

# Fitting Models of VIX on Realized Volatility

Using Cross Validation

Last Updated: 2025-04-22

## 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.

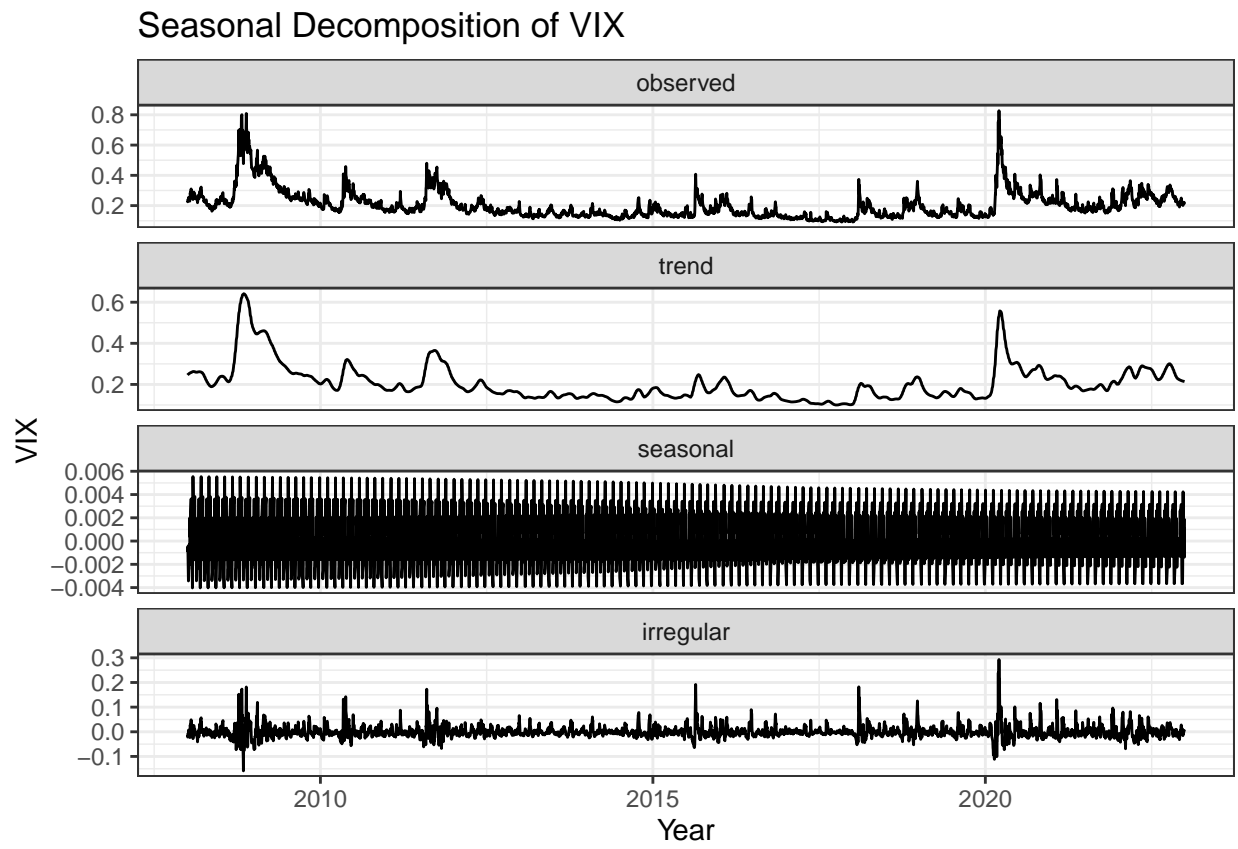
- **tidyverse**: for nice data transformation functions and for making clean plots
- **tidyquant**: for retrieving data about stock prices from Yahoo Finance easily
- **ggseas**: for plotting seasonal decomposition
- **zoo**: for dealing with time series data (ex. rolling averages)
- **np**: for fitting 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.

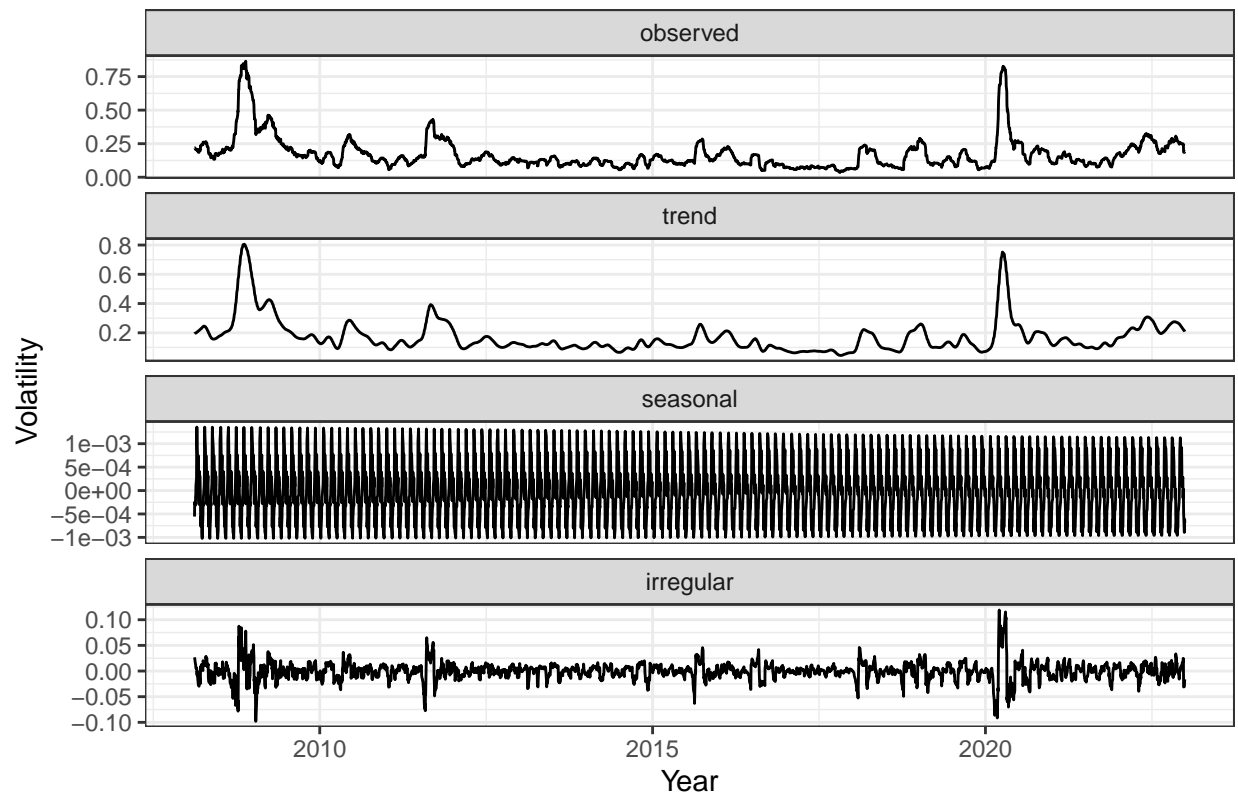
## About This File & Code

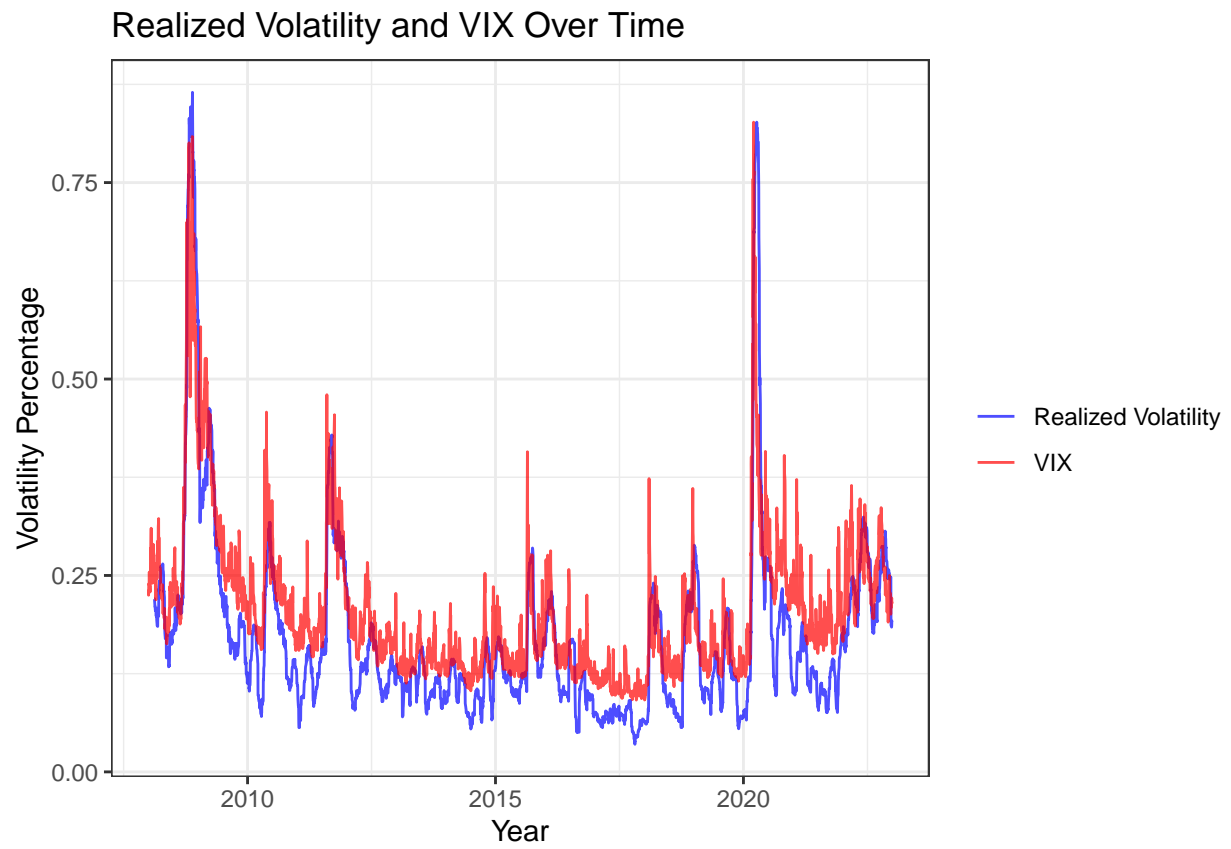
Here, we study the relationship between the VIX index and the realized volatility of the S&P500 (SPY) stock prices by fitting regression models where the realized volatility is the predictor variable and the VIX is the response. We want to examine whether we can predict the VIX by solely regressing on the realized 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 realized volatility. Some of these models include ordinary linear regression models, generalized additive models (GAM), and kernel regression. Note that when fitting these models, we will be using cross validation to minimize the root mean squared error (RMSE) of the model.

## The Data

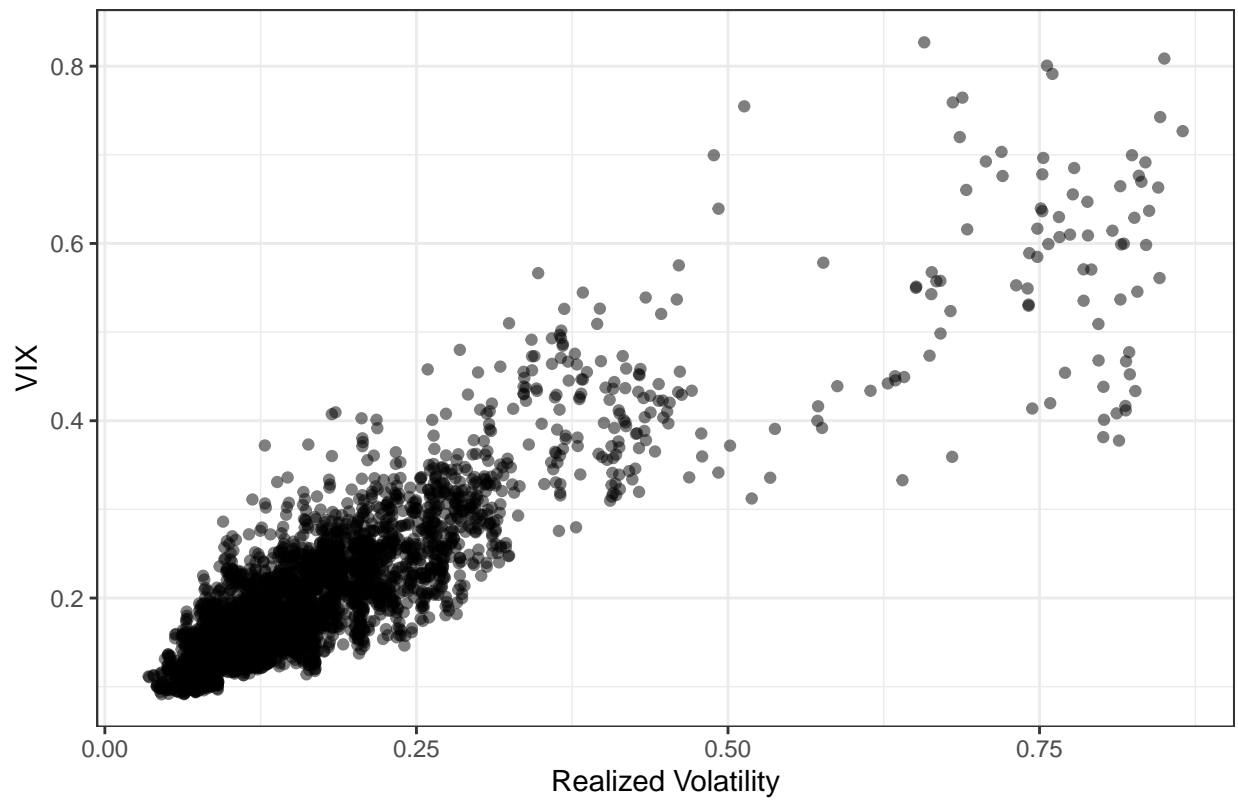


## Seasonal Decomposition of Realized Volatility





VIX vs Realized Volatility



## Fitting the Models

The following is a table of the cross-validated root mean squared errors (RMSE) from each of the models fitted:

Model	Cross Validated RMSE
Linear Model	0.04522
Kernel	0.0431
GAM	0.04517