FORECASTING STOCK PRICES

Instructor: Mikołaj Kasprzak

Members:

Prashant ARYA
Jeremy GARNEAU
Fatin MARWAT
Kaushnav ROY

Table of Contents:

1. Introduction: Continuous Time Series Forecasting in Finance

Traditional time series forecasting methods, such as ARIMA or exponential smoothing, are widely used and perform well when working with data collected at regular intervals like daily or monthly observations. These models assume that the data is evenly spaced and are designed to make predictions for predefined, fixed periods. While this approach is successful in many domains, it often falls short when more precise or irregular predictions are required. For instance, predicting a stock price or index at a specific time, like 1:00 PM next Wednesday, requires models that can go beyond the constraints of discrete data.

In financial markets, this need for precision is even more critical. Stock prices and indices like the S&P 500 are updated every 15 seconds during trading hours. These frequent updates reflect the market's real-time dynamics and require the adoption of forecasting models that can capture and predict values at any given time. Continuous time series modeling provides a way to achieve this by analyzing the continuous evolution of data rather than relying solely on discrete observations. The final dataset used for this demonstration was the closing price (New York time) of daily trading time of the S&P 500 Index.

Methods like the Ornstein–Uhlenbeck (OU) process and Geometric Brownian Motion (GBM) are particularly suited for financial applications. The OU process, often used to model stochastic volatility, helps capture mean-reverting behaviors, which are common in financial systems. GBM, on the other hand, is a foundational model in financial mathematics for simulating asset price dynamics. These models address challenges such as irregular data intervals and volatility clustering—where high-volatility periods follow one another—key characteristics of financial time series. This clustering reveals patterns of heightened market risk and uncertainty, which traditional models may overlook.

The application of continuous time series modeling extends beyond just predicting prices. These models allow analysts to understand the underlying patterns in market behavior, such as how volatility reacts to events like market crashes or recoveries. For example, through Euler discretization, the Ornstein–Uhlenbeck process simulates multiple trajectories of volatility, highlighting the market’s tendency to fluctuate around a central mean. This enables better risk management by providing insights into potential extreme market conditions.

While continuous models are more complex and require advanced techniques to implement, the benefits they offer are significant. They are well-suited for handling irregular data, making them highly relevant in real-world scenarios where traditional methods fall short. Whether forecasting stock prices, managing portfolio risks or understanding volatility, continuous models provide the granularity and flexibility needed for precise and actionable insights.

2. Stochastic Processes in Continuous Time

A Continuous stochastic process is a phenomenon that can be thought of as evolving in time influenced by randomness. Common examples are the location of a particle in a physical system, the price of a stock in a financial market, interest rates, etc. Stochastic processes are fundamental in fields like physics, finance, biology, and engineering, in predicting systems with uncertainty.

Continuous-time stochastic processes can be thought of as a natural extension of discrete-time processes. It is denoted as $\{X_t\}_{t\geq 0}$, and is described using the probability density function function $f_X(\omega)$, which defines the likelihood of the process over time (t).

A basic example is the erratic motion of pollen grains suspended in water, the so-called Brownian motion. The mathematical model of the Brownian motion will be our central object of study when modeling stock prices. Just like pollen grains' motion is determined by the infinitesimal interactions with water molecules, the stock price is influenced by buying and selling actions, which either push the price up or down. These price movements exhibit random fluctuations that resemble Brownian motion.

2.1. Brownian Motion

Brownian motion, also referred to as the Wiener Process, is a continuous-time stochastic process that satisfies the following properties:

1. $P(W(0)=0)=1$: The process starts at zero.
2. Stationarity and Independent Increments: The distribution of $W(t)-W(s)$ depends only on the time difference $(t-s)$. Increments over non-overlapping intervals are statistically independent.
3. $W(t) - W(s) \sim N(0, t - s)$ for $t > s$: Increments follow a normal distribution with mean zero and variance proportional to time.

Brownian motion is the 'limit' of a simple random walk as the step size decreases and the number of steps increases infinitely. The process is highly oscillatory and crosses the time axis infinitely often. Brownian motion paths are continuous but not differentiable anywhere. It requires a specialized mathematical framework (Ito's calculus) to handle its non-differentiability.

The **quadratic variation** of a stochastic process measures the accumulated squared increments of the process over time. For a standard Brownian motion W(t)

$$[W]_t = \lim_{n\to\infty}\sum_{i=0}^{n-1}(W(t_{i+1}) - W(t_i))^2$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[0,t]$, such that $\max(t_{i+1} - t_i)$ goes to zero. For a continuous and differentiable function, the quadratic variation equals 0. But for standard Brownian motion, it converges to t:

$$[W]_t = t.$$

$$(dW(t))^2 = dt,$$

The intuition is that Brownian motion is continuous but nowhere differentiable. This makes its paths highly irregular, and the notion of variation becomes inadequate. Its irregular path leads to non-zero quadratic variation. This result is central to the derivation of Ito's lemma.

3. Prediction Model

3.1. Continuous Autoregressive Model

A continuous time white-noise process, ξ(t), with mean zero and variance σ^2 is defined such that for s > r:

$$\mathbb{E}\left[\int_r^s \xi(t)dt\right] = 0, \quad \operatorname{Var}\left[\int_r^s \xi(t)dt\right] = (s-r)\sigma^2$$

$$\mathbb{E}\left[\int_r^s \xi(t)\,dt \cdot \int_p^q \xi(t)\,dt\right] = 0, \quad \text{for } r < s < p < q$$

where the integral of ξ(t) from time 'r' to 's', is equivalent to the increment of a Brownian motion, W(s) - W(r), as mentioned in the previous section (Property-2,3).

$$\mathbb{E}\left[\int_r^s f(t)\xi(t)dt\right] = 0 \qquad \mathbb{E}\left[\left(\int_r^s f(t)\xi(t)\,dt\right)^2\right] = \sigma^2\int_r^s f^2(t)\,dt$$

Using the above definition of continuous time white noise, a first-order continuous time process can be represented as the stochastic differential equation:

$$\frac{dy(t)}{dt} = \gamma + \alpha y(t) + \xi(t)$$

Strictly speaking, the differential equation above is not valid as ξ(t) does not exist, and y(t) is not mean-square differentiated. An alternative way of writing the equation is:

$$dy(t) = [\gamma + \alpha y(t)]\,dt + dW(t)$$

where W(t) is a Wiener process and dW(t) mean equals zero, variance $\sigma^2 dt$, and is uncorrelated in different periods. This stochastic process is called the Vasicek Model, an extension of the Ornstein-Uhlenbeck Model. It is also considered as the continuous-time analog of the discrete-time AR(1) process. The solution to this stochastic process can be written:

$$y(t) = \left[y(0) + \frac{\gamma}{\alpha}\right]e^{\alpha t} - \frac{\gamma}{\alpha} + \int_0^t e^{\alpha(t-r)}\xi(r)\,dr \qquad -(1)$$

For discrete observation:

$$y(t) = \left[(e^{\alpha} - 1)\frac{\gamma}{\alpha}\right] + e^{\alpha}y(t-1) + \int_0^1 e^{\alpha(1-r)}\xi(r)\,dr \qquad -(2)$$

Observations ($y_n = y(t_n)$) follow an AR(1) process:

$$y_n = \phi y_{n-1} + \beta + \epsilon_n$$

where,

$$\phi = e^{\alpha}, \beta = \frac{\gamma(e^{\alpha}-1)}{\alpha} \qquad \operatorname{Var}(\epsilon_n) = \sigma^2\frac{e^{2\alpha}-1}{2\alpha} \qquad -(3)$$

The stationarity condition for the continuous-time model is that α is negative; it is easy to see from the result (3), that α < 0 corresponds to Φ < 1. Since α = -∞ corresponds to Φ = 0, negative values of Φ are not permitted in the discrete-time AR(1) process. Furthermore, the assumption of a Gaussian distribution for integrals of ξ(t) means that ε_t also has a Gaussian distribution. If y_l is treated as fixed, the discrete-time parameters Φ, β, and Var(ε_t^2) can be estimated by linear regression, and the corresponding estimators of the continuous time parameters α,γ, and σ^2 can be obtained from the three equations (in 3).

Now suppose that observations are spaced at irregular intervals, but assume, for simplicity, that γ = 0. This solution is applied to our forecasting weekly stock prices. Let the τ^{th} observation be denoted by y_τ for τ = 1,..., T, and suppose that this observation is made at time t_τ. Let $t_\tau - t_{\tau-1} = \delta$ as indicated and let $t_0 = 0$. Then equation (1) becomes

$$y(t_\tau) = \exp(a\delta_\tau)y(t_{\tau-1}) + \int_0^{\delta_\tau} \exp[a(\delta_\tau - s)]\xi(t_{\tau-1} + s)ds$$

$$y_\tau = \phi_\tau y_{\tau-1} + \xi_\tau \qquad \phi_\tau = \exp(a\delta_\tau) \qquad \mathrm{Var}(\xi_\tau) = \sigma_\tau^2 \qquad \sigma_\tau^2 = \sigma^2(\phi_\tau^2 - 1)/2a\delta_\tau$$

3.2. Modeling volatility using Ornstein-Uhlenbeck Process

Ornstein-Uhlenbeck is a stochastic process that is a stationary Gauss-Markov process, exhibiting mean-reverting behavior.

$$dX_t = \theta(\mu - X_t)\,dt + \sigma\,dW_t$$

X_t: The state of the process at time t.
μ: The long-term mean to which the process reverts.
$\Theta > 0$: The rate of mean reversion (how quickly the process reverts to μ).
$\sigma > 0$: The volatility or intensity of the random noise.
W_t: A standard Wiener process (Brownian motion)

The OU process is stationary if it starts in its equilibrium distribution. Over time, the process stabilizes around μ with constant variance. The solution to the Ornstein-Uhlenbeck SDE can be written as:

$$X_t = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \sigma \int_0^t e^{-\theta(t-s)}\,dW_s$$

For numerical purposes, the OU process is discretized as:

$$X_{t+\Delta t} = X_t + \theta(\mu - X_t)\Delta t + \sigma\sqrt{\Delta t}\,Z,$$

where $Z \sim N(0,1)$ is a standard normal random variable. This equation is called Euler-Maryuama discretization, and we have used this to model daily volatility.

We have utilized this result to model volatility and incorporated it into the prediction of stock prices using the Geometric Brownian Motion (GBM) model (section: 3.3).

Implementation:

- Calculated historical volatility from log returns using a rolling window (determined by ACF). For a financial asset with price S_t, let $X_t = \ln(S_t)$ represent its logarithm. The log return over a time scale Δ is defined as:

$$r_t = X_{t+\Delta} - X_t = \ln\left(\frac{S_{t+\Delta}}{S_t}\right)$$

- Estimated parameters (θ, μ, σ) for the OU process using Maximum Likelihood Estimation (MLE).
- Simulate volatility paths using the OU stochastic differential equation.

3.3. Forecasting using Geometric Brownian Motion Model

Geometric Brownian Motion is a stochastic process widely used to model stock prices and other financial variables. It is defined by the stochastic differential equation (SDE):

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

S_t: The value of the variable (e.g., stock price) at time t
μ: The drift term (average rate of return)
σ: The volatility term
W_t : A standard Wiener process (Brownian motion)

The solution to this stochastic differential equation:

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$$

To simulate the GBM numerically, the SDE is discretized using the Euler-Maruyama method over a small time interval Δt:

$$S_{t+\Delta t} = S_t \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t} Z_t\right]$$

where $(\mu - \sigma^2/2)\Delta t$ represents the growth of the asset price, adjusted for the drift correction (due to Ito's lemma), and $\sigma\sqrt{\Delta t}$. Z represents random fluctuations due to market volatility.

We utilized this discretization, combined with the volatility results derived from the OU process, to predict stock prices and the weekly forecast.

4. Results and Analysis

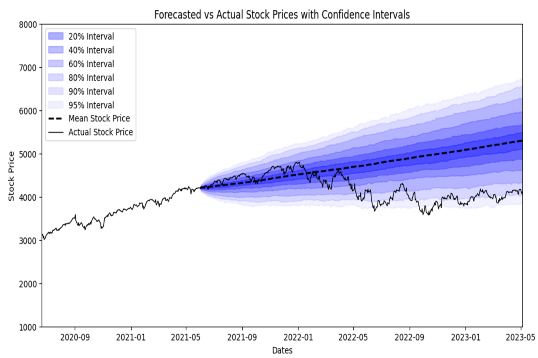


Fig 4.1.: Forecasted Stock Price (AR 1 process)

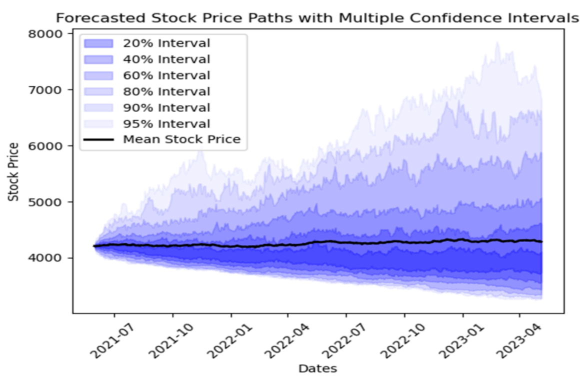


Fig 4.2.: Forecasted Stock Price (Geometric Brownian process)

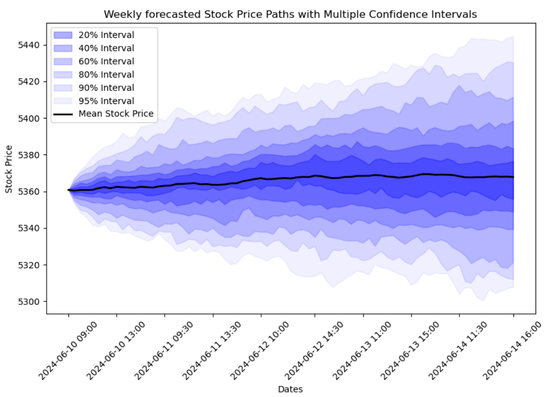


Fig 4.3.: Weekly Forecast (AR 1 Process)

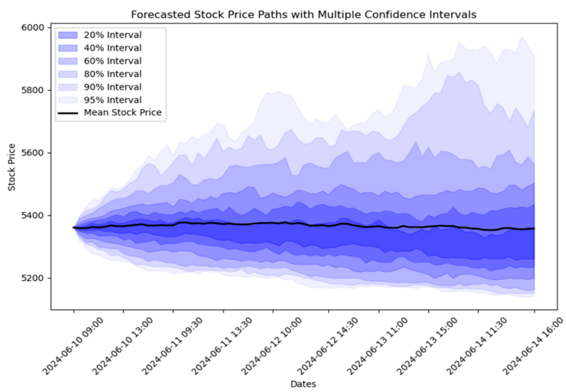


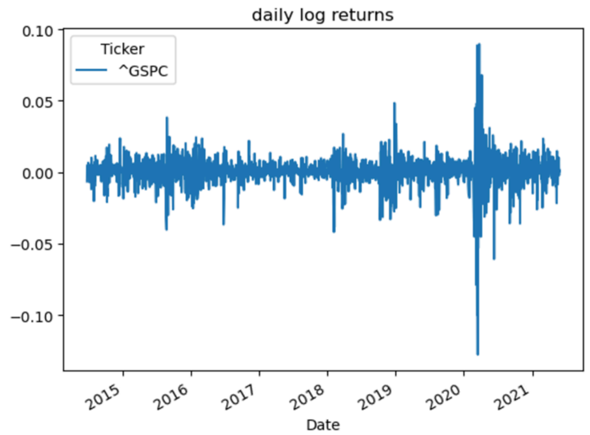
Fig 4.4.: Weekly Forecast (Geometric Brownian Process)

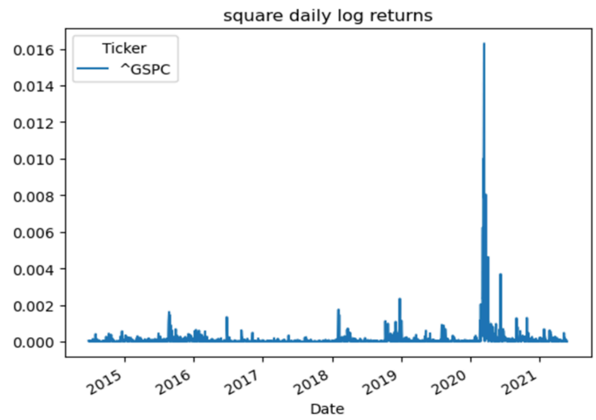
We ran both the AR(1) generalization and GBM models 1000 times and used both models to predict volatility as well as stock prices. This was done with both the original step size of the data set (every trading day) and in half-hour intervals to showcase the granularity of the models. After running the AR process over 1000 times, we were able to see a strong variability in the projections with the projections being highly skewed towards the upper side, with some of the scenarios being very extreme.

The actual forecasted price was well within our predicted interval. When we see the histogram distribution of the stock price after one year, it is very close to a Gaussian distribution with a slight skew towards the left. When we keep on seeing the forecasted histograms for further periods, we see that the variance increases the further we get away from the start of the forecast. With the increasing mean, the left tail (below mean) stays mostly constant and the right tail increases in scedasticity.

When we run the forecasted prices with multiple confidence intervals, we can see that the distribution is pretty flat except at the 95% interval, where it is more stretched. Also, the mean becomes pretty close to the median. The process's mean-reverting behavior is visible in our weekly and daily step discretizations for GBM and OU.

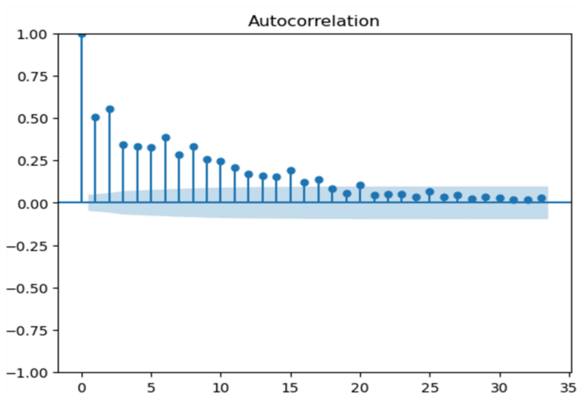
Additionally, we ran some tests on the dataset to explain it a bit more.

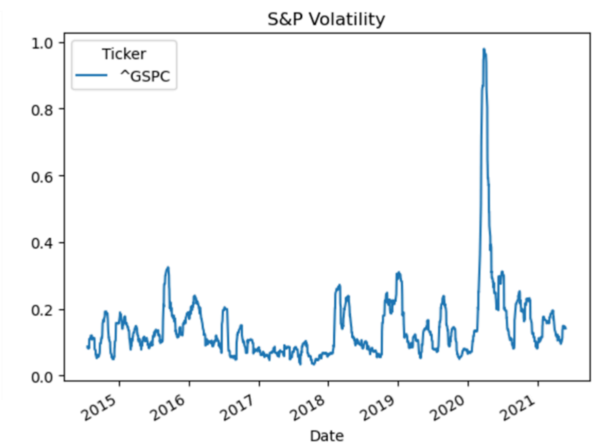




The historical log returns capture daily price changes, where high volatility periods support each other. This supports market randomness with minimal autocorrelation.

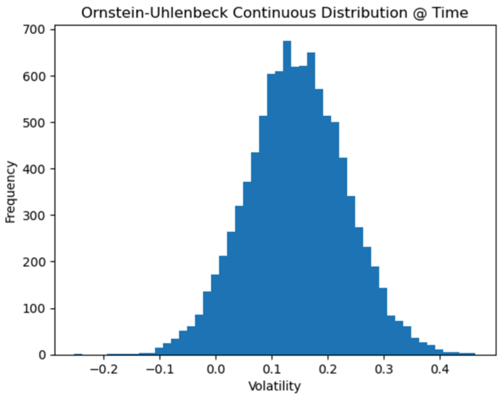
The squared log returns remove directionality to measure magnitude. It captures patterns of persistent high or low volatility. This is a key input for the stochastic processes we are trying to model.

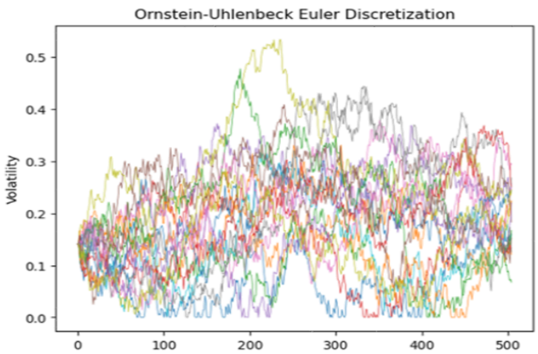




The ACF plot of squared log returns validates this conjecture, showing a tapering-off influence in which at around 20 lags there is still some influence that would indicate clustering. We can therefore see predictable periods of heightened market risk which is important for risk management. This would indicate that volatility is a good way to model the stochastic process rather than opting directly for returns.

The historical volatility highlights market reactions to events (crashes, recoveries). Volatility spikes indicate uncertainty or shock periods. Smoothed trend gives an idea of better long-term insights.





The Ornstein-Uhlenbeck Euler model was used to predict future volatility. Several trajectories of the volatility were simulated. There was a tendency to fluctuate around a central value, consistent with the mean reverting nature. The stationary process ensures paths are bounded and stable.

The histogram of simulated volatility shows the distribution of simulated volatility values after 4 months. A mean-reverting behavior is captured with the concentration near the long-term mean. The tails show that there are occasional extreme deviations in volatility.

All charts spoken about above as well as additional ones not shown here may be found in our accompanying Jupyter notebook.

5. Conclusion

We can conclude that continuous time models like GBM are powerful for capturing stock price dynamics. They handle irregular data and volatility clustering well.

Some Key takeaways are that the predictions align with the actual trend over short intervals and are great for risk management. The caveat is that these models are designed with the idea of stationarity. Stock index prices are decidedly not stationary data (E(y)<>0), and so are not mean-reverting, therefore predictions are not valid over long intervals where the mean changes significantly, but we believe that it would be a good way to demonstrate the practical usefulness of this kind of modeling, especially with our week-sized forecasts with half-hour steps in periods.

6. References

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

- https://statmath.wu.ac.at/~frey/lecture_notes/CTF-lecture-notes.pdf

- https://github.com/TheQuantPy/youtube-tutorials/tree/main/2022/001%20Jan-Mar

- Forecasting, structural time series models and the Kalman filter, Chapter 9 (Continuous time) by Andrew C. Harvey, 1990