STAT520 Project Final Report - Exploring Foreign Reserve Scale of China

Jiexin Duan, Min ren, Sirui Wang

Friday, April 29, 2016

Contents

1 Data Description

1.1 Data

- foreign reserve scale (reserve)
- short term debt (SD)
- foreign direct investment (FDI)
- export (EX)
- import (IM)
- exchange rate (exrate)

All of the above economic indicators are based on China and annual data from 1985 to 2015. So the total length of the time series is $n = 31$. For reserve,SD,FDI,EX,IM, the units are 100 million US dollars. And for exrate, the unit is RMB/$.

1.2 Data Source

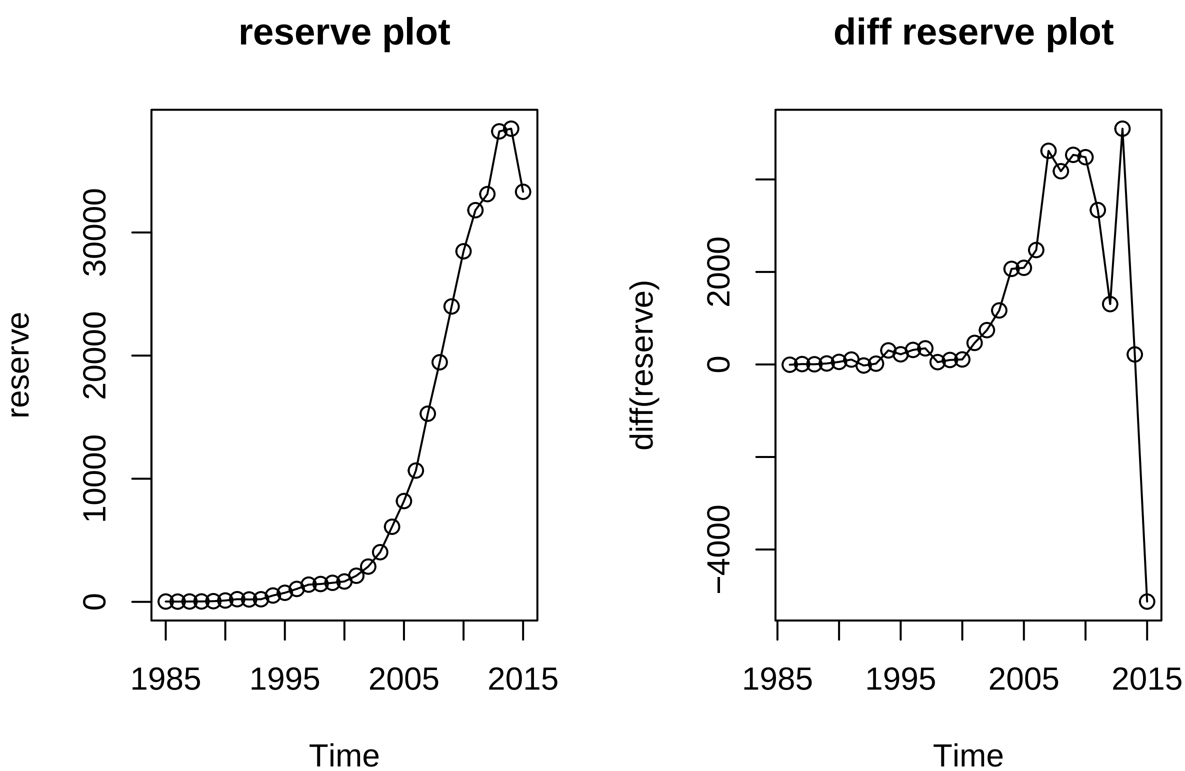
The data in this project come from:

- State of Administration of foreign exchange (http://www.safe.gov.cn/)
- Ministry of Commerce of the People's Republic of China Comprehensive Department (http://zhs.mofcom.gov.cn/tongji.shtml)
- The People's Bank Of China (http://www.pbc.gov.cn/)

2 ARIMA Modeling of Foreign Reserve Scale

In this part, we will find the appropraite ARIMA model to fit the Foreign Reserve Scale (reserve) of China.

2.1 Exploratory Analysis



From the left graph, we find that the reserve has an upward pattern, so we take the first difference and get the right plot. But the right plot still seems to be not very stable. So, we need further analysis.

2.2 Stationary Test

To test the stationarity of the reserve, we use the CADFtest() function (Covariate- ADF test) in R. This this function is more powerful than ordinary adftest() function, as it can consider about the trend term, different lags, and selection criterion for the unit root.

If we don't add the covariate term in the command, CADFtest will be equivalent to ADFtest in essence but the hypothesis set-up is a little different in expression:

H_0 : δ is equal to 0, the series is non-stationary.

H_1 : δ is smaller than 0, the series is stationary.

In the codes, we use take trend option, set the max of lag to be 9 and use BIC criterion to select the optimal lag.

```
##
##  ADF test
##
## data:  reserve
## ADF(8) = 0.56007, p-value = 0.9988
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## 0.2454846
```

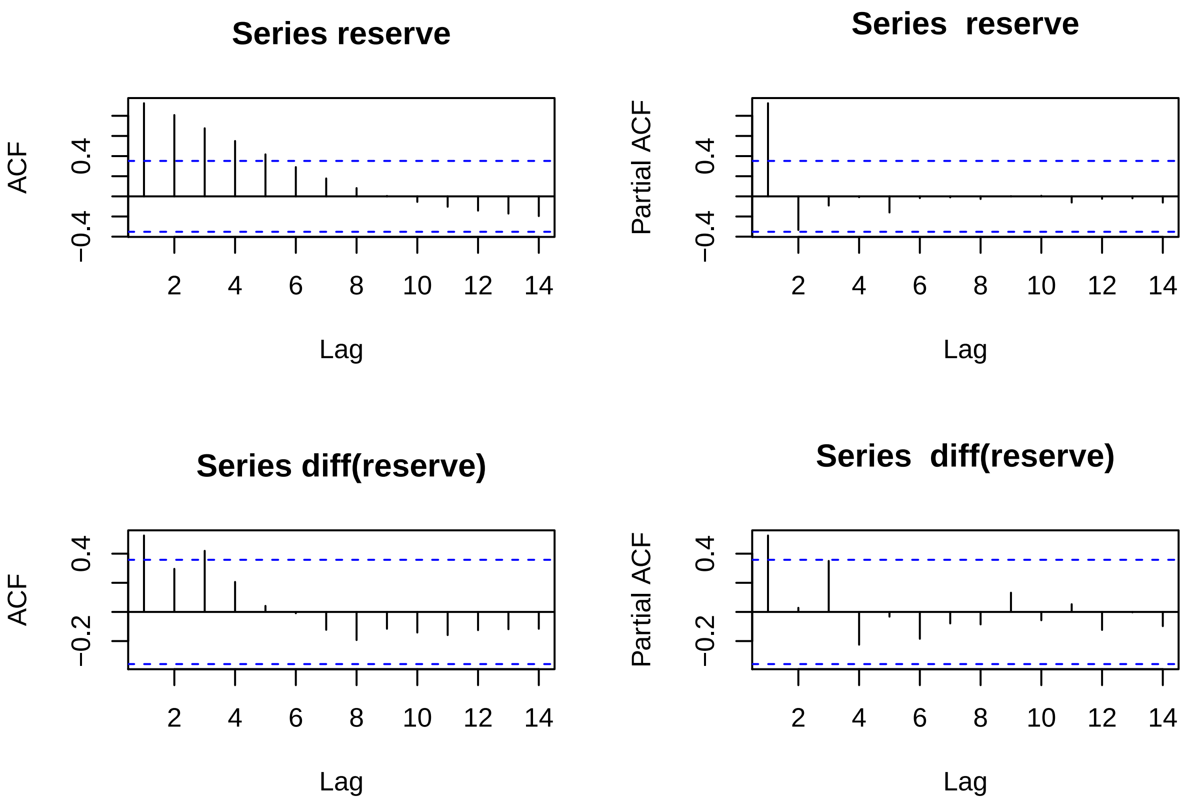
```
##
##  ADF test
##
## data:  diff(reserve)
## ADF(7) = -4.7331, p-value = 0.005853
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## -2.658761
```

From the above results of unit-root test, we can find that: For series of reserve, the p-value=0.9988, the null hypothesis is accepted, it means that the series of reserve is non-stionary.

For the series of first difference of reserve, the p-value=0.005833, the null hypothesis is rejected, it means that the series of first difference of reserve is stationary.

2.3 Autocorrelation Plots

After test the stationarity of the series, we consider about the Autocorrelation plots.



From the previous part we know the series of reserve is non-stationary, and the series of first difference of reserve(diff(reserve)) is stationary. So will focus more on the autocorrelation functions of the first differnce of reserve (diff(reserve)).

From the ACF plot of diff(reserve), we can find that the lag(1) and lag(3) are significant, we use select ARIMA(0,1,3) as a candidate model for series of reserve.

From the PACF plot of diff(reserve), we can find that the lag(1) and is significant, we use select ARIMA(1,1,0) as a candidate model for series of reserve.

```
## AR/MA
##   0 1 2 3 4 5
## 0 x o x o o o
## 1 o o o x o o
## 2 o o o o o o
## 3 o x o o o o
## 4 x x o o o o
## 5 o o o o o o
```

From the EACF plot, we can use ARIMA(1,1,0) or ARIMA(2,1,1) as a candidate model for the series of reserve.

2.4 Model Selection

To select optimal model for series of reserve, we first fit three models of ARIMA(1,1,0). ARIMA(0,1,3) and ARIMA(2,1,1):

```
##
## Call:
```

```
## arima(x = reserve, order = c(1, 1, 0))
##
## Coefficients:
##          ar1
##       0.7721
## s.e.  0.1242
##
## sigma^2 estimated as 2458955:  log likelihood = -263.75,  aic = 529.5
```

```
##
## Call:
## arima(x = reserve, order = c(0, 1, 3))
##
## Coefficients:
##          ma1     ma2     ma3
##       0.7087  0.3919  0.5084
## s.e.  0.2594  0.2920  0.2924
##
## sigma^2 estimated as 2405792:  log likelihood = -263.7,  aic = 533.41
```

```
##
## Call:
## arima(x = reserve, order = c(2, 1, 1))
##
## Coefficients:
##          ar1     ar2     ma1
##       0.2560  0.3121  0.6796
## s.e.  0.4517  0.3796  0.3680
##
## sigma^2 estimated as 2316475:  log likelihood = -262.92,  aic = 531.84
```

2.4.1 Coefficients Comparison

For ARIMA(1,1,0), we can find that at 95% confidence level, the coefficient of ar1 is significant.

For ARIMA(0,1,3), at 95% confidence level, coefficients of ma1 and ma3 are siginificant but the coefficient of ma2 is not significant.

For ARIMA(2,1,1), at 95% confidence level, coefficient of ar1 is significant, but coefficients of ar2 and ma1 are insignificant.

So, based on coffcients comparison, we can conclude that ARIMA(1,1,0) is more appropriate than ARIMA(0,1,3) and ARIMA(2,1,1).

2.4.2 AIC Criteria

From the fitting results, we can find the ARIMA(1,1,0) has the lowest AIC=529.5 than ARIMA(0,1,3)'s 533.41 and and ARIMA(2,1,1)'s 531.84. So based on AIC Criteria, we can conclude that ARIMA(1,1,0) is more appropriate than ARIMA(0,1,3) and ARIMA(2,1,1).

2.5 ARIMA Model (ARIMA(1,1,0))

Based on the model selction conclusion in previous, we use ARIMA(1,1,0) model to estimate and forecast.

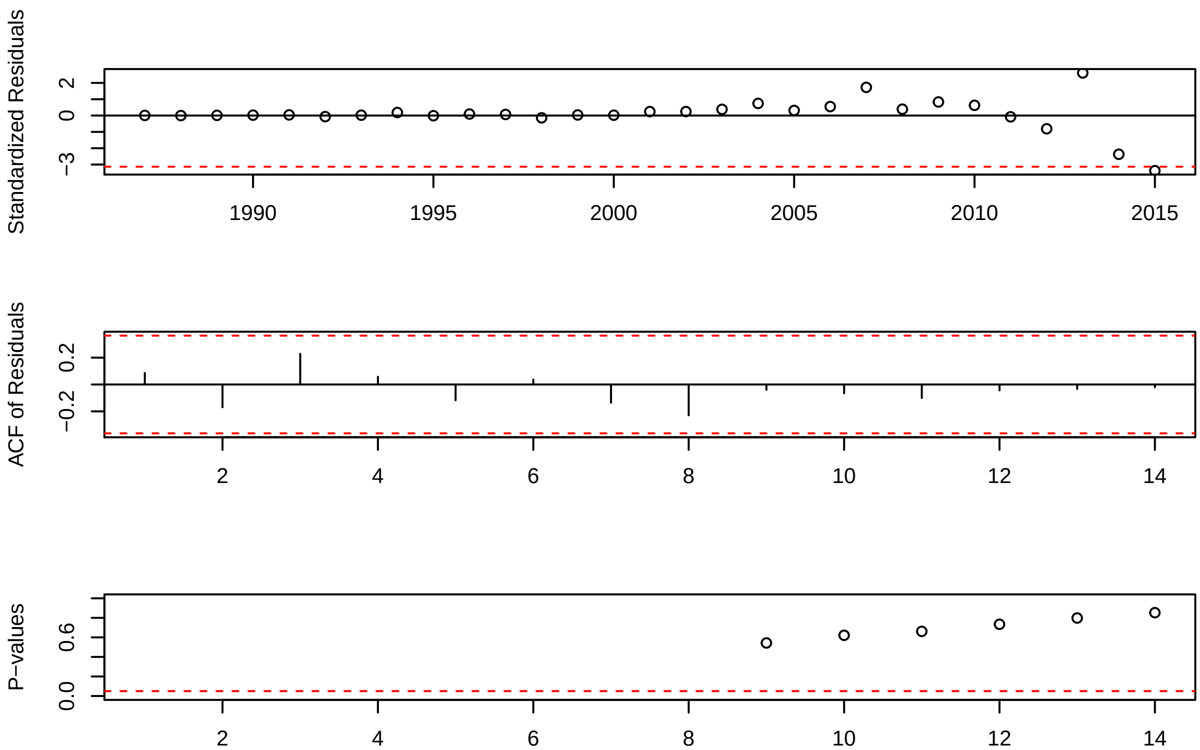
```
Call:
arima(x = reserve, order = c(1, 1, 0))

Coefficients:
         ar1
      0.7721
s.e.  0.1242

sigma^2 estimated as 2458955:  log likelihood = -263.75,  aic = 529.5
```

From the above fitting result of ARIMA(1,1,0), we can get the estimated model for this series:

$$(1 - 0.7721B)(1 - B)Y_t = e_t$$

2.6 Model Diagnostics

After fitting the model of ARIMA(1,1,0), we will do the model diagnostics.



Form the above plots, we can find that:

For standardized residuals, there is only one outlier. That is 2015, this may due to the bad situation of China's economy in 2015.

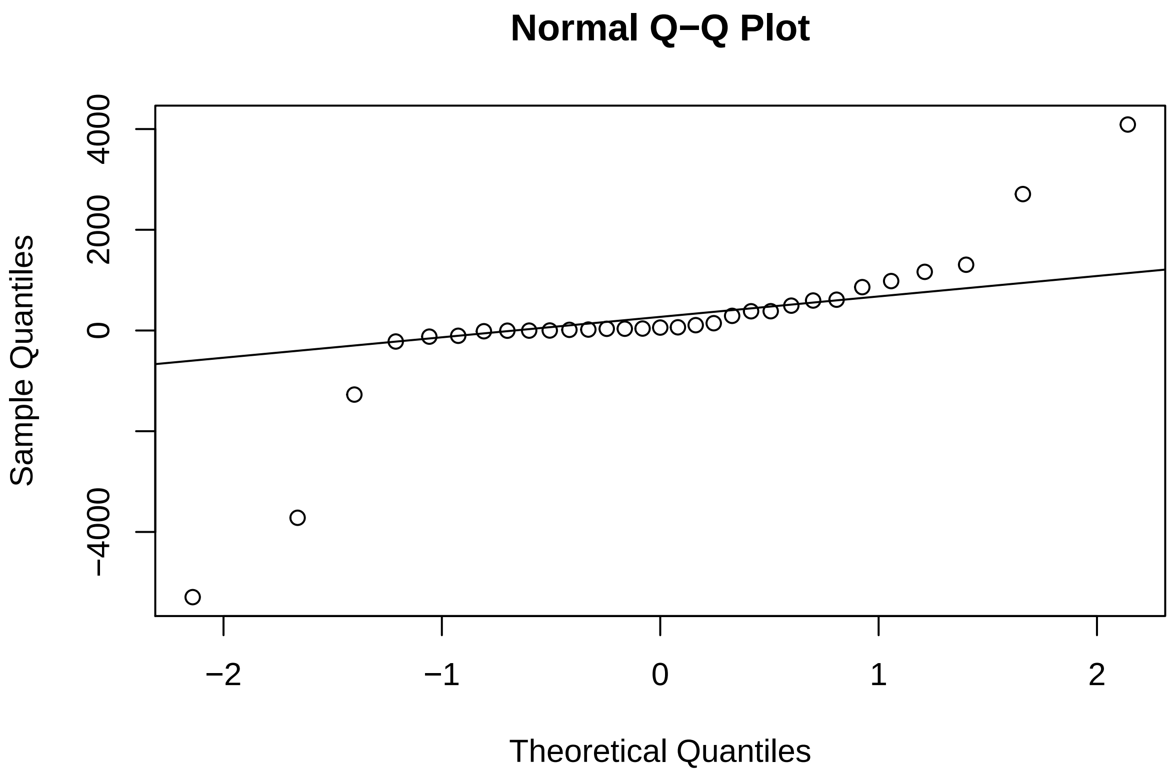
For ACF residuals, all value fall within the CI, it is OK.

For Ljung-Box test, there is no siginificant p-values. So we can accept the null hypothesis: the ARIMA(1,1,0) model is approrpaite for this time series: reserve.

```
##
##  Runs Test
##
```

```
## data:  factor(fitresid > 0)
## Standard Normal = 0.10085, p-value = 0.9197
## alternative hypothesis: two.sided
```

From the runs test, the p-value is 0.0197, so we can accept the null hypothesis: there is no autocorration for the residuals.




```
##
##  Shapiro-Wilk normality test
##
## data:  fitresid
## W = 0.74707, p-value = 6.231e-06
```

From the qq plot, we can find that there are some fat-tails on both side.

And from the Shapiro wilk test, the residuals is not quite normal.

In conclusion, though the residuals in not quite normal, but the other diagnotics all show that ARIMA(1,1,0) is a good model for series of reserve. So, we conclude that ARIMA model fits quite well.

2.7 Forecasting for Next Five Years

After we find that appropriate model to fit the series reserve, we use the ARIMA(1,1,0) model to forecast for the next five years.

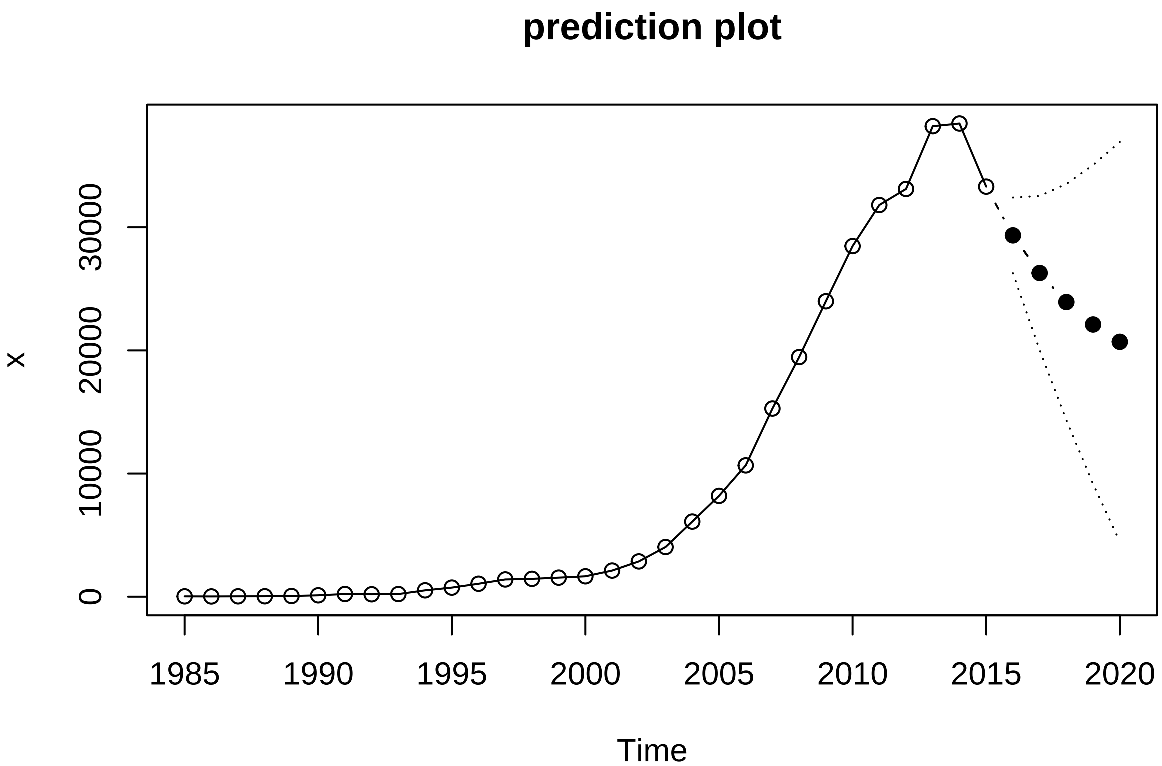
```
## Time Series:
## Start = 2016
## End = 2020
```

```
## Frequency = 1
## [1] 29345.24 26288.86 23928.92 22106.75 20699.79
```

The predicted value for the next five years are listed as follows:

- 2016, predicted reserve is 2.9 trillion dollar
- 2017, predicted reserve is 2.6 trillion dollar
- 2018, predicted reserve is 2.4 trillion dollar
- 2019, predicted reserve is 2.2 trillion dollar
- 2020, predicted reserve is 2.0 trillion dollar

And we also draw the forecasting plot:



For the above forecasting plot, we can find that there will be a downward trend in the next five year. And the confidence inteval widens very quickly as the time goes.

3 Cointegration (Engle-Granger Two Step Method)

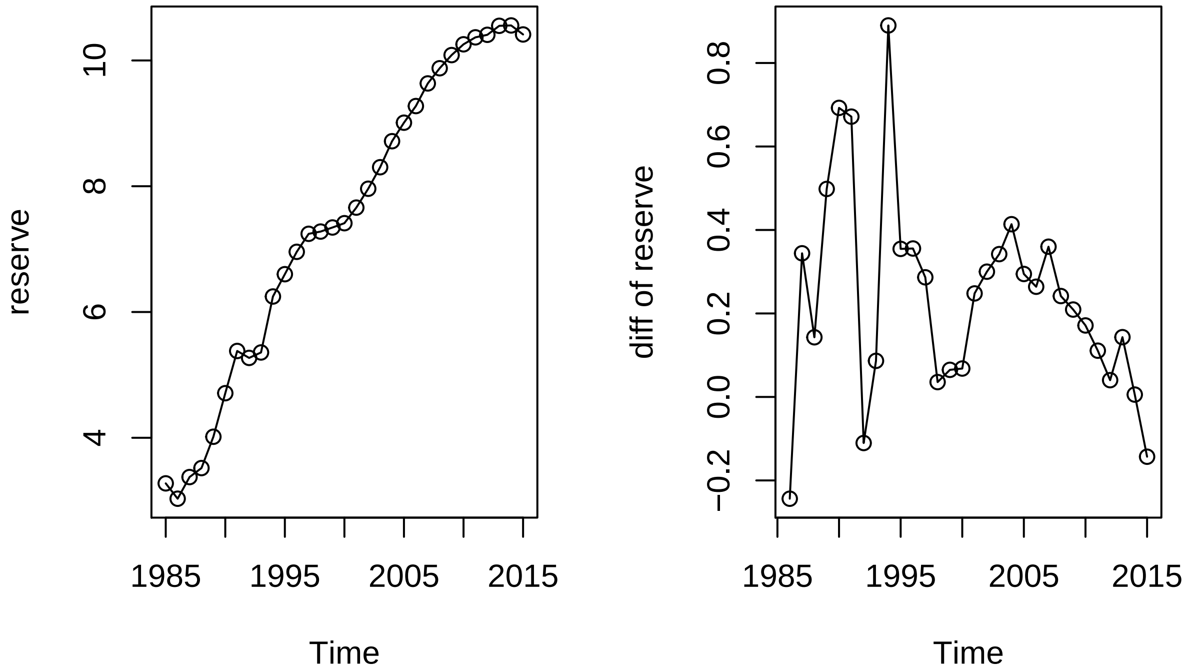
In this part, we will use the Cointegration model (Engle Granger Two-step method) to find the relationshop of Chinese foreign exchange reserve and some other economic indicator such as exports, import, foreign direct investment and so on.

3.1 Integration(1) Checking

Cointegration analysis requires all variables have the same integration order, i.e the order of unit roots.

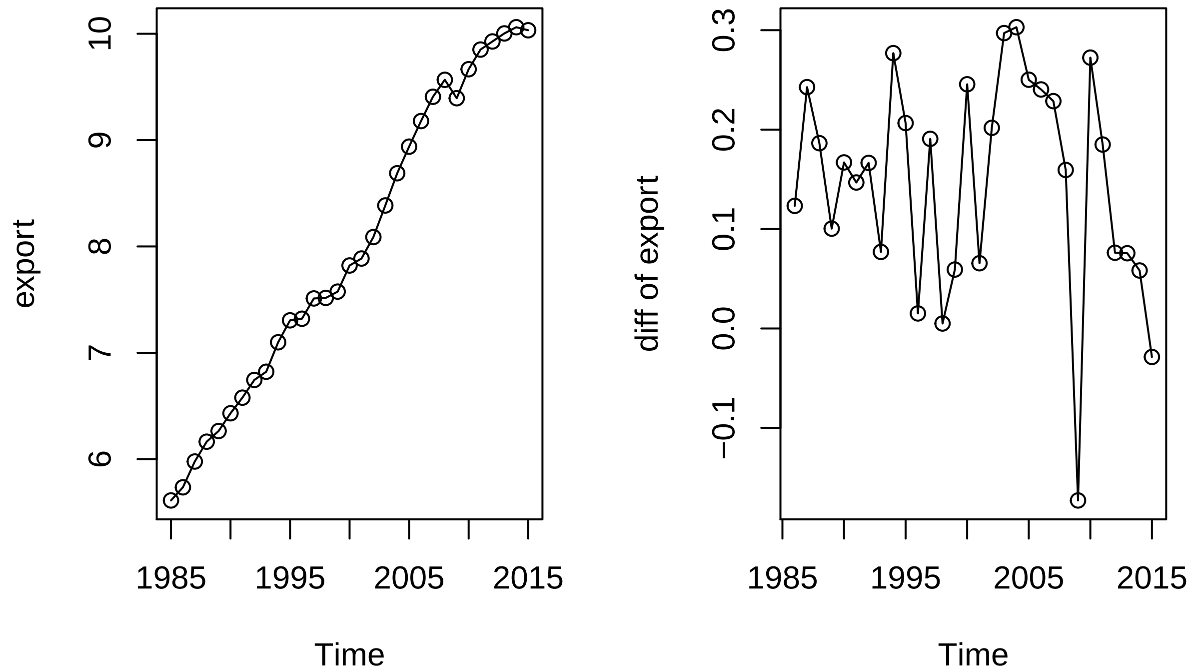
Due to the relatively small sample and the trend effect, we use more powerful Hansen's Covariate-Augmented Dickey Fuller (CADF) test for unit roots.

Since, different economic indicators may have different unit or scales, so we take logarithms for all of the indicators.



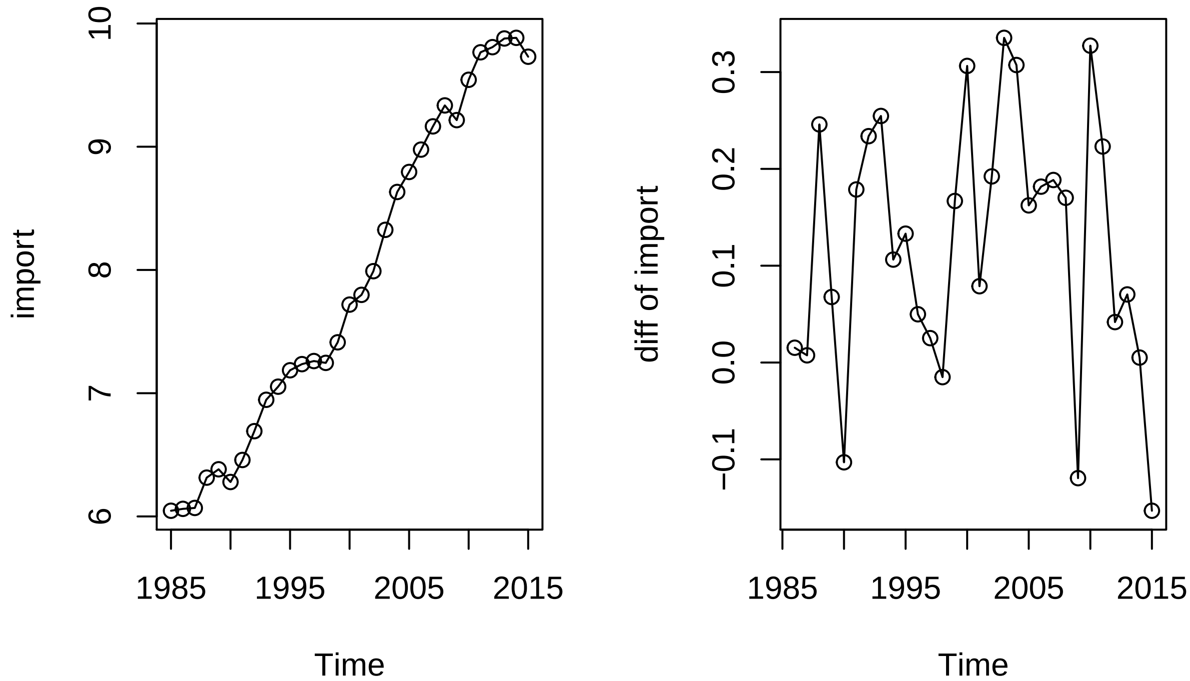

```
##
##  ADF test
##
## data:  reserve
## ADF(8) = 0.56007, p-value = 0.9988
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## 0.2454846
```

```
##
##  ADF test
##
## data:  diff(reserve)
## ADF(7) = -4.7331, p-value = 0.005853
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## -2.658761
```



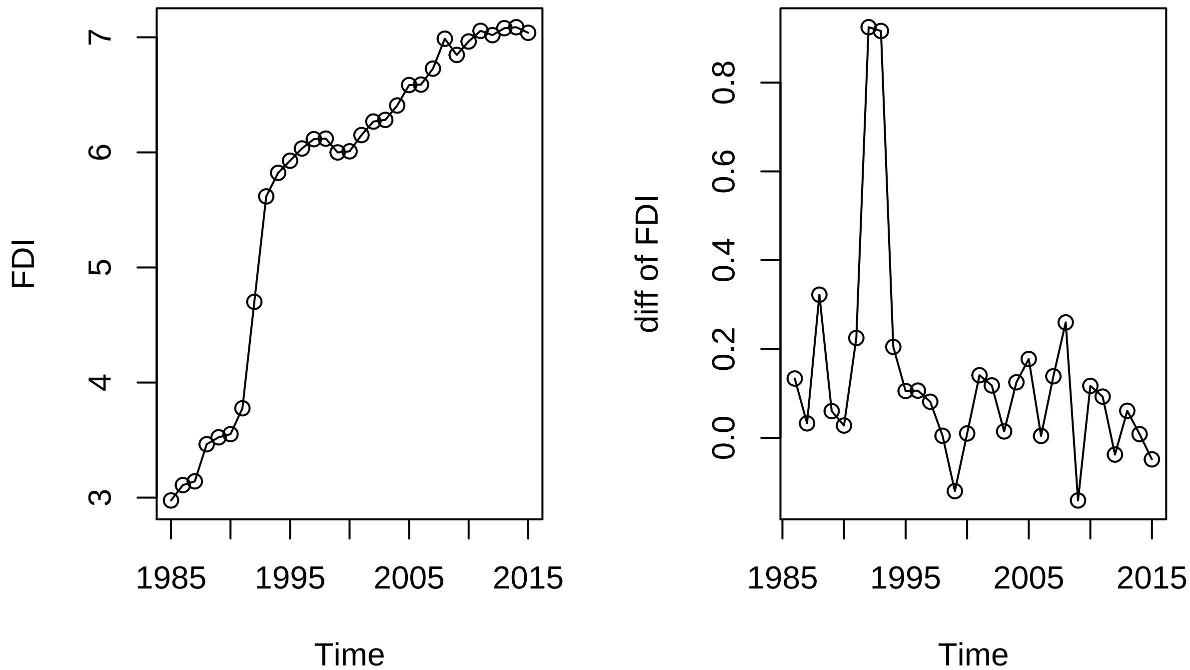

```
##
##  ADF test
##
## data:  ecodatats[, "lnEX"]
## ADF(0) = -0.89637, p-value = 0.9385
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.1516517
```

```
##
##  ADF test
##
## data:  diff(ecodatats[, "lnEX"])
## ADF(0) = -3.8466, p-value = 0.03412
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## -0.920624
```



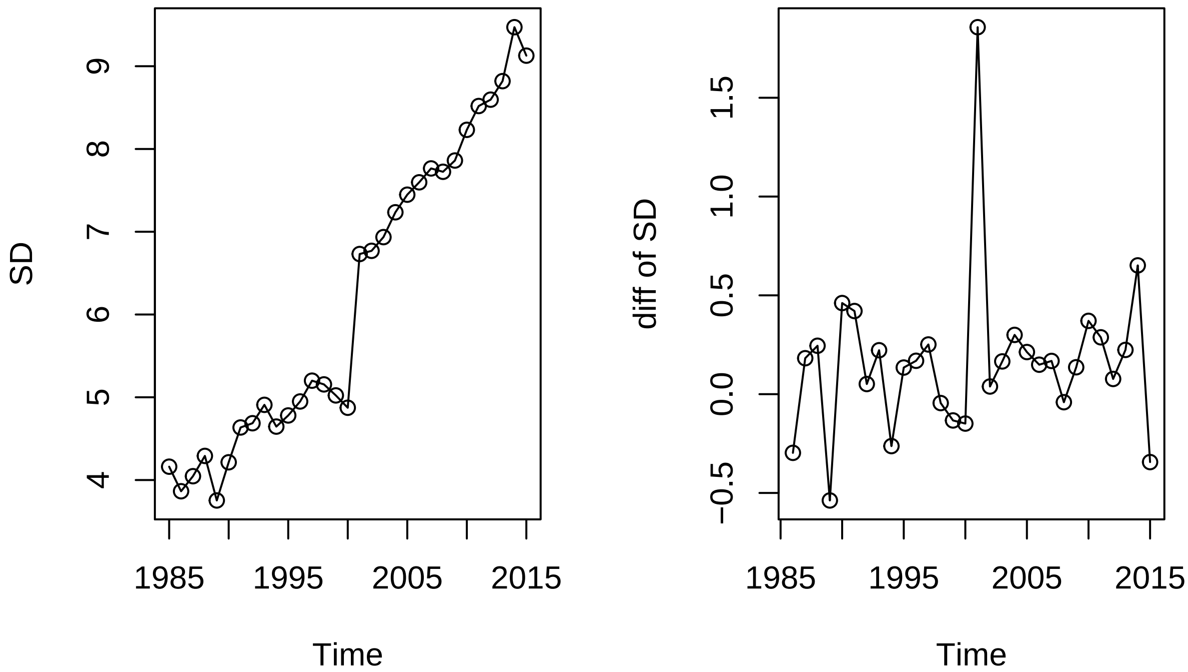

```
##
##  ADF test
##
## data:  ecodatats[, "lnIM"]
## ADF(0) = -0.76943, p-value = 0.9535
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.1369851
```

```
##
##  ADF test
##
## data:  diff(ecodatats[, "lnIM"])
## ADF(0) = -2.0831, p-value = 0.03844
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.3649516
```



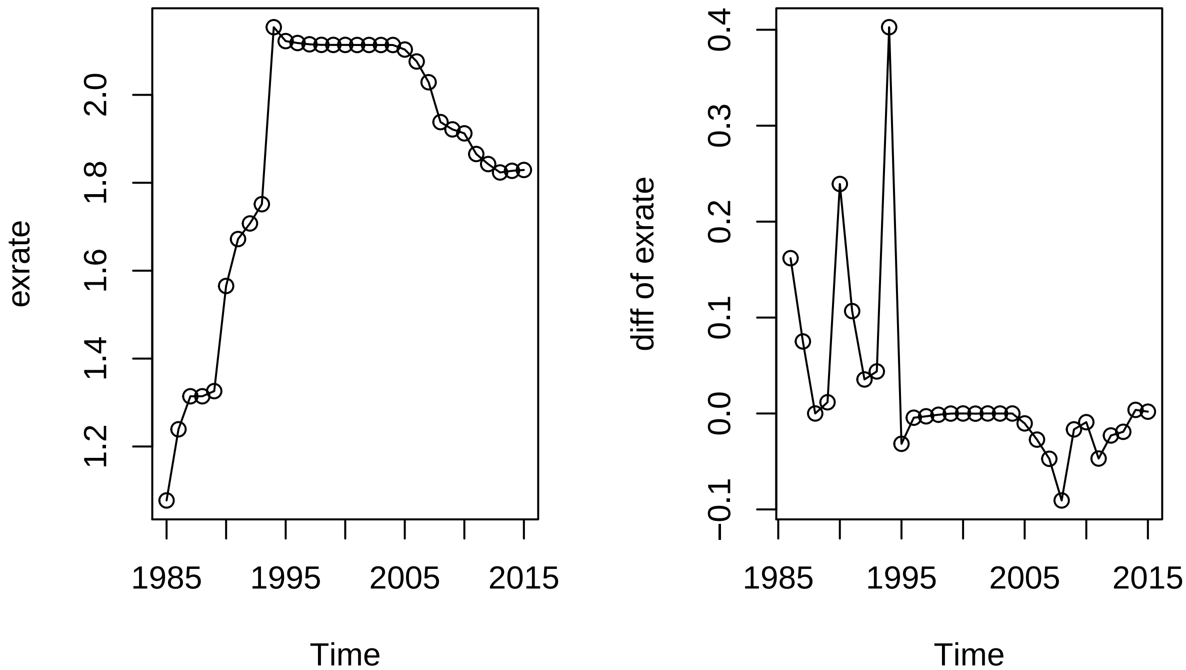

```
##
##  ADF test
##
## data:  ecodatats[, "lnFDI"]
## ADF(1) = -1.7939, p-value = 0.6816
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.1272793
```

```
##
##  ADF test
##
## data:  diff(ecodatats[, "lnFDI"])
## ADF(1) = -3.9752, p-value = 0.02176
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.7951649
```




```
##
##  ADF test
##
## data:  ecodatats[, "lnSD"]
## ADF(0) = -3.2612, p-value = 0.09886
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.6718966
```

```
##
##  ADF test
##
## data:  diff(ecodatats[, "lnSD"])
## ADF(0) = -5.2538, p-value = 0.002026
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## -1.236072
```




```
##
##  ADF test
##
## data:  ecodatats[, "lnexrate"]
## ADF(0) = 0.057747, p-value = 0.6907
## alternative hypothesis: true delta is less than 0
## sample estimates:
##        delta
## 0.0005604721

##
##  ADF test
##
## data:  diff(ecodatats[, "lnexrate"])
## ADF(0) = -15.309, p-value = 1e-04
## alternative hypothesis: true delta is less than 0
## sample estimates:
##     delta
## -1.017937
```

From the above code, after taking the logarithm, the plot of difference of log-value seems more stable than log-value.

And from the unit-roots test, all of the log-value of indictors has unit root of order 1. And after taking the first difference, they become stable.

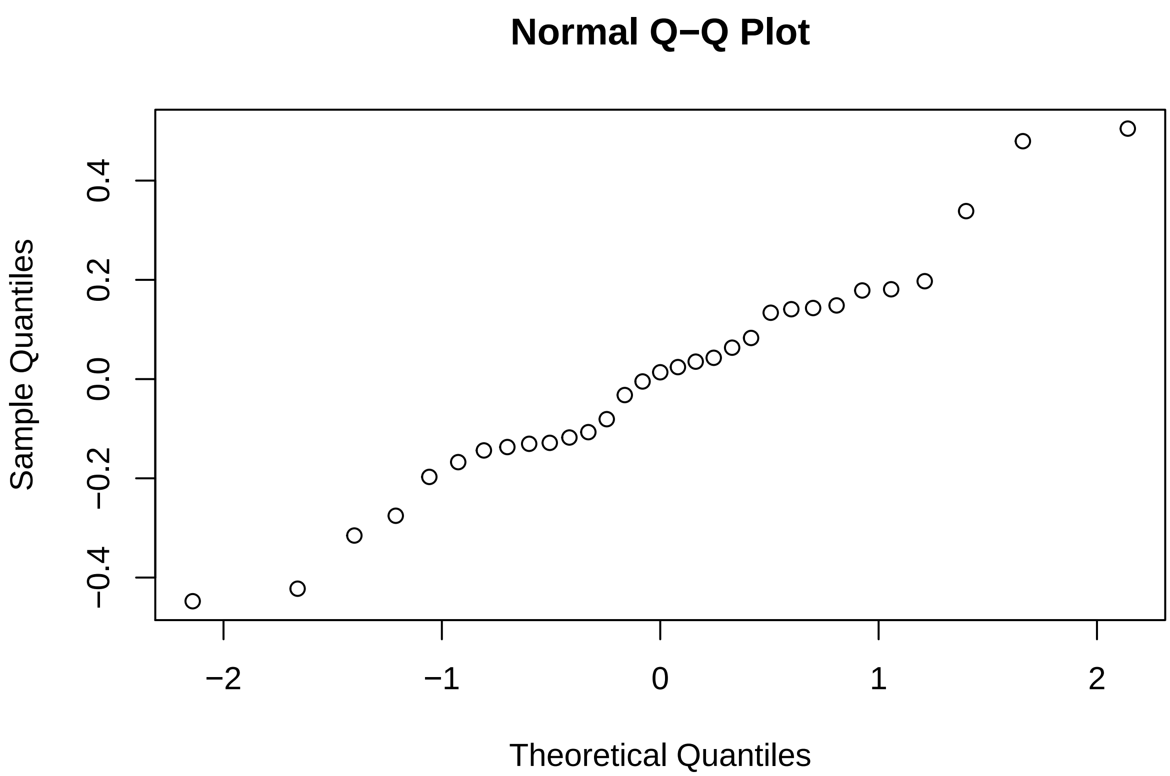
So, in conclusion found all the variables)(log-value) are Integration(1), i.e, has just one unit root of order 1.

3.2 Long-term Model (First Step)

Since all the series has the same order of integration, so we can run the linear regression of lnreserve on the other incators. And we find lnEX, lnIM, lnFDI are signifant for lnreserve in the liner model. So we keep the following regression:

```
##
## Call:
## lm(formula = lnreserve ~ lnEX + lnIM + lnFDI, data = ecodatats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44769 -0.13372  0.01372  0.14202  0.50479
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.91955    0.31067 -15.835 3.44e-15 ***
## lnEX         2.18770    0.39212   5.579 6.46e-06 ***
## lnIM        -0.98155    0.38710  -2.536   0.0173 *
## lnFDI        0.46170    0.08143   5.670 5.07e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2376 on 27 degrees of freedom
## Multiple R-squared:  0.9917, Adjusted R-squared:  0.9908
## F-statistic:  1074 on 3 and 27 DF,  p-value: < 2.2e-16
```

From the above regression result, we can find the all of the coefficients or lnEX, lnIm and lnFDI are significant at 95% confidence interval level.




```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  fit_lm$residuals
## W = 0.97534, p-value = 0.6752
```

```
##
##  ADF test
##
## data:  fit_lm$residuals
## ADF(0) = -3.449, p-value = 0.001469
## alternative hypothesis: true delta is less than 0
## sample estimates:
##      delta
## -0.6507431
```

From the above diagnostic of the residuals for the linear regression, we can find that the residuals is normal and stationary. So we can conclude that lnreserve, lnEx, lnIm and lnFDI has the Cointeration relationshop.

3.3 Short Term Transitory Model (a.k.a Error Correction Mdel,ECM, Second Step)

The we sue the ECM to fit the short-term model of these indicator. We regress the difference of lnreserve on the difference of lnEX, difference of lnIm, difference of lnFDI and the lag(1) of residual from previous model:

```
##
## Call:
## lm(formula = lnreserve.d ~ lnEX.d + lnIM.d + lnFDI.d + res.l1,
##     data = dataFull)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30796 -0.09398 -0.02193  0.08639  0.50596
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.08024    0.05702   1.407 0.171669
## lnEX.d       1.44139    0.41448   3.478 0.001867 **
## lnIM.d      -0.33814    0.37724  -0.896 0.378617
## lnFDI.d     -0.04536    0.15371  -0.295 0.770373
## res.l1      -0.67198    0.15586  -4.311 0.000222 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1767 on 25 degrees of freedom
## Multiple R-squared:  0.5534, Adjusted R-squared:  0.482
## F-statistic: 7.746 on 4 and 25 DF,  p-value: 0.0003328
```

In this way, we can get the short-term ECM model for these for variables. From the above result, we can find the the lnEX.d and residual.l1 are significant and other terms are not significant.

3.4 Final Cointegration Model

In this section, we list the above longrun and short-term model as follows:

- longrun model

$$lnReserve_t = -4.9 + 2.19 * lnEX_t - 0.98 * lnIM_t + 0.46 * lnFDI_t$$

- short term transitory model

$$\nabla lnReserve_t = 0.08 - 0.67 * \epsilon_{t-1} + 1.4 * \nabla lnEX_t - 0.3 * \nabla lnIM_t - 0.05 * \nabla lnFDI_t$$

From the longrun model, we can fint the lnEX and lnFDI has positive effects on lnreserve, and lnIM has negative effect on lnreserve. THis relatioship is also correspondent to the real economic situation: increase in export and FDI will increase the reserve while the increase in import will decrease the reserve.

From the short-term model, most coefficients have the same signal except for FDI. THis may due to the unstable propery of the relations of economic indicators in the short term.

4 Appendix A: All Codes

```
require(urca)
require(TSA)
require(CADFtest)
ecodata <- read.csv("I:/data/ecodata.csv")
ecodata<- ecodata[,c("reserve","EX","IM","FDI","SD","exrate")]
ecodata<- apply(ecodata,2,log)
ecodatats<- ts(ecodata,start = 1985,end = 2015,frequency = 1)
reserve<- ecodatats[,"reserve"]
reserve<- exp(reserve)
colnames(ecodatats)<- c("lnreserve","lnEX", "lnIM","lnFDI", "lnSD", "lnexrate")


oldpar=par
par(mfrow=c(1,2))
plot(reserve,type="o",main="reserve plot")
plot(diff(reserve),type="o",main=" diff reserve plot")
par=oldpar


CADFtest(reserve,type = "trend",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(reserve),type = "trend",max.lag.y = 8,criterion = "BIC")


oldpar=par
par(mfrow=c(2,2))
acf(reserve)
pacf(reserve)
acf(diff(reserve))
pacf(diff(reserve))
par=oldpar


eacf(diff(reserve),ar.max=5,ma.max=5)
```

```r
fit_reserve_1<- arima(reserve,order = c(1,1,0))
fit_reserve_1
fit_reserve_2<- arima(reserve,order = c(0,1,3))
fit_reserve_2
fit_reserve_3<- arima(reserve,order = c(2,1,1))
fit_reserve_3


# arima model arima(1,1,0)
fit_reserve<- arima(reserve,order = c(1,1,0))
fit_reserve


oldpar=par
tsdiag(fit_reserve)
par=oldpar


fitresid<- resid(fit_reserve)
runs.test(factor(fitresid>0))


par(mfrow=c(1,1))
qqnorm(fitresid)
qqline(fitresid)
shapiro.test(fitresid)


# predictions
predict(fit_reserve,n.ahead=5)$pred


plot(fit_reserve,n.ahead=5,pch=19,main="prediction plot")


# reserve
par(mfrow=c(1,2))
plot(ecodatats[,1])
plot(diff(ecodatats[,1]))

CADFtest(reserve,type = "trend",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(reserve),type = "trend",max.lag.y = 8,criterion = "BIC")

# EX
par(mfrow=c(1,2))
plot(ecodatats[,2])
plot(diff(ecodatats[,2]))
CADFtest(ecodatats[,"lnEX"],type = "trend",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(ecodatats[,"lnEX"]),type = "trend",max.lag.y = 8,criterion = "BIC")

# IM
par(mfrow=c(1,2))
plot(ecodatats[,3])
```

```
plot(diff(ecodatats[,3]))
CADFtest(ecodatats[,"lnIM"],type = "trend",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(ecodatats[,"lnIM"]),type = "none",max.lag.y = 8,criterion = "BIC")

# FDI
par(mfrow=c(1,2))
plot(ecodatats[,4])
plot(diff(ecodatats[,4]))
CADFtest(ecodatats[,"lnFDI"],type = "trend")
CADFtest(diff(ecodatats[,"lnFDI"]),type = "trend")

# SD
par(mfrow=c(1,2))
plot(ecodatats[,5])
plot(diff(ecodatats[,5]))
CADFtest(ecodatats[,"lnSD"],type = "trend",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(ecodatats[,"lnSD"]),type = "trend",max.lag.y = 8,criterion = "BIC")

# exrate
par(mfrow=c(1,2))
plot(ecodatats[,6])
plot(diff(ecodatats[,6]))

CADFtest(ecodatats[,"lnexrate"],type = "none",max.lag.y = 8,criterion = "BIC")
CADFtest(diff(ecodatats[,"lnexrate"]),type = "none",max.lag.y = 8,criterion = "BIC")

par(mfrow=c(1,1))
fit_lm<- lm(lnreserve ~lnEX+lnIM+lnFDI,data = ecodatats)
summary(fit_lm)


qqnorm(fit_lm$residuals)
qqline(fit_lm$residuals)
shapiro.test(fit_lm$residuals)
CADFtest(fit_lm$residuals,type = "none",max.lag.y = 8,criterion = "BIC")


dataFull<- data.frame(apply(ecodatats,2,diff),res=fit_lm$residuals[-31])
colnames(dataFull)<- c("lnreserve.d","lnEX.d","lnIM.d","lnFDI.d","lnSD.d","lnexrate.d","res.l1")
fit_ecm<- lm(lnreserve.d ~lnEX.d+lnIM.d+lnFDI.d+res.l1,data = dataFull)
summary(fit_ecm)
```

5 Appendix B: Statement of Contribution

1. Jiexin Duan: Did the statistical analysis and wrote the R code together, wrote the report, present the Cointegration part in the presentation.

2. Min Ren: Did the statistical analysis and wrote the R code together, wrote the presentation slides, present the ARIMA part in the presentation.

3. Sirui Wang: Collect the economic data, did the statisical analysis and wrote the R codes together.