Fitting Models of VIX on Realized Volatility

Using Cross Validation

Last Updated: 2025-04-28

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.

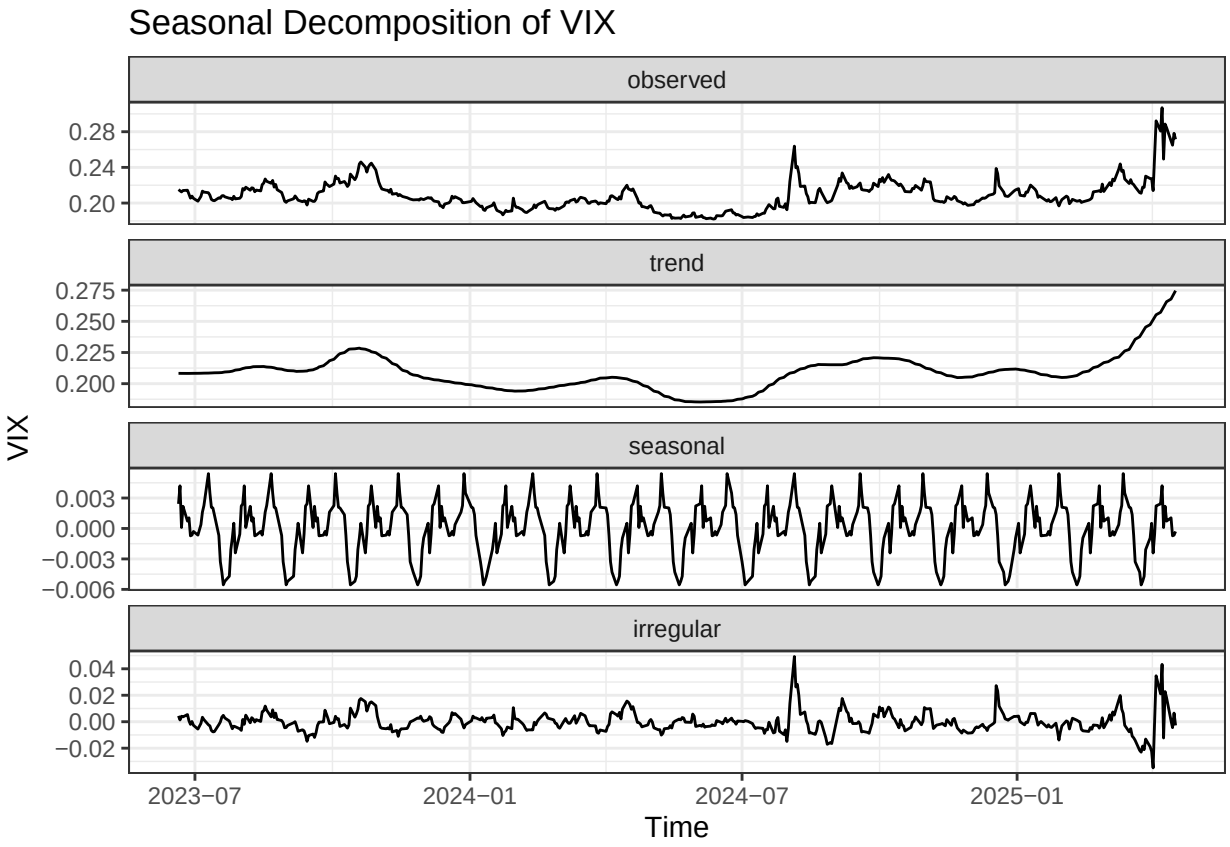
- **tidyverse**: for nice data transformation functions and for making clean plots
- **tidyquant**: for retrieving data about stock prices from Yahoo Finance easily
- **ggseas**: for plotting seasonal decomposition
- **zoo**: for dealing with time series data (ex. rolling averages)
- **np**: for fitting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

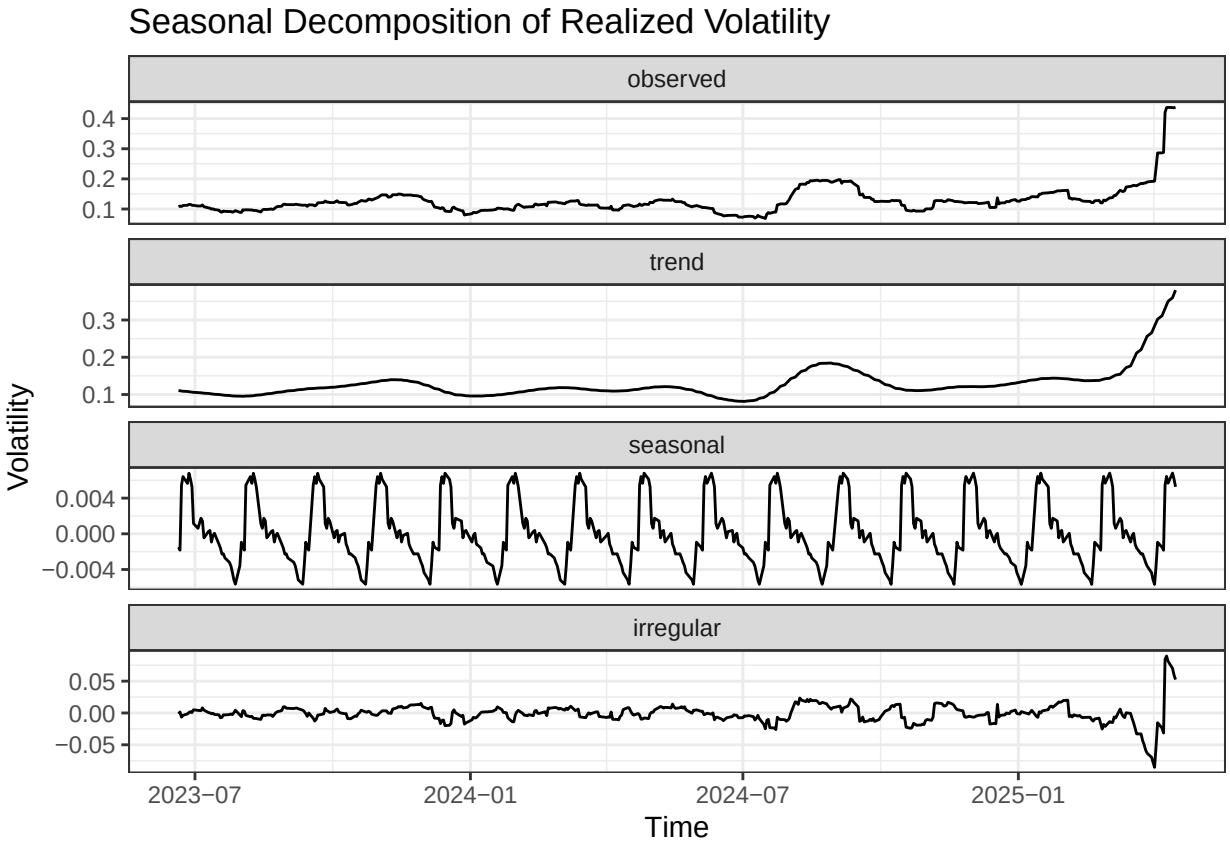
To see any documentation about the built-in functions used, you can use the `help()` command in the Console in RStudio.

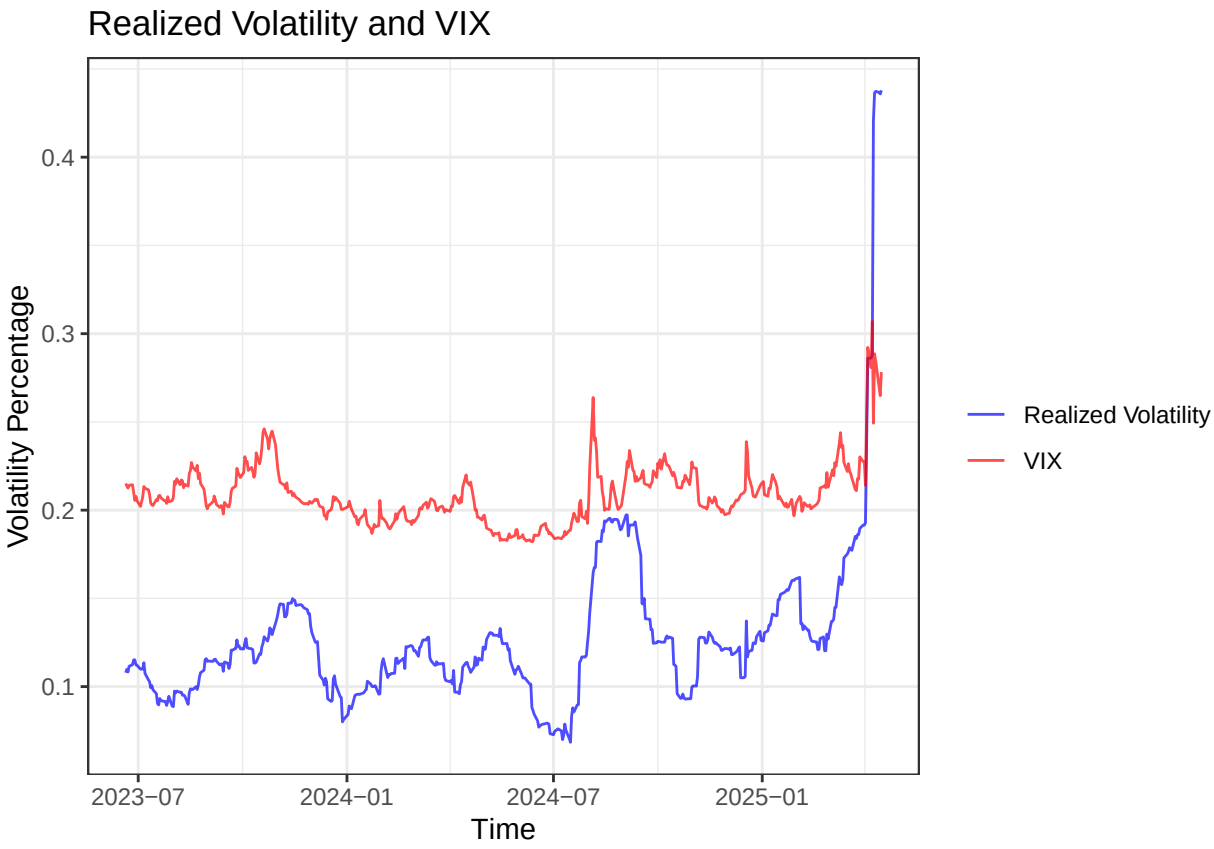
About This File & Code

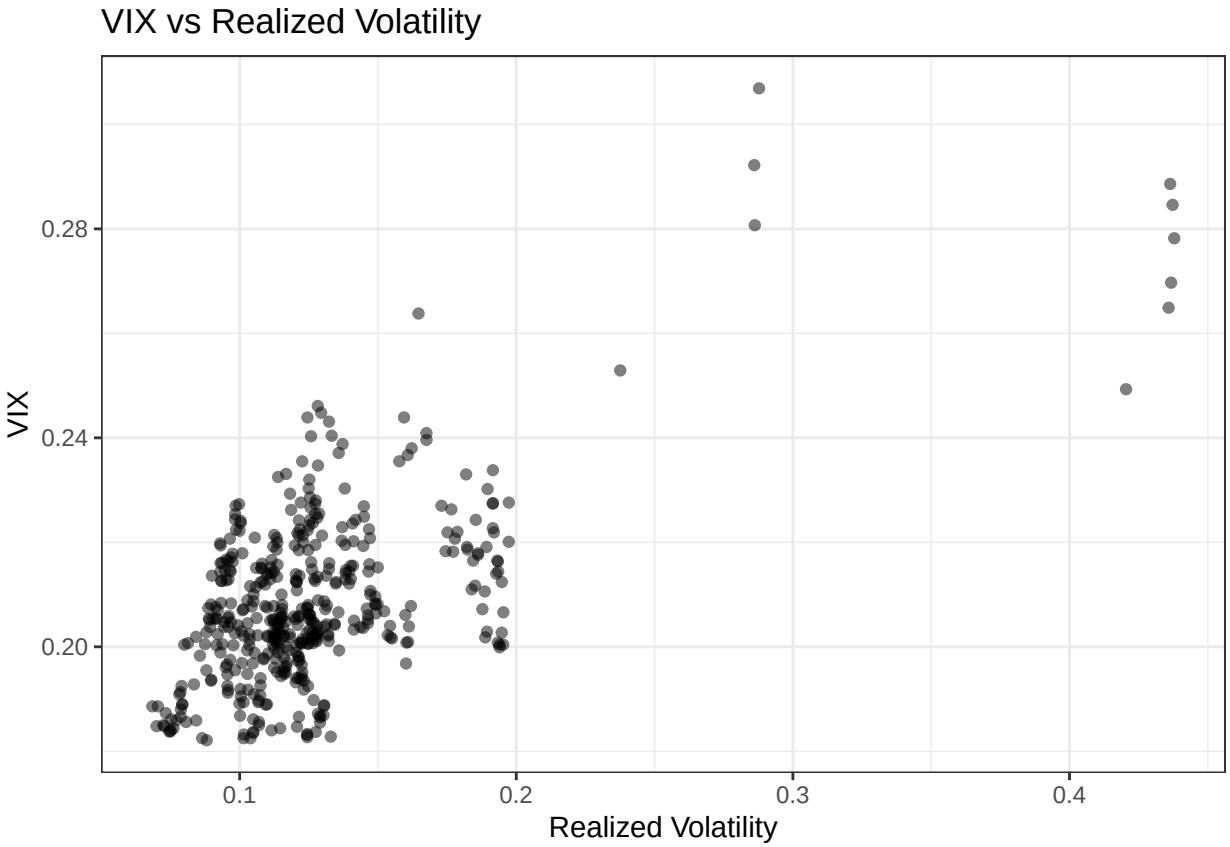
Here, we study the relationship between the VIX1Y index and the realized volatility of the S&P500 (SPY) stock prices by fitting regression models where the realized volatility is the predictor variable and the VIX1Y is the response. We want to examine whether we can predict the VIX by solely regressing 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 VIX
observed
trend
seasonal
irregular
VIX
0.28
0.24
0.20
0.275
0.250
0.225
0.200
0.003
0.000
−0.003
−0.006
0.04
0.02
0.00
−0.02
2023−07
2024−01
2024−07
2025−01
Time


Seasonal Decomposition of Realized Volatility
observed
trend
seasonal
irregular
Volatility
0.4
0.3
0.2
0.1
0.004
0.000
−0.004
0.05
0.00
−0.05
2023−07
2024−01
2024−07
2025−01
Time


Realized Volatility and VIX
Volatility Percentage
0.4
0.3
0.2
0.1
2023−07
2024−01
2024−07
2025−01
Time
Realized Volatility
VIX


VIX vs Realized Volatility
VIX
0.28
0.24
0.20
0.1
0.2
0.3
0.4
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	0.01307
Kernel	0.01183
GAM	0.01209

Summary Outputs for linear regression

```
##
## Call:
## lm(formula = vix ~ rolling_vol_annual, data = filter(vix_volatility,
##     date %in% output_cv.lm$train))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027481 -0.008669 -0.001837  0.007934  0.060911
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.179205   0.001912   93.72   <2e-16 ***
## rolling_vol_annual 0.232075   0.014063   16.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01327 on 411 degrees of freedom
## Multiple R-squared:  0.3985, Adjusted R-squared:  0.3971
## F-statistic: 272.3 on 1 and 411 DF,  p-value: < 2.2e-16
```

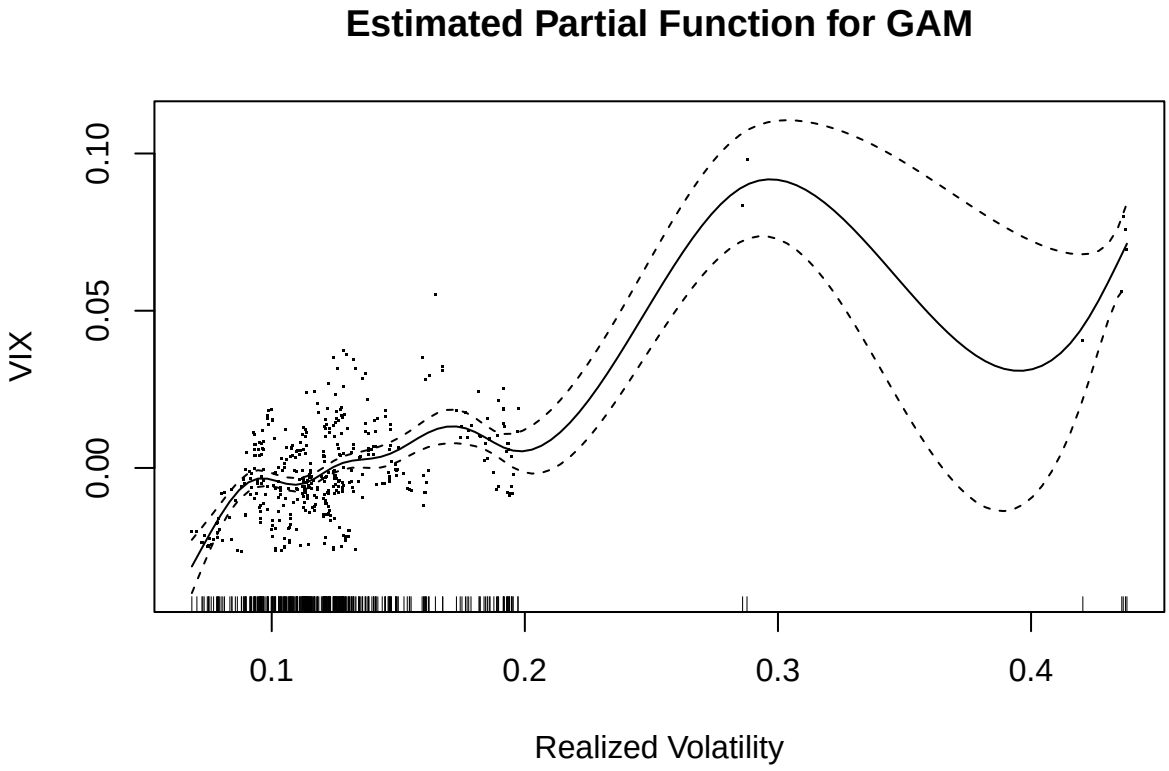
Summary Outputs for kernel regression

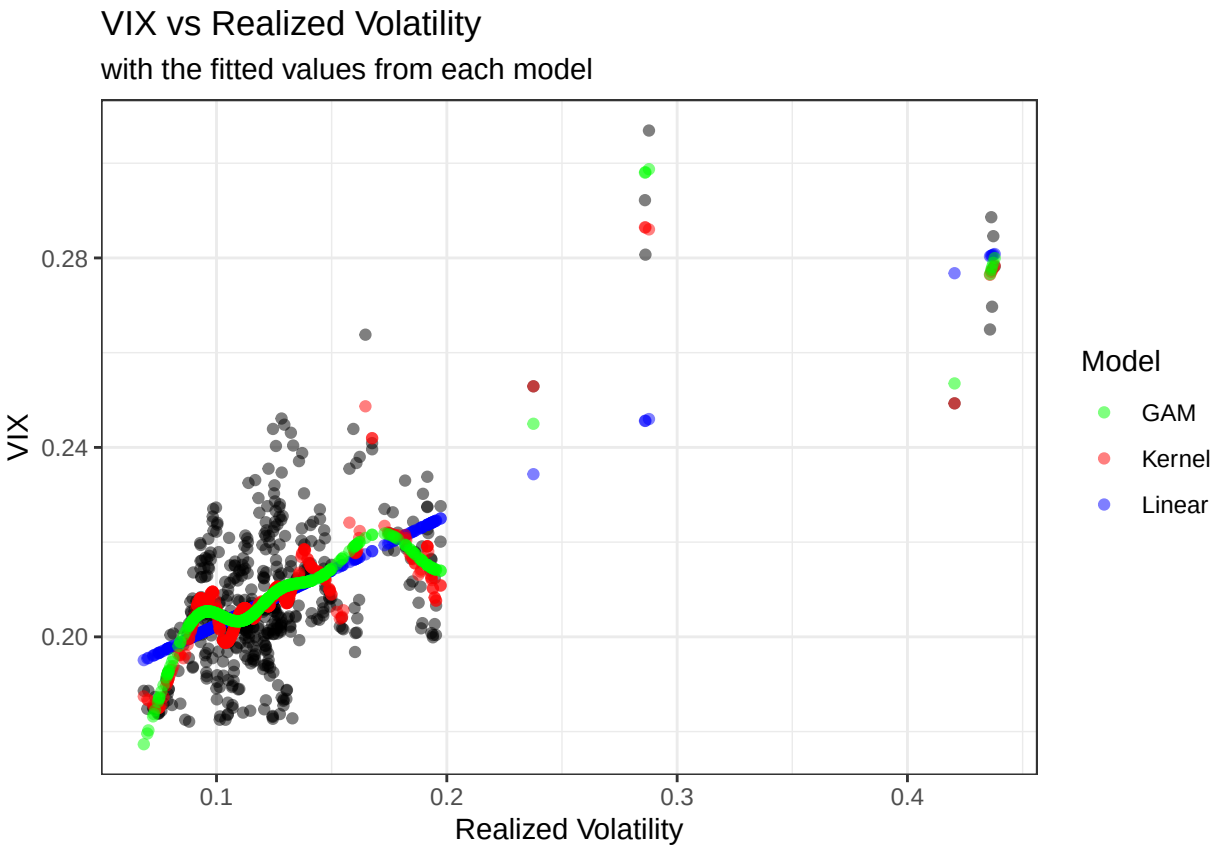
```
##
## Regression Data: 413 training points, in 1 variable(s)
##               rolling_vol_annual
## Bandwidth(s):         0.00150338
##
## Kernel Regression Estimator: Local-Constant
## Bandwidth Type: Fixed
## Residual standard error: 0.01108715
## R-squared: 0.5657679
##
## Continuous Kernel Type: Second-Order Gaussian
## No. Continuous Explanatory Vars.: 1
```

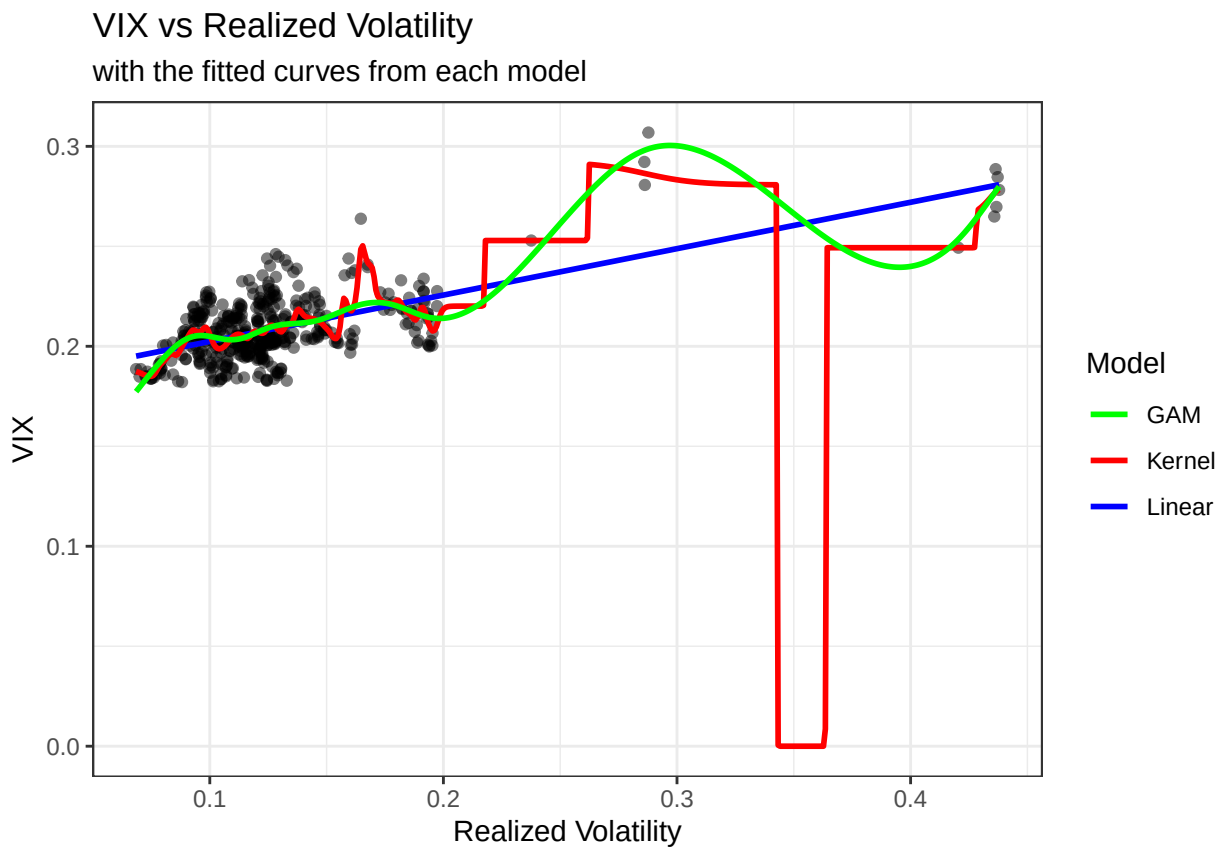
Summary Outputs for GAM

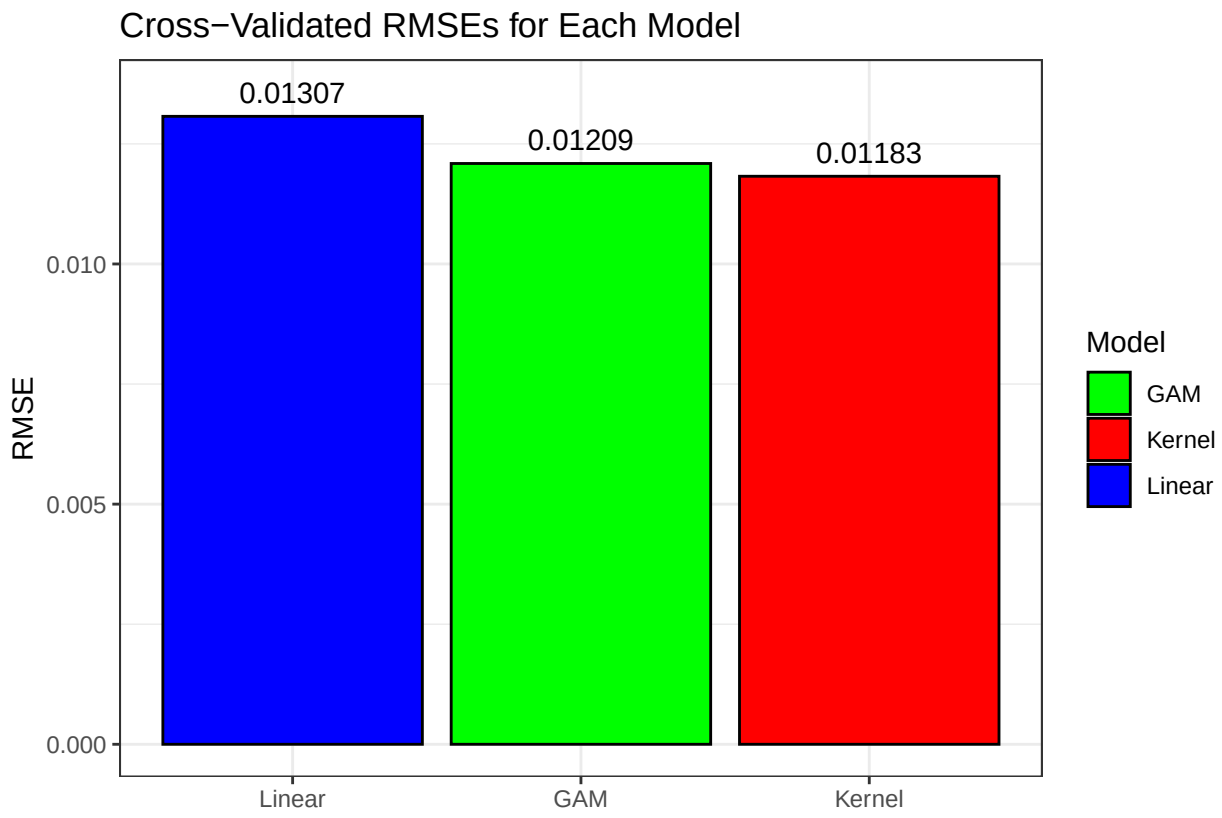
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## vix ~ s(rolling_vol_annual)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 0.2086552  0.0006023   346.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                          edf Ref.df     F p-value
## s(rolling_vol_annual) 8.662  8.967 41.55  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.471   Deviance explained = 48.2%
## GCV = 0.0001534  Scale est. = 0.00014981  n = 413
```




VIX vs Realized Volatility
with the fitted values from each model
VIX
0.28
0.24
0.20
0.1
0.2
0.3
0.4
Realized Volatility
Model
GAM
Kernel
Linear


VIX vs Realized Volatility
with the fitted curves from each model
VIX
0.0
0.1
0.2
0.3
0.1
0.2
0.3
0.4
Realized Volatility
Model
GAM
Kernel
Linear


Cross−Validated RMSEs for Each Model
0.01307
0.01209
0.01183
RMSE
0.010
0.005
0.000
Linear
GAM
Kernel
Model
GAM
Kernel
Linear