Understanding Interest Rates Through the Vasicek Model

By Aditya Vardhan Veerathu

Probability, Expectation, Variance and Other Statistical Tools

Before understanding the intricacies of Interest Rates and the Vasicek Model, one has to understand how basic probability and expectation work. Probability is the branch of mathematics that deals with measuring the likelihood of an event happening. It gives a numerical value between 0 and 1 (or 0% and 100%) that represents how likely it is that a particular outcome will occur. Meanwhile, the expectation of a random variable is the long-run average value it would take after many repetitions of the experiment.

Consider a random variable X which as a probability $\mathbb{P}$(x) associated with each event. Then, in a discrete case this would be equal to:

$$\mathbb{E}[X] = \sum_x x * \mathbb{P}(x)$$

While the continuous case would be

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x * \mathbb{P}(x)dx$$

Essentially, you sum up all the individual events multiplied by the probability of those events.

Expectation tells you what outcome to "expect" on average over many repetitions, even if any single outcome is unpredictable. It helps choose between uncertain options by comparing their expected values. Rational decisions often favor the option with the higher expected value. It is especially useful in calculating expected losses or returns, which is crucial for pricing insurance, designing financial portfolios, and assessing risk.

Expectation is great at telling the average or central value of a random variable, but it tells you nothing about how reliable or stable that average is. That is where variation comes into play. Variance or Variation measures how much the outcomes deviate from the expected value, on average. Its square root, standard deviation, is often used because it's in the same units as the data.

The formula often used to for variance is given by:

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2)]$$

If variance is small, the outcomes are tightly clustered around the mean. If variance is large, the outcomes are more spread out and unpredictable. A common application variance is used for risk analysis. Investors want to maximize expected returns while minimizing risk. Fortunately, variance and standard deviation measure how volatile an investment is helping choose optimal portfolios for a given time period.

Other statistical tools that will be used throughout the discussion are Markov Processes, Martingales, and the Central Limit Theorem.

A Markov process (or Markov chain in discrete time) is a stochastic process where the future state of the system depends only on the current state, and not on the sequence of events that preceded it. This property is known as the Markov property or memoryless property. In simpler terms, the future is independent of the past, given the present.

Meanwhile, the martingale property is a concept from probability theory that describes a specific kind of stochastic process. A process with the martingale property has a "fair game" characteristic, meaning its future value, given all past information, is expected to be equal to its present value. In other words, a martingale is a process where, at any point in time, the best prediction for the next value is the current value—there's no "drift" or trend over time. What this means is that the expected value of a process’s next step is equal to the value of the current step. Essentially, there is expected gain or loss. In a financial market, a fair stock price is an example of a martingale, assuming no additional factors like dividends or interest rates. If you know the current stock price, the best prediction for its future price (ignoring trading costs, etc.) is simply the current price—there's no built-in tendency for it to go up or down over time.

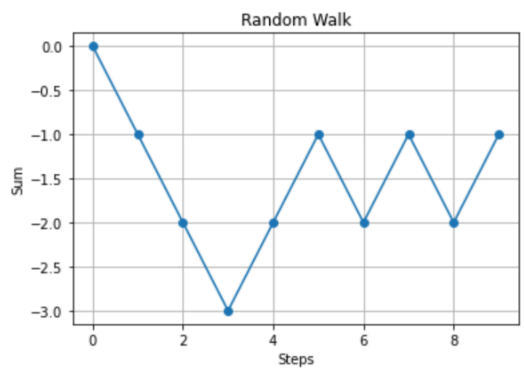
Another important concept that becomes extremely important is the Central Limit Theorem. The Central Limit Theorem (CLT) is a fundamental concept in statistics that explains why the distribution of sample means tends to be normal, even when the underlying population distribution is not. Specifically, it states that if you take sufficiently large random samples from any population with a finite mean and variance, the distribution of the sample means will approximate a normal (bell-shaped) distribution. This holds true regardless of whether the original data is skewed, uniform, or follows any other shape, as long as the samples are independent and the sample size is large enough (typically $n \geq 30$ is sufficient).

Brownian Motion

Building on the Central Limit Theorem, we can now understand how it lays the foundation for more advanced concepts in probability and stochastic processes—one of the most important being Brownian motion. While the CLT describes the behavior of the average of many random variables, Brownian motion models the continuous and cumulative effect of countless tiny random movements over time.

The origin of Brownian motion traces back to 1827, when the botanist Robert Brown observed through a microscope that pollen grains suspended in water moved in a jittery, erratic manner, even though there were no visible forces acting on them. Initially, Brown thought the motion might be a property of life, but he later observed the same behavior in inorganic particles like dust, confirming it was a physical phenomenon rather than a biological one.

Even though the phenomenon was first observed in nature, it is possible to model this motion using statistical principles. In order to do so, first consider the idea of a random walk. To construct the random walk, take a random variable X_i such that after a fair coin flip, $X_i = 1$ if it lands heads and $X_i = -1$ if it lands tails. Now consider a different random variable M such that it takes the sum of each X_i from 0 to some i. To start off the random variable M, $X_0 = 0$. M represents the random walk on a smaller scale. The graphs below are examples of the random walk.



As you can see its long term average is expected to be 0, but there can be moments of time where either is far below 0 or rises far above 0.

A simple random walk is a discrete-time process where at each step, an entity moves randomly in one of two directions (say, left or right) with equal probability. Over time, the path of the random walk becomes a jagged sequence of steps, and the position of the entity at any given time is the sum of these random steps. In a simple random walk, the position of the particle at each time step is a discrete integer value (e.g., step 1, step 2, etc.). As the time intervals between steps become smaller and smaller (i.e., as the walk becomes more continuous), the behavior starts to resemble Brownian motion. One more thing needs to change before the random walk can accurately represent brownian motion which is that the time steps shrink to zero and the number of steps increases to infinity.

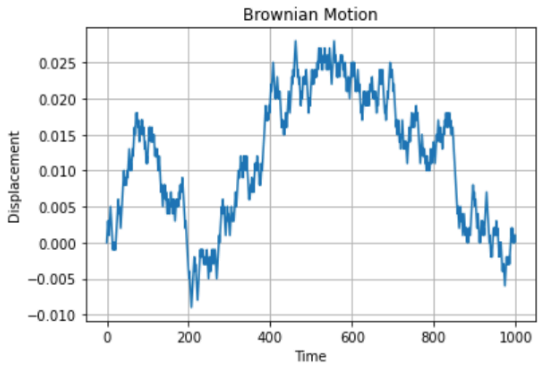
The actual Brownian Motion[1] equation can be represented by

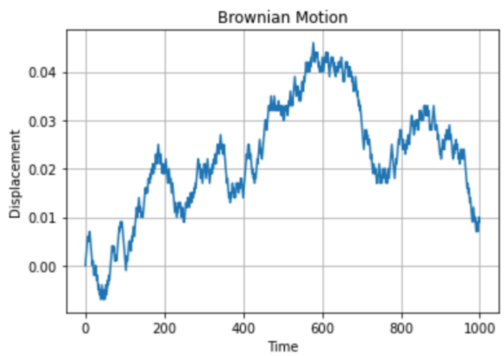
$$W^n(t) = \frac{1}{\sqrt{n}} M_{nt}$$

Where W is Brownian Motion, sometimes also called the Wiener Process.

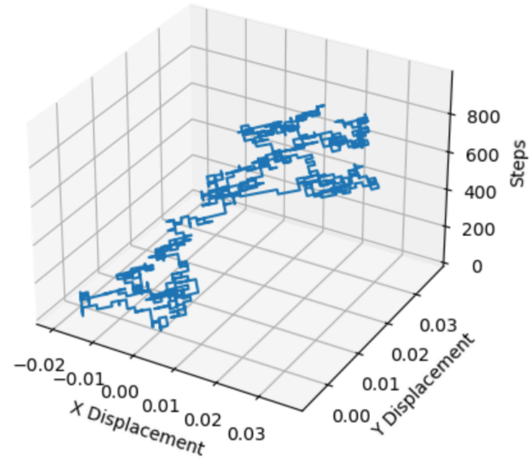
[1] Quick Disclosure here about the Wiener Process. While the graphs generated here and future graphs will use the traditional approach of Brownian Motion of taking large sums of random variables, it is also possible to generate graphs using sin and cosine curves infinitely overlaid on top of each other and taking a small subsection of that.

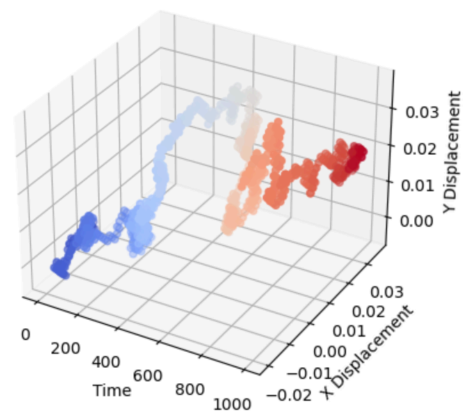
Below are examples of these graphs generated.

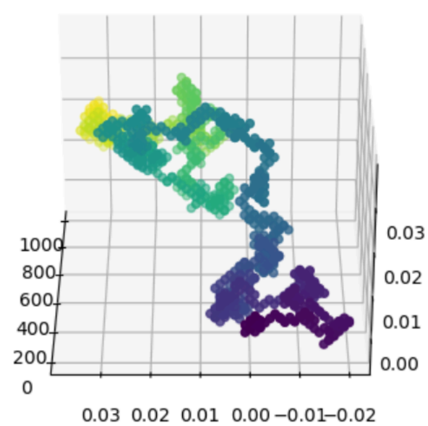




The idea of random walk and Brownian Motion can also be expanded to higher dimensions like the graphs below and on the side.







Central Limit Theorem (CLT) plays a crucial role in this transition. If you think of each step in a random walk as an independent random variable, the CLT tells us that as the number of steps grows large, the distribution of the sum of these steps will approach a normal distribution. When we look at the random walk at increasing time scales, the distribution of the position after many steps will tend to a normal distribution, and the trajectory of the walk will start to resemble a continuous path.

In general for a random motion to be consider Brownian Motion, it has to follow five different properties:

1. Being finite
2. Being Continuous
3. Satisfying both the Markov and Martingale Properties
4. Having unbounded variation
5. Being strongly normally distributed

Geometric Brownian Motion

While Brownian motion models random movement of a particle in space, Geometric Brownian motion incorporates the concept of growth or decay, making it more suitable for modeling processes where the underlying variable (like a stock price or population size) is subject to both random fluctuations and exponential growth.

Geometric Brownian motion is a stochastic process that models the evolution of variables, like stock prices or asset prices, where the value of the process is influenced by both a drift (representing growth or decay) and random fluctuations (captured by Brownian motion). Mathematically, GBM is described by the stochastic differential equation (SDE): $dS(t)=\mu S(t)dt+\sigma S(t)dW(t)$.

Before jumping into how these geometric models behave, it is worthwhile to explore what each of the variables and constants in the stochastic differential equation represent. The first of which is the random variable S which is used to model an asset price. Next we have μ which is the drift term. Then, we have σ which represents the volatility of the given asset. Finally have dW(t) which is the differential of the Brownian Motion.

Here is the general setup for Geometric Brownian Motion (assuming no drift).

Let H represent the number of Heads and T be the number of tails in nt trails which means both of which are scaled appropriately:

$$nt = H_{nt} + T_{nt}$$
$$M_{nt} = H_{nt} - T_{nt}$$

Consider a Stock Price S(t) such that it moves up by a factor of u and moves down by a factor of d depending on whether the coin flip landed on heads or tails. So that means the following is true:

$$S(t) = S(0)u^H d^T$$

Now we will let u and d be:

$$u = 1 + \frac{\sigma}{\sqrt{n}} \text{and } d = 1 - \frac{\sigma}{\sqrt{n}}$$

The sqrt(n) term is there to ensure that the overall variance for the scaled walk will remain t. The sigma term on the top represents the volatility of a stock which is usually calculated by finding the standard deviation of the stock up to time t. This means that the volatility is also a function of t which will become important later.

$$S(t) = S(0)(1 + \frac{\sigma}{\sqrt{n}})^{H}(1 - \frac{\sigma}{\sqrt{n}})^{T}$$

$$H = \frac{1}{2}(nt + M_{nt}) \text{ and } T = \frac{1}{2}(nt - M_{nt})$$

$$S(t) = S(0)(1 + \frac{\sigma}{\sqrt{n}})^{\frac{1}{2}(nt+M_{nt})}(1 - \frac{\sigma}{\sqrt{n}})^{\frac{1}{2}}$$

We want S(t) as n approaches infinity. This can be solved when you take the natural log of both sides and find a series approximation for ln(1+x). In the end, it simplifies to:

$$S(t) = S(0)e^{\sigma W(t) - \frac{1}{2}\sigma^2 t}$$

When differentiating S, it resembles the Geometric Brownian Motion that was originally presented.

Another term that was presented in the Geometric Brownian Motion is dW which can be solved using quadratic variation.

The quadratic variation of M up to time k which is defined as the following:

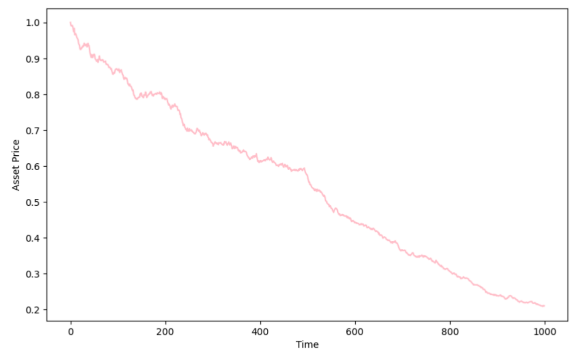
$$\sum_{i=1}^{k}(M_{i+1} - M_i)^2$$

For Brownian Motion, the quadratic variation is:

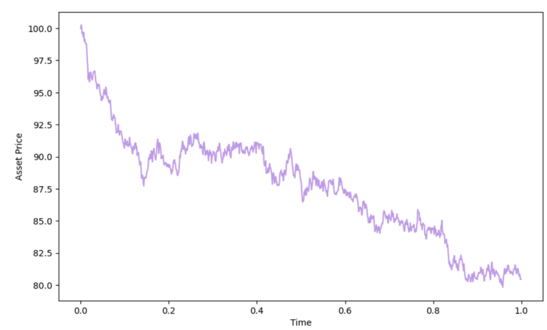
$$\sum_{i=1}^{nt}(W_{i+1} - W_i)^2 = t$$

Through some calculations we will also find that ΔW^2 = dt or more informally written as dW*dW = dt. Which is why often we will see dW = sqrt(dt).

Each factor of Geometric Brownian Motion is now properly defined. Below are graphs of asset prices using Geometric Brownian Motion:

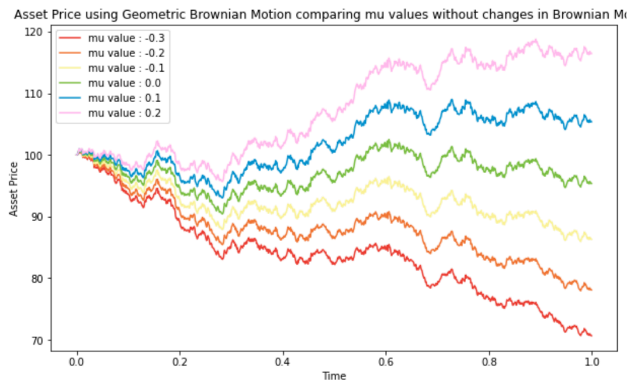


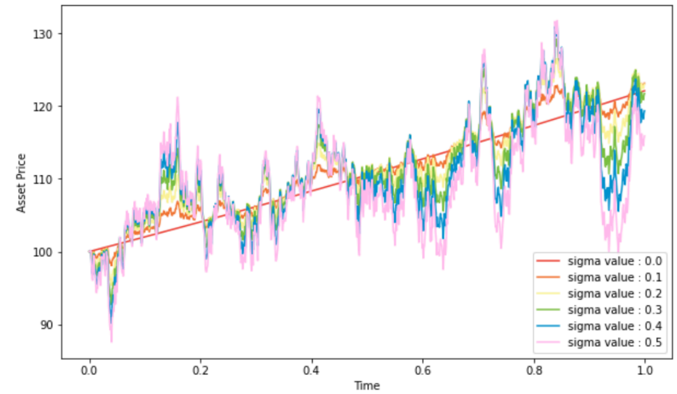
μ=-0.01 and σ = 0.05

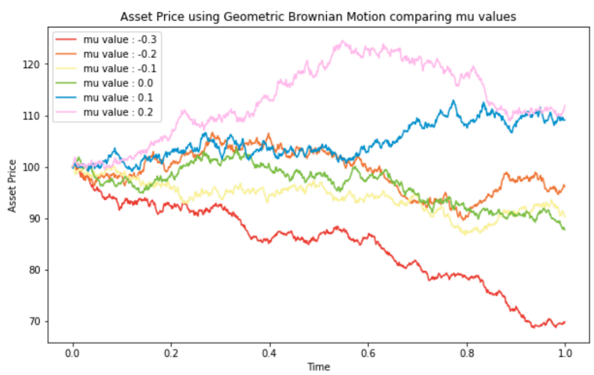


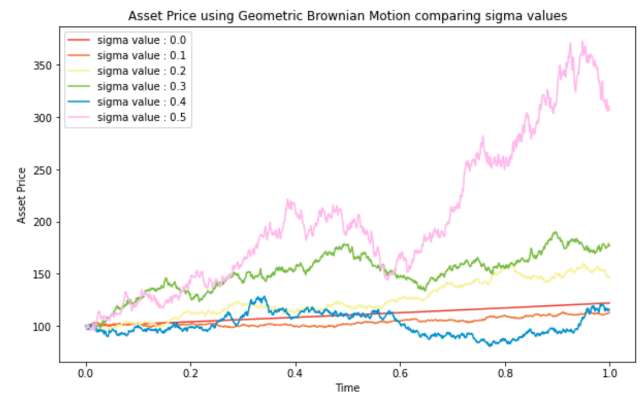
μ=-0.2 and σ = 0.1

The examples hint at the fact that the drift factor determines how much the asset price changes within one time period while the volatility factor describes how much the graph spikes. To give a better understanding of how each of these parameters affect the overall Asset Price, they are better explored below.









From the graphs above, it becomes clear that μ determines how much an asset grows or shrinks in a single time period. For example, a μ = 0 resulted in the Asset Price staying relatively around its original price while μ=0.2 managed to grow about 20% and μ=-0.3 managed to decrease its original value by around 30%.

Patterns can also be observed in σ which represents an asset's underlying volatility. Because the volatility ends up getting squared in as part of the dW term, there was no reason to compare negative values. As such, all of the sigma values here are positive.

The general trend is that the higher the σ term, the more volatile the graphs tend to get and more unpredictable the overall Asset Price is. This in turn means that if the σ value is high enough, it could dominate the μ term and lead us to believe a false trend when there is not one.

<u>Vasicek Model</u>

While Geometric Brownian Motion (GBM) is widely used to model asset prices—particularly stocks—it has limitations when applied to interest rates, especially because GBM allows values to grow indefinitely and cannot go negative. In practice, interest rates are mean-reverting (they tend to fluctuate around a long-term average) and must be modeled differently.

This leads us to the Vasicek model, which was one of the first interest rate models to introduce mean reversion into a stochastic framework.

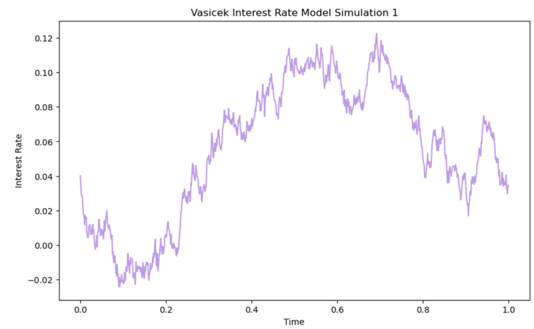
Whereas GBM assumes that the percentage change in an asset's value is proportional to its current level and influenced by random noise, the Vasicek model modifies this by having the process drift toward a long-term average rate. This makes it more realistic for modeling short-term interest rates, which typically don’t wander off to infinity or drop below zero in the same way that GBM might allow.

Thus, the transition from GBM to the Vasicek model represents a shift from modeling exponential, unconstrained growth with randomness (ideal for stock prices) to modeling bounded, mean-reverting behavior with randomness—critical for modeling interest rates, bond yields, and other financial quantities that behave differently from equities.

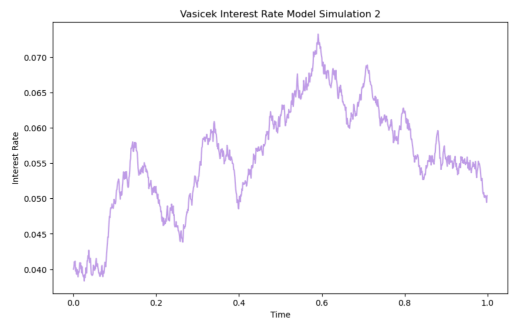
The standard equation of the Vasicek model is:

$$dr = a(b - r)dt + \sigma dW$$

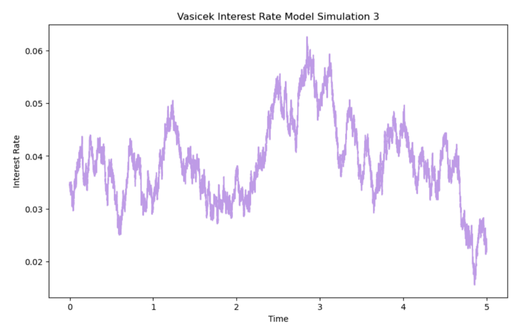
Where a is speed of mean reversion which how quickly the interest rate *r* tends to move back toward the long-term average *b*. The long term average b is the target level or "equilibrium rate" that the interest rate is drawn toward over time. The larger the gap between the current rate *r* and b, the stronger the drift back toward *b*. The following are simulations of how the model behaves given different constraints of a, b, and σ.



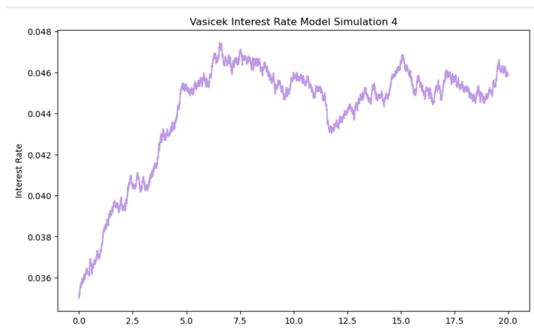
a=0.1, b=0.041, σ=0.1, initial=0.04



a=0.1, b=0.041, σ=0.02, initial=0.04



a=0.3, b=0.045, σ=0.1, initial=0.04



a=0.3, b=0.045, σ=0.001, initial=0.035

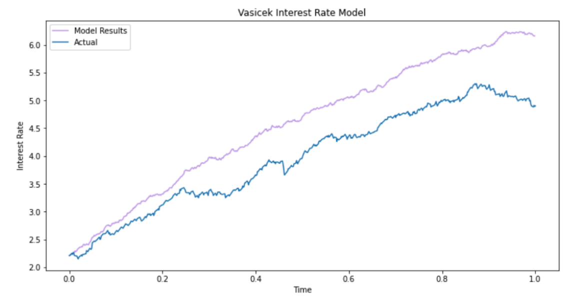
Now that we've introduced the Vasicek model as a framework for modeling interest rates with mean-reverting behavior, the next step is to apply this model to real-world data. To do that effectively, we need a dataset that reflects the kind of interest rate dynamics the Vasicek model is designed to capture—rates that fluctuate over time but tend to revert toward a long-term average.

A highly relevant and widely used dataset for this purpose is the U.S. Treasury bond interest rates, which provide historical yields across various maturities and economic conditions. These rates are ideal for calibrating and testing the Vasicek model because they span multiple decades, monetary regimes, inflation cycles, and central bank policies. Specifically, we'll be using U.S. Treasury interest rate data from the 1960s to 2025, which offers a rich, long-term view of how interest rates evolve over time.

By analyzing this dataset, we can estimate the Vasicek model's parameters—such as the average rate (b), the speed of mean reversion (a), and volatility (σ)—and evaluate how well the model fits actual market behavior. This transition from theory to data grounds the model in empirical reality and allows for both simulation and forecasting of future interest rate movements.

By applying least squares regression to the observed changes in interest rates against the current rates, it is possible to obtain parameters to calibrate the Vasicek Model.

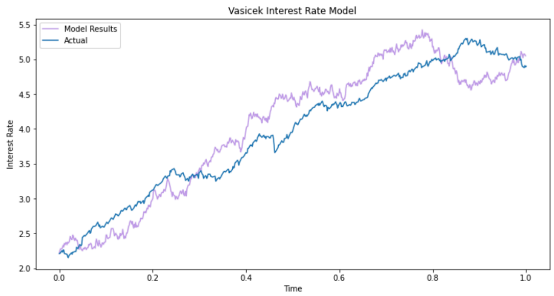
The first attempt at this is shown below where I took a random sample of 500 data points and performed a least squares regression[2].



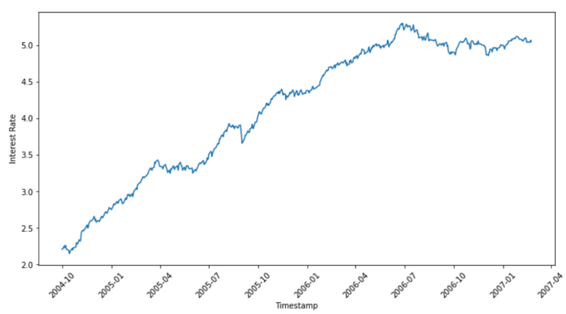
I believe the regression was relatively successful. However something interesting is happening when comparing the model and the actual graph. The regression for σ seems off in the sense that the actual graph displays more volatility than the one predicted by the model. Originally, the graph seemed quite smooth for an interest rate curve. After increasing the volatility of the model by an order of magnitude, it has given a result that somewhat resembles our original interest rate graph. Thus, I believe there needs to be a correctional factor for the parameter sigma. Additionally, the regression believes that the actual parameter for "long-term mean" or the b-term is much higher than displayed. All of this is visible in the graph above

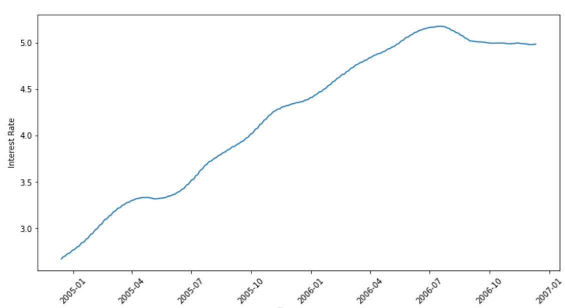
After adding some correctional factors where I increase the σ value by a factor of 3 and decrease the value for b by 25%, a new more accurate graph is achieved below.

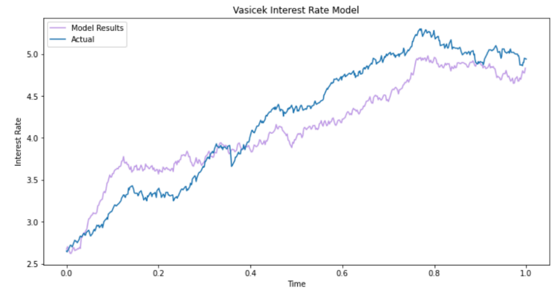
[2] When doing the regression, I actually ended up calculating the speed of mean reversion, then the speed of mean reversion times the long-term mean, and finally the volatility. The reason is that least squares regression does not let me isolate all of the parameters. However, conveniently, these parameters can be isolated after the regression.



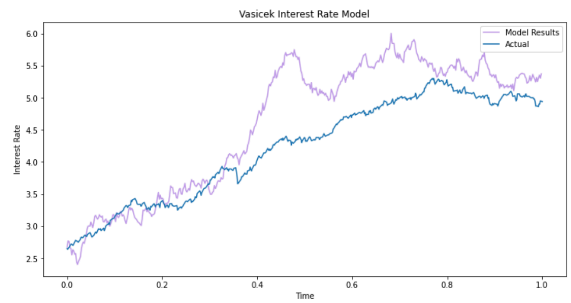
The question became how to capture a more accurate representation of interest rates. The method that I used was to first smooth the graph to remove any artifacts from the volatility and capture better values for "a" and "b". Then, the average mean absolute error between the smooth graph and actual graph would capture the volatility. Below is the actual graph of the 500 data points, the smoothened graph as well the new model against the original graph.





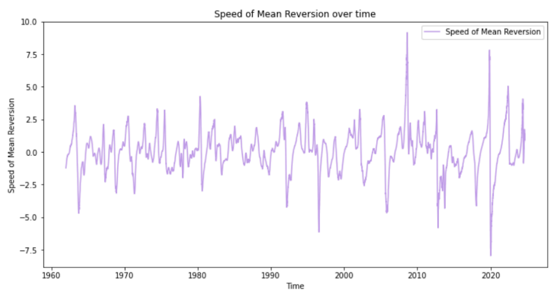


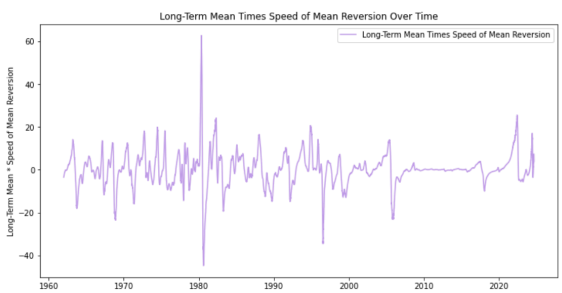
As it can be seen, the new model with these changes performs excellently. However, because of the nature of Brownian Motion, the model does not always produce results this accurate. An example of a different random sample and the model results are shown below.

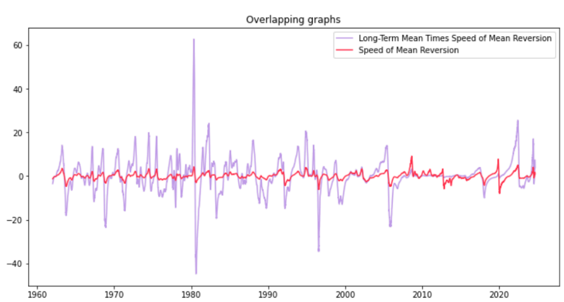


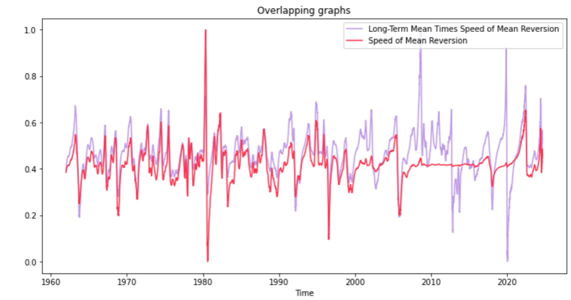
While estimating Vasicek parameters from a single regression provides a snapshot of interest rate dynamics over the entire dataset, it assumes that the underlying behavior of rates—such as the mean reversion speed, long-term average, and volatility—has remained constant over time. However, in reality, interest rate regimes have shifted significantly from the 1960s to 2025, influenced by changes in monetary policy, inflation expectations, and macroeconomic conditions. To capture this evolution, I will extend the analysis by applying the regression method over sequential time intervals, allowing me to observe how the parameters a, b, and σ have changed across decades. This dynamic approach will provide deeper insight into how market behavior and central bank policy have influenced interest rate movements over time, revealing periods of high volatility, strong mean reversion, or structural shifts in long-term rate expectations.

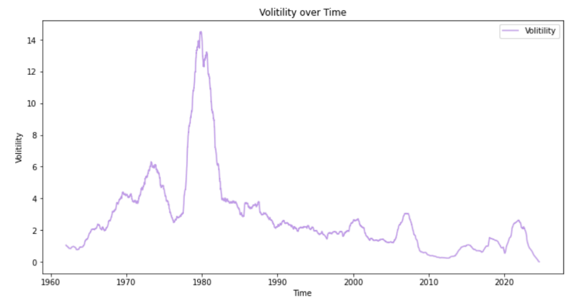
Below are the parameters measured from the one-year treasury bonds from the 1960s to 2025. Additionally, the graphs are overlaid on top of each other and normalized to better illustrate the trends present.











As mentioned in one of the footnotes, I could not isolate the “a” and the “b” when doing the regression leading to the graphs seen above. I could not separate the two values after the regression because there are a number of times that resulted in a division by 0 error. The coincidental part of this is that both of them tend to go to 0 at the same time making me believe that there exists some numeric limit. However, the graphs lead to some interesting conclusions.

These values do not remain constant over time but instead follow repeating patterns or oscillations that correspond to broader economic cycles. For instance, during periods of economic expansion or high inflation, central banks may raise rates, leading to an increase in the estimated long-term mean b. Conversely, during recessions or low-inflation environments, b might fall, reflecting market expectations of lower average rates. Similarly, the mean reversion speed "*a*" could vary with policy regimes: it might be higher during aggressive monetary policy periods (when central banks respond quickly to deviations from target rates) and lower during more passive or uncertain periods. These patterns suggest that the rate-setting environment and the market's perception of interest rate stability change over time in a way that aligns with macroeconomic trends.

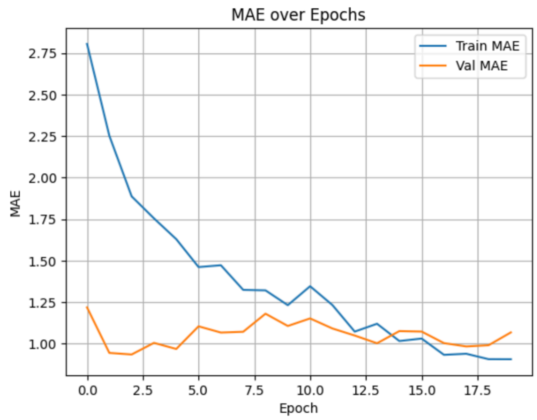
Understanding this cyclicality is important because it highlights the non-stationary nature of financial markets—the assumption that model parameters are constant may be overly simplistic for long-term interest rate modeling. Instead, this finding supports the idea that interest rate behavior is influenced by structural shifts in the economy, such as inflation cycles, fiscal policy, geopolitical events, and changes in central banking philosophy. By tracking how they evolve, we gain insight not only into the technical behavior of rates but also into the broader economic environment across decades. This kind of analysis is especially useful for risk management, forecasting, and macroeconomic modeling, where understanding how quickly markets revert to normal and what 'normal' even means over time is critical.

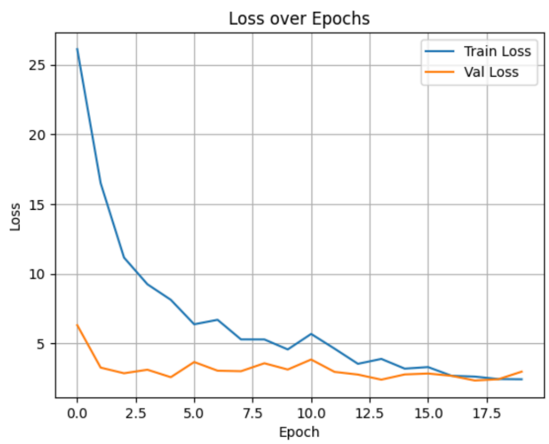
Machine Learning

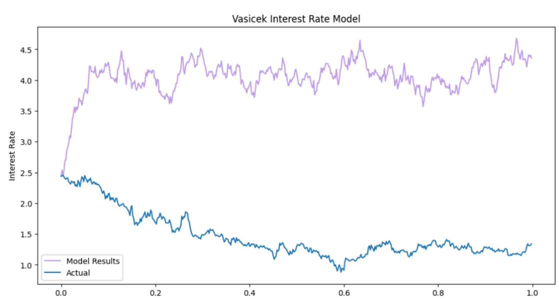
Recognizing that the parameters of the Vasicek model exhibiting cyclical behavior over time raised a key question: could these cycles be predicted or characterized more effectively? While traditional regression provides a useful historical lens, it may not fully capture the nonlinear relationships and structural shifts in the economy that influence interest rates. To explore this further, I turned to machine learning as a tool to uncover deeper patterns and possibly learn the mapping between economic conditions and Vasicek parameters without relying on strict model assumptions.

To begin exploring whether a machine learning model could predict the Vasicek parameters more effectively than traditional regression, I implemented a deep learning model using TensorFlow. The model was designed to take in a window of 500 consecutive interest rate observations as input, under the assumption that a sufficiently long history of rate movements might contain enough information to infer the evolving behavior of the Vasicek parameters a, b, and σ.

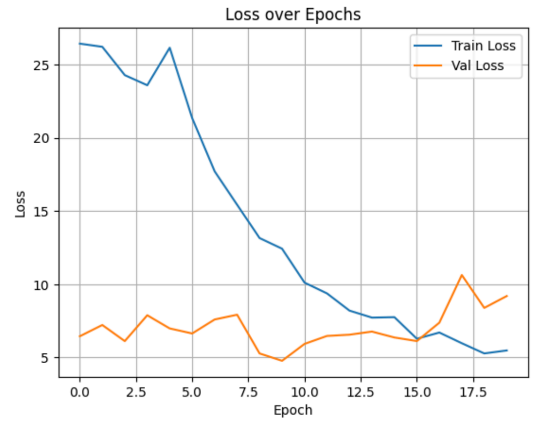
The architecture of the neural network was straightforward. I used a sequence of fully connected (dense) layers with 256, 128, and 64 neurons, respectively. Each layer applied the ReLU (Rectified Linear Unit) activation function to introduce non-linearity, allowing the model to capture more complex relationships between historical rates and parameter behavior. The final output layer produced three values, corresponding to the estimates for the parameters. I trained the model using mean squared error (MSE) as the loss function, comparing its predictions against the parameters calculated from regressions over the same data. This setup aimed to teach the model how to internalize the structure of interest rate dynamics and generalize those insights into more flexible parameter estimation. The following graphs show the loss over time as well as the final model results.

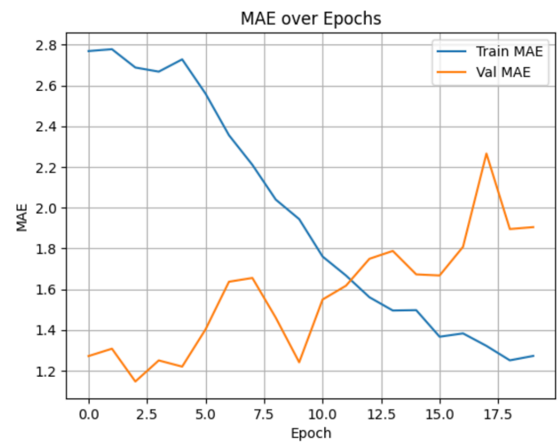


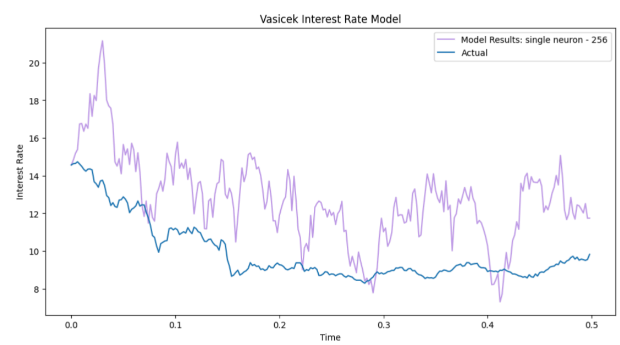




However, despite the depth and complexity of the initial deep learning model, it did not successfully predict the Vasicek parameters with sufficient accuracy or stability. As a result, I decided to go back to the drawing board and simplify the model to improve generalization. I reduced the architecture to just one dense layer with 256 neurons. On the side are the results of the new model.







The simplified model with just one dense layer of 256 neurons performed noticeably better than the deeper architecture.

Conclusion

This journey of exploring Interest Rates through the Vasicek Model has been incredibly rewarding granting me the opportunity to learn a tremendous amount not only about interest rates but also probability and expectation which can be used almost anywhere. This project also marked a significant expansion of my coding skills, particularly in Python and the use of Jupyter Notebooks. As I moved from theoretical understanding to hands-on implementation, I learned how to structure experiments, process and visualize large datasets, and efficiently manage rolling window computations for regression. Jupyter Notebooks proved especially valuable as an environment for exploratory analysis.

Sources:

- https://www.quantstart.com/articles/The-Markov-and-Martingale-Properties/
- https://www.quantstart.com/articles/Geometric-Brownian-Motion/
- https://www.quantstart.com/articles/Brownian-Motion-and-the-Wiener-Process/
- https://en.wikipedia.org/wiki/Wiener_process
- Book: Stochastic Calculus for Finance II by Steven E. Shreve
- https://www.youtube.com/watch?v=6A1-yXwE7CE
- Efficient Methods for Valuing Interest Rate Derivatives
- Stochastic Calculus for Finance II, Continuous-Time Models
- https://en.wikipedia.org/wiki/Ornstein%E2%80%93Uhlenbeck_process