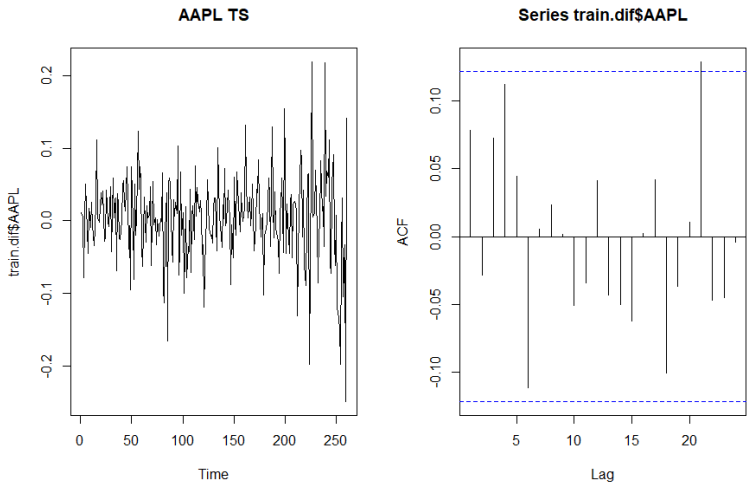
ISYE 6402 – Spring 2020
Georgia Institute of Technology

Final Exam Solutions

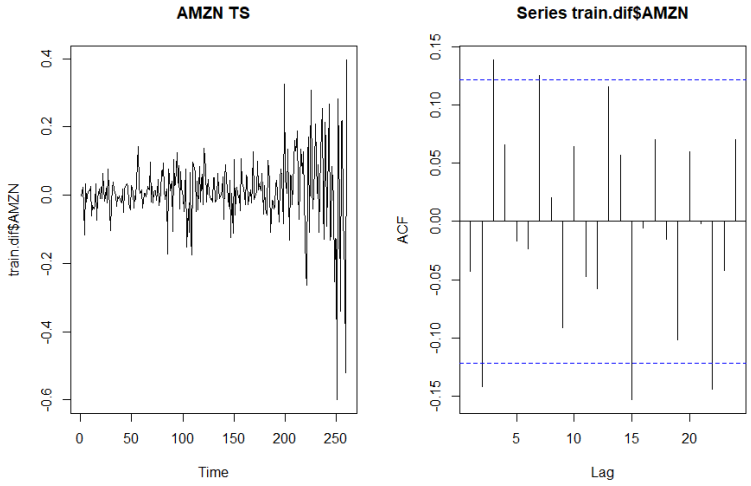
Question 1 – 12 points

A.

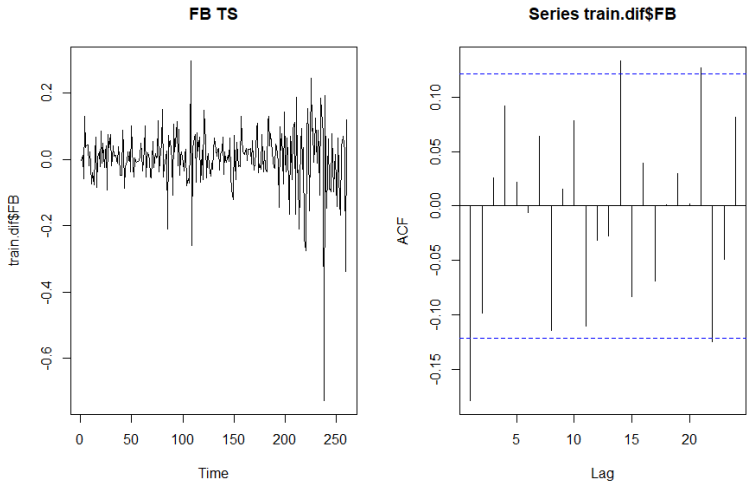
Note: ACF plots below built using library that starts charts at 1st lag



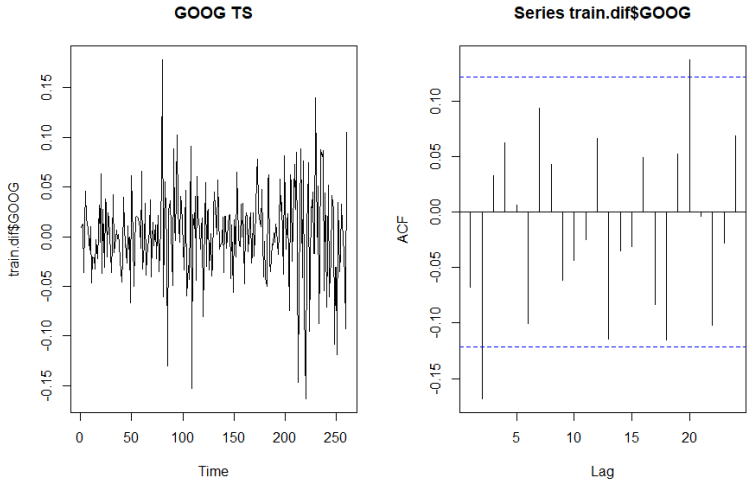
Here there is some possibility for heteroscedasticity at the end, but otherwise appears weakly stationary.



Clear heteroscedasticity near the end of the recorded series but mean appears constant. ACF results inconclusive.



Despite a few isolated spikes, appears to have constant mean and variance. Single significant ACF lag.



Variance stability and ACF interpretation inconclusive, mean appears constant.

All reasonable conclusions with proper charts should receive credit, even if not identical to above.

B.

AMZN Quadratic Polynomial:

```
Call:
lm(formula = train$AMZN ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3211 -0.1359 -0.0212  0.1420  0.9143

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.13422    0.05064  -2.651  0.00853 **
x1          -0.50613    0.23484  -2.155  0.03207 *
x2           4.40131    0.22820  19.287  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2748 on 258 degrees of freedom
Multiple R-squared:  0.948,     Adjusted R-squared:  0.9476
F-statistic:  2352 on 2 and 258 DF,  p-value: < 2.2e-16
```

Mean Absolute Error: .194

Mean Squared Error: .0746

FB Splines:

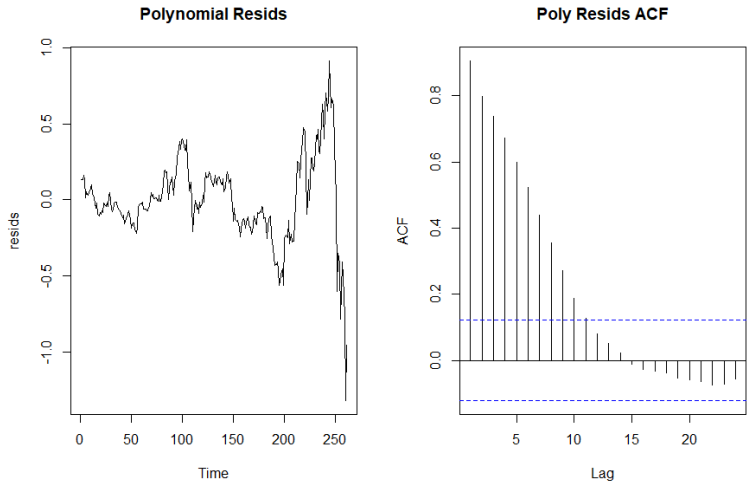
```
Family: gaussian
Link function: identity

Formula:
train$FB ~ s(time.pts)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.078687   0.007937   135.9   <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(time.pts) 8.757  8.983 864.5  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.968   Deviance explained = 96.9%
GCV = 0.017082  Scale est. = 0.016444  n = 261
```

Mean Absolute Error: .0875

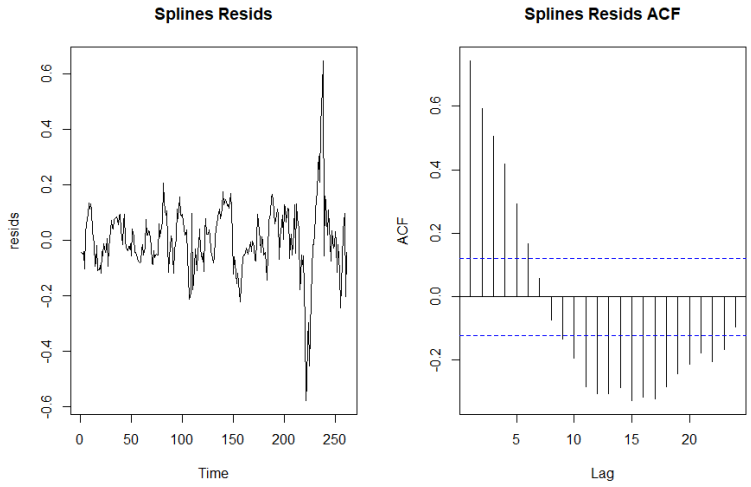
Mean Squared Error: .0158

The models map similar units (change in decimalized percentages) for different stocks. Because units are similar, this means it is possible to compare the two in the sense that we see the polynomial model has more average error than the splines model. Because they are fit on different data; however, no conclusions can be made about the relative quality of either fitting method.

C.



From the above plots, it is clear the residual process for the Quadratic fit on AMZN is non-stationary with a clearly present non-zero trend.

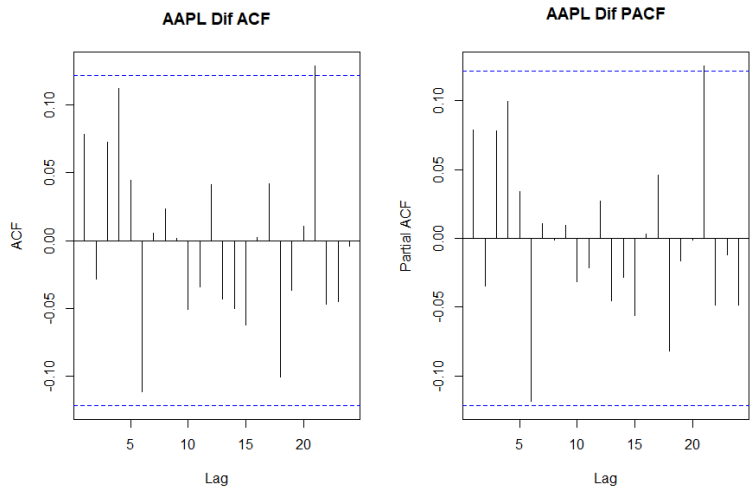


Though mean appears constant and variance is mostly constant, just with a couple spikes, ACF plots show clear patterns of significance and so this process is non-stationary.

ISYE 6402 – Spring 2020
Georgia Institute of Technology

Question 2 – 20 Points

A.



The above charts do not have clear patterns of cut-off/trail-off and so no conclusions can be drawn about ARMA orders.

B.

AAPL:

```
> tail(orders)
    p d q       AIC
69 4 1 4 -733.1826
68 4 1 3 -734.0249
55 3 1 4 -734.0685
41 2 1 4 -735.2308
67 4 1 2 -735.4337
54 3 1 3 -735.6907
```

GOOG:

```
> tail(orders)
    p d q       AIC
52 3 1 0 -872.9408
54 3 1 2 -873.6940
40 2 1 3 -873.7427
9  0 1 2 -873.8059
37 2 1 0 -874.8510
39 2 1 2 -875.5016
```

Using an AIC threshold of 2, the orders selected are (3,1,3) and (2,1,0) respectively.

C.

AAPL Roots

AR: .61, 1.77, 1.77

MA: 1.00, 1.00, 2.53

One AR root within unit circle, two MA roots on unit circle. Process is stationary but neither causal nor invertible.

ISYE 6402 – Spring 2020
Georgia Institute of Technology

GOOG Roots:

AR: 2.22, 2.64

MA: N/A

All (AR) roots outside unit circle; process is causal, invertible, and stationary.

D.

AAPL Forecast Accuracy

Mean Absolute Percentage Error: .117

Precision: 1.79

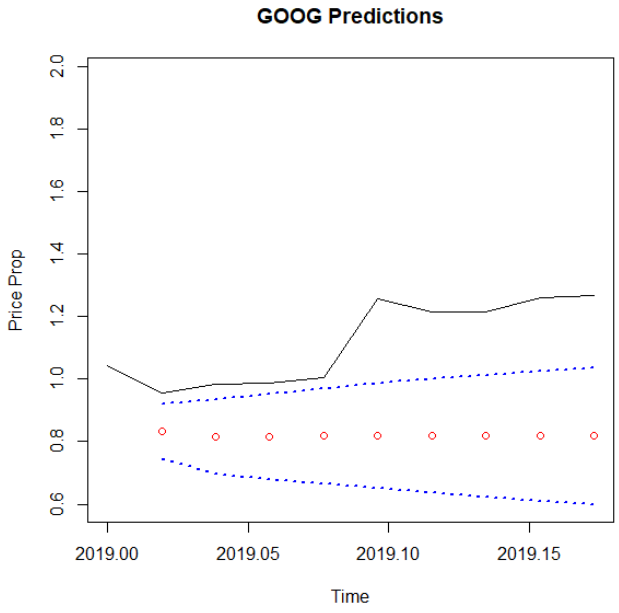
GOOG Forecast Accuracy

Mean Absolute Percentage Error: .126

Precision: 7.12

(Above MAPE's are decimalized, percent format acceptable as well)

E.



Forecasts are quite poor fits, real values not even contained within confidence intervals.

ISYE 6402 – Spring 2020
Georgia Institute of Technology

3 – 18 Points

A.

```
> vs$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     2      1      1      2
```

Order 2 by AIC, 1 by HQ

```
        ARCH (multivariate)

data:  Residuals of VAR object mod
Chi-squared = 1027.2, df = 500, p-value < 2.2e-16
```

```
        JB-Test (multivariate)

data:  Residuals of VAR object mod
Chi-squared = 4561.8, df = 8, p-value < 2.2e-16
```

```
        Portmanteau Test (asymptotic)

data:  Residuals of VAR object mod
Chi-squared = 306.88, df = 224, p-value = 0.0001952
```

The above tests (all significant) indicate significant evidence for heteroskedasticity, non-normality, and correlation of errors.

B.

AAPL Accuracy Measures

MAPE: 1.72

Precision: 1.22

GOOG Accuracy Measures

MAPE: 2.98

Precision: .981

C.

While differing frames of differencing makes exact value comparisons of MAPE not appropriate methodology, the VAR(2) model has MAPE values an order of magnitude greater. This would allow us to conclude the VAR(2) model is less accurate on the differenced data than the ARIMA models are.

Precision metrics indicate that forecasts for the VAR model have relatively lower variance than the actual data, particularly for GOOG data.

Results indicate the data is likely not reliably cross-correlated.

Question 4 – 20 Points

A.

Stepwise results using provided A-G Order Selection Code:

Initial GARCH Order: 1,1 BIC = -3.38

ARMA Update: 3,4 BIC = -3.40

Final GARCH: 1,1 BIC = -3.40

Final A-G Order: (3,4)x(1,1)

B.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
mu	5.842e-03	5.765e-03	1.013	0.31083	
ar1	8.769e-01	3.741e-02	23.442	< 2e-16	***
ar2	-8.155e-01	5.839e-02	-13.966	< 2e-16	***
ar3	9.329e-01	4.937e-02	18.895	< 2e-16	***
ma1	4.982e-02	7.950e-02	0.627	0.53089	
ma2	8.581e-01	8.424e-02	10.186	< 2e-16	***
ma3	-3.157e-02	7.573e-02	-0.417	0.67681	
ma4	1.563e-02	7.090e-02	0.221	0.82548	
omega	2.582e-04	1.965e-05	13.139	< 2e-16	***
alpha1	2.999e-01	1.060e-01	2.830	0.00466	**
beta1	6.071e-01	7.516e-02	8.078	6.66e-16	***

MA coefficients 1, 3, 4 not significant at any standard level, rest are at a level of .01 or better.

Equation:

ARIMA:

$Y_t = 5.84e\text{-}3 + .88*Y_{t-1} - .82*Y_{t-2} + .93*Y_{t-3} + .050*E_{t-1} + .86*E_{t-2} - .032*E_{t-3} + .016*E_{t-4} + E_t$

GARCH:

$\sigma^2_t = 2.58e\text{-}4 + .30*E^2t_{-1} + .61*\ \sigma^2_{t-1}$

C.

Standardised Residuals Tests:

			Statistic	p-Value
Jarque-Bera Test	R	Chi^2	21.82058	1.82693e-05
Shapiro-Wilk Test	R	W	0.9825854	0.002804552
Ljung-Box Test	R	Q(10)	1.650976	0.9983831
Ljung-Box Test	R	Q(15)	4.865108	0.9932082
Ljung-Box Test	R	Q(20)	9.065561	0.9821354
Ljung-Box Test	R^2	Q(10)	3.270009	0.974337
Ljung-Box Test	R^2	Q(15)	12.47468	0.6428024
Ljung-Box Test	R^2	Q(20)	14.07963	0.8264358
LM Arch Test	R	TR^2	4.902221	0.9611653

Arch Test as well as residual and squared residual L-B Tests all have large p values. This indicates no significant evidence for non-constant variance of error and dependence in residuals and squared residuals (weak independence). J-B Test has a small p, indicating significant evidence for non-normality of error.

D.

ARIMA-GARCH Google Accuracy Measures

MAPE: .057

Precision: .308

The MAPE is roughly half of that of the ARIMA model, and the precision is substantially smaller. Because the ARIMA model is asked to forecast nine points ahead (the variance of each subsequent forecast increasing) and the rolling forecast is an aggregation of single forecasts ahead, the comparison is not fair. Additionally, the is substantially more measurements taking for the A-G model accuracy, so variance is much less. Rolling forecasts are likely reliable due to the larger sample size, but the ARIMA forecasts are likely not very accurate and highly variable.