Regressions on VIX

Last Updated: 2025-04-17

How to Run The R Markdown (RMD) File

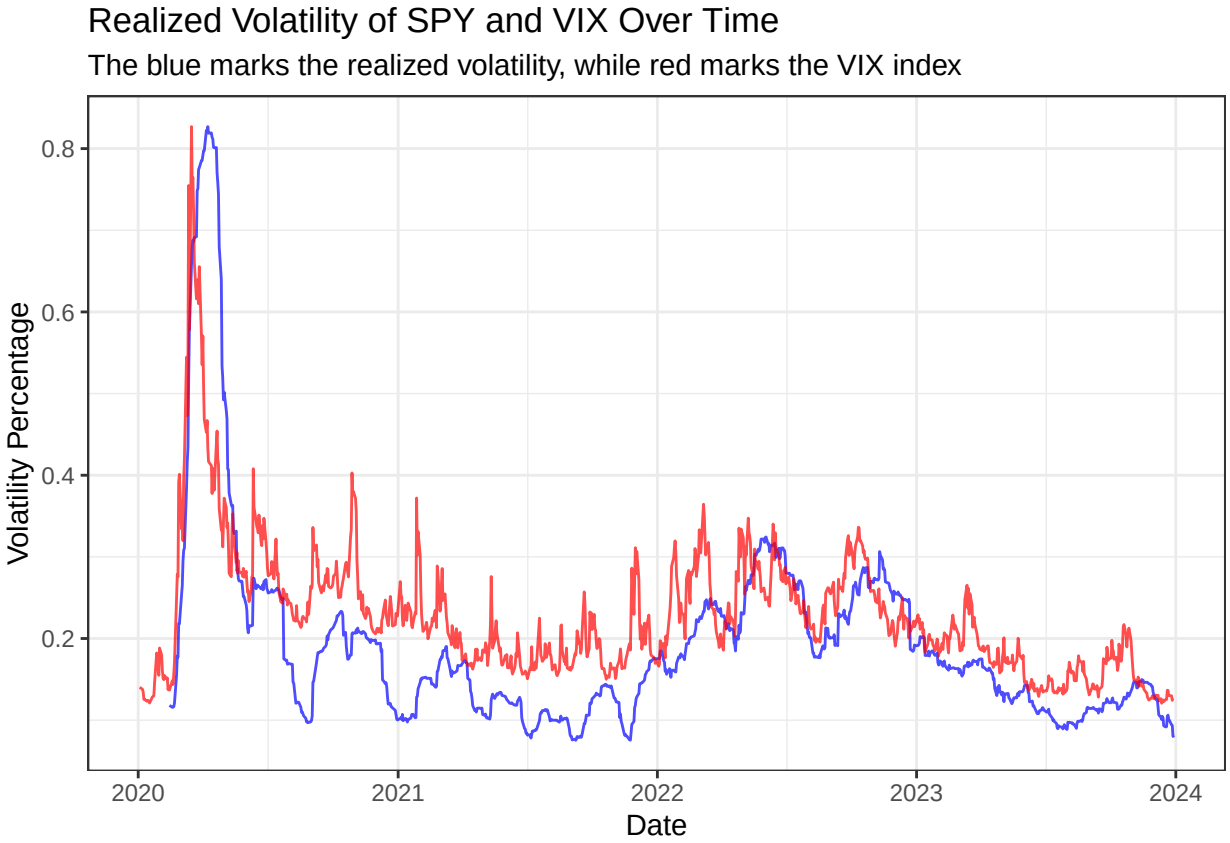
For the R Markdown file, it is recommended to be opened using RStudio with the latest version of R.

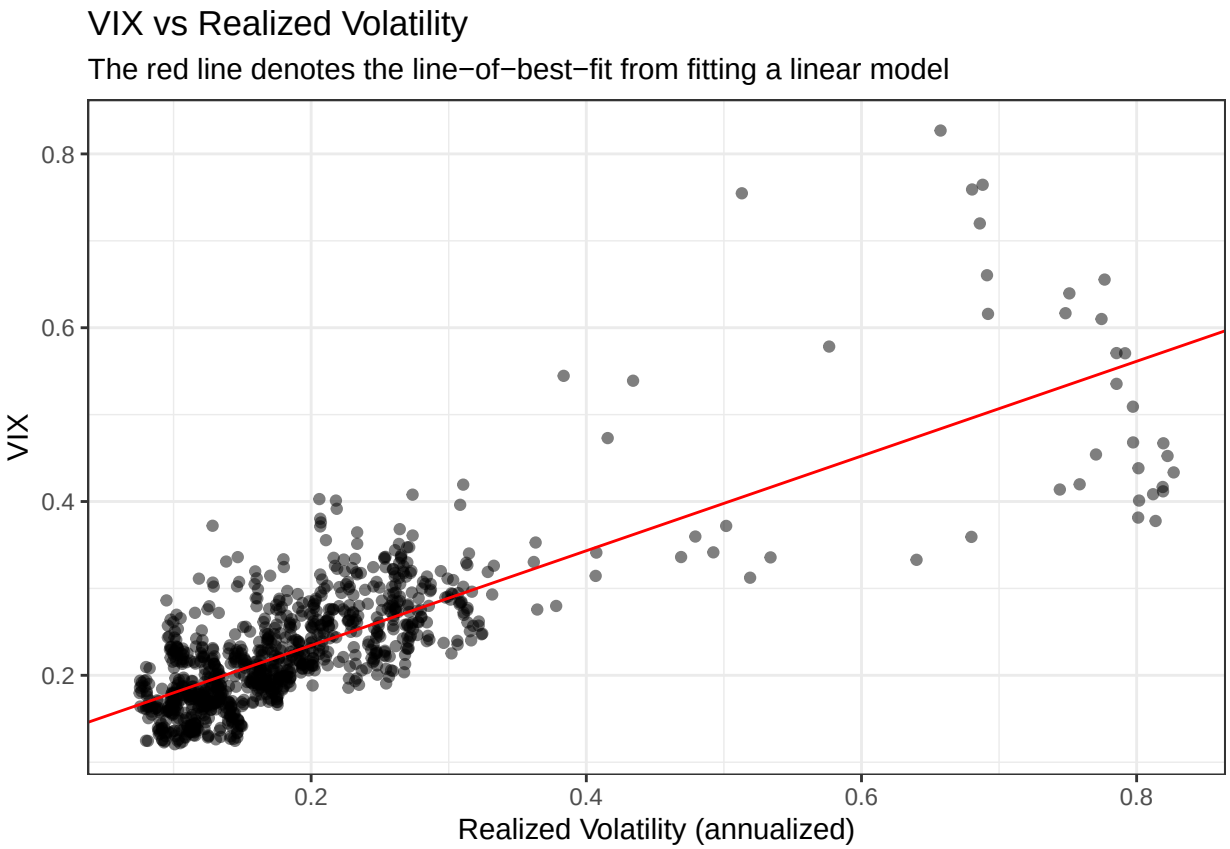
Each of the following items listed is an R package that needs to be installed by running the command `install.packages("package-name")` in the Console before running the file. The commands `library(package-name)` in the RMD file then loads the package into the session.

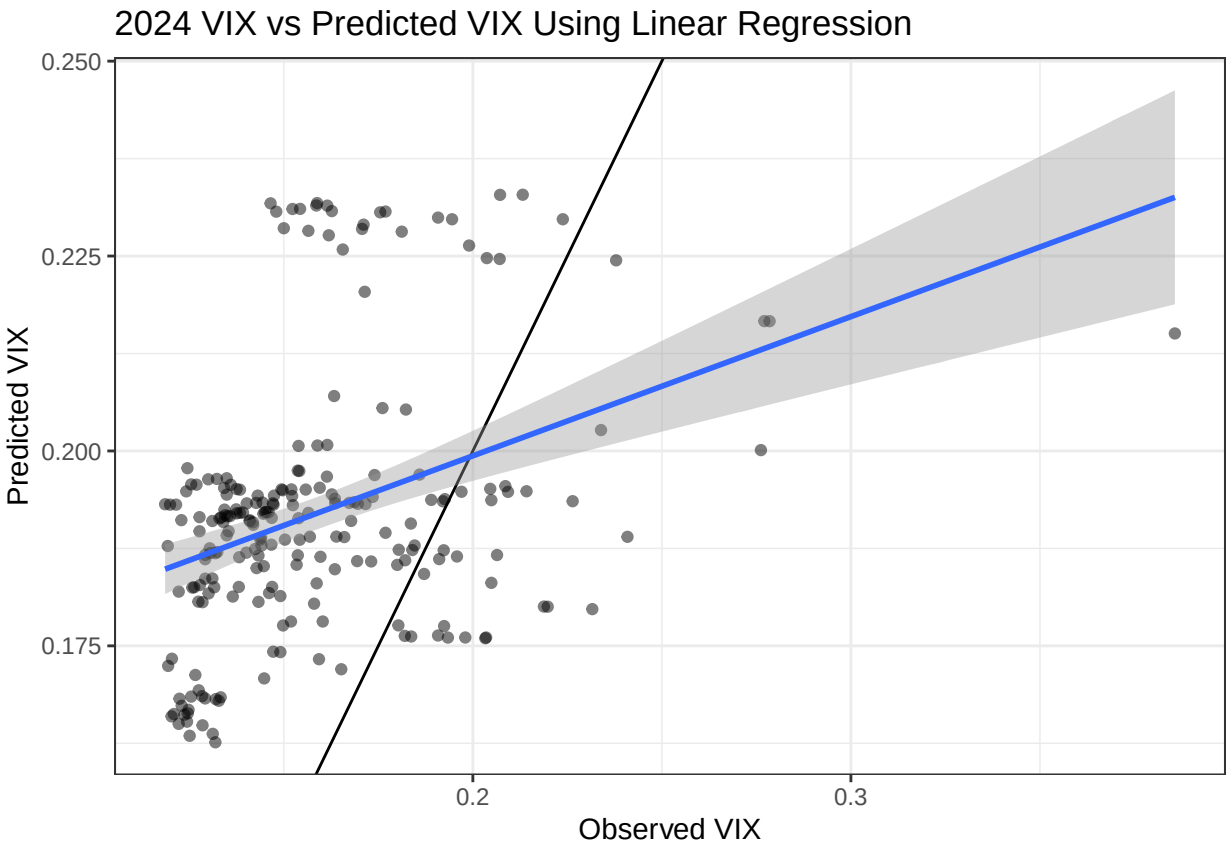
- **tidyverse**: for nice data transformation functions and for making clean plots
- **tidyquant**: for retrieving data about stock prices from Yahoo Finance easily
- **zoo**: for dealing with time series data (ex. rolling averages)
- **np**: for fitting kernel regressions
- **mgcv**: for fitting generalized additive models

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

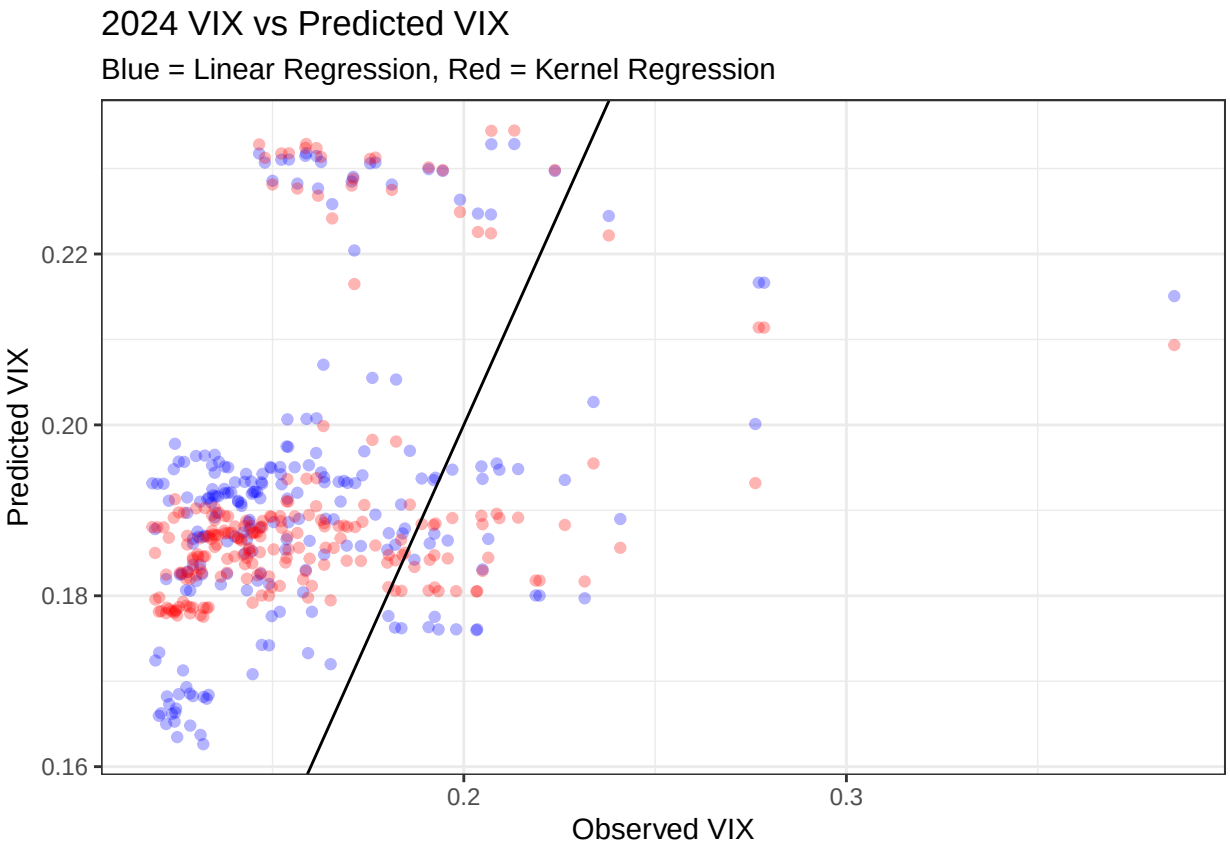
The following code and graphs are about the relationship between the realized volatility of the S&P500 and the VIX index. The realized volatility, also known as the historical volatility, measures how much a stock changed in the past, while the VIX attempts to measure how much a stock will change in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

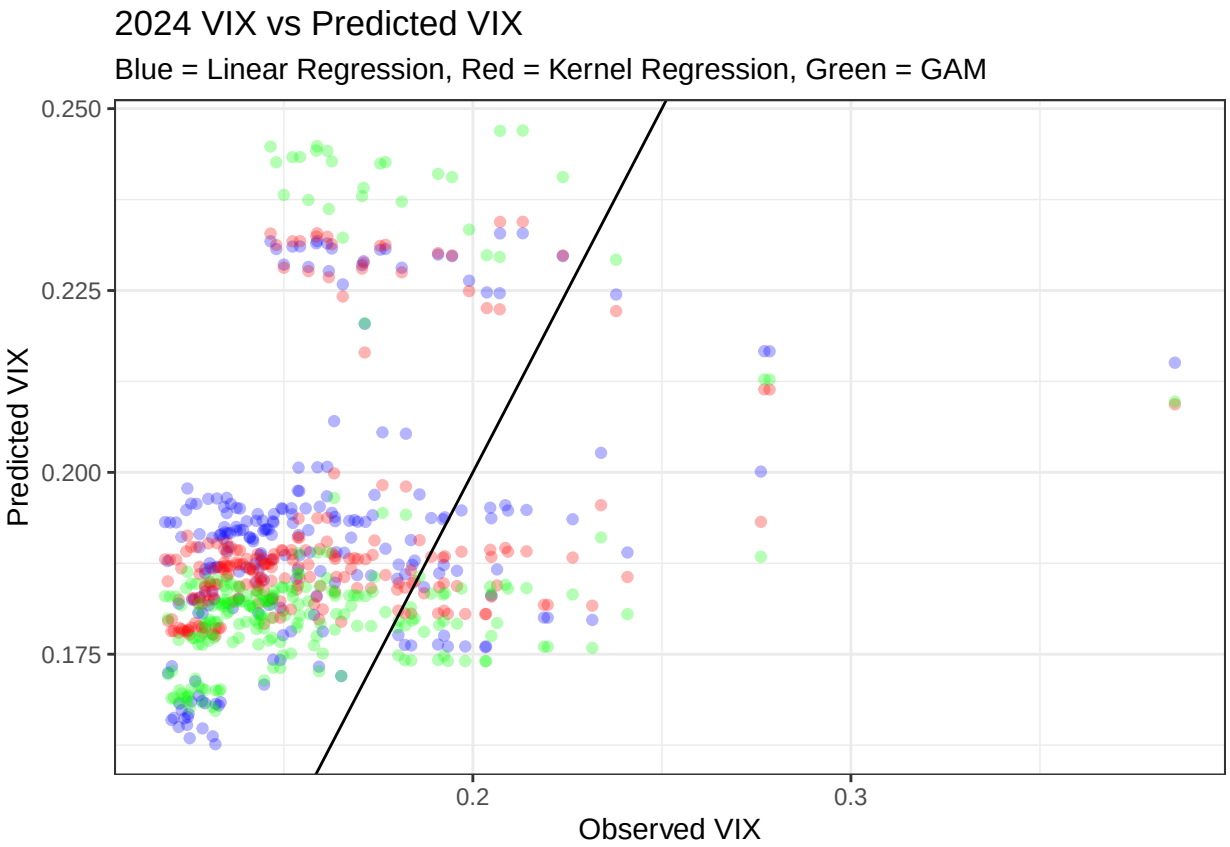



VIX vs Realized Volatility
The red line denotes the line−of−best−fit from fitting a linear model
VIX
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
Realized Volatility (annualized)



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.


2024 VIX vs Predicted VIX
Blue = Linear Regression, Red = Kernel Regression
Predicted VIX
0.22
0.20
0.18
0.16
0.2
0.3
Observed VIX


2024 VIX vs Predicted VIX
Blue = Linear Regression, Red = Kernel Regression, Green = GAM
Predicted VIX
0.250
0.225
0.200
0.175
0.2
0.3
Observed VIX