KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Technology

First cycle, 15 credits

Challenges with Using the Black-Scholes Model for Pricing Long-Maturity Options

WILHELM SIGURD
JARL ERIKSSON

Stockholm, Sweden, 2024

Degree Programme in Engineering Physics
Supervisor: Linn Engström
Examiner: Martin Viklund

Abstract

This thesis investigates the application of the Black-Scholes model for pricing long-maturity options, primarily utilizing historical data on S&P500 options. It compares prices computed with the Black-Scholes formula to actual market prices and critically examines the validity of the Black-Scholes model assumptions over long time frames. The assumptions mainly focused on are the constant volatility assumption, the assumption of normally distributed returns, the constant interest rate assumption and the no transaction cost assumption. The results show that the differences between computed prices and actual prices decrease as options get closer to maturity. They also show that several of the Black-Scholes model assumptions are not entirely realistic over long time frames. The conclusion of the thesis is that there are several limitations to the Black-Scholes model when it comes to pricing long-maturity options.

Keywords: Black-Scholes, option pricing, long-maturity options, S&P500, implied volatility, financial model assumptions

Acknowledgements

We would like to thank our supervisor Linn Engström, whose expertise was invaluable in shaping the methodology and aim of this thesis. Her constructive feedback encouraged us to refine our ideas and significantly enhanced the quality of our work.

Stockholm, May 2024
Wilhelm Sigurd & Jarl Eriksson

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Problem . . . 1
- 1.3 Purpose . . . 1
- 1.4 Delimitations . . . 2

2 Theoretical Background **3**
- 2.1 Option Markets . . . 3
- 2.2 Overview of the Black-Scholes Model . . . 3
- 2.3 Presentation of the Black-Scholes Model Assumptions . . . 4

3 Methodology **6**
- 3.1 Data Collection . . . 6
  - 3.1.1 Option Prices . . . 6
  - 3.1.2 S&P500 Prices . . . 7
  - 3.1.3 Interest Data Collection . . . 7
- 3.2 Data Analysis . . . 7
  - 3.2.1 Obtaining Inputs for the Black-Scholes Formula . . . 7
  - 3.2.2 Calculations Involving Historical Option Prices . . . 10

4 Results **11**
- 4.1 Historical Tests . . . 11
- 4.2 Critical Examination of the Black-Scholes Assumptions . . . 14
  - 4.2.1 Constant Volatility Assumption . . . 14
  - 4.2.2 Assumption of Normally Distributed Returns . . . 15
  - 4.2.3 Constant Interest Rate Assumption . . . 17
  - 4.2.4 No Transaction Cost Assumption . . . 18

5 Discussion **19**
- 5.1 Interpretation of Results . . . 19
  - 5.1.1 Historical Tests . . . 19
  - 5.1.2 Critical Examination of the Black-Scholes Assumptions . . . 19
- 5.2 Validity Analysis . . . 20

6 Conclusions and Future Work **22**
- 6.1 Conclusions . . . 22
- 6.2 Future Work . . . 22

References **23**

A Data Collection Code **24**

1 Introduction

This introductory section presents the topic for the thesis. It begins with a brief background on options and the Black-Scholes model for pricing them before introducing the research questions and explaining why they are relevant.

1.1 Background

Derivatives have become increasingly important in finance over the past couple of decades. Derivatives involve two parties agreeing to future transactions and are used by a wide range of traders including hedgers, speculators and arbitrageurs. A common derivative type is the European option. These give holders the right, but not the obligation, to buy or sell a certain asset at a specified date (called maturity date) in the future. In the early 1970s, the Black-Scholes model for pricing European options was developed and has since had a huge influence on option trading [1].

1.2 Problem

Although the Black-Scholes model has been very successful, it has faced criticism. One example is Warren Buffet questioning its reliability for pricing long-maturity options. In his 2008 letter to Berkshire Hathaway shareholders, he stated: "I believe the Black-Scholes formula, even though it is the standard for establishing the dollar liability for options, produces strange results when the long-term variety are being valued" [2]. This thesis delves deeper into precisely this topic and aims at answering the following two questions:

1. How has the Black-Scholes model historically performed on long-maturity options?

2. How realistic are the assumptions of the Black-Scholes model over extended periods?

1.3 Purpose

After the 2008 financial crisis many people argued that financial engineering techniques were partly to blame. Although it is impossible to determine to what extent this is true, financial modelers definitely played a part in the crash [3]. This highlights the importance of being aware of limitations with different financial models like the Black-Scholes model. Hopefully, this thesis can contribute to such awareness.

1.4 Delimitations

A lot of historical option data has been needed for answering the research questions presented above. We have limited ourselves to using data only on European options with the S&P500 as underlying asset and decided not to use any more than two years of historical data due to limited availability of such data.

2 Theoretical Background

This section explains what options are and what role they play and historically have played in financial markets. It then gives an overview of the Black-Scholes model before finally presenting the underlying model assumptions.

2.1 Option Markets

Options are financial derivatives that provide the buyer the right, but not the obligation, to buy (call option) or sell (put option) an underlying asset at a predetermined price, known as the strike price, before or on a specified expiration date. Essentially, buying an option allows the holder to speculate on the future price movement of an asset without actually owning the asset. The cost of buying an option is called the premium, which is paid to the seller of the option. Call options are typically purchased when the buyer anticipates that the asset's price will rise above the strike price by the expiration date, while put options are bought when the expectation is that the price will fall below the strike price. Options are used for various purposes, including hedging risk, increasing leverage, and generating income through premium collection, making them versatile tools in financial markets [4].

2.2 Overview of the Black-Scholes Model

Before the Black-Scholes model came to be, option trading was less structured and more based on simple supply and demand dynamics, along with some rudimentary principles of risk and reward. In the early 1900, prices were determined more by intuition and the direct negotiation between buyer and seller rather than any formal pricing model. Traders used their judgment to estimate how much a particular option should cost based on their expectations of the underlying asset's price movements. There were often no globally standardized contracts, therefore terms such as expiration dates, strike prices, and option sizes could vary widely from one contract to another. This resulted in a chaotic trading environment, where the price of options often had an even more unpredictable pattern than today [5].

Since the Black-Scholes model was developed by Fischer Black, Myron Scholes, and Robert C. Merton, it has been an essential building block in modern financial theory, providing a strong framework for pricing European-style options (options which can only be exercised at the expiration date). The model has allowed traders to develop their economic understanding and rely less on intuition. It has also created growth

of financial markets by enabling the valuation of options with greater precision than before [6]. Below, in equation 1, the formula for a call option is presented:

$$C(S_t, t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \tag{1}$$

To aid understanding of the building blocks of the formula, it may be broken down into its essential components. Below each element is explained [1]:

- $C(S_t, t)$ represents the price of the call option, dependent on the stock price S_t and the time t. It's the value we seek to calculate.
- S_t symbolizes the present market value of the underlying asset.
- K is the strike price, the predetermined price at which the holder of the option can buy the underlying asset.
- T is the maturity time, the point in time when the option expires.
- r denotes the risk-free interest rate, assumed to be constant over the option's life.
- N is the cumulative distribution function of the standard normal distribution and d_1 and d_2 are intermediary calculations within the formula, incorporating the variables S_t, K, r, σ (volatility), and $T - t$ (time to expiration), where:
  - $d_1 = \frac{\ln(\frac{S_t}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$
  - $d_2 = d_1 - \sigma\sqrt{T-t}$
- σ represents the volatility of the underlying asset, a measure of the uncertainty or risk associated with the underlying asset's price changes.

2.3 Presentation of the Black-Scholes Model Assumptions

In the original Black-Scholes paper [7], the Black-Scholes model assumptions are presented and the formula is derived. The assumptions are given as follows:

a) The short-term interest rate is known and is constant through time.

b) The stock price follows a random walk in continuous time with a variance rate proportional to the square of the stock price. Thus the distribution of possible stock prices at the end of any finite interval is lognormal. The variance rate of the return on the stock is constant.

c) The stock pays no dividends or other distributions.

d) The option is "European," that is, it can only be exercised at maturity.

e) There are no transaction costs in buying or selling the stock or the option.

f) It is possible to borrow any fraction of the price of a security to buy it or to hold it, at the short-term interest rate.

g) There are no penalties to short selling. A seller who does not own a security will simply accept the price of the security from a buyer, and will agree to settle with the buyer on some future date by paying him an amount equal to the price of the security on that date.

In this thesis, assumptions a, b and e are particularly focused on. We have split assumption b into two separate assumptions, which means that the thesis primarily investigates four different assumptions. We have chosen to call these:

- The constant volatility assumption (from assumption b)
- The assumption of normally distributed returns (from assumption b)
- The constant interest rate assumption (from assumption a)
- The no transaction cost assumption (from assumption e)

In the second assumption, returns refer to log returns. This is going to be the case for the rest of the paper. See the part about volatility in section 3.2.1 for an explanation of how log returns are calculated.

3 Methodology

The methodology section is split into two parts. The first part focuses on how data has been collected. The second part focuses on how the collected data has been used for making various calculations and analyses.

3.1 Data Collection

All our data was collected with Python through different APIs. The code which was written for collecting data can be seen in appendix A. In total, three different types of data was needed:

1) Data on historical S&P500 option prices
2) Historical price data on the S&P500 itself
3) Interest rate data

Information regarding collection of data for each of these is presented below, including what data sources have been used and specifics on exactly what data has been obtained.

3.1.1 Option Prices

Polygon.io was used for collecting data on historical option prices. Two years of historical end of day data was available for free, but a payment was required in order to get unlimited API calls. The following steps were carried out each time historical S&P500 option prices were collected:

- An expiration date was chosen (could be for example 2024-04-19)
- Option contracts expiring at that date and having existed for at least one year (could be for example since 2023-04-19) were listed
- Daily closing prices and volumes from the year before the expiration date (could be for example between 2023-04-19 and 2024-04-19) were collected

The expiration date which was used for generating most of the final figures was 2024-04-19. For this expiration date call and put options with strike prices ranging from 200 to 6600 existed.

3.1.2 S&P500 Prices

Finding data on historical S&P500 prices was much easier than finding data on historical option prices. The data source utilized for this was the Yahoo Finance API. It was used to collect daily closing prices for the S&P500 for each trading date since 1930-01-01. It was also used for collecting historical data on the VIX index. The VIX index gives an indication of the market's expectation on future S&P500 volatility and was therefore of interest. VIX data was collected for all trading days since 1990-01-02.

3.1.3 Interest Data Collection

Several sources were relied upon when collecting interest rate data. Data on SOFR (*secured overnight financing rate*) was collected from the Federal Reserve Bank's FRED API. SOFR is a comprehensive measure reflecting the cost of borrowing cash overnight and is commonly used for pricing US derivatives [1]. SOFR data was available from back to 2018-04-03. In addition to the SOFR data, rates on various US Treasury bills (1, 3, 6 and 12 months) were collected via API through the US Treasury's website to reflect the cost of borrowing over longer periods of time. Treasury rates since the beginning of 2014 were collected.

3.2 Data Analysis

Once data had been gathered, the project could move into the data analysis stage. Appropriate inputs to the Black-Scholes formula had to be determined and calculations involving historical option prices had to be made.

3.2.1 Obtaining Inputs for the Black-Scholes Formula

To use the Black-Scholes formula, many inputs are needed. Prices for the underlying asset (in this case the S&P500) and option strike prices were easy to get directly from the raw collected data. In order to obtain the remaining parameters, however, calculations had to be made. Information regarding how appropriate time to maturities, interest rates and volatilities have been calculated is presented below.

Time to Maturity To calculate time left to maturity either trading days or calendar days could be used. Historical volatility (see below) was calculated using trading days and ideally the same approach should have been used for the maturity calculations. However, calendar days were used at the start of the project and this

approach was kept throughout the project for simplicity. This means, for example, that the time left to maturity 2024-04-10 for an option expiring 2024-04-19 was set to $9/365 \approx 0.025$ years instead of $7/252 \approx 0.028$ years.

Interest Rate Appropriate interest rates have been calculated using a combination of SOFR and US Treasury rates according to the function $i(t)$ shown in figure 1. For example, $i(t)$ would provide an interest rate between the 6 month and 12 month US Treasury rate for an option with 7 months left to expiration.

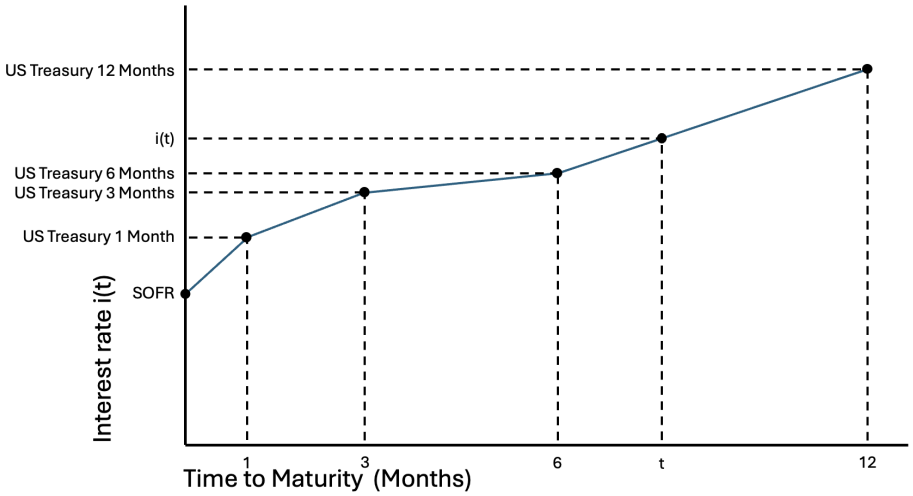


Figure 1: Shows how SOFR and US Treasury rates have been combined into a continuous function $i(t)$ outputting an interest rate for a given time t

Volatility To determine appropriate volatility inputs to the formula, historical volatilities have been used. The first step when calculating historical volatilities has been to calculate log returns for the underlying asset on a daily basis according to equation 2.

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{2}$$

where:

- R_t is the log return on day t

- P_t is the closing price of the asset on day t
- P_{t-1} is the closing price of the asset on the previous day, $t-1$

Different time periods (1, 3, 6 and 12 months) have then been used to calculate historical volatilities. It has been assumed that a trading month consists of 21 trading days, which means, for example, that to calculate the 1 month historical volatility, data from the past 21 trading days have been used. The standard deviation of daily returns was calculated according to equation 3.

$$\sigma_{daily} = \sqrt{\frac{1}{N-1} \sum_{i=t-N+1}^{t} (R_i - \bar{R})^2}, \tag{3}$$

where:

- σ_{daily} is the standard deviation
- N is the number of trading days being used
- R_i is the daily log return on day i
- $\bar{R}$ is the average daily return over the window $[t-N+1, t]$

To obtain the annualized volatility, the assumption that daily returns are independent and identically distributed together with the assumption that a US trading year consists of 252 trading days lead to equation 4

$$\sigma_{annualized} = \sigma_{daily} \times \sqrt{252} \tag{4}$$

where:

- $\sigma_{annualized}$ represents the annualized volatility
- σ_{daily} stands for the standard deviation of daily returns
- 252 signifies the approximate number of trading days within a year

Once 1, 3, 6 and 12 months historical volatilities had been calculated for each date in the dataset, the appropriate volatility to use as input was determined in a similar way to how the appropriate interest rate was determined. The function $\sigma(t)$ which was used can be seen in figure 2.

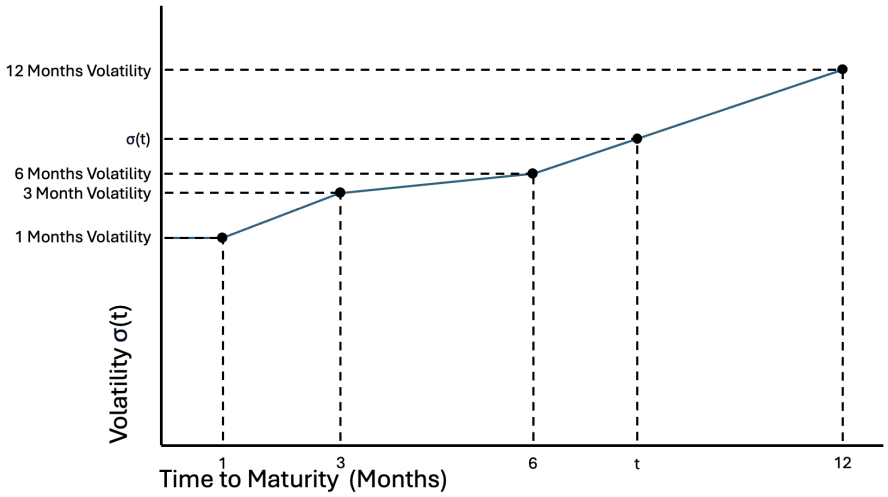


Figure 2: Shows how historical volatilities for different time durations have been combined into a continuous function $\sigma(t)$ outputting a volatility for a given time t

3.2.2 Calculations Involving Historical Option Prices

Two key calculations involving historical option prices were made. Firstly, historical option prices were compared to prices given by the Black-Scholes model with the parameters calculated above. Secondly, implied volatilities were calculated. The implied volatility is the volatility that causes the calculated Black-Scholes price to equal the observed market price of the option.

Price Comparisons To compare calculated Black-Scholes prices with actual prices, absolute price differences have been used in various ways. The absolute price differences were computed according to:

$$\text{Absolute price difference} = |\text{Calculated price} - \text{Actual closing price}| \tag{5}$$

Calculating Implied Volatility To calculate implied volatilities, an iterative numerical process was used. If the market price is denoted by C_{Market} and the Black-Scholes price by $C_{BS}(\sigma)$ (which is a function of the volatility σ), finding the implied volatility is equivalent to finding the root to $f(\sigma)$ in the following equation:

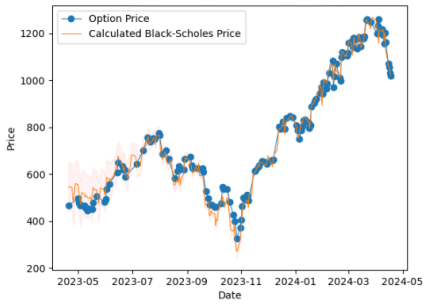
$$f(\sigma) = C_{BS}(\sigma) - C_{Market} \tag{6}$$

4 Results

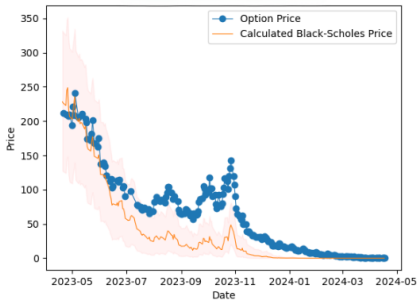
The result section is divided into two parts. The first part presents the results from comparisons of historical option prices to prices calculated by the Black-Scholes formula. The second part consists of a critical examination of some of the underlying assumptions of the Black-Scholes formula.

4.1 Historical Tests

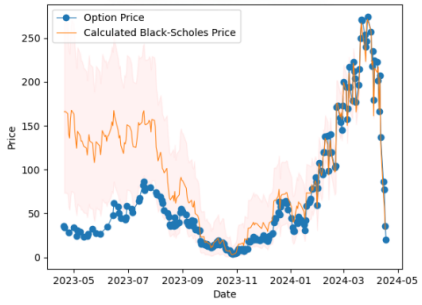
To evaluate the performance of the Black-Scholes model in predicting option prices over time, calculated prices have been compared to actual closing prices for options expiring at 2024-04-19. This has been done over the course of a year for both call and put options with strike prices of 4000$ and 5000$. The resulting graphs can be seen in figure 3.



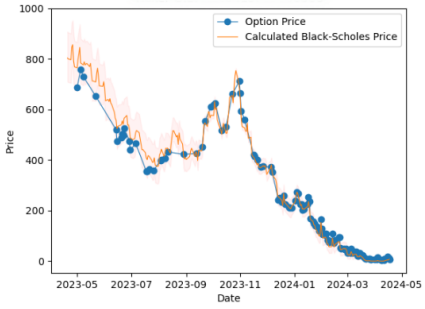
(a) Call option with strike price 4000



(b) Put option with strike price 4000



(c) Call option with strike price 5000



(d) Put option with strike price 5000

Figure 3: Comparison between Black-Scholes prices and actual closing prices for options with expiration date 2024-04-19

In figure 3a and 3d, there is a close relationship between the option close price in blue and the calculated Black-Scholes price in orange. In figure 3b and 3c, however, there are initially some differences between calculated and actual closing prices. The shaded red area shows what the calculated Black-Scholes prices could have been if the volatility input would have been increased or decreased by up to 30%. Despite the initially large differences between calculated and actual closing prices in figure 3b and 3c, the actual option prices are almost always within the red shaded area. It is very clear that the impact of volatility is more pronounced for options with long times to maturity. This highlights the importance of volatility in option pricing models, particularly over extended durations, where the sensitivity to the estimated volatility is very significant.

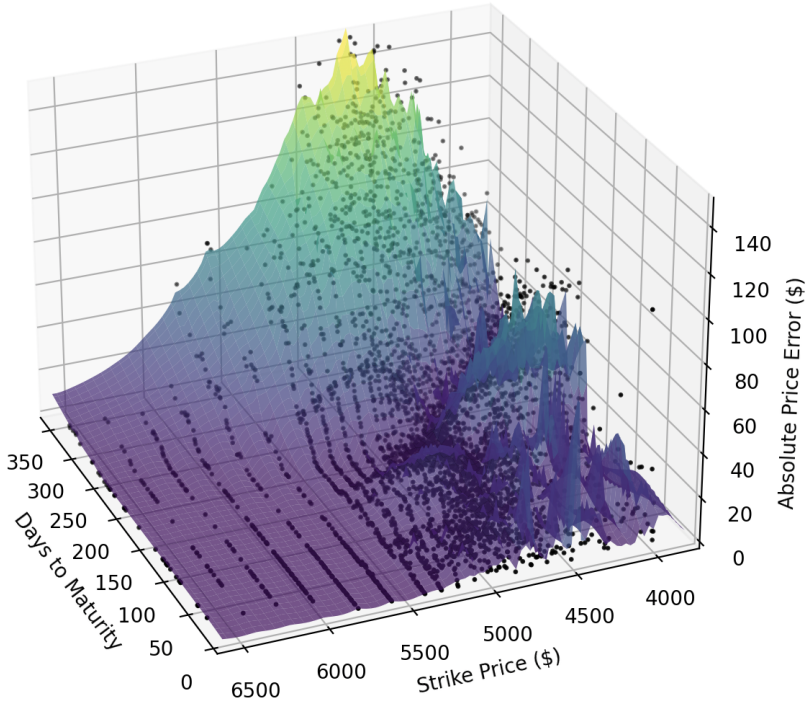


Figure 4: Absolute price difference against strike price and days to maturity for call options with expiration date 2024-04-19

The absolute price differences between the calculated prices and the actual prices have been computed and plotted in a 3D graph for different strike prices and time to maturities. Only call options with a strike price between $3800 and $6600 have been

included in the graph for visual clarity. The results are presented in figure 4. An interpolation between the data points has been included to further improve visual clarity. This interpolation gives an indication that the absolute price differences approach zero for all strike prices as time to maturity goes towards zero. Once again, all data points in the figure come from options with 2024-04-19 as expiration date.

Figure 4 also gives some indications regarding the distribution of the available data. There are many options with strike prices close to the price of the underlying asset and there are many data points near the expiration date.

To further analyze how absolute price differences develop as time to maturity goes towards zero, the mean absolute price differences for all options in the dataset was calculated for each time to maturity. This resulted in one data point per time to maturity and with these data points, a one month rolling average (21 trading days) was computed. At first this was done for the expiration date 2024-04-19. The process was then repeated for many different expiration dates. The resulting data points and rolling averages for all the tested expiration dates are displayed in figure 5. For all tested expiration dates, there is a very clear decreasing trend.

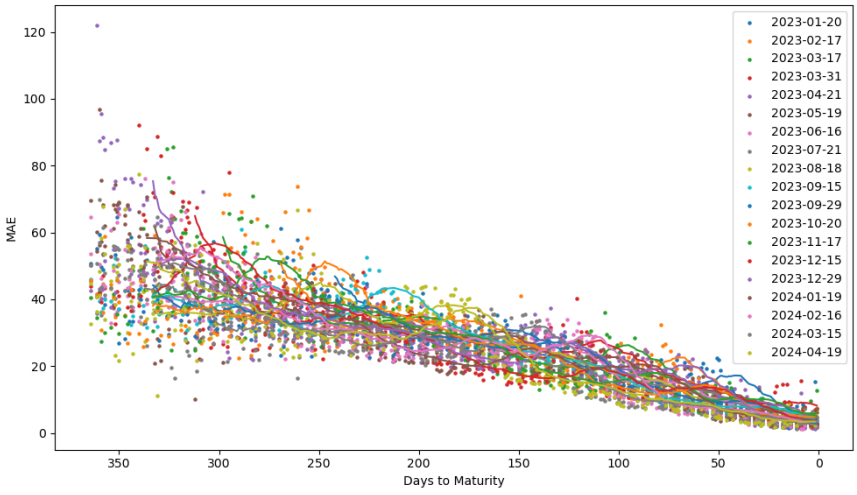


Figure 5: Mean absolute price difference over time for different expiration dates

4.2 Critical Examination of the Black-Scholes Assumptions

In this part of the result section, four of the key assumptions of the Black-Scholes model are examined. The goal is to analyze how realistic they are, particularly over long time frames. The four assumptions are:

1) Constant volatility assumption
2) Assumption of normally distributed returns
3) Constant interest rate assumption
4) No transaction cost assumption

4.2.1 Constant Volatility Assumption

To test how well the constant volatility assumption holds, historical volatility based on data from one month (21 trading days) has been plotted alongside the VIX index over roughly a ten year period. The resulting graph can be seen in figure 6. Two observations can be made. Firstly, the VIX index in general seems to be higher than the historical volatility. Secondly, peaks in the VIX index seem to be followed by peaks in the historical volatility. It is very clear that both time series exhibit significant variations and that volatility can not be considered constant in practice.

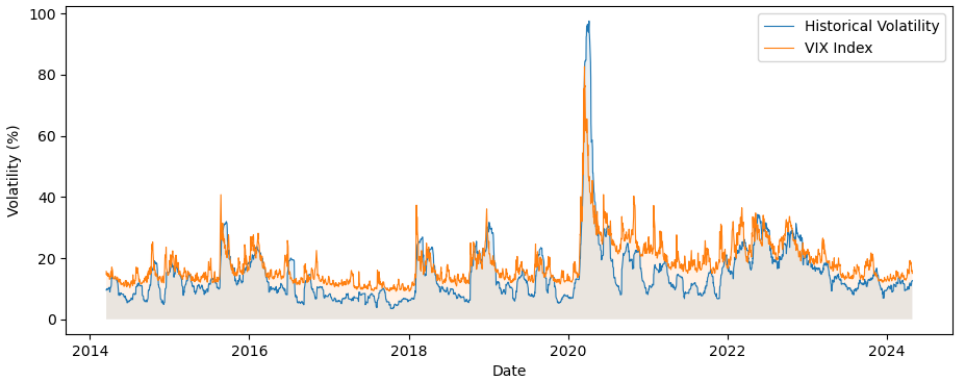


Figure 6: Comparison between historical volatility and the VIX index

To test whether the constant volatility assumption is more accurate for short time frames than long ones, tests regarding how well historical volatility predicts future volatility for various time frames have been made. For each trading day between 2000-01-01 and 2023-01-01, both historical and future volatilities have been calculated with 1, 3, 6 and 12 months of data. The mean absolute difference between

historical and future volatility has then been calculated for each time frame. Note that the future volatility at one date is the historical volatility at a later date. The reason 2023-01-01 was chosen as the end date was to ensure it was possible to calculate future volatilities for all included dates. The results from the tests can be seen in table 1. It is clear that, for the tested time windows, the predictive power of historical volatility decreases the longer the time frame used is.

Table 1: Mean absolute difference between historical and future volatility

Time window (months)	Mean absolute difference (percentage points)
1	5.52
3	5.97
6	6.54
12	7.38

4.2.2 Assumption of Normally Distributed Returns

To investigate the validity of the assumption of normally distribution returns, a histogram visualizing the historical distribution of daily S&P500 returns since 1930-01-01 was created. The mean daily return was 0.023% and the standard deviation was 1.2%. The resulting histogram can be seen in blue in figure 7. The orange curve is the normal distribution with the same mean and standard deviation as the data. It is clear that the actual distribution of returns is not normal. This is partly due to the fact that returns close to zero are more probable than the normal distribution would suggest, but also because extreme returns occur more often than the normal distribution would suggest. A close-up on the tail of the distribution can be seen in figure 8.

One way to investigate whether the market expects future returns to be normally distributed is to look at implied volatilities. When implied volatilities are plotted as a function of strike price, it is very clear that the result is not a straight line. Figure 9 shows such a graph for week 31 2023 (the expiration date used is 2024-04-19). Each data point represents the average implied volatility during the week for call options with a certain strike price. The result is what is known as a volatility smile. The implied volatility gives an indication of how probable it is that a price will be reached according to the market. Implied volatility is greatest for very high and low strike

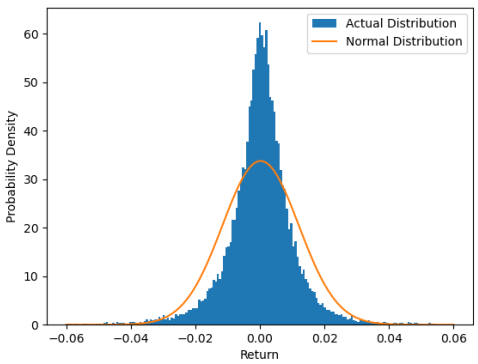


Figure 7: Distribution of returns

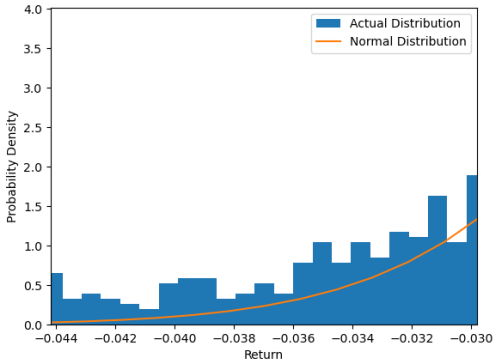


Figure 8: Tail of distribution

prices, suggesting that extreme price movements are more likely than the normally distributed returns used in the Black-Scholes formula would suggest.

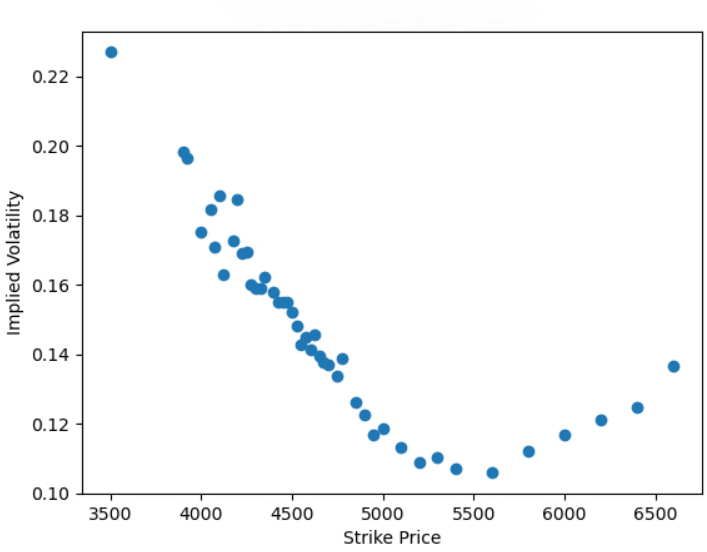


Figure 9: Implied volatility for call options week 31 2023

4.2.3 Constant Interest Rate Assumption

Figure 10 shows the interest rates on US treasury bills from 2014-04-18 to 2024-04-18. It is clear that all plotted rates have varied significantly over the ten year period. A test of to what extent interest rates have varied within differently sized time windows (over the same ten year period) has also been conducted by calculating average standard deviations. For example, the mean standard deviation of the one month interest rate over all possible one month time windows during the ten year period has been computed. The results are presented in table 2.

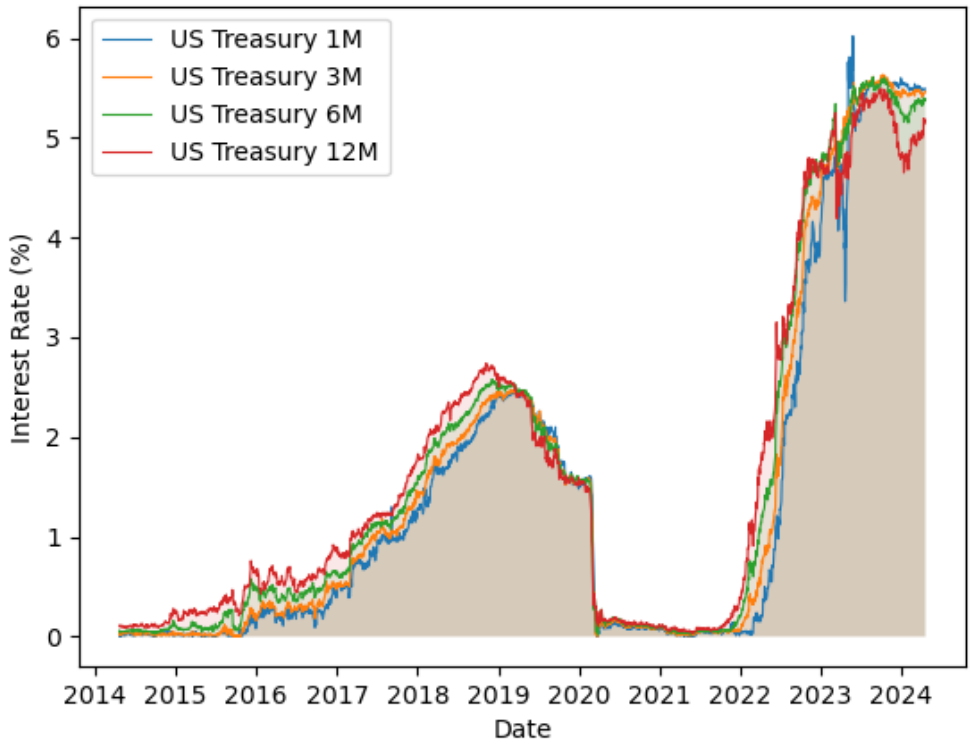


Figure 10: Rates on treasury bills over time

Table 2: Mean interest rate standard deviation for different time window sizes

Time window size (months)	Mean standard deviation (percentage points)
1	0.054
3	0.096
6	0.19
12	0.37

4.2.4 No Transaction Cost Assumption

Thoroughly investigating the amount of transactions costs for different time to maturities would require a lot of data. However, looking at trading volumes could give an indication of liquidity and thereby transaction costs. The same expiration dates used for generating figure 5 were used for calculating average total trading volumes for different number of days to maturity. The resulting graph, including a one month rolling average in blue, can be seen in figure 11. There is a very clear increase in trading volumes as options get closer to their maturity date.

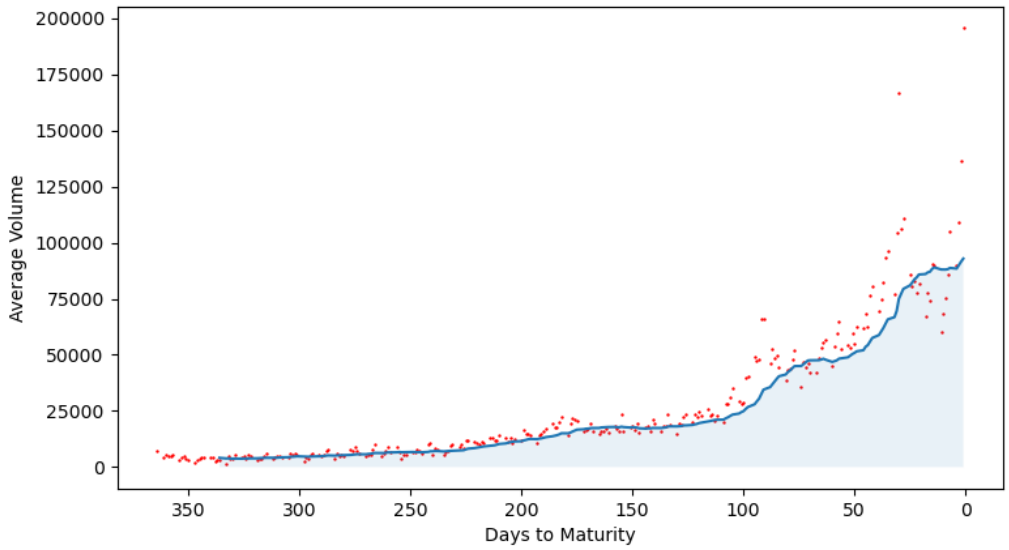


Figure 11: Average total trading volumes for different times to maturity

5 Discussion

This discussion section will first interpret the results presented in section 4 and then present possible sources of error as well as how these may have impacted the results.

5.1 Interpretation of Results

Here the results presented in section 4.1 and 4.2 are interpreted. The results are used to answer the two research questions from the introduction.

5.1.1 Historical Tests

The goal of section 4.1 was to answer how the Black-Scholes model has performed historically on long-maturity options. The findings of the section are clear. Absolute price differences between calculated Black-Scholes prices and actual closing prices decrease as time to maturity decreases and this is seen in figure 3, 4 and 5.

These results should not be surprising. At the expiration date the price of the option is known and consists purely of the option's intrinsic value, but when there is time left to maturity the price of the option will also include a time value. As time to maturity increases, uncertainty grows and determining the correct time value for the option becomes more difficult. This should be the case not only for the Black-Scholes model, but for all option pricing models.

In the Black-Scholes model, volatility is the main parameter of interest when it comes to determining the time value of an option. The sensitivity to the volatility input for different time to maturities is very clearly illustrated by the red shaded area in figure 3. For short times to maturity, changing the volatility by up to 30% barely matters, but for long times to maturity, it has a significant impact.

5.1.2 Critical Examination of the Black-Scholes Assumptions

Section 4.2 aimed to answer how realistic the Black-Scholes model assumptions are over long time frames. The findings show that none of the four tested assumptions seem to fully hold in practice. For all the three assumptions where comparisons were made between different time durations (every single one except the assumption of normally distributed returns), there were also indications that they seemed to hold better for shorter times. Below the tests regarding each of the four assumptions are discussed in more detail.

Constant Volatility Assumption It is clear from figure 6 that historical volatility has not been constant over time. It is also clear, from the VIX index, that market expectation for future volatility has varied significantly over time. Table 1 gives an indication that the constant volatility assumption might be more realistic for short time frames than long ones. However, additional tests with more data (on assets other than the S&P500) would have been useful since the current differences are not huge. Overall, it is evident that the constant volatility assumption is not realistic and that this perhaps could especially impact the pricing of long-maturity options.

Assumption of Normally Distributed Returns Figure 7 and 8 show that historical daily returns on the S&P500 have not been normally distributed, partly because of the fat tails of the distribution. Figure 9 also gives a clear indication that the market generally does not expect future S&P500 returns to be normally distributed. It is clear that the assumption of normally distributed returns does not hold in practice. However, it is also easy to see from the data how this assumption can be a useful starting point for modeling purposes.

Constant Interest Rate Assumption Figure 10 shows that interest rates have varied significantly over time and table 2 shows that the variations have been bigger for long time frames. This is in line with what was expected and does not come as a surprise. The interest rate clearly is not a constant in the real world. Investigating how this could impact option prices would be interesting. Perhaps it barely impacts them as long as the average interest rate up until maturity is known.

No Transaction Cost Assumption It is well-known that in the real world, transaction costs exist. If trading volumes serve as a good indicator of transaction costs, the no transaction cost assumption holds up worse for long times to maturity, as seen in figure 11. Once again, it would be interesting to investigate how the validity of this assumption could impact option prices.

5.2 Validity Analysis

There are a few possible sources of error which are worth mentioning. Many of the options in the dataset have not been traded frequently. During some days they have not been traded at all and during others their trading volumes have been very low. Because of this there is a risk that the reported closing prices of options in reality come from trades which have been executed during the middle of the day. Using such prices together with actual closing prices of the S&P500 could be a possible

source of error. However, most likely it is not a significant source of error. This is because sometimes the reported closing prices will be lower than what the actual closing prices would have been and sometimes they will be higher.

Another source of error is the way we have calculated inputs to the Black-Scholes model. Estimation of future volatility could have been done in various way and there are probably many analysts who could have provided better estimates than we were able to do with historical data. The same thing can be said about the approach for determining the appropriate interest rate to use as input to the model. The fact that we used calendar days instead of trading days for calculating time to maturity is also a possible source of error.

Finally, the fact that all the analyzed options have the S&P500 as underlying asset and that all the option data comes from a two-year span is definitely worth a mention. Ideally, to increase the reliability of the results, historical option data from many different points in time with many different underlying assets should have been used.

6 Conclusions and Future Work

This section wraps up the report by presenting conclusions and ideas for future work.

6.1 Conclusions

This study illustrates the complexities involved in accurately pricing long-maturity options. The differences between Black-Scholes prices and actual prices grow as time to maturity increases and it is evident that several of the assumptions underlying the Black-Scholes model do not hold especially well over long time frames. Overall, this means that the thesis goals have been successfully met. However, there are definitely things which would have been done differently if the project was repeated or if there would have been more time. Some such improvement suggestions and other ideas for future work are presented below.

6.2 Future Work

The first improvement suggestion would be to use data from options with many different underlying assets and from different points in time. Our original thought was to use options on Swedish stocks or indices, but due to limited data this was unfortunately not a possibility this time.

Another idea for improvement would be to modify how the historical testing has been done. Option pricing models aim at valuing the time value of options. Because of this, comparing predicted time values with actual time values would have been interesting and perhaps a better way of assessing the performance of an option pricing model over time. However, specifics regarding how to carry out this testing (for example how to handle negative time values) would have had to be figured out.

The testing could also have been done by somehow looking at actual payoffs of options at expiry. This is the ultimate test of the performance of an option pricing model since it does not assume that market prices are correct. Conducting such tests would have been very challenging, but interesting.

A final suggestion for future work would be to investigate alternative models which have improved upon the Black-Scholes model. There exist many such variations, for example ones which do not assume constant volatility. Exploring such models would have been a natural next step after this thesis.

References

[1] John C. Hull. *Options, Futures, and Other Derivatives*. Pearson, 11th edition, 2021.

[2] Berkshire Hathaway Inc. 2008 shareholder letter. `https://www.berkshirehathaway.com/letters/2008ltr.pdf`, 2008. Accessed: June 5, 2024.

[3] Emanuel Derman and Michael B. Miller. *The Volatility Smile*. Wiley, 2016.

[4] Tomas Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 3rd edition, 2009.

[5] Lyndon Moore and Steve Juh. Derivative pricing 60 years before Black-Scholes: Evidence from the Johannesburg stock exchange. *Journal of Finance*, 61(6):3069–3099, 2006.

[6] Veritasium. The trillion dollar equation, Feb 2024. Accessed: June 5, 2024.

[7] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

A Data Collection Code

```
import pandas as pd
import numpy as np
import ustreasurycurve
from polygon import RESTClient
from datetime import date, datetime, timedelta
import yfinance as yf
from fredapi import Fred


def collect_option_data(expiration_date, api_key="..."):
    """
     Collects historical price data for options with a
        specified expiration date ranging back to one year
         before expiry. Saves the data to an Excel file
        called "option_data.xlsx".

     Parameters:
         expiration_date (str): The expiration date for
            the options, formatted as "YYYY-MM-DD".
         api_key (str): API key for the Polygon API.
     """

    client = RESTClient(api_key)

    expiration_date = datetime.strptime(expiration_date, "
       %Y-%m-%d").date()
    as_of_date = expiration_date.replace(year=
       expiration_date.year-1)

    options = client.list_options_contracts(
       underlying_ticker="SPX", expiration_date=
       expiration_date, as_of=as_of_date)

    data = {
        "Date": [],
        "Close": [],
```

```
        "Volume": [],
        "Ticker": [],
        "Strike": [],
        "Type": [],
        "Style": [],
    }

    for option in options:
        aggregates = client.get_aggs(ticker=option.ticker,
            multiplier=1, timespan="day", from_=as_of_date
           , to=date.today())
        for aggregate in aggregates:
            data["Date"].append(aggregate.timestamp)
            data["Close"].append(aggregate.close)
            data["Volume"].append(aggregate.volume)
            data["Ticker"].append(option.ticker)
            data["Strike"].append(option.strike_price)
            data["Type"].append(option.contract_type)
            data["Style"].append(option.exercise_style)

    df = pd.DataFrame(data)
    df["Date"] = pd.to_datetime(df["Date"], unit="ms").dt.
       normalize()
    df["Days to Maturity"] = (pd.Timestamp(expiration_date
       ) - df["Date"]).dt.days
    df["Years to Maturity"] = df["Days to Maturity"] / 365

    df["Date"] = df["Date"].dt.date
    df.set_index("Date", inplace=True)
    df.to_excel("option_data.xlsx")


def collect_underlying_data(start_date):
    """
    Collects and processes underlying stock market data
       and volatility index (VIX) data starting from a
       specified date. Saves the data to an Excel file
       called "underlying_data.xlsx".
```

```
    Parameters:
        start_date (str): The start date from which to
            begin collecting data, formatted as "YYYY-
            MM-DD".
"""

start_date = datetime.strptime(start_date, "%Y-%m-%d")
    .date()
sp500 = yf.Ticker("^GSPC")
price_data_underlying = sp500.history(start=start_date
    -timedelta(days=400), end=date.today())
price_data_underlying.reset_index(inplace=True)
price_data_underlying["Date"] = price_data_underlying[
    "Date"].dt.date
price_data_underlying.set_index("Date", inplace=True)
price_data_underlying = price_data_underlying[["Close"
    ]]
price_data_underlying["R"] = np.log(
    price_data_underlying["Close"] /
    price_data_underlying["Close"].shift(1))

for number_of_months in [1, 3, 6, 12]:
    trading_days = number_of_months * 21
    price_data_underlying[f"{number_of_months}M Vol"]
        = price_data_underlying["R"].rolling(window=
        trading_days).std() * (252 ** 0.5)

price_data_underlying = price_data_underlying[
    price_data_underlying.index >= start_date]
price_data_underlying = price_data_underlying.rename(
    columns={"Close": "Close Underlying"})

VIX = yf.Ticker("^VIX")
VIX_data = VIX.history(start=start_date, end=date.
    today())
VIX_data.reset_index(inplace=True)
VIX_data["Date"] = VIX_data["Date"].dt.date
```

```
    VIX_data.set_index("Date", inplace=True)
    price_data_underlying = pd.merge(price_data_underlying
       , VIX_data["Close"], on="Date", how="left")
    price_data_underlying.rename(columns={"Close": "VIX"},
        inplace=True)
    price_data_underlying["VIX"] = price_data_underlying["
       VIX"] / 100

    price_data_underlying.to_excel("underlying_data.xlsx")


def collect_interest_rate_data(start_date, api_key="..."):
    """
    Collects and processes interest rate data. Saves the
       data to an Excel file called "interest_rate_data.
       xlsx".

    Parameters:
        start_date (str): The start date for data
           collection, formatted as "YYYY-MM-DD".
        api_key (str): API key for the FRED API.
    """

    treasury_data = ustreasurycurve.nominalRates(
       start_date, date.today().strftime("%Y-%m-%d"))
    treasury_data = treasury_data[["date", "1m", "3m", "6m
       ", "1y"]]
    treasury_data.rename(columns={"date": "Date", "1m": "1
       M Rate", "3m": "3M Rate",
                                  "6m": "6M Rate", "1y": "
                                     12M Rate"}, inplace=
                                     True)
    treasury_data["Date"] = treasury_data["Date"].dt.date
    treasury_data.set_index("Date", inplace=True)

    fred = Fred(api_key=api_key)
    sofr_data = fred.get_series("SOFR")
    sofr_data.index.name = "Date"
```

```
    sofr_data = sofr_data.reset_index()
    sofr_data.columns = ["Date", "SOFR"]
    sofr_data["SOFR"] = sofr_data["SOFR"]/100
    sofr_data["Date"] = sofr_data["Date"].dt.date

    df = pd.merge(treasury_data, sofr_data, on="Date", how
       ="left")
    df.set_index("Date", inplace=True)
    df.to_excel("interest_rate_data.xlsx")


def main():
    collect_option_data(expiration_date="2024-04-19")
    collect_underlying_data(start_date="1930-01-01")
    collect_interest_rate_data(start_date="2014-04-19")


if __name__ == "__main__":
    main()
```