**MScFE 610 Econometrics**

**September 22, 2019**

**Group project: Group 1C**

**Regression and Univariate Analysis**

**Submission 1**

* Nadeem Anwaar: [Nadeem\_anwaar@yahoo.com](mailto:Nadeem_anwaar@yahoo.com)
* Cemre Pehlivanoglu: [cemrepeh@gmail.com](mailto:cemrepeh@gmail.com)
* Okechukwu Okigbo: [okigbookey@gmail.com](mailto:okigbookey@gmail.com)
* Tamirat Terefe: [miranet34@yahoo.com](mailto:miranet34@yahoo.com)
* Paras Jandwani: [paras.jandwani@gmail.com](mailto:paras.jandwani@gmail.com)

­­­­­­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3.1.1 Basic Statistics

Download JP Morgan stock historical prices from Yahoo Finance

Period: February 1, 2018 – December 30, 2018

Frequency: Daily

Price considered in the analysis: Close price adjusted for dividends and splits

3.1.1.1 Calculate in R: [see R script for codes]

1.1. Average Stock value

[Ans] 107.2015

1.2. Stock volatility

[Ans] 4.56665

1.3. Daily stock return

Where = Stock price at time t

= Stock prince at time t – 1 (i.e. a day prior)

Below is head sample:

JPM.simpleReturns

2018-02-02 -0.022161369

2018-02-05 -0.047952357

2018-02-06 0.030422834

2018-02-07 0.006779018

2018-02-08 -0.044210207

2018-02-09 0.020022238

3.1.1.2 Calculate in Excel:

1.4. Average stock value [Ans: see Excel file]

1.5. Stock volatility [Ans: see Excel file]

1.6. Daily stock return [Ans: see Excel file]

1.7. Show JP Morgan stock price evolution using scatter plot

1.8. Add a trendline to the graph (trendline options – linear)

3.1.2 Linear Regression

3.1.2.1 Implement a two-variable regression in R

Explained variable: JP Morgan stock (adjusted close price)

Explanatory variable: S&P500

Period: February 1, 2018 – December 30, 2018

Frequency: Daily

R code:

**linear\_model <- lm(JPM.Adjusted ~ GSPC.Adjusted, data=allData)**

**summary(linear\_model)**

Result:

**Call:**

**lm(formula = JPM.Adjusted ~ GSPC.Adjusted, data = allData)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-6.7551 -2.3973 0.4835 2.3838 5.6483**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 13.751225 5.294731 2.597 0.01 \***

**GSPC.Adjusted 0.034065 0.001929 17.662 <2e-16 \*\*\***

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 2.97 on 227 degrees of freedom**

**Multiple R-squared: 0.5788, Adjusted R-squared: 0.5769**

**F-statistic: 311.9 on 1 and 227 DF, p-value: < 2.2e-16**

Interpretation of linear regression result:

The coefficient of S&P500 [0.034] is statistically significant in the linear regression. For the linear regression model Adjusted R-squared of 0.578 and F-statistic of 311 with p-value < 0.0001. The linear regression model is not the appropriate model for this analysis as the JP Morgan stock prices are not independent but serially autocorrelated. Also, we need more explanatory variables as S&P500 index prices alone will not account for all the variations in the JP Morgan stock prices. Thus, we need a better model with more explanatory variable. An ARMA model will be better.

3.1.2.2. Implement a two-variable regression in Excel using LINEST

function and Analysis ToolPak:

[Ans: see Excel file for entire output]



3.1.3 Univariate Time Series Analysis

1. Forecast S&P/Case-Shiller U.S. National Home Price index suing ARMA model.

Data source: <https://fred.stlouisfed.org/series/CSUSHPINSA>

Period considered in the analysis: January 1978 – latest data

Frequency: monthly data

* 1. Implement the Augmented Dickey-Fuller Test for checking the existence of a unit root in Case-Shiller Index series

**adf.test(CSUSHPINSA)**

**Augmented Dickey-Fuller Test**

**data: CSUSHPINSA**

**Dickey-Fuller = -2.3243, Lag order = 7, p-value = 0.4402**

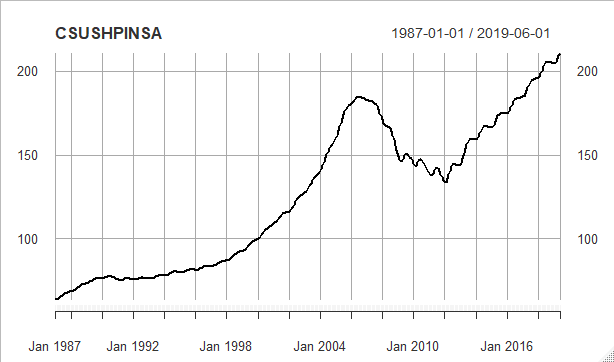
**alternative hypothesis: stationary**

The p-value for the ADF test is 0.4402, which suggests that we cannot reject the null hypothesis that a unit root is present in the time series data even at the 90% confidence level.

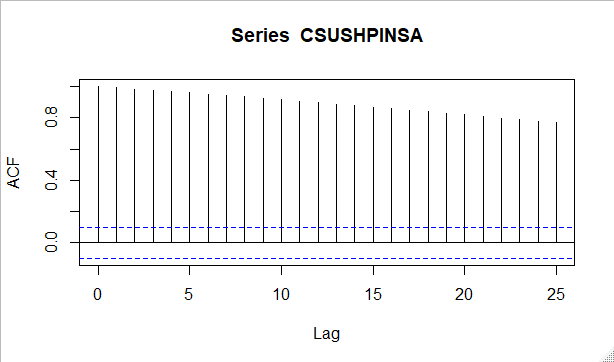
* 1. Implement an ARIMA(p,d,q) model. Determine p, d, q using Information Criterion or Box-Jenkins methodology. Comment results.

We first plot the series as well as the acf and pacf outputs.

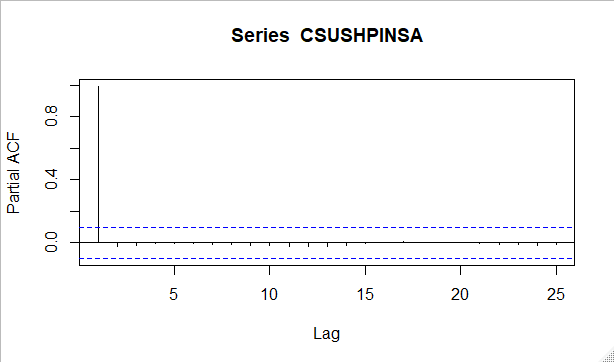
**plot(CSUSHPINSA)**



**acf(CSUSHPINSA)**

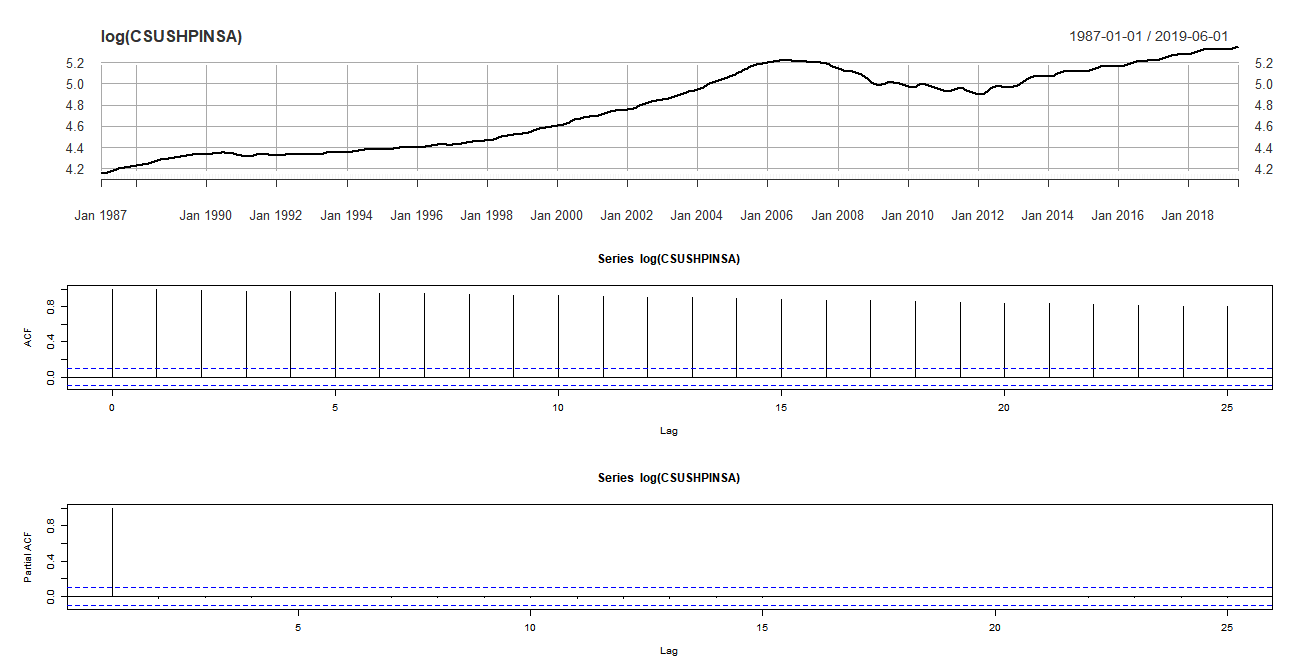


**pacf(CSUSHPINSA)**

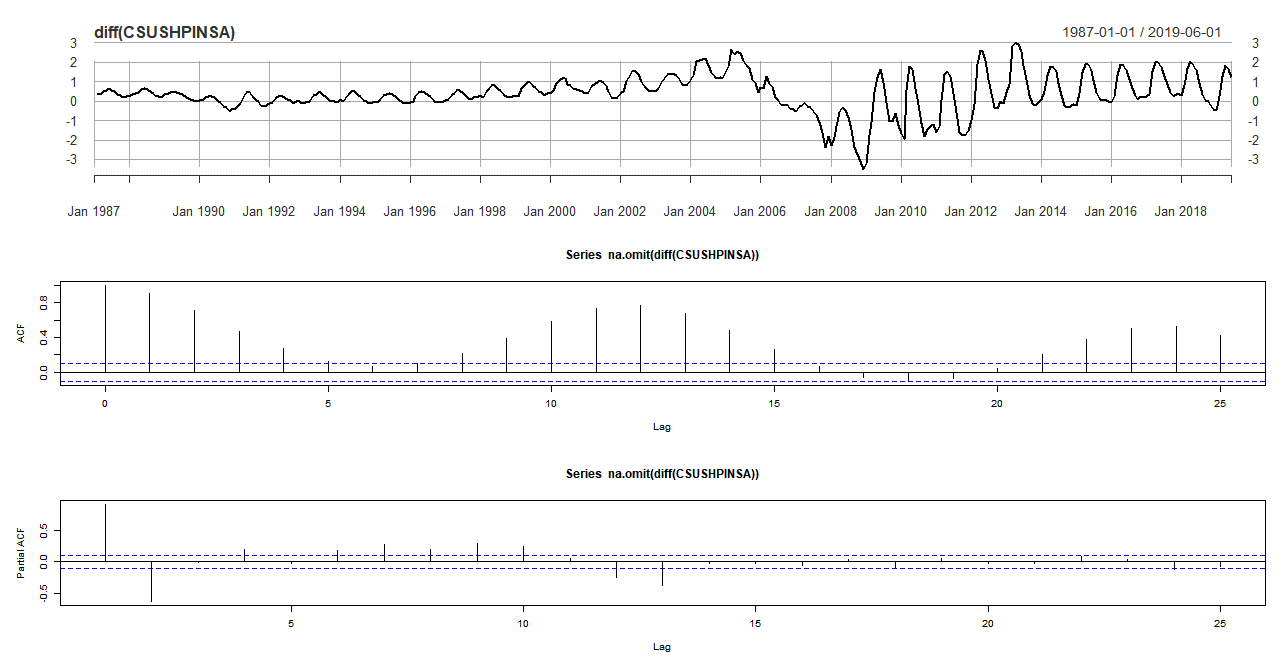


The CSUSHPINSA series plot shows trend and growth over time. Also, acf and pacf exhibited sustained and persistent autocorrelation over multiple lags. Both supporting Augmented Dickey-Fuller Test that the CSUSHPINSA series is not stationary. Thus, differencing or log transformation or both might improve the series stationarity. We plot the CSUSHPINSA series plot for different types of transformations.

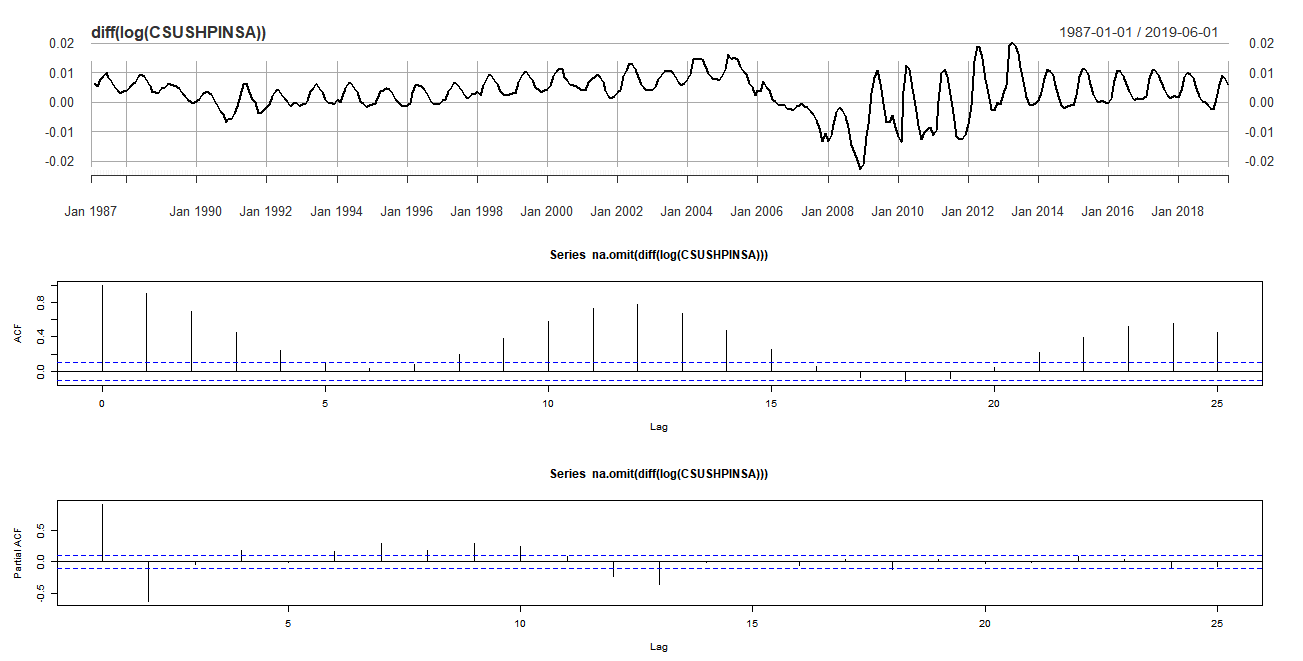
For log transformation:



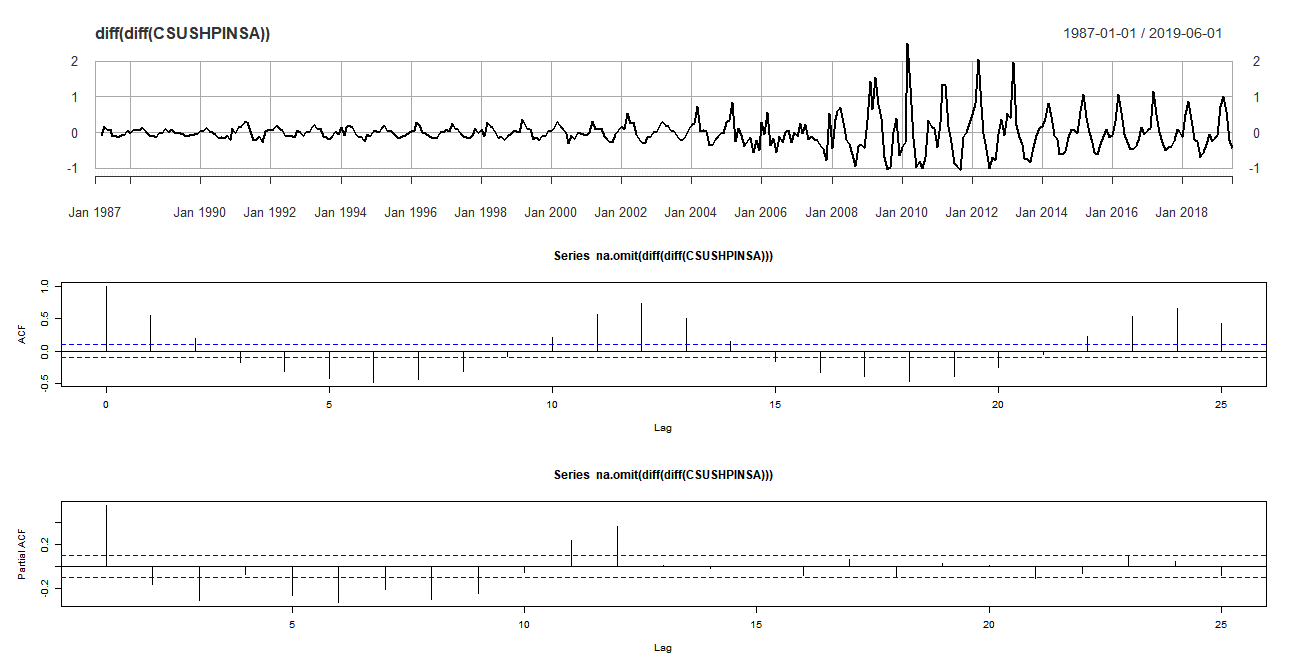
For first differencing transformation:



For differencing log transformation:



For second differencing transformation:



We choose to combine Information Criterion and Box-Jenkins methodology to determine p,d and q. Hence, we implement the ARIMA model with the auto function of forecast package in R.

The auto.arima() function in forecast package used sequential iterative process to search for combinations of arima model parameter that results in lowest AIC. The process final returns the arima model with the lowest AIC. This process is fast and, in most cases, reliable but usually used as a guide to arima model selection.

**auto.arima(CSUSHPINSA)**

**Series: CSUSHPINSA**

**ARIMA(3,1,2) with drift**

**Coefficients:**

**ar1 ar2 ar3 ma1 ma2 drift**

**0.8592 0.1036 -0.2281 0.6294 0.2962 0.3785**

**s.e. 0.1446 0.2154 0.1135 0.1412 0.0789 0.1134**

**sigma^2 estimated as 0.09747: log likelihood=-97.53**

**AIC=209.07 AICc=209.36 BIC=236.81**

##########################################

Series: CSUSHPINSA

ARIMA(3,1,2) with drift

Coefficients:

ar1 ar2 ar3 ma1 ma2 drift

0.8592 0.1036 -0.2281 0.6294 0.2962 0.3785

s.e. 0.1446 0.2154 0.1135 0.1412 0.0789 0.1134

sigma^2 estimated as 0.09747: log likelihood=-97.53

AIC=209.07 AICc=209.36 BIC=236.81

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 1.225712 0.055041 22.2690 < 2.2e-16 \*\*\*

ar2 -0.360678 0.080632 -4.4731 7.709e-06 \*\*\*

ar3 -0.225372 0.052390 -4.3018 1.694e-05 \*\*\*

ma1 -0.835304 0.029221 -28.5854 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:

arima(x = CSUSHPINSA, order = c(3, 2, 1))

Coefficients:

ar1 ar2 ar3 ma1

1.2257 -0.3607 -0.2254 -0.8353

s.e. 0.0550 0.0806 0.0524 0.0292

sigma^2 estimated as 0.09024: log likelihood = -84.75, aic = 179.49

Training set error measures:

ME RMSE MAE MPE MAPE MASE

Training set 0.0005232603 0.2996659 0.1719374 0.003480077 0.1267506 0.2290857

ACF1

Training set -0.005799614

[1] 2.526616

[1] 2.179081

Model RMSE: 2.18

Model RMSE: 3.53