

# Regressions on VIX

Last Updated: 2025-04-17

## How to Run This File

For the R Markdown file, it is recommended to be opened using RStudio with the latest version of R.

This file uses the `tidyverse` package in R, which needs to be installed in the Console using the command `install.packages("tidyverse")` prior to running the following code that loads the package into this session.

```
library(tidyverse)
```

This file also uses the `tidyquant` package in R for retrieving data about stock prices from Yahoo Finance. Again, this needs to be installed in the Console using the command `install.packages("tidyquant")` prior to running the following code that loads the package into this session.

```
library(tidyquant)
```

This also uses the `zoo` package in R for dealing with time series. Again, use `install.packages("zoo")` prior to running the following code.

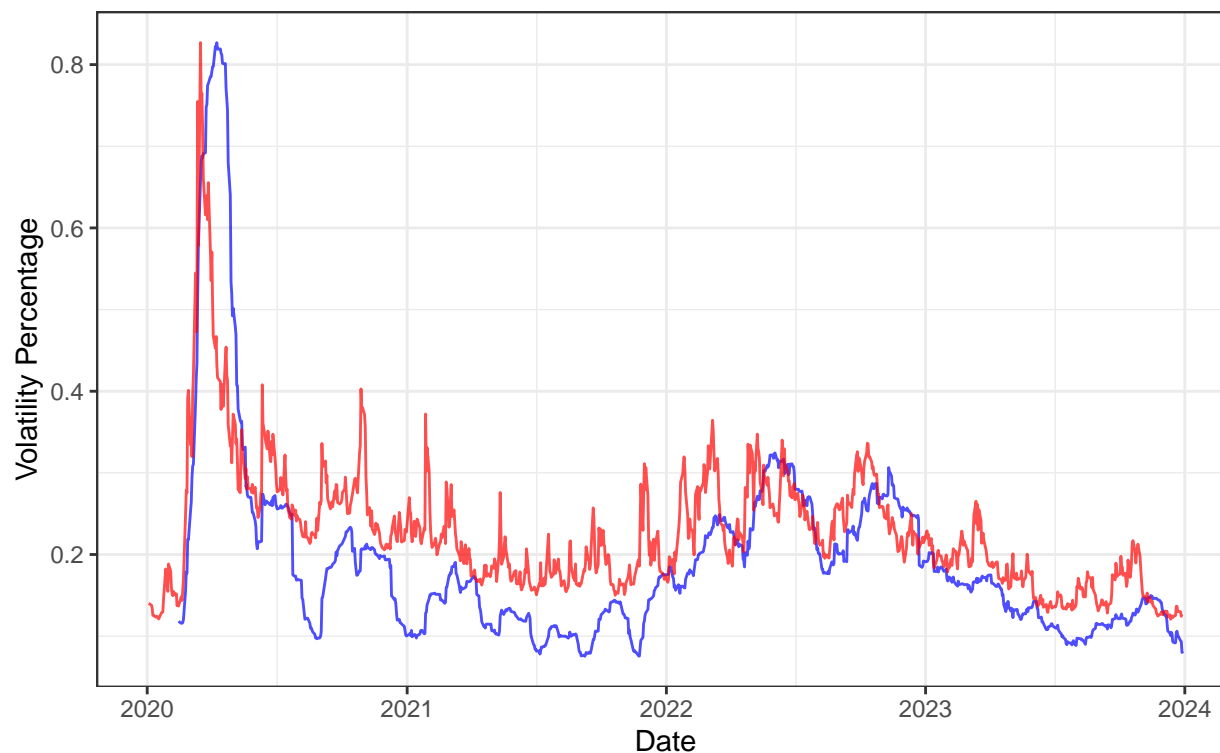
```
library(zoo)
```

To see any documentation about the built-in functions used, you can use the `help()` command in the Console in RStudio.

The following code and graphs are about the relationship between the realized volatility of the S&P500 and the VIX index. The realized volatility, also known as the historical volatility, measures how much a stock changed in the past, while the VIX attempts to measure how much a stock will change in the future.

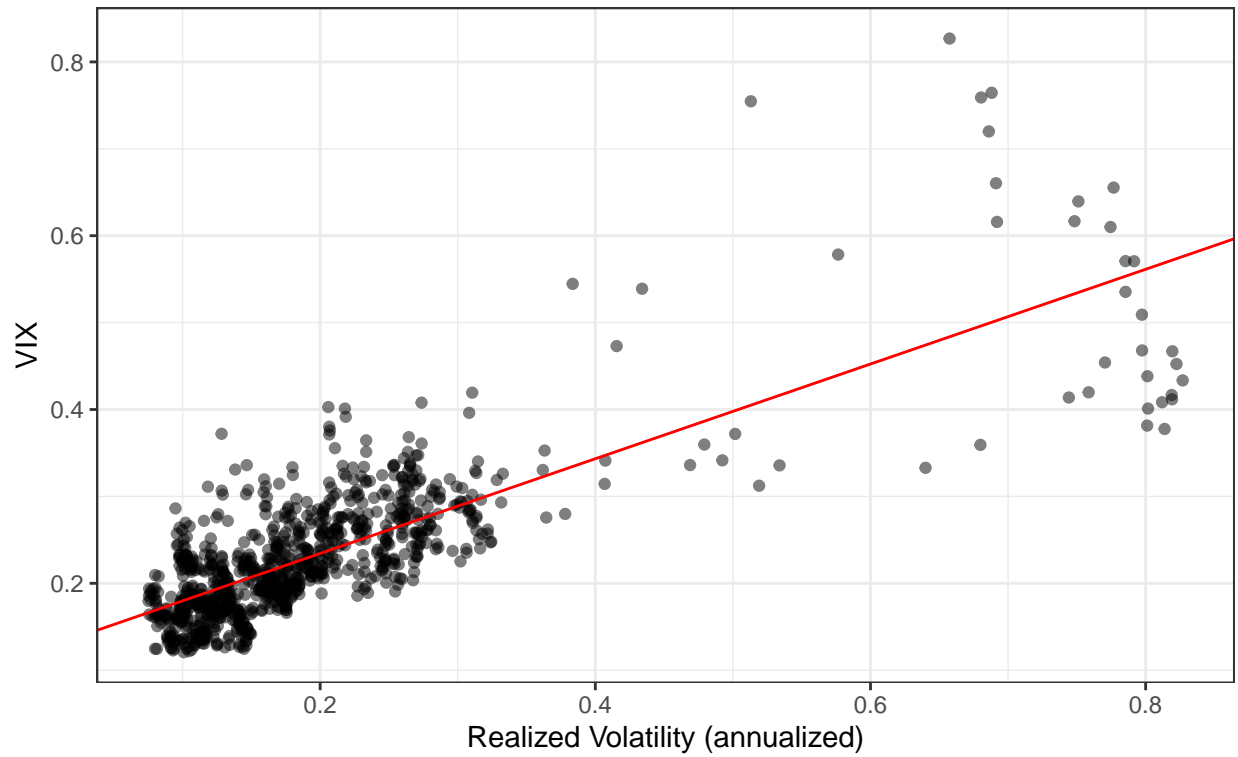
## Realized Volatility of SPY and VIX Over Time

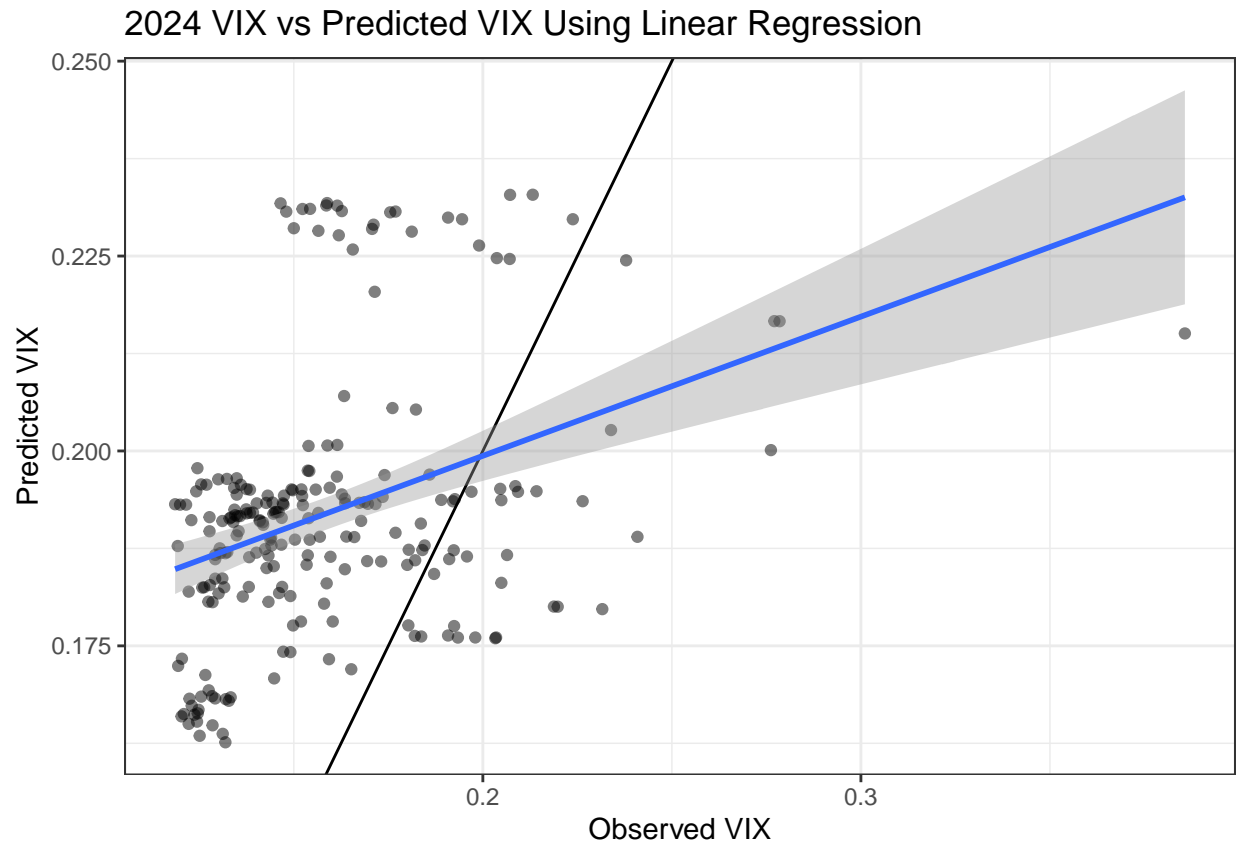
The blue marks the realized volatility, while red marks the VIX index



## VIX vs Realized Volatility

The red line denotes the line-of-best-fit from fitting a linear model





If the predicted value matches the actual observed, then the data would appear centered around the black line in the plot. However, when we fit a linear regression of the predicted values on the observed (i.e. find the line-of-best-fit) and plot it as the blue line, we do see a difference between the two plotted lines. Something to keep in mind is that we do see some outliers on the right where we see a drastically different VIX value from the rest of the data which heavily influences the fitting of the blue line. Regardless, we can see from observing the data points that they don't appear to be centered around the black line. It appears that the linear model tended to over-estimate the VIX index since most points lied above the black line.