Predicting Volatility

Group 3
Lanli Feng, Xiwei Liu, Shiyi Liu, Viktor Aghajanyan, Shaoyu Cheng

Abstract
This paper proposes to study VIX predicting based on multiple linear regression model and discrete-time GARCH model. Our results from multiple linear regression indicate that the unemployment rate, oil price, interest rate on required reserves, PPI, and Bitcoin price will influence the VIX. The results of the ARCH/GARCH model indicate that ARCH (2,0) performs best in predicting VIX volatility. We use ARCH (2,0) to design a Bollinger band trading strategy that trades the rolling predicted VIX. This strategy gets a profit of 80% with a maximum drawdown of -3.48%, which is way much higher than the simple buy and hold strategy (-46.20%).

1. Introduction

The volatility index has become more important given the more frequently fluctuated global financial markets, especially since the financial crisis. The VIX uses S&P 500 option prices to determine expectations of the stock market 30-days-ahead volatility and has become the gauge of the level of risk, fear, or stress in the market. Based on how much money traders are willing to spend on options, the VIX decides where volatility should be headed in the future ("The VIX", 2021). The line graph below shows the VIX close price over the past 20 years. As evidenced by Appendix 1, spikes of VIX were often accompanied by the financial crisis. Therefore, there are clearly a lot of benefits in predicting the future VIX movements using advanced time-series techniques. We aim to compare across various models to find out the best predictive model for VIX.

While only one paper constructs a regression model with explanatory variables that are exogenous to the index and examines the model prediction errors (Saha et al. 2019), we found it necessary to analyze what factors negatively or positively influence the VIX using a multiple linear regression model. Our study provides some explanatory variables for the index. Finding out the correlation between explanatory variables and dependent variables can help us have a more comprehensive understanding of the index, better manage the

risk and portfolio, and make right decisions.
The volatility index prediction has been widely researched. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model seems to be a natural choice for studying VIX since it is appropriate for time series data where the variance of the error term is serially autocorrelated following an autoregressive moving average process ("GARCH", 2021). For instance, Majmudar and Banerjee (2004) forecast the revised VIX using a variety of GARCH models and select the EGARCH model as the most accurate one. Hao and Zhang (2013) compare the various GARCH models and conclude that the GARCH implied VIX is lower than the CBOE VIX. Liu et al. (2015) forecast VIX under GARCH-type models, and find the risk-neutral parameters forecast the one-day out-of-sample VIX accurately. Therefore, we choose the GARCH model to forecast market volatility and model future rolling predictions. We use ARCH (2,0) to design a Bollinger band trading strategy that trades the rolling predicted VIX.

2. Data Collection and Data Description

We used VIX and S&P 500 indices data from January 1, 2015 till now at Yahoo Finance. Referring to Appendix 2, there are 1710 pieces of information in total with a mean of 17.76, a median of 15.6, and a standard deviation of 7.89. The Minimum number in this data is 9.14 whereas the maximum number reaches 82.69, which shows a large difference.
Then, the overall trending of these indices during this period is shown to give a more straightforward visualization in Appendix 3. The blue line represents VIX and the red line represents S&P 500. As to VIX, it displays a relatively smaller fluctuation from 2015 to 2019 with about three non recurring peaks. The smallest figure during this period is about 10, which corresponds to stable, stress-free periods in the markets. At the beginning of 2020, however, the index sharply increased to over 80. Generally, a high figure is linked to large volatility resulting from increased uncertainty, risk, and investors' fear. In this case, this abnormal figure is significantly affected by the outbreak of Covid-19 in the United States. Afterward, it exhibits a dramatic decrease and then approaches the previous figures till now.
Then, incorporating S&P 500 into our analysis, a negative relationship between VIX and S&P 500 is clearly displayed. Whenever S&P 500 shows an upward trend, VIX displays a

downward slope correspondingly. This relationship is also compatible with the intuition of VIX in terms of the market's expectations. Other data such as Covid-19 deaths and CPI are collected in Appendix 11 from public information sources like FRED, which are further explained in regression analysis.

Historical Volatilities

It is notable in Appendix 4 that when we change our window length, the historical volatility exhibits a very different movement. When we extend the rolling window, the line is much more steady. On the contrary, when the rolling window is short, the line is more volatile. It is mainly because a longer rolling window contains more samples of population data. Integrating all data above, further analysis of predicting the future movement of VIX is conducted.

3. Model Selection and Analysis (GARCH)

First, we need to grab the VIX data from Yahoo Finance. In Appendix 5 we have the percent change of VIX close price between one day and the next from the time period starting from 2015 and ending in current data (17 October 2021).

VIX is a good candidate for the GARCH process because there are periods of much higher volatility compared to general volatility in the chosen period. Before we apply the GARCH process to this price change, we need to figure out which order GARCH process we need to use. To do that we apply Partial Autocorrelation Function to get a good hint.

We can see in Appendix 6 that in PACF it is pretty strong until 2 and then shuts off. This hinges us to choose GARCH (2,2) model.

To fit the actual GARCH model, we used ARCH python library with:

```
model = arch_model(returns, p=2, q=2)
```

In Appendix 7 we summarize the results of fitting the model to VIX data. We can see that the choice of GARCH (2,2) model was wrong given that alpha [2], beta [1], and beta [2] all have p values larger than 0.05.

Next, we dropped the complexity of the model from GARCH (2,2) to just ARCH (2,0). In Appendix 8 we can see that both alpha [1] and alpha [2] are significant as evidenced by the p-value. So we're good to keep all these parameters.

To measure how good is the GARCH prediction model on actual data we perform rolling forecast origin method. We are building a GARCH model for each of 365 day forecasting periods and then predicting the next day out. This blue line is still the returns of VIX close price and the orange line is the prediction of the ARCH process. We were hoping to see that the predicted volatility gets higher exactly when the actual returns get jumpier. This turns out to be exactly what we observe in the graph (Appendix 9).

The last thing we want to accomplish with GARCH process volatility prediction is to explain how in practice we would utilize it. For that, we constructed a volatility prediction model for the next 7 days. If we look at the volatility prediction for the next 7 days (Appendix 10), we see that it starts off high, then dips low, and then goes back up for the rest of the days. The reason we chose 7 days is that when we choose longer time periods, we become less certain about what exactly it's telling us.

4. Multiple Linear Regression Analysis

We choose multiple linear regression models to analyze the relationship between variables and VIX index. (Variable details shown in Appendix 11)

$$\boldsymbol{VIX = \beta 0 + UNEMPLOYMENT * \beta 1 + CPI * \beta 2 + SPY * \beta 3 + OIL * \beta 4}$$
$$\boldsymbol{+ IORR * \beta 5 + PPI * \beta 6 + T10Y3M * \beta 7}$$

We assume at the first stage that the VIX index will be affected by seven factors. First, When the economy is in poor shape, job supply will decrease, and the unemployment rate can be expected to rise. Second, increasing inflation is not a good sign for the investors, since it will reduce the income that can be invested in the market. Third, the S & P 500 index shows the performance of the underlying assets of VIX. Fourth, crude oil is widely used in industries, so crude prices can affect the prices of many assets and therefore the expectation of the economy. Fifth, if the Fed raises the reserve requirement, or in other words, executes a tightening monetary policy, it releases the sign that the market is overheating and needs to cool down. Finally, the spread between long-term and short-term treasury rates is an indicator of the yield curve implying investors' expectations. If investors feel fear in the market, they may turn to other assets with higher stability such as treasuries.

From the regression summary (Appendix 11, Table 1), there are four factors that have significance larger than 5%. The seven-factor regression model provides an adjusted R-square of 67.8%. Then we test a new model Only keeping significant factors.

$$VIX = \beta 0 + UNEMPLOYMENT * \beta 1 + OIL * \beta 2 + IORR * \beta 3 + PPI * \beta 4$$

Although the number of factors is reduced to four, we keep a similar adjusted R-square value (Appendix 11, Table 2). Obviously, the unemployment rate has a positive relationship with VIX. The oil price has a negative relationship with VIX, which is opposite to our intuition. However, higher oil prices can also be a sign of strong demand for crude oil and good economic growth. Required reserves rate and PPI index are positively related to VIX. Now we include other factors in the model. Since 2019, the outbreak of COVID had a serious impact around the world, blowing the global economy terribly. We include the COVID death daily data in the model, as well as the Bitcoin prices that play an important role in the financial market now.

$$VIX = \beta 0 + SPY * \beta 1 + OIL\_PRICE * \beta 2 + COVID\_DEATH * \beta 3 + BITCOIN * \beta 4$$

The model still has an adjusted R-square of 68% (Appendix 11, Table 3). Covid death is not significant in the model. Since the death data shows many stages and peaks, a linear model may not be proper to explain the relationship between COVID and VIX. Bitcoin price moves along with VIX. The raising of Bitcoin price is not a good sign for the stability of the financial market because it is still an emerging currency and faces many challenges and risks.

5. Trading Strategy

We introduced the Bollinger Band into our strategy (See Appendix 12). After we got the historical VIX value and predicted daily returns, we can generate the predicted VIX value day by day (See Appendix 13). The strategy applies the principle of mean reversion. Standing at today, if we expect the VIX price tomorrow is: (1) smaller than the lower Bollinger Band, (2) do not have a long position in the last order (which means the price is not continuously dropping below the band), then we place a long position on the VIX index. In other words, because VIX is mean-reverting, we predict the price will increase and "cross" the lower band and long VIX. Similarly, when we predict the price will decrease and "cross" the upper band, we short VIX.

As shown in Appendix 14, The green points are our long positions and the red ones are short positions. Our initial investment was $10,000 and the trading period was 1 year. We placed 10 longs and 10 shorts in total, we also canceled the orders when the price of the next buy was higher than the last sell. Eventually, we got a profit of 212.66% with a maximum drawdown of -3.48%, which is way much higher than the simple buy and hold strategy(-46.20%).

6. Conclusion

Overall, on the subject matter of this project, our goal was to come up with predictive and explanatory models to essentially design near-term forecasts for VIX to then design a profitable VIX trading strategy and explain the market variables (features) that can help explain the VIX market behavior.

The results of the ARCH/GARCH model indicate that ARCH (2,0) was the best in predicting VIX volatility given our partial autocorrelation test as well as GARCH experiments. We used ARCH (2,0) to design a Bollinger band trading strategy that trades the rolling predicted VIX given the bands that are 2 standard deviations away from the SMA 20. To explain the features that have the most explanatory power, we used multiple linear regression using 9 different economic and market variables as summarized in the "Data Collection" and "Multiple Linear Regression" sections. To avoid repetitions, we would refer to the "Multiple Linear Regression" section for the explanations of significance of features and signs of their coefficients.

References

[1] Kuepper, J. (2021, October 14). Cboe volatility index (VIX). Investopedia. Retrieved October 17, 2021, from https://www.investopedia.com/terms/v/vix.asp.

[2] The Investopedia Team. (2021, June 11) GARCH. Investopedia. Retrieved from https://www.investopedia.com/terms/g/garch.asp

[3] The VIX: My biggest moneymaker in 20 years of trading. The Profit Takeover. (2021, June 15). Retrieved October 17, 2021, from https://theprofittakeover.com/the-vix-my-big gest-moneymaker-in-20-years-of-trading/

[4] Hao J., Zhang J.E. (2013) GARCH option pricing models, the CBOE VIX and variance risk premium. Journal of Financial Econometircs, 11, pp. 556-580

[5]Majmudar, U., & Banerjee, A. (2004, April 21). Vix forecasting. SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=533583.

[6] Liu Q., Guo S., Qiao G., (2015) VIX forecasting and variance risk premium: A new GARCH approach. North American Journal of Economics and Finance, 34, pp. 314-322

[7]Saha, A., Malkiel, B.G. & Rinaudo, A. (2019) Has the VIX index been manipulated?. J Asset Manag 20, 1–14 (2019). https://doi.org/10.1057/s41260-018-00102-4

[8] FRED. (2021) Unemployment Rate [Data file]. Retrieved from https://fred.stlouisfed.org/series/UNRATE

[9] FRED. (2021) CPI [Data file]. Retrieved from https://fred.stlouisfed.org/series/CPIAUCSL

[10] U.S. Bureau of Labor Statistics. (2021) PPI [Data file]. Retrieved from https://www.bls.gov/ppi/#data

[11] FRED. (2021) 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity (T10Y3M) [Data file]. Retrieved from https://fred.stlouisfed.org/series/T10Y3M

[12] FRED. (2021) IORR Interest Rate Spread [Data file]. Retrieved from https://fred.stlouisfed.org/series/IORR

[13] Github. (2021) COVID death data[Data file].

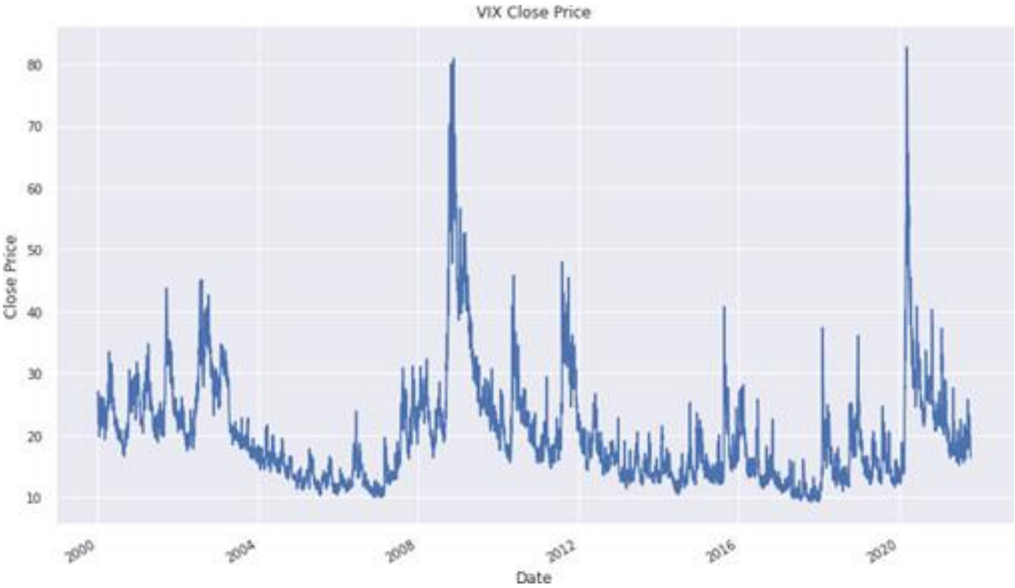
[14] Yahoo Finance.(2021) Bitcoin Closing Price [Data file]. Retrieved from https://finance.yahoo.com/quote/BTC-USD/

[15] Yahoo Finance.(2021) Crude Oil prices [Data file]. Retrieved from https://finance.yahoo.com/quote/CL=F/

[16] Yahoo Finance.(2021) S&P Close [Data file]. Retrieved from https://finance.yahoo.com/quote/%5EGSPC?p=%5EGSPC

Appendix

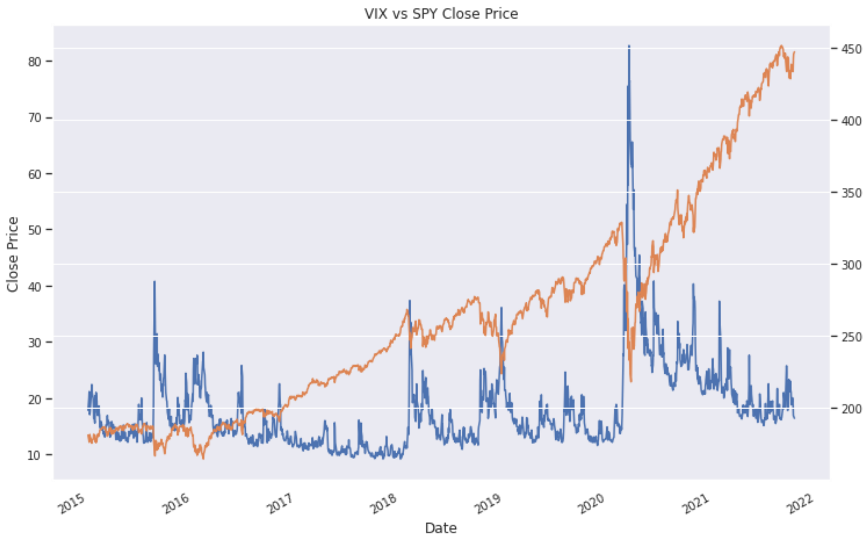
Appendix 1



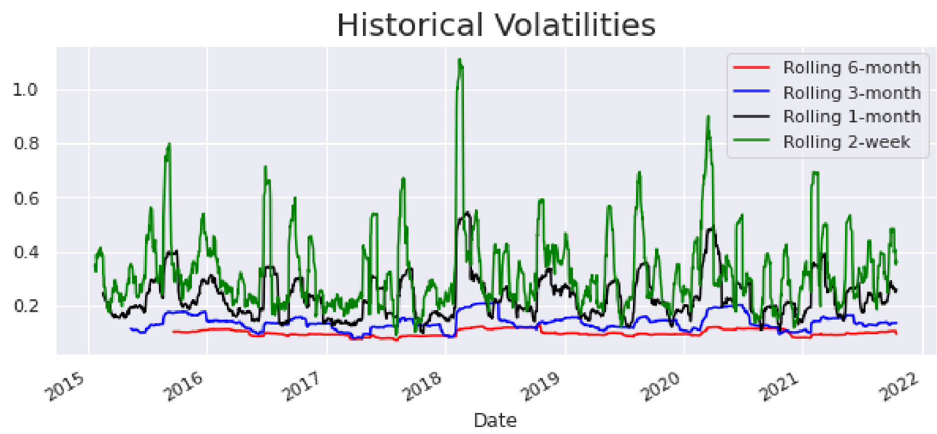
Appendix 2

	count	mean	std	min	25%	50%	75%	max
VIX_Close	1710.0	17.757158	7.892246	9.14	12.8325	15.6	20.557499	82.690002

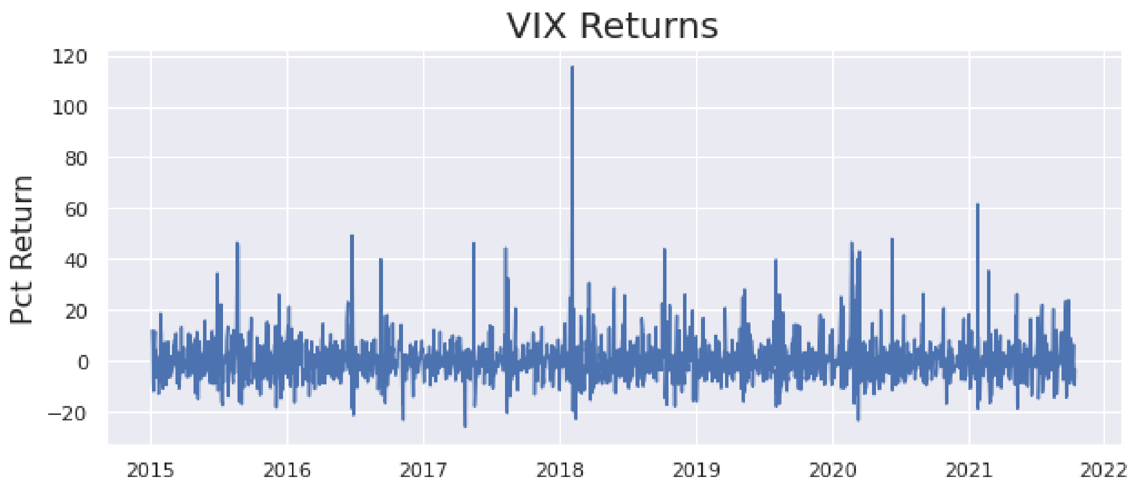
Appendix 3



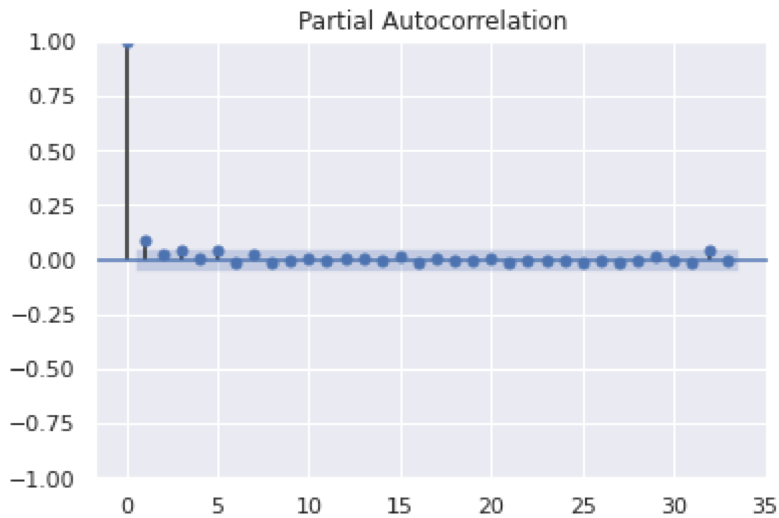
Appendix 4



Appendix 5



Appendix 6



Appendix 7

Constant Mean - GARCH Model Results

Dep. Variable:	VIX_Close	**R-squared:**	0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	0.000
Vol Model:	GARCH	**Log-Likelihood:**	-6060.31
Distribution:	Normal	**AIC:**	12132.6
Method:	Maximum Likelihood	**BIC:**	12165.3
		No. Observations:	1709
Date:	Sun, Oct 17 2021	**Df Residuals:**	1708
Time:	22:52:10	**Df Model:**	1

Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	0.0911	0.195	0.466	0.641	[-0.292, 0.474]

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	23.0295	12.345	1.865	6.212e-02	[-1.167, 47.225]
alpha[1]	0.2338	0.106	2.216	2.671e-02	[2.699e-02, 0.441]
alpha[2]	0.0338	0.189	0.178	0.858	[-0.337, 0.405]
beta[1]	0.4715	0.540	0.873	0.383	[-0.587, 1.530]
beta[2]	0.0000	0.273	0.000	1.000	[-0.535, 0.535]

Covariance estimator: robust

Where: Alpha represents how volatility reacts to new information

Beta represents persistence of the volatility

Appendix 8

Constant Mean - ARCH Model Results

Dep. Variable:	VIX_Close	**R-squared:**	0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	0.000
Vol Model:	ARCH	**Log-Likelihood:**	-6077.09
Distribution:	Normal	**AIC:**	12162.2
Method:	Maximum Likelihood	**BIC:**	12184.0
		No. Observations:	1709
Date:	Sun, Oct 17 2021	**Df Residuals:**	1708
Time:	22:58:35	**Df Model:**	1

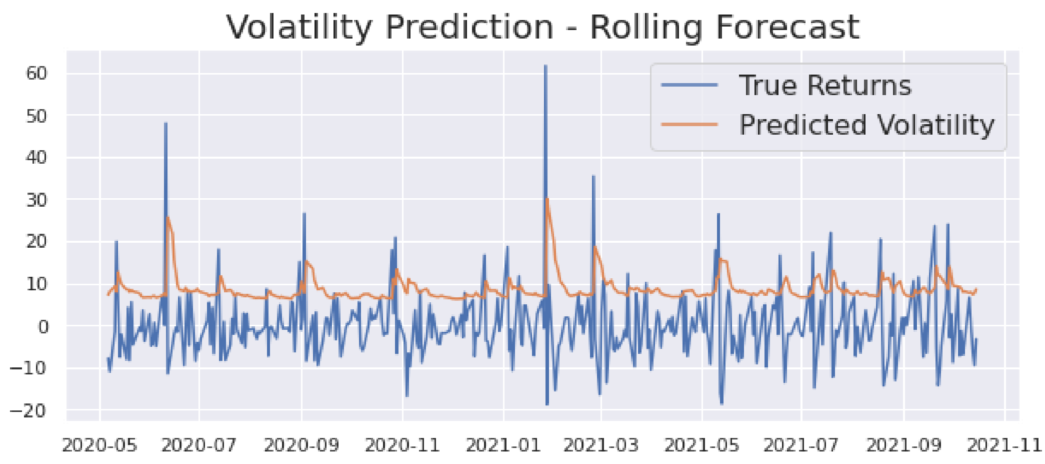
Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	0.1023	0.209	0.488	0.625	[-0.308, 0.513]

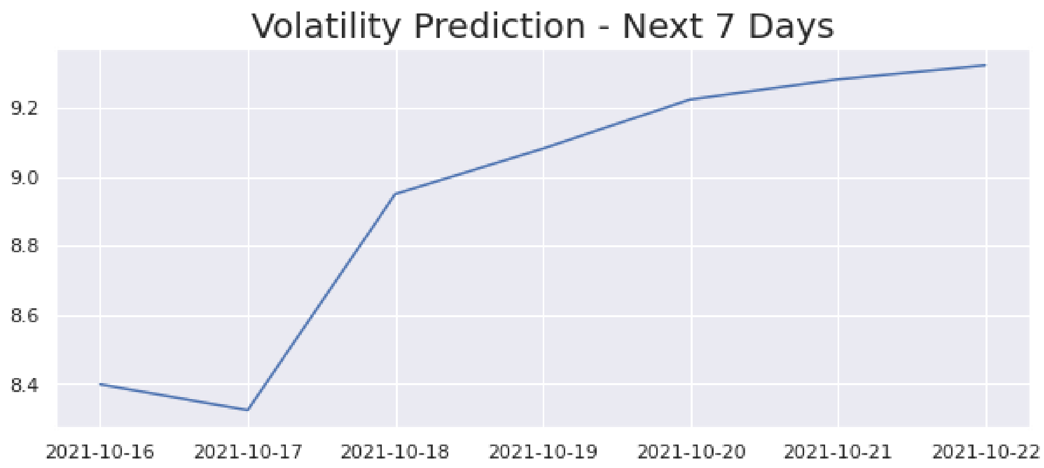
Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	50.0774	6.691	7.485	7.169e-14	[36.964, 63.191]
alpha[1]	0.2407	0.110	2.184	2.896e-02	[2.470e-02, 0.457]
alpha[2]	0.1892	9.044e-02	2.092	3.648e-02	[1.190e-02, 0.366]

Appendix 9



Appendix 10



Appendix 11

1. Monthly Unemployment Rate data – this represents the number of unemployed as a percentage of the labor force. Data from the FRED (Federal Reserve Economic Data).
2. Monthly CPI -– The Consumer Price Index for All Urban Consumers CPI from the FRED (Federal Reserve Economic Data).
3. Monthly PPI – Data from U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: All Commodities.
4. 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity (T10Y3M) – data from FRED.
5. IORR Interest Rate Spread – data from FRED.
6. COVID death data – data from The COVID Tracking Project API (Github)
7. Bitcoin Closing Price – this is self-explanatory. Data from Yahoo Finance.

8. Crude Oil prices – this is self-explanatory. Data from Yahoo Finance.

9. S&P Close – this is self-explanatory. Data from Yahoo Finance.

Regression Table 1 - multiple linear regression model with seven factors

	names	coef	se	T	pval	r2	adj_r2	CI[2.5%]	CI[97.5%]
0	Intercept	-47.015892	38.048221	-1.235692	2.187843e-01	0.694601	0.678282	-122.284348	28.252565
1	UNRATE	3.208102	0.226905	14.138555	3.899811e-28	0.694601	0.678282	2.759231	3.656973
2	CPI	-0.024114	0.442116	-0.054542	9.565868e-01	0.694601	0.678282	-0.898724	0.850496
3	SPY	-0.020285	0.024168	-0.839317	4.028206e-01	0.694601	0.678282	-0.068095	0.027526
4	OIL	-0.248770	0.046572	-5.341610	3.953055e-07	0.694601	0.678282	-0.340900	-0.156639
5	IORR	1.922145	0.825111	2.329558	2.135901e-02	0.694601	0.678282	0.289878	3.554412
6	PPIACO	0.346042	0.101324	3.415198	8.489769e-04	0.694601	0.678282	0.145599	0.546485
7	T10Y3M	-0.724245	0.894544	-0.809624	4.196243e-01	0.694601	0.678282	-2.493867	1.045377

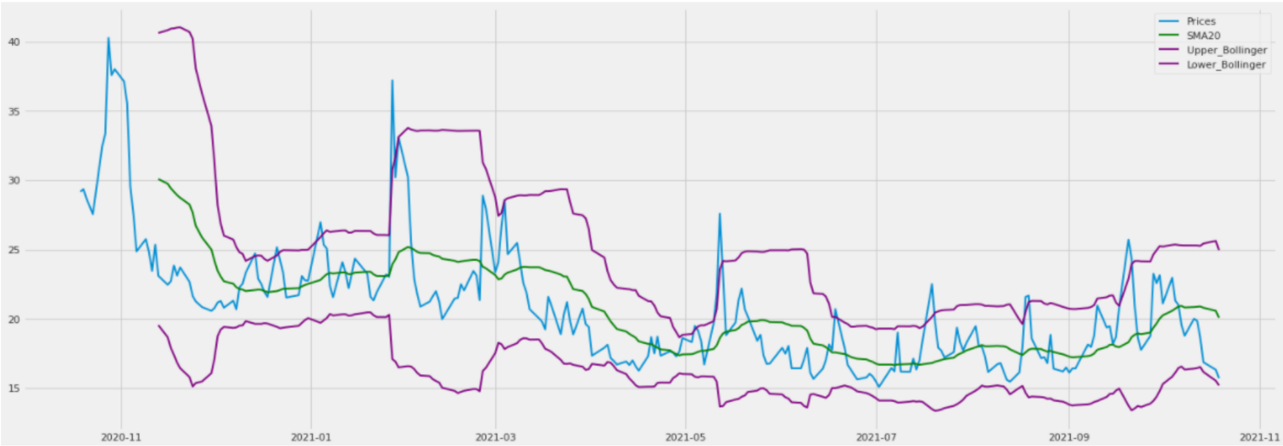
Regression Table 2 - multiple linear regression model with four factors

	names	coef	se	T	pval	r2	adj_r2	CI[2.5%]	CI[97.5%]
0	Intercept	-33.317766	8.726085	-3.818180	2.046856e-04	0.682987	0.673524	-50.576440	-16.059091
1	UNRATE	3.258084	0.218131	14.936352	2.614369e-30	0.682987	0.673524	2.826658	3.689509
2	OIL	-0.192827	0.019057	-10.118506	3.153379e-18	0.682987	0.673524	-0.230518	-0.155136
3	IORR	2.237775	0.630719	3.547976	5.357330e-04	0.682987	0.673524	0.990323	3.485227
4	PPIACO	0.215589	0.044764	4.816076	3.899207e-06	0.682987	0.673524	0.127053	0.304125

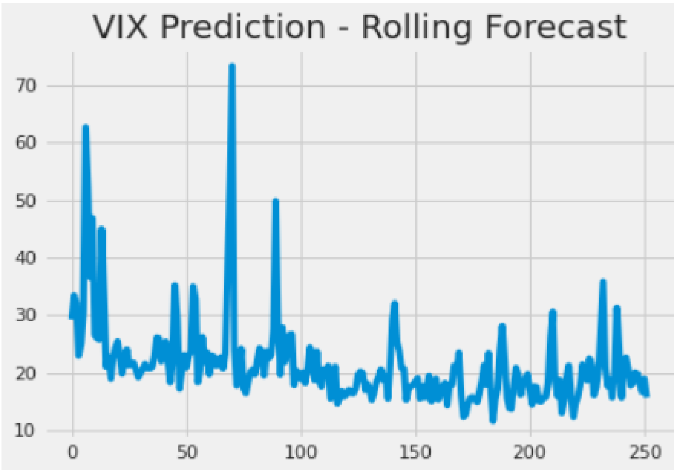
Regression Table 3 - multiple linear regression model with COVID & Bitcoin

	names	coef	se	T	pval	r2	adj_r2	CI[2.5%]	CI[97.5%]
0	Intercept	96.376106	3.475960	27.726475	2.937146e-95	0.684216	0.681066	89.542726	103.209486
1	SPY	-0.223946	0.016046	-13.956770	2.315879e-36	0.684216	0.681066	-0.255490	-0.192402
2	OIL	0.144315	0.054559	2.645118	8.487164e-03	0.684216	0.681066	0.037058	0.251572
3	US_covid	-0.000486	0.000320	-1.518935	1.295668e-01	0.684216	0.681066	-0.001114	0.000143
4	Bitcoin	0.000147	0.000032	4.555036	6.959621e-06	0.684216	0.681066	0.000084	0.000210

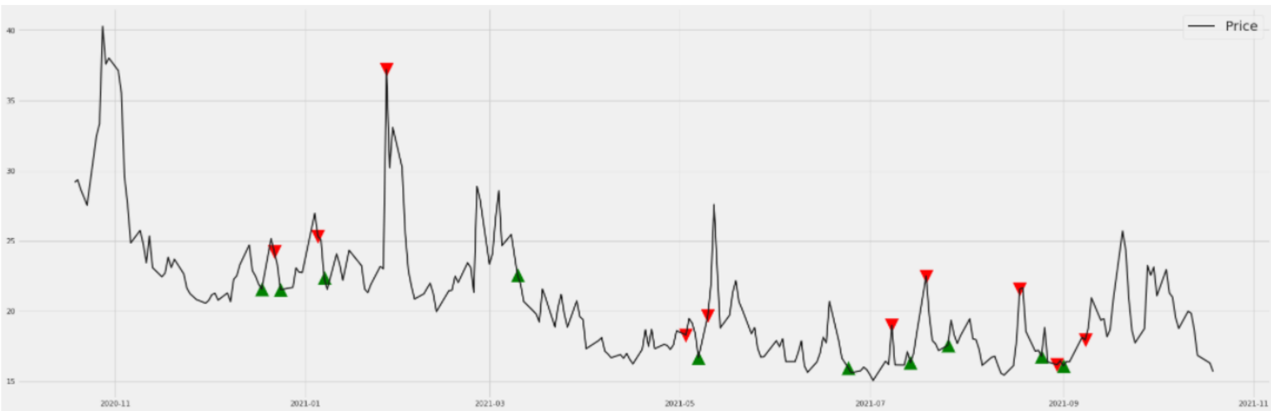
Appendix 12 Bollinger Bands



Appendix 13



Appendix 14 Order Simulation



	Type	Date	Price	Shares	Amount	Comission
19	Buy	2020-12-21	24.250000	463	11227.750000	112.277500
18	Sell	2020-12-23	23.490000	463	10875.869894	108.758699
17	Buy	2020-12-28	22.110001	448	9905.280273	99.052803
16	Sell	2021-01-06	25.480000	448	11415.039795	114.150398
15	Buy	2021-01-08	22.430000	498	11170.140152	111.701402
14	Sell	2021-01-28	33.250000	498	16558.500000	165.585000
13	Buy	2021-03-11	22.500000	730	16425.000000	164.250000
12	Sell	2021-05-04	18.160000	730	13256.799889	132.567999
11	Buy	2021-05-10	17.340000	799	13854.660122	138.546601
10	Sell	2021-05-11	21.170000	799	16914.830061	169.148301
9	Buy	2021-06-25	16.040001	1025	16441.000938	164.410009
8	Sell	2021-07-09	17.879999	1025	18326.999140	183.269991
7	Buy	2021-07-15	16.700001	1117	18653.900852	186.539009
6	Sell	2021-07-20	20.889999	1117	23334.129318	233.341293
5	Buy	2021-07-27	18.620001	1301	24224.621092	242.246211
4	Sell	2021-08-19	23.120001	1301	30079.121092	300.791211
3	Buy	2021-08-26	17.459999	1707	29804.218437	298.042184
2	Sell	2021-08-31	15.980000	1707	27277.859219	272.778592
1	Buy	2021-09-02	16.270000	1624	26422.480743	264.224807
0	Sell	2021-09-09	19.440001	1624	31570.560867	315.705609

Appendix 15 Python codes

Trading Strategy:

https://colab.research.google.com/drive/11XpTl4wWikqVHWsHZHIA_8KiSv02-x2B#scrollTo=5VHs-CO4pSD4