Predicting Volatility Using Autoregressive and Regression Models

Cindy Xu and Rohan Shukla

GitHub : https://github.com/rohanshukla5/time-series-analysis

1 Project Overview

Our project intends to use statistical models to predict the volatility of the stock market using two different measures of volatility: the Chicago Board Option Exchange's (CBOE) Volatility Index (VIX) and the actual implied volatility of the S&P500 option. VIX is a popular measure of the expected volatility of the stock market in 30 days based on the S&P500 option which can be found as a ticker symbol under the same name. Similarly, the implied volatility is a measure of the future expected volatility of an asset.

We will fit separate autoregressive models for predicting VIX and for predicting the implied volatility and draw comparisons between the two. For predicting VIX, we will fit a seasonal autoregressive integrated moving average (SARIMA) model and for the implied volatility, we will use a generalized autoregressive conditional heteroskedasticity (GARCH) model.

We also wish to study the relationship between the two by fitting regression models where the implied volatility is the predictor variable and the VIX is the response. We want to examine whether we can predict the VIX by solely regressing on the implied volatility of the S&P500 and if so, which model is able to make the best predictions. We will fit regression models that don't take the time dependency into consideration to see whether time plays a role in the relationship between the VIX and the implied volatility. Some of these models include ordinary linear regression models, generalized additive models (GAM), and kernel regression.

Note: We are still trying to determine how exactly to compute the implied volatility, so as of right now, we are instead using the realized (historical) volatility in its place. This is a measure of the volatility of an asset over a period of time that has already passed. In other words, it summarizes the volatility that has already occurred while the implied volatility attempts to make predictions about future volatility.

2 Autoregressive Models

Autoregression is when we try to regress a time dependent variable X_t on its past values. While there are many different types of autoregressive models, we will be focusing primarily on the SARIMA and GARCH models for this project.

2.1 Generalized Autoregressive Conditional Heteroskedasticity Model

The generalized autoregressive conditional heteroskedasticity (GARCH) model is one that takes the dependency of the conditional second moment, which is also known as the variance, into consideration. If X_t is our variable of interest, then for the GARCH model, we are assuming our data is of the form

$$\begin{aligned} X_t &= \sigma_t^2 \epsilon_t \\ \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i X_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \end{aligned}$$

where $a_0 > 0$ and $a_i, b_j \geq 0$. We also have that the ϵ_t are independently and identically distributed with mean zero and variance one and each ϵ_t is independent of X_s for $s < t$.

3 Other Regression Models

Other than fitting autoregressive models to predict VIX, we also used linear regression as well as some non-parametric regression models. Non-parametric models are ones that don't assume a predetermined form of the relationship between the predictor variables and the response variable. This allows for more flexibility and complex relationships without imposing as many or any assumptions about the data. While these regression models don't take the time dependency of the data into consideration, they are still powerful tools we can use to analyze the relationship between VIX and the volatility of the S&P500 prices.

3.1 Generalized Additive Model

One of the non-parametric regression models used in this project is a generalized additive model (GAM). Recall that in a linear regression model, if we have predictor variable $X = \begin{pmatrix} X_1 & \cdots & X_n \end{pmatrix}^T$ and $\mu(x) = \mathbb{E}[Y \mid X = x]$ is the true regression function where $Y = \mu(X) + \epsilon$, then we are imposing the assumption that μ is of the form $\mu(X) = \beta_0 + \sum_{t=1}^{n} \beta_t X_t$. In other words, we are assuming that we have a linear and additive relationship between our predictors and the response variable. In a GAM, we loosen these assumptions by only assuming additivity. In other words, we are now assuming that there exists some smooth functions $f_1, \ldots, f_n$ such that $\mu(X) = \sum_{t=1}^{n} f_t(X_t)$. So while in linear regression, we are trying to estimate the coefficients β_t, in a GAM, we are instead trying to estimate the functions f_t which may or may not be linear functions. Note that this implies that a linear regression model is a specific type of GAM, but not all GAMs are of the same form of a linear regression model.

When used in practice, we aren't able to output an estimated formula or definition of $\widehat{f_t}$ the same way we are able to output an estimate of $\widehat{\beta_t}$ in linear regression. However, we are able to plot the estimated function $\widehat{f_t}$ for each variable X_t to visualize the relationship of each X_t with Y.

3.2 Kernel Regression

Another non-parametric regression model we used is kernel regression, otherwise known as kernel smoothing or Nadaraya-Watson smoothing. This is a type of linear smoother.

Linear smoothers are a popular way of estimating the true regression function. By definition, $\widehat{\mu}$ is a linear smoother that estimates μ if there exists a function $w : \mathbb{R}^2 \to \mathbb{R}$ such that $\widehat{\mu}(x) = \sum_{i=1}^{n} y_i w(x_i, x)$. In other words, $\widehat{\mu}$ is estimating μ by taking a weighted average of the observed data and w is known as the weight function. Ordinary linear regression is a famous example of a linear smoother where

$$w(x_i, x) = \frac{1}{n} + \frac{(x_i - \bar{x})(x - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

Kernel regression is another example of a linear smoother where we define its weight function to be

$$w(x_i, x) = \frac{K(\frac{x - x_i}{h})}{\sum_{j=1}^{n} K(\frac{x - x_j}{h})}$$

Here, $h > 0$ is the bandwidth which helps scales the distance $x_i - x$ that we are inputting and K is some kernel function that satisfies the following properties:

1. $K(u) \geq 0$ for all $u \in \mathbb{R}$
2. $\lim_{|u| \to \infty} K(u) = 0$

In practice, K is typically chosen to be a probability density function (PDF). The most commonly used kernel function is the standard Gaussian PDF, which is

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Note that for kernel smoothing, we don't make any assumptions about the data.

References

[1] V. Lyubchich and Y. R. Gel. *Time Series Analysis: Lecture Notes with Examples in R*. 2023-09 edition, 2023.

[2] C. R. Shalizi. *Advanced Data Analysis from an Elementary Point of View*. 2025.