Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD36678

Individual Stock Versus Index: A Comparison of Black-Scholes Calibration and Delta Hedging on TSLA and SPY (2019–2022)

Yifan Guo

Department of Mathematics, London School of Economics and Political Science, London, UK
Y.Guo71@lse.ac.uk

Abstract. The Black–Scholes option pricing model assumes that the volatility of the underlying asset is constant, yet observed market option prices imply a volatility that changes with the strike price. This study investigates whether adjusting the volatility input to match market prices improves option pricing and delta hedging, and whether the benefit differs between a high-volatility individual stock and a broad market index. Daily option data for Tesla and the SPDR S&P 500 exchange-traded fund from 2019 to 2022 are used, covering calm markets, the 2020 market crash, and the subsequent monetary tightening. Two settings are compared: one using historical volatility estimated from past returns, and one using an implied volatility obtained by fitting the Black–Scholes price to market option prices through least squares. Pricing accuracy is assessed by the pricing residuals across strike prices, and hedging effectiveness by the variability of the hedging profit and loss over forty-eight monthly positions. The results show that a single fitted volatility cannot match option prices at all strike prices, confirming that the constant-volatility assumption does not hold. Using the fitted volatility reduces the variability of the hedging profit and loss by about twelve percent for the individual stock and twenty-three percent for the index. The larger improvement for the index indicates that the value of calibration depends on the volatility structure of the underlying asset.

Keywords: Black–Scholes, stock analysis, delta hedging

1. Introduction

Options are among the most widely used financial derivatives nowadays, and play an important role in investment decision-making and risk management. For the pricing of European options, the Black–Scholes (BS) model [1] is one of the most fundamental and widely recognized approaches. BS formula lays a solid foundation for modern derivatives markets, which provides a comprehensive framework for option pricing. A key assumption underlying the Black-Scholes model is that the volatility of the underlying asset is constant throughout the life of the option.

Among the inputs of the Black-Scholes model, volatility is the only one which cannot be observed directly from the market. Therefore, how to determine the volatility becomes quite significant. Two measures are commonly used in practice [2]: Historical Volatility (HV) uses the

DOI: 10.54254/2754-1169/2026.LD36678

past prices of the underlying asset, while Implied Volatility (IV) is backed out from the market prices of options. Calibration refers to the process of adjusting the volatility parameter so that the BS prices match the market prices. In addition, choice of the volatility also directly influences the process of delta hedging, which depends on the model-implied sensitivity of the option price to the underlying.

In its standard application, the Black–Scholes model is implemented with a single constant volatility for a given underlying. However, the assumption of constant volatility does not hold in practice. Empirical studies have shown that the implied volatilities backed out from options with different strike prices are not constant, but instead form a pattern known as the volatility smile [3]. This indicates that a single volatility parameter cannot simultaneously match the market prices of all options on the same underlying asset. To obtain a volatility input that better reflects the market, calibration is widely used to fit the model to observed option prices. Prior research has further compared the effectiveness of implied and historical volatility in delta hedging, generally finding that implied-volatility-based hedging provides more stable results [4]. However, most existing studies focus on a single underlying asset, and it is still unclear whether the benefit of calibration is consistent across assets with different volatility characteristics, such as a high-volatility individual stock and a broad market index.

In this regard, this paper investigates whether the benefit of calibration differs between a high-volatility individual stock and a broad market index. TSLA was selected as a representative high-volatility individual stock and SPY as a proxy for the broad market index. Using daily option data from 2019 to 2022, a period that covers calm markets, the COVID-19 crash, and the subsequent monetary tightening, this study compares the pricing and delta-hedging performance before calibration and after calibration [5]. Options are priced with the Black–Scholes formula; the pre-calibration case uses historical volatility as the volatility input, whereas the post-calibration case uses an implied volatility obtained by fitting the Black–Scholes price to observed market option prices. The hedging performance is evaluated by the variability of the hedging profit and loss.

2. Method

2.1. Dataset preparation

The primary data used in this study are End-Of-Day (EOD) option chains for TSLA and SPY, which are obtained from publicly available datasets on the Kaggle platform [6, 7]. Each record contains the quote date, expiration date, strike price, bid and ask quotes for both calls and puts, implied volatility, and the full set of Greeks. TSLA is selected as a representative high-volatility individual stock, whereas SPY, an exchange-traded fund tracking the S&P 500, serves as a proxy for the broad market index. This contrast allows the effect of calibration to be examined across two underlyings with different volatility characteristics.

The sample covers the period from January 2, 2019 to December 30, 2022. After aligning the trading days common to both underlyings, 1,006 trading days remain. The daily closing price of each underlying is extracted directly from the UNDERLYING_LAST field of its own option data, rather than from an external source, since this guarantees that the underlying price and the option quotes are from the same source and are aligned by date. A uniqueness check confirms that the UNDERLYING_LAST value is unique on every trading day for both underlyings, so the closing-price series can be obtained by deduplication.

The data are preprocessed in several steps. First, invalid quotes are removed, including options with zero or missing trading volume and those with abnormal bid–ask spreads. Second, a moneyness

DOI: 10.54254/2754-1169/2026.LD36678

filter is applied to exclude deep in- and out-of-the-money contracts that are illiquid, retaining the range [0.4, 2.0] for TSLA and [0.6, 1.5] for SPY; the wider range for TSLA reflects its higher volatility and broader moneyness distribution. After cleaning, the dataset contains approximately 1.56 million option records for TSLA and 2.89 million for SPY. Finally, because TSLA underwent a 5-for-1 stock split in August 2020 and a 3-for-1 split in August 2022, its underlying price series is forward-adjusted into a continuous series for computing historical volatility and hedging profit and loss, while the option chain retains its original scale for daily pricing. The adjusted series matches the external stock data with an average relative error of 0.03%, confirming the correctness of the adjustment. Figure 1 shows the distribution of moneyness and the corresponding implied volatility for the two underlyings over the sample period.

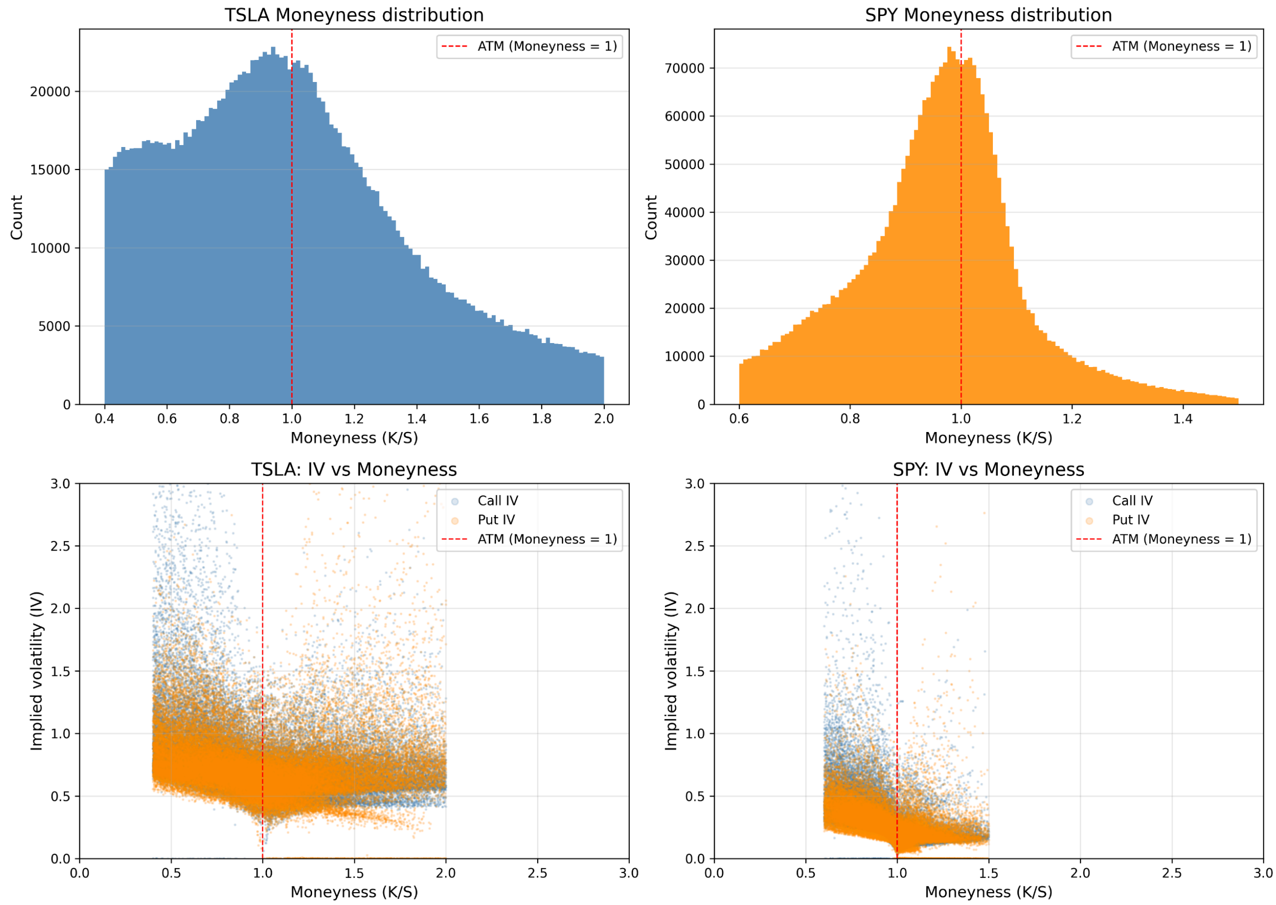

Figure 1. Moneyness distribution (top) and implied volatility versus moneyness (bottom) for TSLA and SPY, 2019–2022 (picture credit: original)

2.2. Black-scholes model

The Black–Scholes (BS) model prices a European option under the assumptions that the underlying price follows a geometric Brownian motion with constant volatility and constant risk-free rate, and that the market is frictionless and free of arbitrage. Under these assumptions, the price of a European call option has the closed-form expression as follows [1]

$$C = S \cdot N(d_1) - K \cdot e^{-rT} \cdot N(d_2) \tag{1}$$

DOI: 10.54254/2754-1169/2026.LD36678

$$d_1 = \frac{\left[\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) \cdot T\right]}{\left(\sigma \cdot \sqrt{T}\right)}, \quad d_2 = d_1 - \sigma \cdot \sqrt{T} \tag{2}$$

where S is the current price of the underlying, K is the strike price, T is the time to maturity (in years), r is the risk-free rate, σ is the volatility, and N( ) denotes the cumulative distribution function of the standard normal distribution.

The sensitivity of the option price to the price of the underlying, known as the delta, is given by [1]

$$\Delta = \partial C / \partial S = N(d_1) \tag{3}$$

The delta plays a central role in the delta-hedging strategy discussed in Section 2.4. Among all the inputs of the model, only the volatility σ cannot be observed directly from the market. The choice of this input—historical volatility or calibrated implied volatility—is the central question addressed in this study.

2.3. Volatility measures

This study compares two types of volatility inputs for the Black–Scholes model: historical volatility, used before calibration, and calibrated implied volatility, used after calibration.

2.3.1. Historical volatility

Historical Volatility (HV) is estimated from the past returns of the underlying and serves as thevolatility input before calibration. The daily log return is defined as follows [8].

$$r_t = \ln\left(\frac{S_t}{S_{\{t-1\}}}\right) \tag{4}$$

where S_t is the closing price of the underlying on day t. For each trading day, HV is computed as the annualized standard deviation of the log returns over a rolling window of the past w days as follows [8].

$$HV_t = \sqrt{252} * \sqrt{\left[\left(\frac{1}{w-1}\right) \cdot \Sigma_{\{i=t-w+1\}}^{\{t\}(r_i - \bar{r})^2}\right]} \tag{5}$$

where r̄ is the mean log return over the window and the factor √252 annualizes the daily standard deviation, assuming 252 trading days per year. Three window lengths are used, w = 20, 60, and 120 days, to capture volatility over different horizons. The window is right-aligned, so that the HV on day t uses only returns up to and including day t, which avoids look-ahead bias.

2.3.2. Implied volatility calibration

The calibrated implied volatility (σ_{cal}) is the volatility input after calibration, obtained by fitting the Black–Scholes model to observed market option prices. For each trading day, a single volatility is calibrated by minimizing the sum of squared pricing errors over a basket of options as follows [5].

DOI: 10.54254/2754-1169/2026.LD36678

$$\sigma_{cal} = \arg\min_{\sigma} \Sigma_i [\, C_{BS(\sigma;\, S,\, K_i,\, T_i,\, r)} - (C_{market}, i)]^2 \tag{6}$$

where C_BS (σ; S, K_i, T_i, r) is the Black–Scholes price of option i computed with volatility σ, and (C_{market}, i) is its observed market mid price. The basket consists of call options with days to expiration between 20 and 40 and moneyness between 0.85 and 1.15, so that only liquid, near-the-money contracts are used. The minimization is solved with the L-BFGS-B algorithm, with σ constrained to a bounded interval and initialized at the same day's 20-day historical volatility. The risk-free rate r is fixed at 2% throughout the study. By construction, σ_{cal} is the single constant volatility that best fits the market prices of the basket; whether a single volatility can adequately match options across all strikes is examined in Section 3.

2.4. Delta-hedging backtest

To evaluate how the choice of volatility input affects hedging, a delta-hedging backtest is conducted for both underlyings. On the first trading day of each month from 2019 to 2022, a near-the-money call option with days to expiration close to 30 is selected and sold, giving 48 positions for each underlying. Each position is delta-hedged daily and held until expiration.

For each position, two hedging paths are compared. Both paths sell the same option at the same market price and differ only in the volatility used to compute the delta: one path uses the 20-day historical volatility (the pre-calibration input), and the other uses the calibrated implied volatility σ_{cal} (the post-calibration input). All other elements, for example the underlying price, the risk-free rate, and the daily rebalancing, are identical, so that any difference in performance can be attributed solely to the volatility input.

The hedging portfolio consists of a short call, a position of Δt units in the underlying, and a cash account [1]. On each trading day the delta is recomputed from the Black–Scholes formula using the corresponding volatility, the underlying position is adjusted to the new delta, and the cash account reflects the net cash flow while accruing interest at the risk-free rate. At expiration, the option is settled, the remaining underlying position is liquidated, and the residual value of the portfolio is recorded as the hedging Profit and Loss (P&L). Under a perfect hedge, this residual P&L would be zero; a smaller dispersion of the P&L therefore indicates a more effective hedge [9]. To avoid look-ahead bias, the delta on day t is computed using only the volatility and price information available up to day t.

The hedging performance of each volatility input is assessed by the distribution of the hedging P&L across the 48 positions, using its standard deviation as the primary measure of hedging effectiveness [4], together with its mean, mean absolute value, and root mean squared error.

3. Results and discussion

3.1. Calibration results

Table 1 summarizes the calibrated implied volatility (σ_{cal}) and the 20-day historical volatility (HV) for both underlyings. The average σ_{cal} is 0.637 for TSLA and 0.177 for SPY, so TSLA is about 3.6 times as volatile as SPY. For TSLA, σ_{cal} slightly exceeds the HV (ratio 1.035), whereas for SPY it is slightly lower (ratio 0.954). Calibration converges on about 95% of the trading days for both. Since σ_{cal} and HV are of the same order of magnitude, it confirms that the calibration is reasonable. The higher σ_{cal} for TSLA reflects a volatility risk premium, while the ratio below one for SPY is mainly caused by the extreme realized volatility during the March 2020 crash, which lifts its average HV.

DOI: 10.54254/2754-1169/2026.LD36678

Table 1. Summary of calibrated implied volatility (σ_{cal}) and 20-day historical volatility (HV), 2019–2022

Underlying	σ_{cal} mean	HV_20 mean	σ_{cal} /HV_20 ratio	Success rate
TSLA	0.637	0.616	1.035	95.920%
SPY	0.177	0.186	0.953	95.224%

3.2. Pricing residuals and the constant-volatility assumption

Figure 2 shows the pricing residuals across moneyness for both underlyings on three representative dates, after pricing all options with a single calibrated volatility σ_{cal} .

The residuals are not randomly scattered around zero but follow a systematic pattern: they are negative for low moneyness and turn positive as moneyness increases. The residual amplitude grows sharply from the calm period (within about ±2 for TSLA in 2019) to the March 2020 crisis (up to ±15). Across all dates, the residual range is much wider for TSLA than for SPY.

This systematic pattern shows that a single constant volatility cannot match option prices at all strikes simultaneously, which is a direct consequence of the volatility smile and confirms that the constant-volatility assumption of Black–Scholes does not hold. The larger residuals for TSLA indicate that its volatility structure is more complex than SPY's, so the assumption fails more severely for the high-volatility stock. The widening of the residuals during the 2020 crisis further shows that the failure becomes more significant in turbulent markets.

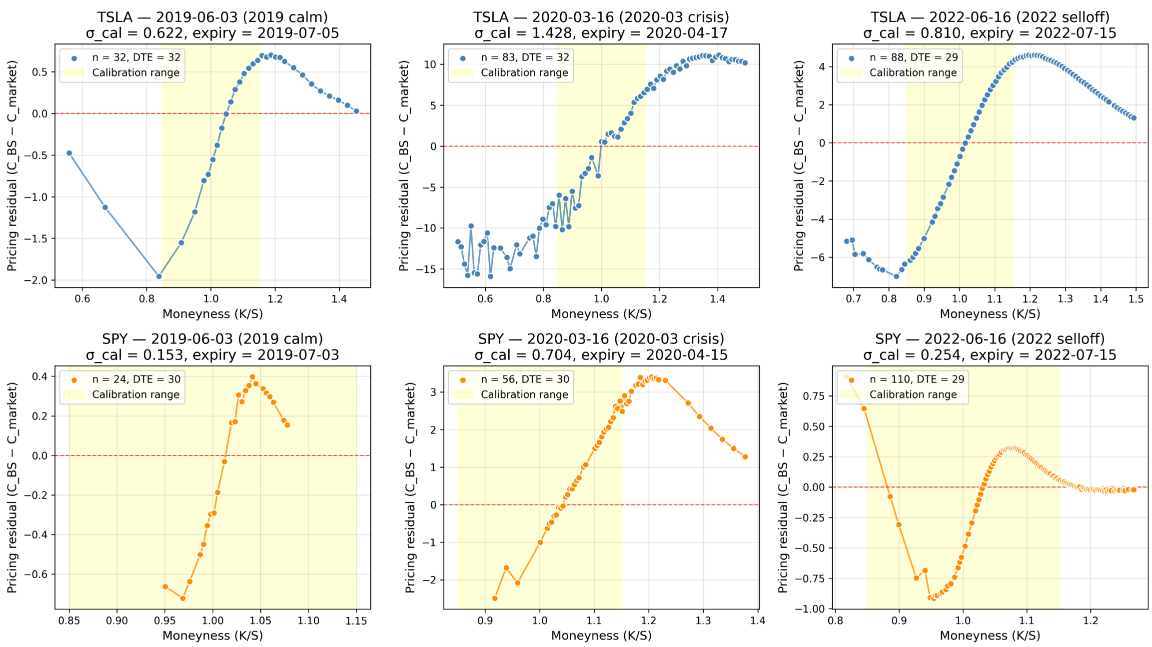

Figure 2. Pricing residuals (C_BS − C_market) against moneyness for TSLA (top) and SPY (bottom) on three dates: 2019 calm, 2020 crisis, and 2022 sell-off. All options priced with a single calibrated volatility σ_cal; shaded band = calibration range, red line = zero residual (picture credit: original)

DOI: 10.54254/2754-1169/2026.LD36678

3.3. Delta-hedging performance

Table 2 reports the hedging P&L statistics for the two volatility inputs, and Figure 3 shows the corresponding P&L distributions.

For both underlyings, using σ_{cal} lowers the standard deviation of the hedging P&L relative to HV. For TSLA, the std falls from 3.544 to 3.110, a reduction of about 12%. For SPY, it falls from 3.713 to 2.855, a reduction of about 23%. Figure 3 shows the same effect: the σ_{cal} boxes are visibly narrower than the HV boxes.

Because a smaller P&L dispersion means the option's risk is more consistently offset, these results indicate that calibration improves hedging effectiveness for both a high-volatility stock and a market index. The improvement is larger for SPY than for TSLA. A possible explanation is that SPY's more liquid option market produces more informative implied volatilities, so σ_{cal} tracks its true volatility more closely; for TSLA, the more complex volatility structure means a single calibrated volatility still leaves larger residual risk.

Table 2. Hedging P&L statistics across 48 monthly positions (2019–2022). Std(P&L) is the primary measure of hedging effectiveness (smaller = better)

Underlying	Volatility input	Mean	Std	RMSE
TSLA	20-day HV (pre-calibration)	1.299	3.544	3.740
TSLA	Calibrated IV（ σ_{cal} ）	1.181	3.110	3.297
SPY	20-day HV (pre-calibration)	-0.677	3.713	3.736
SPY	Calibrated IV（ σ_{cal} ）	-0.403	2.855	2.854

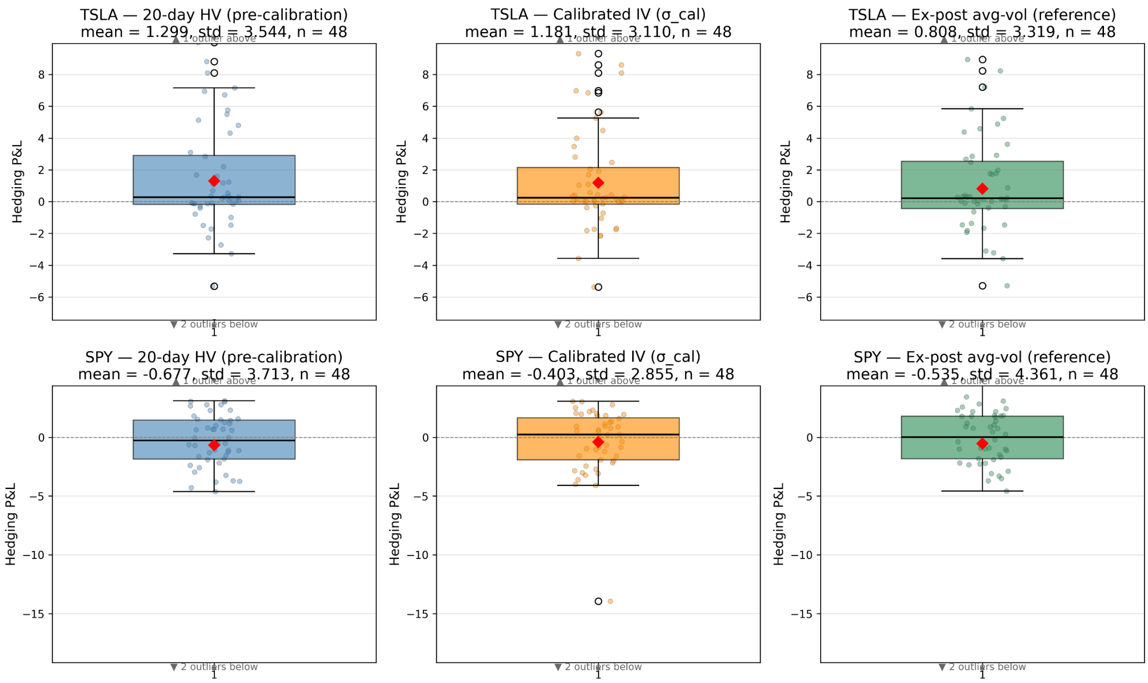

Figure 3. Distribution of hedging P&L across 48 monthly positions for TSLA (top) and SPY (bottom), under three volatility inputs: 20-day HV (pre-calibration), calibrated IV (σcal), and ex-post average volatility (reference). Red diamond = mean; dashed line = zero (picture credit: original)

DOI: 10.54254/2754-1169/2026.LD36678

3.4. Discussion

The three sets of results are consistent with one another and support a coherent interpretation. First, the calibrated volatility σ_{cal} is of a reasonable magnitude. Then, a single constant volatility cannot match option prices across all strikes, which confirms that the Black–Scholes constant-volatility assumption fails. Finally, replacing historical volatility with σ_{cal} reduces the hedging P&L dispersion for both underlyings. Therefore, calibration improves both pricing consistency and hedging effectiveness, but the benefit is not uniform: it is larger for SPY than for TSLA.

The ex-post average-volatility path provides a further insight. Even though this path uses realized volatility computed with future information, its P&L dispersion is not the smallest—for SPY it is in fact the largest (std 4.361). This indicates that a single constant volatility is limited when the true volatility varies over time. The failure of the constant-volatility assumption is thus twofold: across strikes (the volatility smile) and across time (time-varying volatility). This helps explain why calibration, which updates the volatility input daily, outperforms a static historical estimate.

Several limitations should be noted. The risk-free rate is held constant at 2%, transaction costs are ignored, and the hedging is rebalanced only daily rather than continuously. The study also covers a single period (2019–2022) and two underlyings, so the findings may not generalize to other assets or market conditions. Finally, a single Black–Scholes volatility cannot capture the volatility smile by construction; stochastic-volatility models, such as the Heston model [10], which explicitly allow volatility to vary over time and can reproduce the volatility skew, would be a natural direction for future work.

4. Conclusion

This paper set out to examine whether calibrating the volatility input improves option pricing and delta hedging, and whether the benefit differs between an individual stock and a market index. Using Tesla and the S&P 500 exchange-traded fund over 2019–2022, this study compared a historical-volatility setting with a calibrated implied-volatility setting through pricing residuals and hedging profit and loss. The results showed that a single volatility cannot match option prices across all strike prices, and that calibration reduced the variability of the hedging profit and loss for both underlyings, with a larger improvement for the index than for the individual stock. This study is limited by a constant risk-free rate, the exclusion of transaction costs, and a single sample period. Future work could adopt stochastic-volatility models, such as the Heston model, which capture both the volatility smile and its variation over time.

References

[1] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637–654.
[2] Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics, 50*(2), 125–150.
[3] Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *The Journal of Finance, 53*(6), 2059–2106.
[4] Zhao, Y. (2025). Implied volatility vs. historical volatility: Evaluating the effectiveness of delta-neutral hedging strategies. In *ICEMED 2025 Proceedings*. Atlantis Press.
[5] Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance, 52*(5), 2003–2049.
[6] TSLA option chains (2019–2022). [Data set]. *Kaggle*. Retrieved July 2026, from https://www.kaggle.com/datasets/kylegraupe/tsla-daily-eod-options-quotes-2019-2022

[7] SPY EOD options chains (2019–2022). [Data set]. *Kaggle*. Retrieved July 2026, from https://www.kaggle.com/datasets/dudesurfin/spy-options-eod-volatility-surface-2010-2023
[8] Figlewski, S. (1997). Forecasting volatility. *Financial Markets, Institutions & Instruments, 6*(1), 1–88.
[9] Boyle, P. P., & Emanuel, D. (1980). Discretely adjusted option hedges. *Journal of Financial Economics, 8*(3), 259–282.
[10] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies, 6*(2), 327–343.