Statistical and Stochastic Modeling of the Dynamics of Stock Prices in Biopharmaceutical Industry

Austin Li Newbury High School, Newbury Park, CA

Abstract

This study analyzes the dynamics of Amgen's stock prices using statistical and stochastic models. Change-point detection, anomaly detection, and Granger causality are applied to examine stock behavior. Time series models are used to forecast future trends. Geometric Brownian motion and the Ornstein-Uhlenbeck process are fitted to the data. The results suggest potential Granger causality between Amgen and a competitor, underscoring the value of statistical methods in understanding stock market behavior and offering actionable insights.

Key words: pharmaceutical industry, stock price, change-point detection, anomaly detection, Granger causality, time series modeling, stochastic modeling, geometric Brownian motion, Ornstein-Uhlenbeck process

1 Introduction

1.1 Background

In the rapidly evolving biopharmaceutical market, the complexities among stock prices of major companies can provide crucial insights into market dynamics. This article explores various relationships and models that can offer a deeper understanding of the factors driving stock market movements. Most of the methods employed will focus on the large biotech company Amgen, a biopharmaceutical company established on April 8th, 1980, in Thousand Oaks, California. Today, Amgen ranks among the largest companies in the industry, competing with major players like Johnson & Johnson, Pfizer, Moderna, Eli Lilly, and others. This paper chose to focus primarily on Amgen, while including comparisons with other big companies such as Johnson & Johnson and Pfizer to add depth to this research.

1.2 Literature Review

Stock represents ownership or equity in a company or corporation. The stock market, made up of many companies, acts as a trading platform for investors. However, understanding the stock market involves complex concepts and requires considerable expertise. Extensive research has been done to uncover the factors that influence stock prices, providing valuable insights. For example, Straehl and Ibbotson [1] concluded that total payouts play a crucial role in determining stock market prices in their work on the long-term drivers of stock returns.

One sector that has seen significant growth in recent years is the pharmaceutical industry. As one of the world's most research-intensive industries, the pharmaceutical sector continually develops new products that save lives and enhance the quality of life [2]. Recent shifts in the pharmaceutical market, driven by the rising demand for drugs, have led to more attention on the industry [3].

This review focuses on modeling Amgen's stock, as accurate stock modeling offers valuable insights and practical benefits. Amgen's stock has been the subject of previous studies using techniques similar to those employed in this article. For instance, Sander Bos [4] used Granger causality to analyze a collection of stocks, including Amgen's. Similarly, Hansen [5] applied Granger causality tests to Amgen's stock, highlighting the importance of this method. Other research, such as that by Olaleye [6], demonstrated how machine learning and stochastic simulation techniques could be effectively applied to model Amgen's stock, leading to improvements in inventory management. Additionally, Heckens and Guhr [7] used time series analysis on Amgen's stock to identify long-term financial crisis precursors, showcasing the effectiveness of advanced modeling approaches. Much of the information about stochastic modeling was referenced from an textbook on stochastic processes by Dr. Olga Korosteleva [8]. Additional information on time series analysis in R can be found in Hyndman's textbook as well [9].

1.3 Data Description

The data used in this analysis was obtained from the Yahoo Finance website *yahoofinance.com*. The dataset includes daily closing stock prices for companies such as Amgen, Johnson & Johnson (J&J), and Pfizer. The data cover a specific time period, ranging from January 1st, 2020 to July 23rd, 2024. In Figure 1 below, we present the time series plots for each company, allowing for a visual comparison of their stock price movements over the specified duration.

Figure 1: Closing price for Amgen, J&J, and Pfizer stock

1.4 Article Outline

The article begins with anomaly detection, which identifies unusual observations in the data, followed by change-point detection, used to find shifts in data distribution. Next, it covers Granger causality, a method for exploring cause-and-effect relationships between time-dependent variables.

The article then discusses time series models, including the autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models, used for analyzing and forecasting temporal data. Finally, it addresses stochastic modeling, focusing on geometric Brownian motion and the Ornstein-Uhlenbeck process for modeling random and mean-reverting behavior in time series data. The R code used for the analysis in this article is available on this website:

$$https://github.com/greem1372/Amgen-Code/blob/master/.Rhistory$$

2 Methodology and Applications

2.1 Change-point Detection

2.1.1 Theoretical Foundation

Change-point detection is a machine learning method designed to identify abrupt changes in the statistical characteristics of time series data, such as shifts in mean, variance, or both. To illustrate how this method operates, let's consider a straightforward example involving normally distributed observations with a single shift in mean. Suppose we obtain a sequence of n observations, where the first k observations follow a normal distribution with mean μ_1 and variance σ^2, while the remaining observations are normally distributed with a different mean μ_2 and same variance σ^2. The primary objective of change-point detection in this context is to determine the value of k at which the shift in mean occurs. To achieve this, the method of maximum likelihood estimation is used to find the value k, along with the estimates of μ_1, μ_2, and σ^2.

When the analysis requires the identification of multiple change points, the binary segmentation method is utilized. This technique operates by first identifying a single change point, which divides the dataset into two segments: one above the change point and one below it. After this initial split, the method recursively applies the same change-point detection process to both segments, searching for additional change points. The binary segmentation approach continues this recursive splitting until a specified number of change points is detected or until no further significant changes are found in the data.

2.1.2 Application

By using the daily closing prices of Amgen's stock, we can conduct an analysis to identify shifts in Amgen's average closing price, with five change points specified. The results are displayed in Figure 2 below.

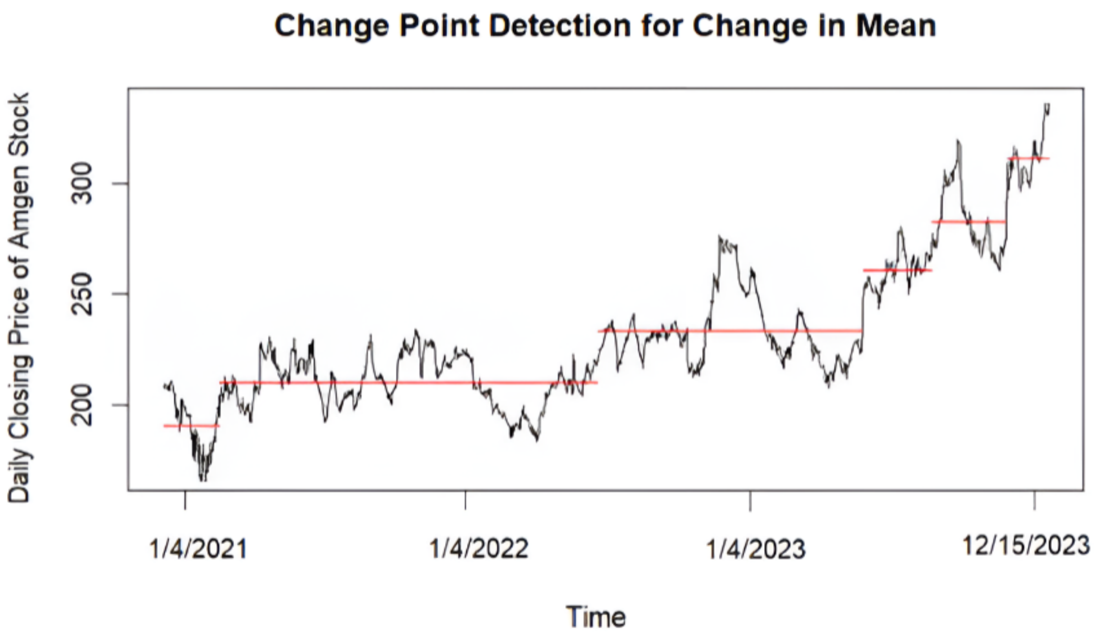

Figure 2: Change-point detection for the mean for Amgen stock

The horizontal red lines represent distinct time intervals during which Amgen's average stock price shifted. These six red lines correspond to five specific dates of stock price changes: April 15, 2020, March 23, 2022, August 7, 2023, December 12, 2023, and May 2, 2024.

The first notable change occurred on April 15, 2020, likely influenced by the looming impact of COVID-19. Just a month later, in May, the U.S. entered lockdown, which brought significant volatility and uncertainty to the market. Amgen, like many companies, experienced stock price fluctuations due to the pandemic-driven turmoil.

On March 23, 2022, a second shift was observed, possibly tied to Amgen's announcement on March 2 of a 10% increase in their second-quarter dividend to $1.94 per share. Furthermore, Amgen's presentation at the Oppenheimer Healthcare Conference on March 10 may have attracted additional investor interest, contributing to the stock price change.

Another change took place on August 7, 2023, shortly after Amgen released its second-quarter financial results and announced that the FDA had approved its leukemia treatment, BLINCYTO. These two developments likely played a role in the stock's upward movement.

By December 12, 2023, Amgen had delivered six presentations in the preceding month and announced an increase in their dividend to $2.25 per share, both of which may have fueled further stock price growth.

The final shift occurred on May 2, 2024, following Amgen's first-quarter results, which reported a 22% increase in sales, driven by a 25% growth in volume. This impressive performance likely attracted more investors, pushing the stock price higher.

Next, a change-point detection was applied to identify the dates with the most significant shifts in variance, occurring on April 15, 2020, March 23, 2022, August 7, 2023, December 12, 2023, and May 2, 2024. These dates are similar to those described above. Figure 3 below illustrates these points on the graph.

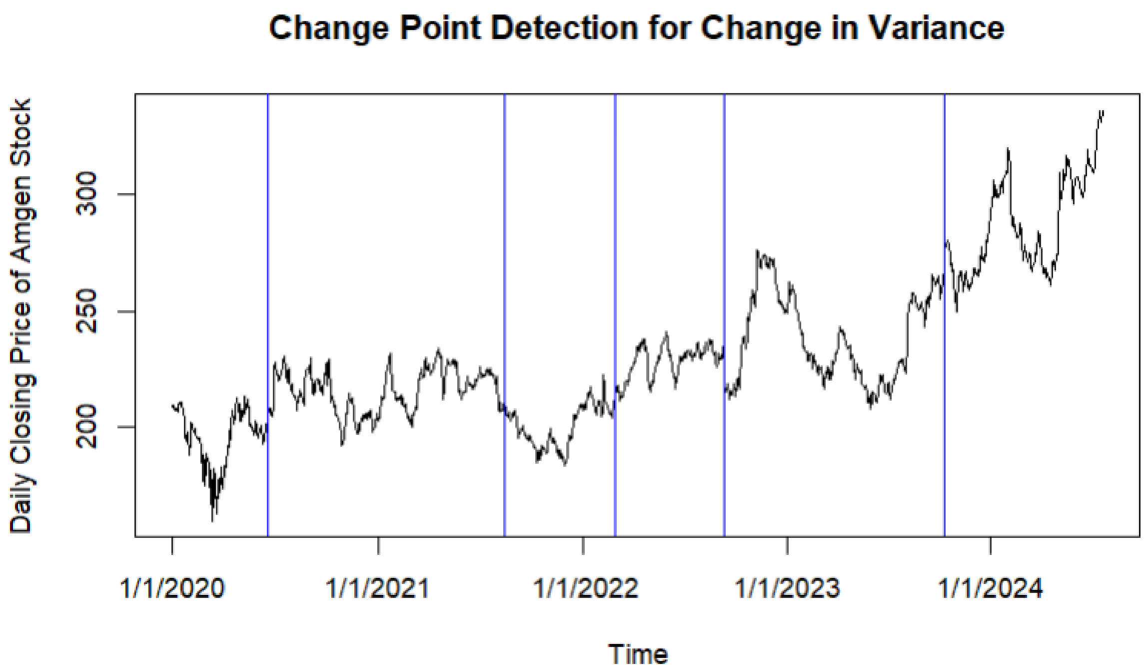

Figure 3: Change-point detection for the variance for Amgen stock

2.2 Anomaly Detection

2.2.1 Theoretical Foundation

Anomalies in time series data are typically identified as outliers in the residuals once linear trends and seasonal patterns have been removed. This process begins with de-trending the data, which involves subtracting the estimated trend component and any seasonal variations to isolate the residuals that reflect the underlying fluctuations in the data. After this adjustment, values that fall outside the expected range can be considered potential outliers.

In the context of the de-trended values, outliers are defined as those that lie below the lower limit $Q1 - 3 \cdot IQR$ or above the upper limit $Q3 + 3 \cdot IQR$, Here $Q1$ represents the first quartile (25th percentile), which marks the value below which 25% of the data falls, and $Q3$ denotes the third quartile (75th percentile), indicating that 75% of the data is below this value. The interquartile range, calculated as $IQR = Q3 - Q1$, measures the spread of the middle 50% of the data, providing a robust statistic that is less sensitive to extreme values. The default threshold of $3 \cdot IQR$ is commonly used to determine which observations are flagged as outliers. However, this threshold is not fixed. Analysts have the flexibility to modify it according to the specific characteristics of the dataset and the goals of the analysis. For instance, a smaller multiplier might be employed to identify more observations as outliers, which could be particularly useful in datasets where subtle deviations from the trend are important to detect. Conversely, increasing the multiplier can reduce the number of flagged outliers, allowing for a focus on only the most extreme deviations.

2.2.2 Application

In this section, we apply anomaly detection to the Amgen stock data to identify outliers, using the default threshold of $3 \cdot IQR$. The results, displayed in Figure 4, reveal clusters of outliers near the start and end of the dataset, with more significant ones appearing around the center. Notably, the central anomalies occurred on January 25-26, 2021, and November 8, 2022.

Figure 4: Anomaly detection for Amgen stock

Upon further investigation online into the reasons behind the stock drops following the detected anomalies, we can refer to Amgen's reports to help explain these declines.

The first anomaly, observed on January 25-26, 2021, remains difficult to explain despite efforts to pinpoint a cause. The closest related event is Amgen's fourth-quarter 2020 financial results, released on February 2, 2021. The report included some negative factors, such as a 3% decline in the generally accepted accounting principles (GAAP) earnings per share and a 2% decrease in GAAP operating income.

The second anomaly, noted on November 8, 2022, appears to be linked to Amgen's third-quarter financial results released on November 3, 2022. Key findings that may have contributed to the stock's decline include a 1% drop in revenue, an 8% decrease in volume growth, and a 2% negative impact on foreign trade.

2.3 Granger Causality

2.3.1 Theoretical Foundation

Granger causality is a statistical hypothesis test used to determine whether one time series can predict another time series by examining the temporal relationships between them. To test for Granger causality, we work with two stationary time series $\{X_t,\, t \geq 0\}$ and $\{Y_t,\, t \geq 0\}$, meaning their statistical properties (like mean and variance) do not change over time. The analysis involves constructing two models: the full model and the reduced model. The full model can be expressed as

$$Y_t = \sum_{k=1}^{p} a_k Y_{t-k} + \sum_{k=1}^{q} b_k X_{t-k} + \varepsilon_t.$$

In this equation, a_k and b_k are coefficients representing the influence of past values of Y and X, respectively, p and q are the number of lags included, and ε_t represents the error term.
The reduced model, on the other hand, is defined as

$$Y_t = \sum_{k=1}^{p} a_k Y_{t-k} + \varepsilon'_t.$$

In this model, the current state of Y is regressed only on its past values, without considering the influence of X. The primary objective of comparing these two models is to assess whether the inclusion of past values of X significantly improves the model's ability to predict Y. Formally, we use an F-test to check the null hypothesis that the reduced model has a better fit against the alternative hypothesis that the full model better fits the data. If the p-value is less than 0.05, we accept the alternative and conclude that the time series X Granger-causes the time series Y. Conversely, if the p-value is above 0.05, we do not reject the null hypothesis, indicating that X does not Granger-cause Y.

It is important to note that Granger causality does not imply true causation; it simply indicates that there is a predictive relationship based on historical data.

2.3.2 Application

We performed F-tests to determine whether the stock prices of Amgen, Johnson & Johnson (J&J), and Pfizer Granger-cause one another. The analysis was carried out for the four combinations involving Amgen stock: Amgen on J&J, J&J on Amgen, Amgen on Pfizer, and Pfizer on Amgen. We conducted the tests for models with equal lag values p and q, ranging from 1 to 5. Table 1 presents the p-values for all the tests conducted.

Lag (p = q)	J&J ∼ Amgen	Amgen ∼ J&J	Amgen ∼ Pfizer	Pfizer ∼ Amgen
1	0.93668	0.27094	0.19739	0.22875
2	0.53173	0.14657	0.29022	0.33416
3	0.07013	0.00961	0.47525	0.29471
4	0.15564	0.02754	0.61962	0.36136
5	0.17515	0.00006	0.72912	0.47476

Table 1: P-values for Granger causality F-tests for Amgen, Johnson & Johnson, and Pfizer stock prices

The table shows that Amgen stock prices do not Granger-cause J&J stock prices, and similarly, Amgen and Pfizer stock prices do not Granger-cause each other, with all p-values exceeding 0.05. However, for the Amgen on J&J pairing, significantly small p-values are observed for lags 3 through 5, particularly for lag 5, where the p-value is as low as 0.00006. This indicates that J&J stock Granger-causes Amgen's stock price.

There have been previous observations of a connection between J&J and Amgen stock. For instance, in 2023, Reuters reported that Amgen proposed a biosimilar version of J&J's top-selling treatment, Stelara. This may serve as an example of how market developments involving J&J could impact Amgen. However, this connection could also be coincidental.

2.4 Time Series Modeling: Auto-regressive Model

2.4.1 Theoretical Foundation

A time series can be modeled using an autoregressive model with a lag of p, known as $AR(p)$ model, represented by the following equation:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t.$$

In this model, ϕ_0 is the intercept, $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients, and ε_t is the random error, assumed to follow a normal distribution with a mean of zero and constant variance.

To estimate the parameters $\phi_0, \ldots, \phi_p$, methods like maximum likelihood estimation or ordinary least squares are commonly used.

Additionally, the partial autocorrelation function (PACF) is often employed to identify the most appropriate lag p of an AR model. The PACF measures the correlation between a time series and its lagged values, after accounting for the correlations explained by the intermediate lags. Essentially, PACF isolates the direct relationship between a time series and a specific lag, removing the influence of shorter lags. A large spike at a particular lag in the PACF plot suggests a strong direct relationship between the time series and that lag, which should be chosen as optimal.

2.4.2 Application

We start by plotting the partial autocorrelation function for the autoregressive model (see Figure 5 below).

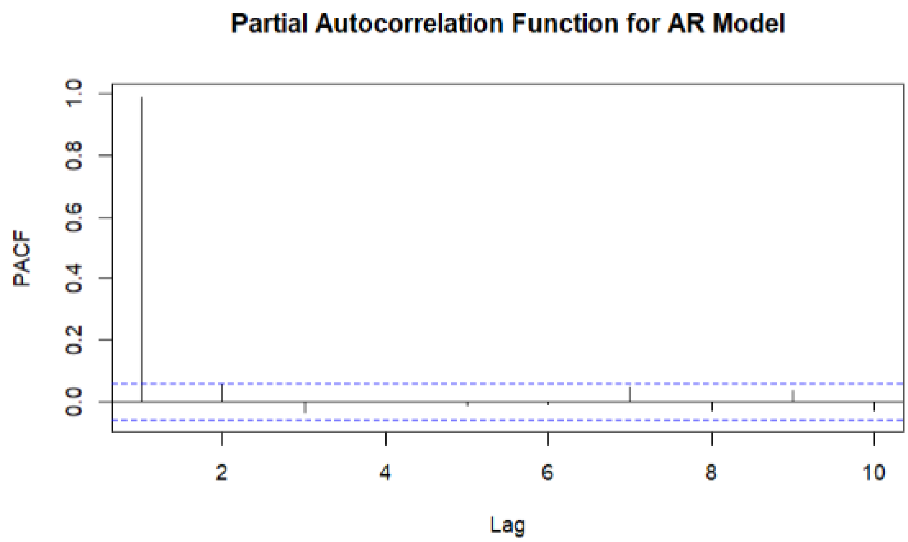

Figure 5: Partial autocorrelation function plot for autoregressive model

The peak at 1 indicates that a lag of 1 provides the best-fitting model. The fitted $AR(1)$ model has the form

$$X_t = 0.1106 + 0.998\, X_{t-1} + \varepsilon_t.$$

2.5 Time Series Modeling: Moving Average Model

2.5.1 Theoretical Foundation

A time series can be modeled using a moving average model with a lag of q, referred to as an $MA(q)$ model. This is expressed by the equation:
In this equation, τ_0 is the intercept, τ_j represents the $j$the moving average coefficient, ε_{t-j} denotes a random process with a mean of zero and constant variance, and ε_t is the random error, assumed to follow a normal distribution with a mean of zero and constant variance.

As in an autoregressive model, the PACF plot aids in determining an optimal value of q. The parameters $\tau_0, \ldots, \tau_q$ are estimated using maximum likelihood estimation or ordinary least squares.

To determine which model fits the data best, a Bayesian Information Criterion (BIC) is used, which is calculated as:

$$\text{BIC} = k\ln(n) - 2\ln(\hat{L})$$

where k is the number of model parameters, n is the sample size, and $\hat{L}$ is the maximum likelihood. A lower BIC indicates a better model fit, while a higher BIC suggests a worse fit.

2.5.2 Application

We create another plot of a partial autocorrelation function (PACF), this time for a moving average model (see Figure 6). The PACF plot shows that lags 7, 9, and 10 are strong fits, suggesting that it takes about one to two weeks for Amgen's actions to impact its stock.

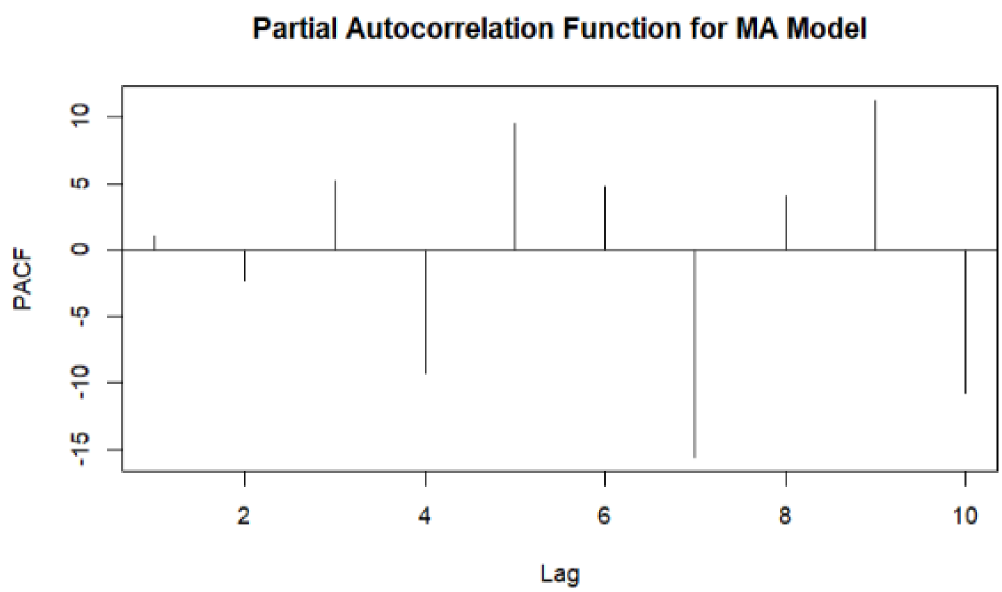

Figure 6: Partial autocorrelation function plot for moving average model

We proceed to fitting the three models, $MA(7), MA(9)$, and $MA(10)$, estimating parameters using maximum likelihood estimation. The BIC values for these three models are respectively, 7099.22, 6906.83, and 6825.09, indicating that $MA(10)$ is the best-fitted model. The estimated form of this model is as follows:

$$X_t = 231.15 + \varepsilon_t + 1.693\,\varepsilon_{t-1} + 1.892\,\varepsilon_{t-2} + 2.015\,\varepsilon_{t-3} + 1.941\,\varepsilon_{t-4} + 1.653\,\varepsilon_{t-5} + 1.348\,\epsilon_{t-6}$$
$$+0.921\,\varepsilon_{t-7} + 0.611\,\varepsilon_{t-8} + 0.269\,\varepsilon_{t-9} + 0.102\,\varepsilon_{t-10}.$$

2.6 Time Series Modeling: Autoregressive Moving Average Model

2.6.1 Theoretical Foundation

A time series can be modeled using an autoregressive moving average (ARMA) model with parameters p and q, denoted as $ARMA(p,q)$. The model is represented by the equation:

$$X_t = \delta + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \tau_j \varepsilon_{t-j} + \varepsilon_t.$$

In this equation, δ represents the stationary component of the model, ϕ_i is the ith autoregressive coefficient, and τ_j is the jth moving average coefficient. Here, ε_{t-j} denotes a random process with a mean of zero and constant variance, while ε_t is the random error, assumed to follow a normal distribution with a mean of zero and constant variance. The parameters of this model are estimated based on the maximum likelihood method or the method of ordinary least squares.

2.6.2 Application

We fit the simplest ARMA model with lags (1,1), and, for comparison, we fit the ARMA models using the lag combinations from the previously fitted AR and MA models. Specifically, we tested ARMA model with (p,q) values of (1,1), (1,7), (1,9), and (1,10). Based on the BIC, the $ARMA(1,1)$ model has the best fit (lowest BIC value). The fitted model is: $X_t = 235.47\ 0.9973\, X_{t-1} - 0.06798\, \varepsilon_{t-1} + \varepsilon_t$.

2.7 Time Series Modeling: Forecasting

Now, we use our fitted models to forecast future values. Since time series models follow algebraic forms, they can be extended beyond the range of the data they were trained on to predict future values. For this analysis, we split the first 1,000 observations of the Amgen stock data as the training set and used the remaining data for testing. We then applied the fitted models to forecast the testing set (see Figure 7).

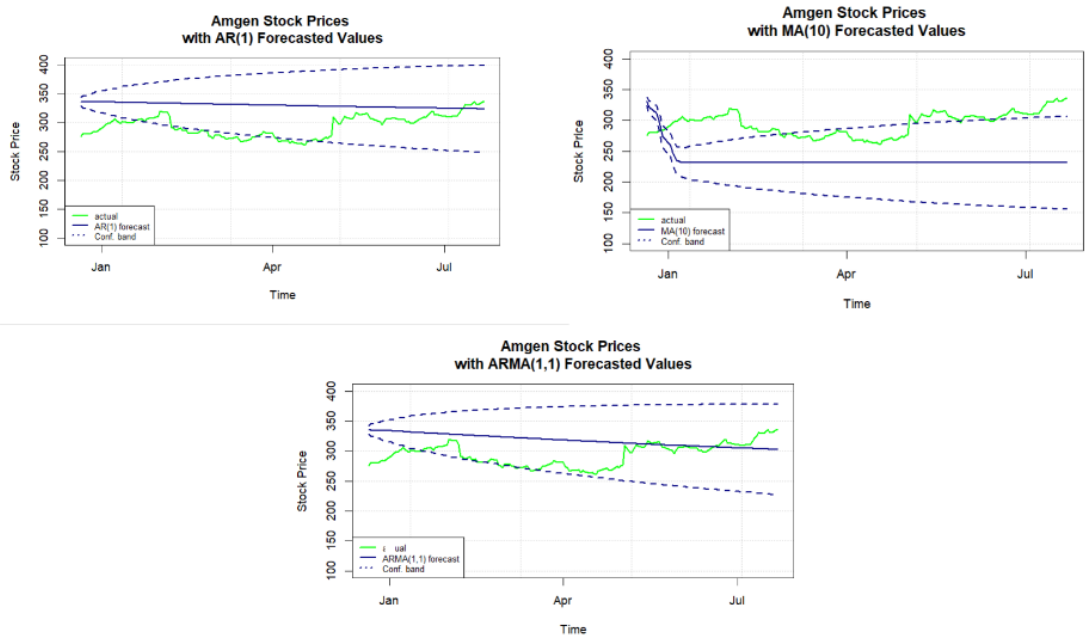

Figure 7: A plot of Amgen stock prices in the testing set along with forecasted values for AR(1), MA(10), and ARMA(1,1) models

We can see that $AR(1)$ and $ARMA(1,1)$ models capture most of the testing data within their confidence bands, indicating potential for real-world application. To evaluate model performance, we compared their BIC values. The AR(1) model had the best fit, with the lowest BIC of 6250.008, closely followed by the $ARMA(1,1)$ model at 6251.341. The $MA(10)$ model performed the worst, with the highest BIC of 6825.086.

2.8 Stochastic Modeling: Geometric Brownian Motion

2.8.1 Theoretical Foundation

A standard Brownian motion $\{B(t), t \geq 0\}$ is a stochastic process that is characterized by the following properties: (1) it starts at zero, that is, $B(0) = 0$; (2) it has independent increments, that is, for any disjoint intervals, the increments are independent random variables; (3) it has stationary increments, meaning that on any intervals of equal lengths, the increments have the same distribution; and (4) $B(t)$ has a normal distribution with mean 0 and variance t.
A geometric Brownian motion $\{X(t), t \geq 0\}$ with drift constant μ and a volatility constant $\sigma > 0$ can be written as:

$$X(t) = X(0) \exp\big(\mu t + \sigma B(t)\big).$$

To estimate the parameters μ and σ from observations, we note that

$$\ln X(t+1) - \ln X(t) = \mu + \sigma B(1).$$

Thus, the ratio $\ln X(t+1)/X(t)$ has $N(\mu, \sigma^2 t)$ distribution. This leads to the likelihood-driven estimators of the form

$$\hat{\mu} = \frac{1}{n-1} \sum_{i=2}^{n} \ln\Big(\frac{X_{n+1}}{X_n}\Big),$$

and

$$\hat{\sigma} = \sqrt{\frac{1}{n-2} \sum_{i=2}^{n} \left[\ln\left(\frac{X_{n+1}}{X_n} \right) - \hat{\mu} \right]^2}.$$

2.8.2 Application

In this section, we analyze Amgen's closing stock price data, treating it as a stochastic process modeled by geometric Brownian motion. First, we generate a histogram of the log-price increments and note that it appears approximately bell-shaped (refer to the figure below). However, the Shapiro-Wilk normality test yields a p-value of less than 0.000001, indicating that the distribution is not normal.

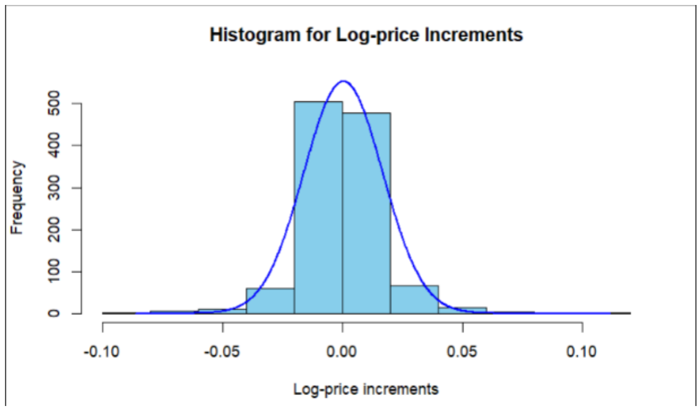

Figure 8: Histogram for log-price increments for Amgen stock daily closing prices

Despite this, we continue by estimating the model parameters under the assumption of normality, which gives us $\hat{\mu} = 0.0004141028$ and $\hat{\sigma} = 0.01649394$. However, the geometric Brownian motion model does not seem to capture the data trend accurately, showing large fluctuations and drifting toward infinity. To further evaluate the model's fit, we conduct ten simulations to gain a clearer understanding of how well geometric Brownian motion represents the stock data. In Figure 9, the observed trajectory is compared with ten simulated trajectories. It is evident that the simulated paths exhibit greater fluctuations and tend to rise more rapidly.

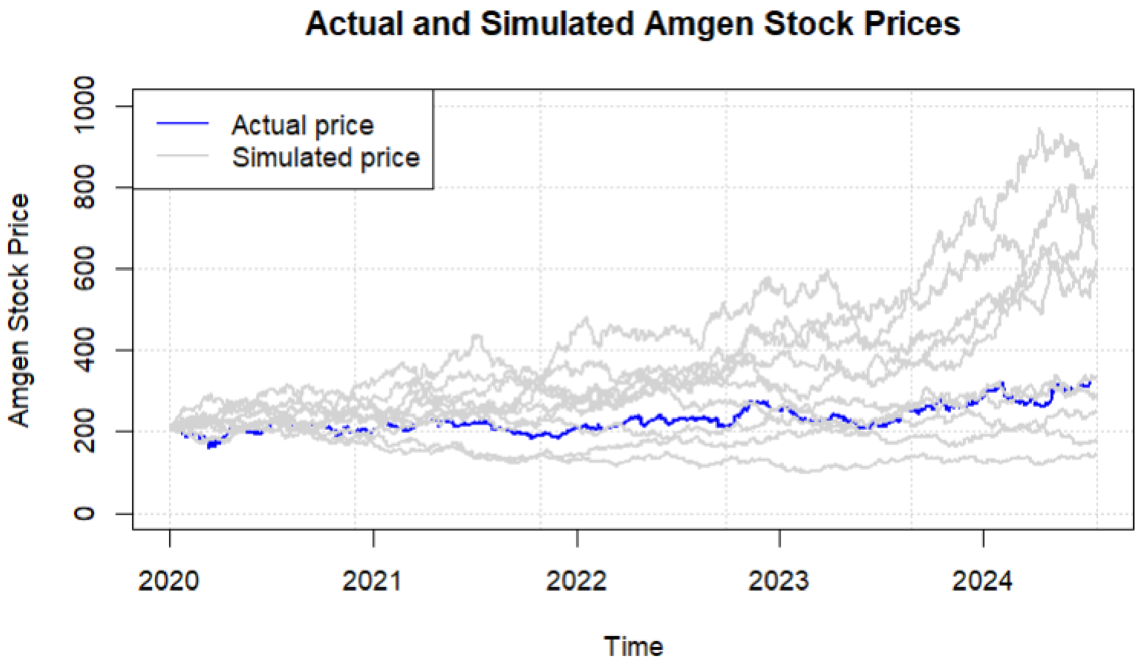

Figure 9: Plot of Amgen stock price and ten simulated trajectories of the geometric Brownian motion

.

2.9 Stochastic Modeling: Ornstein-Uhlenbeck Process

2.9.1 Theoretical Foundation

The Ornstein-Uhlenbeck (OU) process $\{X(t),\, t \geq 0\}$ is another type of stochastic model that, unlike Brownian motion, doesn't drift away but tends to revert to a mean. This process has a long-term mean, denoted by μ, and a volatility constant σ. It also includes a parameter θ, which represents how quickly the process returns to the mean μ.

The OU process is the solution to the following stochastic differential equation:

$$dX(t) = \theta(\mu - X(t))\, dt \,+\, \sigma\, dB(t),$$

and can be expressed in closed form as:

$$X(t) = X(0)e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta t} W\left(e^{2\theta t} - 1\right).$$

To estimate the parameters of the model when fitting it to a dataset, we observe that the process satisfies the following differential equation:

$$X(t + \Delta t) = X(t) + \theta(\mu - X(t))\Delta t + \sigma\sqrt{\Delta t} B(1).$$

By setting $\Delta t = 1$, we can simplify the equation to:

$$X(t+1) - X(t) = \theta\mu - \theta\, X(t) + \sigma\, B(1).$$

This expresses a linear relationship where $X(t+1) - X(t)$ is regressed on $X(t)$. Denoting the estimated intercept as $a = \hat{\theta}\hat{\mu}$ and the slope as $b = -\hat{\theta}$, we find the parameter estimates: $\hat{\theta} = -b$ and $\hat{\mu} = a/\hat{\theta} = -a/b$. The volatility σ can be estimated as the sample standard deviation of the residuals (the error term).

2.9.2 Application

To fit an Ornstein-Uhlenbeck process to Amgen's daily closing stock price, we regress the increments on the lagged observations, obtaining the parameter estimates: $\hat{\theta} = 0.01435791, \hat{\mu} = 249.9769$, and $\hat{\sigma} = 9.353279$. To assess visually how well the Ornstein-Uhlenbeck process captures the overall trend of Amgen's stock, we plot ten simulated trajectories, similar to what was done with the geometric Brownian motion model (see Figure 10). The Ornstein-Uhlenbeck process provides a better fit, as the trajectories revert to the mean, do not drift towards infinity, and show less variability.

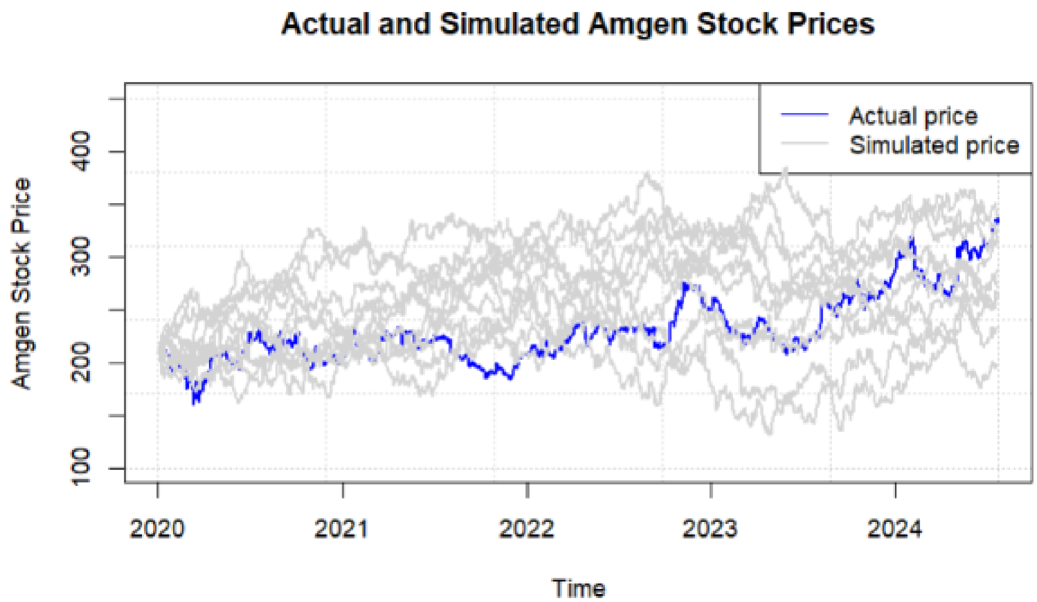

Figure 10: Plot of Amgen stock price and ten simulated trajectories of the Ornstein-Uhlenbeck process

3 Summary and Discussion

This paper investigated various methods for modeling and testing data to draw meaningful conclusions. Techniques such as anomaly detection and change-point detection were applied to Amgen's stock, successfully pinpointing significant price shifts. Granger causality was employed to explore relationships between Amgen's stock and those of other companies, while time series analysis proved valuable for forecasting future stock prices. Stochastic modeling offered another effective approach to capturing the dynamics of Amgen's stock. Altogether, the methods used in this study show considerable promise for future applications, both in stock market analysis and in broader data analysis contexts.

Additionally, other techniques were explored but not included in this paper. For example, nonparametric models, such as the locally estimated scatterplot smoothing (LOESS) model and thin-plate smoothing, were used to analyze Amgen's data. However, these models did not yield significant results and were therefore excluded. Nonetheless, the outcomes of these models can still be accessed via the GitHub link provided in the introduction.

For future research, integrating more variables, such as macroeconomic indicators, could enhance stock price predictions. Combining machine learning techniques with traditional methods may improve the accuracy of anomaly detection and forecasting. Testing these models across different industries would help evaluate their robustness, and exploring real-time applications, like automated trading systems, could offer practical benefits for stock market analysis.

References

[1] Straehl, Philip U. and Ibbotson, Roger G., The Long-Run Drivers of Stock Returns: Total Payouts and the Real Economy (2017). Financial Analysts Journal, Vol. 73, No. 3.

[2] Scherer, F. M. "Chapter 25 the Pharmaceutical Industry." In *Handbook of Health Economics*, Edited by Anthony J. Culyer and Joseph P. Newhouse, Vol. 1, Part B, pages 1297-1336, 2000.

[3] González Peña, et al. "Pharmaceuticals Market, Consumption Trends and Disease Incidence Are Not Driving the Pharmaceutical Research on Water and Wastewater." *MDPI*, Multidisciplinary Digital Publishing Institute, 4 Mar. 2021.

[4] Bos, Sander. "Discovery of multivariate Granger causalities in temporal data." Master Thesis, Eindhoven University of Technology, July 2023

[5] Hansen, Johan Widding. Can Attention Explain the Abnormal Dynamics of the Stock Market?, June 2023

[6] Olaleye, Ololade. "Machine Learning and Stochastic Simulation for Inventory Management." Master Thesis,*Massachusetts Institute of Technology*, 25 June 2024.

[7] Heckens, Anton J, and Thomas Guhr. "A New Attempt to Identify Long-Term Precursors for Endogenous Financial Crises in the Market Correlation Structures." IOPScience, 5 Apr. 2022

[8] Korosteleva, Olga. Stochastic Processes with R: An Introduction. 1st ed., Chapman and Hall/CRC, 2022.

[9] Hyndman, Rob J., and George Athanasopoulos. Forecasting: Principles and Practice. 3rd ed., OTexts, 2021, https://otexts.com/fpp3/.