Critical Trading: Final Stakeholder Update

Team 2: Aditya Dhanotia, Tom Do, Jishnu Channamsetti, Shantanu Trivikram, Levon Hakobyan, Zhixing Zhu, Colin Zhou

Table of contents

1

Introduction

Experiment Overview

Objective

Assess the predictability of the realized volatility of the QQQ ETF using a set of macro features. This experiment aims to build a robust model that can assist in making informed investment decisions.

Hypothesis

Macro features such as GDP, equities, credit spread, futures, implied volatility, and returns can provide significant predictive power for the realized volatility of QQQ.

Experiment Overview

Predictors

Top 25 variables from PCA + QQQ_IV

Target Variable

rv_22d_lead(22+1)d
Realized Volatility
22 Days Later

2

Recap

Dataset Overview

783
Total predictors

4,160
Recorded trading dates (2005-2021)

182
Original predictors surviving since 2005

881
Rolling dates used to build/test daily models

To predict
Realized Volatility 22 Days Later

Data Cleaning & Windsorization

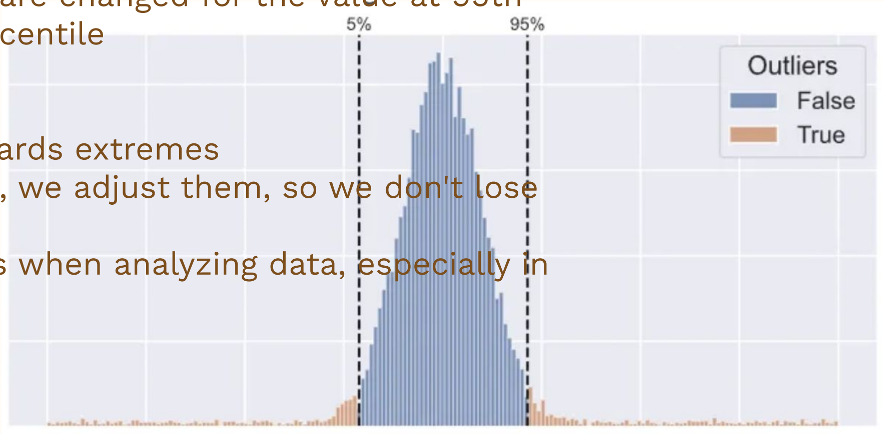
- Instead of removing extreme values that distort the overall picture of data, we adjust them

- 16 variables have max value greater than 1^10

- 95% of distribution is taken - high extreme values are changed for the value at 95th percentile, low values changed for value at 5th percentile

Winsorization:

- Reduces Skewness: Makes our data less tilted towards extremes
- Preserves Data: Instead of deleting extreme values, we adjust them, so we don't lose information
- Robust Analysis: Helps in producing reliable results when analyzing data, especially in statistical analysis



Feature Selection (PCA)

- For each date, we fit a PCA model to obtain the best model for that date.
- We select the Top 25 variables by calculating the mean contribution to principal components
- We will use these 25 variables and variables used in baseline model as the predictors to train the model.

	PC_Weighted1	PC_Weighted2	PC_Weighted3	PC_Weighted4	PC_Weighted5
BofA_EM_Corp_OAS_lag1d	0.577266	-0.000299	0.001991	0.002153	0.001422
BofA_EM_HY_Corp_OAS_lag1d	0.576554	0.000517	-0.000992	0.002574	0.001976
BofA_US_HY_Corp_OAS_lag1d	0.574741	-0.000971	0.004149	0.000360	-0.000224
BofA_Euro_HY_Corp_OAS_lag1d	0.574022	-0.000212	0.001837	0.001653	-0.000148
Equities_XLP_CLOSE_lag1d_1_period_ret_100_period_std_dev_volatility	0.577475	0.000290	0.001638	-0.000319	-0.000418
...	...	...	...	...	...
Macro_UMCSENT	-0.434871	0.005713	-0.025735	0.004042	0.004347
NIKKEI_PB_Ratio_lag1d	-0.503428	0.002925	-0.015384	0.000377	0.001555
SPX_CAPE_lag1d	-0.533853	0.006028	-0.001720	-0.002887	0.002279
MSCI_Europe_PB_Ratio_lag1d	-0.532223	0.004550	-0.012312	-0.000475	0.003019
SPX_PB_Ratio_lag1d	-0.539984	0.006187	-0.007059	-0.002029	0.001753

3

Model

Approach

Rolling Forward Technique: Use 881 past trading days for training to predict RV of 22 days after

	Feature 1	Feature 2	Feature n-1	Feature n
Day 1				
Day 2				
Day 3				
Day 881				
Day 882				
Day 883				

Train Model 1 ➔ Predict: $RV_{882 + 22d}$

Approach

Rolling Forward Technique: Use 881 past trading days for training to predict RV of 22 days after

	Feature 1	Feature 2	Feature n-1	Feature n
Day 1				
Day 2				
Day 3				
Day 881				
Day 882				
Day 883				

Train Model 2 ➔ Predict: $RV_{883 + 22d}$

Approach

Rolling Forward Technique: Use 881 past trading days for training to predict RV of 22 days after

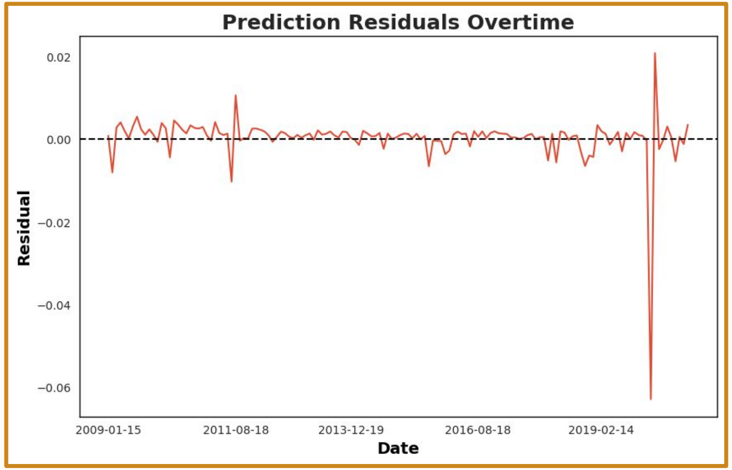
	Feature 1	Feature 2	Feature n-1	Feature n
Day 1				
Day 2				
Day 3				
Day 881				
Day 882				
Day 883				

Train Model 3 ➔ Predict: $RV_{884 + 22d}$

Baseline Model – Linear Regression

rv_22d_lead(22+1)d ~ IV_QQQ + rv_MA_5d_lag1d + rv_MA_10d_lag1d + rv_MA_22d_lag1d

Linear Regression Model



- **Avg. In-sample R^2:** -7.12657
- **Avg. In-sample RMSE:** 0.00396
- **Out-of-sample R^2:** 0.01702
- **Out-of-sample RMSE:** 0.00612

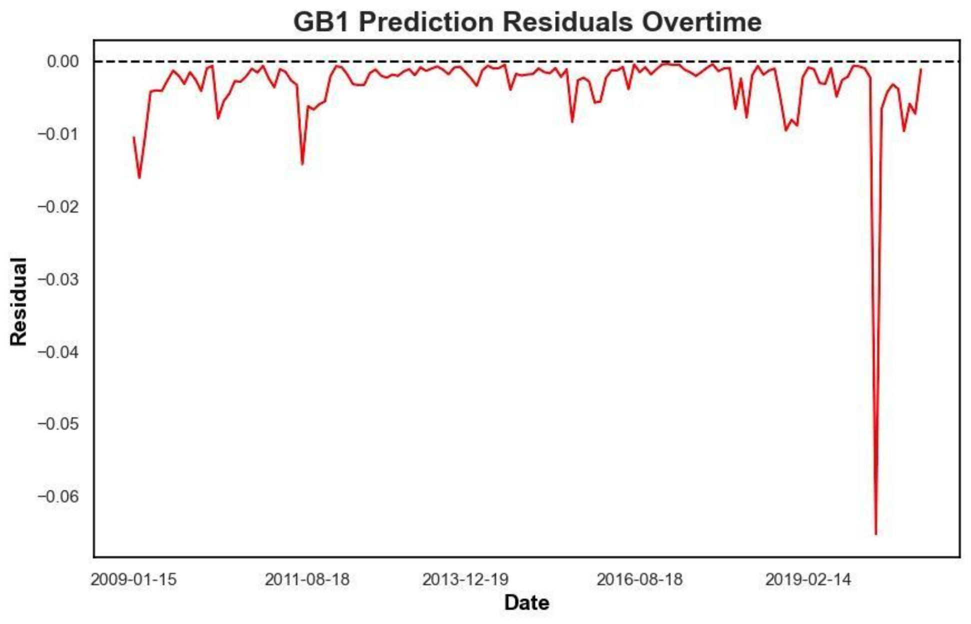
Best Model– Gradient Boost

Hyperparameter:

- learning_rate=0.05
- min_samples_split=80
- min_samples_leaf=50
- max_depth=5
- max_features='sqrt'
- subsample=0.8
- random_state=10

Performance:

Out-of-sample R^2: 0.0067404
Out-of-sample RMSE: -0.285294



Results Comparison

Model	RMSE	R2
RF (30)	0.00674606	-0.2867
RF (100)	0.00674627	-0.2889
RF (200)	0.00702362	-0.2943
GradientBoosting (Default)	0.00674412	-0.2881
GradientBoostingRegressor(learning_rate=0.1, min_samples_split=80,min_samples_leaf=50,max_depth=5,max_features='sqrt',subsample=0.8,random_state=10)	0.00674383	-0.2880
GradientBoostingRegressor(learning_rate=0.2, min_samples_split=80,min_samples_leaf=50,max_depth=5,max_features='sqrt',subsample=0.8,random_state=10)	0.00674636	-0.2890
GradientBoostingRegressor(learning_rate=0.05, min_samples_split=80,min_samples_leaf=50,max_depth=5,max_features='sqrt',subsample=0.8,random_state=10)	0.00674041	-0.2853
GradientBoostingRegressor(learning_rate=0.1, min_samples_split=80,min_samples_leaf=50,max_depth=5,max_features='log2',subsample=0.8,random_state=10)	0.00674343	-0.2879
GradientBoostingRegressor(learning_rate=0.05, min_samples_split=80,min_samples_leaf=50,max_depth=8,max_features='sqrt',subsample=0.8,random_state=10)	0.00674102	-0.2869

Trading Strategy

We have options data for every **3rd friday of month**

Filtered our forecast of IV to only these dates, build trading strategies around these dates

RETURN - bias adjusted dynamic hedged straddle return

Straddle Position: Simultaneously buy/sell call & put options with same strike price & expiry.

Dynamic Hedging: Regularly adjust delta to near zero by trading underlying asset/derivatives.

Return Calculation:

- Net gain/loss from options & hedging activities.
- Includes premiums, hedging costs/gains, transaction fees.

Bias Adjustment:

- Account for biases (e.g., volatility risk premium).
- Adjust hedging strategy based on expected biases.

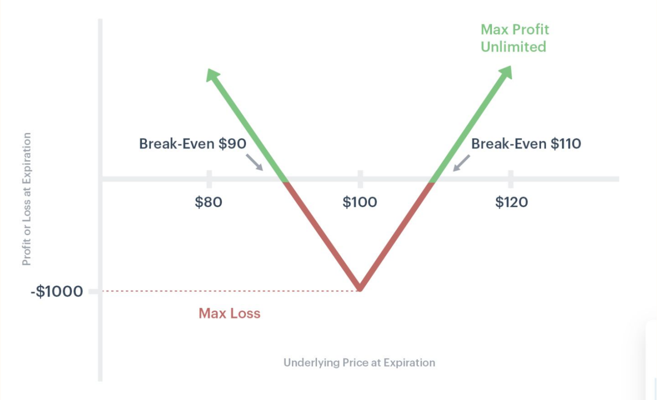
Key Aspects:

- Exploits volatility rather than price direction.
- Requires sophisticated risk management.
- Involves complex financial modeling.

Strategy 1 – Direction Trading

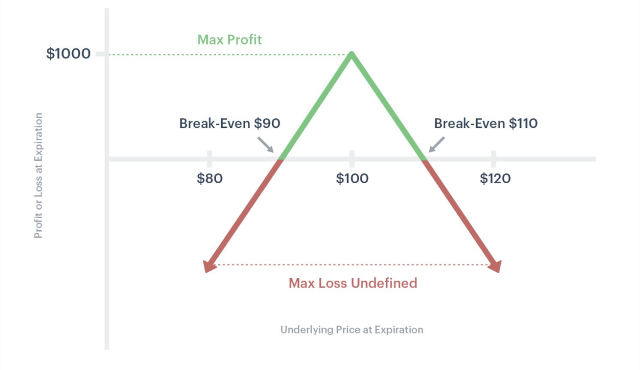
Every month we compare our Forecast Implied Volatility to Realized Volatility:

If, IV > RV, expect more volatility in the market and go long (buy) the straddle portfolio with a weight of +1



Every month we compare our Forecast Implied Volatility to Realized Volatility:

If, IV < RV, expect less volatility in the market and go short (sell) straddle portfolio with a weight of -1



Strategy 2 – Kelly Weights

Kelly Criterion is an allocation technique that helps investors decide how much of capital to place on each trade.

We used Kelly to obtain portfolio weights

Strategy 1:

Weight for each month = Rolling mean of ret / Rolling SD of ret squared

Return = Kelly Weight * Monthly Return

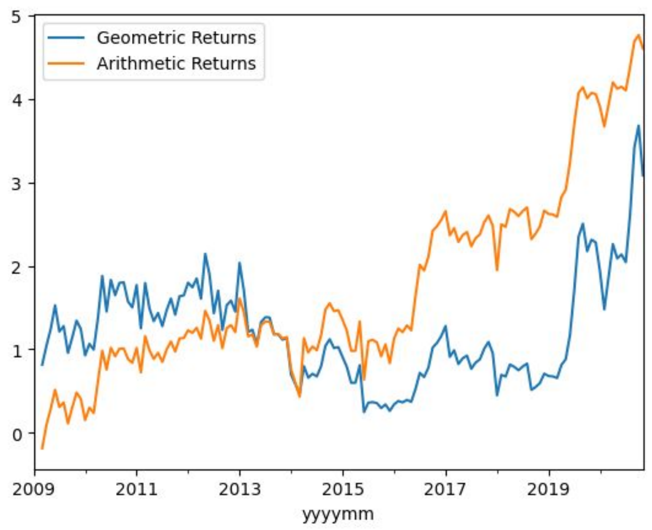
Strategy 2:

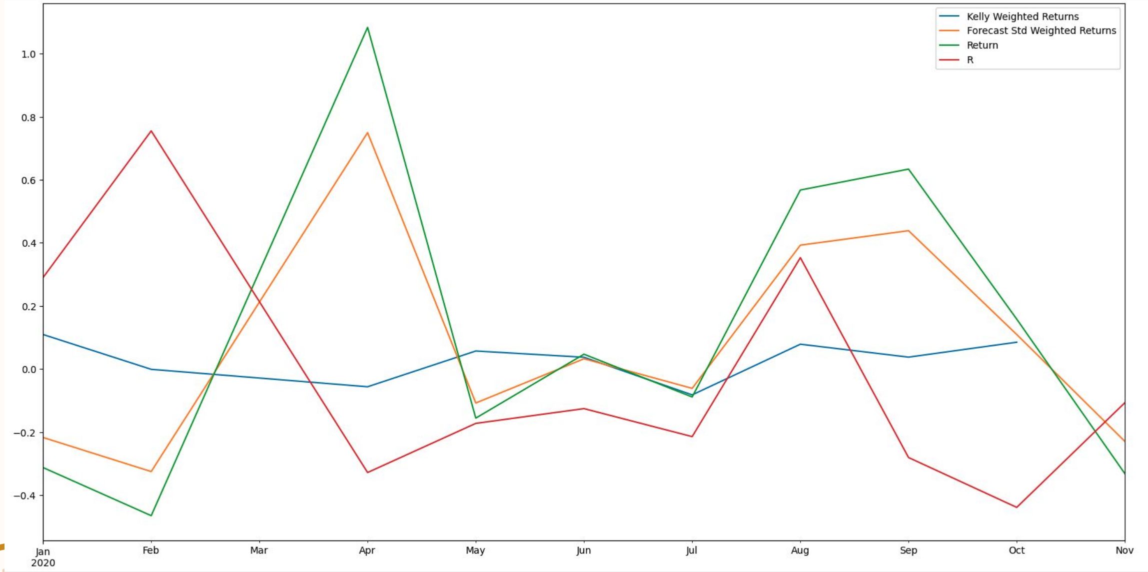
Weight for each month = Rolling mean of ret / Rolling SD of forecast Squared

Return = Kelly Weight * Monthly Return

Summary Stats

strategy	model	stats	R_long	R_short	RC	pctspread	Return Forecasts	Return Forecasts (Kelly)	Return Forecasts (Forecast Std)
biasadj dynamic hedged straddle return	PCA_Filtered_Features_Random_Forest	std	0.327190	0.327190	0.002090	0.008360	0.448677	6.595246	0.311875
	PCA_Filtered_Features_Random_Forest	mean	-0.042447	0.042447	-0.004731	0.018924	0.070011	0.733442	0.044026
	PCA_Filtered_Features_Random_Forest	ann Sharpe	-0.449403	0.449403	-7.841677	7.841677	0.540533	0.385235	0.489007





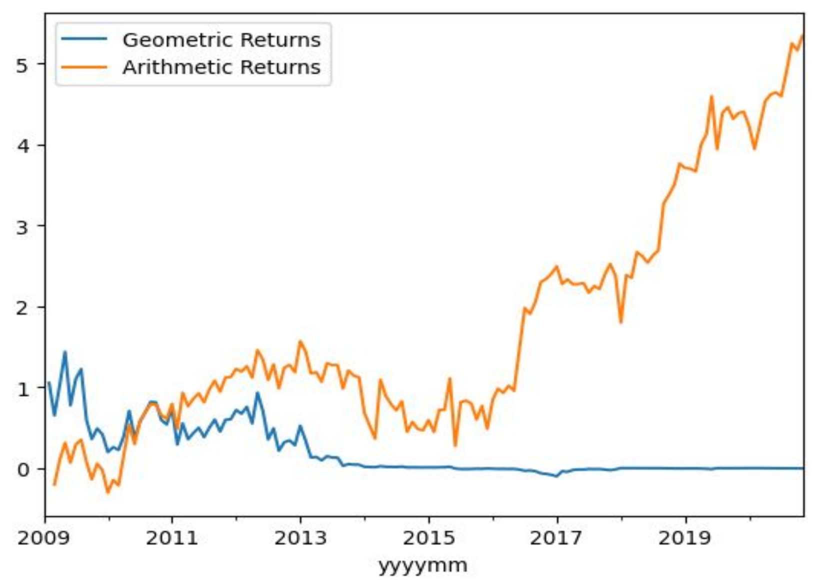
Trading Strategy Returns

Model Name	Stats	Short Returns	Return	Kelly Weighted Returns	Forecast Weighted Returns
Baseline Model (IV~RV)	Mean	0.04	0.06	0.17	0.03
	Standard Deviation	0.32	0.44	0.52	0.24
	Sharpe Ratio	0.44	0.53	1.18	0.53

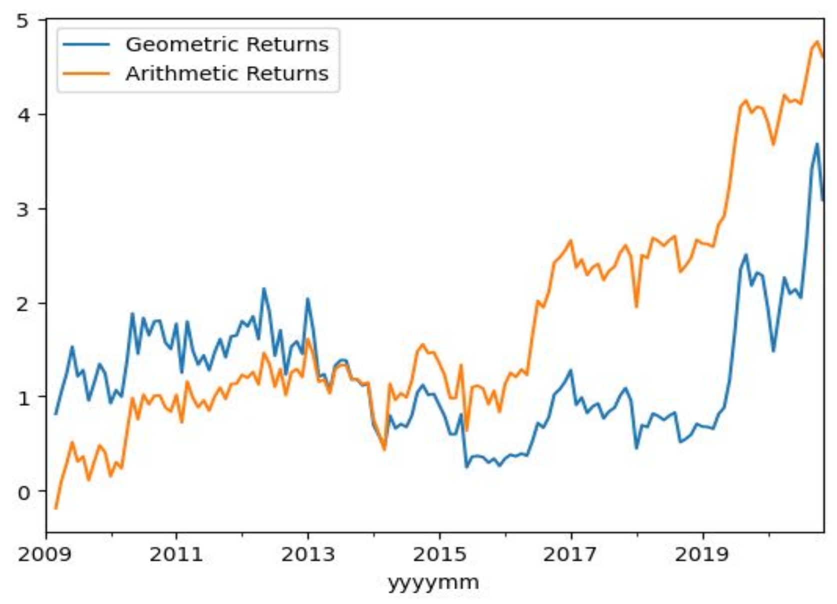
Trading Strategy Returns

Model Name	Stats	Short Returns	Return	Kelly Weighted Returns	Forecast Weighted Returns
Best Model Gradient Boosting IV + PCA Variables	Mean	0.04	0.07	0.17	0.03
	Standard Deviation	0.32	0.44	0.52	0.21
	Sharpe Ratio	0.44	0.57	1.18	0.52

Value of $1 over the years

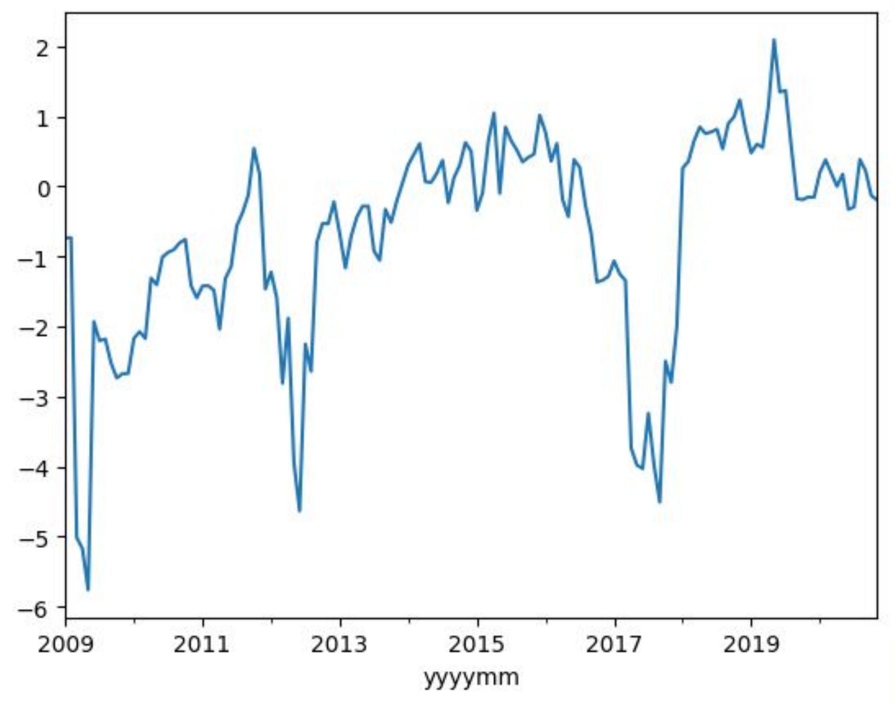


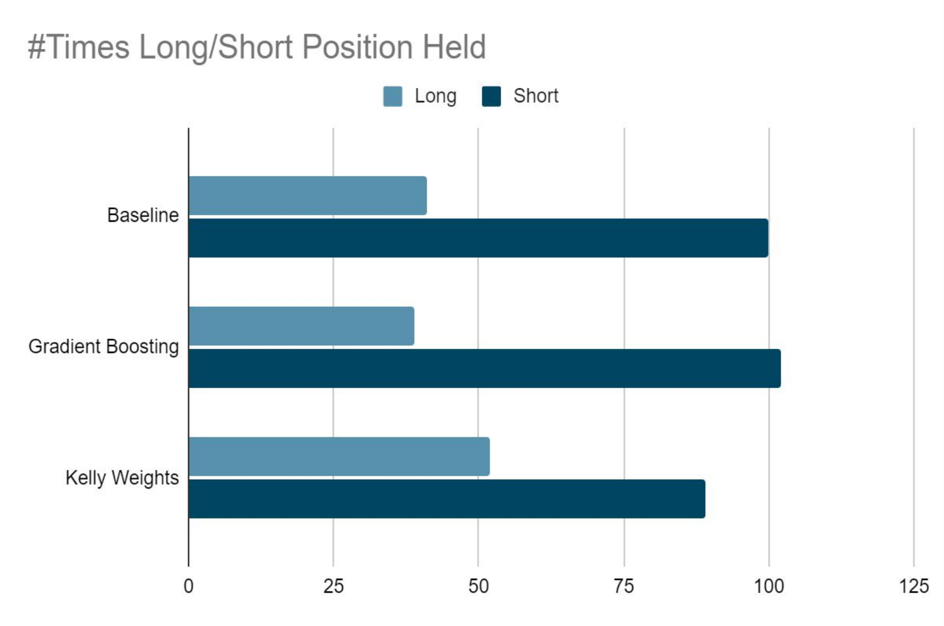
Baseline Model



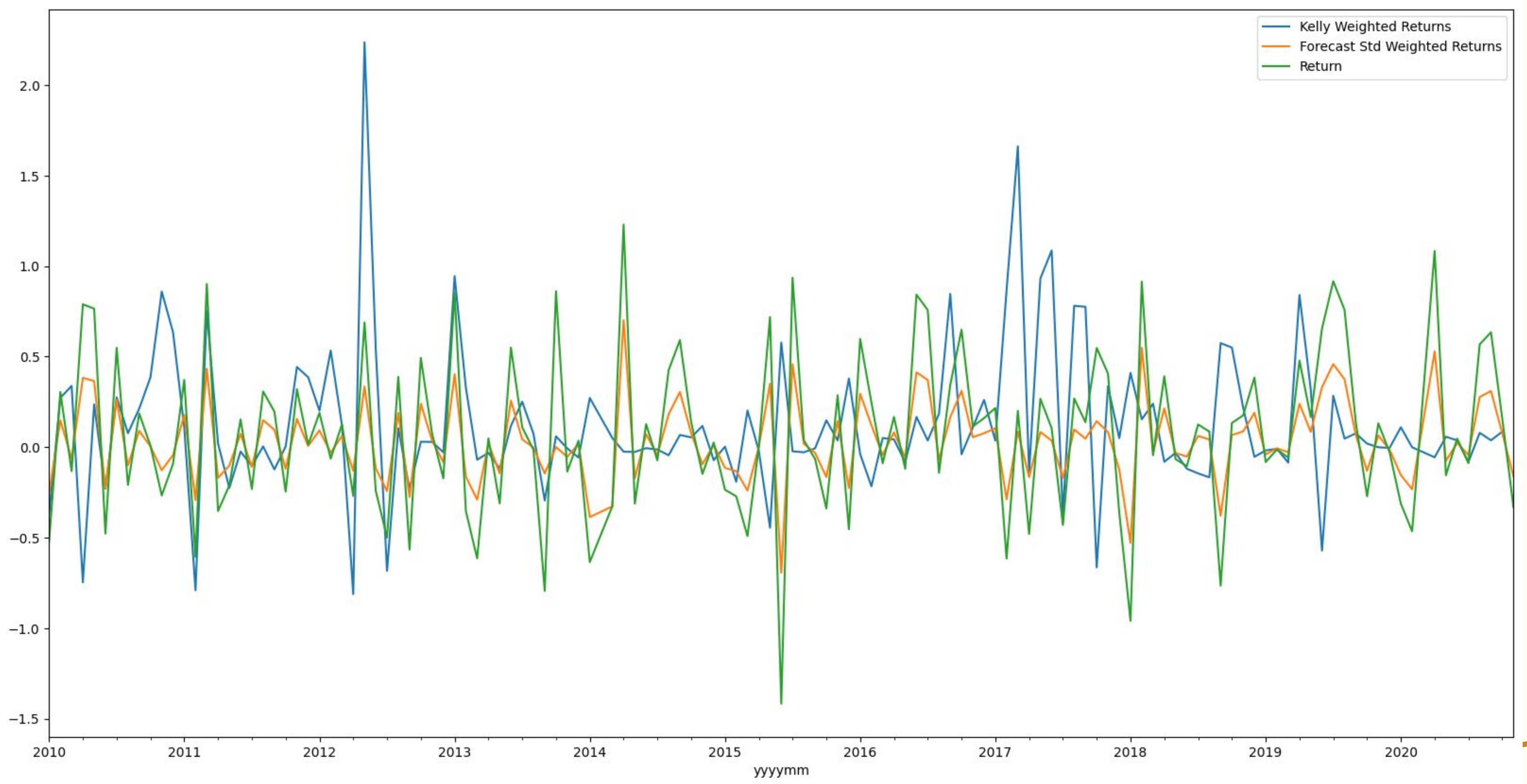
Gradient Boosting Model

Kelly Weights

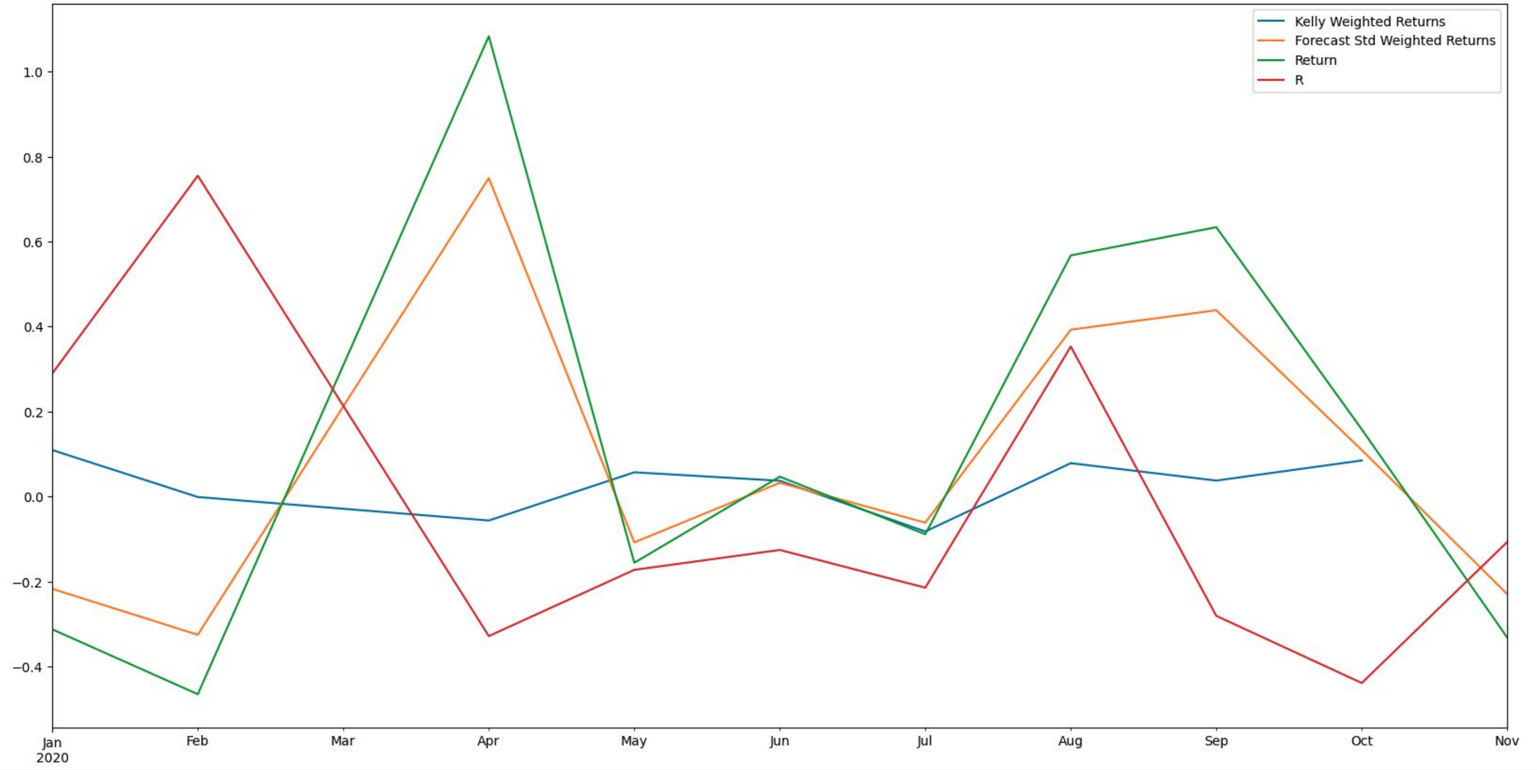




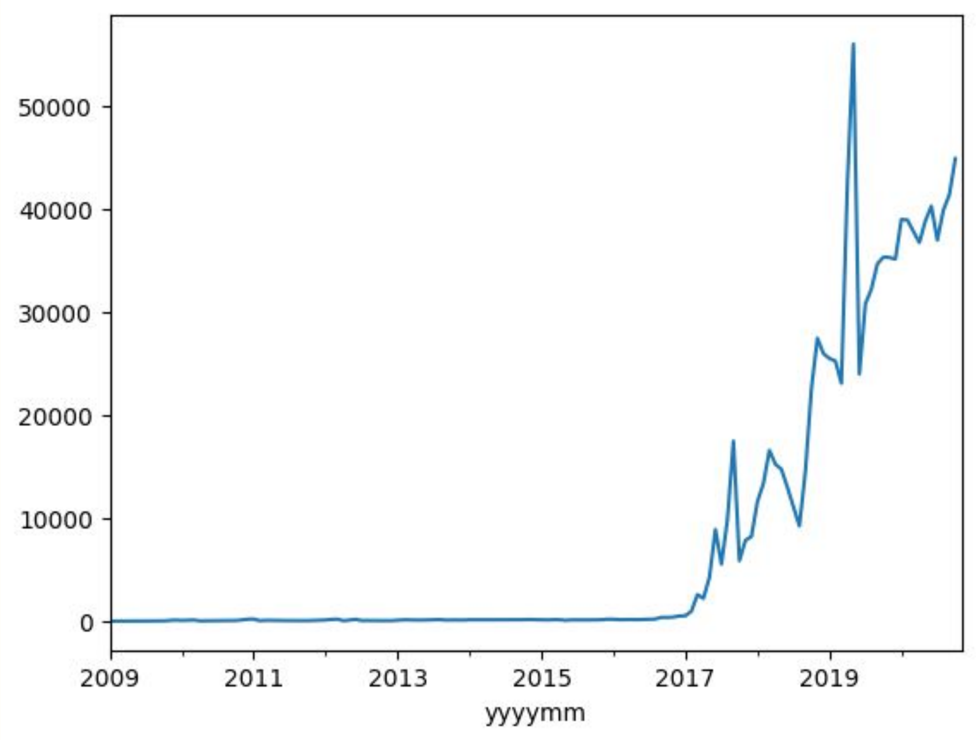
Time Series of Returns

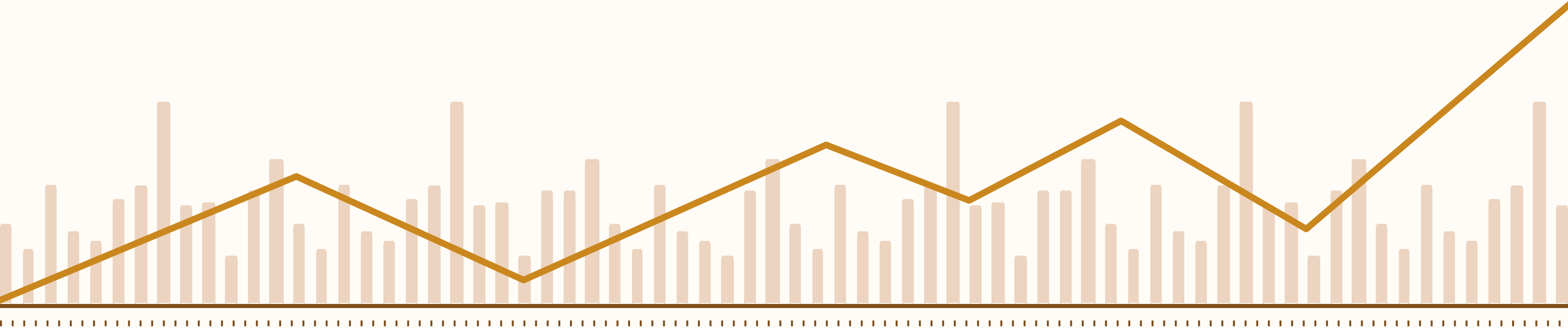


Time Series of Returns



Kelly Weighted Value of $1





Thank You!