Time Series Project

Analysis of House Price Index of East South Central Division

k530005008 HE Yujia
k530012043 TANG Yihan
k530005037 XIE Weiyan
k530005039 XIONG Feng
k530013026 WANG Siyi
**k530005042 YANG Lu**

2019/1/1

Contents

Introduction

By using the traditional standard mortgage transaction data obtained from the Federal Home Loan Mortgage Corporation named Freddie Mac and the Federal National Mortgage Association named Fannie Mae, Quarterly house price indexes for single-family are estimated and published by The Office of Federal Housing Enterprise Oversight (OFHEO). OFHEO's published house price indexes, called "HPI", are the revised version of the weighted-repeat sales (WRS) method brought up by Case and Shiller (1989). In the nine U.S. Census divisions, the fifty states and the District of Columbia, the quarterly HPI are estimated and reported. Since January 1975, Freddie Mac and Fannie Mae have provided and purchased at least two mortgages on individual single-family residential properties and the HPI is recorded by using repeated observations of housing values. The HPI provides broad geographic coverage through the national operations of two government-sponsored housing enterprise data. This is the reason why an index based on enterprise data is of importance. However, some limitations exist in the coverage of the HPI. Therefore, in order to fix the limitations, the HPI is updated in every quarter because the enterprises purchase additional mortgages which are used to identify additional repeat transactions for the most recent quarter and all previous quarters.

The data used in this project is the Federal Housing Finance Agency HPI. It is a weighted, resale index which measures average price changes of a resale on the properties. By examining repeat mortgage transactions on single-family homes that have been bought or securitized by Fannie Mae or Freddie Mac, we gain the information and data.

```r
ts <-read.csv("~/Desktop/seansonal.csv")
ts=ts[1:333,]
names(ts)<-c("month","index")

index_ts<-ts$index
index_ts<-log(index_ts)

index_ts2<-ts(index_ts, start=1991,frequency =12)
plot(index_ts2,ylab='House Price Index',
main="Time Series Plot of House Price Index")
```

ts(ts, start = c(1991, 1, 1), frequency = 12)
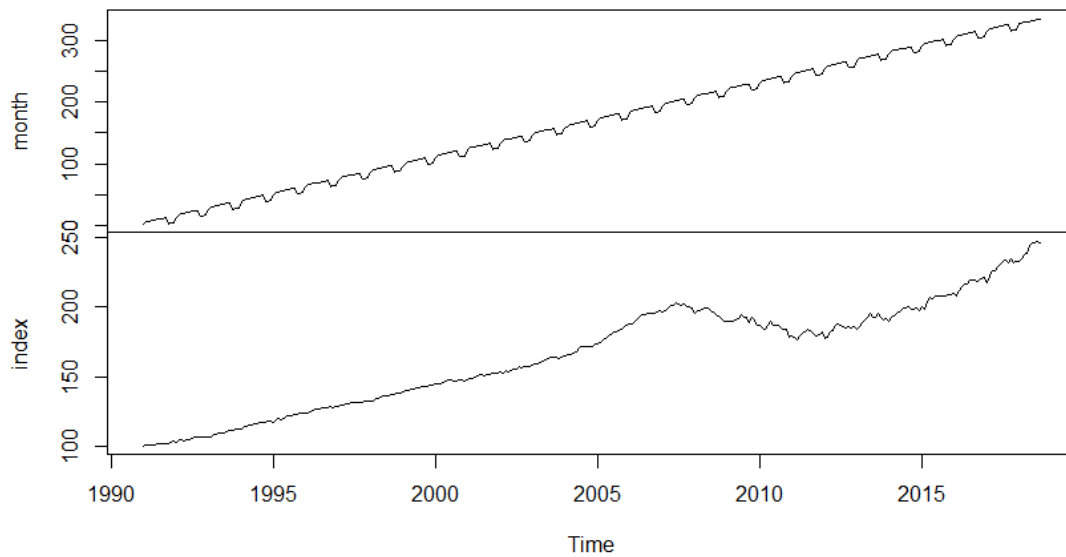
We firstly draw the scatter plot for all the data. From the figure of the raw data, we can easily notice that there exists an upward trend in the scatter plot in the interval from 1990 to 2007 approximately. However, a drop appears in the interval from 2007 to 2011 approximately. Then after 2011, the plot regains the upward trend. We also notice that there exist seasonal changes. Most of data are closely related and therefore there exists a strong correlation among these points.

```
acf(index_ts2,main="ACF Plot of House Price Index",lag.max =72)
```

ACF Plot of House Price Index
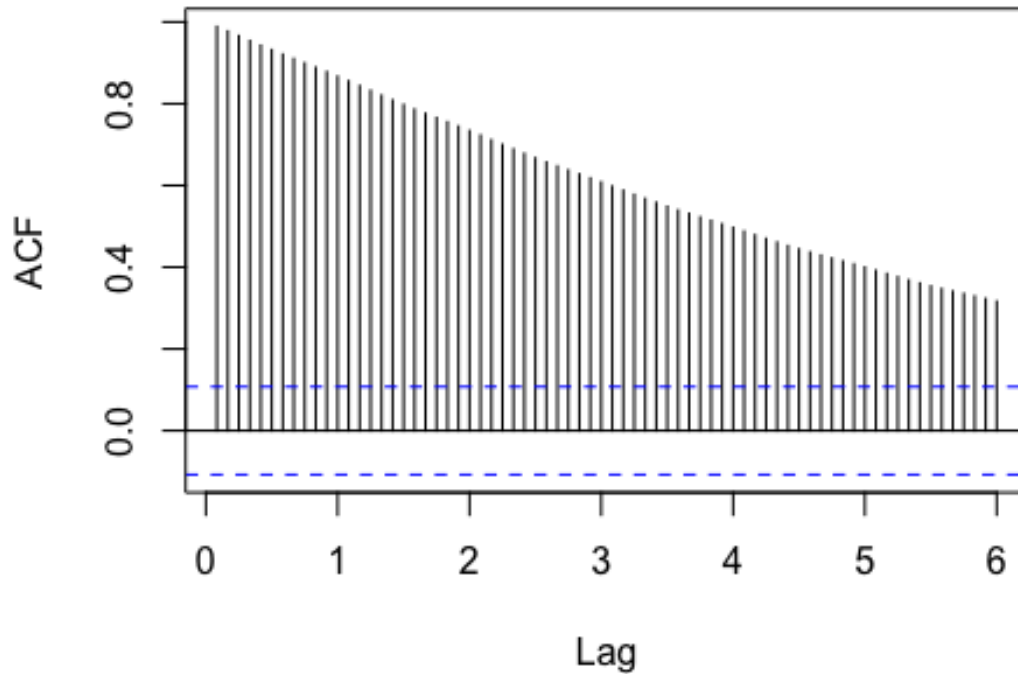
Then we plot the autocorrelation function plot and partial autocorrelation function plot of the raw data. The ACF plot indicates that the sequential points have strong correlation while the raw data is non-stationary. Therefore, we need to make the raw data stationary to continue analyzing the time series data.

```
pacf(index_ts2,main="PACF Plot of House Price Index")
```

PACF Plot of House Price Index
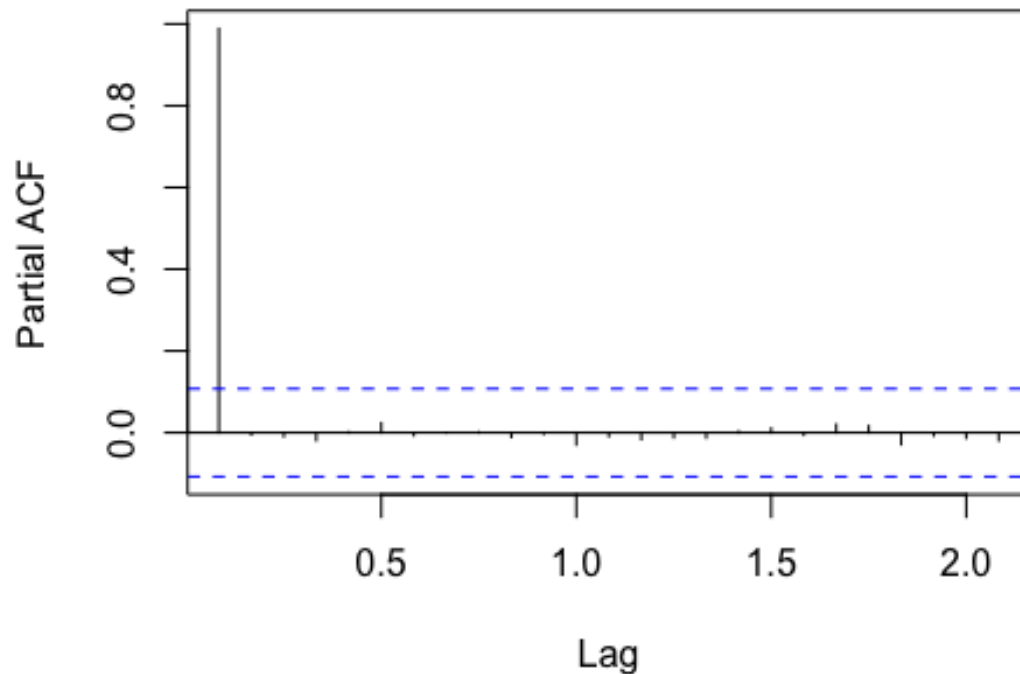

```
#Calculation of the degree of correlation on Ist Lag
y=a_deseasonal
x=zlag(a_deseasonal)
index =2:length(x)
cor(y[index],x[index])

## [1] 0.9472112
```

We also calculate the correlation between every two points and the result shows that there exists high correlation among sequential points because the value of correlation between any two points is significantly high.

```
a_deseasonal=diff(index_ts2, lag=12)

plot(a_deseasonal,ylab="seasonal=12")

index_ts2_d1 = diff(a_deseasonal,differences = 1)
adf.test(index_ts2_d1, alternative ="stationary")
```
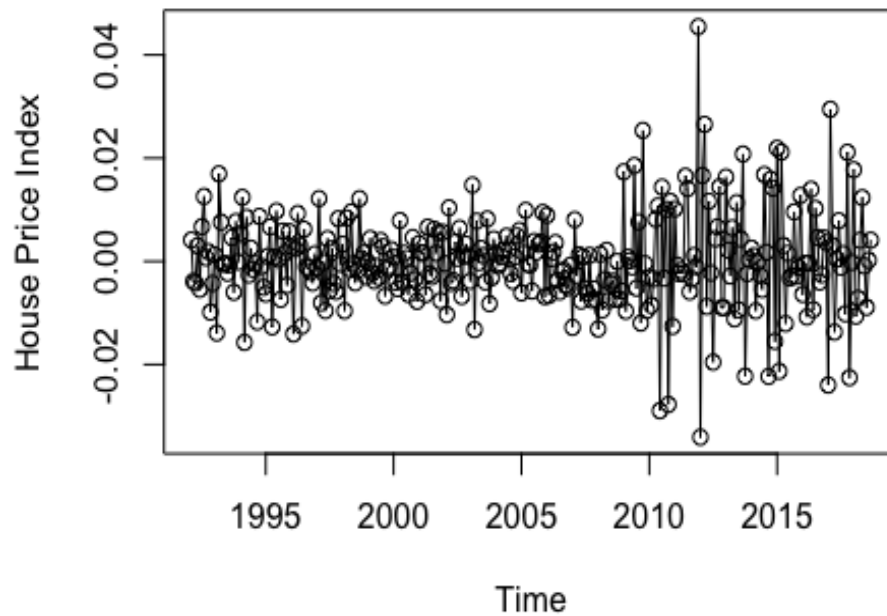
```
## Warning in adf.test(index_ts2_d1, alternative = "stationary"): p-val
ue
## smaller than printed p-value

##
##   Augmented Dickey-Fuller Test
##
## data:  index_ts2_d1
## Dickey-Fuller = -5.4242, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary

plot(index_ts2_d1,type='o',ylab='House Price Index',
main="Time Series Plot of Ist difference HPI")
```
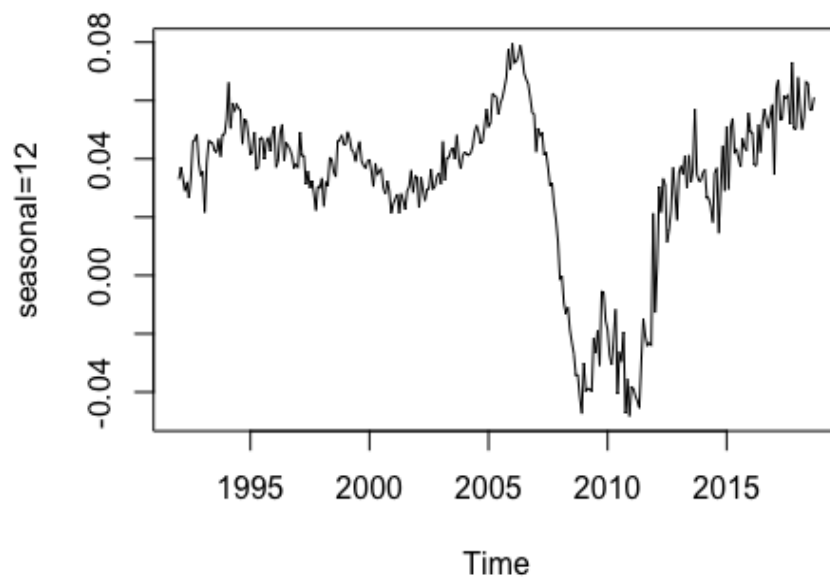
Time Series Plot of Ist difference HPI

From the figure, we can notice that the plot is almost on one horizontal line which means that the time series may become stationary. The Time versus House Price Index plot shows a change in variance.

We plot the ACF and PACF plot again to see more characteristics of this time series.

```
##ACF & PACF plot of the Stationary Time Series
acf(index_ts2_d1,lag.max=36)
```

Series index_ts2_d1
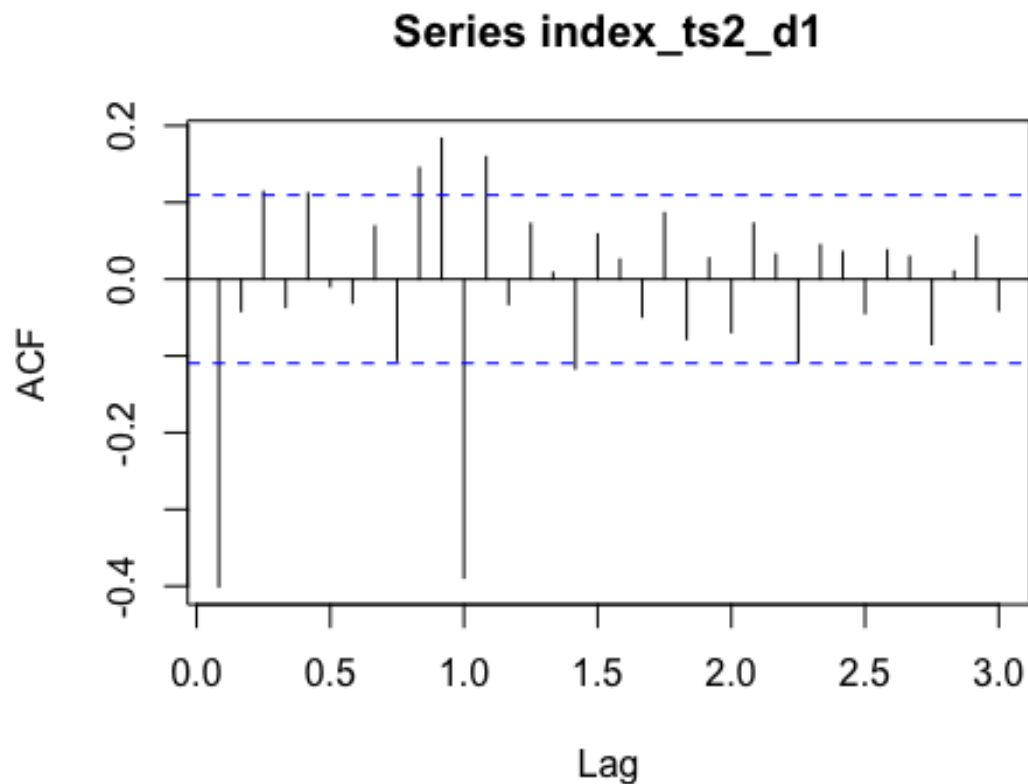
From the ACF plot, we can see that the first lag and the twelfth lag are negatively significant and the tenth lag, eleventh lag, thirteenth lag and seventeenth lag are slightly significant. Therefore, we suspect the seasonality of the time series. However, we ignore the further influence of the seasonal aspect because the values are seasonally adjusted.
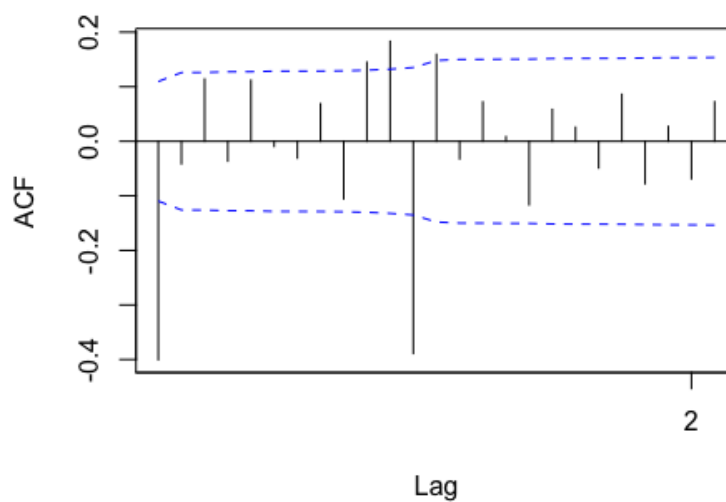
```
pacf(index_ts2_d1,main="PACF Plot of House Price Index")
```

PACF Plot of House Price Index
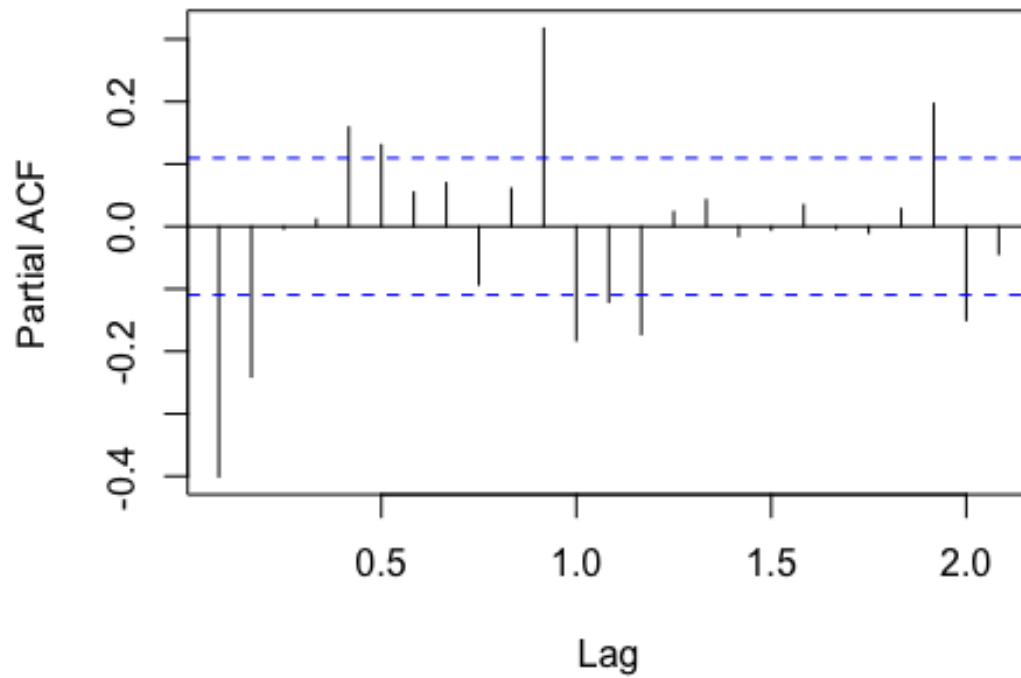

```
acf(index_ts2_d1,ci.type='ma',xaxp=c(0,20,10),main="Alternative Bounds
for IInd Difference HPI")
```

Alternative Bounds for IInd Difference HPI

Model Specification

```
eacf(index_ts2_d1)

## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o x o o o o o o x x  x  x  o
## 1 x x x o x o o o o o x  x  o  o
## 2 o o x o x o x o o o o  o  x  x  o
## 3 x o x o x o o o o o o  x  x  o
## 4 o o x x o o o o o o o  x  o  o
## 5 x x x x x x o o o o o  x  o  o
## 6 x x o x o o o o o o o  x  o  o
## 7 x x x o x o o o o o o  x  o  o

res =armasubsets(y=index_ts2_d1,nar=8,nma=8,y.name='test',ar.method='ol
s')
plot(res)
```
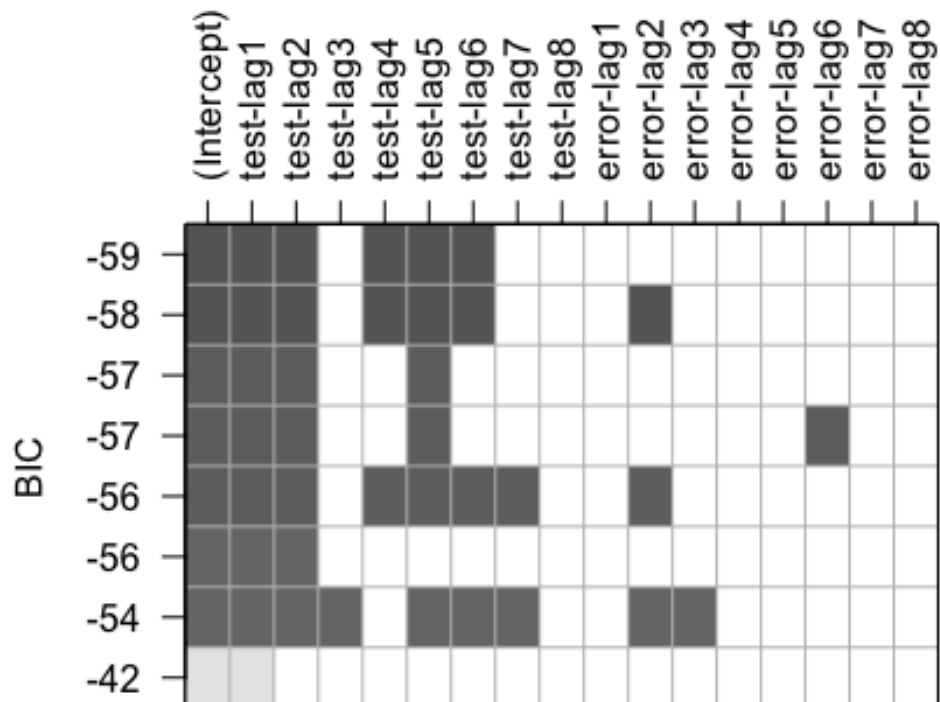

The eacf plot shows the candidate models selected could be ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,1) & ARIMA(1,1,2).

For further investigation of the candidate models, let's plot the BIC table.

The BIC table shows slight higher orders for AR and MA components at -220. We will consider the lower orders only for our analysis. So the BIC table shows the candidate models selected could be ARIMA(4,1,1)

Overall, the candidate models selected will be ARIMA(0,1,0), ARIMA(0,1,1) ARIMA(1,1,1), ARIMA(1,1,2) & ARIMA(4,1,1).

Model Estimation for ARIMA(0,1,0)

```
m010_hpi =arima(a_deseasonal,order=c(0,1,0))
AIC(m010_hpi)

## [1] -2097.762

BIC(m010_hpi)

## [1] -2093.994
```

Model Estimation for ARIMA(0,1,1)

```
m011_hpi =arima(a_deseasonal,order=c(0,1,1))
AIC(m011_hpi)

## [1] -2163.709

BIC(m011_hpi)

## [1] -2156.172

coeftest(m011_hpi)

##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1 -0.445607   0.042216 -10.555 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Estimation for ARIMA(1,1,1)

```
m111_hpi =arima(a_deseasonal,order=c(1,1,1))
AIC(m111_hpi)

## [1] -2163.447
```

```
BIC(m111_hpi)

## [1] -2152.142

coeftest(m111_hpi)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.125397   0.093186 -1.3457    0.1784
## ma1 -0.363553   0.079626 -4.5658 4.977e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Estimation for ARIMA(1,1,2)

```
m112_hpi =arima(a_deseasonal,order=c(1,1,2))
AIC(m112_hpi)

## [1] -2178.382

BIC(m112_hpi)

## [1] -2163.309

coeftest(m112_hpi)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.818641   0.066083  12.388 < 2.2e-16 ***
## ma1 -1.395997   0.071910 -19.413 < 2.2e-16 ***
## ma2  0.565215   0.054608  10.350 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Estimation for ARIMA(4,1,1)

```
m411_hpi =arima(a_deseasonal,order=c(4,1,1))
AIC(m411_hpi)

## [1] -2171.007

BIC(m411_hpi)

## [1] -2148.397

coeftest(m411_hpi)
```
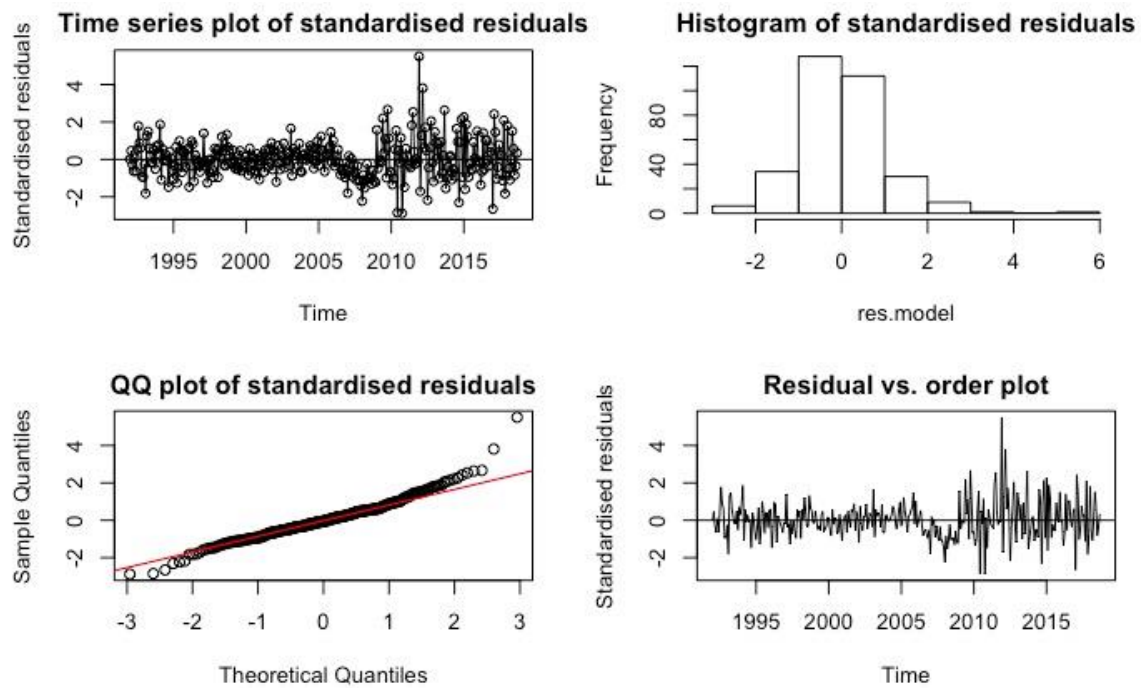
```
##
## z test of coefficients:
##
##        Estimate Std. Error z value  Pr(>|z|)
## ar1  0.267968   0.102189   2.6223  0.008734 **
## ar2  0.163188   0.067843   2.4054  0.016155 *
## ar3  0.231728   0.057102   4.0581 4.947e-05 ***
## ar4  0.103537   0.058012   1.7847  0.074304 .
## ma1 -0.784528   0.088197  -8.8952 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the significance of the coefficients and the AIC and BIC values the candidate models would be ARIMA(0,1,1) ARIMA(1,1,1), ARIMA(1,1,2) & ARIMA(4,1,1) with seasonal period=12.

In order to filter these models further, we try to analyze the residuals of the model selected above. In this part, the 4-in-1 residual plot, the Shapiro-wilk test and the Ljung-Box test are involved to verify whether our assumptions of residual of each time series models are satisfied. To remind you, the 4-in-i residual plot is composed with normal probability plot of residuals, residuals versus fits plot, residuals versus order plot and histogram of residuals. The histogram of the residuals shows the distribution of the residuals for all data. The normal probability plot of the residuals displays the residuals versus their expected values in the condition that the distribution is normal. The residuals versus fits graph plots the residuals and fitted values on the y-axis and on the x-axis accordingly. The residual versus order plot displays the residuals in the order that the data were recorded. In addition, the Shapiro-wilk test is a way to determine if a random sample comes from a normal distribution. Last but not the least, we use the Ljung-box test to determines whether or not errors are independent and identical distribution, white noise, or whether there still exist something useful for us to examine. Whether or not the autocorrelations for the errors or residuals are non-zeros. In addition, it is a test of lack of fit which means if the autocorrelations of the residuals are very small, then we say that the model doesn't show 'significant lack of fit'.

Time series plot of standardised residuals

Histogram of standardised residuals

QQ plot of standardised residuals
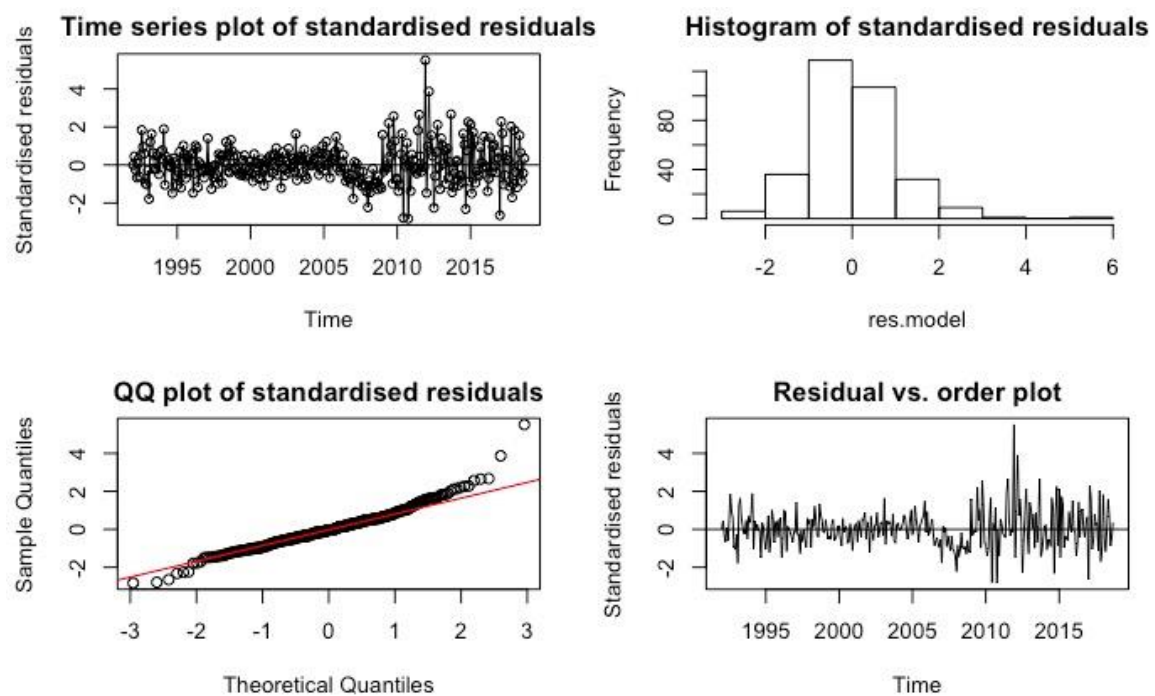
Residual vs. order plot

```
## 
##   Shapiro-Wilk normality test
## 
## data:  res.model
## W = 0.96345, p-value = 3.269e-07
```


Ljung-Box Test
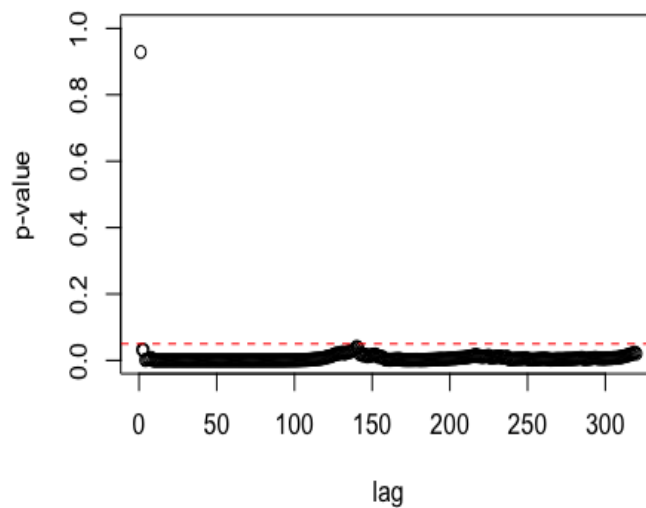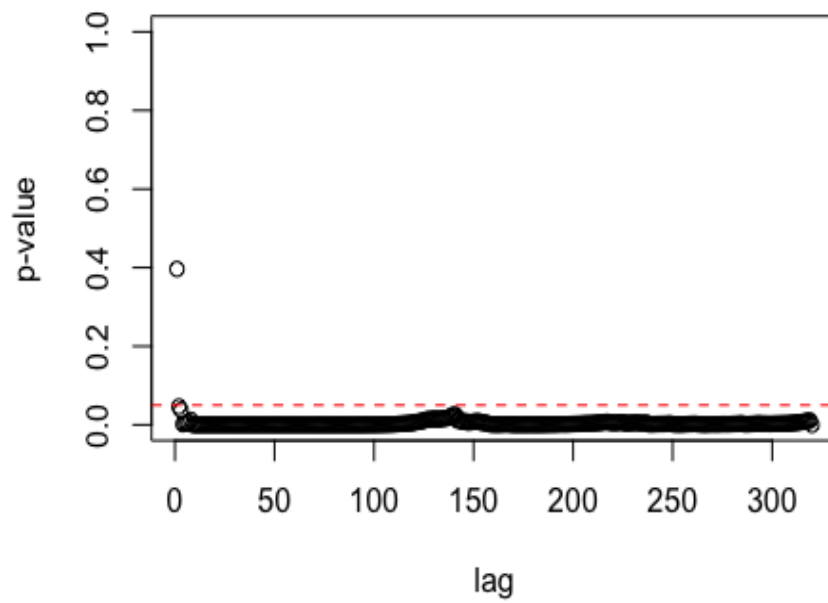
```
##
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.96345, p-value = 3.269e-07
```

ARIMA (1, 1, 1)


```
##
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.9625, p-value = 2.388e-07
```

Ljung-Box Test


```
Box.test (m111residuals, type ="Ljung")
##   Box-Ljung test
##
## data:  m111residuals
## X-squared = 0.028342, df = 1, p-value = 0.8663
```

Time series plot of standardised residuals

Histogram of standardised residuals

QQ plot of standardised residuals

Residual vs. order plot

```
##
##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.97214, p-value = 7.283e-06
```

Ljung-Box Test

17

```
Box.test (m112residuals, type ="Ljung")
##   Box-Ljung test
## 
## data:  m112residuals
## X-squared = 0.3693, df = 1, p-value = 0.5434
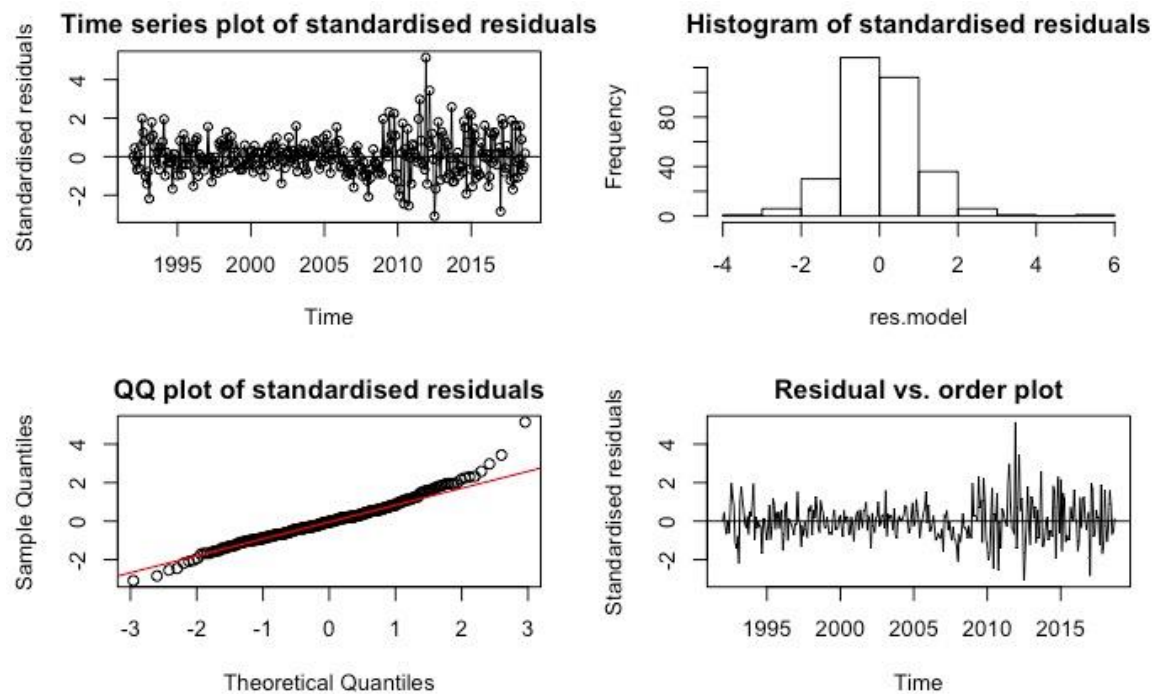```

ARIMA (4, 1, 1)



```
## 
##   Shapiro-Wilk normality test
## 
## data:  res.model
## W = 0.96742, p-value = 1.277e-06
```

Ljung-Box Test

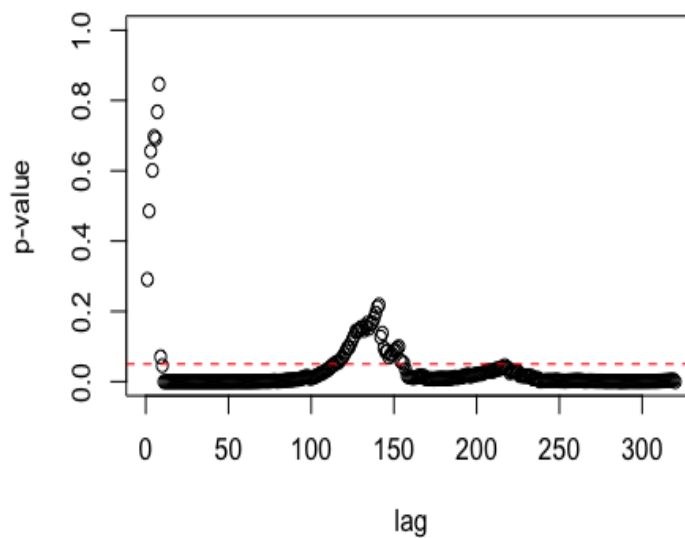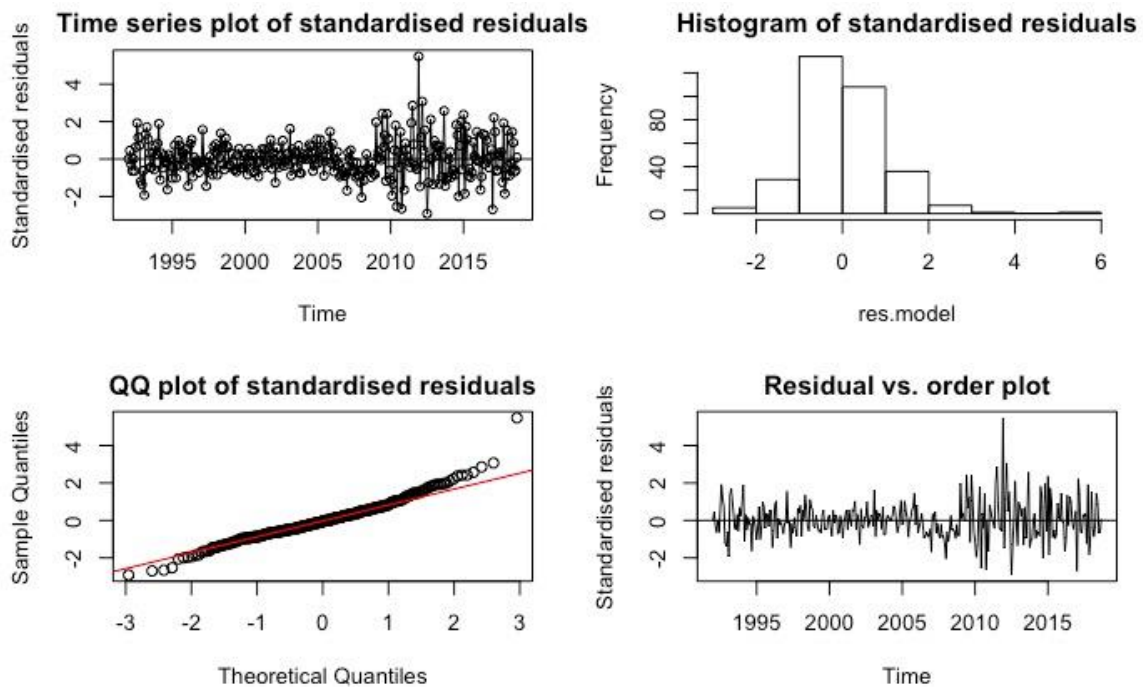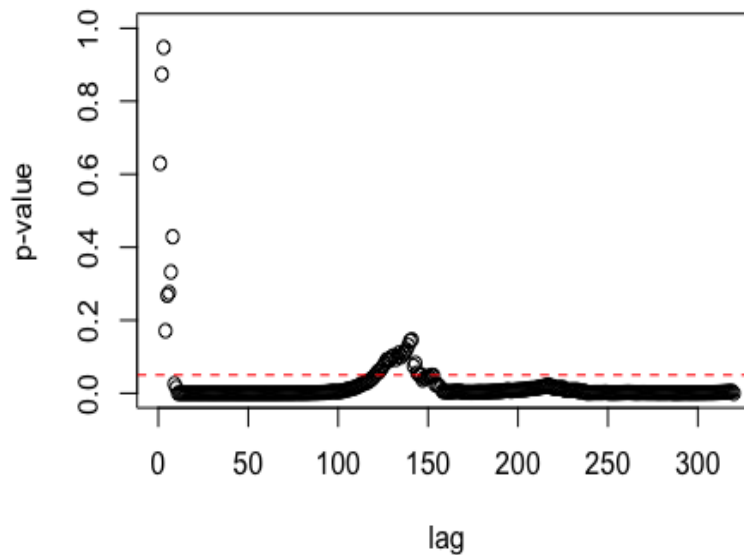```
Box.test (m411residuals, type ="Ljung")
Box-Ljung test
##
data: m411residuals
X-squared = 0.063403, df = 1, p-value = 0.8012
```

Calling the residual analysis function, all the models ARIMA (0,1,1), ARIMA (1,1,1), ARIMA (1,1,2) &ARIMA (4,1,1)doesn't satisfy the assumptions of residual which are independent and identical distribution, zero mean, constant variance and follow normal distribution. The residuals of all the models failed the Shapiro Wilk test. Hence, the residuals are not normally distributed. The Ljung-Box test plot shows a good separation from the p-value reference line for the models, and the plot is not good. The QQ-Plot showing a fat tails makes us suspect conditional heteroscedasticity in the time series.

Garch Model

   From the L-jung test and normity test we can see that this ARMA model is not perfect to fit the model, so we try to use the GRACH model, that is consider the vairance of the model is not the same. We consider the heteoscedasticity in the time series. Because the GRACH model is very difficult to find the parameter of the model, therefore, we consider using the GRACH(1,0), GRACH(1,1),GRACH(2,1),GRACH(2,2) one by one. Therefore, we try to analysis every model one by one to see if there are better model to fit the data. First we use the model that we choose from the previous calculation, and we apply GARCH model one by one.

```
m011residuals =residuals(m011_hpi)
m111residuals =residuals(m111_hpi)
m112residuals =residuals(m112_hpi)
m411residuals =residuals(m411_hpi)


a112garch10=garch(m112residuals,order=c(0,1),trace=F)
a112garch11=garch(m112residuals,order=c(1,1),trace=F)
a112garch21=garch(m112residuals,order=c(2,1),trace=F)

a112garch22=garch(m112residuals,order=c(2,2),trace=F)
```

Case 1: ARIMA (0, 1, 1) $\longrightarrow$ GARCH (1,0)

```
garchresidual.analysis(a011garch10)

Call:
garch(x = m011residuals, order = c(0, 1), trace = F)

 Model:
 GARCH(0,1)

Residuals:
      Min        1Q    Median        3Q       Max
-2.78218  -0.59745  -0.06361   0.54094   5.99417

## Coefficient(s):
```
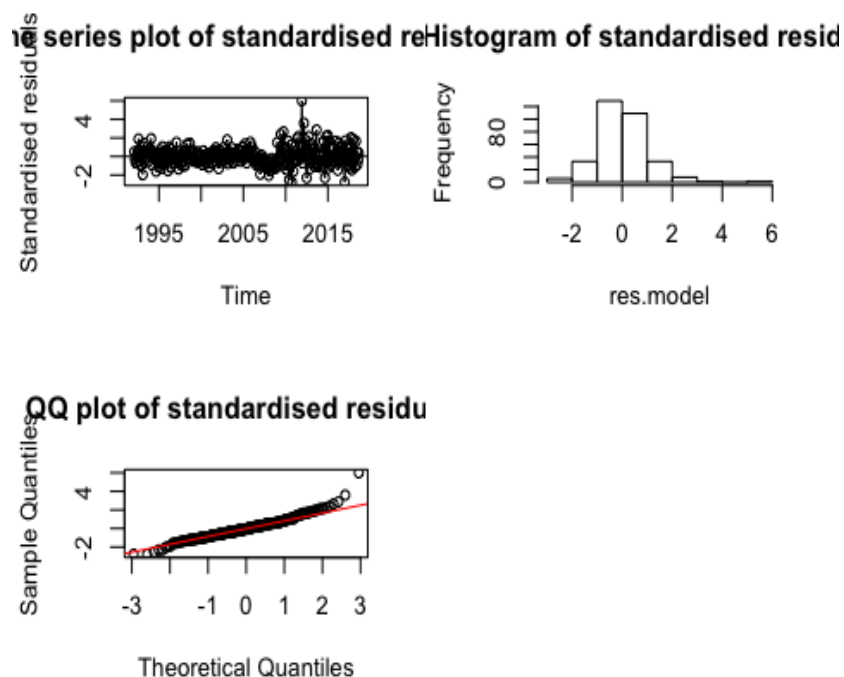
```
##      Estimate  Std. Error  t value Pr(>|t|)
## a0 5.636e-05   2.924e-06   19.273   <2e-16 ***
## a1 1.732e-01   7.195e-02    2.408    0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 256.53, df = 2, p-value < 2.2e-16
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.13392, df = 1, p-value = 0.7144

##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.95664, p-value = 3.938e-08
```

Shapiro-Wilk normality test has p-value smaller than 0.05, it fails to prove the normality of the data. And Box-Ljung test has p-value equals to 0.7144, so it accept original assumption that the error are uncorrelated.

series plot of standardised re / Histogram of standardised resid

QQ plot of standardised residu

Case 2: ARIMA (1,1,2) ⟶ GARCH(1,1)

```
garchresidual.analysis(a112garch11)

##
## Call:
## garch(x = m112residuals, order = c(1, 1), trace = F)
##
## Model:
## GARCH(1,1)
##
## Residuals:
##       Min        1Q     Median        3Q        Max
## -2.75153  -0.67813  -0.03247   0.57275   4.05947
##
## Coefficient(s):
##      Estimate  Std. Error  t value  Pr(>|t|)
## a0  3.723e-07   4.022e-07    0.926   0.35466
## a1  6.252e-02   2.225e-02    2.811   0.00495 **
## b1  9.338e-01   2.435e-02   38.358   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 11.968, df = 2, p-value = 0.002519
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.096133, df = 1, p-value = 0.7565

##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.99155, p-value = 0.06427
```

Shapiro-Wilk normality test has p-value greater than 0.05, so we deem that this result is not bad to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.7565, so it accept original assumption that the error are uncorrelated. Therefore we can use the GARCH (1, 1) to fit the model.
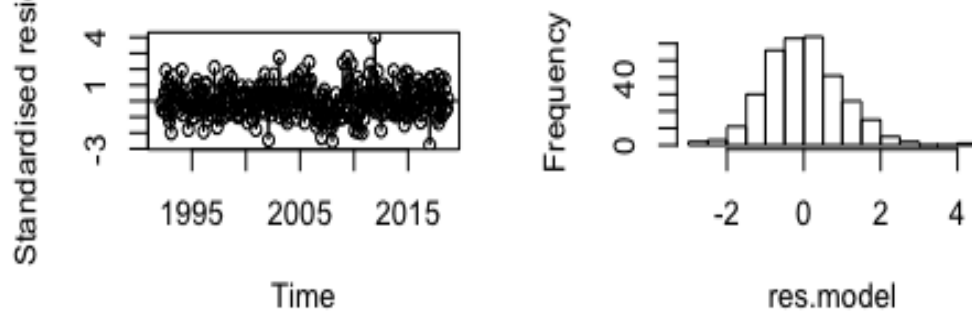
Case 3: ARIMA (1,1,2) ⟶ GARCH(2,1)

```
garchresidual.analysis(a112garch21)

##
## Call:
## garch(x = m112residuals, order = c(2, 1), trace = F)
##
## Model:
## GARCH(2,1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -2.76722 -0.67388 -0.03354   0.57569   4.03458
##
## Coefficient(s):
##      Estimate  Std. Error  t value  Pr(>|t|)
## a0 6.460e-07         NA       NA        NA
## a1 1.052e-01         NA       NA        NA
## b1 2.282e-01         NA       NA        NA
## b2 6.597e-01         NA       NA        NA
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 11.26, df = 2, p-value = 0.003589
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.48412, df = 1, p-value = 0.4866

##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.99137, p-value = 0.05901
```
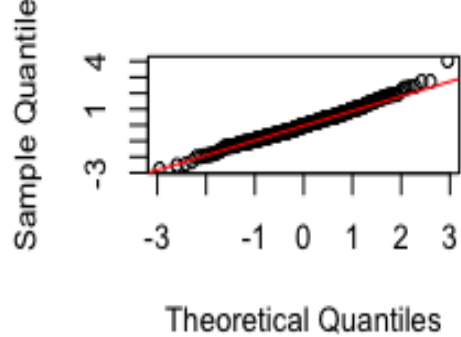
   This result is satisfy the normality of the model since the Shapiro-Wilk normality test has p value is greater than 0.05, therefore, we think this model is okay to fit the data. And also the Box-Ljung test also indict the error are uncorrelated.

ᵢₑ series plot of standardised re Histogram of standardised resid

QQ plot of standardised residu

Case 4: ARIMA (1, 1, 2) ⟶ GARCH (2,2)

```
##
## Call:
## garch(x = m112residuals, order = c(2, 2), trace = F)
##
## Model:
## GARCH(2,2)
##
## Residuals:
##      Min       1Q    Median        3Q       Max
## -3.26495 -0.61362 -0.02483   0.54663   5.16489
##
## Coefficient(s):
##      Estimate  Std. Error  t value Pr(>|t|)
## a0 1.774e-06   1.881e-06     0.943   0.3454
## a1 3.601e-01   1.549e-01     2.325   0.0201 *
## a2 4.302e-07   1.917e-01     0.000   1.0000
## b1 2.351e-01   3.501e-01     0.671   0.5019
## b2 4.660e-01   2.674e-01     1.743   0.0813 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 112.64, df = 2, p-value < 2.2e-16
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.2747, df = 1, p-value = 0.6002
##
##
##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.96986, p-value = 3.659e-06
```

Shapiro-Wilk normality test has p-value fails to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.6002, so it accepts original assumption that the error is uncorrelated.

Case 5: ARIMA (4, 1, 1) ⟶ GARCH (0, 1)

```
##
## Call:
## garch(x = m411residuals, order = c(0, 1), trace = F)
##
## Model:
## GARCH(0,1)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -3.13356 -0.61095 -0.05415  0.55960   5.79071
##
## Coefficient(s):
##      Estimate  Std. Error  t value Pr(>|t|)
## a0 5.429e-05   3.392e-06   16.006   <2e-16 ***
## a1 1.589e-01   8.456e-02    1.879   0.0602 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 189.21, df = 2, p-value < 2.2e-16
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.059032, df = 1, p-value = 0.808
##
##
##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.96172, p-value = 1.92e-07
```

Shapiro-Wilk normality test has p-value fails to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.808, so it accepts original assumption that the error is uncorrelated.

Case 6: ARIMA (4, 1, 1) ⟶ GARCH (1, 1)

```
##
## Call:
## garch(x = m411residuals, order = c(1, 1), trace = F)
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -3.25706 -0.58473 -0.05052  0.56816   4.37559
##
## Coefficient(s):
##      Estimate  Std. Error  t value Pr(>|t|)
## a0 1.956e-06   1.247e-06    1.569  0.11664
## a1 2.528e-01   9.059e-02    2.791  0.00526 **
## b1 7.779e-01   5.535e-02   14.056  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 79.96, df = 2, p-value < 2.2e-16
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.2491, df = 1, p-value = 0.6177
##
##
##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.9708, p-value = 5.348e-06
```
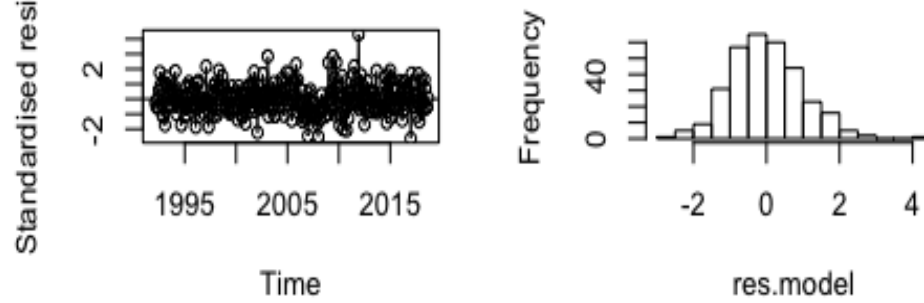
Shapiro-Wilk normality test has p-value fails to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.6177, so it accepts original assumption that the error is uncorrelated.

Case 7: ARIMA (4, 1, 1) $\longrightarrow$ GARCH (2, 1)
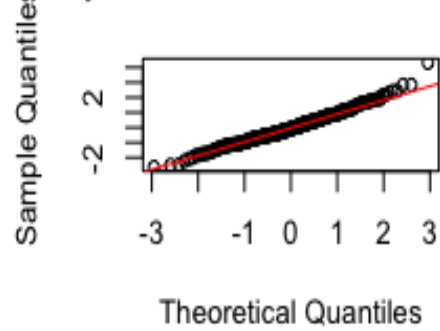
```
##
## Call:
## garch(x = m411residuals, order = c(2, 1), trace = F)
##
## Model:
## GARCH(2,1)
##
## Residuals:
##       Min       1Q    Median       3Q      Max
## -2.63262 -0.65198 -0.08552  0.59544  4.30705
##
## Coefficient(s):
##      Estimate  Std. Error  t value  Pr(>|t|)
## a0 6.314e-07          NA       NA        NA
## a1 1.044e-01          NA       NA        NA
## b1 2.553e-01          NA       NA        NA
## b2 6.343e-01          NA       NA        NA
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 19.448, df = 2, p-value = 5.984e-05
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.42545, df = 1, p-value = 0.5142
##
##
##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.98802, p-value = 0.009805
```

Shapiro-Wilk normality test has p-value fails to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.5142, so it accepts original assumption that the error is uncorrelated.

series plot of standardised re **Histogram of standardised resid**

QQ plot of standardised residu

Standardised residuals

Time

Frequency

res.model

Sample Quantiles

Theoretical Quantiles

Case 8: ARIMA (4, 1 , 1) ⟶ GARCH (2, 2)

```
garchresidual.analysis(a411garch22)

##
## Call:
## garch(x = m411residuals, order = c(2, 2), trace = F)
##
## Model:
## GARCH(2,2)
##
## Residuals:
##      Min       1Q    Median       3Q       Max
## -3.27992 -0.61849 -0.08308  0.58078  4.53043
##
## Coefficient(s):
##      Estimate  Std. Error  t value Pr(>|t|)
## a0 4.259e-06   2.692e-06    1.582  0.11369
## a1 3.219e-01   1.334e-01    2.414  0.01579 *
## a2 5.819e-10   1.488e-01    0.000  1.00000
## b1 1.672e-01   2.665e-01    0.627  0.53051
## b2 4.961e-01   1.881e-01    2.638  0.00835 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##   Jarque Bera Test
##
## data:  Residuals
## X-squared = 41.759, df = 2, p-value = 8.552e-10
##
##
##   Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 1.7795, df = 1, p-value = 0.1822

##   Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.98178, p-value = 0.0004476
```

Shapiro-Wilk normality test has p-value fails to satisfy the normality of the model. And Box-Ljung test has p-value equals to 0.5142, so it accepts original assumption that the error is uncorrelated.

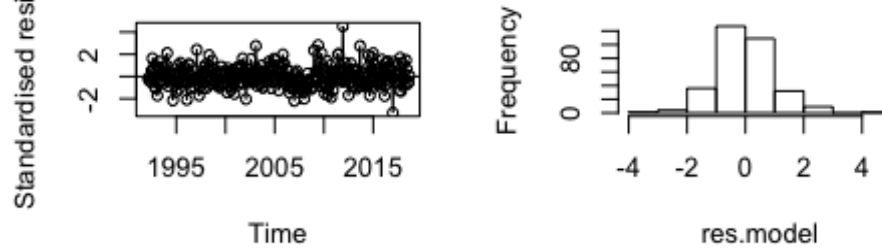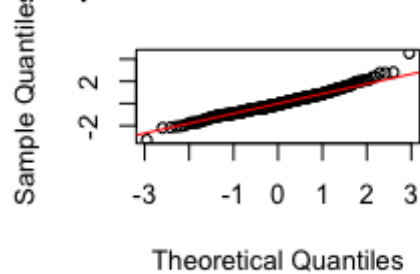## Time series plot of standardised re Histogram of standardised resid

QQ plot of standardised residu

Conclusion

Because we know that ARIMA model is not satisfy the test of normality. Then we think that there may exist some pattern of residuals. Therefore, we perform the GARCH model into the residual of the ARIMA model to get a better model to deal with the non-normal problem. From the result we know that only case 2 ARIMA(1,1,2) to GARCH (1, 1) and case 3 ARIMA(1,1,2) to GRACH (2, 1) are good from the view of the test, so we believe that using the GRACH model can actually improve the performance of the ARIMA, then we conclude, it is better to consider the ARIMA(1,1,2) and GARCH (1, 1) model together to perform the prediction of the model.

Forecast on GARCH model

From the precious experiments, we have already known that GARCH (1,1) and GARCH (2,1) have a good performance. Hence, we use ARMA-GARCH model to do prediction based on the differential sequence. Because here we do not know which kind of combination will achieve the best result, we choose ARMA (1,1), ARMA (1,2), ARMA (2,1), GARCH (1,1) and GARCH (2,1) to fit models. Our purpose is forecasting the next three years' house price which is also 36 months, so it uses 'forecast' function to do that. After getting these groups' result, comparing with the real house price index, we find that the model combined ARMA (1,2) and GARCH (1,1) has the best forecast result. As shown in the forecast graph, the overall foresting capability is good by using ARIMA (1,1,2)- GARCH (1,1) in original dataset.

```
model<-ugarchspec(variance.model = list(model="sGARCH",garchOrder=c(1,
1)),
                  mean.model = list(armaOrder=c(1,2),include.mean=FALS
E),
                  distribution.model = "norm")
modelfit<-ugarchfit(model,data=index_ts2_d1)


forecast_36=ugarchforecast(modelfit,n.ahead=36,data=index_ts2_d1)
forecast36ts = ts(fitted(forecast_36)[,1],start = c(2018,9),frequency =
 12)
fcast = c(index_ts2_d1,forecast36ts)
forecast_36_2db = diffinv(fcast,xi = index_ts2_d1[1:2],differences = 2)
forecast_36_2db = ts(forecast_36_2db,start=1991,frequency = 12)
plot(forecast_36_2db,type='o',ylab= 'House Price Index',
     main="Time Series Plot of House Price Index with the forecasted va
lues")
```

Time Series Plot of House Price Index with the forecasted values
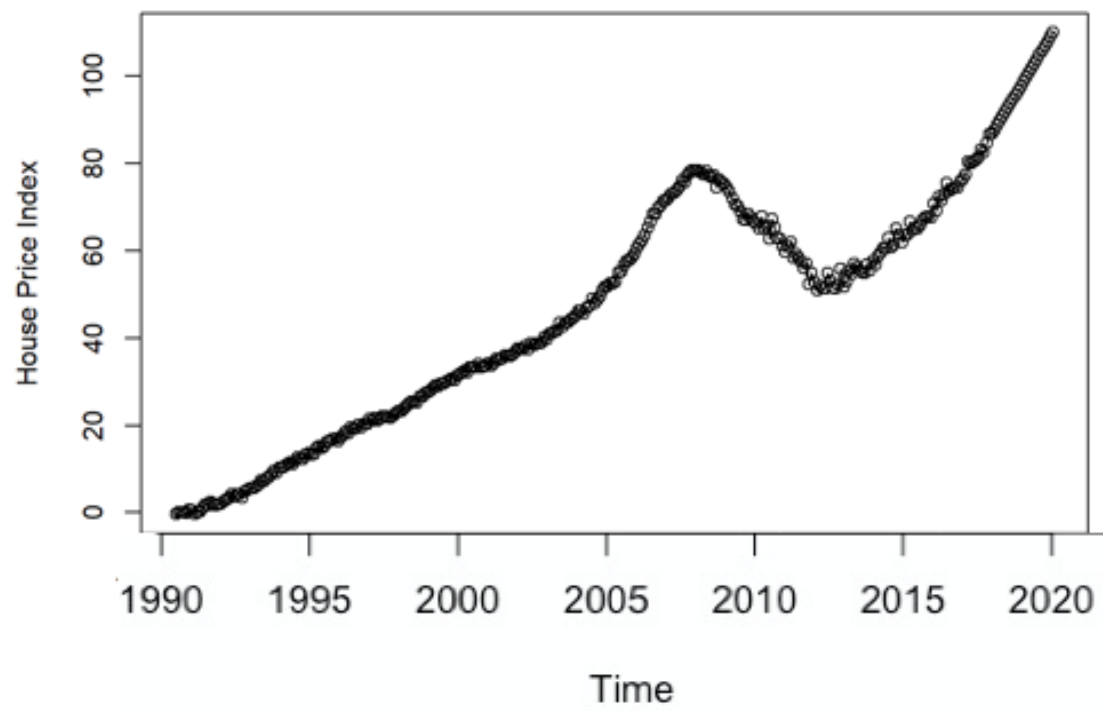
Conclusion

   House price forecast is a popular topic getting more and more people attention to study it because it is of importance to human living condition. What's more, house price forecasting is a real difficult issue due to the price series is changing with so many influential factors that are complex, including seasonal, nature and manmade factors. Also in our simple time series analysis, we used many functions in this project, for instance, ADF test to decide series is stationary or not, 'codefest' to test estimated coefficients and so on. In the process of 'LINE' residual analysis which means linear, independent, normal and equal variance, however, it found that the model had unequal variance. Therefore, we cannot use AR, MA, ARMA, ARIMA model to get satisfied predicted result. In this report, we used GARCH model to do house price forecasting. By comparing different ARIMA-GARCH model prediction, it can conclude that most of these models have the good result. In addition, ARIMA (1,1,2)- GARCH (1,1) has the best forecast accuracy. Therefore, ARIMA (1,1,2)- GARCH (1,1) model issued in this report is a wise method for house price index of east-south central division forecasting.

Appendix

```r
# Function that performs residual analysis
residual.analysis<-function(model, std =TRUE){
library(TSA)
library(FitAR)
if (std ==TRUE){
res.model =rstandard(model)
  }else{
res.model =residuals(model)
  }
par(mfrow=c(2,2))
plot(res.model,type='o',ylab='Standardised residuals',
main="Time series plot of standardised residuals")
abline(h=0)
hist(res.model,main="Histogram of standardised residuals")
qqnorm(res.model,main="QQ plot of standardised residuals")
qqline(res.model, col =2)
plot(res.model,type='l',ylab='Standardised residuals',
main="Residual vs. order plot")
abline(h=0)
print(shapiro.test(res.model))
}




residual.analysis<-function(model, std =TRUE){
if (std ==TRUE){
res.model =rstandard(model)
  }else{
res.model =residuals(model)
  }
  k=0
LBQPlot(res.model, lag.max =length(model$residuals)-1, StartLag = k +1,
 k =0,
SquaredQ =FALSE)
par(mfrow=c(1,1))
}
residual.analysis(m011_hpi)


garchresidual.analysis <-function(model, std =TRUE){
library(TSA)
library(FitAR)
  res.model =residuals(model)
```

```r
shapiro.test(res.model)
print(summary(model))
par(mfrow=c(2,2))
plot(res.model,type='o',ylab='Standardised residuals', main="Time serie
s plot of standardised residuals")
abline(h=0)
hist(res.model,main="Histogram of standardised residuals")
qqnorm(res.model,main="QQ plot of standardised residuals")
qqline(res.model, col =2)
print(summary(model))
print(shapiro.test(res.model))
par(mfrow=c(1,1))
}
```