Fundamentals of Computational Mathematics - Numerical calculus

Project: Constructing forward curves of Futures prices

Names and Identification Numbers

Tong Su - 2105697

Riccardo Niero - 2125392

Gianvittorio Parolin - 2120003

Francesco Landi - 2103413

School of Mathematics
University of Padua

16/02/2024

Statements

We declare that this project entitled *"Constructing forward curves of Futures prices"* was carried out by us four in the Department of Mathematics. Information from reference works and specialized publications has been duly acknowledged in the text in which a list of references is provided. No part of this project has been previously submitted to another degree by this institution or any other.

Contents

Abstract

Constructing Forward Curves is a critical process in evaluating *market-exposed assets*. These curves establish the prices at which future delivery contracts can be settled today, providing valuable insights into expected price movements. In the construction of forward curves, various interpolation methods are employed to estimate missing values along interest rate or financial product price curves.

The aim of the project is to estimate the price of crude oil's futures contract value. A future contract is a standardized contract with which the parties agree to exchange a certain asset at a *pre-established* price and with settlement deferred to a future date.

At fist is reported a brief introduction on forward curves, forward and futures contracts and their differences. Then, after a Data description, the Polynomial model is presented along a detailed graphic representation of the results. Subsequently, an alternative model, Geometric Brownian Motion, is developed and its results are compared with the ones extracted from the previous model and some others found from online reliable sources.

Chapter 1

Introduction

1.1 Forward and Futures Markets

Over 90% of electricity[1] trades occur in advance of delivery between generators and suppliers on what's known as forward or futures markets. Forward and futures markets allow participants to **contract a price today** for electricity that will be delivered at some point in **the future**. These contracts come in all different shapes and sizes.

But what's the point in contracting to sell electricity now that won't actually move through wires and cables for another, let's say, two years? Well, these contracts essentially **reduce risk**. Generators get the **security** of knowing that someone will **buy their electricity**, and suppliers have the **security** of knowing that they'll have some of the electricity they need to **meet their customers' demands**. And by locking in the price way ahead of delivery, generators and suppliers reduce their exposure to continually changing and often volatile spot market prices.

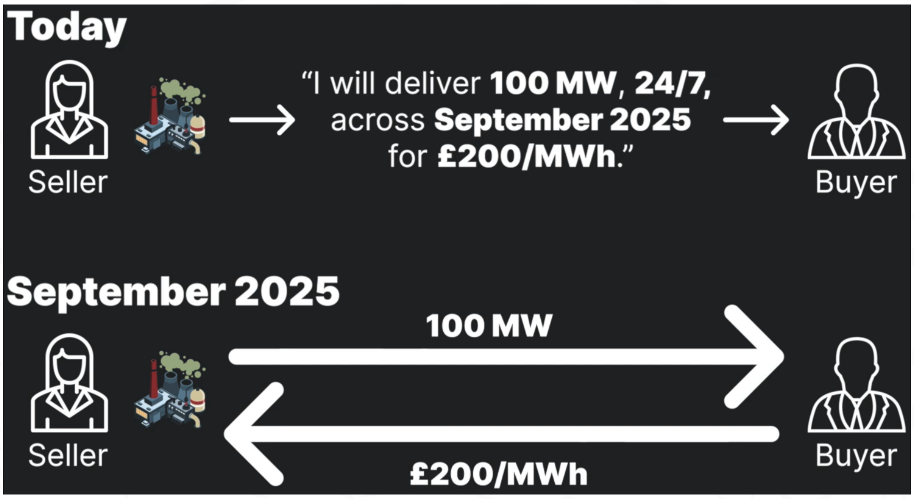
But how do they actually work? Let's start by outlining the difference between forward and futures trades. These are often mixed up due to their similar nature and outcomes of the markets, but there are some important distinctions to be made. While both **forward** and **futures** include an **agreement to deliver** energy at some point in the future, they differ based on how **standardized** the contracts are.

So let's start with forward trades. These are the trades with less standardized contracts. A forward contract is simply a contract between two parties who agree to buy and sell electricity at an **agreed price**, on a **set date**. This is sometimes referred to as "over-the-counter trading". Once agreed, the price is locked in. So it doesn't

[1]We use the example of the electricity market just to explain the different types of contracts

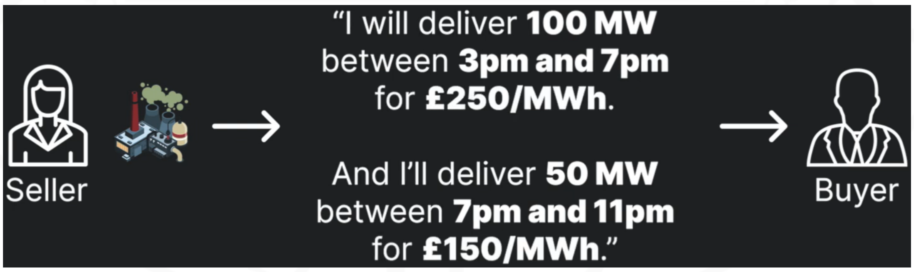
matter how the price changes between agreement and delivery, or what the price is on the day of delivery. Then once the date arrives the agreed volume of electricity is delivered for the pre-agreed price. This form of trading allows the customization of certain contract terms such as contract duration and delivery period. Let's set out an example. Here's what a forward contract might look like:

Figure 1.1: Example of Forward Contract n°1



The seller or generator agrees to deliver 100 megawatts of electricity 24/7 across the whole month of, say, September 2025, and is paid 200 pounds per megawatt hour to do so by the buyer or supplier. This continuous generation of consistent capacity across 24 hours is known as baseload. When September 2025 rolls around, the seller generates 100 megawatts of electricity at all times, and is paid to the agreed 200 pounds per megawatt hour to do so. Alternatively:

Figure 1.2: Example of Forward Contract n°2



The seller may agree to deliver the energy at more specific granular times. Instead of delivering "baseload" i.e. the same amount of electricity across 24 hours, the seller may agree to deliver certain volumes at certain times. As an example, he may agree to deliver 100 megawatts during certain four hour periods, and then 50 megawatts during others, and agree to different prices for both.

But what do we mean by futures? Well, like forward contracts, futures trades are for electricity to be delivered at some point in the future. However, the contracts are more **standardized** and less **flexible**. That's because they're traded on an **exchange**. There are a variety of products that you can buy and sell on futures exchanges. For example on the **BORSA ITALIANA**:

Figure 1.3: Futures Products that can be traded in Italy

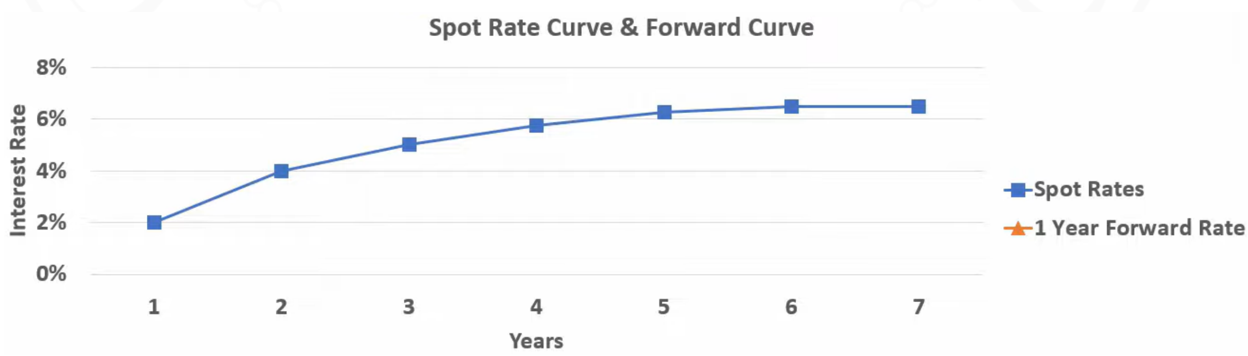
Characteristic	Description
Underlying	**Baseload**: one Megawatt (MW) supplied during the delivery month. **Peakload**: one Megawatt (MW) supplied from 8.00am to 8.00pm CET during the delivery month, excluding Saturday and Sunday

There are two futures products that can be traded in Italy. The first is Italy base Electricity Futures. This is electricity traded 24 hours a day, so baseload, seven days a week. The other is Italy Peak Electricity Futures. This is electricity traded during peak hours from 8:00 AM to 8:00 PM, Monday to Friday. In both cases, delivery is required across a **full month**, and can be traded as far as five years in advance of delivery. Generally, the further in advance that something is traded, the longer the agreed delivery period.

1.2 The Spot Curve and Forward Curve

Figure 1.4: Spot Rates for each Year

Time (in Years)	1	2	3	4	5	6	7
Spot Rates	2%	4%	5%	5.75%	6.25%	6.50%	6.50%



To understand the *Constructing of Forward Curves of Futures Prices* first we have to talk about the Spot Rate yield curve, we've got an example above where we can see the **Time** in years (one year in the future, two years in the future, three years...), below that we got the **Spot Rates**. In the discussion about Spot Rates, the time in years is plotted against the Spot Rates, for instance, a 2% Spot Rate indicates that if we purchase a thousand euros bond today, it would yield a 2% interest, resulting in a €20 coupon payment plus the principal at the end of the year. The pattern continues for various maturities, up to seven years:

Figure 1.5: 1 Year Forward Rate

Time (in Years)	1	2	3	4	5	6	7
Spot Rates	2%	4%	5%	5.75%	6.25%	6.50%	6.50%
1 Year Forward Rate	2.00%	6.04%	7.03%	8.03%	8.27%	7.76%	6.50%

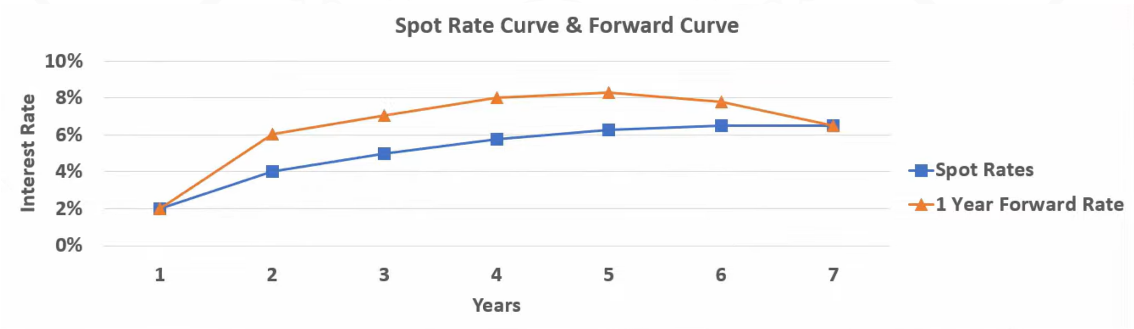
Now, turning our attention to **Forward Rates** (*this is the rate that makes us indifferent*) an **arbitrage** principle between holding a one-year bond and entering a one-year forward contract starting a year from now versus acquiring a two-year bond. So let's take a look at what this looks like:

Time (in Years)	1	2	3	4	5	6	7
Spot Rates	2%	4%	5%	5.75%	6.25%	6.50%	6.50%
1 Year Forward Rate	2.00%	6.04%	7.03%	8.03%	8.27%	7.76%	6.50%

The first one-year Forward Rate is initially equal to the Spot Rate, but subsequent Forward Rates can be calculated using the formula:

$$ForwardRate = \frac{(1+s(t))^t}{(1+s(t-1))^{(t-1)}} - 1$$

Basically the Forward Rate is equal to 1 plus that year's Spot Rate discounted by that many years into the future, divided by 1 plus the previous year Spot Rate, discounted by the number of years into the future the previous year was and at the end we just subtract by 1. That gives us the one year spot rate that ends two years from now. Calculating Forward Rates up to year seven:



We observe an interesting trend. While the spot rate curve is ascending, the forward rate curve increases at a higher rate. However, after the spot rate curve plateaus, the forward rate curve starts to decline.

1.3 Using Curves

Constructing Forward Curves is a critical process in evaluating market-exposed assets. These curves establish the prices at which future delivery contracts can be settled today, providing valuable insights into expected price movements. Not only do they reflect the market's collective expectations regarding future prices or rates, but they also play a pivotal role in risk management, enabling traders to hedge against potential price fluctuations.

In the construction of forward curves, various **interpolation** methods are employed to estimate missing values along interest rate or financial product price curves. The choice of interpolation method is crucial, considering factors such as continuity, stability, and coherence. Constant interpolation is a straightforward approach, while linear interpolation offers continuity but may lack differentiability. Cubic spline interpolation provides greater flexibility but may lead to oscillatory behavior. Polynomial interpolation, though unstable with large datasets, is another option.

After interpolation, the curves for discounting and forwarding are set using two key methods: Bootstrapping and Optimization. Bootstrapping involves generating interest rate curves using market data organized by maturity, starting with shorter durations and progressively developing the curve. Optimization uses pricing algorithms and curve yield data to ensure consistency with market rates. Both methods may require additional data, and interpolation methods are employed accordingly.

The construction of interest rate curves, particularly through methodologies like bootstrapping or calibration, necessitates consideration of the **Turn-Of-Year** (TOY) effect. This phenomenon, characterized by spikes in rates at specific times of the year, requires a method to estimate its impact. Typically, this involves temporarily excluding the TOY effect from initial quotes and reintroducing it into the modeling process to ensure accurate representation.

This comprehensive approach to constructing forward curves, blending the considerations of interpolation methods, curve construction methodologies, and the impact of the TOY effect, ensures a thorough and accurate representation of market expectations and dynamics.

Chapter 2

Data Description

Crude oil futures represent standardized contracts for the future delivery of a specific quantity and quality of crude oil. In this context, the crude oil futures we are referring to, have expiration dates ranging from February 2024 to January 2025. These contracts are an important part of the global commodities market allowing participants to hedge against price volatility or speculate on future price movements.

The data on these crude oil futures comes from **Refinitiv**, a well-known financial data provider known for delivering accurate and up-to-date information to financial professionals around the world.

The specific details on crude oil futures, including pricing, trading volumes and historical performance, are derived from the *New York Mercantile Exchange* (**Nymex**). Nymex is a leading commodities exchange providing a platform for trading energy, metals and other commodities contracts. As a key player in the derivatives market, Nymex facilitates trading in crude oil futures allowing market participants to price and manage risk. The information we got from Nymex is critical as it enables us to make informed decisions based on **real-time** and **historical data**.

The **diversity** of participants in the crude oil futures market, from producers and consumers to speculators and institutional investors, contributes to the overall liquidity and efficiency of the market. The different expiration dates of futures contracts from February 2024 to January 2025 provide market participants with the flexibility to align their trading strategies with specific **time horizons**, taking into account factors such as geopolitical events, economic indicators and supply and demand dynamics.

Through, the combination of Refinitiv's data accuracy and Nymex's market infrastructure we were able to present the following dataset:

Figure 2.1: Dataset - March 2024

Date	Open	High	Low	Close
08/1/2024	73.6	73.98	72.08	72.18
05/1/2024	72.51	74.3	72.35	73.86
04/1/2024	73.2	74.12	71.24	72.36
03/1/2024	70.77	73.36	69.56	72.89
02/1/2024	71.9	73.8	70.32	70.62
29/12/2023	72.15	72.8	71.47	71.84
28/12/2023	74.04	74.62	71.92	71.97
27/12/2023	75.57	75.81	74.01	74.34
26/12/2023	73.74	76.31	73.32	75.71
22/12/2023	74.07	75.11	73.55	73.73

This dataset contains historical data on crude oil futures from September 9, 2023, to February 8, 2024. *"The dataset taken into account is just one of twelve to show how it is structured"*. The data includes daily opening, high, low, and closing prices of crude oil futures contracts (*the models that we provide later solely focus on the closing prices of crude oil futures contracts*). Each row in the dataset corresponds to a specific trading day, and the columns represent the **Date** (date of the trading day), **Open** price (opening price of crude oil futures on that day), **High** price (highest price reached during the trading day), **Low** price (lowest price reached during the trading day), and **Close** price (closing price of crude oil futures on that day), respectively:

- **Format**: the data is presented in a tabular format with columns for the variables and rows for the observations.
- **Granularity** (refers to the level of detail or precision of the data): daily prices are provided, capturing the fluctuations in crude oil futures on a day-to-day basis.

Chapter 3

The Models

3.1 Polynomial Approximation

3.1.1 The Model

In our model we implemented at first the approximation via **polynomials**, where the deviation vector r, i.e. *the vector of the squares of the distance between the value given at a point and the value of the approximating polynomial at that point*, trying to find a polynomial (generally of low degree) that approximates the data points with a **minimum error**.

The **least square** approach is used to approximate solutions of **overdetermined** systems of equations, i.e., systems where the number of equations is greater than the number of unknowns: $Ax = b, A \in R\ m{\times}n\ , b \in R\ m, m > n$. In general, it is impossible to find an exact solution of the linear system above (i.e., an x such that $b - Ax = 0$), but we can look for an approximate solution x such that $b - Ax \approx 0$. The name "least squares" means that the solution minimizes the sum of the squares of the errors made in every single equation. In data fitting, the best fit in the least square sense minimizes the sum of the squares of the **residuals**, each residual being the difference between the observed value and the value provided by the model used.

For the Python implementation we used the following function from the module *numpy* for the esteem of the model:

```
numpy.polyfit(x, y, deg, rcond=None, full=False, w=None, cov=False)
```

3.1.2 Least Squares Polynomial Fit

Fit a polynomial $p(x) = p[0] * x^{deg+\ldots+p[deg]}$ of degree deg to points (x, y). Returns a vector of coefficients p that minimizes the squared error in the order $deg, deg-1, \ldots 0$:

```
numpy.polyval(p, x)
```

Evaluate a polynomial at specific values. If p is of length N, this function returns the value:

$$p[0] * x^{(N-1)} + p[1] * x^{(N-2)} + \ldots + p[N-2] * x + p[N-1]$$

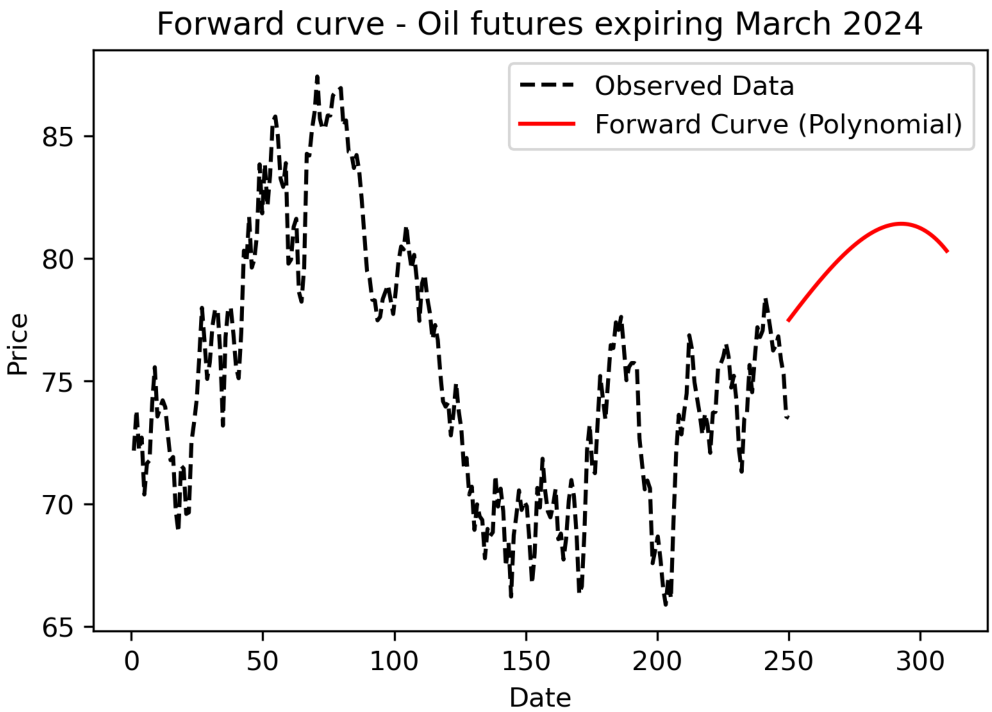
If x is a sequence, then $p(x)$ is returned for each element of x. If x is another polynomial then the composite polynomial $p(x(t))$ is returned.

3.1.3 Choice of the Degree for the Polynomial

Model selection is the task of selecting a model from among various candidates, on the basis of performance criterion, to choose the one which best fits the data. We tried some of the most popular selection criteria for the model like **AIC** (Akaike information criterion). AIC balances the **goodness of fit** of the model (represented by the likelihood function) with the complexity of the model (represented by the number of parameters). It penalizes models with more parameters to prevent overfitting. We also tried **BIC** (Bayesian information criterion), that uses the same assumptions of AIC but with a different approach.

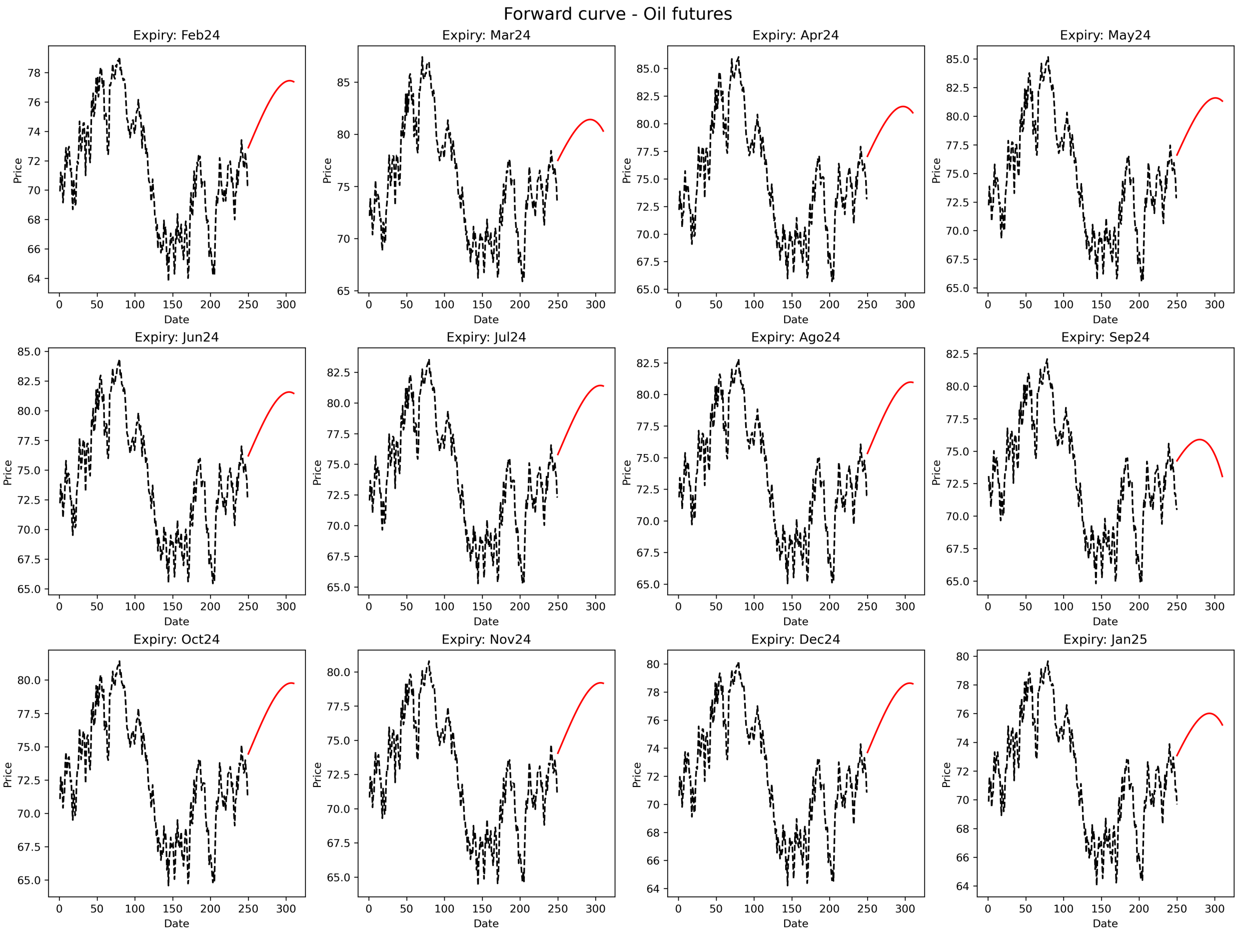
The best degree that results from both tests is 9.But, we decided to discard 9 following the *Occam's razor*, which states: *"presented with competing hypotheses about the same prediction and both theories have equal explanatory power one should prefer the hypothesis that requires the fewest assumptions"*. So, basing our model on this principle, we found that an effective degree for the polynomial is 4 because it assumes the most **reality-like trend** in the interval examined of 60 days. The interval is kept short because the limited amount of data available on Refinitiv would have reduced the effectiveness of longer-term predictions.

Figure 3.1: Forward Curve - March 2024

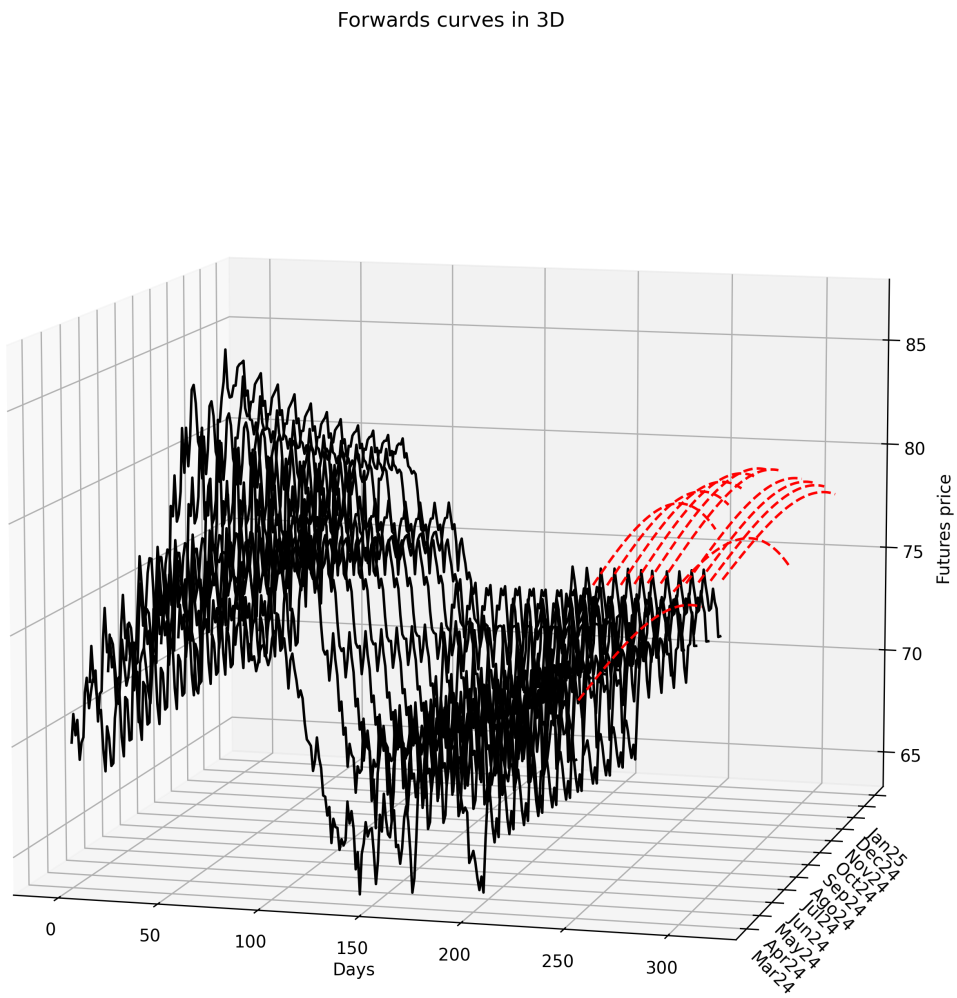


The graphical representation provided in our analysis showcases a comprehensive comparison between observed data and predictions. The observed data are delineated by a black dashed line, while the predictions are illustrated with a continuous red line.

Figure 3.2: Forward Curves - 12 Months



This visual presentation is structured into a grid format of 4 rows by 4 columns, each representing one of the twelve months in two dimensions. The grid layout allows for easy comparison of the trends across different months, facilitating a nuanced understanding of the data dynamics.



This representation offers a three-dimensional perspective, where all observations and the forward curve are positioned one near each other, enhancing the clarity of comparison throughout the entire year. This multidimensional view provides a more holistic understanding of the data trends.

Upon closer examination, it becomes evident that the trends across most months exhibit a **similar trajectory**, with notable exceptions observed in **September**[1]. While the overall trend in the predicted data shows a growth disposition over the examined time horizon, there is a discernible attenuation in intensity towards the end of the two predicted months.

[1]The divergence in trend observed in the month of September could be attributed to seasonal demand patterns: September often marks a transition period between seasons, which can significantly influence the consumption of energy sources like crude oil, primarily utilized for combustion purposes.

3.2 GBM Model

The **Geometric Brownian Motion (GBM)** is a mathematical model that describes the evolution of a variable, such as the price of a financial instrument, over time. It is widely used in finance, particularly for modelling the price movements of stocks, futures, and other assets. When applied to the price of futures, it helps in simulating the random fluctuations and trends observed in financial markets. Geometric Brownian motion is employed to model futures prices and it is most widely used for futures price behaviour.

Some arguments favoring the use of Geometric Brownian Motion (GBM) to model futures prices include its independence of expected returns from the process value, mirroring real-world expectations. Moreover, GBM only assumes positive values akin to actual futures prices. Its path exhibits a comparable level of *'roughness'* as seen in real futures prices, while calculations with GBM processes are relatively straightforward.

However, GBM lacks complete realism in modeling futures prices, primarily due to two limitations. Firstly, while **volatility** in real futures prices often fluctuates over time, possibly in a stochastic manner, GBM assumes **constant** volatility. Secondly, real-world futures prices frequently experience abrupt shifts or **"jumps"** caused by unforeseeable events or news, a feature absent in GBM paths, which remain continuous throughout.
Composition of GBM:

- 1. Stochastic Process:
  - GBM is a stochastic (random) process that describes the **continuous-time evolution** of a variable, often denoted as $S(t)$, where t is **time**.
  - It is characterized by two main components: a deterministic drift (μ) and a stochastic diffusion (σ).
- 2. Drift and Diffusion:
  - Drift (μ): Represents the average rate of return per unit of time. It determines the overall **trend** in the price.
  - Diffusion (σ): Represents the volatility or randomness in the price movement. It accounts for the unpredictable **fluctuations**.
- 3. Stochastic Differential Equation (SDE):

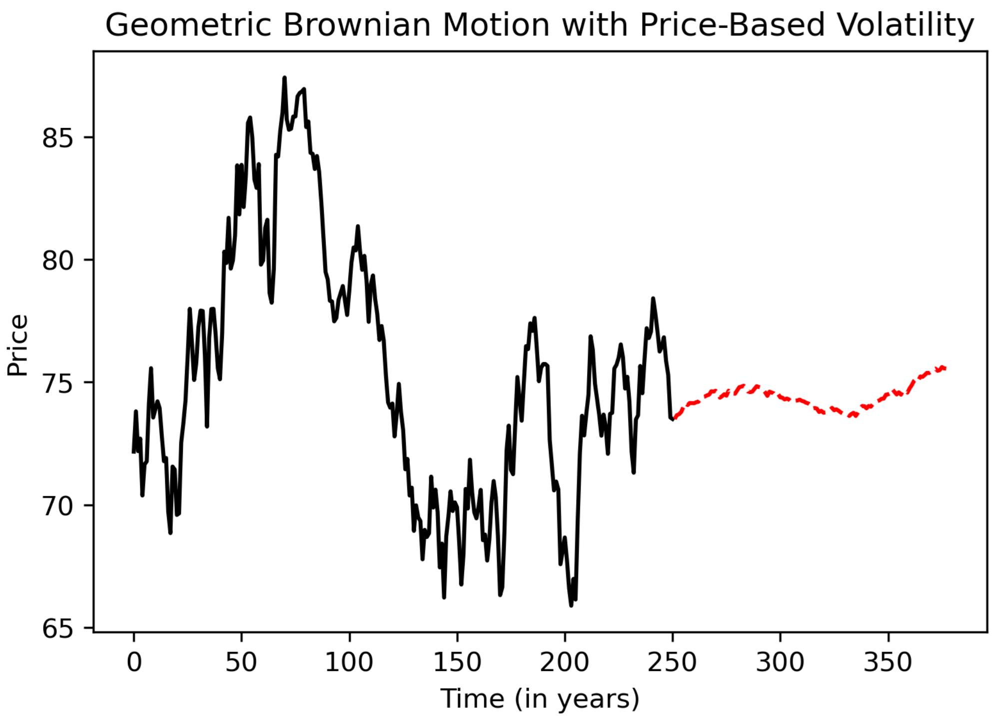
– The dynamics of GBM are described by the following **stochastic differential equation**:

$$dS(t) = (\mu)S(t)dt + (\sigma)S(t)dW(t)$$

where $dW(t)$ is a Brownian motion representing the random component.

Stochastic Differential Equations ($SDEs$) do have analogies with both Ordinary Differential Equations ($ODEs$) and Partial Differential Equations ($PDEs$), but there are some important distinctions:

- $ODEs$ are used to model deterministic systems with no inherent randomness, whereas $SDEs$ incorporate stochastic components to model random processes.
- $PDEs$ often involve spatial dimensions and describe systems evolving in both space and time. $SDEs$ primarily model processes evolving in time.
- While $ODEs$ and $PDEs$ are typically solved for a unique solution given initial or boundary conditions, $SDEs$ often provide a distribution of possible paths due to the stochastic component.



In the graph above, the black line represents the **historical price data**, while the dashed red line represents the **prices predicted** using the geometric Brownian motion (GBM) model. Since the GBM model assumes that historical volatility stays constant over time, this feature is reflected in the graph. Indeed, we observe that prices estimated by the GBM model stay relatively constant (75 euros approximately).

3.3 Monte Carlo Simulation

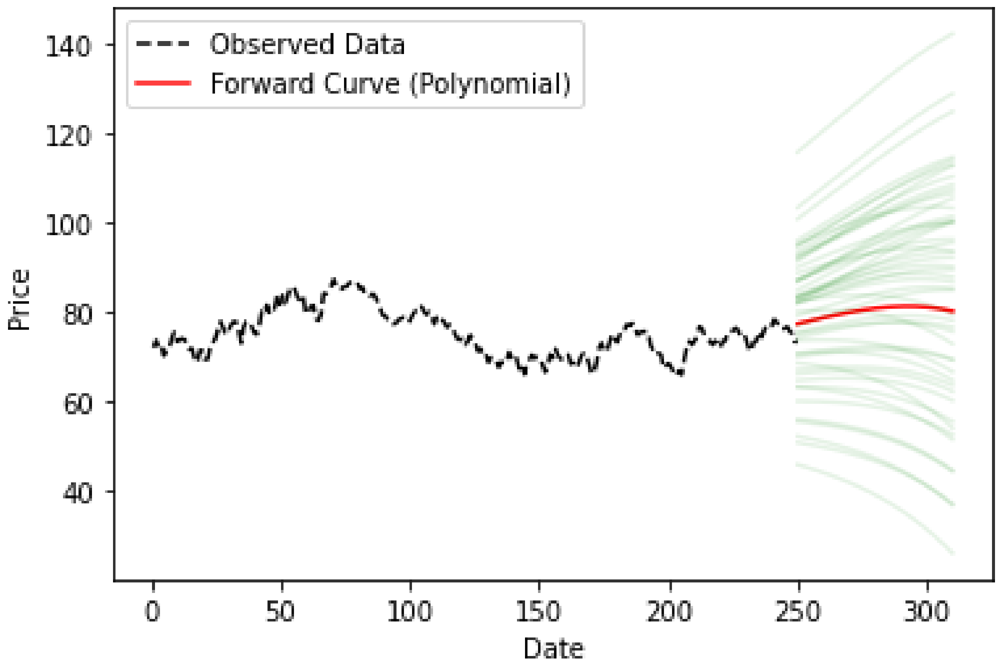
Monte Carlo simulation, also known as the *Monte Carlo method* or *multiple probability simulation*, is a mathematical technique that is used to estimate the possible outcomes of an **uncertain event**. It was named after a well-known casino city, Monaco, as the element of chance is the heart of the modeling approach, similar to a roulette game.

During the years Monte Carlo simulations have evaluated the impact of risk in many real-life scenarios, such as sales forecasting, project management and pricing, as we are dealing with in our project. They also provide several advantages over **predictive models** with fixed inputs, such as being able to conduct a **sensitivity analysis** or calculate input correlation.

Sensitivity analysis allows decision makers to see the impact of individual inputs on a specific outcome. Financial analysts often make **long-term predictions** about prices and then advise their clients on appropriate strategies. In doing so, they must consider market factors that could cause drastic changes in the value of the investment. Accordingly, using Monte Carlo simulation to predict likely outcomes to support their strategies.

In our script, in the Montecarlo function's part, we started with the assumption of **volatility**, which is implemented by **standardizing** the logarithm of the price at *date* i minus the price at *date* $i-1$. This simple esteem of volatility is often used in finance and influences the perturbations computed in Montecarlo's function, making use of a **normal distribution** with *mean* equal to 0 and *variance* equal to the historical volatiliy. We preferred not to wield *"a random variance"* because volatility is a more consistent parameter and small variation in a random variance drives to enormous differences in outputs.

Figure 3.3: Monte Carlo Method - March 2024



We have tried 50 **simulations**, which usually are not suitable for simulations but give us a clearer graph. We can see that unless using a consistent parameter as volatility, in doing such a small amount of simulations, results are so diverse and a small perturbation could deeply influence the forward curve, causing a **wide spread** between the higher and the lower curve.

The basic principle of Monte Carlo simulation lies in **ergodicity**, which describes the statistical behavior of a moving point in a closed system. In an ergodic system the moving point will eventually pass through every possible position. This becomes the basis of Monte Carlo simulation, in which the computer runs enough of them to produce the eventual outcome of several inputs.

Chapter 4

Empirical Results

4.1 Numerical Results

To compare the results of the forecasts made by our model, we can look at the data provided by the *"Oil Price Forecast By Day"* and *"WTI Oil Price Forecast By Day"*. Sources we will refer to as **"OIL"** to indicate **"Oil Price Forecast By Day"** data and **"WTI"** to indicate **"WTI Oil Price Forecast By Day"** data.
Link to the website: `https://30rates.com/oil-price-forecast-and-predictions`

Brent Crude and **West Texas Intermediate** (WTI) are the two primary types of oil in the global market. Brent serves as the standard for light crude oil in Europe, Africa, and the Middle East, while WTI is the **benchmark** for the United States. Despite Brent being the more commonly used benchmark for oil prices, WTI is gaining noteworthiness. Brent is typically sourced from the North Sea near Europe, which reduces transportation costs, whereas WTI originates from inland areas of the United States, mainly Texas. Both are classified as **light sweet crude oils**, traded on U.S. exchanges, and exhibit correlated prices, indicating they generally move together. In the last years, WTI's price falls below Brent's due to geopolitical factors and increased U.S. oil production. Moreover, during periods of political turmoil, Brent's price tends to rise more rapidly than WTI. These **distinctions** illustrate the intricate dynamics of the global oil market. Consequently, while our data primarily aligns with Brent, we've opted to include both Brent Crude and West Texas Intermediate (WTI) for comparison in our estimates.

These sources are specialized in predicting oil prices and provide **daily** estimates based on a range of factors such as demand and supply, geopolitical conditions, and more. The provided data resemble **estimates** which include minimum, maximum, and average prices, where the average price represents the mean between the minimum and maximum expected price for each day (where the price itself is an average of the maximum and minimum prices that can occur day by day).

WTI, means West Texas Intermediate, it is a type of crude oil mainly extracted in Texas and is recognized as one of the key global indicators of oil prices. WTI price forecasts, along with generic oil price data, give us the possibility to understand the **trend** of the oil market movements.

However, it is important to note that the data provided by these sources only cover a **short period**, typically 24 days. With respect to our polynomial model that can estimate oil prices for a much longer period, such as 60 days, or the GBM for half year.

As a consequence of this consideration, the mean, the standard deviation, the root mean square error and other statistics, are all consider for a time period of 25 days in order to standardize the comparison.

Another remark to take into account is that these data, *gathered from two different sources*, estimate the data for 24 days, starting from the next day, without saving the historical series of the data estimated. Following this information, which was not stated on the website were the data was taken, estimated prices are missing for the first few days of February. Indeed, if the models make an estimate starting from February 1st, the prices of "OIL" and "WTI" begin from February 5th since they were downloaded on February 3rd.

The following table reports statistics such as the mean, the standard deviation and the root mean square error of the models and the estimated prices. *(model with 25 days estimate)*:

Data	Mean	Dev.Standard	Root Mean Square Error
Observed Data	75.36	5.17	
Polynomial Data	79.13	0.90	45.08
"OIL Forecast" data	76.12	2.05	23.05
"WTI OIL Forecast" data	71.11	2.02	8.93
GBM Estimated Data	74.40	0.49	

The estimated average values from the Polynomial model and GBM model differ slightly from the observed average data. The Polynomial model yields an average of approximately 79.13, slightly higher than the observed average of 75.36, while the GBM model's average closely matches the observed data. However, when considering the average from both sources, we observe a slightly lower decrease for crude oil, resulting in an average of 71.11 compared to the observed data. Conversely, for "OIL," the difference from the observed data is not significant.

The standard deviation of the estimated data decreases notably for both the Polynomial and GBM models. In details, it is approximately 0.9 for the Polynomial model and 0.49 for the GBM model, compared to an observed value of around 5. This decrease suggests that the forecasts exhibit less variability compared to the observed data. A lower standard deviation indicates higher precision in the forecasts, as the data's variability is **constrained**. While this could imply high forecast quality, it's worth noting that the limited availability of data for the forecasts might also contribute to this low standard deviation. Additionally, both "Oil" and "WTI" exhibit similar standard deviations of about 2, akin to the lowest values observed in the models.

The Mean Squared Error (MSE) serves as a metric for gauging the disparity between the model's predicted values and the actual observed data. Typically, a lower MSE indicates a more accurate prediction of future prices. However, in this instance, the MSE is remarkably high. Specifically, for the Polynomial model, the MSE is around 45.08, considerably higher but slightly lower for "OIL" at approximately 23, and the lowest is represented by "WTI" with a value of 8.

Examining the next graphs below, both the GBM and polynomial models, along with the estimated data from the sources, offer insight into potential future price growth. To further examine this trend, let's compute the average price growth using the following formula:

$$growthp\_percentage = (final\_price - initial\_price)/initial\_price * 100$$

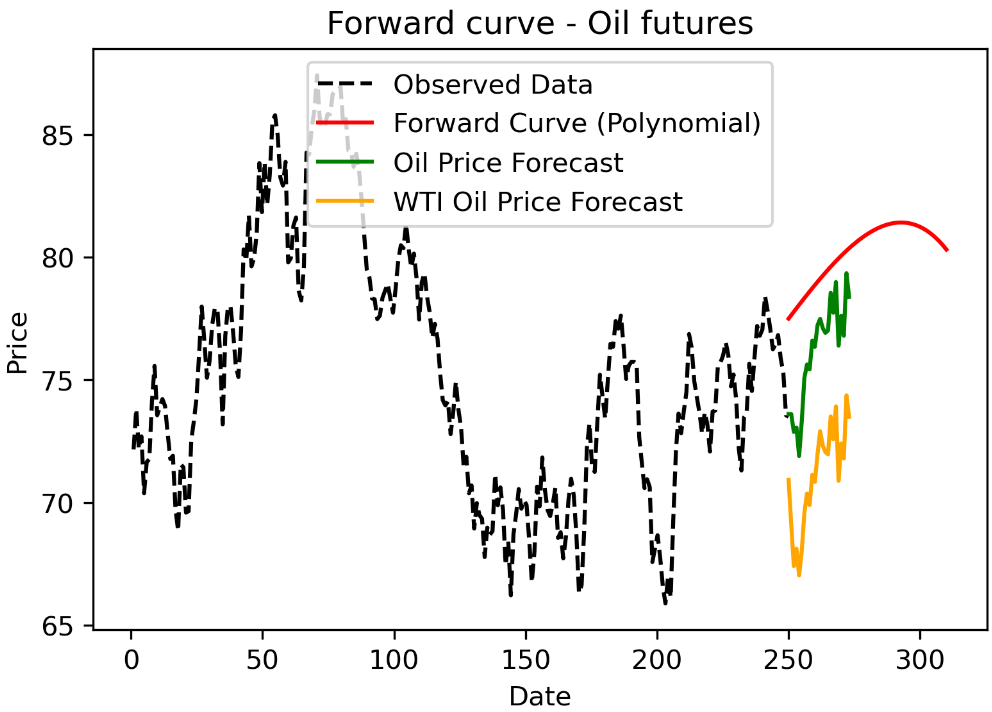
Notably, the polynomial model suggests a growth of 3.87%, considering an initial price of 77.5 and a final price of 80.5. On the other hand, the GBM model shows a growth of 1.57% when evaluating future data at 25 days, with a slightly lower growth of 1.42% over a longer period of 60 days.

Upon analyzing the data provided by the sources, it becomes clear that they predict the highest growth rates. In particular, the predicted price of oil exhibits an estimated growth of 6.49%, while the predicted price of WTI crude oil shows an estimated growth of 3.62%.

4.2 Graphical Results

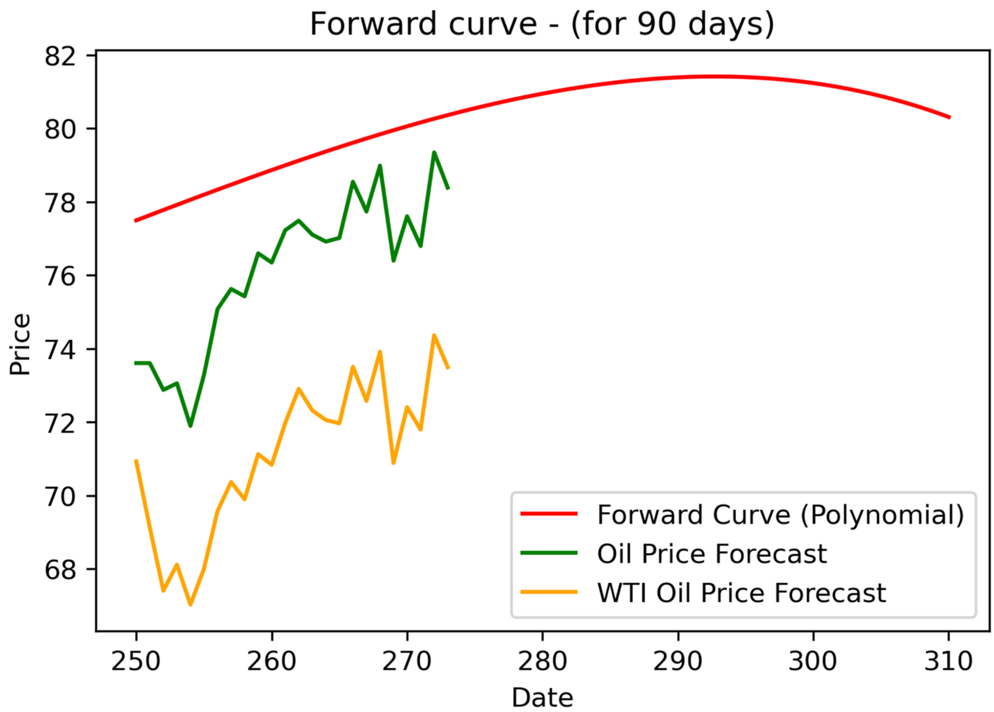
The following graphs represent a series of graphical analyses that have been conducted to provide a better understanding of the topic. They consist of **observed data** represented in black, the **polynomial curve** shown in red, while the **future prices** of "OIL" are depicted in green and the **prices** of "WTI" are shown in orange.

Figure 4.1: Observed data, polynomial model estimated at 90 days, prices "OIL" and "WTI"



We have a general representation that allows us to see at first how the three forecasts seem to present very different predictions from each other, although all three are increasing. For this reason, we have implemented the next two graphs in such a way to display only the forecasts.

Figure 4.2: Polynomial model estimated at 90 days, prices of "OIL" and "WTI"



It considers all price forecasts of the polynomial model up to 90 days, while the graph below considers only a forecast for 25 days. We can see how the polynomial model predicts a more constant increase in prices up to approximately 60 days, followed by a slight decrease towards the end of this period. Regarding "OIL" and "WTI," it is interesting to note, despite the prices being different, they follow a very similar trend. Indeed, prices rise and fall more or less in the same period and exhibit a volatility very much alike to the observed data.

Figure 4.3: Polynomial model estimated at 25 days, prices of "OIL" and "WTI"

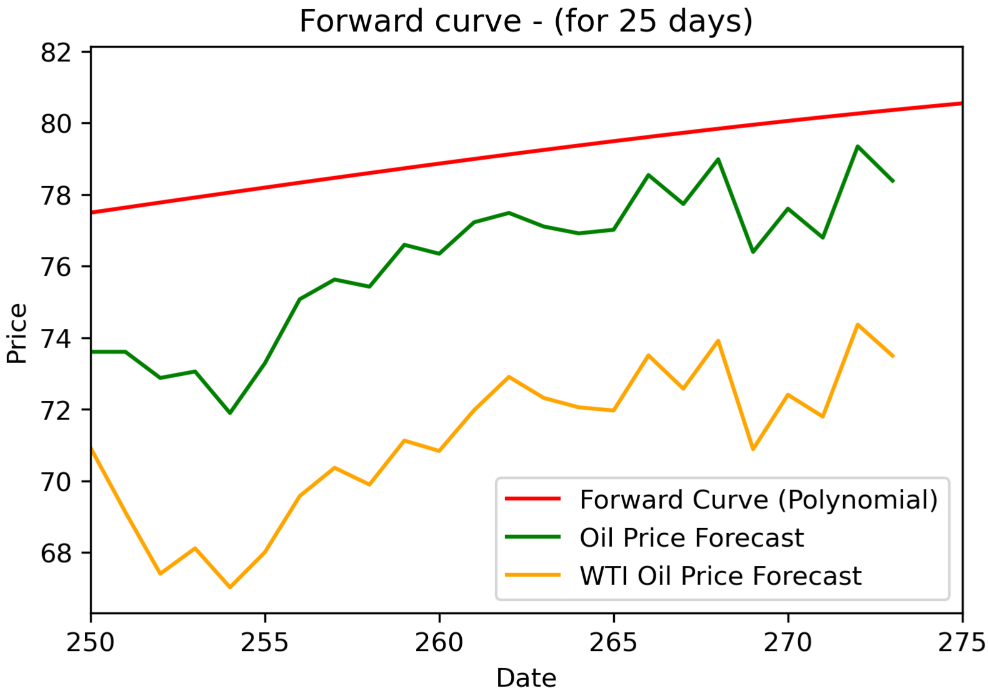
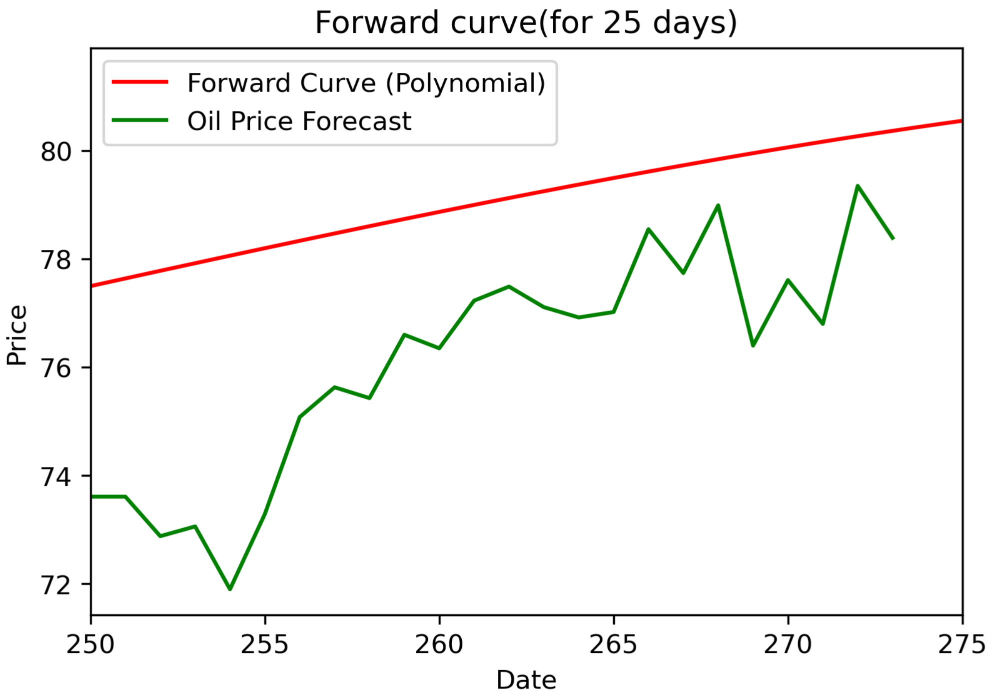


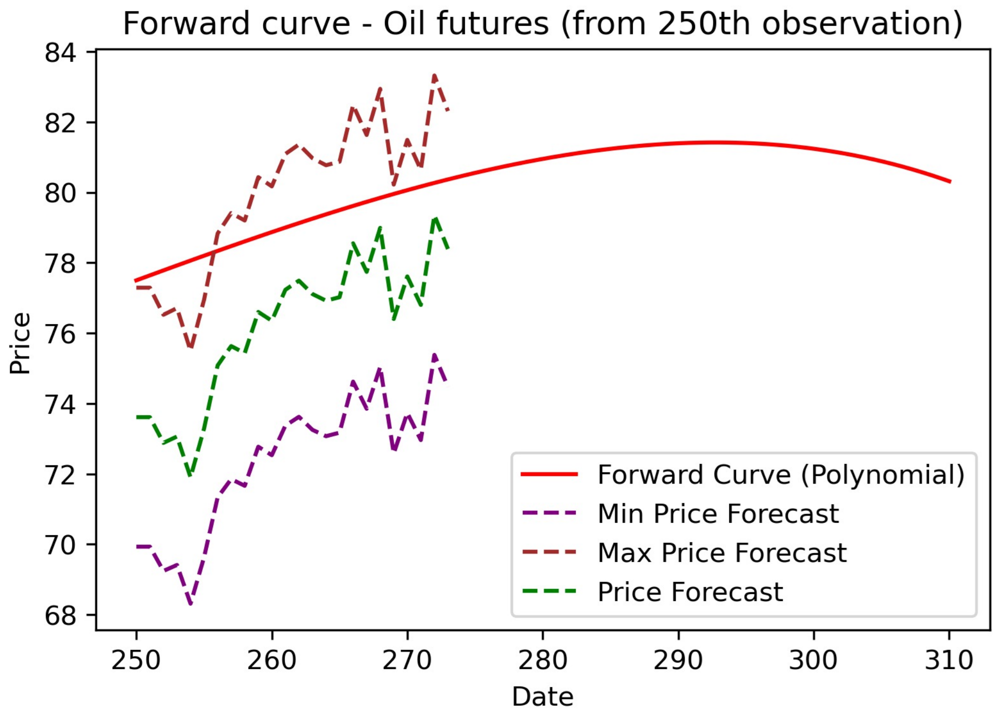
Figure 4.4: Comparison of only "OIL" and the polynomial model estimated at 25 days



Here the estimated prices from the polynomial model and those of "OIL" are the only ones presented. This choice was made to visually highlight the difference between the two models. Considering that the prices of "WTI" are regarded as the crude oil prices in the USA, while Brent better fits our data due to its proximity to the same geographical area.

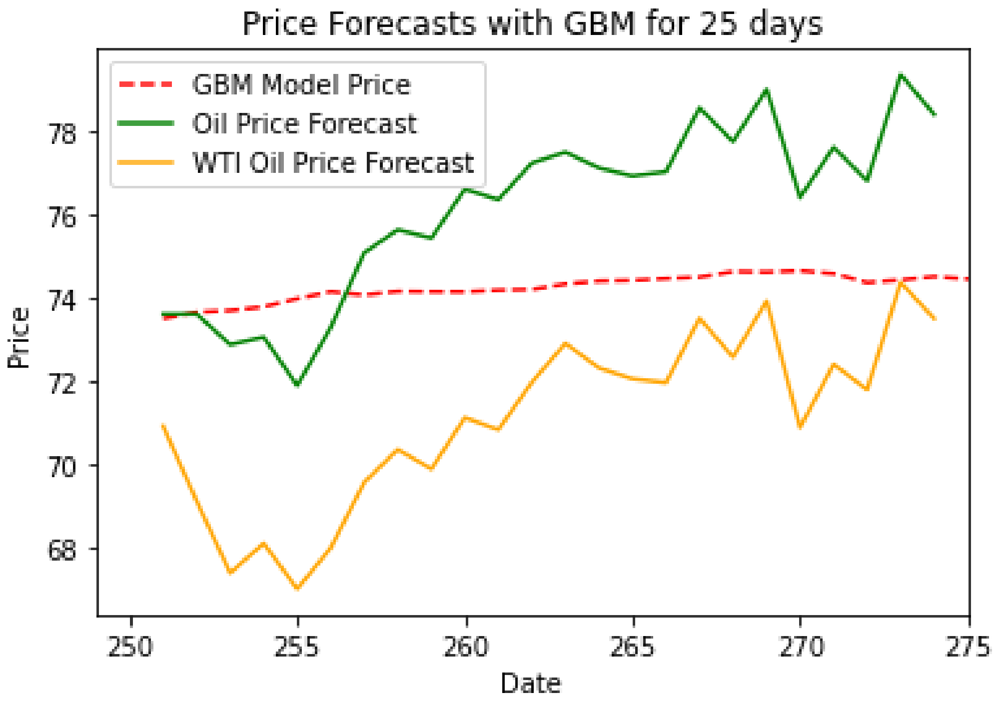
Upon analyzing the graph, it becomes crystal clear that the two datasets don't differ much. The mean from the polynomial data stands at 79.13, slightly lower than the mean of the "OIL Forecast" data, which registers at 76.12. However, the notable distinction lies in their growth rates. The polynomial model shows a growth of 3.87%, while the predicted oil price exhibits an estimated growth of 6.49% — almost double that of the former. This suggests that although the overall price levels are comparable, there's a significant **discrepancy** in the rate of increase between the two models.

Figure 4.5: Polynomial model estimated at 90 days, comparison with main statistics of "OIL" prices.



We observe the estimated maximum prices day by day tend, on average, to exceed the prices estimated by the polynomial curve. However, these stand higher than the estimated minimum and average prices.

Figure 4.6: Comparison of the GBM model with prices of 'OIL' and 'WTI' for 25 days



The prices of "OIL" and "WTI" are again represented, but this time compared with the GBM model. It is noticeable that the GBM model tends to have a much more stable trend compared to the prices of "OIL" and "WTI", which exhibit a more **volatile pattern**.

From these graphs, it emerges that the GBM model generally positions itself between the price estimates of "OIL" and "WTI", while the polynomial model tends to show higher price estimates overall, with the only exception being the maximum price of "OIL". The polynomial model provides higher average price estimates, which stand stable and unaffected by fluctuations. This stability can be attributed in part to the construction of the models, with the GBM being frequently used and better suited for estimating prices of financial instruments.

4.3 DATA: (Downloaded on February 3rd)

4.3.1 Oil Price Forecast By Day

Date	Weekday	Min	Max	Price
5-Feb	Monday	69.93	77.29	73.61
6-Feb	Tuesday	69.93	77.29	73.61
7-Feb	Wednesday	69.24	76.52	72.88
8-Feb	Thursday	69.41	76.71	73.06
9-Feb	Friday	68.31	75.5	71.9
12-Feb	Monday	69.63	76.95	73.29
13-Feb	Tuesday	71.33	78.83	75.08
14-Feb	Wednesday	71.85	79.41	75.63
15-Feb	Thursday	71.66	79.2	75.43
16-Feb	Friday	72.77	80.43	76.6
19-Feb	Monday	72.53	80.17	76.35
20-Feb	Tuesday	73.37	81.09	77.23
21-Feb	Wednesday	73.62	81.36	77.49
22-Feb	Thursday	73.25	80.97	77.11
23-Feb	Friday	73.07	80.77	76.92
26-Feb	Monday	73.17	80.87	77.02
27-Feb	Tuesday	74.62	82.48	78.55
28-Feb	Wednesday	73.85	81.63	77.74
29-Feb	Thursday	75.04	82.94	78.99
1-Mar	Friday	72.58	80.22	76.4
4-Mar	Monday	73.73	81.49	77.61
5-Mar	Tuesday	72.96	80.64	76.8
6-Mar	Wednesday	75.38	83.32	79.35
7-Mar	Thursday	74.47	82.31	78.39

4.3.2 WTI Oil Price Forecast By Day

Date	Weekday	Min	Max	Price
5-Feb	Monday	67.38	74.48	70.93
6-Feb	Tuesday	65.67	72.59	69.13
7-Feb	Wednesday	64.04	70.78	67.41
8-Feb	Thursday	64.71	71.53	68.12
9-Feb	Friday	63.68	70.38	67.03
12-Feb	Monday	64.61	71.41	68.01
13-Feb	Tuesday	66.1	73.06	69.58
14-Feb	Wednesday	66.85	73.89	70.37
15-Feb	Thursday	66.41	73.4	69.9
16-Feb	Friday	67.57	74.69	71.13
19-Feb	Monday	67.3	74.38	70.84
20-Feb	Tuesday	68.38	75.58	71.98
21-Feb	Wednesday	69.26	76.56	72.91
22-Feb	Thursday	68.7	75.94	72.32
23-Feb	Friday	68.46	75.66	72.06
26-Feb	Monday	68.37	75.57	71.97
27-Feb	Tuesday	69.83	77.19	73.51
28-Feb	Wednesday	68.95	76.21	72.58
29-Feb	Thursday	70.22	77.62	73.92
1-Mar	Friday	67.35	74.43	70.89
4-Mar	Monday	68.79	76.03	72.41
5-Mar	Tuesday	68.21	75.39	71.8
6-Mar	Wednesday	70.65	78.09	74.37
7-Mar	Thursday	69.83	77.18	73.5

Chapter 5

Conclusions

After thoroughly examining the fundamentals of forward curves and collecting pertinent data on crude oil, we proceeded to construct these curves using a polynomial model. This approximation involves seeking a fourth-degree polynomial to minimize the error concerning observed data. The choice of a fourth-degree polynomial was motivated by its perceived superior adaptability to the trend in our data. Utilizing this model allowed us to estimate a 60-day period.

Subsequently, we conducted a Monte Carlo simulation with 50 iterations, generating diversified results attributable, in part, to the limited number of simulations, which could impact result variability. Concurrently, the Geometric Brownian Motion (GBM) model, a stochastic process widely recognized for its applicability in financial market studies, was successfully employed to construct forward curves extended up to 127 future days, this model is characterized by its ability to capture random movements in a continuous-time setting, making it particularly suitable for modelling the unpredictable fluctuations observed in financial markets.

The comparison between the polynomial model and the GBM, along with data regarding Brent and WTI prices, did not provide a clear indicator of which model is more suitable for estimations. While the polynomial model exhibits a steady growth with a higher average price and a growth rate of 3.87%, significantly lower than Brent's 6.49%, Brent's price hights appear to be superior. On the other hand, the GBM, while presenting a mean of estimated data closer to the observed data, shows a significantly lower percentage growth than Brent, though with greater, though limited, price fluctuations.

Both models show a significant reduction in standard deviation compared to observed data, suggesting greater precision in forecasts. However, the Mean Squared

Error (MSE) is high for both models, indicating a substantial discrepancy between predictions and observed data. The polynomial model predicts a growth of 3.87%, while the GBM model shows lower growth rates, both at 25 days (1.57%) and 60 days (1.42%). Both models exhibit inaccuracies in precisely predicting observed data, as highlighted by the high MSE in both cases. It is noteworthy that a lower standard deviation suggests greater precision, but it is crucial to consider the effect of limited data availability for forecasts.

Bibliography

Jorg Kienitz. *Interest Rate Derivatives Explained*, volume 1: Products and Markets. Palgrave Macmillan, 2014.

Fridthjof Ollmar Steen Koekebakker. Forward curve dynamics in the nordic electricity market. 31(6):73–90, 2005.

Steven Shreve. Stochastic calculus and finance, stochastic calculus for finance ii: Continuous-time models.

Fractional calculus and fractional processes with applications to financial economics theory and application.

John C. Hull. Options, futures, and other derivatives.

https://www.borsaitaliana.it/homepage/homepage.htm.

https://fastercapital.com/startup-topic/Crude-Oil-Trading.html.

https://fastercapital.com/content/The-Ultimate-Guide-to-Understanding-WTI-Crude-Oil-Prices.html.

https://www.fortrade.com/a/blog/brent-crude-vs-wti-crude-key-distinctions/.

https://www.schwab.com/learn/story/energy-investing-basics-wti-vs-brent-crude-oil.

https://tradingeconomics.com/commodity/crude-oil.

https://www.investopedia.com/ask/answers/052615/what-difference-between-brent-crude-and-west-texas-intermediate.asp.

https://ocw.mit.edu/courses/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/pages/lecture-notes/.

https://medium.com/@ilyurek/gradient-boosting-machines-gbm-with-python-example-b65421d10f2a.

https://it.mathworks.com/help/finance/gbm.html.

https://www.sciencedirect.com/topics/computer-science/brownian-motion.

https://numpy.org/.

https://pandas.pydata.org/.

https://matplotlib.org/stable/gallery/mplot3d/index.html.