

A MOVING VARIANCE (2D) TIME SERIES MODEL

Mian Adnan & Silvey Shamsi
Indiana Univ. Bloomington, Ball State Univ.
Indiana
maadnan@iu.edu

Abstract

Financial turmoil is a fear or a lucrative feature for a latent or a set of latent reasons to the investors or money makers respectively. But obtaining a quick prediction of a threshold or volatility is a big challenge from a huge data that continuously gets bigger with its frame. A step by step approach of quickly identifying the model for the most important latent variable has been inaugurated for demonstrating the capricious behavior of the time series pattern of the original data using the optimum number of predictor(s). There are several methods in Time Series Analyses viz Moving Average Method, etc. Attempts have been made here to develop a time series model (along with the optimum number of set of characteristics or parameters) that can predict the stock prices' pattern as well as a volatility (or volatilities). It is also called the 2D time series model since it adopts the Moving Variance Approach.

Key Words: Adjusted R-square, BIC, Correlation, Granger's causality test, Spectrum, VARMAX Model.



Why?

Volatility in Stock Prices?

It signifies financial turmoil.

How?

To measure Volatility in Stock Prices?

Observing Volatility Index (VIX) as introduced by Chicago Board Option Exchange (CBOE).

Observed by Whom?

Investors since VIX is often referred as the “investor fear gauge.”

Widely being followed by financial companies, news articles in the Wall Street Journal, Barron’s and other leading financial publications.

What to Follow?

Wide range of strike prices of S&P 500 index option prices.

Volatility in these Strike Prices as measured by VIX heuristic, demonstrates the market's expectation and latest thinking of industry practitioners.

Solution?

How and Where Does S&P 500 index Stock Price Dance Much Over Time?

A Moving Variance - Time Series Analyses of the Volatility of the Stock Prices

1. Introduction

All financial organizations want to predict the financial turmoil or volatility for implementing its short run or long run derivatives and/or prerequisites as early as possible. Volatility in Stock Prices signifies the financial turmoil. Knowing the behavior of S&P 500 index Stock Price over time demonstrates the market's expectation and latest thinking of industry practitioners. However, at every time span, the changing behavior of the huge

density of stock prices indicates new direction(s) in the stock market.

Volatility in Stock Prices can be measured observing the Volatility Index, VIX heuristic, as introduced by Chicago Board Option Exchange (CBOE). It is observed by the investors since VIX is often referred as the “investor fear gauge.” The VIX is widely being followed by financial companies, news articles in the Wall Street Journal, Barron’s and other leading financial publications. Wide range of strike prices of S&P 500 index option prices are seriously followed by money makers.

A time series analyses of the of the S&P 500 index strike price-changes over time can give a message of volatility. As such the topic of this paper is the time series analyses of the S & P stock prices. The resultant time series model checks the moving variances of the residuals to identify which set of few time points contribute the highest variation in the prices. The money makers want to predict these time points.

2. Methods and Methodology

It is better to first assess the time series pattern like trend or cyclic or seasonal or irregular variation of the data for each of the key variables of the data. The scatter plot or any other graph can be some options to do so. From the summary and descriptive statistics of the data, the relationship among the key variables can be guessed. Standardization and detrending of the key variables should be performed for increasing the compatibility of them. Optimum lag period of the VARMAX (Vector Auto Regressive Model) (Shumway *et al* [2]) model can be found for the data along-with the lowest BIC.

The best response variable and the best overall set of predictors for the VARMAX model having the predefined lag with the highest adjusted R-square for the specific regressed variable can be obtained. All of the variables within the VARMAX model should be checked whether all have significant contribution in prediction using Granger's individual causality test. All possible combinations of the set of predictors and the subset of the predictors for various lags lower than or

equal to the optimum lag obtained (from the optimum VARMAX model) for the selected response variable should be generated.

Later all possible combination of the simple type model like ARIMA (p,r,q) (or SARIMA if there is any cyclic pattern in the time series data evident from graph, etc) for all possible combination of lags of the lower than or equal to the optimum lag obtained (from the optimum VARMAX model) for each of the values of p,r,q for the ARIMA (or SARIMA) models can be generated. For each of the set of the combination of predictors, the ARIMA model which has the lowest BIC should be selected. The final ARIMA model having the lowest BIC (Bayesian Information Criterion) ARIMA models should be selected among the least BIC ARIMA models obtained from each combination of the set of predictors.

The Auto Correlation Function Plot of the residual (data – fitted model) should be checked whether it diminishes for lags other than 0. The Spectrum plot should be checked whether it has the uniform spikes pattern showing the evidence of the white noise pattern of the residuals.

The overall steps of finding a model to fit can be demonstrated in the following 12 steps which are as below.

Step 1: Recognize the time series pattern of the data.

Step 2: Assess the relationship among the key variables.

Step 3: Detrend and standardize each key variable.

Step 4: Find the optimum lag period of the VARMAX model having the lowest BIC.

Step 5: Find the set of response variable and the set of predictors for the VARMAX model having the highest adjusted R-square.

Step 6: Check all of the variables have significant contribution in prediction using Granger's Individual Causality Tests.

Step 7: Generate all possible combinations of the set of predictors lower than or equal to the optimum lag.

Step 8: Generate all possible combination of ARIMA (p, r, q) (or SARIMA, etc) for all possible combination of lower than or equal to the optimum lag.

Step 9: Find the ARIMA having the lowest BIC for each of the set of the combination of predictors.

Step 10: Find the final ARIMA model which has the grand lowest BIC of the lower BIC ARIMA models.

Step 11: Check the Auto Correlation Function Plot, the Spectrum plot, Residual plot of the residuals are similar to those of the white noise.

Step 12: Check the Moving Variance for various (time) intervals to identify the smallest set of time points that contribute the highest variation.

3. Data Structure and Pattern Recognition

Data of Daily Opening, Closing, Higher and Lower S & P 500 Index Stock Prices from January 2, 2004 to April 20, 2018 have been collected from the public data hub (addressed as <https://datahub.io/core/finance->

vix#readme[1]). There are 3600 time points. From the summary of the data (Appendix: Table A1), each of the four variables', of VIX opening price (vo), VIX closing price (vc), VIX high price (vh), VIX low price (vl), values range from 8 to 90 and have locations around 19. They have linear pairwise-relation (Figure A1) and roughly similar shapes over time (Figure A2).

Now, to select first the optimum lag and secondly the most reliable response variable and the set of best predictors, several Vector Auto Regressive Models (ARIMAX) having various lags ranging from 1 to 13 have been constructed. For lag period 3, the VARMAX model gives the lowest BIC (-26581) and the VARMAX (x, p = 3) model gives the highest adjusted R-square (0.9898) for the response variable VIX opening price (vo) (Table A2). So, VIX opening price is the best response variable out of the four available variables. From the related correlation matrix of the residuals (from Table A3) for the VARMAX (x, p = 3) model, the residuals for variable vo has moderate linear relationship with those for other predictors. However, from the individual Granger's causality test also shows that each of the variables has significant contribution in prediction using the VARMAX model (displayed in Table A4).

12 possible combinations of the set of predictors and the subset of the predictors have been considered for various lags lower than or equal to the optimum lag obtained (from the optimum VARMAX model) for the selected response variable. Those combinations are addressed below. Here, as for example vc_1 means the predictor vc with lag 1. T and $T2$ are the two new predictors to see whether time and time-square have any effect in prediction.

$$X1 = \text{cbind}(vc_1,vc_2,vc_3,vh_1,vh_2,vh_3,vl_1,vl_2,vl_3,T,T2)$$

$$X3 = \text{cbind}(vc_1,vc_2,vh_1,vh_2,vl_1,vl_2,T,T2)$$

$$X4 = \text{cbind}(vc_1,vh_1,vl_1,T,T2)$$

$$X5 = \text{cbind}(vc_1,vh_1,vl_1)$$

$$X6 = \text{cbind}(vc_1,vc_2,vc_3,vh_1,vh_2,vh_3,vl_1,vl_2,vl_3)$$

$$X7 = \text{cbind}(vc_1,vc_2,vh_1,vh_2,vl_1,vl_2)$$

$$X8 = \text{cbind}(vc_1,vh_1)$$

$$X9 = \text{cbind}(vc_1,vl_1)$$

$$X10 = \text{cbind}(vl_1,vh_1)$$

$$X11 = \text{cbind}(vc_1)$$

$$X12 = \text{cbind}(vl_1)$$

$$X13 = \text{cbind}(vh_1).$$

The following 42 possible combination of ARIMA (p, r, q) [for different combination of p, r, q] have been considered for lag 3 ($p = 3$) of the lower than or equal to

the optimum lag obtained (from the optimum VARMAX model).

```
m1 = arima(vo, xreg=X1, order=c(1,0,1))#ARIMA(1,0,1)
m2 = arima(vo, xreg=X1, order=c(0,0,1))#ARIMA(0,0,1)
m3 = arima(vo, xreg=X1, order=c(1,0,0))#ARIMA(1,0,0)
m4 = arima(vo, xreg=X1, order=c(1,1,1))#ARIMA(1,1,1)
m5 = arima(vo, xreg=X1, order=c(2,0,2))#ARIMA(2,0,2)
m6 = arima(vo, xreg=X1, order=c(0,2,2))#ARIMA(0,2,2)
m7 = arima(vo, xreg=X1, order=c(2,2,0))#ARIMA(2,2,0)
m8 = arima(vo, xreg=X1, order=c(2,1,2))#ARIMA(2,1,2)
m9 = arima(vo, xreg=X1, order=c(2,2,1))#ARIMA(2,1,3)
m10 = arima(vo, xreg=X1, order=c(1,2,2))#ARIMA(1,2,2)
m11 = arima(vo, xreg=X1, order=c(1,2,2))#ARIMA(1,2,2)
m12 = arima(vo, xreg=X1, order=c(2,2,2))#ARIMA(2,2,2)
m13 = arima(vo, xreg=X1, order=c(3,0,0))#ARIMA(3,0,0)
m14 = arima(vo, xreg=X1, order=c(0,3,0))#ARIMA(0,3,0)
m15 = arima(vo, xreg=X1, order=c(0,0,3))#ARIMA(0,0,3)
m16 = arima(vo, xreg=X1, order=c(3,0,1))#ARIMA(3,0,1)
m17 = arima(vo, xreg=X1, order=c(3,1,0))#ARIMA(3,1,0)
m18 = arima(vo, xreg=X1, order=c(1,3,0))#ARIMA(1,3,0)
m19 = arima(vo, xreg=X1, order=c(0,3,1))#ARIMA(0,3,1)
m20 = arima(vo, xreg=X1, order=c(1,0,3))#ARIMA(1,0,3)
m21 = arima(vo, xreg=X1, order=c(0,1,3))#ARIMA(0,1,3)
m21 = arima(vo, xreg=X1, order=c(3,0,2))#ARIMA(3,0,2)
m22 = arima(vo, xreg=X1, order=c(3,2,0))#ARIMA(3,2,0)
m23 = arima(vo, xreg=X1, order=c(2,3,0))#ARIMA(2,3,0)
m24 = arima(vo, xreg=X1, order=c(0,3,2))#ARIMA(0,3,2)
m25 = arima(vo, xreg=X1, order=c(2,0,3))#ARIMA(2,0,3)
```

```

m26 = arima(vo, xreg=X1, order=c(0,2,3))#ARIMA(0,2,3)
m27 = arima(vo, xreg=X1, order=c(3,1,2))#ARIMA(3,1,2)
m28 = arima(vo, xreg=X1, order=c(3,2,1))#ARIMA(3,2,1)
m29 = arima(vo, xreg=X1, order=c(2,3,1))#ARIMA(2,3,1)
m30 = arima(vo, xreg=X1, order=c(1,3,2))#ARIMA(1,3,2)
m31 = arima(vo, xreg=X1, order=c(2,1,3))#ARIMA(2,1,3)
m32 = arima(vo, xreg=X1, order=c(1,2,3))#ARIMA(1,2,3)
m33 = arima(vo, xreg=X1, order=c(3,0,3))#ARIMA(3,0,3)
m34 = arima(vo, xreg=X1, order=c(3,3,0))#ARIMA(3,3,0)
m35 = arima(vo, xreg=X1, order=c(0,3,3))#ARIMA(0,3,3)
m36 = arima(vo, xreg=X1, order=c(1,3,3))#ARIMA(1,3,3)
m37 = arima(vo, xreg=X1, order=c(3,1,3))#ARIMA(3,1,3)
m38 = arima(vo, xreg=X1, order=c(3,3,1))#ARIMA(3,3,1)
m39 = arima(vo, xreg=X1, order=c(3,3,2))#ARIMA(3,3,2)
m40 = arima(vo, xreg=X1, order=c(3,2,3))#ARIMA(3,2,3)
m41 = arima(vo, xreg=X1, order=c(3,3,2))#ARIMA(3,3,2)
m42 = arima(vo, xreg=X1, order=c(3,3,3))#ARIMA(3,3,3)

```

For each of the set of the combination (from X1 to X13) of predictors, the 42 ARIMA models have been constructed. The BIC for each of the 42 ARIMA model have been found and compared among themselves (42 models) to figure out which ARIMA model gives the lowest BIC for a specific combination of any of the X1 to X13 set of predictors. This process has been repeated for all possible combination of the set of predictors. Out of 504 (= 42*12) ARIMA models, 12 ARIMA models

with the corresponding lowest BIC have been found (from Table A5). These are as follows.

`arima(vo, xreg=X1, order=c(1,0,3))#ARIMA(1,0,3)` with BIC -6197,
`arima(vo, xreg=X3, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6228,
`arima(vo, xreg=X4, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6238,
`arima(vo, xreg=X5, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6245,
`arima(vo, xreg=X6, order=c(1,0,3))#ARIMA(1,0,3)` with BIC -6222,
`arima(vo, xreg=X7, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6244,
`arima(vo, xreg=X8, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6146,
`arima(vo, xreg=X9, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6262,
`arima(vo, xreg=X10, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -4573,
`arima(vo, xreg=X11, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -6153,
`arima(vo, xreg=X12, order=c(1,0,1))#ARIMA(1,0,1)` with BIC -3734,
`arima(vo, xreg=X13, order=c(2,0,2))#ARIMA(2,0,2)` with BIC -3885,

Out of these 12 lower BIC ARIMA models, `arima(vo, xreg=X9, order=c(1,0,1))#ARIMA(1,0,1)` has the lowest BIC -6262. This is the ARIMA (1, 0, 1) model with the predictors VIX Closing Price (`vc`) for lag 1 and VIX Low Price for lag 1. The mathematical form of the ARIMA (1, 0, 1) model for the aforesaid predictors for the response VIX Opening Price (`vo`) is

$$VIXOpen(t) = \beta_0 + \beta_c VIXClose(t-1) + \beta_l VIXLow(t-1) + \text{Noise},$$

where, Noise = *ARIMA* (1,0,1) Colored Noise = $U(t) = \alpha U(t - 1) + \theta_1 w(t) + \theta_2 w(t - 1)$. The predicted *ARIMA* (1, 0, 1) model is

$$vo(t) = 0.86 vc(t-1) + 0.14 vl(t-1) + 0.75 u(t-1) + \theta_1 w(t) - 0.83 w(t - 1),$$

where, $\theta_1 < \sqrt{0.01013}$. Here, $u(t-1)$ is the colored noise for lag 1 and $w(t)$ and $w(t-1)$ are the white noises for no lag and lag 1 respectively.

The auto correlation function of the residuals obtained from the difference of the data and the fitted *ARMA* (1,1) model has been found diminishing as lag increase (in Figure A3). This is a sign of the residuals that are white noise means the fitted model is a good fit. The spectrum plot of the residuals also shows that they have uniform spikes except in the frequency at 0.32 and 0.48 (observed from Figure A5). There is a cycle for every 3 ($= 1/0.32$) days and a moderate cycle for every 2 ($=1/0.48$) days in the market. 1595, 1596 and 1597 are the time points where the variance is the highest and money makers to know these time points ahead of the turmoil or volatility (Figure A9 to Figure A14).

Conclusion

If we have the same data frame for the next 14 years in future, the predicted model for this data set will generate a time series graph (Figure A7) which is similar to that of the original data set for past 14 years (Figure A6). So, the predicted model seems to be a good fit to represent the original time series data. A quick step by step approach of finding a proper time series model that uses Moving Variance Method has been addressed here.

APPENDIX

Table A1: Summary of Data

time	VIXOpen	VIXHigh	VIXLow
VIXClose			
Min. : 1.0	Min. : 9.01	Min. : 9.31	Min. : 8.56
Min. : 9.14			
1st Qu.: 900.8	1st Qu.:13.07	1st Qu.:13.53	1st Qu.:
12.61	1st Qu.:12.98		
Median :1800.5	Median :15.88	Median :16.55	Med
ian :15.22	Median :15.77		
Mean :1800.5	Mean :18.62	Mean :19.46	Mean
:17.77	Mean :18.51		
3rd Qu.:2700.2	3rd Qu.:21.18	3rd Qu.:22.21	3rd Q
u.:20.23	3rd Qu.:21.04		
Max. :3600.0	Max. :80.74	Max. :89.53	Max. :
72.76	Max. :80.86		

Figure A1: Scatter Plot

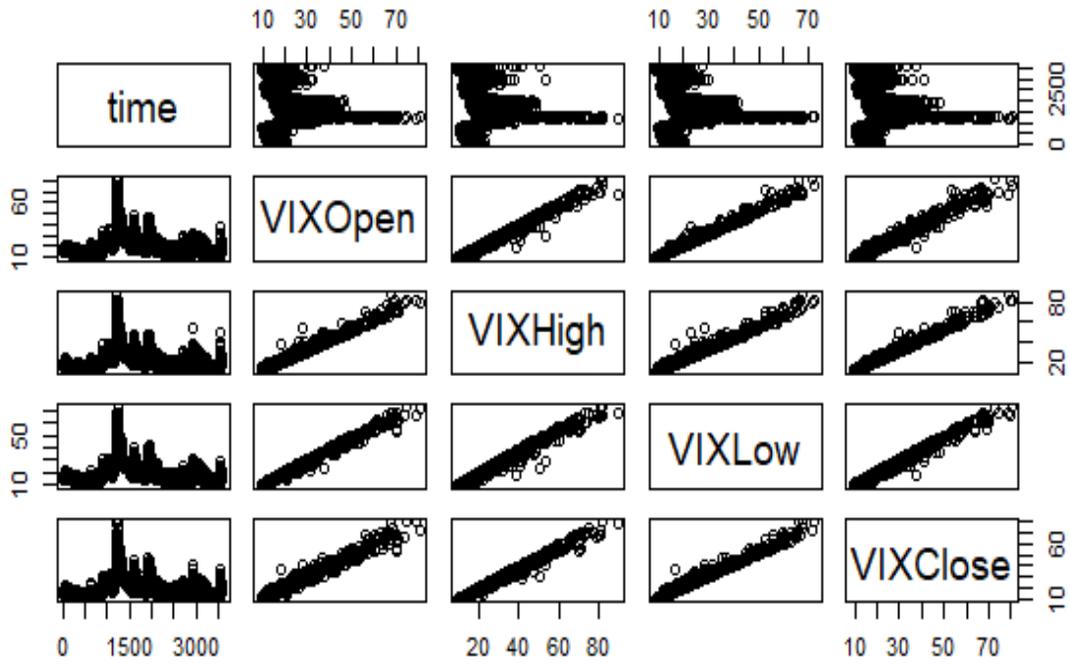


Figure A2: Time Series Plot

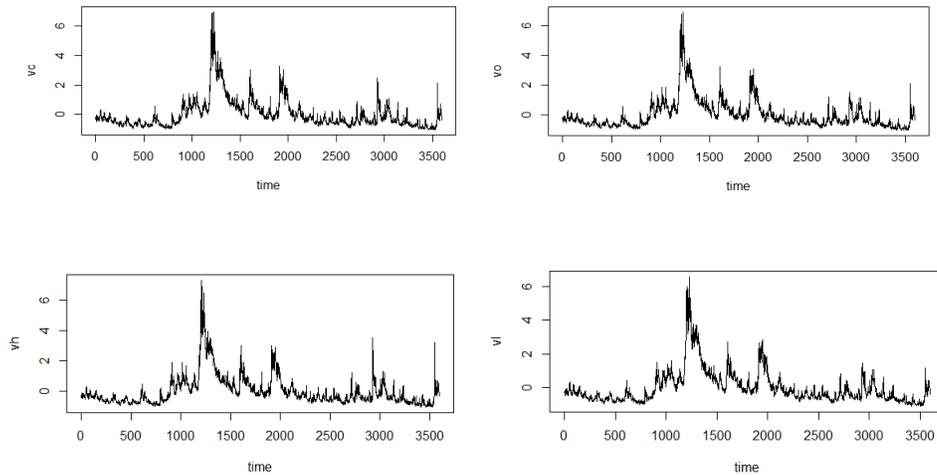


Table A2: BIC of 11 Vector Auto Regressive Models (VARs (x, p)) have been found for p = lags = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Dependent Variable	p=1	2	3	4	5	6	7	8	9	10	p=11
vo			0.98								
vc			0.96								
vl			0.98 27								
vh			0.97								
BIC	- 26 39 7	- 26 52 6	- 265 81	- 26 56 5	- 26 54 1	- 26 45 7	- 26 38 8	- 26 33 2	- 26 26 4	- 26 19 6	- 26 08 5

Table A3: Correlation Matrix of residuals from VARMAX model ($x, p = 3$)

Correlation matrix of residuals:

	vo	vc	vh	vl
vo	1.0000	0.5355	0.6487	0.6832
vc	0.5355	1.0000	0.8559	0.8145
vh	0.6487	0.8559	1.0000	0.7159
vl	0.6832	0.8145	0.7159	1.0000

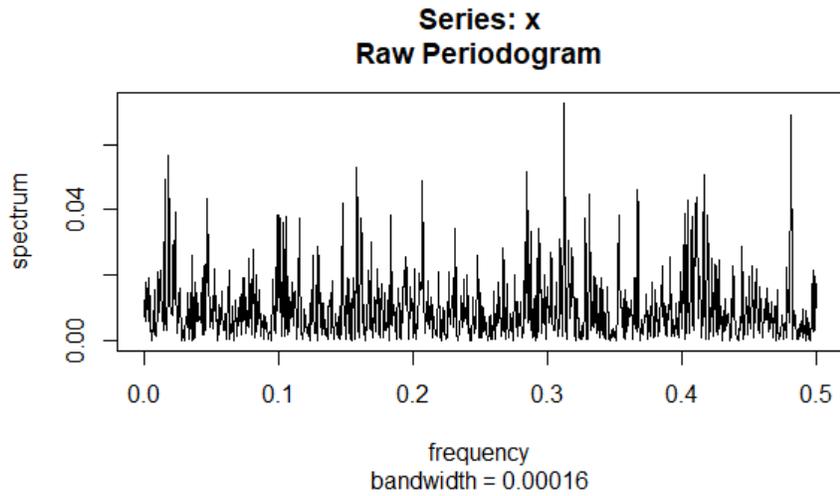
Table A4: Granger's Tests

Granger test for the predictor	p-value	Decision
causality(vR, cause = "vo")\$Granger	7.42e-09	vo has significant role in prediction
causality(vR, cause = "vc")\$Granger	2.2e-16	vc has significant role in prediction
causality(vR, cause = "vh")\$Granger	2.2e-16	vh has significant role in prediction
causality(vR, cause = "vl")\$Granger	2.2e-16	vl has significant role in prediction

Table A5: BICs of $42 \times 12 = 504$ Time Series Models

	X1	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
arma(vo, xreg, order=c(1,0,1))#ARIMA(1,0,1)		- 6228	- 6238	- 6254		- 6244	- 6146	- 6262	- 4573	- 6153	- 3734	
arma(vo, xreg, order=c(0,0,1))#ARIMA(0,0,1)												
arma(vo, xreg, order=c(1,0,0))#ARIMA(1,0,0)												
arma(vo, xreg, order=c(1,1,1))#ARIMA(1,1,1)												
arma(vo, xreg, order=c(2,0,2))#ARIMA(2,0,2)												- 3885
arma(vo, xreg, order=c(0,2,2))#ARIMA(0,2,2)												
arma(vo, xreg, order=c(2,2,0))#ARIMA(2,2,0)												
arma(vo, xreg, order=c(2,1,2))#ARIMA(2,1,2)												
arma(vo, xreg, order=c(2,2,1))#ARIMA(2,1,3)												
arma(vo, xreg, order=c(1,2,2))#ARIMA(1,2,2)												
arma(vo, xreg, order=c(1,2,2))#ARIMA(1,2,2)												
arma(vo, xreg, order=c(2,2,2))#ARIMA(2,2,2)												
arma(vo, xreg, order=c(3,0,0))#ARIMA(3,0,0)												
arma(vo, xreg, order=c(0,3,0))#ARIMA(0,3,0)												
arma(vo, xreg, order=c(0,0,3))#ARIMA(0,0,3)												
arma(vo, xreg, order=c(3,0,1))#ARIMA(3,0,1)												
arma(vo, xreg, order=c(3,1,0))#ARIMA(3,1,0)												
arma(vo, xreg, order=c(1,3,0))#ARIMA(1,3,0)												
arma(vo, xreg, order=c(0,3,1))#ARIMA(0,3,1)												
arma(vo, xreg, order=c(1,0,3))#ARIMA(1,0,3)	- 6197				- 6222							
arma(vo, xreg, order=c(0,1,3))#ARIMA(0,1,3)												
arma(vo, xreg, order=c(3,0,2))#ARIMA(3,0,2)												
arma(vo, xreg, order=c(3,2,0))#ARIMA(3,2,0)												
arma(vo, xreg, order=c(2,3,0))#ARIMA(2,3,0)												

Figure A4: Spectrum of the residuals



Highest spike at $\omega = 0.32$. One cycle in every 3 days.

Figure A5: Residual Plot

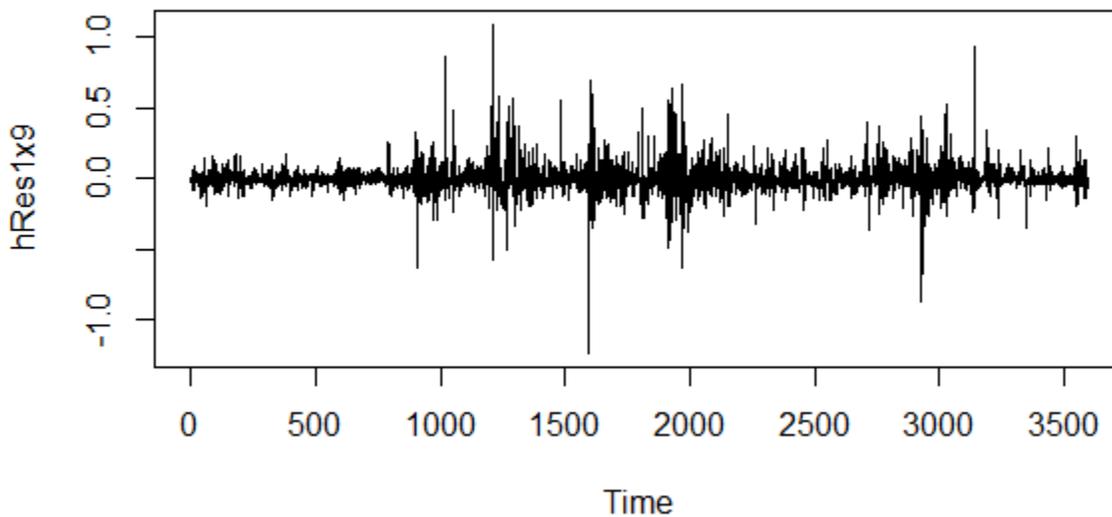


Figure A6: Time Series Plot of the Daily Opening Price of S&P 500 Stock Index Option Prices from January 2, 2004 to April 20, 2018

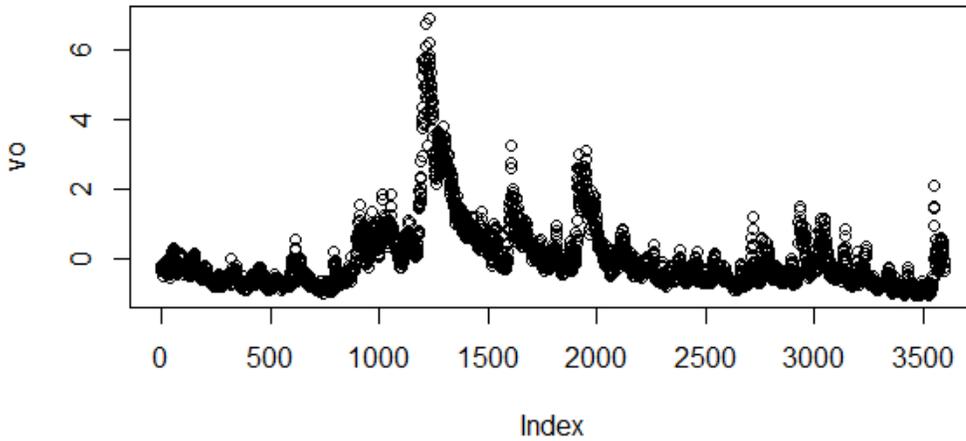


Figure A7: Predicted Time Series Plot of the Daily Opening Price of S&P 500 Stock Index Option Prices

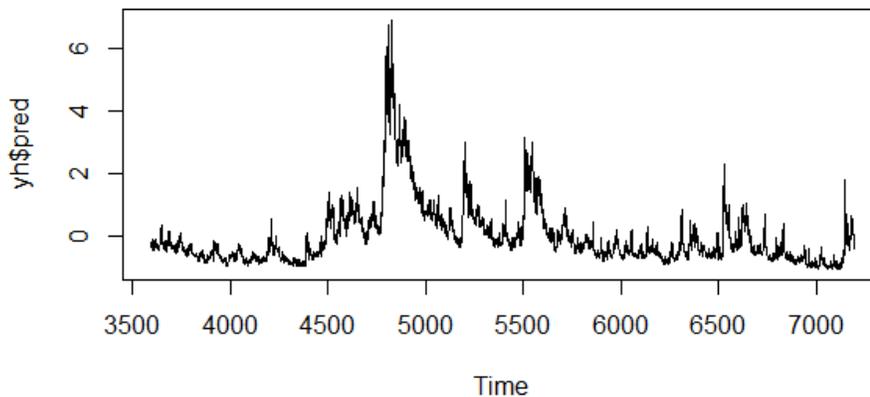


Figure A8: 3-point Moving Average Plot

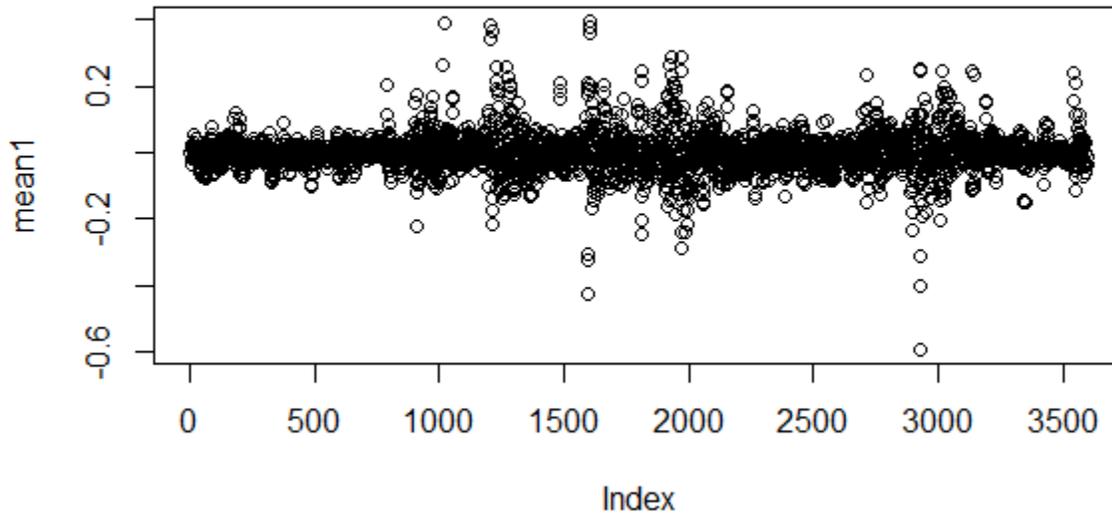


Figure A9: 3-point Moving Variance Plot

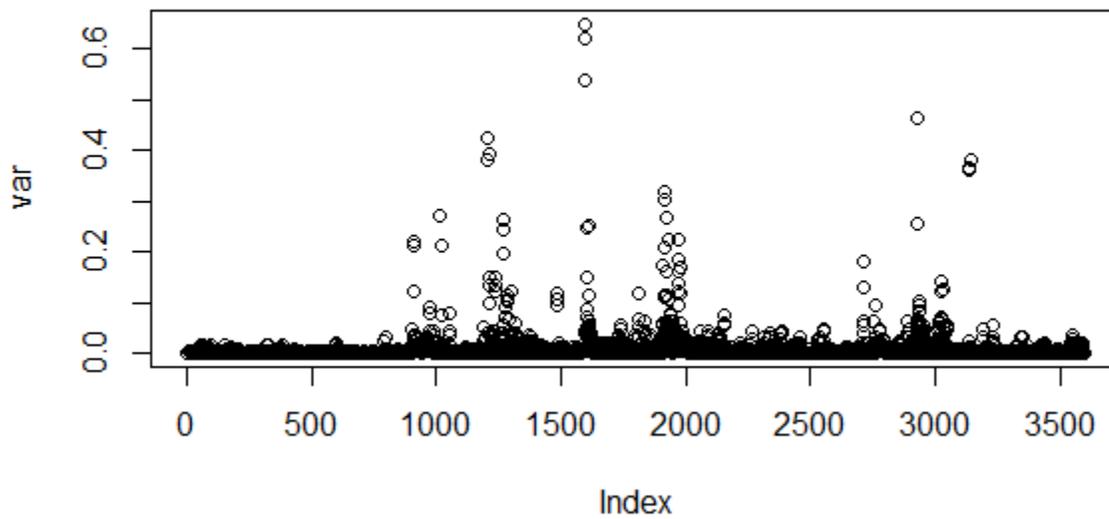


Figure A10: 5-point Moving Variance Plot

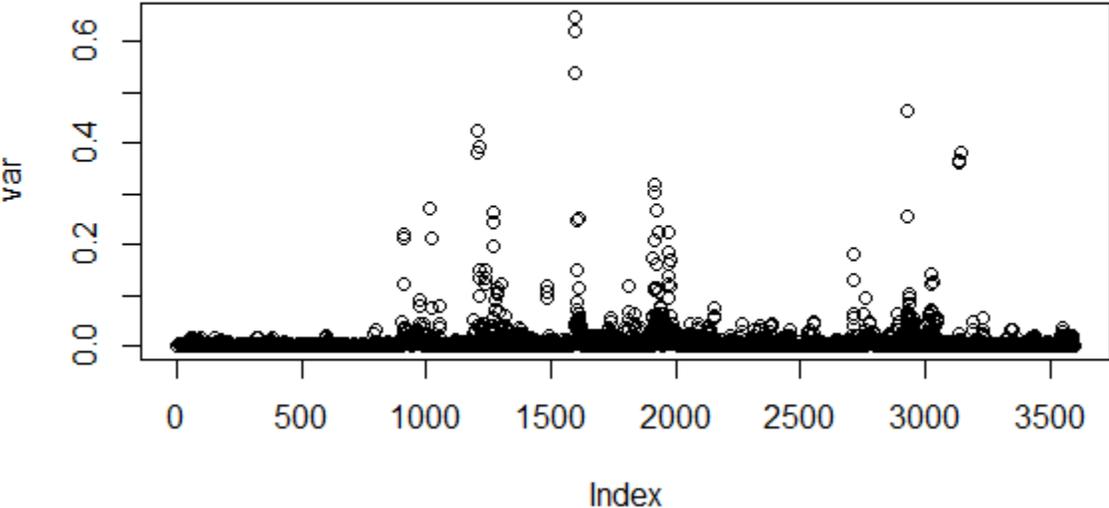


Figure A11: 21-point Moving Variance Plot

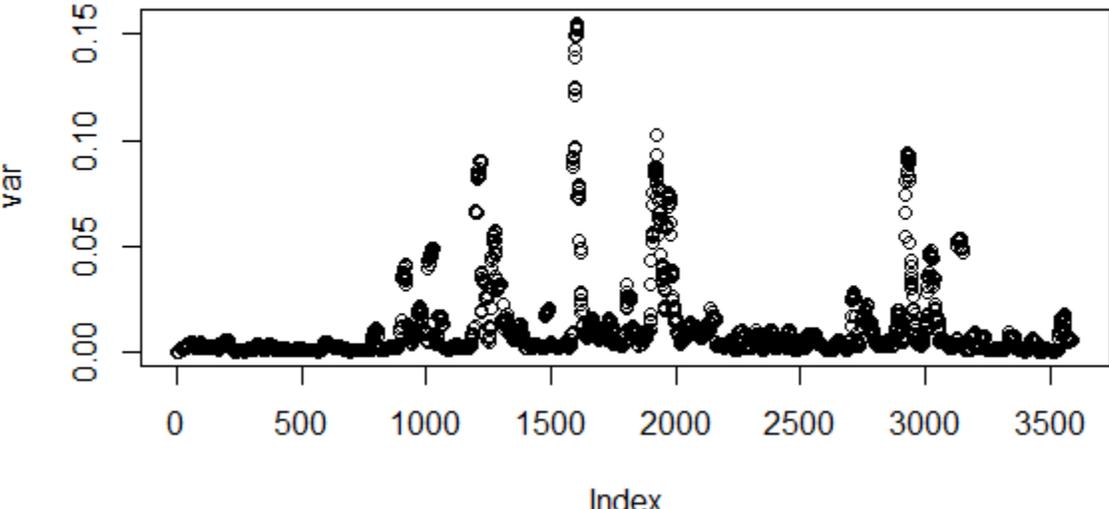


Figure A12: 201-point Moving Variance Plot

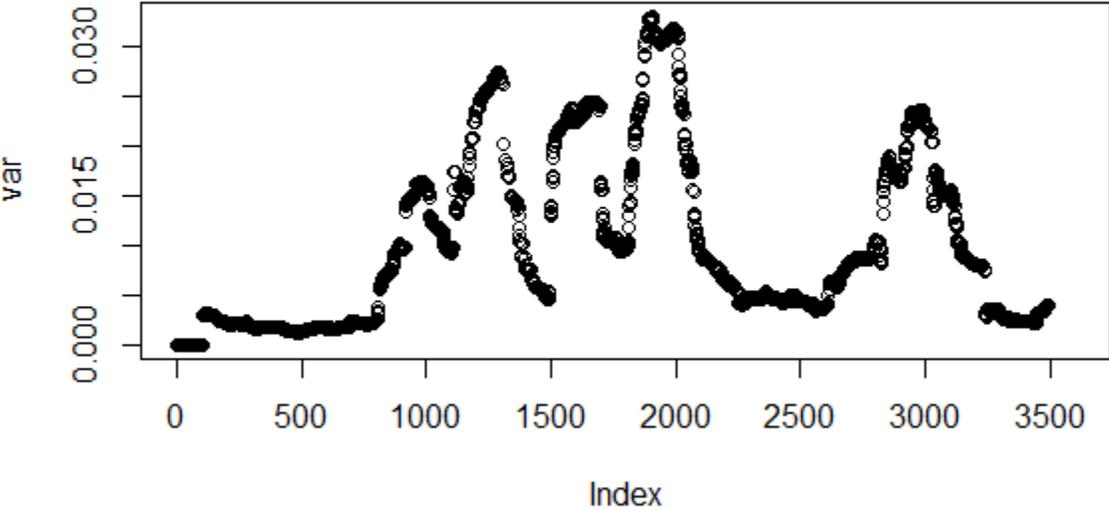


Figure A13: 1001-point Moving Variance Plot

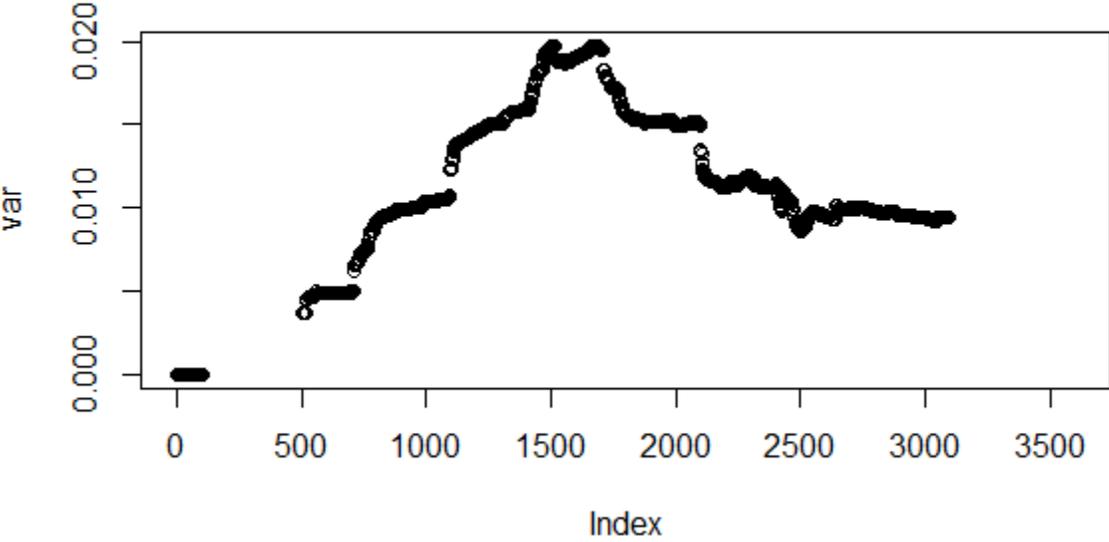


Figure A14: Auto Correlation Function Plot of First Order Difference vo

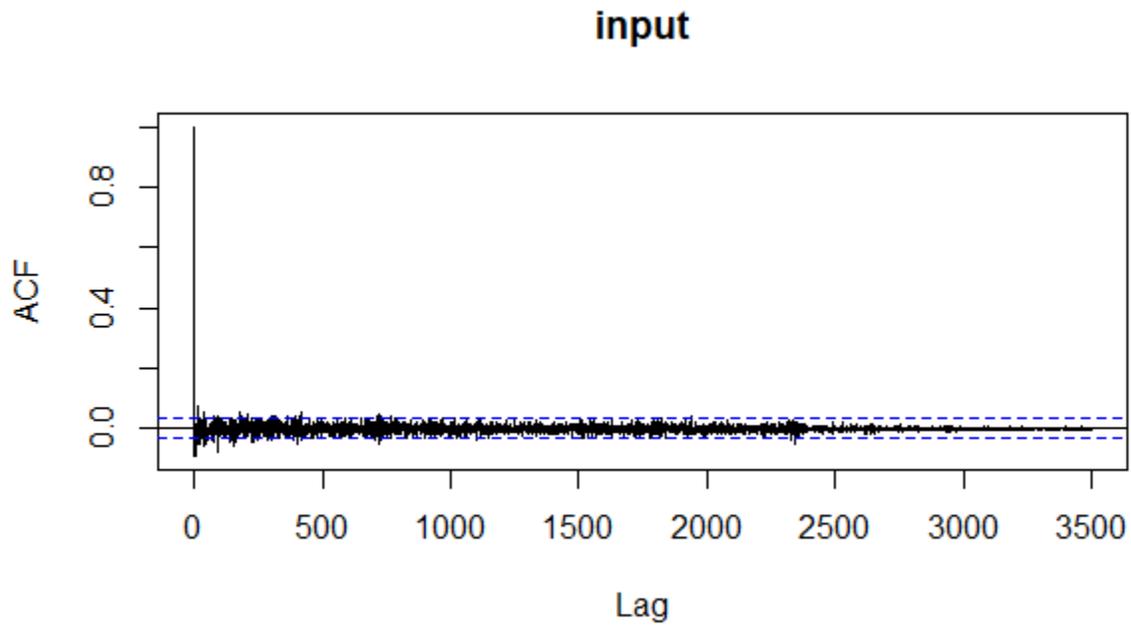
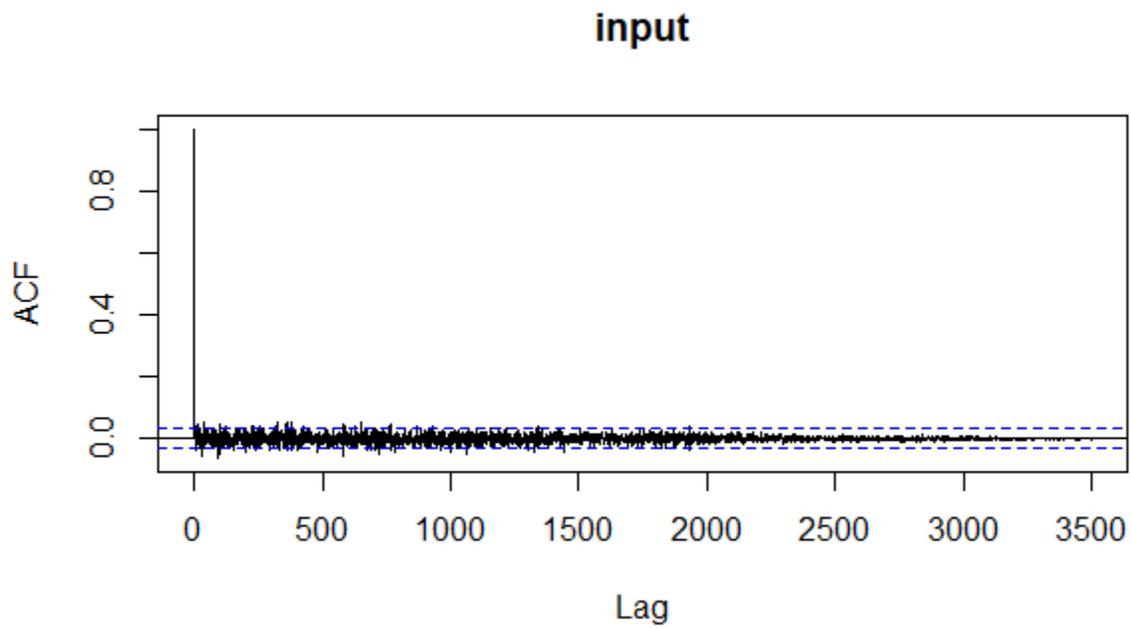


Figure A15: Auto Correlation Function Residual



References

- [1] <https://datahub.io/core/finance-vix#readme>
- [2] Shumway, *et al.* (2016). Time Series Analysis and its Applications. Springer. 3rd Edition.

THANK YOU!!

