Group Work Project - Econometrics - Group 6 - D

Members:

- Leila Tugarayeva
- Shiqi Zhang
- Mateusz Dubaniowski
- Long Son Nguyen
- Matthew Ang

Contents

1. Basic Statistics

This section was contributed by Leila Tugarayeva.

In this section we performed basic analysis of daily JP Morgan historical stock prices. Period under review is February 1, 2018 – December 30, 2018.

All calculations are shown on the pictures below.

Calculations in R:

1.1. Average stock value is equal to 106.3496 USD (picture 1).

1.2. Daily stock volatility – 0.0144 (picture 1).

Annualized stock volatility – 0.2289 (picture 1).

1.3. Daily stock returns (first several values are shown on picture 2).

Calculations in Excel

1.4. Average stock value – 106.3496 USD (picture 3).

1.5. Daily stock volatility – 0.0144 (picture 4).

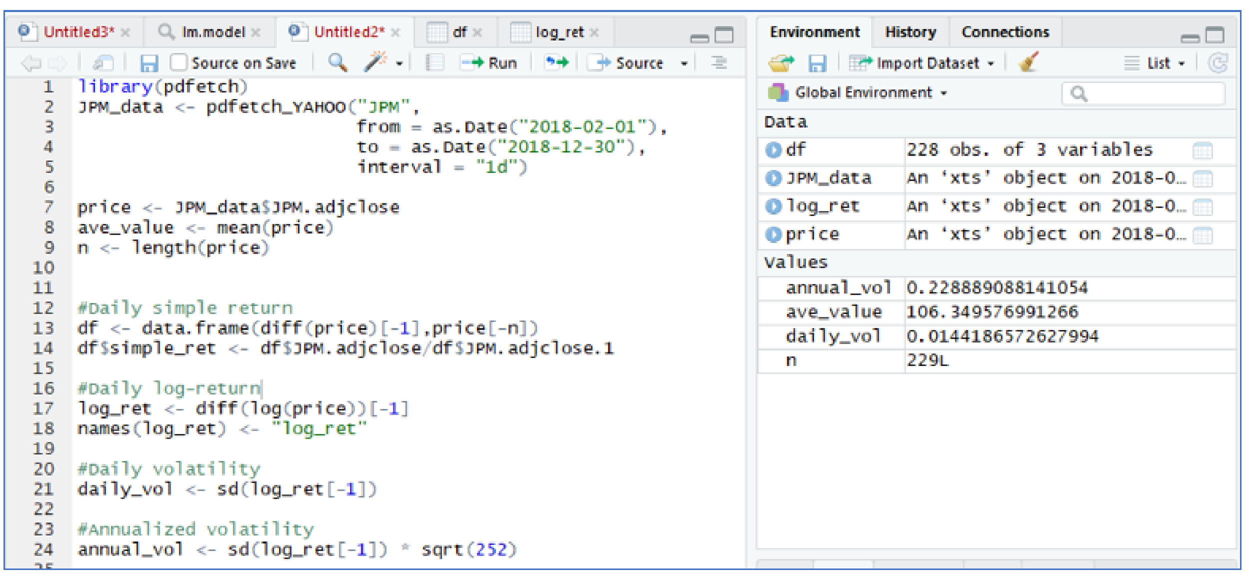
Annualized stock volatility – 0.2289 (picture 5).

1.6. Daily stock returns (first several values are shown on picture 6).

1.7. JP Morgan stock price evolution using a scatter plot (picture 7).

1.8. Trendline on the graph in p. 1.7 (picture 7).

We obtained same results using R and Excel for the average stock value, stock volatility and daily returns. Scatterplot and the trendline exhibit general downwards slope of the closing adjusted prices.



Picture 1 – R code and results of calculations for subsections 1.1, 1.2 and 1.3

	JPM.adjclose	JPM.adjclose.1	simple_ret		log_ret
2018-02-02	-2.468124	111.37041	-2.216140e-02	2018-02-02	-2.241065e-02
2018-02-05	-5.222122	108.90228	-4.795236e-02	2018-02-05	-4.914021e-02
2018-02-06	3.154243	103.68016	3.042282e-02	2018-02-06	2.996922e-02
2018-02-07	0.724236	106.83440	6.779052e-03	2018-02-07	6.756178e-03
2018-02-08	-4.755188	107.55864	-4.421019e-02	2018-02-08	-4.521725e-02
2018-02-09	2.058357	102.80345	2.002226e-02	2018-02-09	1.982445e-02
2018-02-12	1.619995	104.86181	1.544886e-02	2018-02-12	1.533074e-02
2018-02-13	0.657539	106.48180	6.175130e-03	2018-02-13	6.156142e-03
2018-02-14	2.477654	107.13934	2.312553e-02	2018-02-14	2.286219e-02
2018-02-15	0.457405	109.61700	4.172756e-03	2018-02-15	4.164074e-03
2018-02-16	-0.790932	110.07440	-7.185431e-03	2018-02-16	-7.211370e-03
2018-02-20	0.028580	109.28347	2.615217e-04	2018-02-20	2.614875e-04
2018-02-21	0.457413	109.31205	4.184470e-03	2018-02-21	4.175740e-03
2018-02-22	-0.200112	109.76946	-1.823021e-03	2018-02-22	-1.824684e-03
2018-02-23	2.220352	109.56935	2.026435e-02	2018-02-23	2.006176e-02

Picture 2 – First several results of subsection 1.3 – simple daily return and log-return

=AVERAGE(B3:B231)

	A	B	C	D	E	F	G
1		Data	Calculated values				
2	Date	Adj Close	Squared deviation	Daily return	Daily log-return	Average stock value	106.349577
3	01-02-18	111.370407	25.20873398			Daily volatility	0.014418657
4	02-02-18	108.902283	6.516307967	-0.741505941	-0.022410652	Annualized volatility	0.228889088
5	05-02-18	103.680161	7.125781734	0.093530535	-0.049140208		

Picture 3 – Average stock value calculations in Excel

G3 =STDEV.S(E4:E231)

	A	B	C	D	E	F	G
1		Data	Calculated values				
2	Date	Adj Close	Squared deviation	Daily return	Daily log-return	Average stock value	106.349577
3	01-02-18	111.370407	25.20873398			Daily volatility	0.014418657
4	02-02-18	108.902283	6.516307967	-0.741505941	-0.022410652	Annualized volatility	0.228889088
5	05-02-18	103.680161	7.125781734	0.093530535	-0.049140208		

Picture 4 – Daily volatility calculations in Excel

G4 =SQRT(252)*G3

	A	B	C	D	E	F	G
1		Data	Calculated values				
2	Date	Adj Close	Squared deviation	Daily return	Daily log-return	Average stock value	106.349577
3	01-02-18	111.370407	25.20873398			Daily volatility	0.014418657
4	02-02-18	108.902283	6.516307967	-0.741505941	-0.022410652	Annualized volatility	0.228889088

Picture 5 – Annualized volatility calculations in Excel

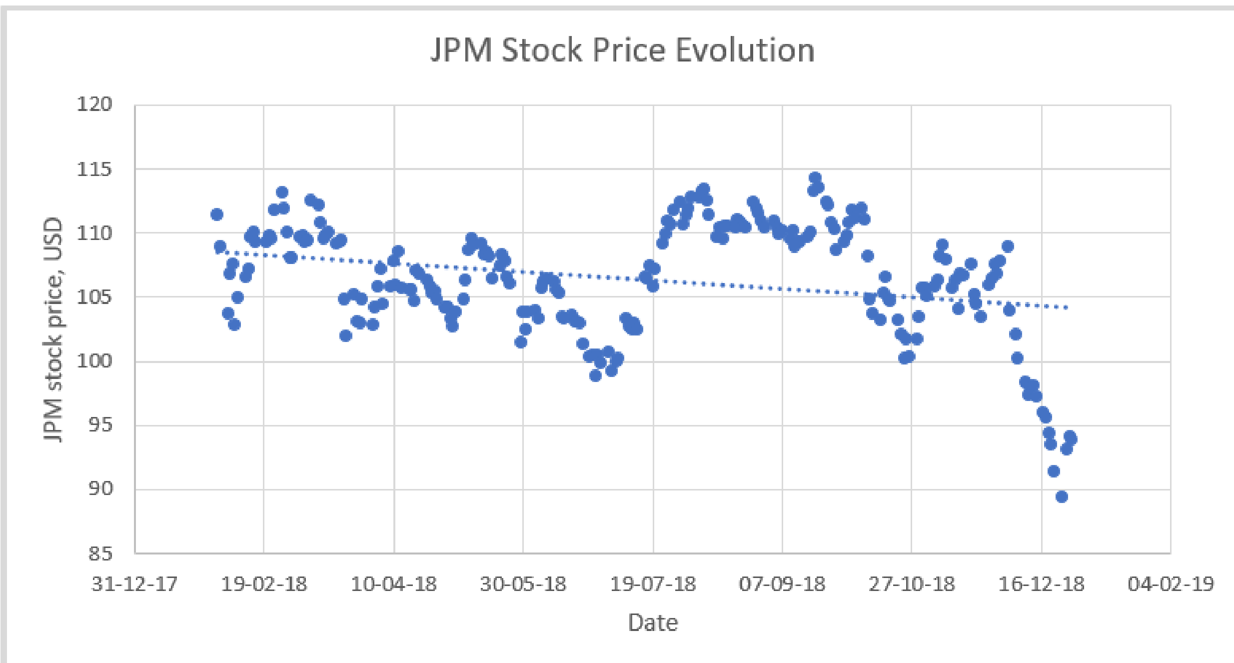
C4 | =(B4-B3)/B3

	A	B	C	D
1	Data			
2	**Date**	**Adj Close**	**Daily return**	**Daily log-return**
3	01-02-18	111.370407		
4	02-02-18	108.902283	-0.022161399	-0.022410652
5	05-02-18	103.680161	-0.047952365	-0.049140208
6	06-02-18	106.834404	0.030422821	0.029969224
7	07-02-18	107.55864	0.006779052	0.006756178
8	08-02-18	102.803452	-0.044210191	-0.045217255
9	09-02-18	104.861809	0.020022256	0.019824446
10	12-02-18	106.481804	0.015448856	0.015330738
11	13-02-18	107.139343	0.00617513	0.006156142
12	14-02-18	109.616997	0.023125529	0.022862186
13	15-02-18	110.074402	0.004172756	0.004164074
14	16-02-18	109.28347	-0.007185431	-0.00721137
15	20-02-18	109.31205	0.000261522	0.000261488

D4 | =LN(B4/B3)

	A	B	C	D
1	Data			
2	**Date**	**Adj Close**	**Daily return**	**Daily log-return**
3	01-02-18	111.370407		
4	02-02-18	108.902283	-0.022161399	-0.022410652
5	05-02-18	103.680161	-0.047952365	-0.049140208
6	06-02-18	106.834404	0.030422821	0.029969224
7	07-02-18	107.55864	0.006779052	0.006756178
8	08-02-18	102.803452	-0.044210191	-0.045217255
9	09-02-18	104.861809	0.020022256	0.019824446
10	12-02-18	106.481804	0.015448856	0.015330738
11	13-02-18	107.139343	0.00617513	0.006156142
12	14-02-18	109.616997	0.023125529	0.022862186
13	15-02-18	110.074402	0.004172756	0.004164074
14	16-02-18	109.28347	-0.007185431	-0.00721137
15	20-02-18	109.31205	0.000261522	0.000261488

Picture 6 – First several results of subsection 1.6 – simple daily return (left) and log-return (right)



Picture 7 – JPM stock price evolution and its trend line.

2. Linear Regression

This section was contributed by Leila Tugarayeva.

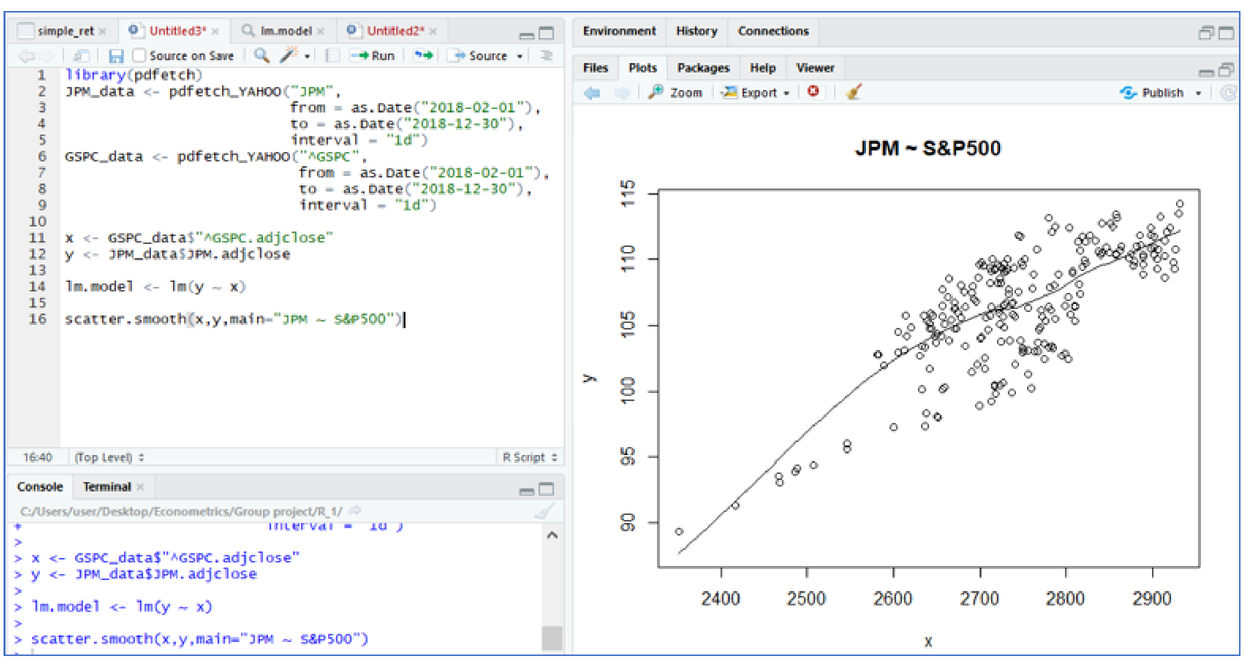
In this section, we implement a two-variable regression in R and Excel.

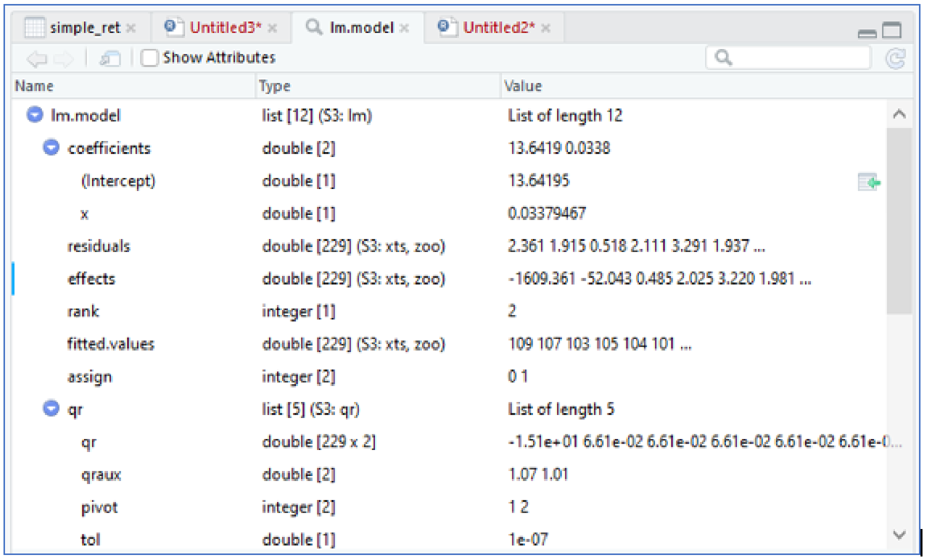
Explained variable: Daily JP Morgan adjusted close prices.
Explanatory variable: Daily S&P500 adjusted close prices
Period under review: February 1, 2018 – December 30, 2018

Calculations performed both in R (picture 8) and Excel (picture 9) yielded the following results:

Intercept	13.6419
SE for intercept	5.2527
Slope	0.0338
SE for slope	0.0019
Multiple R	0.760790769
R Square	0.578802594
F-statistic	311.9397
P-value of F-statistic	1.6594E-44

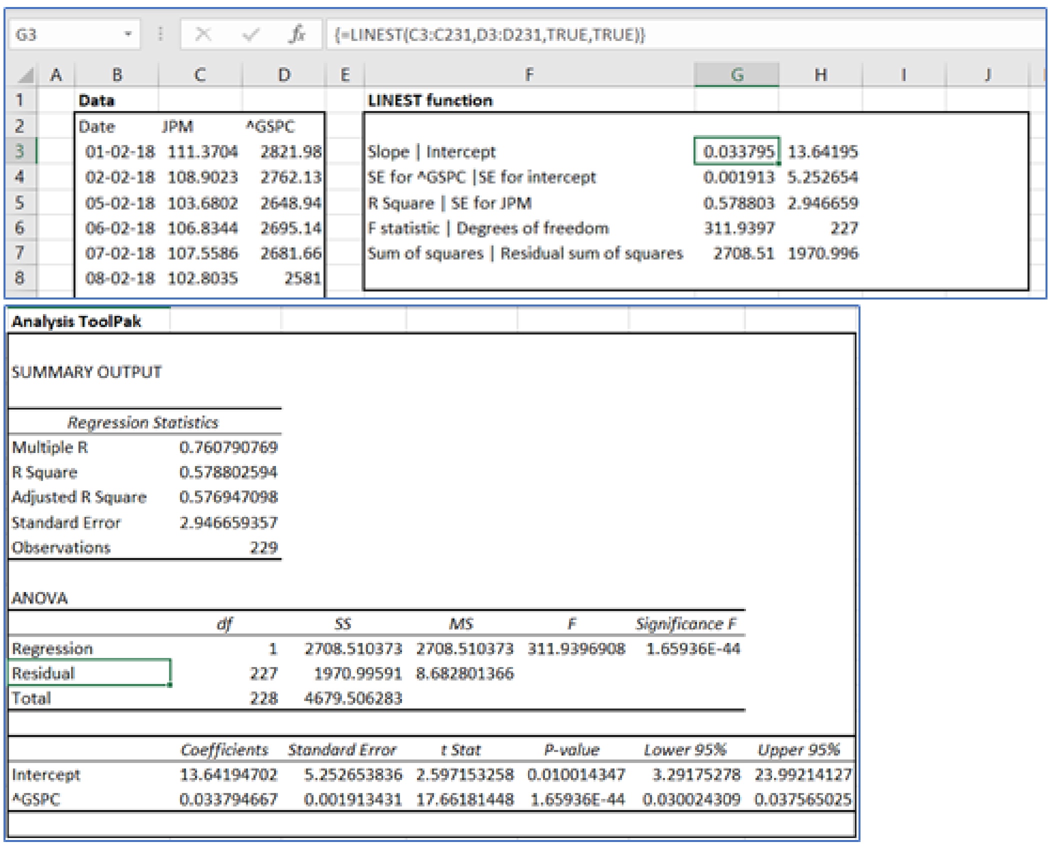
Correlation coefficient of 0.76 indicates that there is relatively strong positive relationship between the two variables. The goodness-of-fit of 0.5788 indicates that around 58% of the dependent variable variance is explained by the model.





Name	Type	Value
lm.model	list [12] (S3: lm)	List of length 12
coefficients	double [2]	13.6419 0.0338
(Intercept)	double [1]	13.64195
x	double [1]	0.03379467
residuals	double [229] (S3: xts, zoo)	2.361 1.915 0.518 2.111 3.291 1.937 ...
effects	double [229] (S3: xts, zoo)	-1609.361 -52.043 0.485 2.025 3.220 1.981 ...
rank	integer [1]	2
fitted.values	double [229] (S3: xts, zoo)	109 107 103 105 104 101 ...
assign	integer [2]	0 1
qr	list [5] (S3: qr)	List of length 5
qr	double [229 x 2]	-1.51e+01 6.61e-02 6.61e-02 6.61e-02 6.61e-02 6.61e-0...
qraux	double [2]	1.07 1.01
pivot	integer [2]	1 2
tol	double [1]	1e-07

Picture 8 – Implementation of a Linear Regression model in R (top) and the resulting model (bottom).



G3 {=LINEST(C3:C231,D3:D231,TRUE,TRUE)}

Data

Date	JPM	^GSPC
01-02-18	111.3704	2821.98
02-02-18	108.9023	2762.13
05-02-18	103.6802	2648.94
06-02-18	106.8344	2695.14
07-02-18	107.5586	2681.66
08-02-18	102.8035	2581

LINEST function

Slope | Intercept	0.033795	13.64195
SE for ^GSPC |SE for intercept	0.001913	5.252654
R Square | SE for JPM	0.578803	2.946659
F statistic | Degrees of freedom	311.9397	227
Sum of squares | Residual sum of squares	2708.51	1970.996

Analysis ToolPak

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.760790769
R Square	0.578802594
Adjusted R Square	0.576947098
Standard Error	2.946659357
Observations	229

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2708.510373	2708.510373	311.9396908	1.65936E-44
Residual	227	1970.99591	8.682801366		
Total	228	4679.506283			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	13.64194702	5.252653836	2.597153258	0.010014347	3.29175278	23.99214127
^GSPC	0.033794667	0.001913431	17.66181448	1.65936E-44	0.030024309	0.037565025

***Picture 9** – Implementation of a Linear Regression model in Excel using (a) LINEST function (top) and (b) the Analysis Tool Pak (bottom).*

3. Univariate Time Series Analysis

In this section, we develop an ARIMA model for the S&P/Case-Shiller U.S. National Home Price Index.

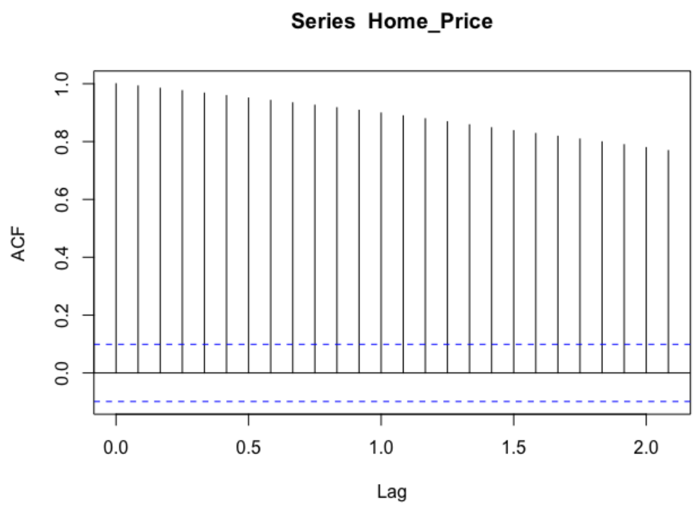
3.1 Augmented Dickey-Fuller (ADF) Test

This sub-section was contributed by Shiqi Zhang.

Below are the steps taken to use Augmented Dickey-Fuller Test for checking the existence of a unit root in Case-Shiller Index series, using the R language.

```
# import packages
library(tidyverse)
library(stats)
library(readxl)
library(tseries)
library(forecast)
library(ggplot2)

# import Home Price data from downloaded xls file
CSUSHPINSA <- read_excel("Desktop/CSUSHPINSA.xls",
                         sheet = "FRED Graph", col_types = c("date",
                                                             "numeric"),
skip = 10)

# view time plots
ggplot(data = CSUSHPINSA, mapping = aes(x = observation_date, y =
CSUSHPINSA)) +
  geom_line(color = "darkblue") +
  labs(x = "month", y = "Home Price")

# convert to Time Series data
Home_Price <- ts(CSUSHPINSA$CSUSHPINSA, start = 1987, frequency = 12)

# ADF & ACF tests
adf.test(Home_Price)
acf(Home_Price)
```

```
> adf.test(Home_Price)

      Augmented Dickey-Fuller Test

data:  Home_Price
Dickey-Fuller = -2.2758, Lag order = 7, p-value = 0.4607
alternative hypothesis: stationary
```



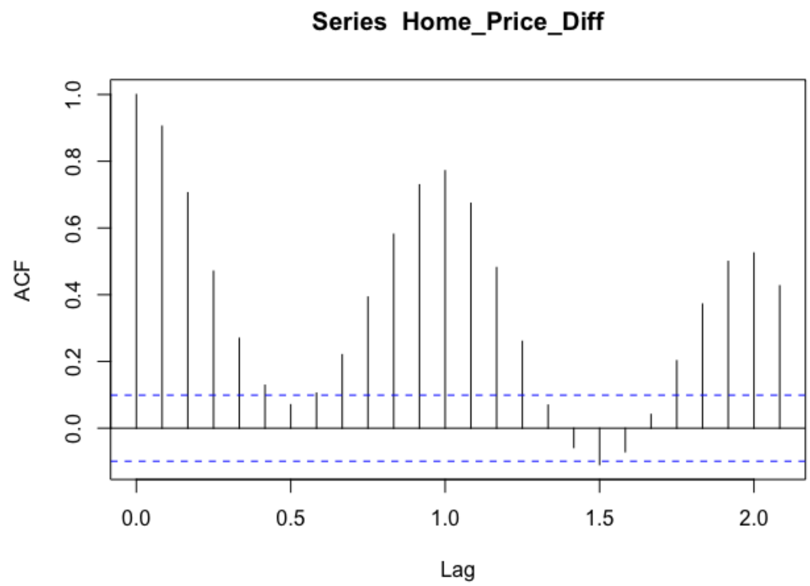
***Picture 10** – ACF plot of Home-Shiller Index*

The null hypothesis of ADF test is that unit root exists, and the time series is non-stationary. The ADF test on the original Price index time series shows a result of -2.2758 with p-value = 0.4607 which is >0.05 significant level. This suggest that we cannot reject the null hypothesis, hence there exists a unit root in Case-Shiller Index series.

Furthermore, by running ACF test we confirmed the existence of unit root and non-stationarity because autocorrelations are very close to 1 and fading very slowly across many lags.

```
# get the first difference of the Home Price Index
Home_Price_Diff <- diff(Home_Price, lag = 1, differences = 1)

# ADF & ACF tests on 1st diff Home_Price
acf(Home_Price_Diff)
adf.test(Home_Price_Diff)
```

```
> adf.test(Home_Price_Diff)

      Augmented Dickey-Fuller Test

data:  Home_Price_Diff
Dickey-Fuller = -2.8115, Lag order = 7, p-value = 0.2345
alternative hypothesis: stationary
```

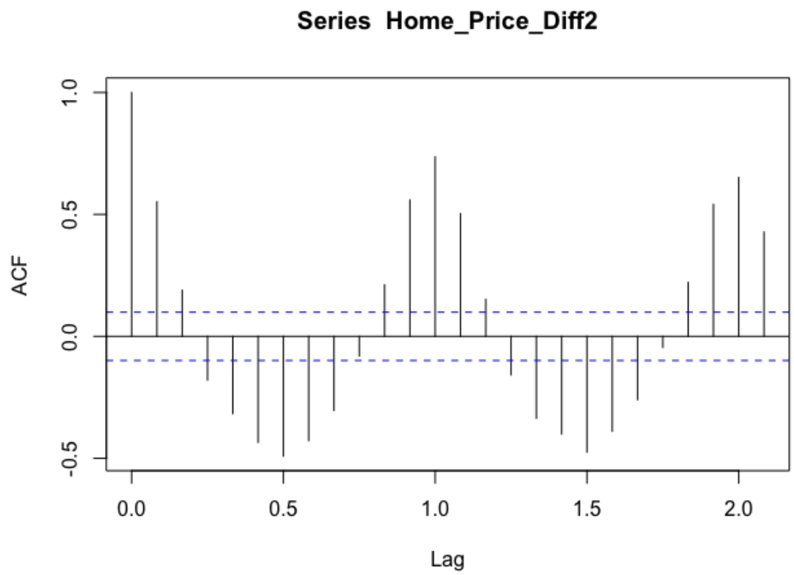


***Picture 11** – ACF plot of differenced Home-Shiller Index*

We then proceed to determine the degree of differencing required to transform the original time series to achieve stationarity. The ADF test on the first difference of Price Index shows a result of -2.8115 and p-value = 0.2345, which is still >0.05 significant level. Therefore, we are still unable to reject the null hypothesis and unit root and non-stationarity remain. The ACF test on first difference time series confirms this result. A shock in some months will still have an impact on the current level more than two months into the future.

```
# get the second difference of the Home Price Index
Home_Price_Diff2 <- diff(Home_Price, lag = 1, differences = 2)

# ADF & ACF tests on 2nd diff Home_Price
acf(Home_Price_Diff2)
adf.test(Home_Price_Diff2)
```

```
> adf.test(Home_Price_Diff2)

      Augmented Dickey-Fuller Test

data:  Home_Price_Diff2
Dickey-Fuller = -16.102, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(Home_Price_Diff2) : p-value smaller than printed p-value
```



***Picture 12** – ACF plot of twice-differenced Home-Shiller Index*

The ADF test on the second difference Price Index yield positive results. DF = -16.102 and p-value < 0.01, which is less than the significant level of 0.05. Hence, we reject the null hypothesis and the transformation was successful in achieving stationarity. ACF results also confirms this.

We will continue to implement an ARIMA model on the Home Price Index in the next section.

3.2 Auto Regressive Integrated Moving Average (ARIMA) model

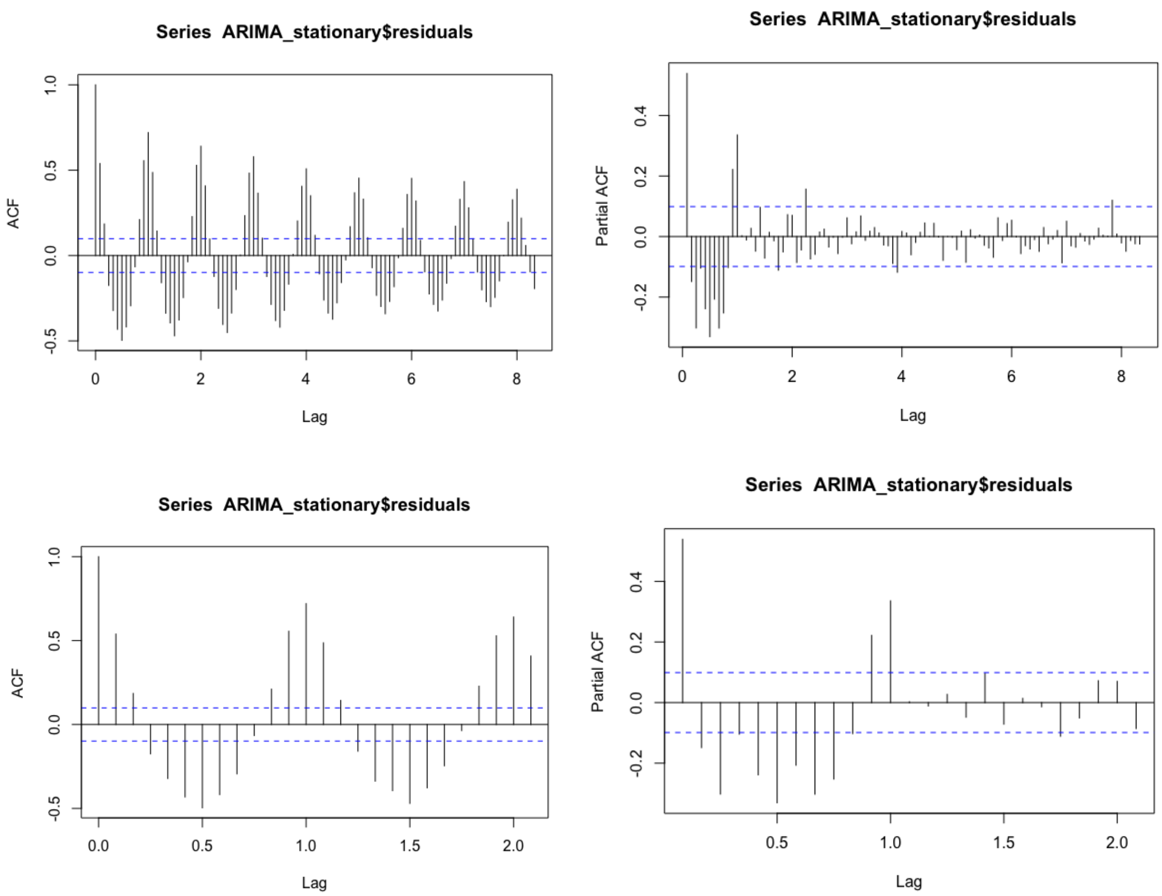
This sub-section was contributed by Mateusz Iwo Dubaniowski.

The task was completed in R.
Extensively commented R code can be found below. Respective graphs are placed below the code with short explanations. The graphs are also referred to in the code. Running the code will produce the graphs.

```
# WQU: Group Work - Group 6-D - Econometrics
# Mateusz Iwo Dubaniowski (iwo@utexas.edu)

#Libraries
library(tidyverse)
library(stats)
library(readxl)
library(forecast)
library(tseries)

#Import data
Data <- read_excel("CSUSHPINSA.xls", col_types = c("date",
                                                  "numeric"), skip = 10)
#Transform the data by taking log and specifying the data as time series
Data$log<-log(Data$CSUSHPINSA)
Data$log<-ts(Data$log, start=1987, frequency = 12)
```

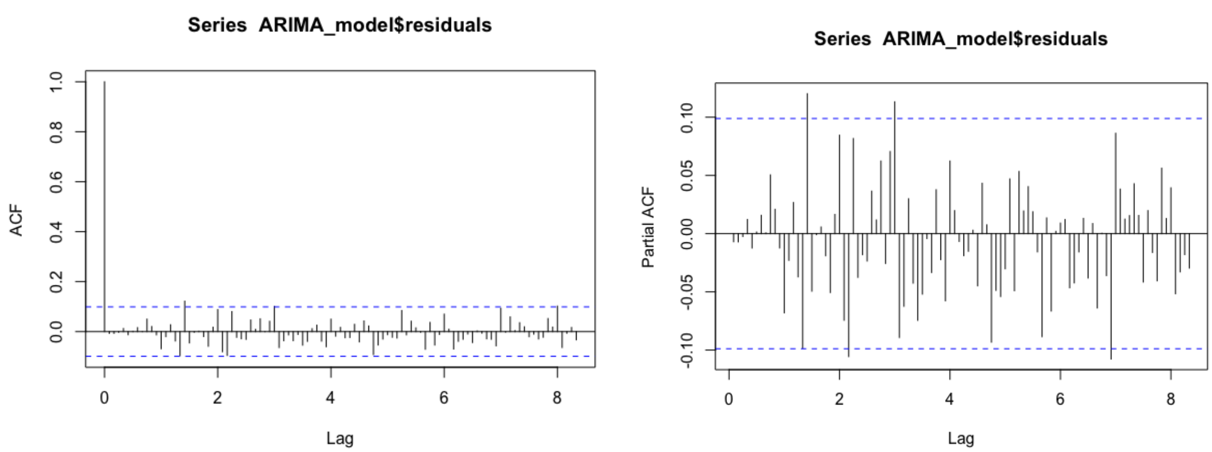
```r
#From Part 3.1.3.1, we know that degree of differencing to achieve
stationarity is 2.
#Hence, we analyze residuals at differencing degree of 2, d=2:
ARIMA_stationary <- arima(Data$log, order=c(0, 2, 0), method="ML")

#We look at ACF and PACF graphs:
#With 100 lags considered:
acf(ARIMA_stationary$residuals, lag.max=100)
pacf(ARIMA_stationary$residuals, lag.max=100)
#Zoom in with default no. of lags:
acf(ARIMA_stationary$residuals)
pacf(ARIMA_stationary$residuals)

#From the graph, we see: ACF is a damped sine wave, PACF cuts of at lag-12.
#This suggests AR signature.
#Therefore, we set parameters according to Box-Jenkins model: p=12, d=2,
q=0
#ARIMA(12, 2, 0)
ARIMA_model <- arima(Data$log, order=c(12, 2, 0), method="ML")
summary(ARIMA_model)

#By looking at summary, we can see that most parameters are significant.
This is good.
#Errors on average are also small. This is good too.

#We use Box-Pierce test to check if residuals are independently
distributed:
print(Box.test(ARIMA_model$residuals, lag=12))
print(Box.test(ARIMA_model$residuals, lag=20))
print(Box.test(ARIMA_model$residuals, lag=30))

#They are independently distributed, as we get high p-values and do not
reject the null hypothesis of Box-Pierce test.
#p-values: lag=12: 0.9929, lag=20: 0.7966, lag=30: 0.5233, lag=40: 0.5513,
lag=50: 0.7285

#Finally just to double check, we also look at the graphs of ACF and PACF
of the residuals of the model.
acf(ARIMA_model$residuals, lag.max = 100)
pacf(ARIMA_model$residuals, lag.max =100)

#The graphs show that residuals are majorly within significance bands.
#The residuals seem to be randomly independently distributed, and are
negligible. This is good.
#Another good feature of the model is that the residuals do not seem to
grow with increasing order of lag.
```

```
#Overall, this presents a good characteristic of the model.
#The model should perform well with a forecast, and it will be tested with
an in-sample forecast to confirm this.
```



Picture 13 *– ACF and PACF plots of ARIMA(0, 2, 0) residuals*

Above are plots for ACF and PACF of stationary data i.e. raw data differenced to second degree, i.e. ARIMA(0,2,0).



Picture 14 *– ACF and PACF plots of ARIMA(12, 2, 0) residuals*

Above are ACF and PACF graphs for residuals of the model. We can see that these are small, and randomly, independently distributed. This is desired, and we confirmed this observation with Box-Pierce test. The residuals also do not seem to increase with increasing lag, which is another desired feature that we can infer from the ACF and PACF graphs of the residuals.

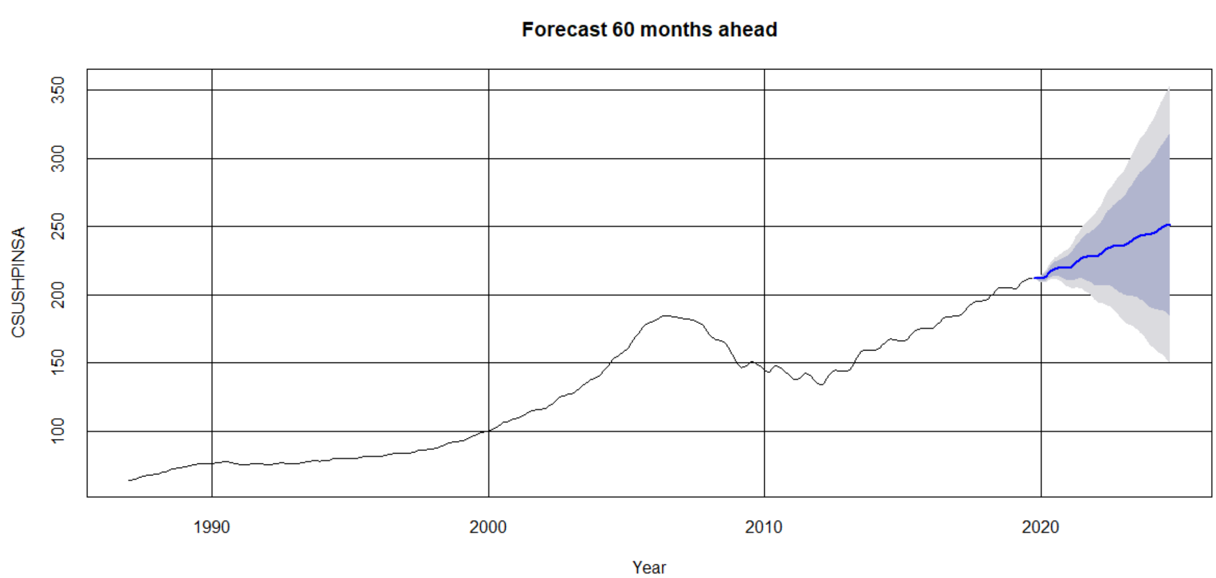
This model also minimizes AIC in the region. It has lower AIC than: ARIMA(13, 2, 0), ARIMA(11, 2, 0), ARIMA(12, 2, 1), ARIMA(12, 2, 2), ARIMA(13, 2, 1), ARIMA(13, 2, 2), ARIMA(11, 2, 1), ARIMA(11, 2, 2). The minimization of AIC shows further that the model is good.

The model should perform reasonably well with a forecast, and it will be tested with an in-sample forecast to confirm this.

3.3 Forecasting

This sub-section was contributed by Nguyen Long Son.

```
#From 3.1.3.1. and 3.1.3.2 we know that our data follows the ARMA(12, 2, 0)
#In this part we will make a forecast for the future evolution of Case-
Shiller Index

tData <- ts(Data$CSUSHPINSA, start = c(1987, 1), frequency = 12)

ARMA_model <- arima(tData, order = c(12,2,0), method = "ML")

plot(forecast(ARMA_model, h = 60), main = "Forecast 60 months ahead",
              xlab = "Year", ylab = "CSUSHPINSA", tck = 1 )

#plot the forecast for 60 months ahead

#the plot is shown below
```



***Picture 15** – 5-year out-of-sample forecast*

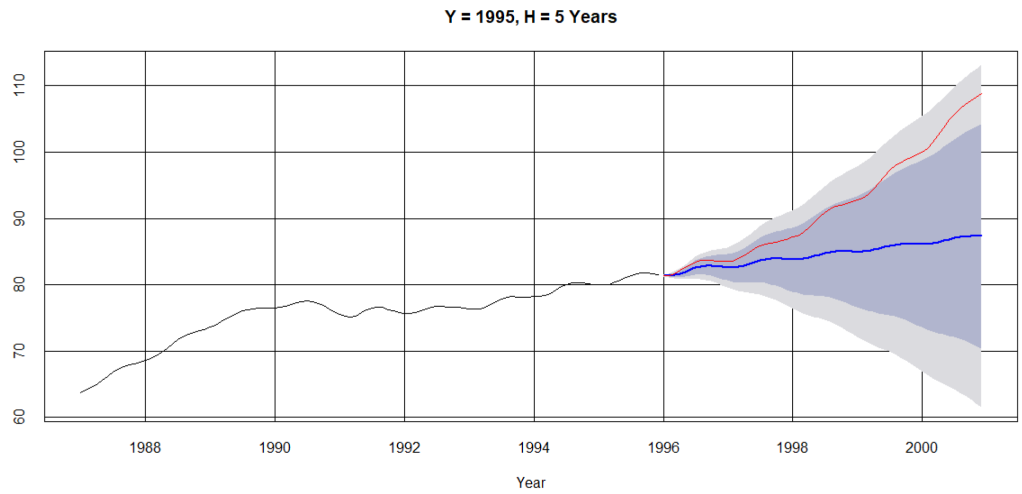
```r
# Explore the standard error as a percentage of the predicted value
ARMA_predict <- predict(ARMA_model, n.ahead = 60, se.fit = TRUE)
100*ARMA_predict$se/ARMA_predict$pred
```

```
            Jan        Feb        Mar        Apr        May        Jun        Jul
2019
2020  0.6522653  0.8537091  1.0552280  1.2527889  1.4464249  1.6361502  1.8247538
2021  3.3211443  3.6386046  3.9495943  4.2536354  4.5542090  4.8545999  5.1568204
2022  7.2640333  7.6481385  8.0176925  8.3769535  8.7344607  9.0967029  9.4661744
2023 11.9694788 12.3943473 12.7990456 13.1916105 13.5842816 13.9859364 14.4000253
2024 17.1769113 17.6275832 18.0547194 18.4692178 18.8856686 19.3145968 19.7604331
            Aug        Sep        Oct        Nov        Dec
2019                        0.1118006  0.2655155  0.4553626
2020  2.0120514  2.2086335  2.4370438  2.7046415  3.0054082
2021  5.4623030  5.7789678  6.1180938  6.4843698  6.8714432
2022  9.8445210 10.2368508 10.6492517 11.0824627 11.5277164
2023 14.8281167 15.2731556 15.7375003 16.2179580 16.7030871
2024 20.2246563 20.7083778
```

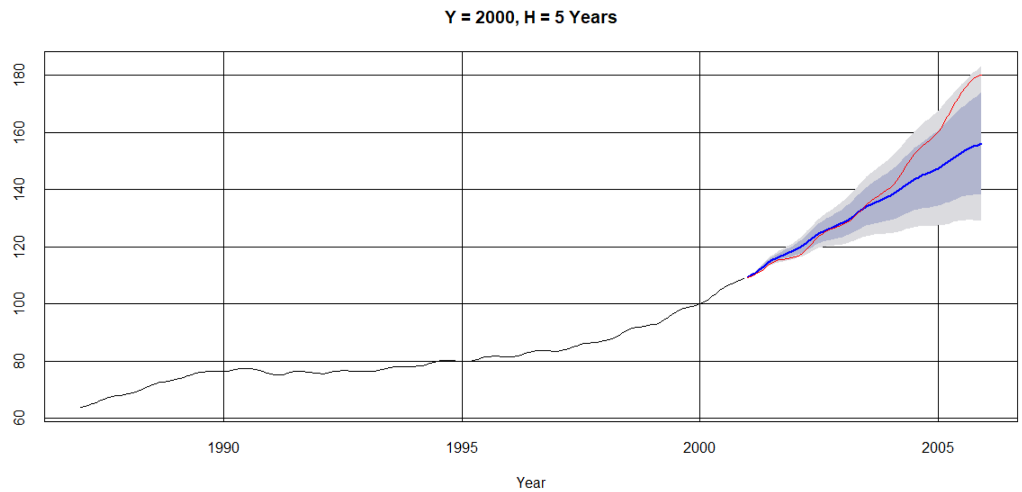
As we can see from the result, starting from 2020 Sep (i.e., 12 months from the last available value) the standard error starts to exceed 10% of the predicted value. Therefore, the model is suitable for predicting up to 1 year ahead

```r
# Now we will test the model using in-sample forecast. We will use a subset of the available data
# from year 1987 until year Y (1990 < Y < 2010) and forecast H year ahead (Y + H < 2019)
# after that we compare the predicted value with the available data of the same period

Y <- 2000 # Year
H <- 5 # Year ahead

tData <- ts(Data$CSUSHPINSA, start = c(1987, 1), frequency = 12)

tData_isp <- window(tData, start = c(1987,1), end = c(Y,12))
# tData_isp is the in-sample data we use to build the model

tData_pred <- window(tData, start = c(Y+1,1), end = c(Y+H,12))
# tData_pred is the known data we use to compare with the predicted value built using tData_isp

ARMA_model <- arima(tData_isp, order = c(12,2,0), method = "ML")

plot(forecast(ARMA_model, h = 12*H), main = "")
lines(tData_pred, col = "red")
#plotting predicted value against known value on the same plot
```

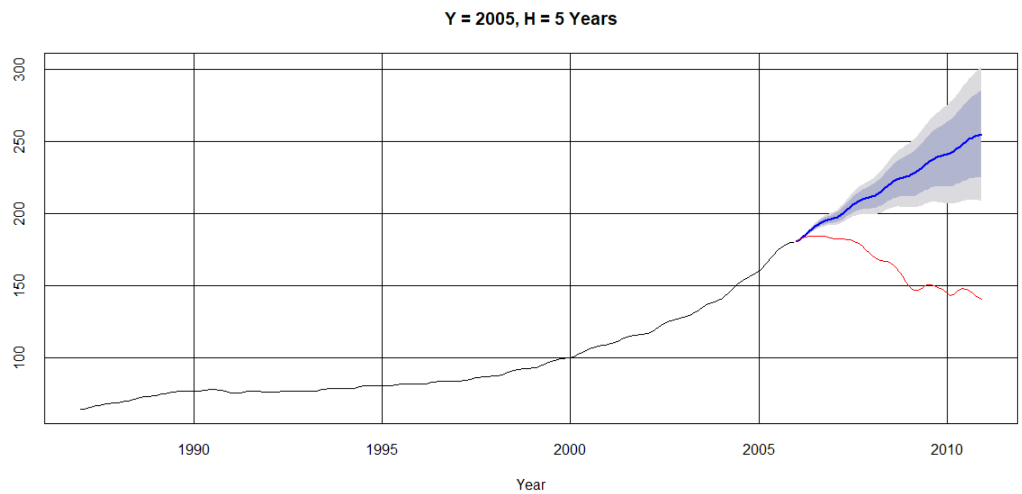
Below are some results with different Y and H. On the plots below, the red line is the known value while the blue line is the predicted value



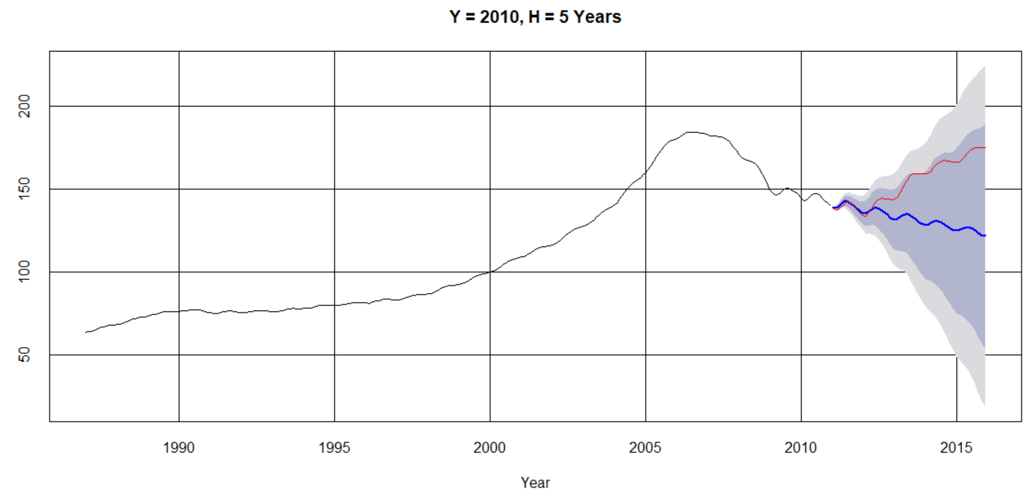
Picture 16 *– 5-year in-sample forecast, beginning 1995*



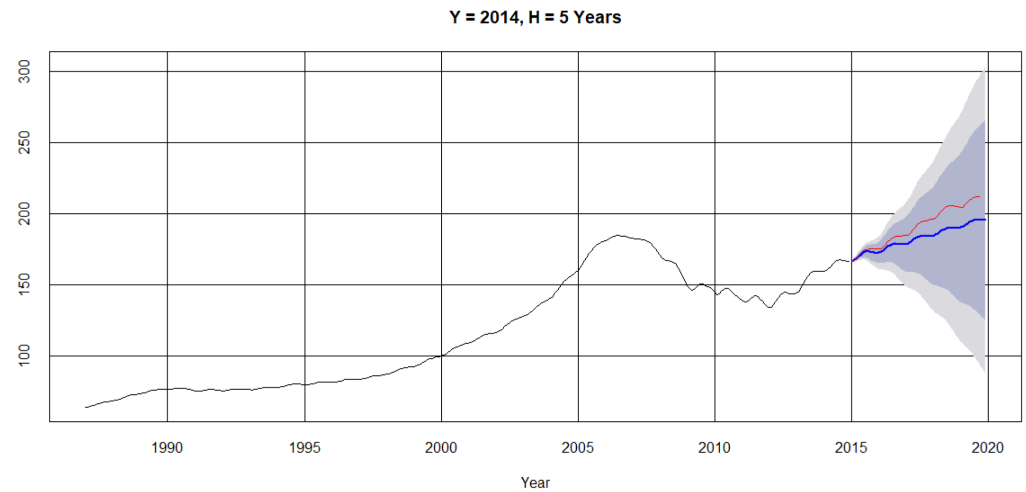
Picture 17 *– 5-year in-sample forecast, beginning 2000*



Picture 18 *– 5-year in-sample forecast, beginning 2005*



Picture 19 *– 5-year in-sample forecast, beginning 2010*



Picture 20 *– 5-year in-sample forecast, beginning 2014*

We can see that, if we choose Y from 1990 to 2000, or Y from 2010 to 2014, the predicted value is well in line with the known value within 1 year. However, if choosing Y from 2004 - 2007, the predicted value significantly differs from the real value. The predicted value of housing index continues to rise while the real index drops largely. The reason for that is the housing bubble during 2007 - 2009, the period of large volatility in house prices.

3.4 Suggested improvement exogenous variables

This sub-section was contributed by Matthew Ang.

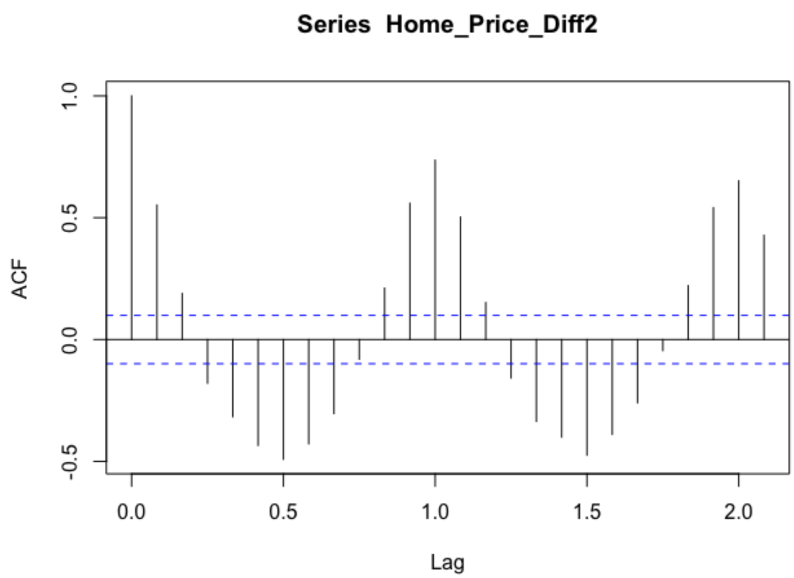
The Case-Shiller Index is a monthly housing price index that measures the value of single-family home price indices using the repeat sales methodology. The index is constructed using a three-month moving average, and attempts to capture robust valuations by e.g. excluding homes with significant physical change, statistical anomalies and high-turnover houses (S&P Global, 2019).

As the housing price index is on an aggregate level, exogenous variables should be on a national level. Indeed, macroeconomic factors like the **job market**, **mortgage market**, and **economic outlook** could be useful in improving our forecast. We dive into each variable in turn:

- **Job market:** Analysis from Zimmermann (2017) suggests *household income growth* leads to higher housing prices. We can also consider *unemployment rate*, a leading indicator to wage growth.
- **Mortgage market:** While mortgage rates from Freddie Mac could be indicative of housing prices, structural variables such as *percentage of banks tightening mortgage lending* and *mortgage originations growth* are better indicators of the trends underlying the mortgage market.
- **Economic outlook:** Producer and consumer outlook indicators (such as the *Purchasing Managers' Index* and the *Consumer Confidence Index*) illustrate citizens' expectations of the future economy, and in turn affect their willingness to purchase houses.

Other housing indices can also be incorporated. While we expect other housing indices to be coincident, Moody's Case-Shiller Index methodology (2018) illustrated that the *FHFA Price Index* Granger-causes the Case-Shiller Index.

While not exogenous, **seasonal terms** can also be included in the model. This reuslts in a seasonal ARIMA. The ACF plot of the second-differenced Case-Shiller Index exhibits a cyclical pattern in the residuals. Thus, the data exhibits a seasonal structure, where *seasonal dummies* and/or *seasonal autoregressors* can be useful.



***Picture 21** – ACF plot of twice-differenced Home-Shiller Index*

Inclusion of such terms changes our implicit assumptions of the data-generating process. It is therefore important that we re-examine ARIMA's *p*, *q* parameters when including these seasonal terms.

References

Carbacho-Burgos, A., & Campbell, B. (2018, September). *Moody’s Analytics Case-Shiller Home Price Index Forecast Methodology*. From Moody's Analytics: https://www.economy.com/home/products/samples/case-shiller-methodology.pdf

S&P Dow Jones Indices LLC. (2019). *S&P CoreLogic Case-Shiller Home Prices Indices Methodology*. From S&P Global: https://us.spindices.com/documents/methodologies/methodology-sp-corelogic-cs-home-price-indices.pdf

S&P Dow Jones Indices LLC. (2019, December). *S&P/Case-Shiller U.S. National Home Price Index [CSUSHPINSA]*. From FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/CSUSHPINSA

Verizon Media. (2019). *Historical stock prices*. From Yahoo Finance: http://finance.yahoo.com/

Zimmermann, C. (2017, October 30). *Incomes determine house prices: An illustration for San Francisco*. From FRED, Federal Reserve Bank of St. Louis: https://fredblog.stlouisfed.org/2017/10/incomes-determine-house-prices/