options and Corporate Liabilities. Journal of Political Economy. 1973, 81(3), pp. 637-654. https://doi.org/10.1086/260062.

[5] Agazzotti, G., Aguilar, J.-P., Aglieri Rinella, C. and Kirkby, J. L. Calibration and option pricing with stochastic volatility and double exponential jumps. Journal of Computational and Applied Mathematics. 2025, 465, p. 116563. https://doi.org/10.1016/j.cam.2025.116563.

[6] Chen, S. and Zhang, Z. Forecasting implied volatility smile surface via deep learning and attention mechanism. arXiv preprint arXiv:1912.11059. 2019. https://doi.org/10.48550/arXiv.1912.11059.

[7] El-Khatib, Y., Makumbe, Z. S. and Vives, J. Approximate option pricing under a two-factor Heston–Kou stochastic volatility model. Computational Management Science. 2023, 21(1), p. 3. https://doi.org/10.1007/s10287-023-00486-8.

[8] Kou, S. G. A Jump-Diffusion Model for Option Pricing. Management Science. 2002, 48(8), pp. 1086-1101. https://doi.org/10.1287/mnsc.48.8.1086.166.

[9] Liu, S., Oosterlee, C. W. and Bohte, S. M. Pricing Options and Computing Implied Volatilities using Neural Networks. Risks. 2019, 7(1), p. 16. https://doi.org/10.3390/risks7010016.

[10] Ding, L., Lu, E. and Cheung, K. Deep Learning Option Pricing with Market Implied Volatility Surfaces. arXiv preprint arXiv:2509.05911. 2025. https://doi.org/10.48550/arXiv.2509.05911.

[11] Van Mieghem, L., Papapantoleon, A. and Papazoglou-Hennig, J. Machine learning for option pricing: an empirical investigation of network architectures. arXiv preprint arXiv:2307.07657. 2023. https://doi.org/10.48550/arXiv.2307.07657.

[12] Aboussalah, A. M., Li, X., Chi, C. and Patel, R. The AI Black-Scholes: finance-informed neural network. arXiv preprint arXiv:2412.12213. 2024. https://doi.org/10.48550/arXiv.2412.12213.

[13] Liu, D. Studies of issues on hybrid neural networks for pricing financial options. Artificial Intelligence and Applications. 2025, 3(3), pp. 331-339. https://doi.org/10.47852/bonviewAIA52023377.

[14] Ruf, J. and Wang, W. Neural networks for option pricing and hedging: a literature review. arXiv preprint arXiv:1911.05620. 2019. https://doi.org/10.48550/arXiv.1911.05620.

[15] Visagie, J. A statistical technique for cleaning option price data. arXiv preprint arXiv:2501.11164. 2025. https://doi.org/10.48550/arXiv.2501.11164.

[16] Andersen, T., Archakov, I., Grund, L., Hautsch, N., Li, Y., Nasekin, S., Nolte, I., Pham, M. C., Taylor, S. and Todorov, V. A Descriptive Study of High-Frequency Trade and Quote Option Data*. Journal of Financial Econometrics. 2021, 19(1), pp. 128-177. https://doi.org/10.1093/jjfinec/nbaa036.

[17] Shi, Y., Teng, B. and Wang, S. Option pricing mechanisms driven by backward stochastic differential equations. Financial Innovation. 2025, 11(1), p. 90. https://doi.org/10.1186/s40854-024-00714-3.

[18] Luo, J. and Yu, C. L. The Application of Symbolic Regression on Identifying Implied Volatility Surface. Mathematics. 2023, 11(9), p. 2108. https://doi.org/10.3390/math11092108.

[19] Chen, H., Didisheim, A. and Scheidegger, S. Deep structural estimation: With an application to option pricing. arXiv preprint arXiv:2102.09209. 2021. https://doi.org/10.48550/arXiv.2102.09209.

[20] Qiu, Z., Kownatzki, C., Scalzo, F. and Cha, E. S. Historical Perspectives in Volatility Forecasting Methods with Machine Learning. Risks. 2025, 13(5), p. 98. https://doi.org/10.3390/risks13050098.

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.