Machine Learning Application to Delta-Gamma Hedging

Machine Learning in Finance

Written at EPFL June 7, 2024

Group Members:
Chady Bensaid
Dorah Borgi
Simon Jonängen
William Olsson
Sara Tegström

Supervised by:
Elise Marie Gourier

Teaching Assistant:
Giuseppe Matera

Abstract

Machine learning (ML) is rapidly transforming the financial industry by enabling sophisticated data analysis, predictive modelling and automation of complex tasks for investors. This projects aims to achieve a predictive model for creating an option portfolio that is Delta-Gamma hedged towards risk. By using the equation for Black-Scholes option pricing model for European options, different measurements of options such as Delta (Δ) and Gamma (Γ), commonly referred to as the Greeks, have been derived. Using these expressions for the Greeks, regressive machine learning algorithms have then used in order to continuously predict future values of Gamma. The main goal was to achieve a Delta-Gamma neutral portfolio. Four different algorithms were used, namely, linear regression, random forest, support vector regression and XGBoost. The algorithms were evaluated under the two performance metrics of R^2 and mean squared error (MSE).

Contents

1 Options as a Financial Derivative

1.1 Background information

Options are financial derivatives that grant the buyer the right, but not the obligation, to buy or sell an underlying asset at a specified price, known as the strike price, on or before a specified date. These instruments are utilized for hedging risk, speculating on future price movements, and generating income, across various underlying assets such as stocks, indices, commodities, and currencies [6] .

Options are versatile tools in the financial markets, allowing for tailored investment strategies. They are contracts with specified key details such as the underlying asset, strike price, expiration date, and the premium. The value of an option is derived from the intrinsic value and the time value, which decreases as the option nears expiration. The use of options spans across:

- **Hedging**: To protect against adverse price movements.
- **Income Generation**: Through premium collection from selling options.
- **Speculation**: Betting on the direction of the market with leveraged potential profits or losses.

Options require a comprehensive understanding of market mechanics and the associated risks, including the potential loss of the entire premium paid.

1.1.1 Call Options

A call option gives the holder the right to buy the underlying asset at the strike price within a specified time period. Investors buy call options if they believe the price of the underlying asset will rise above the strike price before the expiration date. If this happens, the investor can purchase the asset at the lower strike price and potentially sell it at a higher market price, thus realizing a profit [6].

1.1.2 Put Options

Put options give the holder the right to sell the underlying asset at the strike price within a specified time period. Investors buy put options if they believe the price of the underlying asset will fall below the strike price before the expiration date. This allows the investor to sell the asset at the higher strike price, even if the market price has fallen, thus providing a mechanism for hedging against potential losses or for speculative gains [6].

1.2 The Greeks

1.2.1 Delta (Δ)

In the context of call options, "delta" (Δ) is a measure of how much the price of an option is expected to move based on a change of one unit of price of the underlying asset, such as one dollar. It is one of the "Greeks," which are metrics used to assess the risk and potential reward of options positions. Delta is particularly important because it gives investors and traders an idea of how the value of an option might change as the market price of the underlying asset moves, thereby helping in decision-making processes [7].

For call options, delta can range in the interval $[0, 1]$. A delta of 0 means the option's price is not expected to move in response to price changes in the underlying asset. A delta of 1 means the option's price is expected to move one-for-one with the price of the underlying asset, thus being equivalent to holding a share of a stock.

Delta can also be interpreted as the option's sensitivity to price changes in the underlying asset or as an estimate of the probability that the option will expire in-the-money (ITM). A higher delta not only indicates a greater sensitivity to changes in the underlying asset's price but also suggests a higher probability of the option expiring ITM. For instance, a delta of 0.75 suggests a 75% theoretical probability of expiring ITM, making it a more attractive choice for those bullish on the underlying asset.

Options that are at-the-money (ATM) generally have a delta around 0.5, reflecting that they have an approximately equal chance of ending up in or out of the money. As the underlying asset's price moves in the money for a call option, its delta will increase, approaching 1, indicating it is moving deeper ITM and its price is more closely tracking the underlying asset.

Delta is not static; it changes with the underlying asset's price, time to expiration, and volatility. As the expiration date approaches, the delta of in-the-money options increases towards 1 for calls, reflecting the increasing likelihood that the option will remain in the money. Conversely, the delta of out-of-the-money options decreases towards 0 as expiration approaches, reflecting the decreasing likelihood of the option expiring in the money.

Delta is also used in hedging strategies, such as delta-neutral trading, where the goal is to offset potential losses in one position with gains in another by maintaining a delta of zero. This involves adjusting the positions in the underlying asset and options to neutralize the overall delta of the portfolio.

1.2.2 Gamma (Γ)

Gamma is another key metric from the family of "Greeks" in options trading, and it measures the rate of change of an option's delta in response to a one-dollar change in the price of the underlying asset. While delta gives an estimate of how the price of an option changes with movements in the underlying asset, gamma indicates how quickly the delta itself changes. This measure is crucial for understanding the sensitivity of an option's price to market movements, especially for traders managing complex portfolios or employing dynamic hedging strategies [8].

Gamma is particularly significant for options that are near or at the money (ATM), where the delta is most sensitive to price changes in the underlying asset. A high gamma value indicates that the delta of the option is highly responsive to changes in the price of the underlying asset. In practical terms, this means that as the stock price moves, the rate at which the option's price is expected to change (its delta) will also change rapidly. This can lead to larger than expected changes in the option's price, presenting both opportunities and risks for traders.

Gamma is higher for options that are closer to their expiration date. As the time to expiration decreases, the sensitivity of delta to changes in the underlying asset's price increases, making accurate prediction of the option's price movement more critical—and challenging.

Gamma reaches its highest value for ATM options because their delta is most responsive to price changes in the underlying asset. For options deep in or out of the money, gamma tends to be lower, indicating that changes in the underlying price have a less dramatic effect on the option's delta and, consequently, its price.

For portfolio managers and traders, gamma is crucial for managing the delta of a portfolio. A portfolio with a high gamma requires more frequent adjustments to maintain a delta-neutral position, as small changes in the underlying asset's price can cause significant changes in delta. This makes gamma a critical factor in dynamic hedging strategies, where the goal is to neutralize not just the directional risk (delta) but also the risk related to the rate of change of this risk (gamma).

A high gamma can be both a risk and an opportunity. For traders holding options with a high gamma, small price movements in the underlying asset can lead to significant profits but also substantial losses. Therefore, understanding gamma allows traders to better navigate the markets by anticipating changes in the price behavior of their options.

1.2.3 Theta (Θ)

Theta is another important benchmark belonging to the option Greeks. Theta measures the sensitivity of an option's price to the passage of time, indicating how the price of the option decreases as it approaches its expiration date. Generally speaking, the value of theta is more negative for options with a shorter period of time until expiration. As an option approaches full maturity, its theta becomes more pronounced due to the diminishing time value of the option [9].

Furthermore, the value of theta also depends on the strike price of an option. Theta is typically highest when the strike price is close to the actual price of the underlying asset, in other words, within the region of ATM (at-the-money). It then decreases in both directions as the price of the underlying asset moves either ITM (in-the-money) or OTM (out-of-the-money). High theta values for ATM options reflect the accelerated time decay as the expiration date approaches, causing a rapid decrease in the option's price.

It is important to keep in mind that theta, while a useful measure, represents the theoretical rate of time decay and assumes that all other factors remain constant. In reality, the actual time decay can deviate from the projected values due to market fluctuations and changes in volatility. Therefore, continuous monitoring and recalculation of theta are essential to maintain accurate and relevant valuations.

2 Theory

2.1 Black-Scholes Option Pricing Model

The Black-Scholes formula for the price of a European call option is given by [10] :

$$C(S,t) = S_t\Phi(d_1) - Ke^{-rt}\Phi(d_2)$$

where:

– $C(S,t)$ is the price of the call option as a function of the stock price S and time t.

- S_t is the current stock price.
- $\Phi(d)$ is the cumulative distribution function of the standard normal distribution.
- K is the strike price of the option.
- e^{-rt} is the discount factor, where r is the risk-free interest rate and t is the time to expiration.
- d_1 and d_2 are calculated as follows:

$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}}$$

$$d_2 = d_1 - \sigma\sqrt{\tau},$$

- σ is the volatility of the stock's returns.
- $\tau = T - t$.

2.2 Derivation of Black-Scholes Price

Given a European call option with payoff $\max(S_T - K, 0)$, where S_T is the stock price at time T and K is the strike price. Let $C(S_t, t)$ be the price of the call option at time t with stock price S_t.

The stock price dynamics under the Black-Scholes model are given by the stochastic differential equation (SDE) [11] :

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where μ is the drift rate, σ is the volatility, and W_t is a Wiener process.

Using Itō's lemma to find the differential of $C(S_t, t)$, we have:

$$dC = \left(\frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2}\right) dt + \sigma S_t \frac{\partial C}{\partial S_t} dW_t$$

To eliminate risk, construct a portfolio Π by shorting $\Delta = \frac{\partial C}{\partial S_t}$ shares of stock and holding one option. The portfolio value is:

$$\Pi = C - \Delta S_t$$

The differential of Π is:

$$d\Pi = dC - \Delta dS_t$$

Substituting dC and dS_t, and choosing $\Delta = \frac{\partial C}{\partial S_t}$ to eliminate the risk (terms involving dW_t):

$$d\Pi = \left(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2}\right) dt$$

Since Π is risk-free, it must grow at the risk-free rate r:

$$d\Pi = r\Pi dt = r(C - \Delta S_t)dt$$

Equating the expressions for $d\Pi$:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} + rS_t \frac{\partial C}{\partial S_t} - rC = 0$$

This is the Black-Scholes PDE. For a European call option, solving this PDE with the final condition $C(S_T, T) = \max(S_T - K, 0)$ yields the Black-Scholes formula:

$$C(S_t, t) = S_t \Phi(d_1) - Ke^{-r\tau}\Phi(d_2)$$

where

$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}$$

Φ denotes the cumulative distribution function of the standard normal distribution.

In deriving the Black-Scholes formula several assumptions [2][3] have been made, these being the following:

a. The risk-free interest rate is constant.

b. The price of the asset follows a random walk, in accordance with the findings of Bachelier [4].

c. No dividends are paid out during the holding period of the option.

d. The option is European, with the only possibility of utilizing the option at its full maturity.

e. Transaction costs do not exist.

f. Variance is constant over the life of the option.

g. Stock prices are log-distributed.

h. There are no arbitrage opportunities.

Yielding from these assumptions is the fact that the future value of the option is solely based on known values and constants, such as the risk-free rate, the time for holding the option and the option price at the date of acquirement.

2.3 The Greeks for the Black-Scholes

2.3.1 Delta (Δ)

Delta for a European call option can be obtained by differentiating the Black-Scholes formula with respect to the stock price S_t [12]. Analytically, it is given by:

$$\Delta = \frac{\partial C}{\partial S_t} = \Phi(d_1)$$

2.3.2 Gamma (Γ)

Gamma measures, as mentioned earlier, the rate of change of delta with respect to changes in the underlying asset price [12]. It can be derived by differentiating delta with respect to S_t:

$$\Gamma = \frac{\partial^2 C}{\partial S_t^2} = \frac{\partial \Delta}{\partial S_t} = \frac{\varphi(d_1)}{S_t \sigma \sqrt{\tau}}$$

where $\phi(d_1)$ is the probability density function of the standard normal distribution evaluated at d_1.

Note: $\Phi(d)$ represents the cumulative distribution function of the standard normal distribution, and $\varphi(d)$ represents the probability density function of the standard normal distribution.

2.3.3 Derivative of Gamma ($\frac{\partial \Gamma}{\partial S_t}$)

In the context of greeks in the Black-Scholes model, the derivative of Γ not widely used. In this study, further information than what Δ and Γ gives will be obtained by calculating $\frac{\partial \Gamma}{\partial S_t}$. Even though this is not used to hedge the portfolio in particular, this serves one important predictor once Γ is being predicted.

$$\frac{\partial \Gamma}{\partial S_t} = \frac{\partial}{\partial S_t} \frac{\varphi(d_1)}{S_t \sigma \sqrt{\tau}} = \frac{1}{\sigma \sqrt{\tau}} \frac{\frac{\partial \varphi(d_1)}{\partial S_t} S_t - \varphi(d_1)}{(S_t \sigma \sqrt{\tau})^2} = \frac{\frac{\partial \varphi(d_1)}{\partial S_t} S_t - \varphi(d_1)}{S_t^2 \sigma^3 \tau^{3/2}} =$$

$$= \left\{ \frac{\partial \varphi}{\partial S_t}(d_1) = \frac{\partial \varphi}{\partial S_t}(d_1) \frac{1}{\sigma \sqrt{\tau}} \frac{K}{S_t} \right\} = \frac{\frac{\partial \varphi}{\partial S_t}(d_1) \frac{1}{\sigma \sqrt{\tau}} \frac{K}{S_t} S_t - \varphi(d_1)}{S_t^2 \sigma^3 \tau^{3/2}} = \frac{\frac{K}{\sigma \sqrt{\tau}} \frac{\partial \varphi}{\partial S_t}(d_1) - \varphi(d_1)}{S_t^2 \sigma^3 \tau^{3/2}}.$$

Here, d_1 is as defined before and $\frac{\partial \varphi}{\partial S_t}$ is the derivative of the function of a standard Gaussian density.

2.3.4 Theta (Θ)

Given the Black-Scholes formula for a European call option [12]:

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

where:

$$d_1 = \frac{\log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}},$$
$$d_2 = d_1 - \sigma\sqrt{T-t},$$

and Φ is the cumulative distribution function of the standard normal distribution, σ is the volatility of the underlying asset.

To find theta, we need the partial derivatives of d_1 and d_2 with respect to t:

$$\frac{\partial d_1}{\partial t} = \frac{\partial}{\partial t} \left(\frac{\log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \right)$$

$$= \frac{-(r + \frac{\sigma^2}{2})}{\sigma\sqrt{T-t}} + \frac{\left(\log\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})(T-t)\right) \cdot \frac{1}{2}(T-t)^{-\frac{3}{2}}\sigma}{\sigma(T-t)}$$

$$= \frac{-(r + \frac{\sigma^2}{2})}{\sigma\sqrt{T-t}} - \frac{\log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{2\sigma(T-t)^{\frac{3}{2}}}$$

$$= \frac{-(r + \frac{\sigma^2}{2})\sqrt{T-t} - \left(\log\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})(T-t)\right)}{2\sigma(T-t)}$$

Similarly, the partial derivative of d_2 with respect to t is:

$$\frac{\partial d_2}{\partial t} = \frac{\partial d_1}{\partial t} - \frac{\partial}{\partial t}(\sigma\sqrt{T-t})$$

$$= \frac{-(r + \frac{\sigma^2}{2})\sqrt{T-t} - \left(\log\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})(T-t)\right)}{2\sigma(T-t)} - \frac{\sigma \cdot \frac{1}{2}(T-t)^{-\frac{1}{2}}(-1)}{\sqrt{T-t}}$$

$$= \frac{-(r + \frac{\sigma^2}{2})\sqrt{T-t} - \left(\log\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})(T-t)\right)}{2\sigma(T-t)} + \frac{\sigma}{2\sqrt{T-t}}$$

The partial derivative of the Black-Scholes formula with respect to t is:

$$\frac{\partial C}{\partial t} = -rKe^{-r(T-t)}\Phi(d_2) + S\frac{\partial \Phi(d_1)}{\partial d_1}\frac{\partial d_1}{\partial t} - Ke^{-r(T-t)}\frac{\partial \Phi(d_2)}{\partial d_2}\frac{\partial d_2}{\partial t}$$

Given $\Phi'(d) = \phi(d)$, the density function of the normal distribution, we have:

$$\frac{\partial C}{\partial t} = -rKe^{-r(T-t)}\Phi(d_2) + S\phi(d_1)\frac{\partial d_1}{\partial t} - Ke^{-r(T-t)}\phi(d_2)\frac{\partial d_2}{\partial t}$$

$$\Theta = -\frac{\partial C}{\partial t}$$

Thus, theta quantifies the sensitivity of the option's price to the passage of time, indicating the rate at which the price of the option decreases as it approaches its expiration date. This expression, Θ, can be broken down into three components: the loss due to the approaching expiration ($-rKe^{-r(T-t)}\Phi(d_2)$), the time decay effect on d_1 and d_2 ($S\phi(d_1)\frac{\partial d_1}{\partial t}$ and $-Ke^{-r(T-t)}\phi(d_2)\frac{\partial d_2}{\partial t}$).

3 Hedging Strategy & Algorithmic Application

The essence of delta-gamma hedging involves adjusting a portfolio in such a way that it is neutral to both the direction of the market movements (delta-neutral) and the curvature of how option prices change with those movements (gamma-neutral). This dual neutrality helps in managing the risks associated with small and incremental price changes in the underlying asset. Delta-gamma hedging exemplifies the sophistication possible in options trading and risk management, offering a nuanced approach to stabilizing a portfolio's value against minor fluctuations in market prices.

3.1 Delta-Neutrality

The first step in delta-gamma hedging is to neutralize the delta of the portfolio. This is done by adjusting the holdings of the underlying asset or other derivative instruments so that the overall delta of the portfolio is as constant as possible. A delta-neutral portfolio is not affected by small movements in the price of the underlying asset because gains or losses on the options positions are offset by changes in the value of the underlying holdings.

3.2 Gamma-Neutrality

The second step focuses on neutralizing gamma, ensuring that the delta of the portfolio remains stable even if the underlying price continues to move. This is crucial because, in a delta-neutral portfolio, the delta can change with movements in the underlying asset's price, necessitating continuous rebalancing. By also achieving gamma neutrality, the portfolio is insulated against the need for frequent adjustments since the delta will not change significantly with small movements in the underlying price.

3.3 Delta-gamma hedging

Suppose we are at time t and are locked in with a call option on an underlying asset of price S_t. We are unable to sell the option right away and would like to neutralize movements in the option's price at future time steps with a portfolio $\pi_t = (x_t, 1, y_t)$ where x_t, y_t are respectively the number of shares held in the asset and y_t the quantity held in another option at time t. Why is another option needed? Since the gamma of a portfolio is the linear combination of its gammas and that stocks have constant $\Gamma = 1$, we cannot achieve 0 gamma only by shorting a stock so we would need to hold a certain quantity of another call option for the equation $\Gamma_t + y_t\Gamma'_t = 0$ to hold, where Γ, Γ' are the gamma's of respectively our initial option and the one sold to hedge.

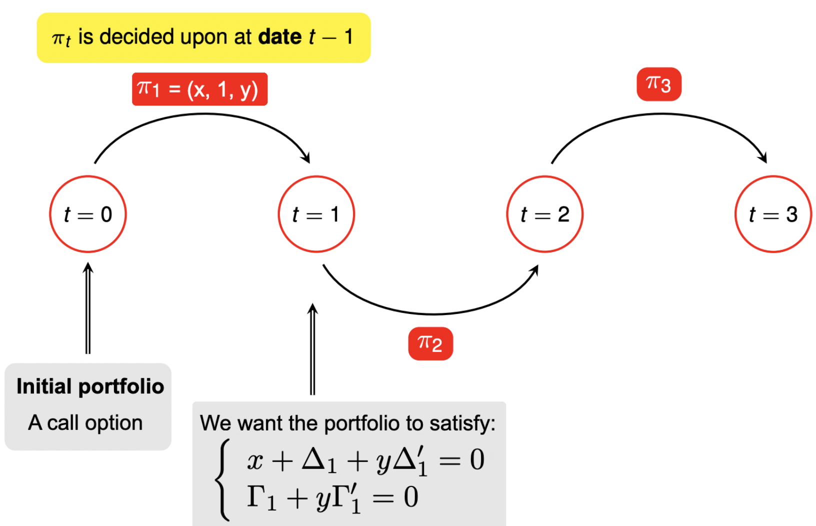


Figure 1: Diagram explaining the time steps in the delta-gamma hedging strategy.

3.4 Machine Learning Algorithms

In this study the purpose is to forecast Γ, which by definition will tell us about Δ. Some of the algorithms used are non-linear regression (with some type of penalization method), random forest, gradient boosting algorithms. This will be complemented by some comparisons of the ARMA-GARCH time series model.

3.4.1 Linear Regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. The simplest form of linear regression is one with two variables that fits a linear equation of a slope which intercept the data points. In multiple linear regression, several independent variables are used to predict the dependent variable.

The linear equation for simple linear regression is represented as:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Where,

- y is the dependent variable,
- x is the independent variable,
- β_0 is the y-intercept,
- β_1 is the slope of the line,
- ϵ is the error term.

In multiple linear regression, the equation expands to accommodate multiple independent variables $(x_1, x_2, \ldots, x_n)$:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon.$$

The coefficients (β) are estimated during the training process using the least squares method, aiming to minimize the sum of the squared differences between the observed values and the values predicted by the linear equation. The algorithms simplicity and comprehensibility are two of its strength while at the same its simple nature being its shortcoming due an under-performance for complex data sets with patterns that do not follow linearity.

3.4.2 Random Forest

The random forest machine learning algorithm is an ensemble method with an extension of bagging. By running multiple regression trees and then averaging over them a *forest* of regression trees are created, hence the name of the algorithm. In order to capture different perspectives the random forest also use random feature selection when creating the trees. Bagging, or *Bootstrap Aggregating*, trains an algorithm by creating different subsets of the same original data set in order to achieve a better performing model. Assume there are B number of data sets that are independent, yielding B number of trees. Then the aggregated model can be written as,

$$f_{agg}(\mathbf{x}) = \sum_{b=1}^{B} f^b(\mathbf{x})$$

The primary goal of baggging is to decrease a model's variance while maintaining low bias. A potential downfall with bagging when using regression trees is the risk of correlated subsets which random forest address by inducing randomness. Learning a random forest model occurs in parallel and large number of B datasets does not lead to overfitting.

3.4.3 Gradient Boosting Machines

Gradient Boosting Machine is an iterative machine learning method with a main idea of aggregating simple models, referred to as *weak learners*, together. A weak learner could be of such nature that it barely outperform random chance such as a decision stump, a shallow decision tree with only one decision node. By sequentially aggregating the weak learners together by assigning adaptive weights, it is possible to achieve a more robust and better performing model. At each iteration the algorithm assess and fit the new model by reweighing focusing on training data points with high past losses, that is points which the model previously performed weakly. The general boosting algorithm idea when aggregating B number of simple models looks as follows:

$$f_{boost}(\mathbf{x}) = \text{sign} \sum_{b=1}^{B} \alpha^b f^b(\mathbf{x})$$

where $\alpha^b =$ are the different adaptive weights for each simple model $f^b(\mathbf{x})$
. However, boosting machines may be prune to overfitting when B, the number of models, increases. To address this issue of choosing an appropriate number of B models, there exist multiple approaches. One is to stop when an additional model does not increase the overall model's performance. Another is to create a performance threshold, at which the algorithm stops when it is reached. B can thus be tuned as a hyperparameter. One type of of gradient boosting machines is a method called *Xtreme Gradient Boosting*, or XGBoost in short. The aim of a boosting algorithm is to reduce bias.

3.4.4 Support Vector Regression

A Support Vector Regression (SVR) is a variant of a Support Vector Machine (SVM). The aim of a SVR is similarly to a SVM, to find and create the best fitting hyperplane which maximises the margin between itself and the data points. However, instead of classification, the data is regressed and the points that fall within the so called tube of margin are considered to be accurately predicted. Data points outside of this margin are added to the models error as seen in algorithms loss function below

$$\mathcal{L}(y, \hat{y}) = \begin{cases} 0 & \text{if } |y - \hat{y}| < \epsilon \\ |y - \hat{y}| - \epsilon & \text{otherwise} \end{cases}$$

Being a regression model it yields a continuous value as an output. The *kernel trick* can be applied for SVR just as for SVM to transform the input feature space in order to model and capture non-linear relationship in the data. The SVR tube does therefore not necessarily have to be linear.

SVRs require hyperparameter tuning one of which is known as the regularization factor c. A larger parameter c equals to a larger margin which implies that the model is less strict towards errors, whilst a smaller one decreases the margin making the model more sensitive to errors. The data points that are on the boundary or outside the ϵ-insensitive tube are the *support vectors* and are important in the constituting the model. Relatively speaking, SVR is more robust towards outliers in comparison to a normal regression, as the primary focus is on data points close to the margin.

4 Data Extraction, (potential) cleaning and exploration

4.1 Data selection

For numerical preidictors, historical values of our response variable gamma as well as volatility, stock price, delta, vega, derivative of gamma, volume, volume change (derivative of volume), time to maturity, inflation rate and risk free interest rate were retrieved daily at the market's closing state. Inflation rate as well as risk free interest rate is updated monthly by the

For categorical values, predictors that identified bull or bear market and weather an option was in, out or at the money were used.

Table 1: Predictor Descriptions and Formulas

Predictors	**Type**	**Formula**	**Note**
Stock Price	Numerical	Retrieved (Yahoo Finance)	
Delta	Numerical	$\frac{\partial C}{\partial S_t} = \Phi(d_1)$	∂C = Change in Call Option Price, ∂S_t = Change in Stock Price
Call Option Price	Numerical	$C(S,t) = S_t\Phi(d_1) - Ke^{-rt}\Phi(d_2)$	
Gamma	Numerical	$\Gamma = \frac{\partial^2 C}{\partial S_t^2}$	Variables explained in section 2.3.2
Theta	Numerical	$\theta = \frac{\partial P}{\partial t}$	
Volume	Numerical	Retrieved (Yahoo Finance)	
Volume Change	Numerical	$\Delta\text{Volume} = \text{Volume}_n - \text{Volume}_{n-1}$	
Bull or Bear	Categorical	Bear < -20% Index decline, Bull > 20% Index increase	Over the time horizons we chose, this variable was most of the time a column full of 1 so we did not use it to avoid multicollinearity with the constant column.
In Out At Money	Categorical	In the Money: $S-C > 0$, Out of the Money: $S-C < 0$, At the Money: $S-C = 0$	S: Stock Price, C: Call Price. Same remark as the one above.
Time to Maturity	Numerical	Current date − Maturity date	Calculated
Lagged variables	Numerical		Lagged variables of the above variables

Below follows a section that explains the reasoning why the chosen predictors where were explored in the machine learning models, as well as information on their time interval.

Predictor	Relevance	Time Interval
Volatility	Relevant for determining risk and pricing of option.	Calculated daily with a periodicity of 21 days over one year (252 trading days).
Stock price	Contains most information about the company, market expectations, trends, performance etc.	Continuous over time, retrieved daily
Delta	Calculates change in option price per change in the underlying asset, thus, crucial for delta-gamma hedging.	Continuous over time, retrieved daily.
Historical gamma	Corresponding variable	Continuous over time, retrieved daily.
Volume	Volume can indicate the capacity and demand of the market as well as pricing of options and change of hedging strategy.	Continuous over time, retrieved daily.
Volume Change	Change of volmme can change option pricing hence bring the need to adjust the hedging.	Continuous over time, retrieved daily.
Bull or Bear	Can predict the direction of a market as a whole.	Retrieved Daily.
In/At/Out of Money	Can be an indicator of high/low Delta as said in the theory section	Continuous over time, retrieved daily.

Table 2: Predictors and their Relevance and time span

4.2 Stock Prices, Options Prices and Greeks

Yahoo Finance offers free access to stock prices over a large time period and which can be easily loaded in Python. It also has prices for some call options but that cannot be loaded directly to Python. The Swiss Federal Institute of Technology, EPFL, grants partial access to WRDS database but it only has samples of options data and contained a significant number of wrong or missing information. OPRA dataset from Databento is very comprehensive, maybe a bit too much and it is not entirely free. Polygon.io is also pretty good but it has lots of missing data. There are also other websites that give live data for options but no access to previous data. Since options data can be retrieved easily from the theory stated above, we chose to load Yahoo Finance stock prices and use Black-Scholes model to find the price of call options with a given maturity, strike at a given moment. Same goes for Greeks. If the data collected from Yahoo Finance is accurate, it can be confidently concluded that the options data independently calculated will be free from missing or incorrect information.

The stock prices for Apple from 1st January 2023 to 16th May 2024 and a call option with strike 185$ and maturity 17th May 2024 may be seen in the following figure:

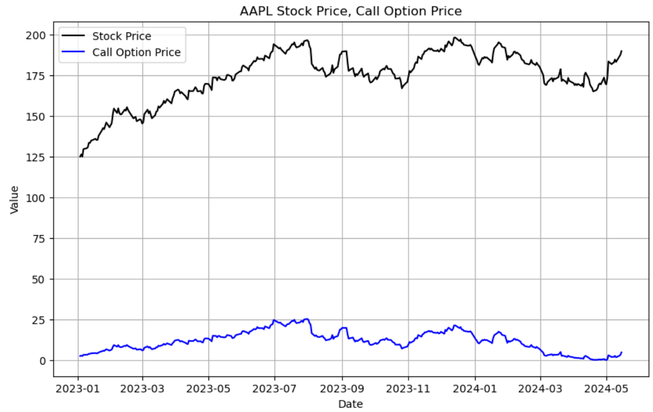


Figure 2: Price of Apple stock and a call option during a 15-month period.

The Greeks for the same period and same option are as follows. We see that Δ and Γ both belong to $[0, 1]$ as they should. The numbers correspond pretty well to the ones we can find in the website Barchart.

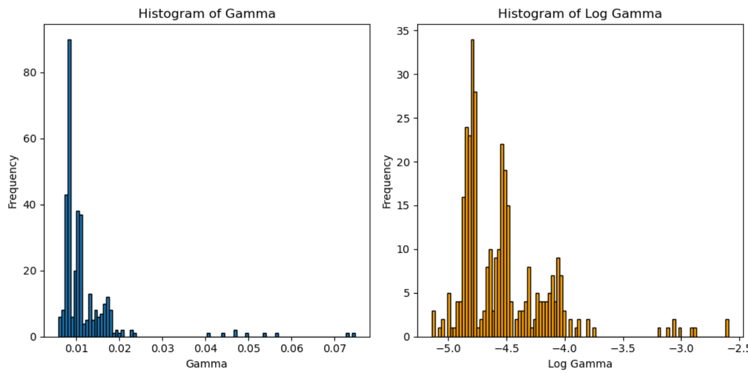
Figure 3: Delta (Δ) and Gamma (Γ) for a call option for Apple stock during a 15 month period.

The historical volatility we chose for stocks comes from Barchart which is calculated over minute data rather than the daily data Yahoo Finance offers.

We rescale the volume variable such that the regression coefficients are not too high or too small.

For our regression, we want to predict Γ at time t and we start by making up our design matrix containing the variables we have at $t-1$. At the latter time, we know the time to maturity and all lagged variables such as prices, volume and Greeks. We also include 2 day lagged variables to look at changes in quantities. Our variables are hence time to maturity, $\Gamma_{t-1}, \Delta_{t-1}, C_{t-1}, S_{t-1}, \Theta_{t-1}, V_{t-1}, \Gamma_{t-2}, \Delta_{t-2}, C_{t-2}, S_{t-2}, \Theta_{t-2}, V_{t-2}$. We drop variables that will have NaN values for lagged variables.

We move on to a closer look to our variables. Our dependent variable Γ has a skewness of 4.93 and it is really right-skewed as shown by the histogram below so we decide to apply a log-transformation to it, which makes it clearly better.



We also look at histograms and boxplots of the independent variables. Some of them are also right-skewed and need log-transformation. The volume is such an example as shown below.

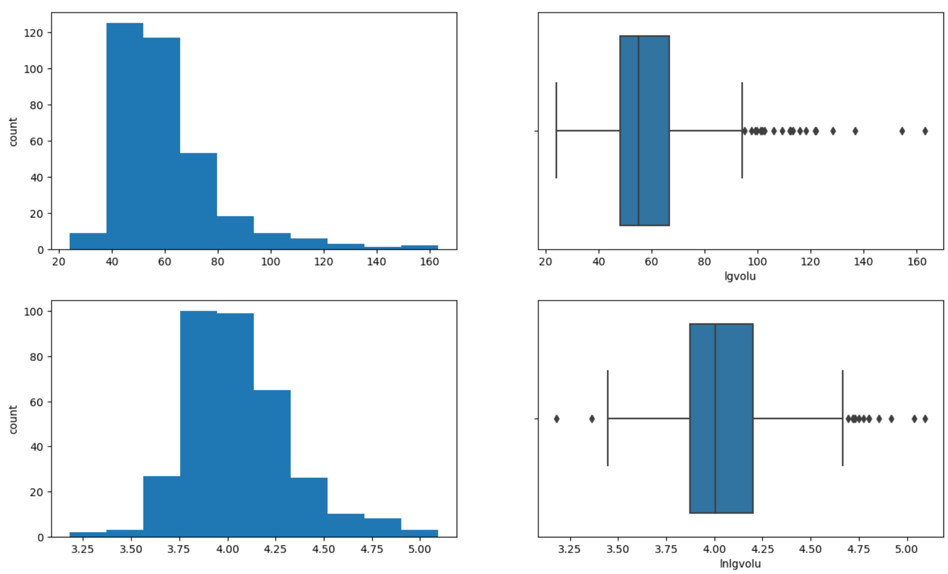


Figure 4: Histogram and boxplot of Lagged Volume before and after log-transformation

We then move on to some bivariate analysis. We plot variables against each other and have some interesting results. We see that some clear relations appear with Γ. The whole picture being too big to export, we only give examples of interesting plots. Here is Gamma against time to maturity, we clearly observe some -log shape.

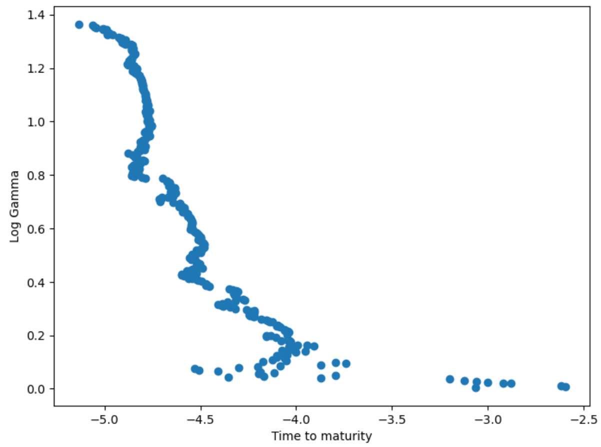


Figure 5: Plot of Log Gamma against time to maturity

Some variables don't seem to have predictive power which matches our expectations about them. It is the case of volume as we can see in the plot below.

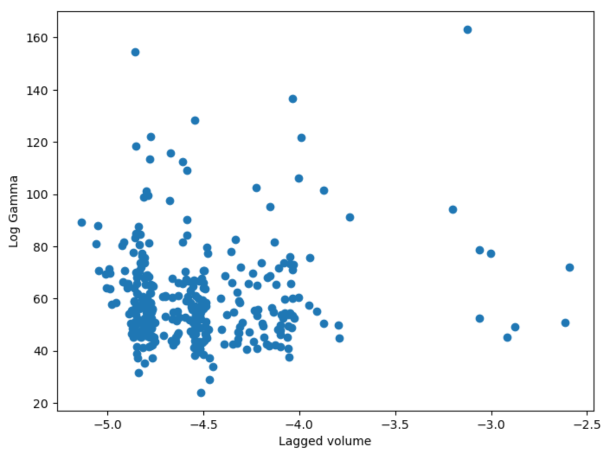


Figure 6: Plot of Log Gamma against lagged volume

There are variables that do not change a lot between time $t-1$ and $t-2$ so to avoid a case of near-multicollinearity, where $\det(X^T X) \approx 0$, we should include one or another, not both.

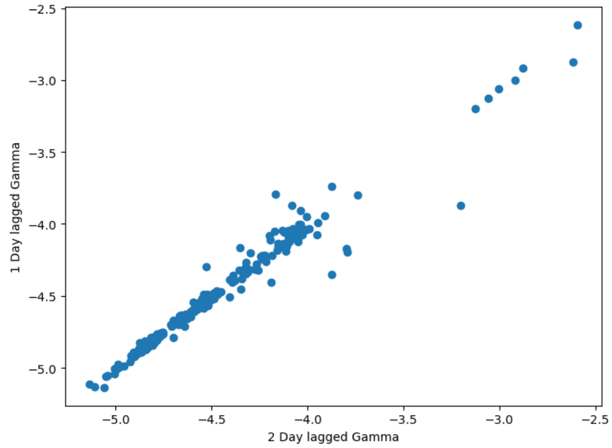


Figure 7: Plot of Log 1 Day lagged gamma vs Log 2 Days lagged gamma

Finally, we have a look at a correlation heatmap of our variables. We see that Γ at time t is very highly correlated with its value at time $t-1$. We also notice as said before that the volume is poorly correlated with Γ so we decide not to keep it for the regression. We can also see that Δ has zero correlation with Γ, we could remove it as we did for volume but it is intuitively linked with Γ in options theory and encodes information about underlying's price so we keep it for now.

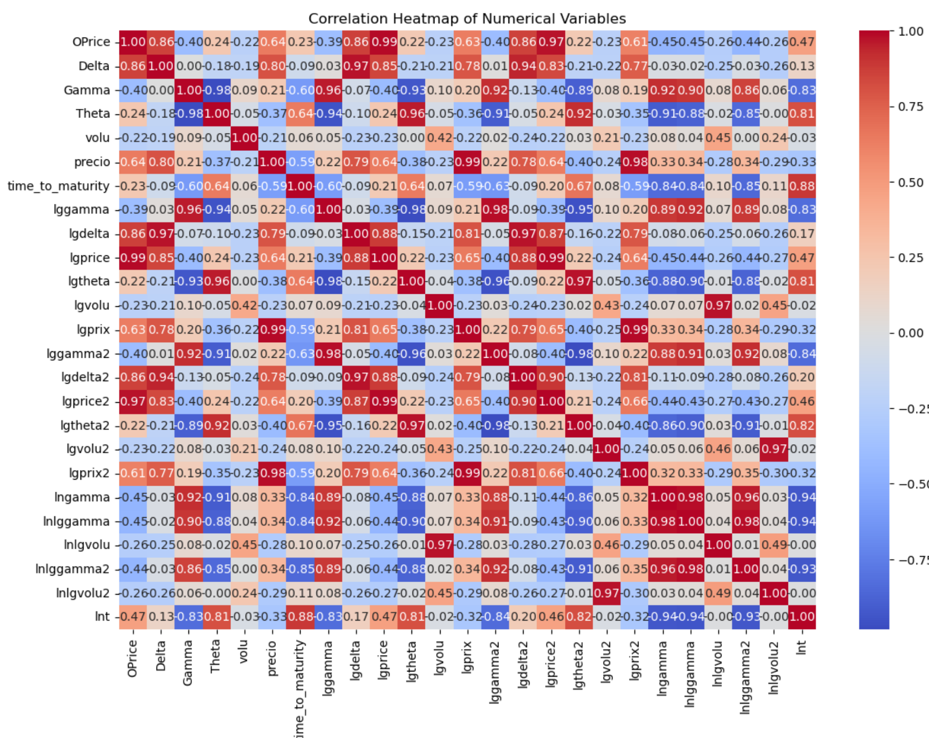


Figure 8: Plot of Log 1 Day lagged gamma vs Log 2 Days lagged gamma

5 Implementation

5.1 Regression

We now delve into the implementation of a linear regression model to predict Gamma at a given time. We must be careful though, as Delta is bounded between 0 and 1, the variation of delta, i.e., Gamma, is also inherently bounded. However, traditional linear regression techniques may overlook this constraint, potentially leading to predictions that exceed the feasible range. To address this challenge, if it arises, we will enforce the bounds of Gamma, by truncating predictions beyond the feasible range or by mapping them to the nearest feasible boundary (0 or 1).

We choose to implement a backward selection to find the right model. We start with a model containing all the regressors and removing one at a time to maximize or minimize the following metrics : R^2, AIC and BIC. The latter procedure leads us to the following table where we removed the worst-performing variable, i.e. the one where the metrics were behaving best.

Table 5.1.1 Metrics of the models through backward selection

Regressors removed	R^2	AIC	BIC
None	0.8256	-1492.579	-1462.764
Lagged Option Price	0.8294	-1494.562	-1468.474
2 Days Lagged Stock Price	0.8278	-1494.935	-1472.574
Lagged Delta	0.8472	-1491.415	-1472.782
Lagged Stock Price	0.8467	-1493.307	-1478.4

Table 5.1.2 Linear regression results of last model

Dependent variable : $\log(\Gamma_t)$	(1)
$\log(\Gamma_{t-1})$	0.93370 ***
	(0.019)
$\log(T)$	-0.04378 ***
	(0.011)
Θ_{t-1}	0.002 *
	(0.001)
Constant	-0.29641 ***
	(0.093)
Out-of-sample R^2	0.8467
Number of observations (in-sample)	307
Number of observations (out-of-sample)	35

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We now plot residuals against the variables we kept to see if there is some missing form that was not captured.

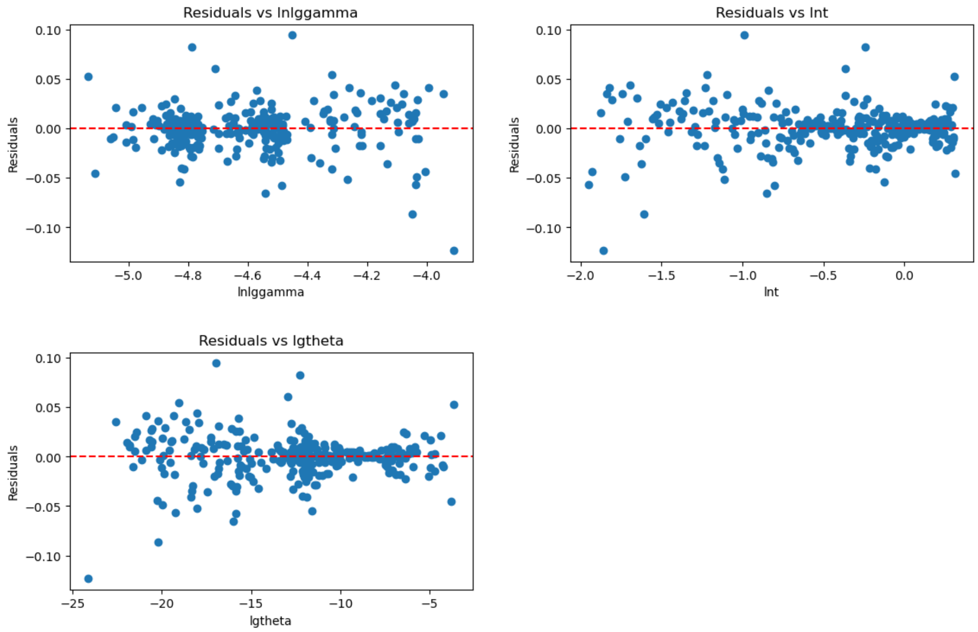


Figure 9: Plot of Log Gamma against lagged volume

We see no significant pattern which is a good sign. We also check for heteroskedasticity by plotting residuals against fitted values.

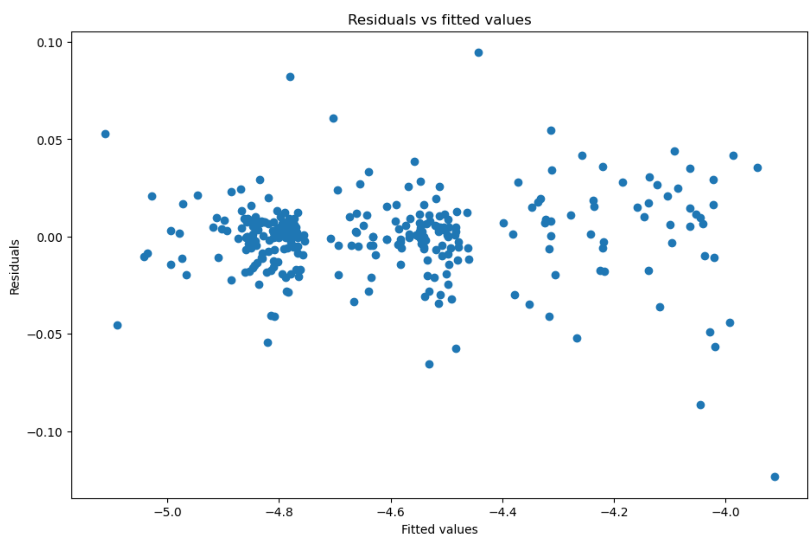


Figure 10: Plot of residuals against fitted values

The plot looks fine. If we compare our predictions with real data, it is actually pretty close!

5.2 Delta-gamma predictions

Suppose we are at date $i-1$ and want a portfolio where we have a hedged position at date i. We have a prediction for the next Gamma values given by :

$$\hat{\Gamma_i} = e^{\ln \hat{\Gamma}_i} e^{\hat{\sigma}^2/2}$$

We still need predictions for the next Delta values. We use a linear approximation the following way:

$$\Delta_i \approx \Delta_{i-1} + \Gamma_{i-1} \cdot (S_i - S_{i-1})$$
$$\frac{\Gamma_i - \Gamma_{i-1}}{(\frac{\partial^3 C}{\partial S})_{i-1}} \approx (S_i - S_{i-1})$$

We then solve the linear system in figure 1 to get the positions we need. Here is an example. Suppose its 16th April 2024 and you are given call options from Apple stock with strike 185 and maturity 04th June 2024. We would like to hedge them with call options of Microsoft stock with strike 410. Volatility details and examples of strike prices can be found either on a excel given of on Yahoo Finance. Suppose you want to hedge the options until the 23rd May. Here are the positions you need to take on the market to be delta-gamma hedged:

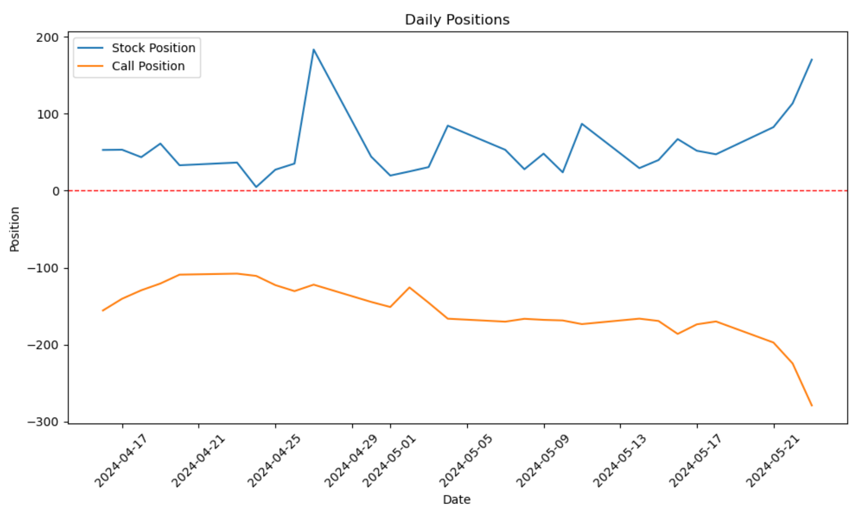


Figure 11: Daily positions on the market to be delta-gamma hedged.

6 Alternative methods results

6.1 Random Forest

Table 3: Feature Importances

Feature	Importance
$\log(\Gamma_{t-1})$	0.468894
$\log(T)$	0.402111
Θ_{t-1}	0.114102
Price_{t-1}	0.005920
C_{t-2}	0.004195
Δ_{t-1}	0.002627
C_{t-1}	0.002151
Mean Squared Error: 0.323615	
R^2 Score: -0.07625	

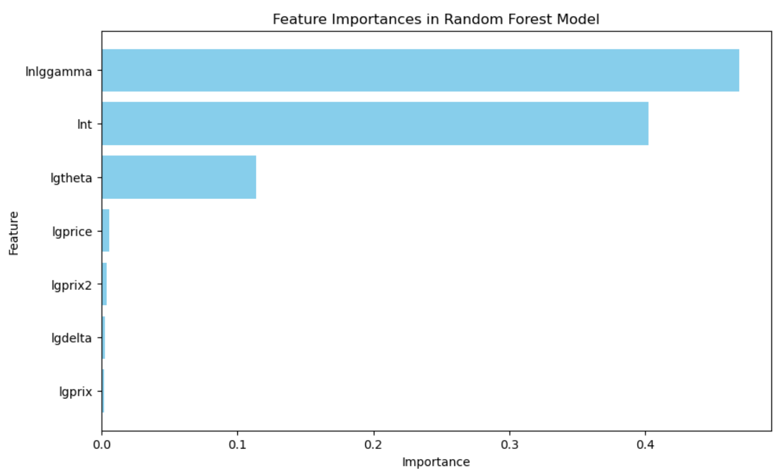


Figure 12: Feature importances of Random forest model

6.2 SVR

Table 4: SVR Model Evaluation with best parameters

Parameter	Value
C	1
Gamma	0.01
Epsilon	0.1
Kernel	linear
Mean Squared Error	0.07667
R^2 Score	0.74502

6.3 XGBoost

Table 5: XGBoost Model Evaluation with Best Parameters

Parameter	**Value**
Regressor	XGBoost
Number of Estimators	200
Learning Rate	0.1
Max Depth	3
Min child weight	1
Subsample	1
Colsample Bytree	0.7
Mean Squared Error	0.32122
$\mathbf{R^2}$ **Score**	-0.06830

7 Methods comparison

We will now compare the different machine learning models used and determine which one is the most performant based on two metrics:R^2 score and the mean squared error. A lower MSE indicates better performance, while a higher R^2 score (closer to 1) signifies a better fit to the data. It corresponds to the proportion of the variance in the dependent variable that is predictable from the independent variables.

First, here is a summary table of the results:

	Linear Regression	Random forest	SVR	XGBoost
Mean Squared Error	0.04612	0.323615	0.07667	0.32122
Out of sample R^2	0.8467	-0.07625	0.74502	-0.06830

Table 6: Summary table of the results

Let's first analyze the results of each model:

- Linear regression: as we can see in Table 6, the mean squared error is relatively low and has pretty good R^2, indicating good performance while having only 4 variables. For prediction purposes, the model does a good job, the best of the 4 methods, but we should watch out for heteroskedasticity depending on stocks.
- Random Forest: The random forest model has a significantly higher mean squared error and a negative R^2 value, indicating poor performance. The negative R^2 suggests that the model is worse than a horizontal line predicting the mean of the target variable. This could be due to overfitting on the training data or insufficient tuning of the hyperparameters. The complexity of random forest, with its numerous decision trees, might not be suitable for this dataset without proper parameter tuning.
- Support Vector Regression (SVR): This method has an MSE that is slightly higher than linear regression but still relatively low compared to the other models. The R^2 value is also fairly high, suggesting that SVR performs well on this dataset. The linear kernel used in SVR gives it similar results to linear regression, though it doesn't outperform linear regression in this case.

- XGBoost: Similar to the random forest, XGBoost has a high mean squared error and a negative R^2 value, indicating poor predictive performance. XGBoost is a powerful algorithm, but its performance heavily relies on the correct tuning of hyperparameters. In this instance, it seems to suffer from either overfitting or inappropriate hyperparameters.

8 Conclusion

In summary, the linear regression model outperforms the other models in this scenario, both in terms of mean squared error and R^2. Random Forest and XGBoost show poor performance, potentially due to inadequate parameter tuning or overfitting. Support Vector Regression, while performing well, does not surpass linear regression. The results suggests that using these linear regressing algorithms would be reasonably feasible for predicting gamma and thus create a Delta-Gamma neutral portfolio in the market.

For future improvement, having access to a better database would be beneficial since the data available from Yahoo Finance is limited. Furthermore, due to the relative underperformance of random forest and XGBoost, a more deliberate parameter tuning for these alternative methods would be well in place. Finally, an implementation of a more robust linear model to potentially remove Theta from regressors would be desirable since heteroskedasticity may arise (depending on stocks), resulting in a falsely low p-value.

For further research or future projects, a next step would be to start a real Delta-Gamma hedging strategy in the market.

References

[1] Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions [Internet]. Vol. 2, SN Computer Science. Springer Science and Business Media LLC; 2021. Available from: http://dx.doi.org/10.1007/s42979-021-00592-x

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

[3] Merton, R. C. (1973). Theory of Rational Option Pricing. The Bell Journal of Economics and Management Science, 4(1), 141–183. https://doi.org/10.2307/3003143

[4] Bachelier, L. (1900). Théorie de la spéculation. Annales Scientifiques de l'École Normale Supérieure, 17, 21–86. https://doi.org/10.24033/asens.476

[5] In Appendix: Prashant Sharma. (2024) Stock Market Prediction Using Machine Learning https://www.analyticsvidhya.com/blog/2021/10/machine-learning-for-stock-market-prediction-with-step-by-step-implementation/

[6] James Chen. (2024) What are Options? Types, Spreads, Example, and Risk Metrics. Investopedia. https://www.investopedia.com/terms/o/option.asp

[7] Jhon Summa.(2024) Options Trading Strategies: Understanding Position Delta. Investopedia. https://www.investopedia.com/articles/optioninvestor/03/021403.asp

[8] James Chen.(2024) What Is Gamma in Investing and How Is It Used?. Investopedia. https://www.investopedia.com/terms/g/gamma.asp

[9] James Chen.(2024) Theta: What It Means in Options Trading, With Examples. Investopedia. https://www.investopedia.com/terms/t/theta.asp

[10] Adam Hayes.(2024) Black-Scholes Model: What It Is, How It Works, and Options Formula. Investopedia. https://www.investopedia.com/terms/b/blackscholes.asp

[11] Scott Guernsey.(2013).AN INTRODUCTION TO THE BLACK-SCHOLES PDE MODEL.University of new Mexico. https://math.unm.edu/~nitsche/mctp/reus/proposals/2013prop_Guernsey.pdf

[12] Martin Haugh.(2016).The Black-Scholes Model.Foundations of Financial Engineering, University of Columbia. https://www.columbia.edu/~mh2078/FoundationsFE/BlackScholes.pdf

A Appendix: Alternative approach

One can say : "Well, since the formula for Gamma is known, what do we actually need to calculate tomorrow's Gamma?"

Γ satisfies the following equation:

$$\log(\Gamma) = \frac{-d_1^2}{2} - \log(S\sigma\sqrt{t}\sqrt{2\pi})$$

$$= \frac{-\log(S/K)^2}{2\sigma^2 t} - \frac{\log(S/K)(r+\frac{\sigma^2}{2})}{\sigma^2} + \frac{(r+\frac{\sigma^2}{2})^2 t}{\sigma^2} - \log(S\sqrt{t}) - \log(\sigma\sqrt{2\pi})$$

Then the only thing we do not know to calculate tomorrow's Γ is the stock price S. If we have a good prediction for the stock price, then we have a good prediction for Γ. We try to predict the stock using an arbitrary LSTM model (its parameters are in the code). Here the input is not the variables used in the methods above but the classic Yahoo Finance dataset. Why? Because the However, as we can see below, sometimes the predicted values can be far from the real ones implying big errors in the Gamma prediction which is usually a small quantity.

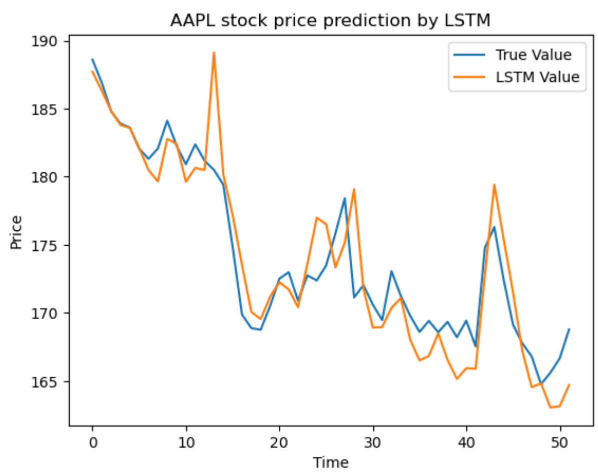


Figure 13: Apple stock price prediction by LSTM