CHALMERS

Comparing Empirical Minimum Variance Hedging with Delta-Vega Hedging and Model-Based Delta hedging (Black-Scholes, GARCH, and SABR) Under Normal and Extreme Market Conditions

Bachelor's thesis in Industrial Economy

SIMON HASHEMI
ALBIN WEGFORS
HUGO DAHLGREN
MARKUS HALVARSON
ARVID JÖNSON
ERIK LARSSON

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Bachelor's thesis TEKX18-VT26-07A

Bachelor’s thesis TEKX18-VT26-07A

Comparing Empirical Minimum Variance Hedging with Delta-Vega Hedging and Model-Based Delta hedging (Black-Scholes, GARCH, and SABR) Under Normal and Extreme Market Conditions

En Jämförelse av empirisk Minimum Variance hedging med delta-vega hedging och modellbaserad delta-hedging (Black-Scholes, GARCH och SABR) under normala och extrema marknadsförhållanden

SIMON HASHEMI
ALBIN WEGFORS
HUGO DAHLGREN
MARKUS HALVARSON
ARVID JÖNSON
ERIK LARSSON

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Comparing Empirical Minimum Variance Hedging with Delta-Vega Hedging and Model-Based Delta hedging (Black-Scholes, GARCH, and SABR) Under Normal and Extreme Market Conditions

SIMON HASHEMI
ALBIN WEGFORS
HUGO DAHLGREN
MARKUS HALVARSON
ARVID JÖNSON
ERIK LARSSON

Bachelor’s thesis TEKX18-VT26-07A
Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Gothenburg, Sweden 2026

Comparing Empirical Minimum Variance Hedging with Delta-Vega Hedging and Model-Based Delta hedging (Black-Scholes, GARCH, and SABR) Under Normal and Extreme Market Conditions

SIMON HASHEMI
ALBIN WEGFORS
HUGO DAHLGREN
MARKUS HALVARSON
ARVID JÖNSON
ERIK LARSSON

Department of Technology Management and Economics
Chalmers University of Technology

SUMMARY

Options are one of the most used instruments in risk management and are used by corporations and investors manage their risk exposure. For the issuer of an option, there is an exposure to the underlying asset. This is usually managed through whats called hedging. The most well-known hedging method is the delta hedge, from the Black-Scholes framework. In the Black-Scholes framework, the position in the underlying asset is continuously adjusted to counter changes in the option value, while assuming constant volatility. In practice, however, the assumption of constant volatility doesn't hold. The relationship between the underlying price and its implied volatility leads to hedging errors that become more pronounced during periods of market stress. The approaches proposed to handle these limitations include, stochastic volatility models such as SABR, time-varying volatility models such as GARCH, the empirical minimum variance hedge developed by Hull and White, and combined delta-vega strategies that aim to neutralise both price and volatility exposure at the same time. Each approach captures a different part of the relationship between price and volatility, and each has its own strengths and weaknesses. This thesis aims to evaluate and compare these hedging strategies under both normal and more stressed market conditions. The goal is to provide insight into which strategies remain robust during both normal and extreme market regimes.

The thesis is conducted as a quantitative replication and extension of the work by Hull and White from 2017. The dataset is sourced from OptionMetrics and covers the period 2015 to 2025. The dataset includes index options, equity options, and ETF options, and each trading day is classified into a normal, elevated, or extreme regime based on the VIX index level. Each methods performance is measured through the sum of squared hedging errors relative to the Black-Scholes benchmark, with Newey-West corrections applied to account for autocorrelation in the daily loss differentials. The theoretical literature is reviewed in order to motivate the construction of each hedge and to interpret the difference in hedging performance. The hedging strategies discussed are the standard Black-Scholes delta hedge, the GARCH plug-in delta hedge, the GJR-GARCH minimum variance correction, the SABR model-based delta hedge, the empirical minimum variance hedge, the model-free delta-vega hedge, and the GARCH-scaled delta-vega overlay.

The thesis shows that it is highly uncertain whether any hedging method dominates across market conditions. The investigated approaches share strengths and weaknesses, and their relative performance depends on the option type, the moneyness, and the volatility regime. The model-free delta-vega hedge delivers the largest and most consistent improvement over Black-Scholes, while the empirical minimum variance hedge also improves on the benchmark, although by a smaller margin than originally reported. The SABR approach yields modest gains and exhibits a clear difference between calls and puts. The GARCH-based delta hedges perform poorly because they discard the smile information embedded in implied volatility. The broader analysis is that effective hedging depends less on forecasting volatility itself and more on capturing the relationship between the underlying price and its implied volatility. Practitioners should therefore prioritise hedging strategies that target volatility exposure directly rather than those that rely on historical return-based forecasts.

SAMMANFATTNING

Optioner är ett av de mest använda instrumenten inom modern riskhantering och spelar en central roll i hur finansiella institutioner, företag och investerare hanterar sin exponering mot prisrörelser. Den som utfärdar en option får en naturlig exponering mot den underliggande tillgången, och denna exponering hanteras vanligtvis genom hedging. Den mest välkända hedgingmetoden är delta hedging, som bygger på Black-Scholes-ramverket. Här justeras positionen i den underliggande tillgången löpande för att kompensera för förändringar i optionens värde, under antagandet att volatiliteten är konstant.

I praktiken håller dock inte detta antagande. Sambandet mellan underliggande pris och implicit volatilitet skapar hedging fel som blir särskilt påtagliga i perioder av marknadsstress. Flera metoder har föreslagits för att hantera detta. Dessa inkluderar stokastiska volatilitetsmodeller som SABR, tidsvarierande volatilitetsmodeller som GARCH, den empiriska minimum variance hedge som utvecklats av Hull och White, samt kombinerade delta-vega-strategier som neutraliserar både pris- och volatilitetsexponering. Var och en av dessa fångar olika aspekter av samspelet mellan pris och volatilitet, och alla har sina styrkor och svagheter.

Rapporten utvärderar och jämför dessa hedgingstrategier under både normala och stressade marknadsförhållanden. Syftet är att klargöra vilka metoder som förblir robusta när precision är som mest avgörande, och både det statistiska och det praktiska perspektivet diskuteras.

Arbetet är en kvantitativ replikering och utvidgning av Hull och White (2017). Datamaterialet hämtas från OptionMetrics och täcker perioden 2015 till 2025. Det omfattar indexoptioner, aktieoptioner och ETF-optioner, och varje handelsdag klassificeras som en normal, förhöjd eller extrem regim baserat på nivån på VIX. Hedgingprestandan mäts som summan av kvadrerade hedging fel relativt Black-Scholes, med Newey-West-korrigeringar för att hantera autokorrelation i de dagliga förlustdifferenser. Som komplement till den empiriska analysen granskas relevant teoretisk litteratur. Detta för att motivera konstruktionen av varje hedge och för att tolka skillnaderna i prestanda. De strategier som behandlas är den klassiska Black-Scholes delta hedge, GARCH plug-in delta hedge, GJR-GARCH minimum variance correction, den SABR-baserade delta hedge, den empiriska minimum variance hedge, den modellfria delta-vega hedge samt den GARCH-skalade delta-vega overlay.

Resultaten visar att ingen enskild metod dominerar över alla marknadsförhållanden. Samtliga strategier har styrkor och svagheter, och deras relativa prestanda beror på optionstyp,

moneyness och rådande volatilitetsregim. Den modellfria delta-vega hedge ger den största och mest konsekventa förbättringen jämfört med Black-Scholes, medan den empiriska minimum variance hedge också slår referensen. Dock med mindre marginal än vad som ursprungligen rapporterats. SABR-metoden ger små vinster och uppvisar en tydlig asymmetri mellan calls och puts. De GARCH-baserade delta hedgarna presterar svagt eftersom de bortser från den information som de implicita volatiliteterna för olika lösenpriser ger. Den bredare slutsatsen är att effektiv hedging beror mindre på att prognostisera volatilitet i sig, och mer på att fånga samvariationen mellan underliggande pris och implicit volatilitet. Praktiker bör därför prioritera strategier som riktar sig direkt mot volatilitetsexponering, snarare än sådana som bygger på prognoser från historiska avkastningar.

Preface

This Bachelor's thesis was carried out in the spring of 2026 at the Department of Technology Management and Economics at Chalmers University of Technology in Gothenburg, Sweden. The work corresponds to 15 credits and concludes our bachelor's studies in Industrial Engineering and Management and in Engineering Physics.

We would like to thank our supervisor, Carl Lindberg, for his guidance, insightful feedback, and continued encouragement throughout this project. His teaching during this last year sparked our interest in quantitative finance and derivatives in the first place. This thesis would not have taken its current shape without the inspiration drawn from those courses.

Glossary

Leptokurtosis (Fat Tails) A property of a distribution characterised by a higher peak and heavier tails than the normal distribution, increasing the likelihood of extreme values.

Homoskedasticity When the variance of the error terms remains constant across all different levels of the independent variable.

Heteroskedasticity When the variance of the error terms varies across levels of the independent variable.

Parsimony The principle of selecting the simplest model that describes the data adequately.

ATM (At The Money) An option whose strike price is approximately equal to the price of the underlying asset.

ITM (In The Money) An option with intrinsic value. For a call, when $S > K$.

OTM (Out of The Money) An option with no intrinsic value. For a call, when $S < K$.

CEV (Constant Elasticity of Variance) Is a model where volatility depends on the level of the asset price through an exponent β.

Delta (Δ) The option's price sensitivity to changes in the underlying asset price.

Gamma (Γ) How delta changes with respect to the underlying price.

Vega (ν) The sensitivity of an option's price to changes in implied volatility.

Theta (Θ) The sensitivity of an option's price to the passage of time.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) A class of models with time-varying volatility.

Hedging Error The difference between the realized change in option value and the change in value of the hedging portfolio.

Implied Volatility The volatility parameter that, when put into the Black-Scholes formula, produces a given market price.

Mark-to-Market The continuous revaluation of a position using current market prices.

MV Hedging (Minimum variance hedging) An empirical approach that selects the hedge ratio that minimizes the variance of the hedging portfolio.

P&L (Profit and loss) The change in value of a portfolio or position over a time interval.

Risk-Neutral Valuation A pricing framework where assets are assumed to grow at a risk-free rate.

SABR (Stochastic Alpha, Beta, and Rho) A stochastic volatility model for the implied volatility smile.

Volatility Smile/Skew The pattern where implied volatility varies with strike price, deviating from the constant-volatility assumption.

Contents

1 Introduction

This chapter introduces the background, purpose, research questions, delimitations, and ethical considerations of the study.

1.1 Background

Options are one of the most used financial instruments in risk management, and it is used by a wide number of businesses in most industries. From investors and hedge funds wanting to minimize their risk exposure in portfolios to large restaurant chains looking to guard themselves from large swings in the price of staple foods like wheat and potatoes. The main purpose of an option is to give the buyer the right, but not the obligation, to buy (call option) or sell (put option) an underlying asset at a predetermined (strike) price on a predetermined expiration (maturity) date (Hull, 2018). This right, but not obligation, to exercise the option creates an asymmetric risk exposure between the buyer and the seller of the option. If the price of the underlying asset is higher than the strike price of a call (ITM, In The Money), the buyer can exercise the option for a profit equal to the difference between the underlying price and the strike price. Meanwhile, if the underlying price is lower than the strike price (OTM, Out of The Money) on the expiration date, the buyer has the choice to neglect exercising the option since its value is voided (Hull, 2018).

For the creator of an option contract, often a market maker or investment bank, there naturally exists an exposure to the movement of the price of the underlying asset. Before the Black-Scholes equation was published in 1973, option pricing was often based on heuristic methods and intuition as well as negotiation between the buyer and seller (Hull, 2018). This would frequently lead to different prices for practically identical contracts, a clear indication of market inefficiency, leading to traders of the options often exposing themselves to too much, or too little risk.

With the creation of the Black-Scholes model (Black and Scholes, 1973), and its subsequent hedging strategies based on the aptly named Greeks, one could theoretically completely eliminate the risk of selling an option. The most famous of these hedging strategies is the delta-hedge, the practice of continuously adjusting one's exposure inside of the underlying asset in order to match the value of the hedging portfolio with that of the value of the option (Hull, 2018).

In practice though, a hedge portfolio can never be perfect. The value of an option is convex and nonlinear, meaning that the value of an option deep out of the money moves very little compared to its underlying asset price (low delta, close to, or equal to 0) since the chance of exercising is very low. A deep in the money option on the other hand moves almost as much as the underlying price (high delta, close to, or equal to 1), since the chance of exercising the option is almost certain (Hull, 2018). Couple this with the spread of an asset, which is the difference between the lowest ask and the highest bid of an asset, one would generate a financial loss by continually buying and selling the same asset. The inability to continuously adjust the hedge portfolio means that its value can only achieve a "pseudo nonlinearity" compared to the value of the option by re-balancing at a reasonable rate e.g. once a week or once a month (Hull and White, 2017).

Another important factor influencing the value of the option is the volatility of the underlying. A higher volatility means that the option has a larger potential upside while its downside remains the same, a potential lowest value of 0. In the traditional delta hedge the volatility is assumed to be constant (Black and Scholes, 1973). This has been shown to not be a perfect approximation of reality, in fact, empirical evidence shows that there is a negative relationship between the price of an equity and its volatility (Black, 1976).

This difference between the value of the option and the hedge is often referred to as the hedging error and in practice it is accounted for by charging a premium on the buyer (Hull, 2018). So if one were to come up with a better model than one's competitors, it would lead to larger margins and bigger profits. But different models could perform differently depending on how the market is behaving during a certain time, creating the need to evaluate how they perform during "normal" market conditions compared to "extreme" ones.

Evaluating hedging models solely under normal market conditions risks painting a misleading picture of their practical value. It is precisely during periods of extreme market stress e.g. sharp downturns, liquidity crises and volatility spikes, that the cost of a suboptimal hedge is the highest. During such episodes, assumptions made in standard delta hedging, most notably the constant volatility assumption is frequently violated (Derman and Kani, 1994). Implied volatility can double or triple within days, the correlation between asset prices and volatility shifts dramatically. The period from 2015 to 2025 is especially informative in this regard, encompassing the Chinese devaluation panic in 2015, the inverse-VIX ETF unwind of February 2018, the COVID-19 market shock of 2020, the Federal Reserve tightening cycle in 2022, the Russian invasion of Ukraine in 2022, and the tariff-driven market downturn of April 2025. Each of these events exposed hedging portfolios to conditions far outside the range on which most models are calibrated. Understanding how different hedging strategies perform under such conditions is therefore not purely an academic exercise, it has direct implications for risk management and capital allocation in derivatives markets.

1.2 Purpose

The purpose of this thesis is to evaluate and compare the hedging performance of four delta hedging approaches and a delta-vega hedging approach. The four delta hedging approaches are the Black–Scholes with constant volatility, Black–Scholes with GARCH-forecasted volatility, SABR model-based delta hedging, and empirical minimum variance hedging following the work of Hull and White (2017). This is done across both normal and extreme market conditions. The study uses data from OptionMetrics covering the period 2015–2025 for index, equity options, and ETF options. This thesis replicates the framework of Hull and White (2017) as well as extends it to a broader set of models, markets, and a time period that includes several episodes of market stress. It aims to provide empirical evidence on which hedging methodology is most robust when it matters most.

1.3 Research Questions

The study aims to answer the following research questions:

1. How does the hedging performance of empirical minimum variance hedging compare to model-based delta hedging (Black–Scholes, GARCH, and SABR) and delta-vega hedging under normal market conditions?
2. How does this comparison change under extreme market conditions, characterized by elevated volatility and rapid price movements?
3. To what extent do the findings of Hull and White (2017) regarding the superiority of minimum variance hedging generalize to different asset classes and a more recent time period?

1.4 Delimitations

This study is subject to several delimitations. Transaction costs are not incorporated into the analysis. Although such costs are relevant in practice, three considerations motivate this choice. The first is that data from OptionMetrics does not contain reliable transaction cost information beyond the bid-ask spread. The spread itself is a rough approximation of the true cost faced by a market maker, who internalises a substantial fraction of order flow. Another consideration is that all hedging strategies are evaluated on the same data with the same daily rebalancing frequency. Transaction costs would therefore affect each strategy approximately in proportion to its number of trades, which preserves the relative comparison that is the focus of this study. The third consideration is that this choice follows Hull and White (2017), the methodological benchmark of the thesis. Strategies that require more frequent or two instrument rebalancing, notably delta and vega hedging, face a higher cost burden in practice. The discussion returns to this point.

Hedging is restricted to first order Greeks, namely delta and vega. Exotic options fall outside the scope, and the analysis is confined to standard European and American options. Market regimes are defined using the VIX index. Levels below 20 are classified as normal, levels between 20 and 30 as elevated, and levels above 30 as extreme. Alternative regime definitions may yield different results.

1.5 Sustainability and Ethics

This study is analytical and quantitative, based on already established models and secondary data. It is also conducted in accordance with good research practice as described by (Vetenskapsrådet, 2024). The study does not involve sensitive personal data nor human participants and therefore does not require ethical review under the Ethical Review of Research Involving Humans (Sveriges riksdag, 2003).

This study is a replication and extension of Hull and White (2017), which serves as the methodological starting point. A clear distinction between the original study and the scope of this work is maintained through consistent referencing and structured data management.

From a sustainability perspective the study's contribution is indirect. Although it may contribute to an increased understanding of how financial models behave under extreme and normal market conditions, therefore support assessment of model robustness in financial markets. There are no recommendations made regarding the actual application of results, reducing the risk that the study's findings have unintended consequences.

2 Theoretical Framework

This chapter introduces and explains the theory required in order to evaluate option hedging performance. It starts by describing option pricing and Black-Scholes delta hedging, and thereafter extends to minimum variance hedging, GARCH-based volatility forecasting, SABR and also delta-vega hedging.

2.1 Introduction to Financial Contracts

Financial contracts are agreements between two parties involving the exchange of money or assets, either immediately or at a future predetermined date. Some contracts involve direct transactions, while others depend on how the value of an underlying asset evolves over time. The latter are known as derivatives, as their value is derived from assets such as stocks, commodities, or interest rates.

A key distinction between financial contracts is whether they create an *obligation* or a *right*. Some contracts require both parties to fulfill the agreement at a future date, while others give one party the flexibility to decide whether to proceed (Hull, 2018). The contracts relevant to this article are forward contracts and, in particular, options.

2.1.1 Forward Contracts

A forward contract is an agreement between two parties to buy or sell an asset at a predetermined price at a future date. Both parties are obligated to carry out the transaction at maturity.

These types of contracts are in most cases privately negotiated and tailored to specific demands. This flexibility, however, introduces counterparty risk, as there is no guarantee that the other party will fulfill their obligation (Hull, 2018).

2.1.2 Options

Options differ fundamentally from forwards because they do not create an obligation for both parties. Instead, an option gives the buyer the *right, but not the obligation*, to buy or sell an underlying asset at the predetermined strike price (Hull, 2018).

The payoff of an option describes how much the holder receives at expiration. The payoff of a call option is given by

$$\max(0, S(T) - K) \tag{1}$$

where $S(T)$ is the price of the underlying asset at expiration and K is the strike price. The payoff of a put option is given by

$$\max(0, K - S(T)) \tag{2}$$

Option pricing models, such as the Black–Scholes model explained in Section 2.2, are used to determine the fair value of these payoffs (Hull, 2018).

Option Payoff Diagrams The payoff structure of options can be visualized graphically. A call option has a payoff that increases when the underlying asset price exceeds the strike price, while a put option increases in value when the asset price falls below the strike price.

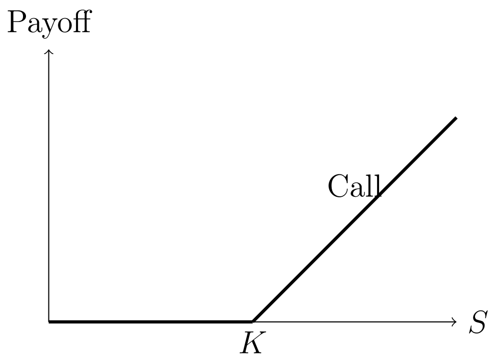

Figure 1: The payoff from a call option, as the stock price S surpasses K, the option can be exercised for a profit of $S(T) - K$, which is equal to its payoff.

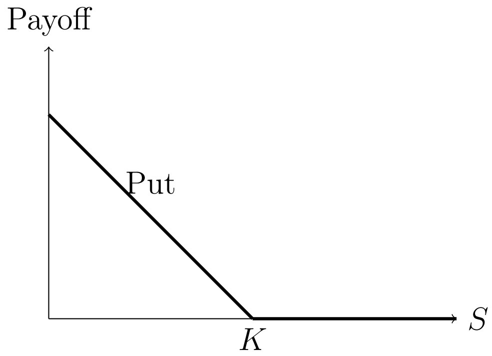

Figure 2: The payoff from a put option, as the stock price S falls short of K, the option can be exercised for a profit of $K - S(T)$, which is equal to its payoff.

American and European Options Options can be broadly classified into two types depending on when the holder is permitted to exercise: *European options* may only be exercised at the expiration date T, whereas *American options* may be exercised at any point up to and including T (Hull, 2018). Since an American option encompasses all the rights of a European option and additionally grants the flexibility to exercise early, it can never be worth less than an otherwise identical European option. Formally, if C and c denote the prices of an American and a European call, respectively, then $C \geq c$, with equality holding whenever early exercise is suboptimal.

For a *call option* on a non-dividend-paying asset, early exercise is never optimal (Hull, 2018). The value of a call can be decomposed into its *intrinsic value* $\max(0, S(T) - K)$ and its *time value*, which is strictly positive prior to expiration. Exercising early captures only the intrinsic value and permanently destroys the remaining time value.

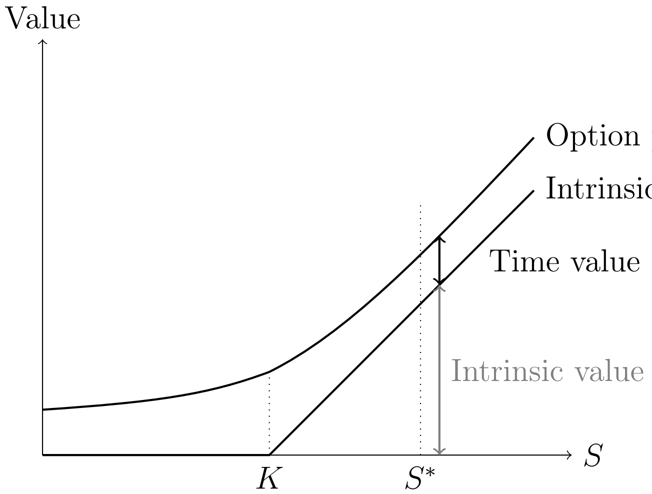

Figure 3: Decomposition of a European call option price into intrinsic value $\max(0, S-K)$ and time value. The option price c lies strictly above the intrinsic value prior to expiration. Early exercise of an American call captures only the intrinsic value and destroys the remaining time value, making early exercise suboptimal for non-dividend-paying assets.

By holding rather than exercising, the holder retains the interest earned on the deferred payment of the strike price K. A formal lower bound confirms this: under the no-arbitrage principle,

$$c \geq \max\big(0,\, S - Ke^{-r(T-t)}\big) > \max(0, S - K), \tag{3}$$

since $Ke^{-r(T-t)} < K$ for $r > 0$ and $T > t$. The market value of the option therefore strictly exceeds its exercise value before maturity, making early exercise a certain loss. Consequently, for non-dividend-paying assets, American and European calls have the same price: $C = c$ (Hull, 2018).

The argument does not extend to *put options*. For a deep in-the-money put it may be optimal to exercise early in order to receive $K - S$ immediately and reinvest the proceeds at the risk-free rate, particularly when interest rates are high and little time value remains. American puts therefore carry a strictly positive early-exercise premium relative to their European counterparts, and $P > p$ can hold (Hull, 2018).

2.2 Black–Scholes Equation

The development of the Black-Scholes model was prompted by the need for a consistent and theoretically grounded method to price options. Prior to its introduction, option pricing was largely based on heuristic approaches and market conventions, which often resulted in inconsistent valuations (Hull, 2018).

With the rapid growth of financial markets, a more rigorous framework became necessary. Black and Scholes addressed this by applying the no-arbitrage principle and demonstrating that an option's value can be derived by constructing a hedged portfolio (Hull, 2018; Bodie et al., 2021).

2.2.1 Derivation of the Black–Scholes Equation

The derivation presented here follows the framework in Hull (2018) and the notation used in Borell and Lindberg (2023). The stock price is assumed to follow:

$$dS = \mu S\, dt + \sigma S\, dB \tag{4}$$

This equation describes how the price of the underlying asset evolves over time. The first term represents the expected growth of the asset, while the second term captures random fluctuations. This means that although the asset tends to grow on average, its actual path is uncertain. The option price is written as $c(t, S(t))$, and applying Itô's lemma (stochastic Taylor expansion) gives

$$dc = c_t\, dt + c_S\, dS + \frac{1}{2} c_{SS} \sigma^2 S^2\, dt \tag{5}$$

This equation describes how the option price changes over time. The term c_t captures how the option value changes as time passes, while c_S reflects sensitivity to the underlying asset price. The second derivative term arises due to the randomness of the stock price and captures curvature in the option value.

The key idea in the Black–Scholes framework is to eliminate risk. This is done by constructing a hedged portfolio consisting of one option and a position in the underlying asset. By choosing the number of shares as $\Delta = c_S$, the random component of the portfolio is removed.

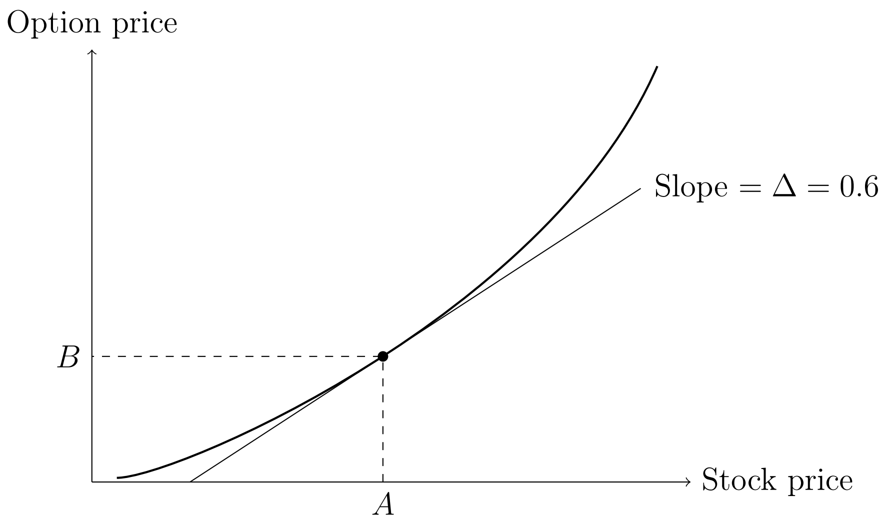

Figure 4: The delta of a call option. The slope of the tangent to the option price curve at point A equals the delta of the option. Here $\Delta = 0.6$, the figure is adapted from Hull (2018).

Figure 4 illustrates how the option payoff is nonlinear, while the hedge creates a locally linear approximation. By continuously adjusting the hedge, the portfolio behaves as if it were risk-free over small time intervals.

Since the portfolio is risk-free, it must earn the risk-free interest rate. Otherwise, arbitrage opportunities would arise. This implies that the return on the portfolio satisfies

$$dc = rc\,dt \tag{6}$$

Substituting the expression for dc and simplifying leads to the Black–Scholes partial differential equation:

$$c_t + rSc_S + \frac{1}{2}\sigma^2 S^2 c_{SS} - rc = 0 \tag{7}$$

This equation describes how the price of the option must evolve over time to ensure that no arbitrage opportunities exist in the market. To obtain a specific solution, a boundary condition is required. For a European call option, the payoff at maturity is:

$$c(T, S(T)) = \max(S(T) - K, 0) \tag{8}$$

Meaning that the option will only be exercised if it is profitable. Solving the differential equation with this condition yields the Black–Scholes pricing formula:

$$c(t, S(t)) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{9}$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{10}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} \tag{11}$$

Here $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution. This final expression gives the fair value of the option as a function of observable market variables. It shows that the price depends on the current asset price, the strike price, time to maturity, volatility, and the risk-free interest rate.

2.2.2 Interpretation of the Black–Scholes Formula

The Black–Scholes formula can be interpreted as the present value of the expected payoff of the option under a risk-neutral probability measure. Specifically, the formula

$$c(t, S(t)) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{12}$$

can be understood as the difference between two components. The first term, $S\Phi(d_1)$, represents the value of holding the underlying asset, weighted by the risk-neutral probability that the option will be exercised. The second term, $Ke^{-r(T-t)}\Phi(d_2)$, represents the present value of the strike price, also weighted by the probability of exercise. The quantities d_1 and d_2 can be interpreted as standardized measures of the distance between the current asset price and the strike price, adjusted for volatility and time to maturity.

This interpretation follows from the concept of risk-neutral valuation, where the option price equals the discounted expected payoff under a probability measure in which all assets grow at the risk-free rate (Hull, 2018).

2.3 Hedging Dynamics

Having established the Black–Scholes pricing formula in the previous section, we now turn to how the model is used in practice to manage the risk of an option position. The core idea of hedging is to construct a portfolio that offsets the sensitivity of an option to changes in market variables. The sensitivities themselves are described by the Greeks. The *Greeks* measure how the value of an option changes in response to small changes in the underlying inputs. Four of them are particularly important for this study.

2.3.1 Delta (Δ)

Delta is the first derivative of the option price with respect to the underlying asset price S:

$$\Delta = \frac{\partial c}{\partial S}. \tag{13}$$

For a European call under the Black–Scholes model, this equals $\Phi(d_1)$, while for a put, $\Delta = \Phi(d_1) - 1$. Delta takes values in $[0, 1]$ for calls and $[-1, 0]$ for puts. It can be interpreted both as the hedge ratio, meaning the number of units of the underlying needed to replicate the option and approximately as the risk-neutral probability that the option expires in-the-money (Hull, 2018).

2.3.2 Gamma (Γ)

Gamma measures the rate of change of delta with respect to S:

$$\Gamma = \frac{\partial^2 c}{\partial S^2} = \frac{\partial \Delta}{\partial S}. \tag{14}$$

Under Black–Scholes, the gamma of a European call and put are equal:

$$\Gamma = \frac{\phi(d_1)}{S\sigma\sqrt{T-t}}, \tag{15}$$

where $\phi(\cdot)$ denotes the standard normal density. Gamma is always non-negative and peaks when the option is at-the-money with short time to expiry. It quantifies the curvature of the option price as a function of S, and therefore governs how quickly the delta changes when the underlying moves. A position with large gamma requires more frequent rebalancing to remain delta-neutral, and is more exposed to large sudden moves in S (Hull, 2018).

2.3.3 Vega (ν)

Vega measures the sensitivity of the option price to changes in volatility:

$$\nu = \frac{\partial c}{\partial \sigma}. \tag{16}$$

For both calls and puts under Black–Scholes:

$$\nu = S\phi(d_1)\sqrt{T-t}. \tag{17}$$

Vega is strictly positive and largest for at-the-money options with longer time to maturity. Even though the Black–Scholes model treats σ as a constant, vega is still practically important because the implied volatility used to price and hedge an option can change over time. A position with large vega is therefore exposed to movements in implied volatility even when it is delta-neutral (Hull, 2018).

2.4 Delta Hedging

Delta hedging is the process of constructing a portfolio that is insensitive to small movements in the underlying asset price by continuously adjusting the hedge ratio. In the Black–Scholes framework, the replicating portfolio for a long call consists of holding

$\Delta = \Phi(d_1)$ units of the underlying asset, financed by borrowing at the risk-free rate. As S and t change, so does Δ, and the position must be rebalanced to maintain neutrality.

Consider a portfolio Π consisting of a long position in one call option and a short position of Δ units in the underlying:

$$\Pi = c - \Delta \cdot S. \tag{18}$$

By construction, $\partial\Pi/\partial S = 0$ to first order. Since the portfolio has no immediate exposure to movements in S, it must earn the risk-free rate to prevent arbitrage. This requirement is exactly what produces the Black–Scholes PDE (Hull, 2018).

The option price as a function of S is illustrated in Figure 4. The curve is strictly convex, meaning its slope increases as S rises. At any given stock price the delta is the slope of the tangent to this curve, and a short position of exactly Δ units in the underlying offsets the first-order risk of the option. Because the curve is not linear, however, this offset is only exact at the point of tangency. For every finite move in S the value of the hedge will drift away from the value of the option, and the portfolio must be rebalanced.

The source of this drift is gamma. From $\Gamma = \partial\Delta/\partial S$, every move in S shifts the delta of the option while the hedge position held in the underlying does not adjust automatically. For a long call with $\Gamma > 0$, a rise in S increases Δ, so the portfolio becomes net long the underlying; a fall in S decreases Δ, leaving the portfolio net short. The deviation grows with the magnitude of the price move and the size of gamma.

To restore delta neutrality the hedge position must therefore be rebalanced: the number of units of the underlying is updated to the new value of $\Delta(S, t)$ at each rebalancing date. In the continuous limit assumed by Black–Scholes this rebalancing happens infinitely often, and the portfolio is risk-free at every instant. In practice trading is discrete and costly, so rebalancing occurs at fixed intervals such as daily or weekly, generating a hedging error at each step (Hull, 2018).

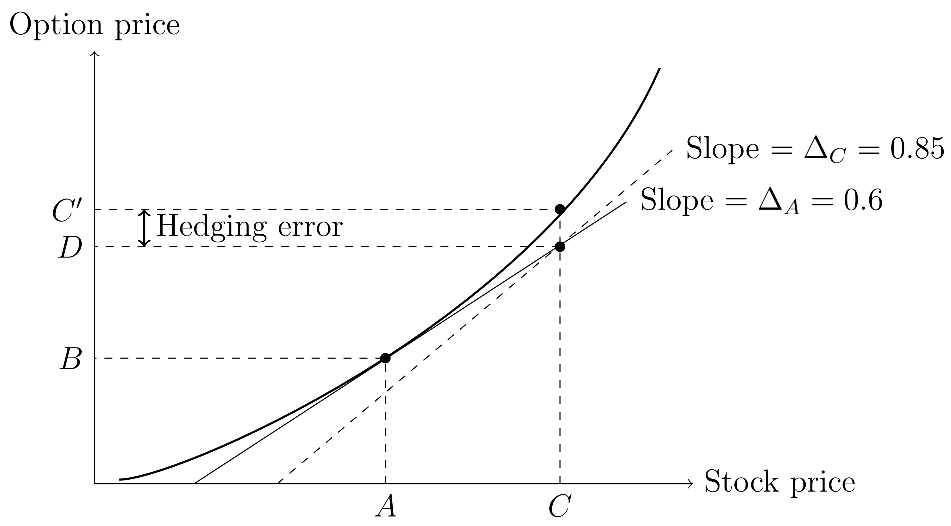

Figure 5: Delta hedging and the hedging error, follow-up to Figure 4. At point A the option is hedged with slope $\Delta_A = 0.6$. As the stock price moves to C inside the chosen re-balancing interval, the option value on the curve rises to C', while the hedged portfolio only reaches D due to the linear approximation. The gap between C' and D on the vertical axis represents the hedging error arising from the convexity of the option price.

2.4.1 Put-Call Parity

Put-call parity is a fundamental relationship in options pricing that links the price of a European call option to the price of a European put option with the same strike price, expiration date, and underlying asset. The relationship was formalized by Stoll (1969) and follows directly from the no-arbitrage principle: if two portfolios generate identical payoffs at maturity, they must have the same price today, otherwise a riskless profit could be extracted by buying the cheaper and selling the more expensive (Hull, 2018).

To derive the relationship, consider two portfolios. The first consists of a long European call option with strike price K and maturity T, combined with a cash position equal to the present value of the strike price, $Ke^{-r(T-t)}$, invested at the risk-free rate r. The second portfolio consists of a long European put option with the same strike and maturity, combined with a long position in the underlying asset at its current price S. At maturity, both portfolios pay $\max(S, K)$, since the call pays off when $S > K$ and the put pays off when $S < K$, and in either case the combined position corresponds to the other portfolio exactly. The absence of arbitrage therefore requires that their prices be equal today, giving the put-call parity relation:

$$c + Ke^{-r(T-t)} = p + S, \tag{19}$$

where c is the price of the European call, p is the price of the European put, S is the price of the underlying asset, K is the strike price, r is the risk-free interest rate, and $T - t$ is the time remaining to maturity (Hull, 2018).

This relationship has several important ramifications for the hedging framework developed in this study. First, it shows that calls and puts on the same underlying are not independent instruments; their prices are linked through the underlying asset price and the risk-free rate. As a consequence, the delta of a European put can be derived directly from the delta of the corresponding call:

$$\Delta_{\text{put}} = \Delta_{\text{call}} - 1. \tag{20}$$

This means that a put with delta -0.4 corresponds to a call with delta $+0.6$ at the same strike and maturity. Second, put-call parity implies that any hedging strategy applied to a call position has a direct counterpart for the corresponding put position. If the delta hedge of a call is improved by accounting for the correlation between the underlying price and implied volatility, the same correction should in principle benefit the put hedge as well, since the two are linked by an arbitrage relationship. In practice, however, the empirical results of Hull and White (2017) show a systematic asymmetry between call and put hedging performance, suggesting that factors beyond put-call parity play an important role in building an optimal hedging strategy.

2.5 Moneyness

Moneyness describes the relationship between the price of the underlying asset and the strike price of an option (Hull, 2018). It indicates whether an option would generate a positive payoff if exercised immediately, and is one of the most important variables in both pricing and hedging .

An option is said to be in the money whenever immediate exercise would give the holder a positive payoff (Hull, 2018). For a call option this happens when the underlying price S is above the strike price K, and for a put option when S is below K. An option is at the money when S is approximately equal to K, and out of the money when exercise would give a negative payoff and the holder would therefore choose not to exercise. Out of the money options have no intrinsic value, only time value, while in the money options have both .

The simplest way to measure moneyness is through the ratio S/K, where values above one indicate an in the money call and values below one an out of the money call. The opposite holds for puts (Hull, 2018). A common alternative is log moneyness, defined as $\ln(S/K)$, which is symmetric around zero and often easier to work with in statistical models. Both measures rely solely on the price of the underlying asset and the strike. It does not take into account other factors such as volatility or time to maturity.

A more refined measure used widely in practice is the option's delta (Hull, 2018). Since delta expresses both moneyness and the time remaining until expiration, it provides a smoother as well as more informative classification than the raw S/K ratio. A deep out of the money call has a delta close to zero, an at the money call has a delta near 0.5, and a deep in the money call has a delta close to one. For puts the delta runs from zero for deep out of the money options to negative one for deep in the money options. Two options with the same S/K ratio but different maturities will have different deltas.

Out of the money puts behave very differently from at the money puts. This is because their deltas change more steeply with movements in the underlying and because the implied volatilities of out of the money puts tend to be substantially higher than those of at the money options. This last feature is known as the volatility skew and is especially pronounced in equity index markets. This means that any hedging method which assumes constant volatility will perform very differently over moneyness regions (Bakshi et al., 1997).

In line with Hull and White (2017), this thesis groups options into nine moneyness buckets based on the absolute value of delta, ranging from 0.1 to 0.9 in increments of 0.1. Options with $|\Delta|$ near 0.1 correspond to deep out of the money calls or deep out of the money puts, those near 0.5 correspond to at the money options, and those near 0.9 correspond to deep in the money options. Reporting hedging performance separately for each bucket makes it possible to detect where each hedging method works well and where it breaks down, which is particularly relevant during periods of market stress when out of the money puts are most actively used as protection.

2.6 Minimum Variance (MV) Hedging

The standard Black-Scholes delta hedge is derived under the assumption of constant volatility. In practice, however, implied volatility varies over time and is empirically correlated with movements in the underlying asset price. As a result, the practitioner Black-Scholes delta does not minimize the variance of the hedged portfolio (Hull and White, 2017).

Minimum variance (MV) hedging deals with this limitation by accounting for the joint dynamics of the underlying price and implied volatility. The objective is to choose a hedge ratio that minimizes the variance of changes in the value of the hedged position. Formally, the MV hedge ratio Δ_{MV} is defined as the value that minimizes the variance of

$$\Delta f - \Delta_{MV} \Delta S, \tag{21}$$

where $\Delta f = f(t+1) - f(t)$ is the change in the option price and ΔS is the change in the underlying asset price.

Using a first-order Taylor expansion, the change in the option price can be approximated as

$$\Delta f \approx \Delta_{BS} \Delta S + \nu_{BS} \Delta \sigma_{imp}, \tag{22}$$

where Δ_{BS} and ν_{BS} denote the practitioner Black-Scholes delta and vega, respectively, and $\Delta\sigma_{imp}$ is the change in implied volatility (Hull and White, 2017).

Minimizing the variance of the hedging error then leads to the approximation

$$\Delta_{MV} = \Delta_{BS} + \nu_{BS} \frac{\partial E(\sigma_{imp})}{\partial S}. \tag{23}$$

This expression shows that the MV delta consists of the standard Black-Scholes delta plus a correction term that conveys the expected change in implied volatility conditional on a change in the underlying price. Intuitively, when the underlying asset price moves, the option value is affected not only directly through delta, but also indirectly through changes in volatility. Since these effects are correlated, ignoring the volatility component leads to less optimal hedging.

Empirical evidence suggests that, for equity options, there is typically a negative relationship between asset prices and implied volatility. Consequently, the MV delta is generally smaller than the Black–Scholes delta for call options and more negative for put options. This implies that a variance-minimizing hedge requires under-hedging calls and over-hedging puts relative to the standard delta hedge (Hull and White, 2017).

2.6.1 The a,b,c parameters

The key challenge in implementing the MV hedge ratio is estimating the sensitivity $\partial E(\sigma_{imp})/\partial S$. Hull and White (2017) show empirically that, for S&P 500 options, this quantity is well approximated by a quadratic function of the practitioner Black–Scholes delta, scaled by the current asset price and the square root of time to maturity. Substituting this approximation into Eq. 23 yields the estimable form of the MV delta:

$$\Delta_{MV} = \Delta_{BS} + \frac{\nu_{BS}}{S\sqrt{T}} \left(a + b\Delta_{BS} + c\Delta_{BS}^2 \right), \tag{24}$$

where a, b, and c are constants to be estimated from data, S is the current asset price, and T is the time to maturity. The scaling by $S\sqrt{T}$ reflects two empirical properties of the implied volatility surface. First, scale invariance implies that what matters for volatility dynamics is the proportional change in the underlying rather than its absolute magnitude. Second, the sensitivity of implied volatility to price changes is observed to decay with the square root of time to maturity. The quadratic form in Δ_{BS} captures the fact that the volatility–price relationship is not uniform across moneyness: it tends to be strongest for out-of-the-money options and weakens as the option moves into the money (Hull and White, 2017).

The three coefficients are estimated separately for puts and calls by fitting a regression of one-day option price changes on the underlying price change, weighted by the model correction term. Rather than estimating a, b, and c once over the full sample, a rolling window approach is adopted in which the parameters are estimated over a historical window of 12-36 months and then applied out of sample to determine hedge ratios over the following month. This assures that the parameters adapt to prevailing market environments while remaining grounded in a sufficiently long history to produce stable estimates.

2.7 GARCH Models

This section introduces the GARCH models and explains how time-varying volatility can be relevant for option hedging.

2.7.1 Transition from OLS to ARCH

As Engle (2001) notes, the standard linear regression model, estimated by Ordinary Least Squares (OLS), has long been a natural choice in econometrics for measuring how much one variable changes when another varies. A core assumption of this model is homoskedasticity: that the variance of the error term ϵ_t is the same at every point in time,

$$E[\epsilon_t^2] = \sigma^2, \quad \forall t \tag{25}$$

In financial applications, the dependent variable is typically the asset return, and the risk level is represented by the variance of that return. Financial data shows that periods of high risk are not random; rather, they exist in connected clusters. An event with high volatility is likely to be followed by another volatile event, and vice versa. This degree of autocorrelation in the volatility of returns is commonly known as volatility clustering and was first documented by Mandelbrot (1963). The ARCH/GARCH framework was developed precisely to relax the homoskedasticity assumption and let the variance evolve through time, which is essential when modelling derivatives and managing financial risk.

2.7.2 The ARCH Model

Before the ARCH model proposed by Engle (1982), a common way to estimate volatility was the rolling variance of historical observations. In this approach, today's variance was taken to be the average of the squared residuals over a fixed number of recent periods, with every observation receiving the same weight. Engle (1982) identified two primary problems with this:

1. It is more reasonable that recent observations have a larger influence on today's variance than older ones, which contradicts the assumption of equal weights.
2. Observations outside the chosen window are given a weight of exactly zero, which is also undesirable.

The Autoregressive Conditional Heteroskedasticity (ARCH) process addresses the first issue by treating the weights as parameters to be estimated, so that the optimal weights for forecasting are determined by the data.

ARCH Model Specifications For a process with zero mean and no explanatory variables, the model is defined as

$$y_t = \epsilon_t h_t^{1/2} \tag{26}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{27}$$

where $y_t \mid \psi_{t-1} \sim N(0, h_t)$ means that, conditional on all information available up to time $t-1$ (denoted ψ_{t-1}), y_t is normally distributed with mean zero and variance h_t. To improve the efficiency of a standard regression, the ARCH regression model is used, where $x_t\beta$ is a combination of lagged (past) values and exogenous (externally given) variables:

$$y_t \mid \psi_{t-1} \sim N(x_t\beta, h_t) \tag{28}$$

$$h_t = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p}, \alpha) \tag{29}$$

$$\epsilon_t = y_t - x_t\beta. \tag{30}$$

Engle (1982) showed that ARCH models are typically estimated using Maximum Likelihood Estimation (MLE) rather than OLS. MLE chooses the parameter values that make the observed data the most probable under the assumed conditional distribution. OLS is avoided here because, although it remains unbiased, it is no longer efficient when the variance changes over time, and it cannot directly model how the conditional variance evolves. In other words, in the presence of heteroskedasticity OLS fails to be the Best Linear Unbiased Estimator (BLUE), the standard benchmark from the Gauss–Markov theorem for the regression estimator with the smallest variance among all linear unbiased estimators.

2.7.3 The GARCH Model

The Generalized ARCH (GARCH) model was proposed by Bollerslev (1986) to extend the ARCH process so that it has a longer memory and a more flexible lag structure. As in ARCH, the conditional variance is a weighted average of squared past residuals; in addition, GARCH lets past variances feed back into today's variance, with weights that decline gradually but never reach exactly zero. Defining h_t as the conditional variance, the GARCH(p, q) process is

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{31}$$

Here the orders p and q specify how far back the model looks: q is the number of lagged squared residuals ϵ_{t-i}^2 included (the ARCH terms, capturing the impact of recent shocks), and p is the number of lagged conditional variances h_{t-i} included (the GARCH terms, capturing how persistent past variance levels are). For example, GARCH(1,1) uses one lag of each, while GARCH(2,1) would use two variance lags and one squared-residual lag.

A major advantage of GARCH over ARCH is parsimony, that is, the ability to describe the data with as few parameters as possible. Bollerslev (1986) notes that a simple GARCH(1,1) model often performs as well as, or better than, a high-order ARCH(q) model, because the lagged variance term $\beta_1 h_{t-1}$ effectively summarises an infinite history of squared residuals through a geometrically declining weight structure.

According to Bollerslev (1986), for the GARCH(1,1) process to be wide-sense stationary (that is, to have a constant unconditional mean and a constant unconditional variance through time), the sum of the ARCH and GARCH coefficients must be less than one:

$$\alpha_1 + \beta_1 < 1 \tag{32}$$

This sum measures the persistence of volatility shocks. When it is close to one, shocks to the conditional variance fade out only slowly, producing prolonged periods of high or low volatility. When the process is stationary, the long-run (unconditional) variance is constant and given by:

$$\sigma^2 = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \tag{33}$$

A further property emphasised by Bollerslev (1986) and Engle (2001) is the model's ability to generate leptokurtosis, the technical name for a return distribution with "fat tails" (kurtosis higher than that of a normal distribution, meaning extreme outcomes occur more often than the normal model predicts). Even when the innovations ϵ_t, that is the random shocks driving the model, are themselves normally distributed, the time-varying variance of GARCH inherently produces a return distribution with thicker tails.

2.7.4 The Leverage Effect

In equity markets it has been well documented that negative returns tend to be followed by an increase in volatility, while positive returns of the same magnitude have a smaller effect. Known as the leverage effect, the term was introduced by Black (1976). It originates from the idea that when a firm's equity value falls, its debt-to-equity ratio rises which in turn makes the firm riskier and more volatile. Later research has shown that the leverage explanation alone is insufficient to describe the full magnitude of the effect, however the pattern itself is robust across equity markets (Black, 1976; Christie, 1982).

The standard GARCH(1,1) model described in Section 2.7.3 treats negative and positive price shocks the same. In Eq. 31 the conditional variance depends on the squared residuals, resulting in upwards and downwards moves being treated symmetrically. As a result, the model cannot capture the leverage effect. This motivated the development of asymmetric extensions, with the most notable being the EGARCH model of Nelson (1991) and the GJR-GARCH model of Glosten et al. (1993), both of which are presented in the following section.

2.7.5 Asymmetric GARCH models

The first GARCH variant made to explicitly account for the leverage effect was the Exponential GARCH (EGARCH) model proposed by Nelson (1991). Instead of modeling the conditional variance h_t directly, EGARCH specifies the logarithm of the conditional variance

$$\ln h_t = \omega + \alpha \left(\left| \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \ln h_{t-1} \tag{34}$$

where $\epsilon_{t-1}/\sqrt{h_{t-1}}$ is the standardised residual, that is, the most recent shock rescaled by its own conditional standard deviation. The asymmetry parameter γ is what allows good and bad news to have different impacts: when $\gamma < 0$, a negative standardised shock raises $\ln h_t$ more than a positive shock of the same size, reproducing the leverage effect. The constant $\sqrt{2/\pi}$ is the expected value of the absolute standardised residual under a standard normal distribution; subtracting it ensures that α measures the deviation of $|\epsilon_{t-1}/\sqrt{h_{t-1}}|$ from its expected size (Nelson, 1991).

An advantage of the EGARCH specification is that, because the model is formulated in terms of $\ln h_t$, the conditional variance $h_t = \exp(\ln h_t)$ is guaranteed to be positive regardless of the values of the parameters. Therefore, the non-negativity constraints ($\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$) required in the standard GARCH model are eliminated (Nelson, 1991).

The GJR-GARCH(1,1) is another alternative approach to modeling asymmetry, proposed by Glosten et al. (1993). The model is an extension of the standard GARCH(1,1) adding a variable that allows negative shocks to have a different impact on future volatility:

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I(\epsilon_{t-1} < 0) + \beta h_{t-1} \tag{35}$$

where $I(\epsilon_{t-1} < 0)$ is an indicator function that equals one when the previous residual is negative and zero otherwise. When $\gamma > 0$, the total impact of a negative shock on the conditional variance is $(\alpha + \gamma)\epsilon_{t-1}^2$, while the impact of a positive shock of the same magnitude is only $\alpha\epsilon_{t-1}^2$. The leverage effect is therefore captured by the asymmetry parameter γ (Glosten et al., 1993).

For the GJR-GARCH(1,1) process to be covariance stationary, the persistence condition becomes:

$$\alpha + \frac{\gamma}{2} + \beta < 1 \tag{36}$$

where the factor 1/2 on the asymmetry term arises because, under a symmetric innovation distribution, $E[I(\epsilon_t < 0)] = 1/2$. When the process is stationary, the unconditional variance is:

$$\sigma^2 = \frac{\omega}{1 - \alpha - \gamma/2 - \beta} \tag{37}$$

The GJR-GARCH model has two useful advantages, when compared to the EGARCH model. First, the asymmetry parameter γ has a direct interpretation as the additional variance contribution from negative shocks, making it easy to quantify the strength of the leverage effect. Second, the model keeps the linear variance structure of standard GARCH, which simplifies estimation and forecasting. For these reasons, the GJR-GARCH specification is widely used in empirical option pricing and hedging studies (Glosten et al., 1993; Engle, 2001).

For this study, the GJR-GARCH(1,1) model is chosen as the primary GARCH specification because it captures the leverage effect while retaining the simplicity and interpretability of the standard GARCH framework. The leverage effect is particularly relevant for equity options, where the negative correlation between returns and volatility changes has explicit effects on the hedge ratio, as discussed in Section 2.8.

2.8 GARCH-Based Volatility Forecasting for Delta Hedging

The standard practitioner Black–Scholes delta hedge uses the market-implied volatility as its volatility input. While implied volatility reflects the market's consensus view, it does not directly account for the time-series dynamics of realised volatility, such as volatility clustering and the leverage effect described in Sections 2.7.3 and 2.7.4. GARCH models offer an alternative by providing a conditional volatility forecast based on the recent history of returns.

The theoretical framework of Duan (1995) shows that option pricing under GARCH dynamics requires a risk-neutral valuation relationship (LRNVR). This is an extension of standard risk-neutral pricing where the variance changes over time. Heston and Nandi (2000) derive a closed-form option pricing formula for a GARCH process, but most GARCH specs require Monte Carlo simulation for option valuation. A simpler approach is adapted in this study. Instead, The GARCH model is used, only as a volatility forecasting tool, and the forecasted volatility is substituted into the standard Black–Scholes delta formula. This approach is consistent with Hull (2015), who presents GARCH(1,1) as a practical method for estimating time varied volatility in risk management.

2.8.1 Plug-In GARCH Delta

Given a GARCH volatility forecast $\hat{\sigma}_t^{\mathrm{GARCH}}$ for trading day t, the plug-in GARCH delta is defined as the Black–Scholes delta evaluated at this forecasted volatility:

$$\Delta_{\mathrm{GARCH}} = \Delta_{\mathrm{BS}}(S_t, K, r, \tau, \hat{\sigma}_t^{\mathrm{GARCH}}), \tag{38}$$

where S_t is the underlying price, K the strike price, r the risk-free rate, and τ the time to maturity from t. This GARCH approach replaces the option implied volatility with a single GARCH-based estimate. As a consequence to this, the plug-in delta discards the information contained in the volatility smile, that is, the empirical pattern in which implied volatility varies systematically across strike prices. Near at-the-money, where the Black–Scholes delta is approximately a linear function of volatility, the plug-in delta is close to the standard practitioner delta. For out-of-the-money and in-the-money options, where implied volatility can differ substantially from the GARCH forecast, the plug-in delta may deviate noticeably from the practitioner delta.

The plug-in approach is therefore included primarily as a baseline that deliberately discards the smile information, rather than as a hedge expected to outperform the practitioner delta on out-of-the-money and in-the-money strikes.

2.8.2 Minimum-Variance GARCH Delta

The plug-in approach described above uses the GARCH forecast to replace the volatility input but does not account for the correlation between price movements and volatility changes. However, the GJR-GARCH model provides precisely this information through its asymmetry structure.

From the fitted GJR-GARCH model, the conditional volatility series $\{\hat{\sigma}_k\}$ over a rolling estimation window can be used to estimate the empirical leverage slope:

$$\hat{\beta} = \frac{\mathrm{Cov}(\Delta\hat{\sigma}_k, r_k)}{\mathrm{Var}(r_k)} \tag{39}$$

where $\Delta\hat{\sigma}_k = \hat{\sigma}_k - \hat{\sigma}_{k-1}$ is the change in annualised conditional volatility and r_k is the daily return. This slope quantifies how much the conditional volatility changes, on average, per unit change in the return. For equity underlyings, $\hat{\beta}$ is typically negative, reflecting the leverage effect: negative returns are associated with increases in volatility.

The minimum-variance GARCH delta is then defined as:

$$\Delta_{\text{GARCH-MV}} = \Delta_{\text{BS}} + \nu_{\text{BS}} \cdot \frac{\hat{\beta}}{S_t} \tag{40}$$

where Δ_{BS} and ν_{BS} are the practitioner Black–Scholes delta and vega evaluated at the market-implied volatility, and S_t is the current underlying price. The correction term $\nu_{\text{BS}} \cdot \hat{\beta}/S_t$ adjusts the hedge ratio for the expected change in implied volatility conditional on a move in the underlying, estimated from the GARCH model's conditional volatility dynamics rather than from the implied volatility smile.

This formulation is structurally identical to the minimum-variance hedge ratio of Hull and White (2017), presented in Section 2.6:

$$\Delta_{\text{MV}} = \Delta_{\text{BS}} + \nu_{\text{BS}} \cdot \frac{\partial E(\sigma_{\text{imp}})}{\partial S}$$

The key difference lies in how $\partial E(\sigma_{\text{imp}})/\partial S$ is estimated. Hull and White (2017) derive this quantity empirically from historical regressions of implied volatility changes on underlying price changes. The GARCH-based approach instead estimates it from the time-series leverage effect captured by the GJR-GARCH model. Similarly, the SABR model (Section 2.9) estimates this quantity from the calibrated implied volatility smile slope η. All three approaches therefore share a common theoretical structure but differ in the source of information used to estimate the volatility-price sensitivity.

2.9 The SABR Volatility Model

This section introduces the SABR model as a stochastic volatility framework that is used to describe the IV smile.

2.9.1 From constant volatility to stochastic volatility

The Black-Scholes model as previously mentioned assumes a single constant volatility σ_B. However, empirical data shows that prices observed on the market for options require significantly different volatilities to match their strike K (i.e. implied volatility). This can be viewed when implied volatility is plotted against the strike price, typically producing a downwards sloping curve known as the "volatility smile" or "volatility skew". Volatility is not constant but instead depends on the option's strike and time to maturity (Hagan et al., 2002).

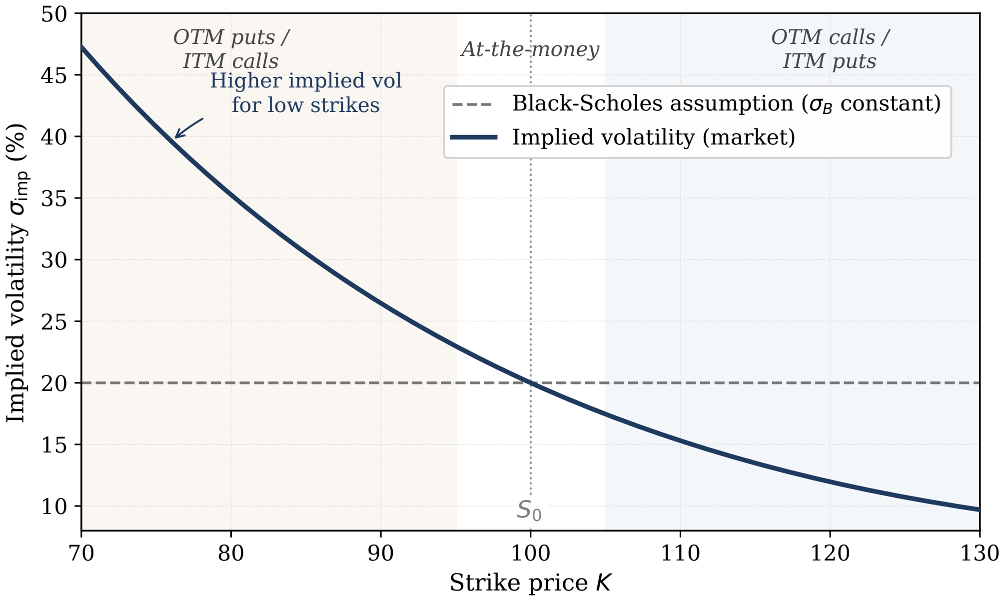

Figure 6: Illustrative example of the volatility skew. The dashed line shows the constant volatility σ_B assumed by Black-Scholes, while the solid curve shows the implied volatilities required to match market prices across strikes.

Hagan et al. (2002) also mention 3 problems with the classic Black-Scholes model. The first is in the pricing of exotic options. Instruments whose payoffs depend on the underlying at multiple price levels (e.g. barrier options) require a volatility input at each level, and there is no principled way to choose between them without a unified model. Second, different implied volatilities are used across strikes. As a result, the corresponding risk measures (delta and vega) are derived from different model specifications. This in turn means that risk exposures across options at different strikes cannot be reliably combined, as there is no single unified model shared by all of them. Third, because the implied volatility itself depends on the current forward price, movements in the underlying can induce systematic shifts in the volatility curve, blurring the distinction between delta risk and vega risk.

2.9.2 The Local Volatility Model

To solve these problems and address the limitations of constant volatility, the local volatility model was developed by Dupire (1994) and Derman and Kani (1994). Rather than treating volatility as a fixed parameter, the local volatility framework replaces the constant σ_B with a deterministic function $\sigma(t, F)$ of both time and forward price, so that the forward dynamics become

$$d\hat{F} = \sigma_{\text{loc}}(t, \hat{F})\, \hat{F}\, dW, \qquad \hat{F}(0) = f. \tag{41}$$

From observed market prices the function can be calibrated. Creating a model that uses a single volatility function across all strikes and maturities, is arbitrage-free, and capable of reproducing the market smile for any given exercise date (Hagan et al., 2002). The model as a result became one of the most widely used approaches for managing exotic equity and foreign exchange options during the late 1990s.

However, despite the model's wide use Hagan et al. (2002) discovered that the model suffered from a critical dynamic flaw. The local volatility model predicts that the market smile moves in the opposite direction to the underlying asset. This is directly contradicted by empirical market behaviour, where smiles and the underlying asset price generally move in the same direction. As a consequence, the delta and vega hedges derived from the local volatility model tend to be disrupted by actual market movements. In many cases these hedges perform worse than the Black-Scholes hedges they were designed to improve upon (Hagan et al., 2002).

This is illustrated by the case where the local vol function depends on $\hat{F}$ only, such that

$$d\hat{F} = \sigma_{\text{loc}}(\hat{F})\, \hat{F}\, dW, \qquad \hat{F}(0) = f. \tag{42}$$

It is found by singular perturbation methods (adding a tiny, structurally critical parameter) that prices of European calls and puts using Black's formula have an implied volatility

$$\sigma_B(K, f) = \sigma_{loc}\left(\tfrac{1}{2}[f + K]\right) \left\{ 1 + \frac{1}{24} \frac{\sigma''_{loc}\left(\tfrac{1}{2}[f + K]\right)}{\sigma_{loc}\left(\tfrac{1}{2}[f + K]\right)} (f - K)^2 + \cdots \right\} \tag{43}$$

The first term in the formula is said to dominate the solution and therefore the behavior of the local volatility model can be understood by just examining this first term.

Retaining only the leading-order term and evaluating it at a reference forward price f_0 (for instance today's observed forward) gives $\sigma_B^0(K) = \sigma_{loc}(\frac{1}{2}[f_0 + K])$, so that $\sigma_{loc}(u) = \sigma_B^0(2u - f_0)$. Evaluating this at $u = \frac{1}{2}(f + K)$ then yields the implied volatility for an option with strike K and current forward price f:

$$\sigma_B(K, f) = \sigma_B^0(K + f - f_0)\{1 + \cdots\} \tag{44}$$

It is easy to conclude that when the forward price f_0 increases, a shift to the left will happen for the implied volatility curve, and if f_0 decreases the opposite will happen. The volatility curve will also experience a shift towards higher levels whether f increases or decreases.

As mentioned previously, the hedges calculated with the local volatility model are also wrong. Differentiating $V_{\text{call}} = \text{BS}(f, K, \sigma_B(K, f), T)$ using the local vol $\sigma_B(K, f)$ with respect to f gives the delta risk under the local volatility model.

$$\Delta = \frac{\partial V_{\text{call}}}{\partial f} = \frac{\partial BS}{\partial f} + \frac{\partial BS}{\partial \sigma_B} \cdot \frac{\partial \sigma_B(K, f)}{\partial f}. \tag{45}$$

The second term is the correction to the normal delta risk (first term) using Black's model. Included in the correction term is the vega risk from Black's model and the anticipated change in σ_B received from changes in f. Since the predictions from the local volatility model and implied volatility in the real markets move in opposite directions, the correction term should have the opposite sign to capture the dynamics of the real markets.

2.9.3 The SABR Model Specification

In 2002 Hagan et al. introduced the SABR (Stochastic Alpha, Beta, Rho) model to resolve the problems with local volatility. Rather than the deterministic local volatility function $\sigma_{\text{loc}}(t, \hat{F})$ used in Equation 42, SABR replaces it with a stochastic process $\hat{\alpha}$ that has its own dynamics. The forward price and its volatility are jointly modelled as correlated stochastic processes. Volatility is therefore random rather than a fixed function of the underlying price.

Under the forward measure, the SABR model is specified by the following system of stochastic differential equations:

$$\begin{aligned} d\hat{F} &= \hat{\alpha}\, \hat{F}^{\beta}\, dW_1, \quad & \hat{F}(0) &= f, \\ d\hat{\alpha} &= \nu\, \hat{\alpha}\, dW_2, \quad & \hat{\alpha}(0) &= \alpha, \end{aligned} \tag{46}$$

with correlation $dW_1\, dW_2 = \rho\, dt$.

where f is the current forward price and α is the level of instantaneous volatility. The first equation describes how the forward price $\hat{F}$ evolves. Driven by a Brownian motion W_1, the price fluctuates randomly over time. The second equation specifies that the volatility itself follows a geometric Brownian motion driven by a second Brownian motion W_2 (i.e. how much price fluctuates also moves randomly). The third equation captures the correlation ρ between movements in the forward price and movements in its volatility. Each of the model's four parameters plays an important role in determining the shape of the implied volatility surface.

The parameter α, as previously mentioned, represents the current level of instantaneous volatility. It acts as a measure of how volatile it is right now and can be thought of as the current level of "fear" in the market. In the implied volatility formula, α primarily controls the overall height of the implied volatility curve. A higher value of α shifts the entire smile upward, while a lower value shifts it downward. In practice, α is often replaced by the at-the-money implied volatility σ_{ATM} since it is directly observable. α is then recovered by inverting the ATM relationship (Eq. 50) (Hagan et al., 2002).

β, lying between 0 and 1, is the CEV (constant elasticity of variance) exponent. It controls the "backbone" of the model, meaning how the at-the-money (ATM) implied volatility responds to changes in the forward price. When $\beta = 1$, the forward dynamics are lognormal and the ATM volatility backbone is approximately flat. When $\beta = 0$, the dynamics are normal (Gaussian), producing a steeper downward-sloping backbone. In practice, β is difficult to determine from a single snapshot of the smile because its effect on the skew closely resembles that of the correlation parameter ρ. As a result, β is typically fixed in advance based on historical analysis of the relationship between ATM volatility and forward price. It could also be set according to market convention, rather than calibrated alongside the other parameters (Hagan et al., 2002).

The parameter ρ is the instantaneous correlation between the forward price process and its stochastic volatility, taking values in the interval $(-1, 1)$. It essentially determines if the underlying asset and volatility move together and if so, how. This parameter primarily controls the skew, i.e. the slope of the implied volatility curve as a function of strike. A negative correlation (normal in equity markets) causes the implied volatility to decrease with strike, producing the commonly observed downward-sloping skew. The more negative ρ is, the steeper this skew becomes (Black, 1976; Christie, 1982).

Finally, the parameter ν, often referred to as the volatility of volatility or volvol, determines how much the stochastic volatility $\hat{\alpha}$ fluctuates over time. So, it measures how unpredictable volatility is itself. In terms of the implied volatility curve, ν primarily controls the curvature or convexity of the smile (i.e. how much the "wings" curve up). A higher ν produces more pronounced curvature, with implied volatilities rising for strikes far from the forward price in both directions. This reflects the intuition that when volatility is itself highly volatile, extreme price movements become more likely, driving up the price of deep out-of-the-money options (Hagan et al., 2002).

2.9.4 Implied Volatility Approximation

One of the main reasons that made the SABR model practically successful is the ability to obtain closed-form approximations for the implied volatility. Using singular perturbation techniques, Hagan et al. (2002) derived an explicit formula for Black's implied volatility $\sigma_B(K, f)$ as a function of the strike K and the current forward price f, given the model parameters α, β, ρ and ν. Whilst obtained through advanced mathematical methods, the approximation only involves elementary functions, making it straightforward to implement. This volatility can then be directly substituted into Black's formula, meaning that the pricing of European options under the SABR model is no more computationally demanding than pricing under the standard Black-Scholes framework.

$$
\begin{aligned}
\sigma_B(K,F) = {} & \frac{\alpha}{(FK)^{(1-\beta)/2}\left[1+\frac{(1-\beta)^2}{24}\ln^2\left(\frac{F}{K}\right)+\frac{(1-\beta)^4}{1920}\ln^4\left(\frac{F}{K}\right)\right]} \\
& \times \left(\frac{z}{x(z)}\right) \\
& \times \left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FK)^{1-\beta}}+\frac{\rho\beta\nu\alpha}{4(FK)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2\right]T\right\}
\end{aligned}
\tag{47}
$$

Where:

$$z = \frac{\nu}{\alpha}(FK)^{(1-\beta)/2} \ln\left(\frac{F}{K}\right) \tag{48}$$

$$x(z) = \ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right) \tag{49}$$

At $K = f$, the SABR implied volatility formula reduces to:

$$\sigma_{\text{ATM}} = \frac{\alpha}{f^{1-\beta}}\left[1+\left(\frac{(1-\beta)^2}{24}\frac{\alpha^2}{f^{2-2\beta}}+\frac{\rho\beta\nu\alpha}{4f^{1-\beta}}+\frac{2-3\rho^2}{24}\nu^2\right)T\right] \tag{50}$$

where T is the time to expiry. The leading term $\alpha/f^{1-\beta}$ indicates that the ATM volatility follows a curve known as the backbone as the forward price f varies.

The model works well for options that are not too far from the money. But it can produce anomalies for deep OTM or ITM options, since the perturbation expansion on which it is based loses accuracy in those regions.

The SABR model introduces two new risk measures corresponding to changes in ρ and ν. The sensitivity of a call option's value to changes in ρ is called vanna, and the sensitivity to changes in ν is called volga (vol gamma). Vanna expresses the risk to an increased skew, and volga expresses the risk of a sharper smile; both can be hedged with away-from-the-money options.

Just as in the local volatility model, the delta in the SABR model consists of the ordinary delta calculated with Black's model and a correction term. Depending on which parametrisation is chosen, the correction term varies quite a bit. Using the parametrisation α, β, ρ, ν so that $\sigma_B(K, f) = \sigma_B(K, f;\, \alpha, \beta, \rho, \nu)$, the correction term is made up of the Black vega multiplied by the change in implied volatility given a change in the forward price:

$$\Delta \equiv \frac{\partial V_{call}}{\partial f} = \frac{\partial BS}{\partial f} + \frac{\partial BS}{\partial \sigma_B}\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \nu)}{\partial f}. \tag{51}$$

If instead the parametrisation $\sigma_{\text{ATM}}, \beta, \rho, \nu$ is used, the delta becomes:

$$\Delta \equiv \frac{\partial BS}{\partial f} + \frac{\partial BS}{\partial \sigma_B}\left\{\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \nu)}{\partial f} + \frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \nu)}{\partial \alpha}\frac{\partial \alpha(\sigma_{ATM}, f)}{\partial f}\right\} \tag{52}$$

Now the correction term consists of the vega times the change in volatility as f changes, plus a term used to keep σ_{ATM} constant with changing f.

In theory, managing risk for option books should be done with the first formula. However, since in many markets changes in volatility are not instantaneous after a large change in f, the equation using σ_{ATM} provides a hedge that is significantly more effective.

2.9.5 Calibration and Practical Use

In practice, the SABR model is calibrated separately for each option expiry date using liquid options. The first step is to fix β, with common choices being 1, 0, or 0.5. The remaining parameters α, ρ, and ν are determined by fitting the SABR implied volatility formula to the volatility smile observed on the market for a given expiry, minimising the difference between market and model implied volatility across all available strikes. It is also highly convenient in practice to use σ_{ATM} as a direct input in place of α, since σ_{ATM} is immediately observable from the market. The SABR formula is then used to evaluate implied volatility at any desired strike, including ones where no liquid market quote exists. The resulting implied volatility is simply plugged into Black's formula to obtain the option price (Hagan et al., 2002).

The SABR model has become the industry standard for interest rate derivatives, especially for European swaptions and caplets/floorlets. This is primarily due to the fact that these instruments are quoted one expiry at a time, which is consistent with the SABR model's single-expiry structure. Equity options, however, are actively traded across many different maturities and require a model that is consistent across the entire volatility surface, making the single-expiry structure more constraining. The standard SABR model does not naturally produce a consistent term structure of volatility, and fitting the volatility surface requires either independent calibration at each expiry or extensions such as the dynamic SABR model (Hagan et al., 2002; Hull, 2018).

2.9.6 Delta Hedging under SABR and Bartlett's Correction

The main purpose of the SABR model in the context of this study is its application to delta hedging. Since the model includes stochastic volatility that is correlated with the forward price, the calculation of the delta is more complex than in the Black-Scholes framework.

The standard delta for the SABR model is given in Eq. 51 and Eq. 52. They do however share a limitation, namely that the correction term depends on the choice of β. This was first observed by Bartlett in 2006 and explored further by Hagan and Lesniewski in 2020. Different values of β can fit the same market smile equally well by adjusting the remaining parameters. Hence, two practitioners calibrating identical market data with different β values will arrive at different hedge ratios, even if their models agree on the price of every option.

Recall that in the SABR model, the stochastic volatility $\hat{\alpha}$ is correlated with the forward price $\hat{F}$ (ρ). This means that when the forward price moves, the volatility level $\hat{\alpha}$ is also expected to shift in a predictable direction. Eq. 51 and Eq. 52 account for how the implied volatility formula changes with f, but neither accounts for this additional expected movement of $\hat{\alpha}$ itself. Bartlett (2006) proposed a correction to address this. By adding the expected change in $\hat{\alpha}$ conditional on a move in f, producing the modified delta:

$$\Delta^{\text{mod}} = \frac{\partial BS}{\partial F} + \frac{\partial BS}{\partial \sigma}\left(\frac{\partial \sigma^{\text{imp}}}{\partial F} + \frac{\partial \sigma^{\text{imp}}}{\partial \alpha} \cdot \frac{\rho\nu}{F^{\beta}}\right), \tag{53}$$

where σ^{imp} is the SABR implied volatility. The additional term $\frac{\partial \sigma^{\text{imp}}}{\partial \sigma} \cdot \frac{\rho\nu}{F^{\beta}}$ captures the expected shift in the SABR volatility level $\hat{\alpha}$ that follows a move in the forward price.

Hagan and Lesniewski (2020) showed that when the two correction terms are combined exactly at the money (K=F), the β-dependent parts cancel, and the expression simplifies to:

$$\Delta^{\text{mod}} \approx \frac{\partial BS}{\partial F} + \frac{\partial BS}{\partial \sigma} \cdot \eta, \tag{54}$$

where $\eta = \frac{\partial \sigma^{\text{imp}}}{\partial K}\Big|_{K=F}$ is the slope of the implied volatility smile evaluated at the money, i.e. the market skew. More precisely, Hagan and Lesniewski (2020) show that this simplification is exact at $K = F$. The residual error grows linearly with the moneyness gap, i.e. it is of order $\mathcal{O}(|F - K|)$ for small departures from the money. The approximation is therefore most accurate near the at-the-money region and progressively coarser in the wings. Both the vega, $\partial BS/\partial\sigma$, and the skew η are directly observable from market prices and do not depend on which value of β was used to calibrate the model. Making the modified delta more robust and simple than the equations seen before (no calibration needed).

$$\Delta_{\text{MV}} = \Delta_{\text{BS}} + \nu_{\text{BS}} \cdot \frac{\partial E(\sigma_{\text{imp}})}{\partial S}, \tag{55}$$

where Δ_{BS} and ν_{BS} are the practitioner Black–Scholes delta and vega, and $\partial E(\sigma_{\text{imp}})/\partial S$ is the expected change in implied volatility per unit change in the asset price. Both these approaches adjust the Black–Scholes delta by a vega-weighted estimate of how implied volatility moves with the underlying. The key difference is in how that estimate is obtained. Hull and White (2017) derive it empirically from historical data, while Bartlett (2006) derives it from the SABR model's calibrated skew η.

2.10 Delta-Vega Hedging

This section introduces delta-vega hedging as an extension of delta hedging. This, to show how a second hedging instrument can be used to neutralise volatility exposure which remains unhedged in a standard delta hedge.

2.10.1 Limitations of Delta-Only Hedging

The delta hedging framework described in Section 2.4 eliminates the portfolio's first-order exposure to movements in the underlying price. However, the P&L of a delta-hedged position still depends on three other risk factors: gamma (the curvature of the option price with respect to the underlying), theta (the loss of value as time passes), and, most importantly here, vega (the sensitivity of the option price to changes in implied volatility). In the Black–Scholes framework, where volatility is assumed constant, this vega exposure is zero by construction. In practice, however, implied volatility changes over time and is correlated with the underlying price, meaning that a delta-neutral portfolio remains exposed to volatility risk (Hull, 2018).

The minimum-variance delta corrections introduced in Sections 2.6, 2.8.2, and 2.9.6 address this problem by adjusting the stock position to partially account for the correlation between price and volatility movements. These adjustments attempt to hedge a two-factor risk, price risk and volatility risk, with a single instrument: the underlying asset. This is inherently limited. A single-instrument hedge can only neutralise one linear risk factor at a time. To simultaneously neutralise both delta and vega exposure, a second hedging instrument is required (Hull, 2018; Taleb, 1997).

2.10.2 Constructing a Delta-Vega Neutral Portfolio

Consider a portfolio consisting of a target option whose risk is to be hedged. Let Δ, ν denote the delta and vega of the target option, and let Δ_H, ν_H denote the delta and vega of a hedge option on the same underlying with the same expiry date. The hedge option is chosen to be approximately at-the-money, since ATM options have the highest vega per unit of premium and therefore offer the most efficient way to buy and sell volatility exposure (Hull, 2018).

To achieve vega neutrality, the portfolio must hold w units of the hedge option such that the total vega of the combined position is zero:

$$\nu + w \cdot \nu_H = 0 \quad \Longrightarrow \quad w = -\frac{\nu}{\nu_H} \tag{56}$$

This vega-neutral ratio w ensures that the portfolio's value is, to first order, insensitive to changes in implied volatility. However, the inclusion of the hedge option introduces additional delta exposure of $w \cdot \Delta_H$. To restore delta neutrality, the stock position must be adjusted. The required number of units of the underlying asset is:

$$a = \Delta + w \cdot \Delta_H = \Delta - \frac{\nu}{\nu_H} \cdot \Delta_H \tag{57}$$

The hedged portfolio then consists of a short position of a units of the underlying and w units of the hedge option, held against the target option. The one-period hedging error is:

$$\varepsilon = \Delta f - a \cdot \Delta S - w \cdot \Delta f_H \tag{58}$$

where Δf is the change in the target option price, ΔS is the change in the underlying price, and Δf_H is the change in the hedge option price (Hull, 2018).

This construction generalises directly. For a whole portfolio of options with aggregate delta Δ_{port} and aggregate vega ν_{port}, the same two-equation system determines the positions in the underlying and the hedge option needed to be simultaneously delta- and vega-neutral (Hull, 2018).

2.10.3 Residual Risks

A delta-vega neutral portfolio is hedged against first-order movements in both the underlying price and implied volatility, but higher-order risks remain. These include gamma risk (the convexity of the option price with respect to the underlying), vanna risk (the sensitivity of delta to changes in volatility), and volga risk (the sensitivity of vega to changes in volatility). In addition, theta, the time decay of the option value, contributes to the P&L at each rebalancing step. These residual exposures are generally smaller than the first-order delta and vega risks, particularly when the hedge is rebalanced daily and the options are not close to expiry (Hull, 2018; Taleb, 1997).

Bakshi et al. (1997) provide empirical evidence that adding vega hedging to a delta hedge substantially reduces hedging errors for equity options, particularly when the underlying volatility process is stochastic (that is, when volatility evolves randomly over time rather than being constant). Their study compares hedging performance across several option pricing models and finds that the improvement from vega hedging is largest for at-the-money and near-the-money options with intermediate maturities, where vega is biggest in absolute terms.

2.10.4 GARCH-Scaled Vega Hedging

The analytical delta–vega hedge described in Section 2.10.2 is, by construction, vega-neutral: it eliminates first-order exposure to changes in implied volatility. A natural question is whether information from a volatility model can be used to do better than a pure vega-neutral hedge. The approach proposed in this study is not a refinement of vega-neutrality but a deliberate heuristic overlay that uses GARCH information to take a directional view on volatility.

The intuition is the following. When the GARCH-forecasted volatility $\hat{\sigma}_{\mathrm{GARCH}}$ exceeds the market-implied volatility $\sigma_{\mathrm{imp}}^{\mathrm{ATM}}$ of the at-the-money hedge option, the GARCH model is signalling that future realised volatility is likely to be higher than the level currently priced by the market. If implied volatility subsequently rises towards the GARCH forecast, a position that is over-hedged in vega will benefit from that rise. The opposite holds when $\hat{\sigma}_{\mathrm{GARCH}} < \sigma_{\mathrm{imp}}^{\mathrm{ATM}}$: the market already prices elevated volatility, and a smaller vega hedge is preferable. The overlay is therefore best understood as a bet on the variance risk premium, on the gap between the physical volatility forecast from GARCH and the risk-neutral expectation embedded in implied vol, rather than as a hedge improvement.

This intuition is implemented by scaling the vega-neutral ratio by the volatility ratio

$$\kappa = \frac{\hat{\sigma}_t^{\mathrm{GARCH}}}{\sigma_{\mathrm{imp}}^{\mathrm{ATM}}}. \tag{59}$$

The GARCH-scaled vega hedge ratio is then

$$w^G = w \cdot \kappa = -\frac{\nu}{\nu_H} \cdot \kappa, \tag{60}$$

and the adjusted stock position is

$$a^G = \Delta - w^G \cdot \Delta_H. \tag{61}$$

When $\kappa > 1$, the position takes on more vega than is needed for vega-neutrality; when $\kappa < 1$, it takes on less. For any constant $\kappa \neq 1$ the position is, by construction, no longer vega-neutral: a residual exposure of $(1 - \kappa)\,\nu\, d\sigma_{\mathrm{imp}}$ to changes in implied volatility is left in the linearised P&L. This residual is intentional, and is precisely the directional bet described above. The implementation details of κ – the bound applied to it and the maturity mismatch between numerator and denominator – are deferred to Section 3.4.4.

2.10.5 Relationship to the Minimum-Variance Delta Correction

A general algebraic observation connects the analytical delta-vega hedge of Section 2.10.2 to the minimum-variance (MV) delta corrections of Sections 2.6, 2.8.2, and 2.9.6. All of those corrections take the form

$$\Delta_{\text{mod}} = \Delta + \nu \cdot \eta, \tag{62}$$

where η is a function of the underlying (and possibly the date) but does not depend on the strike or moneyness of the option. Specifically, $\eta = \hat{\beta}/S$ for the GARCH MV correction and $\eta = \partial\sigma_{\text{imp}}/\partial F$ for the SABR-based correction.

When such a modified delta is used inside the analytical delta-vega hedge with the vega-neutral ratio $w = \nu/\nu_H$, the η term cancels algebraically. Substituting $\Delta_{\text{mod}} = \Delta + \nu\,\eta$ into the adjusted stock delta gives

$$a = (\Delta + \nu\,\eta) - \frac{\nu}{\nu_H}\,(\Delta_H + \nu_H\,\eta) = \Delta + \nu\,\eta - \frac{\nu}{\nu_H}\Delta_H - \nu\,\eta = \Delta - \frac{\nu}{\nu_H}\Delta_H. \tag{63}$$

The leverage correction terms cancel exactly, regardless of what η is, as long as it depends only on the underlying. The same result therefore applies to any single-instrument MV-style correction of this form and not only to the GARCH MV correction.

The intuition behind the cancellation is that the vega hedge instrument absorbs the volatility-price coupling directly, through its own vega exposure, so an additive correction to the stock delta intended to capture the same coupling has nothing left to do. Within the delta-vega framework, the family of single-instrument MV corrections collapses onto the model-free delta-vega hedge: they all produce the same stock position once a second instrument is introduced.

This result motivates the heuristic overlay described in Section 2.10.4. Once the stock-delta channel has been closed by the cancellation argument, the only remaining channel through which GARCH information can enter the delta-vega hedge is the vega ratio itself. The GARCH model's role therefore shifts: rather than adjusting the direction of the stock position, as in the MV framework, GARCH information is used to adjust the magnitude of the vega hedge through the volatility ratio κ.

2.11 Measure of model performance

To evaluate the hedging models' performance, this study uses three related measurements: gain, hedging error and the variance in the hedging error. Together these measurements capture the residual risk post hedging. Following the theory of Hull and White (2017), the best performing hedge, is a hedge that minimizes variance in the value of the hedged position. Specifically, the optimal hedge ratio is the value of Δ which minimizes the variance of $\Delta f - \Delta\Delta S$. Where Δf is the observed change in option price, ΔS is the change in price of the underlying asset and Δ is the hedge ratio given by the used model.

Hull and White express the one-period hedging relationship as:

$$\Delta f = \Delta_{MV}\Delta S + \varepsilon, \tag{64}$$

where hedging error is denoted in this study as the residual term ε. For a model m, the hedging error based on the standard BS hedging (ε_{BS}) can be written as:

$$\varepsilon_{BS} = \Delta f - \Delta_{BS}\Delta S \tag{65}$$

This is a direct consequence of the MV-approach, where the optimal hedge ratio is the ratio that minimizes the residual variance in the hedged position. More generally, for some model m, the hedging error is defined by

$$\varepsilon_M = \Delta f - \Delta_M \Delta S \tag{66}$$

The main risk-based performance measure is variance of the hedging error. Since the means of ΔS and Δf are close to zero, this is equivalent to minimizing the sum of squared errors (SSE). Hence, SSE is used as a practical comparison metric where a lower produced variance gives a lower SSE and a more effective hedge.

To put the measurements in relative terms, Gain is used. Gain is denoted by Hull and White (2017) as the percentage increase in the effectiveness of an MV hedge over the effectiveness of the practitioner Black-Scholes hedge. The definition can be written as:

$$\text{Gain} = 1 - \frac{\text{SSE}(\Delta f - \Delta_{MV}\Delta S)}{\text{SSE}(\Delta f - \Delta_{BS}\Delta S)} \tag{67}$$

A positive gain therefore indicates that the hedge reduces residual risk more relative to the BS benchmark. This in theory can therefore be extended to all models, so the gain for model m can be written as:

$$\text{Gain}_m = 1 - \frac{\text{SSE}(\varepsilon_m)}{\text{SSE}(\varepsilon_{BS})} \tag{68}$$

In turn, this allows for comparison between models like GARCH, SABR against the benchmark BS in an efficient way.

To assess whether the difference in Gain between the comparing models is statistically significant, and not based on noise, Hull and White (2017) derive a t-statistic based on individual squared residuals. Let x_i denote the squared hedging error from the model

m_1, y_i denote the squared hedging error from model m_2 and z_i the squared hedging error from Black-Scholes for observation i. Then, defining

$$Z = \sum_{i=1}^{n} z_i. \tag{69}$$

With these definitions, the difference in Gain can be written

$$\text{Diff} = \frac{1}{Z}\sum_{i=1}^{n}(x_i - y_i) \tag{70}$$

The individual differences $(x_i - y_i)$ are then treated as being drawn from a distribution with mean $\hat{\mu}$ and $\hat{\sigma}$, and with independence, this gives the corresponding t statistic

$$t = \frac{\sqrt{n}\,\hat{\mu}}{\hat{\sigma}} \tag{71}$$

It is further noted by Hull and White (2017) that the assumptions of independence are not likely to hold in practice, and instead it can be expected to exhibit serial correlation in the data. Therefore, a Newey-West adjustment should be applied to the estimated variance. This is seen as especially relevant since the same option may appear on many consecutive trading days. In this study, this procedure is used to assess whether the difference in Gain between models is statistically significant.

The test that connects the gain measure to a hypothesis is the paired forecast comparison test of Diebold and Mariano (1995). The idea is straightforward: for each observation i form the squared-error difference

$$d_i = \varepsilon^2_{BS,i} - \varepsilon^2_{m,i}, \tag{72}$$

which is positive whenever model m produces a smaller squared error than Black–Scholes on that observation. If the two models are equally accurate on average, the mean of d_i should be zero. Testing whether the average is reliably different from zero is therefore a direct test of whether one model is better than the other. The test statistic is the standard t-ratio, $t = \bar{d}/\text{SE}(\bar{d})$.

The complication is the standard error in the denominator. A textbook t-test assumes that observations are independent. In our setting they clearly are not. The same option appears on consecutive trading days. Therefore every option on a given day shares the same market move, and volatility itself moves in clusters. Treating each observation as independent would understate the standard error and overstate significance. Newey and West (1987) introduced a way to fix this. Their HAC (heteroskedasticity and autocorrelation consistent) estimator allows residuals to be both correlated over time and varying in size, and still produces a valid standard error. Intuitively, the adjustment reduces the effective sample size from the nominal count of option-day observations toward the number of approximately independent blocks of trading days. The resulting t-statistic is therefore more conservative whenever the underlying data exhibit serial dependence. Hull and White (2017) also apply this adjustment to their gain comparisons.

In practice, the daily loss differentials are collapsed to one number per trading day before applying the HAC adjustment. The kernel used is the Bartlett kernel (linearly declining lag weights) with a lag length of $L = 30$. A t-statistic of $|t| > 1.96$ is interpreted as significant at the 5 % level, and $|t| > 2.58$ at the 1 % level.

As a complementary measurement, root mean squared error (RMSE) of the hedging error is also used.

$$\text{RMSE}_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{m,i}^2} \tag{73}$$

RMSE allows this study to express the size of the hedging error in the same unit as option price change it is expressed in, and therefore it allows for a more intuitive comparison than SSE alone.

Overall, the performance should therefore be measured from the residual in the hedged position, first through the model specific hedging error, then the variance or SSE and finally through Gain relative to BS as a benchmark. This is both consistent with Hull and White (2017) theoretical framework and also fits the model comparison in this study.

2.12 Asset Selection and Market Suitability

Choosing assets is very important when evaluating a hedge models performance since the suitability of a model is not only dependent on the model itself but also the market it is applied on. This study is based on OptionMetrics data for the period 2015-2025. The markets included are S&P 100, S&P 500 and DJ. But also equity options and ETF options.

Hull and White (2017) use the S&P 500 as their main benchmark and extend to S&P 100 and DJ. Using similar indices but a different timeframe therefore strengthens the comparability with the reference study.

Including equity options and ETF options serves a complementary purpose to the market hedgings. Hull and White's own results suggest that hedging for individual stocks and ETFs are generally weaker performing so including these assets allows for a broader assessment whether the chosen hedging models remains stable outside of indices.

2.13 Market Conditions and Volatility Regimes

This section describes how market conditions are classified in the empirical analysis. The classification is based on volatility regimes, allowing for comparison during both normal and stressed market periods.

2.13.1 VIX: Volatility Index

One way to define market regimes is by using the CBOE Volatility Index (VIX). VIX represents the expected volatility of the S&P 500 over the coming 30 days. The index measures the magnitude of price swings in the S&P 500, meaning that larger movements correspond to higher volatility.

When calculating volatility, there are 2 main methods available. One is using historical data to compute statistical averages, variance, and standard deviation. Another method is using the implied volatility inferred by options prices, and this is what VIX uses. The way it is calculated is by using standard SPX options traded on the CBOE, which all expire on the third Friday of each month, and also the weekly SPX options expiring on all other Fridays. The time to expiry for all options used lie between 23 and 37 days. The volatility is then estimated by weighted prices of puts and calls on the SPX across different strikes. Commonly known as the "Fear Index" because it showcases how stressed the stock markets are. When prices in the market suddenly drop VIX rises and when the market experiences growth VIX decreases (Kuepper, 2025).

By looking at the VIX index it is possible to divide historical data into market regimes. One way to divide the data into buckets is by choosing thresholds for when VIX is deemed low, normal or high. Low levels could be chosen as below 12, high levels as above 20 and normal levels in between. By placing historical data into these buckets, distinct patterns in S&P 500 price changes over subsequent 30-day periods can be found. The low volatility VIX bucket is associated with a tighter, more frequent distribution of small price changes, whereas the high volatility VIX bucket is associated with a much wider distribution, meaning the market is more likely to experience larger (both positive and negative) price swings. These buckets do not necessarily predict the direction of the market but grant some predictive ability with regards to the magnitude of subsequent price movements (Edwards and Preston, 2017).

In other studies such as Woodley et al. (2019), they base regimes on the average VIX over the 30 prior days. A VIX below 20 is seen as low, an average or elevated VIX is seen as between 20 and 30, and a high VIX is above 30.

3 Methodology

This study was conducted as a quantitative replicating study based on empirical data with a modified scope. The analysis was based on Hull and White (2017), which examined how traditional delta hedging can be improved by accounting for the relationship between changes in the price of the underlying asset and its implied volatility. The study served as an analytical framework to investigate whether its conclusions also hold under a different scope, in terms of time period containing differing conditions in regard to volatility. The study was therefore comparative, with the results of the analysis being compared to the patterns identified in the reference study.

The work began with a theoretical background on option pricing, different hedging strategies, and volatility relationships. The Black–Scholes model (Black and Scholes, 1973) was used for option pricing and delta hedging. The study by Hull and White (2017) constituted the main reference and served as a methodological template. Additional literature was used to support and clarify previous findings, assumptions, and mathematical formulations.

3.1 Data Collection and Sources

The study was based on secondary data consisting of options and their underlying Black–Scholes-based Greeks. The dataset was obtained from OptionMetrics, which provides daily closing prices for options and their underlying assets, as well as calculated Black–Scholes measures such as implied volatility and Greeks. The dataset covered the period from 2 January 2015 to 28 August 2025. The asset classes included index options (S&P 500, S&P 100, Dow Jones), equity options, ETF options as well as options on individual stocks included in the Dow Jones during the time frame. The motivation for this data selection was to examine whether the findings of Hull and White (2017) remain valid under new market conditions.

3.2 Data Filtering and Pre-processing

To ensure stable estimates, reduce noise, and maintain consistency with Hull and White (2017), several filtering steps were applied. Observations lacking bid or ask prices, implied volatility, delta, gamma, vega, or theta were excluded. The data was then sorted so that the same option could be identified across two consecutive trading days. Each observation was paired with its subsequent trading day. For each pair the one-period change in the underlying price was expressed relative to the current underlying level, $\Delta S/S$. The one-period change in the option price was also scaled by the same underlying level, $\Delta f/S$. Thus, all empirical hedging errors are computed on a common normalized scale, treating the underlying level as normalized within each of the one-period hedge intervals.

Options with less than 14 days to maturity were excluded. Call options with a practitioner Black–Scholes delta outside the interval 0.05 to 0.95, and put options with delta outside the interval -0.95 to -0.05, were also removed from the dataset to stay true to the method used in Hull and White (2017).

3.3 Regime Classification and Thresholds

Each trading day in the sample is classified into one of three volatility regimes based on the daily closing value of the VIX index. This classification enables a systematic comparison of a hedge performance over various market environments. The convention used in this study is: *Normal* (VIX $<$ 20), *Elevated* (20 $\leq$ VIX $<$ 30), and *Extreme* (VIX $\geq$ 30).

The 20-point cutoff for elevated volatility follows the procedure of S&P Dow Jones Indices (Edwards and Preston, 2017) and is widely adopted in wider literature (Whaley, 2009). The 30-point cutoff is designed to isolate severe stress days, which represent a small fraction of the sample. Financial market stress tends to cluster into specific, identifiable episodes, making this high threshold consistent with the documented concentration of volatility shocks in equity-index markets (Whaley, 2009). While any discrete threshold applied to a continuous variable is inherently arbitrary, this regime classification provides a stable and a replicable structure for evaluating how hedging performance varies with tail risk.

3.4 Implementation

This section describes how each hedging model is implemented in the empirical analysis in order to ensure that the models are evaluated consistently.

3.4.1 Empirical Minimum Variance Hedging

The MV implementation follows the minimum-variance hedging framework in Section 2.6 and uses the performance definitions in Section 2.11. The objective of MV hedging is to choose a hedge ratio that minimizes the variance of the one-period hedging error, which is expressed through the one-period hedging relationship in Section 2.11. The resulting MV hedge ratio is expressed as a correction to the Black-Scholes delta 23

All MV results are computed on one-period option pairs. Each option observation is paired with its subsequent observation, which is the next trading date for the same contract identifier. The changes between the two observations are used as the one-period changes in option value and underlying price described in the hedging error definitions in Section 2.11. Pairs are restricted to short calendar gaps, as described in Section 3.2. Underlying prices S are taken from the delivery-based underlying file and merged into the paired dataset by identifier and date. This ensures that all models use the same underlying series when computing one-period changes.

For each one-period pair, the benchmark hedging error is computed using the Black-Scholes delta, as defined in the performance-measure section. The MV hedge is then implemented using the correction term derived in Section 2.6.1, Eq. 24. In practice, the expected implied-volatility response is not observed directly and must be estimated empirically. In the implementation, this estimated term is applied as a vega-weighted adjustment to the Black-Scholes delta.

To avoid in-sample bias, the MV correction is estimated out of sample using rolling calendar time windows. Three different windows were used: OOS12, OOS24, and OOS36, using 12, 24, and 36 months of training, respectively, applied to the following month. The main results are reported for the OOS12 specification, while OOS24 and OOS36 are used as robustness checks. The estimation is performed separately for calls and puts. For the early months where insufficient or no history exists for a given window length, MV coefficients are not produced, and thus MV deltas and MV hedging errors are missing for those observations.

Once the MV delta is available, the MV hedging error is computed using the same one-period residual definition as for all models, Eq. 66, which in this case means the realized change in option value minus the model hedge ratio times the realized change in the underlying. MV performance is evaluated using the Gain definition in Section 2.11, Eq. 68, which compares the SSE of the MV hedging error to the SSE of the Black-Scholes benchmark error. Because MV deltas are undefined in early periods, before sufficient rolling history exists, all reported MV gains are computed on the common sample. Only observations where both the benchmark error and the MV error are available are included in the SSE ratio. This prevents overstating MV performance due to missing MV errors.

3.4.2 SABR-Based Hedging

The implementation of the SABR model builds on the theory presented in Section 2.9. Namely, the implied volatility approximation of Hagan et al. (2002) and the Bartlett delta correction analyzed by Bartlett (2006) and Hagan and Lesniewski (2020). Two approaches were therefore developed. Approach A calibrates the full SABR model and computes the Bartlett corrected delta numerically (Eq. 53). Approach B avoids calibration entirely and instead estimates the correction from implied volatilities directly observed from the market (Eq. 54). As previously mentioned, the two equations are related. Hagan and Lesniewski (2020) show that the model-dependent parameters cancel when the terms inside the Bartlett correction are combined. The result then simplifies to Eq. 54. This simplification is exact at the money and introduces an approximation error that grows with the distance between the strike and the forward price. In practice, however, the simpler Approach B turned out to be more robust.

The SABR implied volatility was computed using the closed-form approximation from Hagan et al. (2002) with β set to 1 (Eq. 47). The formula contains the term $\ln(F/K)$, meaning that it becomes numerically unstable when the forward price F is close to the strike K (i.e. close to ATM). Whenever $|F - K|/F < 10^{-6}$, the ATM formula was used instead (Eq. 50). Additionally, a floor of 10^{-8} was imposed on the implied volatility output of the SABR formula. This was done to prevent numerical instabilities downstream when the formula returns a near-zero value on degenerate inputs.

As discussed in Section 2.9, Hagan et al. (2002) recommend fixing β in advance based on market convention rather than calibrating it alongside the other parameters. This is because β and ρ have similar effects on the skew, meaning that different combinations of the two can produce the same result.The choice $\beta = 1$ implies lognormal forward dynamics, which is the standard convention for equity index options and is the value adopted by Hull and White (2017) for the S&P 500. Alternative values of $\beta = 0$ and $\beta = 0.5$ were also explored on the S&P 500 sample but produced a poorer fit to the observed implied volatility smile. The value $\beta = 1$ was therefore retained, both for consistency with the reference study and on the basis of this empirical comparison.

For each trading date and expiry pair, the parameters α, ρ, and ν were calibrated. This was done by minimizing the sum of squared differences between SABR implied volatilities and market implied volatilities across all available strikes. The optimization was run from three different starting points to reduce the risk of convergence to local minima, since the objective function is non-convex in (α, ρ, ν). In addition, on each trading date the previous day's calibrated parameters for the same expiry were used as a "warm-start". This was done to reduce the day-to-day estimation noise. The solution with the lowest objective value across starts was retained. The calibration was accepted only if the root mean square error (RMSE) was below 1 % of implied volatility. A stricter threshold of 0.5 % discarded too many calibrations, particularly on days with wide bid–ask spreads. A higher threshold of 2 % produced fits that visibly misrepresented the smile. The 1 % threshold is consistent with Hull and White (2017). Expiry groups with fewer than eight strikes or with very short time to maturity were excluded.

In Approach A, the calibrated SABR parameters were used to compute the modified delta from Eq. 53 numerically. For each option, Black–Scholes delta and vega were evaluated at the SABR-implied volatility. The partial derivatives of the SABR implied volatility with respect to F and α were approximated by central finite differences, bumping each value by ± 0.1 percent (i.e. estimating the slope by looking slightly to the left and right of the point). When the calibration failed for a given expiry, the most recent successful calibration for the same expiry was reused to ensure continuity in the hedging parameters.

For Approach B, the parameter η, i.e. the slope of the implied volatility smile at the money, needed to be estimated. Two choices had to be made for this estimation. The first was what to use as the x-variable and the second was what degree of polynomial to fit. For the x-variable, normalized moneyness $(K - F)/F$ and log-moneyness $\ln(K/F)$ were compared. Since the SABR formula itself is built around $\ln(F/K)$, the smile is more evenly shaped on a logarithmic scale. This allows a simple polynomial to capture it more accurately over a wider range of strikes. For the polynomial degree, a straight line captures the overall tilt of the smile but conflates skew and curvature. A quadratic fit, however, separates the two. Different combinations of moneyness definitions and polynomial specifications were tested. The combination of log-moneyness and a quadratic fit produced the most accurate representation of the volatility smile and resulted in the best hedging performance. It was therefore adopted as the primary method.

Initially, only options with an absolute Black–Scholes delta between 0.25 and 0.75 (near-ATM) were used for the regression, on the grounds that the ATM slope should be estimated from options close to the money. However, removing this restriction and including all available strikes improved performance. This is because the quadratic polynomial is flexible enough to capture the shape of the full smile, and having more data points stabilizes the regression coefficients. A minimum of eight observations per date and expiry group was still required.

Despite being derived from the same theoretical correction, Approach A performed substantially worse than Approach B. On this basis Approach B was adopted as the primary hedging method, and the results section reports only the Approach B numbers. The full bucket-level comparison, the underlying mechanism, and the corresponding tables are presented in the appendix.

3.4.3 GARCH-Based Hedging

The implementation of the GARCH-based hedging models builds on the theory presented in Section 2.7 and Section 2.8. Two approaches were developed, mirroring the structure used for SABR. Approach A treats GARCH as a pure volatility-forecasting tool and feeds the forecast into the standard Black–Scholes delta formula. Approach B uses the asymmetric GJR-GARCH(1,1) specification of Glosten et al. (1993) to derive an empirical leverage slope, which is then used to construct a minimum-variance correction on top of the practitioner Black–Scholes delta. The two approaches are designed to isolate the two ways in which a GARCH model can enter a hedge ratio: through the volatility input itself (Approach A), and through the conditional volatility–return covariance that captures the leverage effect (Approach B).

For each trading date, both models were re-estimated on a rolling window of the 252 most recent daily proportional returns of the underlying asset. A 252-day window corresponds to roughly one trading year and is consistent with Hull (2015), who recommends using one to two years of daily data for GARCH-based volatility forecasts in risk-management applications. Shorter windows produced unstable parameter estimates on volatile dates, while longer windows were slow to adapt to regime changes. Both the plain GARCH(1,1) model and the GJR-GARCH(1,1) model were fitted on the same window in a single pass to keep the two approaches strictly comparable. Model parameters were estimated by maximum likelihood under a zero-mean assumption, in line with Hull (2015) and the standard practice for daily equity-index returns where the conditional mean is negligible relative to the conditional standard deviation. Fits that did not produce a finite one-day-ahead volatility forecast in the interval $(0, 5)$ on an annualised scale were discarded as numerical failures and the corresponding date was excluded from the model's results.

Approach A follows the textbook GARCH plug-in. The one-day-ahead variance forecast from the plain GARCH(1,1) model was annualised using the standard $\sqrt{252}$ scaling, and the resulting volatility $\hat{\sigma}_t^{\text{GARCH}}$ was substituted directly into the Black–Scholes delta formula:

$$\Delta_{\text{GARCH}} = \Phi(d_1), \qquad d_1 = \frac{\ln(F_t/K) + \frac{1}{2}\left(\hat{\sigma}_{\text{GARCH}}\right)^2 T}{\hat{\sigma}_{\text{GARCH}}\sqrt{T}}. \tag{74}$$

This formulation differs from the OptionMetrics reference delta in two ways. First, it uses a single GARCH-based volatility input rather than the option's market-implied volatility; this is the substantive feature of Approach A, since by construction it discards the smile information embedded in the OptionMetrics deltas. Second, the d_1 above omits the cost-of-carry drift term $(r - q)T$ that the OptionMetrics calculation includes. The latter is an implementation simplification: it introduces a small systematic shift on top of the volatility substitution, particularly for options with longer maturities or on dividend-paying underlyings. The shift is acknowledged here as a known confound and is assumed not to dominate the smile-substitution effect that Approach A is designed to expose. The hedge error was computed in the same way as for the Black–Scholes benchmark, $\varepsilon_t^{\text{GARCH}} = \Delta f - \Delta_{\text{GARCH}}\Delta S$, and Approach A is retained primarily as a baseline that isolates the cost of ignoring the smile, since the Black–Scholes delta is highly sensitive to the volatility input for out-of-the-money and in-the-money options.

Approach B was constructed as the minimum-variance correction described in Section 2.8.

From the fitted GJR-GARCH(1,1) model, the conditional volatility series $\{\hat{\sigma}_k\}$ over the estimation window was extracted and converted from the percentage-daily scale used by the optimiser to the decimal-annual scale used by the implied-volatility convention. The empirical leverage slope was computed as

$$\hat{\beta} = \frac{\mathrm{Cov}(\Delta\hat{\sigma}_k,\, r_k)}{\mathrm{Var}(r_k)}, \tag{75}$$

where $\Delta\hat{\sigma}_k$ is the daily change in annualised conditional volatility and r_k is the daily proportional return. To suppress occasional estimation noise on dates dominated by a single extreme observation (for example circuit-breaker days), $\hat{\beta}$ was clipped to the interval $[-1.5,\, 1.5]$. This range comfortably contains the values typically observed for equity indices, and the clip activated on a small fraction of dates in the sample. The clipping is the GARCH analogue of the bound imposed on η in Approach B of the SABR implementation.

The minimum-variance GARCH delta was then defined as

$$\Delta_{\text{GARCH-MV}} = \Delta_{\text{OM}} + \nu_{\text{OM}} \cdot \frac{\hat{\beta}}{S_t}, \tag{76}$$

where Δ_{OM} and ν_{OM} are the practitioner Black–Scholes delta and vega taken directly from OptionMetrics. The structural form is identical to the SABR-based correction in Eq. 54, the difference being that the slope η is here estimated from the conditional volatility–return covariance under the physical measure rather than from the cross-sectional implied volatility smile. By construction, the GARCH-based Approach B and the SABR-based Approach B share the same baseline (the OptionMetrics delta) and the same correction structure $\Delta + \nu\eta$, so that any difference in hedging performance can be attributed to the source of the slope estimate alone.

The hedge error for Approach B was computed in the same way as for all other models, $\varepsilon_{\text{GARCH-MV}} = \Delta f - \Delta_{\text{GARCH-MV}}\Delta S$, and the hedge was rebalanced daily. Performance was assessed using the SSE-based Gain measure relative to the Black–Scholes benchmark introduced in Section 2.11, with the same monthly and delta-bucket reporting structure used for SABR so that the two model families could be placed side by side.

A single methodology caveat on $\hat{\beta}$ should be flagged here. Because the leverage slope is, by construction, a property of the physical return process rather than of the implied volatility surface, Approach B is expected to recover the correct sign of the smile-slope effect but to underestimate its magnitude, since for equity options the volatility risk premium amplifies the spot–vol coupling that the option market prices in. The size of this gap, together with the noise properties of $\hat{\beta}$, is revisited in Section 5.1.

3.4.4 Delta-Vega Hedging

The implementation of the delta-vega hedging models builds on the theory presented in Section 2.10. As argued there, every minimum-variance correction studied in this thesis — Bartlett's correction under SABR (Section 2.9.6), the implied-volatility-smile correction of Hull and White (2017) (Section 2.6), and the GJR-GARCH-based correction of Section 2.8 — attempts to hedge a two-factor risk, exposure to the underlying price and exposure to changes in implied volatility, using a single instrument. Delta-vega hedging removes this restriction by introducing a second instrument that absorbs vega exposure directly, leaving only second-order residuals (gamma, vanna, volga, theta) in the hedge error. Two approaches were developed. Approach A is a model-free analytical hedge that derives both the stock position and the vega-hedge ratio from market-observed Greeks alone, with no calibration. Approach B overlays a GARCH-based scaling on the vega-hedge ratio, deliberately breaking strict vega-neutrality in order to test whether the GJR-GARCH volatility forecast contains information that the implied-volatility quote of the hedge instrument does not.

Both approaches share the same hedge structure and the same selection of the hedge instrument. For each (date, expiry) group in the filtered options data, a single at-the-money option on the same underlying and the same expiry was chosen as the vega-hedge instrument. The selection rule favoured the call option whose practitioner Black–Scholes delta was closest to 0.5; if fewer than three calls were available in the group, the put option whose delta was closest to -0.5 was used instead. At-the-money options were chosen because vega is largest, in absolute terms and per unit of premium, near the money, which makes them the most efficient vega-hedge instruments (Hull, 2018). To avoid degenerate cases on dates where no genuinely at-the-money option was traded, a quality guard rejected the group entirely if the chosen option's delta differed from ± 0.5 by more than 0.15, or if its vega was numerically smaller than 0.01. On rejected dates the hedge was not computed, and the corresponding observations were excluded from the performance measures. The hedge instrument itself was also excluded from the results, since an option cannot serve as its own hedge.

Approach A is the model-free benchmark. For each option in the group other than the hedge instrument (referred to here as the target option), a vega-neutral hedge ratio was computed as

$$w = \frac{\nu^{\mathrm{OM}}_{\mathrm{target}}}{\nu^{\mathrm{OM}}_{H}}, \tag{77}$$

where ν^{OM} denotes the OptionMetrics vega and H denotes the at-the-money hedge option. Holding w units of the hedge option against the target option neutralises the portfolio's first-order exposure to changes in implied volatility but introduces an additional delta of $w \cdot \Delta^{\mathrm{OM}}_{H}$. The stock position required to restore delta-neutrality is therefore the adjusted stock delta

$$a = \Delta^{\mathrm{OM}}_{\mathrm{target}} - w \cdot \Delta^{\mathrm{OM}}_{H}. \tag{78}$$

The hedge error over one trading day was computed as

$$\varepsilon^{\mathrm{DV}} = \Delta f - a \cdot \Delta S - w \cdot \Delta f_H, \tag{79}$$

where Δf and Δf_H are the one-day price changes of the target and the hedge option, and ΔS is the one-day price change of the underlying. No fitting, optimisation, or auxiliary model is involved: all inputs come from the OptionMetrics Greeks on that date, exactly as in Approach A's role for the SABR section. This makes Approach A the strongest theoretically motivated benchmark in the study, since under the linearised P&L $dP \approx \Delta\, dS + \nu\, d\sigma$ the pair (a, w) eliminates both linear sensitivities exactly. Any improvement over Approach A by a competing model can therefore only come from second-order risks or from information outside the linearised one-day P&L.

Approach B introduces GJR-GARCH information into the same hedge through a multiplicative scaling of the vega-hedge ratio. The GJR-GARCH(1,1) model was fitted on the same 252-day rolling window of daily returns used in Section 3.4.3, with identical estimation settings, and the one-day-ahead annualised volatility forecast $\hat{\sigma}_t^{\text{GARCH}}$ was extracted on each date. The scaling factor was defined as

$$\kappa = \text{clip}\left(\frac{\hat{\sigma}_t^{\text{GARCH}}}{\sigma_{\text{imp}}^H},\ 0.5,\ 2.0\right), \tag{80}$$

where σ_{imp}^H is the OptionMetrics implied volatility of the at-the-money hedge option. The scaled vega-hedge ratio and the corresponding adjusted stock delta were then

$$w^G = w \cdot \kappa, \qquad a^G = \Delta_{\text{target}}^{\text{OM}} - w^G \cdot \Delta_H^{\text{OM}}, \tag{81}$$

and the hedge error was

$$\varepsilon^{\text{GARCH-DV}} = \Delta f - a^G \cdot \Delta S - w^G \cdot \Delta f_H. \tag{82}$$

The intuition for the scaling, as set out in Section 2.10.4, is that when the GARCH model forecasts higher one-day volatility than the market currently prices into the at-the-money option ($\kappa > 1$), additional vega-hedging is taken on in anticipation of an upward correction in implied volatility, and conversely $\kappa < 1$ reduces the vega-hedge. The clip to [0.5, 2.0] prevents the heuristic from taking extreme positions on dates where the GARCH fit is unstable or where the at-the-money implied volatility is unusually low; it activates as the binding constraint on a small fraction of dates.

In practice κ is clamped to the interval $[0.5, 2.0]$. This bound is an ad hoc cap, chosen to limit the loss that the directional bet can generate on dates where the volatility ratio is extreme, typically because of unstable GARCH estimates in short rolling windows or transient dislocations between the GARCH forecast and the implied volatility surface. The clip activated as the binding constraint on a small fraction of dates in the sample.

Approach B is included as a directional test of the GARCH–implied volatility ratio rather than as a refinement of vega-neutrality, since for any constant $\kappa \neq 1$ the linearised P&L retains a residual $(1-\kappa)\nu_{\text{target}}\, d\sigma$ term and Approach B does not, in general, dominate Approach A in the linearised model. The conceptual rationale – that the stock-delta channel through which Section 3.4.3 fed GARCH information into the hedge is closed once a vega instrument is introduced – is given in Section 2.10.4, and the empirical question of whether the ratio $\hat{\sigma}_t^{\text{GARCH}}/\sigma_{\text{imp}}^H$ has predictive content for short-horizon implied volatility changes is addressed in Section 5.1.

A known approximation in the construction of κ should be acknowledged. The numerator $\hat{\sigma}_t^{\text{GARCH}}$ is a one-day-ahead forecast under the physical measure, while the denominator σ_{imp}^H is the risk-neutral implied volatility of an option with maturity ranging from a few weeks to several months in the filtered sample. A maturity-matched forecast (an h-step-ahead GARCH variance with h equal to the hedge option's days-to-expiry) would be more consistent. The one-day forecast was retained because (i) it matches the one-day rebalancing horizon of the hedge, and (ii) its computational cost is independent of the cross-section of expiries in the data on each date. The implication of this choice — that Approach B mixes a short-horizon physical forecast with a multi-month risk-neutral expectation — is flagged as a limitation in Section 3.5 and is revisited when interpreting the empirical results.

The hedge was rebalanced daily for both approaches, and performance was assessed using the same SSE-based Gain measure relative to the Black–Scholes benchmark used elsewhere in the study. Reporting was structured by month and by delta bucket, separately for calls and puts, in the same format used for the SABR and GARCH delta hedges, so that all four model families could be compared on a like-for-like basis.

3.5 Method Critique

A thing of note is that the SABR model is derived for European Black-style options (Hagan et al., 2002). The dataset is, however, mixed. The S&P 500 and Dow Jones index options are European, while the individual equity and ETF options are American. For non-dividend-paying stocks, an American call has the same value as an otherwise identical European call (Hull, 2018). American puts are different. They carry a strictly positive early exercise premium even without dividends, because exercising early means that the strike can be invested at the risk-free rate. The SABR implied volatility used to price and hedge them is therefore a further approximation. The size of this approximation is largest precisely where early exercise is most attractive. It is therefore most likely to matter for deep in-the-money puts and dividend-paying assets in stressed markets with high volatility or sharp rate moves (Barone-Adesi and Whaley, 1987). Empirically, the premium is small under normal conditions and often dominated by liquidity effects (Li et al., 2024). The gain for the SABR model for American style options, presented in the appendix, should be read with this in mind.

4 Results

This chapter shows the empirical hedging performance of the previously mentioned models compared to the BS benchmark for options on the S&P500. Results for other tickers are presented in the Appendix. The full samples are first reported, and thereafter decomposed by market regime, option type, and delta bucket.

4.1 Descriptive Statistics

After applying the filtering steps described in Section 3.2, the S&P 500 sample contains 19,844,947 paired option-day observations. The observations are approximately evenly distributed between calls (9,890,552) and puts (9,954,395) over the period from 2 January 2015 to 28 August 2025.

Each observation enters the analysis under a single VIX regime, classified using the scheme described in Section 2.13 (Normal: VIX $<$ 20; Elevated: 20 $\leq$ VIX $<$ 30; Extreme: VIX $\geq$ 30). Table 1 reports the resulting distribution across the three regimes.

The Normal regime accounts for 62.3 % of the observations, the Elevated regime for 29.8 %, and the Extreme regime for 7.9 %. The Extreme regime is also split into 5 well-known events: the China devaluation panic (August 2015), the inverse-VIX ETF unwind ("Volmageddon", February 2018), the COVID-19 outbreak (March–June 2020), the Federal Reserve tightening cycle (2022), the Russian invasion of Ukraine (2022), and the April 2025 tariff shock. The episode of 2022 differs quantitatively from the other mentioned periods. It displays a sustained elevated volatility over an extended interval compared to the sharp spikes characterizing the other events. Figure 7 plots the VIX over the sample with the threshold lines and the named-event windows shaded.

Table 1: Observation counts by VIX regime, S&P 500, 2015–2025.

Regime	N	Calls	Puts	Share of sample
Normal (VIX $<$ 20)	12,358,163	6,137,695	6,220,468	62.3 %
Elevated (20 $\leq$ VIX $<$ 30)	5,919,459	2,967,843	2,951,616	29.8 %
Extreme (VIX $\geq$ 30)	1,567,325	785,014	782,311	7.9 %
Total	19,844,947	9,890,552	9,954,395	100.0 %

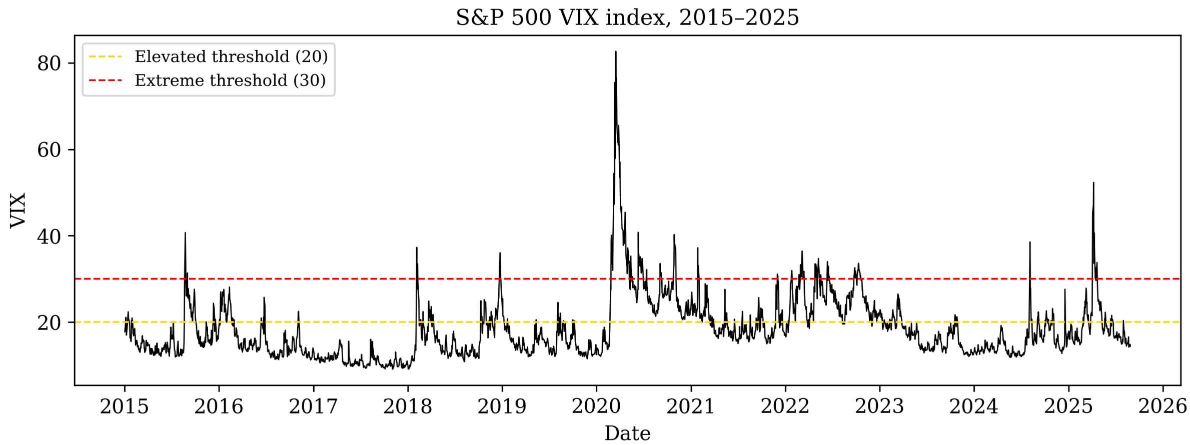

Figure 7: S&P 500 VIX index over the sample period 2015–2025. Dashed lines mark the regime thresholds (20 and 30).

4.2 Overall Performance

Table 2 presents the full-sample results for the primary hedging approaches across signed Black–Scholes delta buckets of width 0.10.

The SABR-B hedge gets a total gain of 3.42 %, a this improvement is not statistically significant after applying the Newey–West HAC correction ($t = 0.27$). The total gain is driven entirely by calls (+14.39 %, $t = 0.88$), meanwhile puts totals a lower gain than Black–Scholes (-8.96 %, $t = -0.99$). For the call surface, gains turn positive for ATM and ITM strikes, and OTM calls lose ground. The put pattern is the mirrored image, positive in the deep OTM wing and increasingly negative toward ITM. Puts reaches the largest underperformance around $\Delta = -0.6$ and $\Delta = -0.7$.

The GARCH-based delta specs, both the plain plug-in delta and the GJR-GARCH minimum-variance correction, produce large negative gains across all buckets, with total gains reaching −89.55 % and −151.950 % respectively. Full bucket-level results for these specs, together with their VIX-regime breakdowns, are reported in Appendix D; the corresponding GARCH-DV regime tables are also gathered there. The interpretation of why the volatility-only specifications fail, and why the GARCH-DV overlay collapses on puts, is referred to Section 5.1.

The model-free delta–vega hedge (DV) produces a total gain of +56.01 %, a significant increase over SABR-B. The call-side gains are notable at the 1 % level from $\Delta = +0.2$ through $\Delta = +0.5$ and at the 5 % level elsewhere. On the put side, gains are positive and notable in the OTM wing but turn negative for deep ITM strikes (−49.68 % at $\Delta = -0.9$, $t = -2.16^{**}$). The GARCH-scaled overlay (GARCH-DV) keeps the call-side gains but collapses on puts. Resulting in a total gain of only +8.13 %. The bucket-level GARCH-DV results and the regime breakdowns are reported in Appendix D.

The empirical MV hedge reduces squared hedging errors by 6.77 % pooled ($t = 1.11$), with the improvement being notable stronger for calls (+9.54 %) than puts (+3.32 %). Call gains are positive across all delta buckets and peak near $\Delta = +0.6$ (+15.29 %), while puts underperform for deep ITM strikes but improve in the OTM wing, reaching +19.34 % at $\Delta = -0.1$. Although economically significant, the overall MV gain does not reach conventional significance levels, unlike the statistically strong improvements documented for the DV hedge.

Table 2: Gain as defined in Eq. 67 for the full sample 2015–2025 for options on the S&P 500 from MV, SABR-B, and DV delta hedging. Results are reported for buckets based on rounding Δ_{BS} to the nearest tenth.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	+5.91	−13.07	−8.69	−0.9	−13.93	−16.79	−49.68
+0.2	+9.87	−10.46	+77.31	−0.8	−7.95	−26.52	−14.56
+0.3	+11.62	−6.94	+90.39	−0.7	−3.94	−30.30	+5.88
+0.4	+12.91	−0.53	+96.06	−0.6	−1.06	−31.20	+27.76
+0.5	+14.68	+7.81	+98.21	−0.5	+4.02	−22.09	+44.53
+0.6	+15.29	+18.49	+95.23	−0.4	+6.78	−5.53	+56.57
+0.7	+10.26	+24.40	+87.71	−0.3	+9.40	+7.69	+63.09
+0.8	+6.88	+24.42	+73.69	−0.2	+17.34	+15.79	+64.85
+0.9	+6.60	+17.76	+43.66	−0.1	+19.34	+20.21	+45.55
All	+9.54	+14.39	+78.05	All	+3.32	−8.96	+31.65

MV pooled: gain = +6.77 %, $t = +1.11$.
SABR-B pooled: gain = +3.42 %, $t = +0.27$.
DV pooled: gain = +56.01 %.

Table 3: Newey–West adjusted Diebold–Mariano t-statistic for the full sample 2015–2025. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	$+2.58^{***}$	+1.29	+1.12	−0.9	−0.79	−0.96	-2.16^{**}
+0.2	+1.59	−0.24	$+3.00^{***}$	−0.8	−0.83	−1.55	-1.91^{*}
+0.3	+1.28	−0.45	$+2.77^{***}$	−0.7	−0.45	-1.81^{*}	+0.61
+0.4	+1.16	−0.21	$+2.69^{***}$	−0.6	0.06	-1.96^{**}	$+1.70^{*}$
+0.5	+1.16	+0.28	$+2.60^{***}$	−0.5	+0.59	-1.70^{*}	$+1.92^{*}$
+0.6	+1.19	+0.78	$+2.54^{**}$	−0.4	+0.94	−0.76	$+1.96^{**}$
+0.7	+1.18	+1.15	$+2.53^{**}$	−0.3	+1.13	+0.26	$+1.95^{*}$
+0.8	+1.21	+1.37	$+2.48^{**}$	−0.2	+1.17	+0.89	$+1.92^{*}$
+0.9	+1.21	+1.50	$+2.40^{**}$	−0.1	+1.02	+0.98	+1.52
All	1.22	0.88		All	0.52	−0.99	

4.3 Hedging Performance Under Normal Conditions

Tables 4 and 5 report the bucket-level gains and Newey–West HAC corrected t-statistics for the Normal VIX regime (VIX < 20, $N = 12{,}358{,}163$, 62.3 % of the sample).

The SABR-B hedge produces a pooled gain of −2.24 % ($t = 0.11$), with a small positive call gain of +1.32 % ($t = 0.25$) and a negative put gain of −6.81 % ($t = -0.05$). None of these differences from Black–Scholes are statistically significant. At the bucket level, call gains turn positive only for ITM strikes ($\Delta \geq +0.7$), while put gains are uniformly negative except in the deep OTM wing.

The DV hedge delivers a pooled gain of +37.26 % ($t = +4.89^{***}$), statistically significant at the 1 % level. The improvement is concentrated on calls (+63.94 %, $t = +7.42^{***}$), where gains are positive and highly significant across all delta buckets, rising from +13.89 % at $\Delta = +0.1$ to a peak of +93.36 % near $\Delta = +0.5$. Put gains are positive in the OTM wing but negative for deep ITM strikes, producing a modest pooled gain of +3.32 % ($t = +1.25$), which does not reach conventional significance.

The MV hedge reduces squared hedging errors by +7.99% relative to Black-Scholes ($t = 2.91^{***}$). The improvement is present for both option types, where calls gain +8.18% ($t = 1.90^{*}$) and puts gain +7.73% ($t = 3.40^{***}$). At the bucket level, call gains are positive across all delta buckets, however there is a decline from +22.68% at $\Delta = +0.1$ to +5.12% at $\Delta = +0.9$. Put gains are also positive across all buckets, ranging from +3.57% at $\Delta = -0.9$ to +12.09% at $\Delta = -0.2$. Thus, under normal conditions, the MV correction improves on Black-Scholes for both calls and puts, although the statistical strength is higher on the put side than on the call side.

Table 4: Gain as defined in Eq. 67 for the Normal VIX regime (VIX < 20) for options on the S&P 500 from MV, SABR-B, and DV hedging. Results are reported for buckets based on rounding Δ_{BS} to the nearest tenth.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	+22.68	+9.79	+13.89	−0.9	+3.57	+3.27	−23.72
+0.2	+18.93	−2.43	+75.36	−0.8	+3.84	−1.74	−20.42
+0.3	+13.88	−12.95	+82.74	−0.7	+4.77	−7.84	−18.10
+0.4	+10.64	−15.04	+88.69	−0.6	+7.05	−18.02	−9.27
+0.5	+10.13	−12.30	+93.36	−0.5	+8.20	−21.16	+3.96
+0.6	+9.70	−5.24	+90.53	−0.4	+9.79	−17.26	+16.15
+0.7	+8.74	+2.96	+79.89	−0.3	+11.07	−8.66	+25.50
+0.8	+7.24	+7.25	+60.77	−0.2	+12.09	+0.33	+34.54
+0.9	+5.12	+6.98	+36.14	−0.1	+10.50	+4.60	+27.54
All	+8.18	+1.32	+63.94	All	+7.73	−6.81	+3.32

MV pooled: gain = +7.99 %, $t = 2.91^{***}$.
SABR-B pooled: gain = −2.24 %.
DV pooled: gain = +37.26 %.

Table 5: Newey–West adjusted Diebold–Mariano t-statistic for the Normal VIX regime (VIX < 20). Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	$+5.23^{***}$	+1.43	$+2.82^{***}$	−0.9	$+3.14^{***}$	$+1.73^{*}$	-3.12^{***}
+0.2	$+4.18^{***}$	−0.33	$+9.06^{***}$	−0.8	$+2.86^{***}$	−0.68	-4.32^{***}
+0.3	$+3.45^{***}$	-1.71^{*}	$+8.95^{***}$	−0.7	$+2.81^{***}$	-2.10^{**}	-2.36^{**}
+0.4	$+2.85^{***}$	-2.06^{**}	$+9.03^{***}$	−0.6	$+2.86^{***}$	-2.99^{***}	−0.19
+0.5	$+2.44^{**}$	-1.67^{*}	$+8.43^{***}$	−0.5	$+3.05^{***}$	-2.98^{***}	+1.44
+0.6	$+2.05^{**}$	−0.78	$+7.72^{***}$	−0.4	$+3.28^{***}$	-2.20^{**}	$+2.34^{**}$
+0.7	$+1.76^{*}$	+0.51	$+7.08^{***}$	−0.3	$+3.36^{***}$	−1.02	$+2.65^{***}$
+0.8	+1.45	$+1.63^{*}$	$+6.66^{***}$	−0.2	$+3.17^{***}$	+0.04	$+2.79^{***}$
+0.9	+0.92	$+2.27^{**}$	$+5.59^{***}$	−0.1	$+2.21^{**}$	+0.52	$+1.86^{*}$
All	$+1.90^{*}$	+0.25	$+7.42^{***}$	All	$+3.40^{***}$	−0.05	+1.25

4.4 Hedging Performance Under Extreme Conditions

Tables 6 and 7 report the bucket-level gains and Newey–West HAC corrected t-statistics for the Extreme VIX regime (VIX ≥ 30, $N = 1{,}567{,}325$, 7.9 % of the sample).

The SABR-B hedge produces a pooled gain of +5.64 % ($t = 0.16$), the largest of the three regimes. The improvement is concentrated on calls (+20.74 %, $t = 0.71$), where bucket-level gains turn positive for ATM and ITM strikes but remain negative in the OTM wing. Puts continue to underperform Black–Scholes (−11.31 %, $t = -1.12$), with losses deepening for ITM strikes. None of these differences are statistically significant.

The DV hedge attains its highest pooled gain in this regime (+64.65 %, $t = +1.67^{*}$), significant at the 10 % level. Call gains are positive across all but the lowest delta bucket, reaching +99.26 % at $\Delta = +0.5$, though the pooled call gain of +84.24 % does not reach conventional significance ($t = +1.60$). On the put side, the pooled gain of +42.85 % is significant at the 10 % level ($t = +1.74^{*}$), with gains positive for $\Delta \geq -0.7$ and large losses confined to the deep ITM buckets (−74.82 % at $\Delta = -0.9$). The GARCH-scaled overlay, by contrast, collapses to a pooled gain of −16.63 % ($t = -0.68$), driven entirely by puts, where the overlay reaches −66.74 %.

The MV hedge produces a pooled gain of +7.73% ($t = 0.57$), which is positive but not statistically significant. The result is strongly asymmetric across option types. For calls the pooled gain is +13.29%, with positive gains across all buckets except $\Delta = +0.1$ (-3.53%). The largest call-side gains occur in the middle and high delta buckets, reaching +17.52% at $\Delta = +0.6$ and +18.30% at $\Delta = +0.9$. For puts the pooled gain is much smaller at +1.47%. Deep ITM puts show losses (-28.63% at $\Delta = -0.9$ and -13.28% at $\Delta = -0.8$), while OTM show stronger gains, reaching +20.95% at $\Delta = -0.2$ and +26.56% at $\Delta = -0.1$. However, no t-statistic for MV exceeds conventional significance thresholds in this regime.

Table 6: Gain as defined in Eq. 67 for the Extreme VIX regime (VIX ≥ 30) for options on the S&P 500 from MV, SABR-B, and DV hedging. Results are reported for buckets based on rounding Δ_{BS} to the nearest tenth.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	−3.53	−28.44	−18.78	−0.9	−28.63	−33.19	−74.82
+0.2	+3.38	−19.32	+75.74	−0.8	−13.28	−53.48	−15.39
+0.3	+9.11	−7.19	+91.60	−0.7	−8.69	−50.42	+19.45
+0.4	+13.26	+4.43	+97.73	−0.6	−6.06	−40.43	+41.25
+0.5	+16.16	+14.77	+99.26	−0.5	+1.24	−24.11	+57.68
+0.6	+17.52	+27.66	+96.72	−0.4	+8.64	−1.35	+70.21
+0.7	+14.16	+34.70	+90.05	−0.3	+18.29	+14.87	+76.86
+0.8	+10.54	+35.30	+79.60	−0.2	+20.95	+23.62	+76.64
+0.9	+18.30	+30.30	+52.94	−0.1	+26.56	+30.80	+57.23
All	+13.29	+20.74	+84.24	All	+1.47	−11.31	+42.85

MV pooled: gain = +7.73 %, $t = 0.57$.
SABR-B pooled: gain = +5.64 %.
DV pooled: gain = +64.65 %.

Table 7: Newey–West adjusted Diebold–Mariano t-statistic for the Extreme VIX regime (VIX ≥ 30). Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Call Options				Put Options			
Δ_{BS}	MV	SABR-B	DV	Δ_{BS}	MV	SABR-B	DV
+0.1	+0.95	−0.72	−0.54	−0.9	−1.18	-2.05^{**}	−1.41
+0.2	+0.51	−0.52	$+1.83^{*}$	−0.8	−1.23	-2.19^{**}	−0.70
+0.3	+0.55	−0.21	$+1.74^{*}$	−0.7	−0.99	-1.89^{*}	+1.27
+0.4	+0.63	+0.13	$+1.70^{*}$	−0.6	−0.63	−1.46	$+1.71^{*}$
+0.5	+0.73	+0.44	$+1.65^{*}$	−0.5	−0.09	−0.86	$+1.71^{*}$
+0.6	+0.82	+0.88	+1.60	−0.4	+0.39	−0.05	$+1.66^{*}$
+0.7	+0.88	+1.23	+1.57	−0.3	+0.67	+0.56	+1.58
+0.8	+0.96	+1.50	+1.52	−0.2	+0.83	+0.97	+1.51
+0.9	+0.97	$+1.96^{**}$	+1.41	−0.1	+0.85	+1.38	+1.23
All	+0.84	+0.71	$+1.67^{*}$	All	−0.17	−1.12	$+1.74^{*}$

4.5 Comparison Across Regimes

Table 8 compares hedging performance across the three VIX regimes for all three models.

For SABR-B, call gains rise monotonically with market stress (+1.32 %, +10.88 %, +20.74 %), while put gains remain negative across all regimes (−6.81 %, −5.53 %, −11.31 %). The pooled gain turns positive and increases with stress (−2.24 %, +2.83 %, +5.64 %), but none of these differences reach statistical significance. The overall performance of SABR-B is thus primarily driven by the call-side gain.

The DV hedge gain also rises monotonically with stress (+37.26 %, +51.25 %, +64.65 %) and is statistically significant at the 1 % level under both Normal and Elevated conditions. Under Extreme conditions the point gain is largest but the HAC-adjusted t-statistic drops to $+1.67^{*}$, indicating only 10 % significance, reflecting the higher volatility of the daily-collapsed loss differential in that regime.

For MV, the regime comparison changes after applying the common VIX regime. The hedge produces positive pooled gains across all three regimes. Gains reach +7.99% under normal conditions, +5.04% under elevated conditions, and +7.73% under extreme conditions. Though, only the normal regime result is statistically significant ($t = 2.91^{***}$). Performance for calls is strongest under extreme conditions, rising to +13.29%, while put-side performance declines from +7.73% under normal conditions to +1.47% under extreme conditions. Thus, MV does not break down during stress. Simoultaneously, however, its statistical reliability is concentrated in the normal regime and its extreme regime performance is driven mainly by calls and OTM puts rather than by puts as a whole.

Table 8: Gain as defined in Eq. 67 by VIX regime for MV, SABR-B, and DV hedging on S&P 500 options, 2015–2025. Newey–West HAC corrected Diebold–Mariano t-statistics in parentheses. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

		MV		SABR-B		DV	
Group	MV N	Gain (%)	t	Gain (%)	t	Gain (%)	t
All options							
Normal	11,708,047	+7.99	+2.91***	−2.24	+0.11	+37.26	+4.89***
Elevated	5,763,997	+5.04	+1.59	+2.83	+0.50	+51.25	+4.97***
Extreme	1,549,156	+7.73	+0.57	+5.64	+0.16	+64.65	+1.67*
Calls							
Normal	5,808,135	+8.18	+1.90*	+1.32	+0.25	+63.94	+7.42***
Elevated	2,889,388	+5.32	+1.58	+10.88	+1.31	+75.67	+7.12***
Extreme	775,791	+13.29	+0.84	+20.74	+0.71	+84.24	+1.60
Puts							
Normal	5,899,912	+7.73	+3.40***	−6.81	−0.05	+3.32	+1.25
Elevated	2,874,609	+4.66	+1.47	−5.53	−0.35	+26.14	+2.55**
Extreme	773,365	+1.47	−0.17	−11.31	−1.12	+42.85	+1.74*

Note: The N column reports the MV OOS12 sample size; SABR-B and DV have model-specific sample sizes.

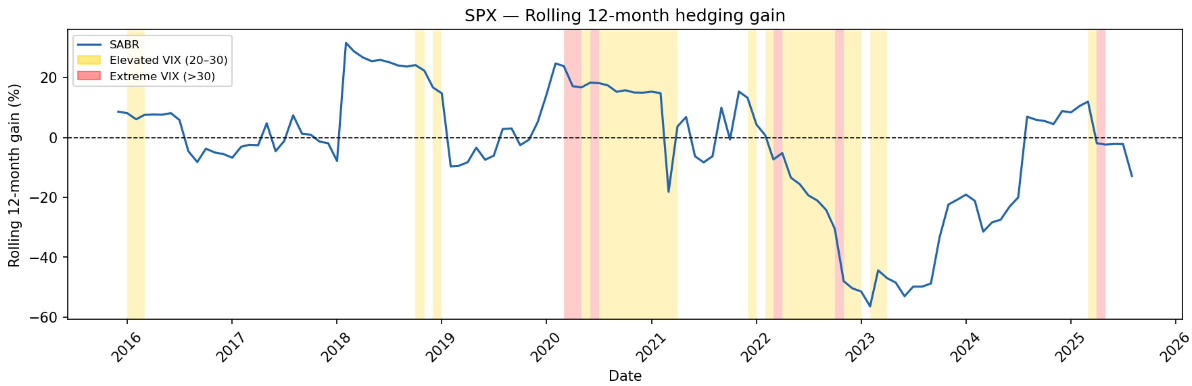

Figure 8: Rolling 12-month SABR-B hedging gain (%) versus Black–Scholes, S&P 500, 2015–2025.

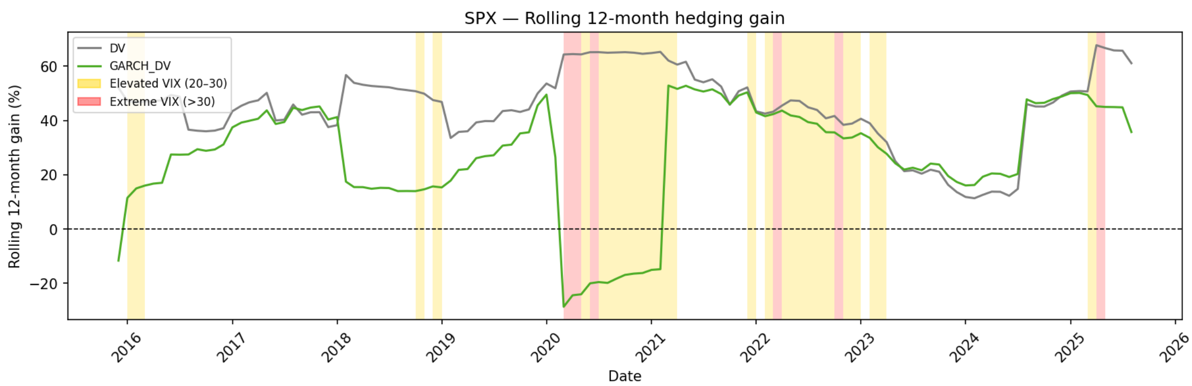

Figure 9: Rolling 12-month DV and GARCH DV hedging gain (%) versus Black–Scholes, S&P 500, 2015–2025.

4.6 Sensitivity Analysis

Figures 10 and 11 plot the rolling 12-month estimates of the quadratic coefficients a, b, and c separately for calls and puts over the full 2015–2025 sample. The coefficient paths remain broadly stable across the sample but exhibit pronounced instability around two episodes. The most severe disruption occurs during the COVID-19 shock of 2020, where both the call and put parameter series reach their most extreme values before partially reverting. The 2022 episode produces a secondary disruption of similar character. Outside of these stress periods, the parameters evolve gradually and maintain consistent signs across the full sample.

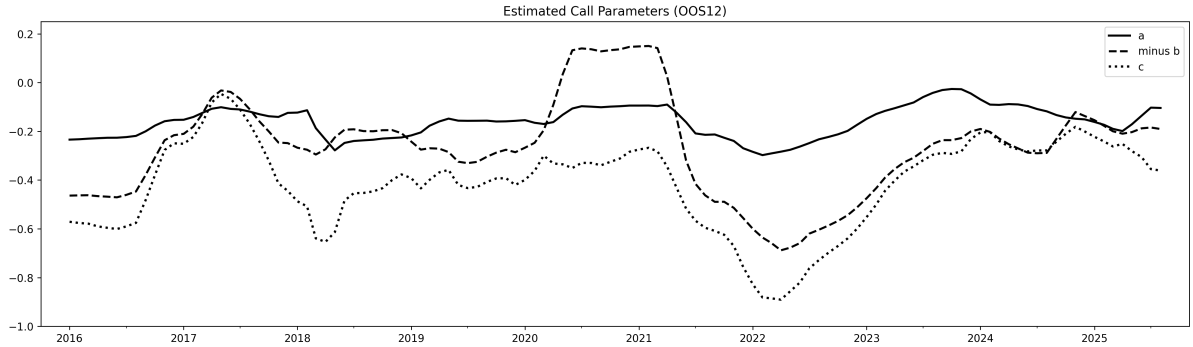

Figure 10: Rolling 12-month for abc calls, S&P 500, 2015–2025.

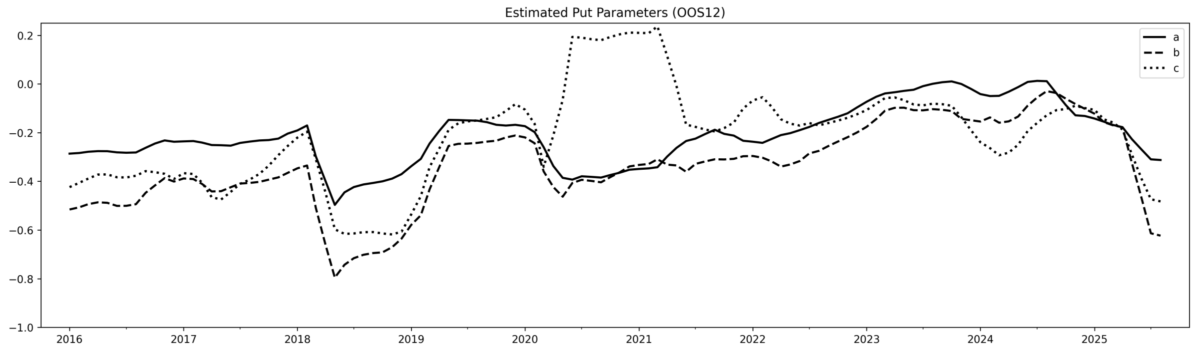

Figure 11: Rolling 12-month for abc puts, S&P 500, 2015–2025.

5 Discussion

The previous chapter showed that hedge ratios incorporating the movement of price and implied volatility jointly outperform BS on the call side. And that the same corrections weaken or reverse on the put side. Also that a model-free delta–vega hedge is the only specification whose gains remain statistically significant across all three VIX regimes. This chapter interprets the results of the study in relation to the study's research questions and also the findings of HullWhite2017. The discussion's focus is to answer why the models perform differently across option types and market regimes rather than restating the result.

5.1 Findings relative to Hull and White (2017)

The findings of this study are not that volatility-based hedging performs universally superior to BS, but that the value of information on volatility is dependent on how it's applied in the hedge ratio. The four model families compared in the last chapter make this clear. MV, SABR-B and DV each adjust the hedge ratios by a quantity proportional to vega, while GARCH only replaces the volatility input. The three models in turn add measurable value on the call side. On calls, the three vega-correction families produce the following: DV with +56.0 %, SABR-B at +14.4 % , and MV with +9.5 %. Where the GARCH plug-in instead had a pooled gain of −89.6 %, and the GJR-GARCH minimum-variance variant deteriorates further to −151.95 %. Models like the GARCH plug-in delta that only replace or forecast volatility perform overall poorly since it does not capture the relevant dependency between the underlying price and implied volatility. Therefore, on the contrary, approaches that adjust the hedge ratio using the dependency mentioned above, such as MV hedging, SABR-B and delta-vega hedging, overall outperform GARCH in parts of the option surface.

The overall SABR-B gain found in this study is +3.42 %, well below what Hull and White (2017) report on the S&P 500 over 2004–2015. This could be due to the fact that the period 2015–2025 contains a number of major events. With volatilities not present in the earlier sample. These events include the inverse-VIX unwind of 2018, the COVID-19 shock of 2020, the Ukraine episode of 2022, and the tariff shock of 2025. The conclusion that a stochastic-volatility correction reliably outperforms Black–Scholes is not statistically supported in this longer and more turbulent sample. Any practical benefit is isolated entirely to call options.

The results following the GARCH model are consistent with the argument by Hull and White (2017) that time-varying volatility does not necessarily improve delta hedging. Both GARCH-based delta hedges perform worse than the Black–Scholes benchmark. The plug-in GARCH delta produces a pooled gain of −89.55 %, while the GJR-GARCH minimum-variance specification deteriorates further to −151.950 % (Tables 18 and 19). Replacing the implied volatility input with a GARCH forecast discards the smile information that the Black–Scholes delta relies on, and the MV correction compounds the problem by applying the estimated leverage slope $\hat{\beta}$ uniformly across all strikes. Although GARCH can describe volatility clustering in the underlying asset, it is based on historical realised return volatility rather than the implied volatility reflected in option prices, so substituting it for the OptionMetrics implied volatility worsens, rather than improves, the hedge.

A second mechanism behind the GJR-GARCH MV underperformance is the noise content of $\hat{\beta}$ itself. In 252-day windows of equity-index returns, the contemporaneous covariance between conditional-volatility changes and returns typically explains only a small fraction of the conditional-volatility dynamics; the remainder is vol-of-vol and idiosyncratic. The leverage slope $\hat{\beta}$ recovered from the rolling fit is therefore a noisy estimate of the true spot-slope even before any clipping, and the magnitude of the correction $\nu \cdot \hat{\beta}/S$ is mechanically modest. Combined with the physical-vs-implied measure gap flagged in Section 3.4.3, this explains why the GJR-GARCH MV correction not only fails to recover the SABR-B gain but actually deteriorates relative to the plug-in: the correction is small, biased toward zero, and applied uniformly across strikes where the true smile slope varies sharply with moneyness.

For the MV hedge, the results provide some support for the mechanism proposed by Hull and White (2017). Across the full SPX OOS12 sample, the MV hedge reduces squared hedging errors by 6.77% relative to Black-Scholes. The improvement is economically meaningful, but the result is not statistically significant. This suggests that the volatility-price correction remains relevant in the more recent 2015-2025 sample, but that its advantage is somewhat less uniform than Hull and White (2017) would suggest. The strongest support can be seen under normal market conditions. In this regime, the MV hedge produces a pooled gain of +7.99% ($t = 2.91^{***}$). The improvement is present for both calls and puts, with calls gaining +8.18% ($t = 1.90^{*}$) and puts gaining +7.73% ($t = 3.40^{***}$). This indicates that the empirical correction works best when the relationship between underlying price movements and implied volatility changes is relatively stable. Thus Hull and White (2017) intuition is partly confirmed. The Black-Scholes delta is incomplete because it ignores the systematic co-movement between the underlying and implied volatility.

Following the results, the only GARCH specification that adds value is the GARCH-scaled delta-vega hedge. However, this improvement only appears when GARCH is applied on top of an already vega-neutral hedge, as shown in Table 21. The findings therefore support the point made by Hull and White (2017): improving hedging is not only about allowing volatility to vary over time, but about modelling the joint movement between the underlying price and volatility.

The fact that the only GARCH specification that adds value is the GARCH-scaled delta–vega overlay is consistent with the algebraic result of Section 2.10.5. Any additive correction to the stock delta of the form $\Delta + \nu\eta$, where η depends only on the underlying, cancels exactly inside the analytical delta–vega framework: the contribution from the target option is offset by the contribution from the at-the-money hedge, weighted by w. This holds for both the SABR and the GARCH minimum-variance corrections, and it implies that the stock-delta channel through which Section 3.4.3 fed GARCH information into the hedge is closed once a vega instrument is introduced. Approach B's scaling of w is therefore the only remaining channel through which a GARCH model can affect the hedge, which reframes the empirical question. Rather than asking whether GARCH improves the delta directly – a channel the cancellation argument shows is closed – the relevant question becomes whether the ratio $\hat{\sigma}_t^{\mathrm{GARCH}}/\sigma_{\mathrm{imp}}^{H}$ has predictive content for short-horizon implied volatility changes. The headline numbers reported in Section 4.2 say it does on calls and does not on puts, and the deterioration is largest precisely in the regimes where the OTM put smile is steepest.

The full-sample MV pooled gain of +6.77 % is qualitatively consistent with Hull and White (2017) but substantially lower in magnitude than the roughly 25 % call gain and 23 % put gain they report on the S&P 500 over 2004–2015. The call–put asymmetry is the same: MV gains +9.54 % on calls against +3.32 % on puts pooled, in line with Hull and White (2017), who attribute weak put performance to a high level of idiosyncratic noise in put implied volatilities not driven by movements in the underlying.

The lower overall gain relative to Hull and White (2017) warrants closer examination.

One contributing factor is the composition of the 2015–2025 sample, which contains several prolonged periods of elevated idiosyncratic noise in implied volatilities. During such episodes the relationship between implied volatility changes and underlying price movements weakens, the quadratic correction captures less of the total variation in option price changes, and the out-of-sample parameters estimated in the preceding window become less representative of the current regime. A second factor is that Hull and White (2017) report their results over a sample that includes the 2008 financial crisis, a single concentrated stress event after which the volatility–price relationship quickly restabilised. The 2015–2025 period instead contains multiple distinct shocks in closer succession, leaving less time between episodes for the rolling estimates to converge to stable values. This repeated parameter instability, visible in the pronounced excursions of a, b, and c in Figures 10 and 11, mechanically suppresses the reported gain by inserting periods where the correction adds noise rather than reducing it. The MV framework remains structurally sound, but its out-of-sample performance is sensitive to the frequency and severity of regime changes in the estimation period.

The same call-versus-put asymmetry is present for SABR-B. Hull and White (2017) themselves describe the put result as a puzzle that is most plausibly explained by idiosyncratic noise in put-option prices that is not driven by movements in the underlying (Hull and White, 2017, p. 17–18), and the same explanation appears to apply here.

A second factor that is specific to Approach B is the structure of the Bartlett correction itself. Hagan and Lesniewski (2020) show that the modified delta reduces to vega multiplied by the at-the-money slope of the smile, $\Delta^{\text{mod}} \approx \Delta^{\text{BS}} + \nu^{\text{BS}} \cdot \eta$, where $\eta = \partial\sigma^{\text{imp}}/\partial K$ evaluated at $K = F$. This expression is exact at the forward and is otherwise an approximation. S&P 500 puts are routinely traded across a wide range of out-of-the-money strikes, where the smile is steep (Benzoni et al., 2011). As a result, the same correction that helps the call hedge tends to be over-applied on the put wing, where the formula is being used away from the strike at which it is exact. The smile in this region is not only steep but also visibly curved, so a linear correction built from the ATM slope alone cannot capture its true shape (Zhang and Xiang, 2008).

5.2 Comparative Hedging Performance Under Normal Market Conditions

Under what has been defined as normal market conditions, the pure GARCH delta approaches do not improve on the Black–Scholes benchmark. This suggests that forecasted volatility using GARCH alone is not a viable way to obtain a more accurate hedge ratio in calmer markets. Instead, the results indicate that the GARCH volatility forecast does not capture the information contained in the market-implied volatility surface well enough to improve the hedge.

The GARCH-scaled delta-vega hedge performs better. It produces a pooled gain of +38.97 %, which is slightly higher than the unscaled delta-vega hedge, which produces a pooled gain of +37.26 %. It also increases the put-side gain from +3.32 % to +12.25 % under the normal VIX regime (Tables 22 and 23) . This indicates that GARCH can add value when it is used as a scaling tool, but only within the vega-hedging framework.

The MV hedge under normal market conditions is the strongest delta-only alternative to Black-Scholes, with a statistically significant ($t = 2.91^{***}$) pooled gain of 7.99%. The improvement is present for both option types, with calls gaining +8.18% ($t = 1.90^{*}$) and puts gaining +7.73% ($t = 3.40^{***}$). This distinguishes it from SABR-B and the GARCH, whose normal-regime gains are either weaker, negative or statistically insignificant. A possible reason behind this is that the MV hedge uses historical option-price data to estimate the volatility-price correction directly, rather than through a parametric volatility model. In relatively stable markets, this rolling empirical relation seems persistent enough for the estimated correction to improve the hedge ratio. This could explain why the MV result is statistically strongest in the normal regime, where volatility dynamics are less abrupt and the relationship between underlying returns and implied-volatility changes is easier to estimate. Though, the comparison with the model-free delta-vega hedge also highlights the limitations of MV. The MV hedge still tries to absorb volatility risk through a single adjusted position in the underlying asset. The delta-vega hedge instead introduces an additional option position and directly neutralises first-order exposure to implied volatility. This helps explain why MV improves considerably on Black-Scholes under normal conditions, while the delta-vega hedge is the strongest performer overall.

The improvement is relatively small, and a likely explanation is that market conditions are already stable in the normal VIX regime. This means that the GARCH scaling factor remains close to one and therefore makes only a limited adjustment to the hedge. The results in the normal regime therefore suggest that GARCH can be used for minor adjustments on an already strong hedge, but not as a standalone improvement to Black–Scholes.

In the Normal regime the smile is shallow and the at-the-money slope η is small in absolute terms, so the Bartlett correction $\nu^{\mathrm{BS}} \cdot \eta$ adds only a small adjustment to the standard Black–Scholes delta. With the correction this small, any structural over-correction on the put side is enough to push the average gain slightly below zero. The pattern in Table 4 is consistent with this reading. The SABR-B pooled gain on calls in the Normal regime is only +1.32 %, the put gain is −6.81 %, and the overall option gain is −2.24 %. SABR-B is therefore essentially indistinguishable from Black–Scholes in calm markets, with a small bias on the put side.

5.3 Model Resilience and Breakdown During Extreme Market Stress

Under extreme market conditions, the GARCH-DV overlay deteriorates sharply. The pooled gain falls from positive territory under Normal and Elevated conditions to −16.63 % in the Extreme regime, while the unscaled model-free DV hedge reaches its highest pooled gain of +64.65 % in the same regime (Table 6 and the regime breakdown in Appendix D). The deterioration is concentrated entirely on puts, where the overlay underperforms DV by tens of percentage points. Notably, this is the regime in which volatility information should be most valuable.

The most plausible mechanism is that GARCH reacts too strongly to realised volatility during stress periods. When realised volatility rises sharply above ATM implied volatility, the scaling factor κ pins to its upper clip on a large share of dates, which amplifies the vega leg in precisely the direction that fails when the implicit variance-risk-premium bet misfires. The result is that GARCH information becomes a liability rather than an asset in stressed markets, even though the underlying intuition — that GARCH should add the most value when volatility itself is most volatile — predicts the opposite.

Overall, GARCH deteriorates in the market conditions where it should be most useful, which makes the GARCH-scaled overlay less robust than the simpler model-free hedge and supports the view that effective hedging requires modelling the joint movement between price and volatility, rather than volatility itself alone.

The Extreme regime is the opposite. Vega exposure is large, the smile is steep, and the leverage effect documented by Black (1976) and Christie (1982) is strongly negative, meaning that implied volatility tends to rise sharply when the spot price falls. Hull and White (2017) point to this same leverage relationship as the standard explanation for why the minimum-variance delta on equity-index calls is consistently below the Black–Scholes delta (Hull and White, 2017, p. 19). Approach B captures the relationship through a single at-the-money slope per (date, expiry) pair, so its correction is largest exactly when delta hedges are most sensitive to volatility moves. The fact that put hedging deteriorates further in this regime mirrors the mechanism described in Section 5.1: a steeper smile slope amplifies the over-correction that already affects the put wing, even as it lifts the call-side hedge.

The MV hedge shows a related but more moderate stress-period pattern. In the extreme regime, the pooled MV gain remains positive at 4.83%, but it is substantially weaker than under normal conditions and is not statistically significant. This suggests that the rolling MV correction does not break down completely during stress, but becomes less reliable when volatility dynamics change rapidly. The bucket results show that this weaker result hides a strong asymmetry across the option surface. Calls remain positive in most delta buckets, while puts combine large losses for deep in-the-money contracts with gains for out-of-the-money contracts. This indicates that MV captures part of the leverage relationship during stress, but may still adjust too much in regions of the put surface where the volatility smile becomes steep. Compared with the GARCH, MV is therefore more robust, but compared with the model-free delta-vega hedge it remains less robust because it still tries to absorb volatility risk through a single adjusted position in the underlying asset.

In conclusion, the results from the extreme market regime suggest that model complexity alone does not guarantee robustness against market movements. It is more dependent on if a model can capture the relevant hedge channel during stress. The MV hedge lies between these cases. It remains positive overall in the extreme regime, with a pooled gain close to the normal regime. However, the gain is not statistically significant and is much more uneven across option types. The extreme regime MV result is driven mainly by calls and OOM puts, while the put side as a whole remains close to neutral. The model-free delta-vega hedge remains robust since it neutralises volatility exposure; meanwhile, GARCH becomes fragile because it transforms volatility forecasting into larger hedge adjustments.

5.4 Generalizability of Hull and White (2017)

The third research question asks to what extent the findings of Hull and White (2017) generalize to different markets, different asset classes and for a different, more recent time period. From the results, Hull and White (2017) broader intuition is supported. The intuition that hedging performance is increased when the hedge ratio takes into account the relationship between price movements in the underlying asset and the changes in volatility.

The most noticeable support for Hull and White (2017) is found in the empirical MV hedge, which generally improves on the BS benchmark across all three VIX regimes. The support is statistically strongest under normal market conditions, where the pooled MV gain is significant. Under elevated and extreme conditions the point estimates remain positive, but the gains are not statistically significant. This indicates that the core mechanism in Hull and White (2017) remains relevant in the more recent sample, but that its reliability depends on regime stability and option type.

The results from this study also show that this conclusion does not generalize uniformly across market regimes and models. SABR-B performs best for call options, and the call gain rises steadily with stress, from $+1.32\,\%$ in the Normal regime to $+20.74\,\%$ in the extreme regime (Table 8). Put performance moves in the opposite direction, deteriorating from $-6.81\,\%$ to $-11.31\,\%$ across the same regimes. This suggests that the effectiveness of volatility-based corrections depends on where the option is located on the implied-volatility surface and that the put side is more difficult to hedge with a single volatility-slope correction.

The performance of the GARCH model further limits the generalizability of Hull and White (2017) because the GARCH-based hedges don't consistently improve performance, even though it is designed to model time-varying volatility. It replaces implied volatility with a volatility forecast and, in turn, ignores important smile information. The GARCH-scaled delta-vega hedge performs overall better in normal conditions but deteriorates during extreme market regimes, when volatility forecasts become unstable and, in turn, amplify hedge adjustment. Which in turn shows how modeling volatility alone is not sufficient.

The model-free delta-vega hedge provides strong evidence that the BS biggest weakness is mainly related to volatility exposure. Unlike GARCH, it does not rely on forecasting realised volatility, but instead neutralises vega exposure. Its strong performance across regimes and in particular its performance in extreme conditions suggests that the most robust extension of the intuition behind Hull and White (2017) is a hedge that can directly target the volatility risk that is left unhedged by the standard delta hedging.

Overall, the answer to this study's third research question is that the findings of Hull and White (2017) generalize in principle. However, the result does not support a claim that all volatility-based or MV inspired models outperform the BS benchmark across markets, option types and regimes. Generalizability depends on mainly three factors, does the model use implied rather than realised volatility information, is the correction stable across moneyness and does it remain robust in market stress. In this study, MV provides the most direct support towards the Hull and White framework meanwhile model-free delta-vega provides additional support for the idea that volatility exposure is central to hedging performance. Meanwhile SABR-B provides some support and GARCH clearly showcases its limitations.

References

Bakshi, G., Charles, C., and Zhiwu, C. (1997). Empirical performance of alternative option pricing models. Journal of Finance.

Barone-Adesi, G. and Whaley, R. E. (1987). Efficient analytic approximation of american option values. The Journal of Finance, 42(2):301–320.

Bartlett, B. (2006). Hedging under SABR model. Wilmott Magazine, pages 2–4.

Benzoni, L., Collin-Dufresne, P., and Goldstein, R. S. (2011). Explaining asset pricing puzzles associated with the 1987 market crash. Journal of Financial Economics, 101(3):552–573.

Black, F. (1976). Studies of stock price volatility changes. Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section, pages 177–181.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3):637–654.

Bodie, Z., Kane, A., and Marcus, A. J. (2021). Investments. McGraw-Hill, 12th edition.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3):307–327.

Borell, C. and Lindberg, C. (2023). Introduction to the Black-Scholes Theory.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. Journal of Financial Economics, 10(4):407–432.

Derman, E. and Kani, I. (1994). Riding on a smile. Risk, 7(2):32–39.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13(3):253–263.

Duan, J.-C. (1995). The GARCH option pricing model. Mathematical Finance, 5(1):13–32.

Dupire, B. (1994). Pricing with a smile. Risk, 7(1):18–20.

Edwards, T. and Preston, H. (2017). A practitioner's guide to reading vix. Accessed: 2026-04-16.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4):987–1007.

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. Journal of Economic Perspectives, 15(4):157–168.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance, 48(5):1779–1801.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. Wilmott Magazine, 1:84–108.

Hagan, P. S. and Lesniewski, A. S. (2020). Bartlett's delta in the SABR model. arXiv:1704.03110v2.

Heston, S. L. and Nandi, S. (2000). A closed-form GARCH option valuation model. The Review of Financial Studies, 13(3):585–625.

Hull, J. and White, A. (2017). Optimal delta hedging for options. Journal of Banking and Finance.

Hull, J. c. (2015). Risk Management and Financial Institutions. Wiley.

Hull, J. C. (2018). Options, Futures, and Other Derivatives. Pearson, 10th edition.

Kuepper, J. (2025). Understanding the cboe volatility index (vix) in investing. Accessed: 2026-04-16.

Li, W., Zhang, J. E., Ruan, X., and Aschakulporn, P. (2024). An empirical study on the early exercise premium of american options: Evidence from oex and xeo options. Journal of Futures Markets, 44(7):1117–1153.

Mandelbrot, B. (1963). The variation of certain speculative prices. Journal of Business, 36(4):394–419.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2):347–370.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3):703–708.

Stoll, H. R. (1969). The relationship between put and call option prices. Journal of Finance, 24(5):801–824.

Sveriges riksdag (2003). Lag (2003:460) om etikprövning av forskning som avser människor. Svensk författningssamling. Accessed: 2026-04-28.

Taleb, N. N. (1997). Dynamic Hedging: Managing Vanilla and Exotic Options. Wiley.

Vetenskapsrådet (2024). God forskningssed. Technical report, Vetenskapsrådet. Accessed: 2026-04-28.

Whaley, R. E. (2009). Understanding the VIX. Journal of Portfolio Management, 35(3):98–105.

Woodley, M., Jones, S. T., and Reburn, J. P. (2019). The "fear index" and the effectiveness of piotroski model. Quarterly Journal of Finance and Accounting.

Zhang, J. E. and Xiang, Y. (2008). The implied volatility smirk. Quantitative Finance, 8(3):263–284.

A Approach A vs. Approach B for the SABR Hedge

This appendix supports our choice of Approach B as the primary SABR specification. Approach A calibrates the full SABR model to the daily smile each day. It then computes the Bartlett delta from the calibrated parameters using Eq. 53. Approach B takes a different route. It reads the at-the-money slope of the smile directly from market quotes and applies the simplified correction in Eq. 54. Hagan and Lesniewski (2020) show that the two specifications coincide at $K = F$. Away from the money, Approach B introduces an error of order $\mathcal{O}(|F - K|)$.

On the full S&P 500 sample, Approach A produces a pooled gain of $-20.88\,\%$ relative to Black–Scholes. Approach B delivers $+3.42\,\%$ over the same sample. Approach A is therefore uniformly worse on the S&P 500 (Table 9). Two factors could be behind this. The first concerns the closed-form formula of Hagan et al. (2002), which is an asymptotic expansion in moneyness. Differentiating it amplifies the expansion error in the wings. The second factor is calibration noise. The analytic correction depends explicitly on the fitted parameters (α, ρ, ν), and these carry day-to-day calibration noise that propagates into the derivatives. Approach B sidesteps both problems by reading a single market-observed slope. The price it pays is the $\mathcal{O}(|F - K|)$ simplification. This finding lines up with the broader argument in Hull and White (2017) that simple corrections built from observable quantities can outperform more elaborate model-based ones.

We ran the same comparison on three other tickers for which a full Approach A calibration was produced. Approach B is the higher-gain specification on every name except IAU, where both approaches are essentially flat. The difference is largest on tickers with the steepest equity-style smiles, namely SPX and MSFT. It is smallest on the gold ETF, where there is little smile curvature for either approach to capture.

Table 9: Approach A vs. Approach B overall gain by ticker.

Ticker	N	Approach A (%)	Approach B (%)	Diff B–A (pp)
SPX	19,844,947	−20.88	+3.42	+24.30
MSFT	2,282,086	−9.69	+2.63	+12.32
DJX	1,269,846	−2.48	+1.14	+3.62
IAU	291,960	−0.13	−0.35	−0.23

The advantage of Approach B over Approach A shows up in every VIX regime, but it is largest under calm conditions. On the S&P 500 the gap is +41.6 percentage points in the Normal regime, +18.6 in the Elevated regime, and +21.1 in the Extreme regime (Table 10). This pattern fits the two mechanisms above. In calm markets the smile is shallow and the calibration is well determined in absolute terms. The analytic derivatives that build Approach A pick up most of their leverage from the wings, however, and that is exactly where the asymptotic formula of Hagan et al. (2002) is least accurate. In stressed markets the smile is steep, so the wings carry more weight in both approaches. The sign of the leverage effect documented by Black (1976) and Christie (1982) is then at least partly captured by Approach A as well, narrowing the gap. The result is that Approach B remains the preferred specification across regimes. Its advantage is largest where Approach A's analytic noise dominates the signal.

Table 10: Approach A vs. Approach B by VIX regime, S&P 500, 2015–2025.

Regime	N	Approach A (%)	Approach B (%)	Diff B–A (pp)
Normal	12,358,163	−43.87	−2.24	+41.63
Elevated	5,919,459	−15.74	+2.83	+18.57
Extreme	1,567,325	−15.50	+5.64	+21.14

B Cross-Ticker Robustness

This appendix collects the Approach B gain across the broader set of tickers in our dataset. We group them into equity indices (SPX, OEX, DJX), individual equities (eight Dow constituents), and commodity ETFs (SLV, IAU, USO). The American exercise feature on the equity and ETF subsets introduces an additional approximation, since SABR is derived for European options (Hull, 2018). This approximation bites hardest where early exercise is most attractive.

Table 11 reports the overall gain. The indices behave similarly to the S&P 500. They show positive call gains and negative or near-zero put gains. The individual equities follow the same pattern across calls and puts. The put underperformance is amplified by the early-exercise mismatch on dividend-paying names such as HON and SHW. On the commodity ETFs the gain is close to zero on both sides. A plausible explanation is that the persistent negative spot–vol correlation that drives the equity-leverage smile (Black, 1976; Christie, 1982) does not generalise to precious metals or oil. Without that structure, there is little for the SABR correction to exploit.

Table 11: SABR-B gain by ticker, full sample 2015–2025.

Ticker	N	Gain (%)	Calls (%)	Puts (%)
Equity indices				
SPX	19,844,947	+3.42	+14.39	−8.96
OEX	3,255,982	−1.36	−0.75	−2.03
DJX	1,269,846	+1.14	+2.01	+0.38
Individual equities				
AAPL	2,011,402	+6.74	+13.71	−1.89
MSFT	2,282,086	+2.63	+10.01	−7.50
IBM	1,110,900	−1.66	+2.51	−5.89
CAT	1,330,829	−3.08	−1.03	−5.31
HON	774,229	−8.48	−4.14	−13.26
JNJ	313,048	+2.63	+5.62	−0.72
RTX	1,008,878	+0.07	+1.93	−1.89
SHW	760,017	−3.03	+1.44	−9.12
Commodity ETFs				
SLV	1,407,312	+0.02	−0.24	+0.29
IAU	291,960	−0.35	−0.18	−0.49
USO	1,343,232	+1.98	+2.99	+1.12

The regime breakdown in Table 12 is informative. Across the equity indices and most individual equities, the SABR-B gain rises from Normal to Extreme conditions. This is consistent with the leverage explanation in Hull and White (2017). They note that the minimum-variance correction is largest precisely when the smile is steepest. On equity markets, that condition coincides with the high-volatility regime captured by the VIX (Whaley, 2009). The commodity ETFs show the same direction but a much smaller magnitude, again consistent with a weaker spot-vol coupling on those underlyings. Two names, HON and JNJ, deteriorate in the Extreme regime. This could reflect the small number of stress-day observations available for these tickers, together with the corresponding sensitivity of the SSE-based gain to a few large hedge errors.

Table 12: SABR-B gain by VIX regime and ticker.

Ticker	Normal (%)	Elevated (%)	Extreme (%)
Equity indices			
SPX	−2.24	+2.83	+5.64
OEX	−1.33	+0.39	−5.18
DJX	−0.31	−1.01	+11.73
Individual equities			
AAPL	+1.35	+4.77	+13.17
MSFT	−4.46	+4.06	+9.96
IBM	−5.22	−0.63	+2.83
CAT	−6.02	−3.64	+2.88
HON	−3.35	−5.68	−14.28
JNJ	−1.40	+9.81	−10.54
RTX	−0.76	−1.76	+3.46
SHW	−8.67	−9.90	+2.56
Commodity ETFs			
SLV	+0.01	−2.10	+4.08
IAU	−0.31	−0.44	−0.61
USO	+1.15	+1.95	+4.70

C Minimum-Variance Robustness and Supplementary Results

This appendix includes supplementary results for the minimum-variance (MV) hedge. The main text focuses on the central OOS12 results and their interpretation, while the appendix provides additional bucket decompositions and rolling-window data. All gains are measured as percentage reductions in squared hedging error relative to the Black-Scholes delta hedge.

C.1 Full-sample MV bucket results

Table 13: Minimum-variance hedging performance by signed delta bucket for SPX, OOS12 evaluation sample.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
0.1	749,858	5.91	2.58***	-0.9	818,484	-13.93	-0.79
0.2	611,399	9.87	1.59	-0.8	623,301	-7.95	-0.83
0.3	607,097	11.62	1.28	-0.7	610,446	-3.94	-0.45
0.4	656,813	12.91	1.16	-0.6	657,080	-1.06	0.06
0.5	747,230	14.68	1.16	-0.5	744,201	4.02	0.59
0.6	892,767	15.29	1.19	-0.4	886,300	6.78	0.94
0.7	1,134,854	10.26	1.18	-0.3	1,125,802	9.40	1.13
0.8	1,569,792	6.88	1.21	-0.2	1,558,044	17.34	1.17
0.9	2,503,475	6.60	1.21	-0.1	2,524,222	19.34	1.02
All calls	9,473,314	9.54	1.22	All puts	9,547,886	3.32	0.52

All options pooled: $N = 19{,}021{,}200$, gain $= 6.77\%$, $t = 1.11$.

C.2 MV bucket results by volatility regime

C.2.1 Normal VIX regime

Table 14: MV hedging gain by signed delta bucket under the Normal VIX regime, defined as VIX < 20, SPX OOS12 evaluation sample.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
0.1	428,280	22.68	5.23***	-0.9	493,240	3.57	3.14***
0.2	348,886	18.93	4.18***	-0.8	360,009	3.84	2.86***
0.3	347,935	13.88	3.45***	-0.7	351,510	4.77	2.81***
0.4	376,375	10.64	2.85***	-0.6	378,233	7.05	2.86***
0.5	429,318	10.13	2.44**	-0.5	429,061	8.20	3.05***
0.6	519,466	9.70	2.05**	-0.4	516,350	9.79	3.28***
0.7	675,849	8.74	1.76*	-0.3	671,792	11.07	3.36***
0.8	978,682	7.24	1.45	-0.2	972,788	12.09	3.17***
0.9	1,703,318	5.12	0.92	-0.1	1,726,926	10.50	2.21**
All calls	5,808,135	8.18	1.90*	All puts	5,899,912	7.73	3.40***

All options pooled: N = 11,708,047, gain = 7.99%, $t = 2.91$***.

C.2.2 Elevated VIX regime

Table 15: MV hedging gain by signed delta bucket under the Elevated VIX regime, defined as $20 \leq \text{VIX} < 30$, SPX OOS12 evaluation sample.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
0.1	238,552	27.15	4.00***	-0.9	240,626	1.55	0.82
0.2	198,853	23.08	3.00***	-0.8	199,660	3.33	1.14
0.3	197,714	17.49	2.32**	-0.7	197,672	2.51	1.70*
0.4	214,965	13.25	1.81*	-0.6	213,798	3.72	1.78*
0.5	244,573	13.03	1.57	-0.5	242,485	7.72	1.61
0.6	289,535	11.93	1.47	-0.4	286,905	3.68	1.46
0.7	360,712	4.80	1.29	-0.3	356,585	3.23	1.42
0.8	475,272	2.83	1.13	-0.2	469,880	12.32	1.29
0.9	669,210	2.82	1.23	-0.1	666,996	12.72	0.82
All calls	2,889,388	5.32	1.58	All puts	2,874,609	4.66	1.47

All options pooled: N = 5,763,997, gain = 5.04%, $t = 1.59$.

C.2.3 Extreme VIX regime

Table 16: MV hedging gain by signed delta bucket under the Extreme VIX regime, defined as VIX $\geq$ 30, SPX OOS12 evaluation sample.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
0.1	83,026	-3.53	0.95	-0.9	84,618	-28.63	-1.18
0.2	63,660	3.38	0.51	-0.8	63,632	-13.28	-1.23
0.3	61,448	9.11	0.55	-0.7	61,264	-8.69	-0.99
0.4	65,473	13.26	0.63	-0.6	65,049	-6.06	-0.63
0.5	73,339	16.16	0.73	-0.5	72,655	1.24	-0.09
0.6	83,766	17.52	0.82	-0.4	83,045	8.64	0.39
0.7	98,293	14.16	0.88	-0.3	97,425	18.29	0.67
0.8	115,838	10.54	0.96	-0.2	115,376	20.95	0.83
0.9	130,947	18.30	0.97	-0.1	130,300	26.56	0.85
All calls	775,791	13.29	0.84	All puts	773,365	1.47	-0.17

All options pooled: N = 1,549,156, gain = 7.73%, t = 0.57.

C.3 MV rolling-window

The main MV results use the OOS12 specification, where the coefficients are estimated using a rolling 12-month window and applied out of sample. To assess whether the coefficient dynamics are sensitive to the choice of estimation window, Figures 12- 15 report the corresponding OOS24 and OOS36 coefficient paths. The longer windows produce smoother coefficient series, but respond more slowly to changes in market conditions. These figures therefore support the interpretation in Section 4.6, claiming OOS12 adapts faster, while OOS24 and OOS36 serve as robustness checks.

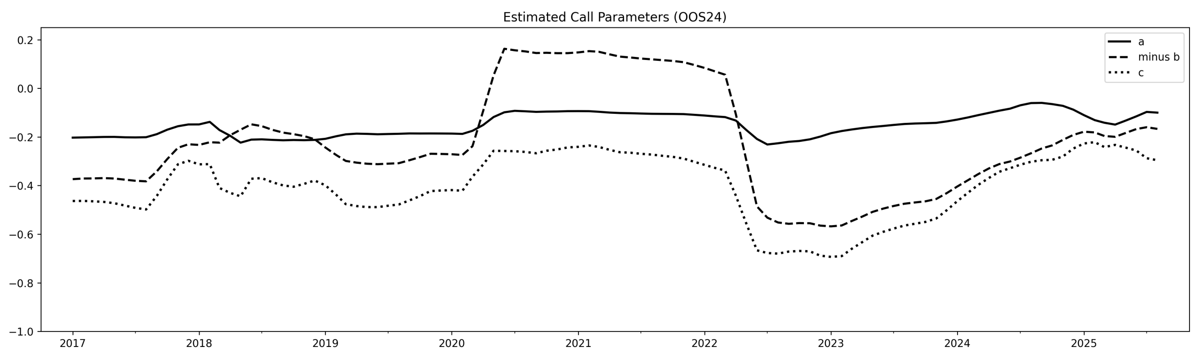

Figure 12: Rolling OOS24 estimates of the MV correction coefficients for SPX calls, 2015-2025.

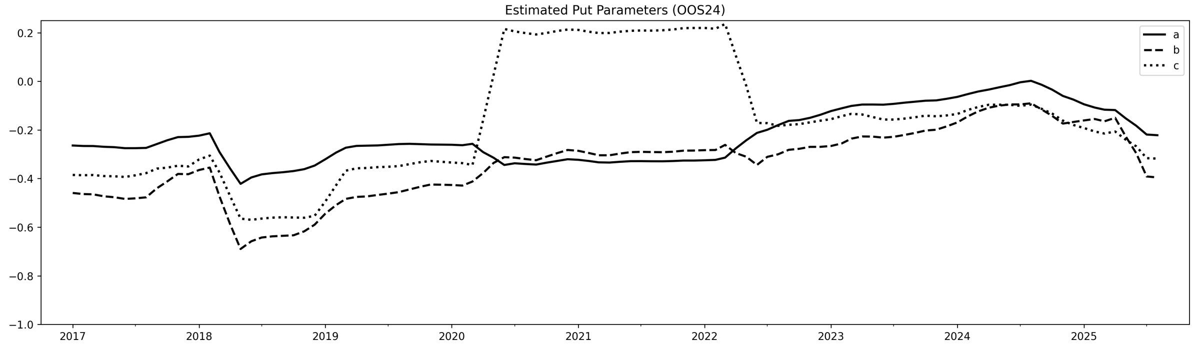

Figure 13: Rolling OOS24 estimates of the MV correction coefficients for SPX puts, 2015-2025.

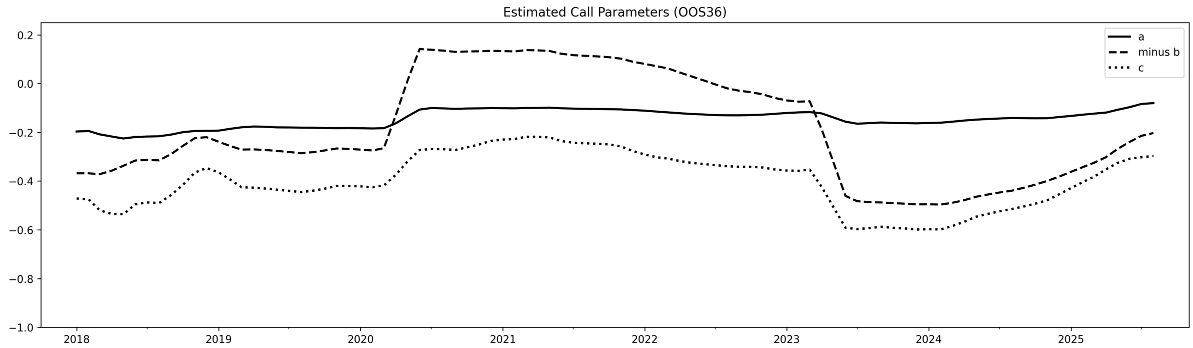

Figure 14: Rolling OOS36 estimates of the MV correction coefficients for SPX calls, 2015-2025.

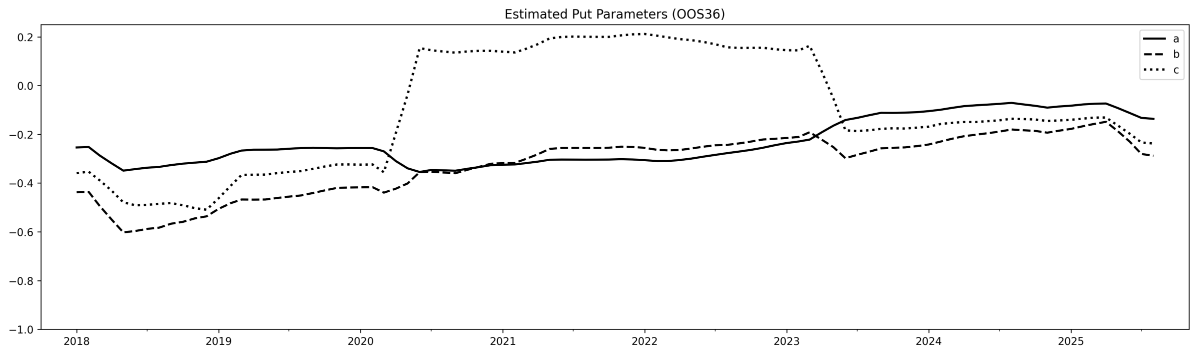

Figure 15: Rolling OOS36 estimates of the MV correction coefficients for SPX puts, 2015-2025.

C.4 Cross-index MV robustness

Table 17: Cross-index MV hedging gains for SPX, OEX and DJX, OOS12, 2018+ common sample. Gains are measured as percentage reductions in squared hedging error relative to the Black-Scholes delta hedge.

Option type	SPX	OEX	DJX
Calls	+9.5	−0.1	+5.5
Puts	+3.2	+1.5	+0.3

Table 17 reports a supplementary cross-index robustness check for the MV hedge. The table uses the OOS12 specification on a common sample across SPX, OEX, and DJX, from 2018-2025. Therefore, it is not directly comparable to the full 2015-2025 SPX results in the main text. The results show that the MV hedge remains positive for SPX and DJX calls, while OEX calls are close to zero. Put gains are positive across all three indices but generally smaller. Overall, the results suggests that the MV correction is not purely an SPX specific result.

D GARCH and DV Models: Supplementary Gains

This appendix collects the full bucket-level and cross-ticker results for the four model families that share the volatility channel: the GARCH plug-in delta (Section 2.8.1), the GJR-GARCH minimum-variance correction (Section 2.8.2), the model-free delta–vega hedge (Section 2.10), and the GARCH-scaled delta–vega overlay (Section 2.10.4). Section D.1 presents the full-sample gains by signed delta bucket; Section D.2 breaks the same statistics down by VIX regime; Section D.3 reports the cross-ticker robustness of the GARCH adn DV specifications.

D.1 Full-sample GARCH and DV bucket results

Tables 18 and 19 confirm that both GARCH delta specifications underperform Black–Scholes uniformly across the option surface. The plug-in losses are largest in the OTM call wing, where d_1 is most sensitive to the volatility input; the GJR-GARCH MV correction makes the situation worse rather than better, because the leverage slope $\hat{\beta}/S$ is applied uniformly across strikes and therefore over-corrects ATM and ITM options where the smile slope priced into the market is much shallower than the physical slope estimated under the rolling window. Tables 20 and 21 show the corresponding pattern for the two delta–vega hedges. The model-free DV hedge is positive and statistically significant in nearly every call bucket and in the OTM put wing, with a deep-ITM put deterioration consistent with the linearised P&L losing accuracy where vega and gamma diverge most. The GARCH-scaled overlay preserves the call-side gains but collapses on puts, in line with the variance-risk-premium interpretation of κ given in Section 2.10.4.

Table 18: GARCH plug-in delta gain by signed Black–Scholes delta bucket. Full sample, 2015–2025. t is the Newey–West HAC corrected Diebold–Mariano paired statistic. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	749,948	−1610.15	−1.38	−0.1	2,524,704	−91.05	-3.30^{***}
+0.2	611,520	−504.19	−1.33	−0.2	1,558,280	−51.41	-4.15^{***}
+0.3	607,194	−215.62	−1.27	−0.3	1,125,963	−20.67	-5.24^{***}
+0.4	656,917	−102.04	−1.20	−0.4	886,395	−17.40	-2.31^{**}
+0.5	747,344	−48.69	−1.19	−0.5	744,306	−33.59	−1.30
+0.6	892,910	−22.71	-1.88^{*}	−0.6	657,185	−63.67	−1.29
+0.7	1,134,933	−20.55	-5.80^{***}	−0.7	610,516	−99.06	−1.33
+0.8	1,569,897	−37.76	-3.30^{***}	−0.8	623,389	−163.65	−1.40
+0.9	2,503,935	−42.86	-2.10^{**}	−0.9	818,568	−274.04	−1.47

Table 19: GJR-GARCH MV delta gain by signed Black–Scholes delta bucket. Full sample, 2015–2025. t is the Newey–West HAC corrected Diebold–Mariano paired statistic. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	749,948	−412.20	−1.35	−0.1	2,524,704	−161.23	−1.27
+0.2	611,520	−351.78	−1.37	−0.2	1,558,280	−197.42	−1.32
+0.3	607,194	−323.03	−1.38	−0.3	1,125,963	−221.25	−1.35
+0.4	656,917	−295.41	−1.38	−0.4	886,395	−240.52	−1.37
+0.5	747,344	−272.13	−1.36	−0.5	744,306	−241.61	−1.38
+0.6	892,910	−234.94	−1.35	−0.6	657,185	−218.60	−1.40
+0.7	1,134,933	−188.62	−1.33	−0.7	610,516	−158.06	−1.40
+0.8	1,569,897	−134.35	−1.30	−0.8	623,389	−102.88	−1.41
+0.9	2,503,935	−64.17	−1.26	−0.9	818,568	−34.06	−1.47

Table 20: Model-free delta-vega (DV) gain by signed Black–Scholes delta bucket. Full sample, 2015–2025. t is the Newey–West HAC corrected Diebold–Mariano paired statistic. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	786,692	−8.69	+1.12	−0.1	2,643,665	+45.55	+1.52
+0.2	639,697	+77.31	$+3.00^{***}$	−0.2	1,620,438	+64.85	$+1.92^{*}$
+0.3	632,924	+90.39	$+2.77^{***}$	−0.3	1,169,665	+63.09	$+1.95^{*}$
+0.4	682,759	+96.06	$+2.69^{***}$	−0.4	920,549	+56.57	$+1.96^{**}$
+0.5	688,408	+98.21	$+2.60^{***}$	−0.5	772,966	+44.53	$+1.92^{*}$
+0.6	927,596	+95.23	$+2.54^{**}$	−0.6	682,979	+27.76	$+1.70^{*}$
+0.7	1,179,649	+87.71	$+2.53^{**}$	−0.7	636,462	+5.88	+0.61
+0.8	1,634,154	+73.69	$+2.48^{**}$	−0.8	651,948	−14.56	-1.91^{*}
+0.9	2,630,706	+43.66	$+2.40^{**}$	−0.9	855,723	−49.68	-2.16^{**}

Table 21: GARCH-scaled delta-vega (GARCH-DV) gain by signed Black–Scholes delta bucket. Full sample, 2015–2025. t is the Newey–West HAC corrected Diebold–Mariano paired statistic. Stars: $^{*}p < 10\,\%$, $^{**}p < 5\,\%$, $^{***}p < 1\,\%$.

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	749,948	−193.04	−1.18	−0.1	2,524,704	+15.60	−0.55
+0.2	611,520	−24.57	−0.13	−0.2	1,558,280	+40.25	+2.67***
+0.3	607,194	+6.35	+0.90	−0.3	1,125,963	+17.56	+2.22**
+0.4	656,901	+27.32	+3.87***	−0.4	886,395	−11.70	−0.12
+0.5	665,855	+44.87	+6.84***	−0.5	744,306	−41.28	−0.66
+0.6	892,896	+59.93	+3.79***	−0.6	657,185	−64.88	−0.93
+0.7	1,134,931	+68.12	+2.97***	−0.7	610,516	−70.86	−1.14
+0.8	1,569,897	+68.33	+2.33**	−0.8	623,389	−81.11	−1.29
+0.9	2,503,935	+39.37	+2.11**	−0.9	818,568	−130.49	−1.45

D.2 GARCH and DV bucket results by volatility regime

The regime tables below confirm that the call–put asymmetry of the DV hedge is present in every regime and that the GARCH delta failures are largest precisely where the smile is steepest. The DV hedge is positive and highly significant on calls in all three regimes; its put gains rise from negative under Normal conditions to clearly positive under Elevated and Extreme conditions, matching the regime ordering documented for SABR-B and consistent with the stronger spot–vol coupling that drives the smile in stress periods.

D.2.1 Normal VIX Regime

In the Normal regime every DV call bucket gains more than +13 % at 1 % significance, while the GARCH plug-in and GJR-GARCH MV corrections lose ground in every bucket. The GARCH-DV overlay closely tracks the DV hedge on calls and recovers most of the put-side gain, confirming that the call channel dominates under calm conditions.

Table 22: Per-bucket SSE gain over Black–Scholes for the model-free delta–vega (DV) hedge, Normal VIX regime (VIX $<$ 20). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided).

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	456 795	+13.89	+2.82***	−0.9	521 929	−23.72	−3.12***
+0.2	370 931	+75.36	+9.06***	−0.8	382 311	−20.42	−4.32***
+0.3	368 112	+82.74	+8.95***	−0.7	371 860	−18.10	−2.36**
+0.4	396 527	+88.69	+9.03***	−0.6	398 362	−9.27	−0.19
+0.5	390 487	+93.36	+8.43***	−0.5	451 455	+3.96	+1.44
+0.6	546 651	+90.53	+7.72***	−0.4	543 083	+16.15	+2.34**
+0.7	711 214	+79.89	+7.08***	−0.3	706 340	+25.50	+2.65***
+0.8	1 029 657	+60.77	+6.66***	−0.2	1 022 142	+34.54	+2.79***
+0.9	1 805 817	+36.14	+5.59***	−0.1	1 822 986	+27.54	+1.86*

Table 23: Per-bucket SSE gain over Black–Scholes for the GARCH plug-in, GJR-GARCH minimum-variance (MV), and GARCH-scaled delta–vega (GARCH-DV) hedges, Normal VIX regime (VIX $<$ 20). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided). N corresponds to the largest model sample within each cell.

		GARCH plug-in		GJR-GARCH MV		GARCH-DV	
Δ	N	Gain (%)	t	Gain (%)	t	Gain (%)	t
Panel A: Calls							
+0.1	428 370	−164.22	−3.97***	−19.09	−3.33***	+23.91	+3.98***
+0.2	349 004	−55.04	−3.83***	−17.63	−4.26***	+75.18	+10.17***
+0.3	348 034	−25.34	−3.93***	−17.12	−4.50***	+80.80	+10.66***
+0.4	376 479	−16.32	−3.07***	−16.04	−4.45***	+83.83	+10.17***
+0.5	429 436	−15.39	−3.80***	−16.30	−4.38***	+88.25	+8.58***
+0.6	519 625	−22.79	−5.52***	−15.12	−4.25***	+84.80	+7.02***
+0.7	676 014	−48.15	−4.95***	−11.49	−3.94***	+73.95	+5.86***
+0.8	978 901	−79.96	−5.67***	−7.64	−3.65***	+56.09	+5.08***
+0.9	1 703 844	−66.16	−6.55***	−4.21	−3.35***	+33.20	+3.94***
Panel B: Puts							
−0.9	493 332	−16.33	−2.98***	−0.14	+0.67	−19.40	−2.42**
−0.8	360 121	−14.40	−3.49***	−1.55	−0.40	−12.99	−3.04***
−0.7	351 607	−13.13	−3.48***	−3.90	−1.73*	−8.84	−1.02
−0.6	378 347	−21.59	−3.24***	−8.45	−2.49**	+1.50	+1.33
−0.5	429 173	−24.17	−3.36***	−11.76	−2.83***	+13.81	+3.14***
−0.4	516 481	−33.82	−6.26***	−13.86	−2.93***	+26.24	+3.57***
−0.3	671 980	−74.58	−5.59***	−14.49	−2.78***	+36.59	+3.15***
−0.2	973 033	−163.69	−5.79***	−13.33	−2.53**	+44.77	+2.74***
−0.1	1 727 408	−221.76	−6.63***	−11.56	−2.45**	+37.30	+1.70*

D.2.2 Elevated VIX Regime

Under Elevated conditions the DV call gains remain large and significant, while the GARCH-only specifications continue to lose value. The GARCH-DV overlay still adds value on calls but begins to diverge from DV on puts, foreshadowing the collapse seen in the Extreme regime.

Table 24: Per-bucket SSE gain over Black–Scholes for the model-free delta–vega (DV) hedge, Elevated VIX regime ($20 \leq \text{VIX} < 30$). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided).

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	245 802	+9.94	+2.22**	−0.9	248 208	−24.17	−4.50***
+0.2	204 421	+82.79	+7.66***	−0.8	205 350	−8.85	−1.25
+0.3	202 742	+91.56	+7.65***	−0.7	202 730	+0.54	+0.47
+0.4	220 113	+95.83	+7.60***	−0.6	218 927	+21.45	+2.22**
+0.5	228 924	+97.82	+7.38***	−0.5	248 140	+36.11	+3.03***
+0.6	296 349	+94.06	+7.12***	−0.4	293 617	+44.84	+3.36***
+0.7	369 136	+87.94	+7.08***	−0.3	364 851	+51.57	+3.56***
+0.8	487 149	+74.33	+7.02***	−0.2	481 407	+57.64	+3.71***
+0.9	691 806	+41.57	+5.47***	−0.1	688 386	+38.85	+2.03**

Table 25: Per-bucket SSE gain over Black–Scholes for the GARCH plug-in, GJR-GARCH minimum-variance (MV), and GARCH-scaled delta–vega (GARCH-DV) hedges, Elevated VIX regime ($20 \leq \text{VIX} < 30$). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided). N corresponds to the largest model sample within each cell.

		GARCH plug-in		GJR-GARCH MV		GARCH-DV	
Δ	N	Gain (%)	t	Gain (%)	t	Gain (%)	t
Panel A: Calls							
+0.1	238 552	−460.25	−3.71***	−53.11	−1.76*	−23.73	−0.49
+0.2	198 855	−151.47	−3.56***	−56.07	−1.96*	+62.15	+6.34***
+0.3	197 713	−48.57	−3.12***	−62.58	−1.96**	+73.87	+7.78***
+0.4	214 963	−12.93	−2.44**	−63.40	−1.94*	+80.02	+8.43***
+0.5	244 572	−3.21	−2.02**	−61.04	−1.92*	+84.00	+8.42***
+0.6	289 533	−12.81	−3.30***	−55.34	−1.91*	+82.38	+7.89***
+0.7	360 677	−46.28	−4.39***	−47.76	−1.87*	+78.88	+7.52***
+0.8	475 214	−89.63	−5.26***	−36.63	−1.82*	+67.42	+7.11***
+0.9	669 165	−79.69	−6.59***	−19.38	−1.80*	+34.68	+3.98***
Panel B: Puts							
−0.9	240 628	−62.58	−2.99***	−5.47	−1.61	−30.27	−4.36***
−0.8	199 660	−37.47	−2.98***	−13.58	−1.73*	−15.85	−2.64***
−0.7	197 661	−16.01	−2.23**	−26.04	−1.82*	−9.00	−1.09
−0.6	213 794	−6.21	−1.64	−43.87	−1.85*	+8.85	+0.75
−0.5	242 481	−0.45	−0.48	−51.12	−1.86*	+22.89	+1.91*
−0.4	286 884	−7.94	−2.10**	−55.16	−1.86*	+32.45	+2.60***
−0.3	356 559	−43.85	−4.23***	−55.94	−1.83*	+40.29	+3.02***
−0.2	469 873	−117.57	−5.17***	−56.61	−1.79*	+47.37	+3.26***
−0.1	666 996	−180.49	−6.03***	−53.59	−1.79*	+21.42	+0.70

D.2.3 Extreme VIX Regime

Under Extreme conditions the GARCH plug-in losses become very large in absolute terms (*e.g.* $-2214.88\,\%$ at $\Delta = +0.1$), reflecting the high sensitivity of d_1 to a misspecified volatility input when the smile is steep. The DV hedge still delivers double-digit positive gains on calls and on most of the put surface; the GARCH-DV overlay underperforms DV on both sides, consistent with the directional bet on the variance risk premium misfiring in stressed markets where the GARCH forecast and the implied volatility surface diverge sharply.

Table 26: Per-bucket SSE gain over Black–Scholes for the model-free delta–vega (DV) hedge, Extreme VIX regime (VIX $\geq$ 30). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided).

Calls				Puts			
Δ	N	Gain (%)	t	Δ	N	Gain (%)	t
+0.1	84 095	−18.78	−0.54	−0.9	85 586	−74.82	−1.41
+0.2	64 345	+75.74	+1.83*	−0.8	64 287	−15.39	−0.70
+0.3	62 070	+91.60	+1.74*	−0.7	61 872	+19.45	+1.27
+0.4	66 119	+97.73	+1.70*	−0.6	65 690	+41.25	+1.71*
+0.5	68 997	+99.26	+1.65*	−0.5	73 371	+57.68	+1.71*
+0.6	84 596	+96.72	+1.60	−0.4	83 849	+70.21	+1.66*
+0.7	99 299	+90.05	+1.57	−0.3	98 474	+76.86	+1.58
+0.8	117 348	+79.60	+1.52	−0.2	116 889	+76.64	+1.51
+0.9	133 083	+52.94	+1.41	−0.1	132 293	+57.23	+1.23

Table 27: Per-bucket SSE gain over Black–Scholes for the GARCH plug-in, GJR-GARCH minimum-variance (MV), and GARCH-scaled delta–vega (GARCH-DV) hedges, Extreme VIX regime (VIX $\geq$ 30). Newey–West HAC corrected paired Diebold–Mariano t-statistics with 30 lags. Stars: * 10%, ** 5%, *** 1% (two-sided). N corresponds to the largest model sample within each cell.

		GARCH plug-in		GJR-GARCH MV		GARCH-DV	
Δ	N	Gain (%)	t	Gain (%)	t	Gain (%)	t
Panel A: Calls							
+0.1	83 026	−2214.88	−1.37	−590.91	−1.44	−282.78	−1.36
+0.2	63 661	−718.36	−1.32	−522.89	−1.47	−75.11	−0.99
+0.3	61 447	−312.57	−1.27	−476.06	−1.47	−32.35	−0.73
+0.4	65 475	−149.99	−1.22	−431.93	−1.47	−2.24	+0.04
+0.5	73 336	−70.19	−1.18	−392.07	−1.45	+24.52	+3.69***
+0.6	83 752	−26.04	−1.23	−344.25	−1.44	+46.83	+1.93*
+0.7	98 242	−2.07	−0.65	−297.62	−1.42	+62.11	+1.63
+0.8	115 782	+8.53	+0.47	−243.64	−1.39	+74.52	+1.44
+0.9	130 926	+11.81	+0.63	−160.48	−1.36	+49.34	+1.20
Panel B: Puts							
−0.9	84 608	−494.43	−1.46	−63.46	−1.57	−230.21	−1.31
−0.8	63 608	−327.48	−1.37	−216.59	−1.52	−161.14	−1.17
−0.7	61 248	−184.10	−1.31	−300.54	−1.51	−133.38	−1.11
−0.6	65 044	−101.69	−1.25	−356.47	−1.50	−117.01	−1.04
−0.5	72 652	−49.31	−1.19	−371.66	−1.48	−80.05	−0.95
−0.4	83 030	−17.68	−1.22	−361.02	−1.46	−37.05	−0.78
−0.3	97 424	+0.98	−0.14	−335.46	−1.45	+3.98	−0.02
−0.2	115 374	+7.58	+0.31	−308.31	−1.42	+35.92	+1.67*
−0.1	130 300	+11.01	+0.40	−280.93	−1.38	+3.55	−1.29

D.3 Cross-ticker robustness of the GARCH and DV hedges

The four tables below extend the GARCH plug-in (Approach A), the GJR-GARCH minimum-variance hedge (Approach B), the model-free delta–vega hedge (Approach A) and the GARCH-scaled delta–vega overlay (Approach B) to the broader set of tickers introduced in Appendix B.

The GARCH-only hedges lose ground on every ticker. The plug-in delta is most damaging on the indices and on the individual equities with the steepest equity-style smiles (SPX, MSFT) and least damaging on the commodity ETFs, where there is little smile to discard in the first place (IAU, SLV). The GJR-GARCH MV correction reduces the loss on most individual equities and on the ETFs, since the leverage slope $\hat{\beta}/S$ at least partially captures the spot–vol correlation that the plug-in throws away, but it deteriorates on the indices, where the implied skew that Approach A's substitution destroys is steeper than the physical slope that Approach B can estimate from a 252-day window of returns. The combined picture is consistent with the main-text finding that volatility-only specifications cannot recover the information embedded in the option-implied smile.

The two delta–vega tables tell a different story. The model-free hedge is positive on every index and on most individual equities, with the call channel doing the bulk of the work and put gains turning negative for tickers whose smile is steepest in the OTM put wing (MSFT, IBM, JNJ, SHW). On the commodity ETFs the gain is small in absolute terms on both sides, mirroring the SABR-B result and consistent with a weaker spot–vol coupling on those underlyings. The GARCH-scaled overlay preserves most of the call-side gain but compresses or reverses the pooled gain across the equity surface, with the deterioration concentrated entirely on puts. This matches the pattern documented for SPX in the main text: scaling the vega leg by $\sigma_{\mathrm{GARCH}}/\sigma_{\mathrm{imp}}^{\mathrm{ATM}}$ is a directional bet on the variance risk premium, and that bet misfires most consistently on equity puts, where the implied volatility of the OTM wing is dominated by skew rather than by the level captured by the GARCH forecast.

The patterns are uniform across asset classes. The plug-in delta is negative on every ticker, with losses largest on the indices and on individual equities with the steepest equity-style smiles (SPX, MSFT) and smallest on the commodity ETFs where there is little smile to discard in the first place (IAU, SLV). Approach B reduces the loss on most individual equities and on the ETFs, since the leverage slope $\hat{\beta}/S$ at least partially captures the spot–vol correlation that the plug-in throws away, but it deteriorates on the indices, where the implied skew that Approach A's substitution destroys is steeper than the physical slope that Approach B can estimate from a 252-day window of returns. The combined picture is consistent with the main-text finding that volatility-only specifications cannot recover the information embedded in the option-implied smile.

Table 28: GARCH plug-in (Approach A) gain by ticker, full sample 2015–2025.

Ticker	N	Gain (%)	Calls (%)	Puts (%)
Equity indices				
SPX	19,023,904	−89.55	−96.91	−81.27
OEX	3,103,917	−13.76	−10.60	−17.23
DJX	1,168,206	−64.25	−62.87	−65.46
Individual equities				
MSFT	2,188,215	−84.40	−91.03	−75.22
AAPL	1,873,122	−87.93	−94.78	−79.44
IBM	1,036,578	−29.03	−38.26	−19.84
CAT	1,226,356	−35.33	−40.39	−29.72
HON	742,310	−38.94	−39.88	−37.92
JNJ	236,820	−8.55	−10.04	−6.99
RTX	934,033	−33.03	−36.91	−28.97
SHW	734,381	−43.05	−40.34	−46.74
Commodity ETFs				
SLV	1,322,932	−8.14	−8.91	−7.34
IAU	304,402	−0.24	−0.35	−0.15
USO	1,235,051	−31.34	−36.84	−26.54

Table 29: GJR-GARCH minimum-variance (Approach B) gain by ticker, full sample 2015–2025.

Ticker	N	Gain (%)	Calls (%)	Puts (%)
Equity indices				
SPX	19,023,904	−151.95	−160.71	−142.10
OEX	3,103,917	−23.48	−21.33	−25.84
DJX	1,168,206	−95.99	−100.52	−92.05
Individual equities				
MSFT	2,188,215	−85.64	−88.84	−81.20
AAPL	1,873,122	−119.32	−126.20	−110.81
IBM	1,036,578	−13.76	−19.19	−8.36
CAT	1,226,356	−12.36	−15.12	−9.30
HON	742,310	−55.11	−58.15	−51.82
JNJ	236,820	−5.88	−5.59	−6.19
RTX	934,033	−9.62	−10.79	−8.40
SHW	734,381	−8.35	−7.42	−9.62
Commodity ETFs				
SLV	1,322,932	−2.14	−1.85	−2.45
IAU	304,402	−0.33	−0.45	−0.24
USO	1,235,051	−12.50	−14.16	−11.06

Table 30: Delta–Vega gain by ticker, Approach A (model-free plug-in), full sample 2015–2025.

Ticker	N	Gain (%)	Calls (%)	Puts (%)
Equity indices				
SPX	19,756,689	+56.18	+78.05	+31.65
OEX	3,218,514	+11.20	+19.42	+2.10
DJX	1,246,899	−9.95	+6.75	−24.19
Individual equities				
MSFT	2,242,631	+21.86	+48.12	−12.78
AAPL	1,970,814	+43.84	+65.86	+17.75
CAT	1,299,785	+22.02	+41.71	+1.72
IBM	1,079,794	+4.67	+27.33	−17.08
RTX	978,968	−3.23	+9.06	−15.54
HON	749,710	+10.07	+29.74	−10.44
SHW	744,003	+21.70	+50.88	−16.57
JNJ	302,140	+4.32	+25.55	−18.23
Commodity ETFs				
SLV	1,368,770	−4.04	−4.60	−3.49
USO	1,314,825	+5.40	+16.27	−3.36
IAU	274,381	−10.77	−11.32	−10.33

Table 31: Delta–Vega gain by ticker, Approach B (GARCH-implied vega), full sample 2015–2025.

Ticker	N	Gain (%)	Calls (%)	Puts (%)
Equity indices				
SPX	18,942,092	+8.32	+43.66	−31.22
OEX	3,069,432	−0.66	+8.78	−11.10
DJX	1,147,282	−15.13	+0.51	−28.35
Individual equities				
MSFT	2,151,532	−17.43	+25.30	−74.21
AAPL	1,835,345	+1.50	+36.82	−40.11
CAT	1,197,703	+3.62	+30.34	−24.56
IBM	1,007,657	−11.26	+15.89	−36.79
RTX	905,544	−27.58	−12.83	−42.19
HON	718,425	−9.39	+14.41	−33.73
SHW	719,163	−10.67	+34.92	−70.68
JNJ	227,703	+10.60	+26.64	−5.02
Commodity ETFs				
SLV	1,285,160	−2.72	−3.40	−2.04
USO	1,208,929	−1.83	+6.52	−8.80
IAU	270,027	−6.91	−7.26	−6.63

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

CHALMERS