Forecasting Project 1

Yuhao Zhao,Yz3085

November 10, 2015

Part A. Introduction

Long-term investors, such as sovereign funds, have more capacity to invest in private assets and have gradually included private assets in their asset allocation. However, the absence of good quality information on the mark-to-market prices, such as lack of trades or confidentiality, on private assets has led the long-term investors to use appraisals for various investment decisions. In this project, I will model the major characteristics of real estate appraisal returns using the techniques we discussed in class.

Part B. Data for analysis

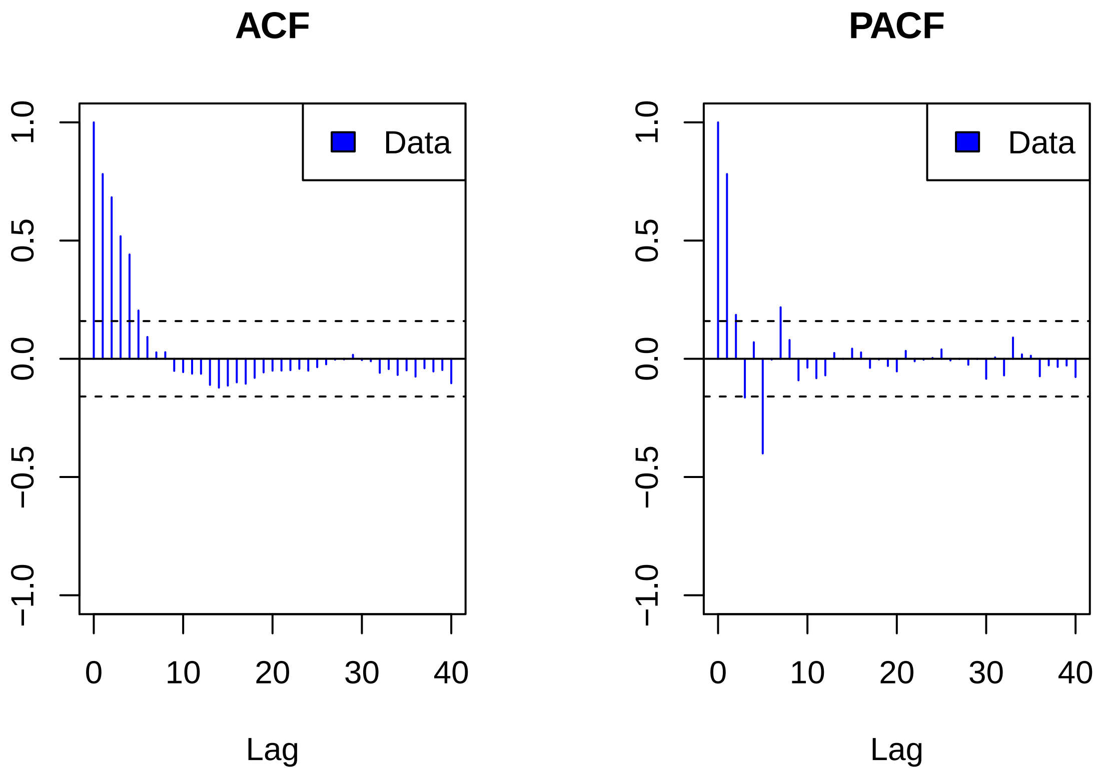
data source: *NCREIF Property Index Returns*

> "The NCREIF Property Index is a quarterly time series composite total rate of return measure of investment performance of a very large pool of individual commercial real estate properties acquired in the private market for investment purposes only. All properties in the NPI have been acquired, at least in part, on behalf of tax-exempt institutional investors - the great majority being pension funds. As such, all properties are held in a fiduciary environment.
> **citted from** *http://www.ncreif.org/property-index-returns.aspx*"

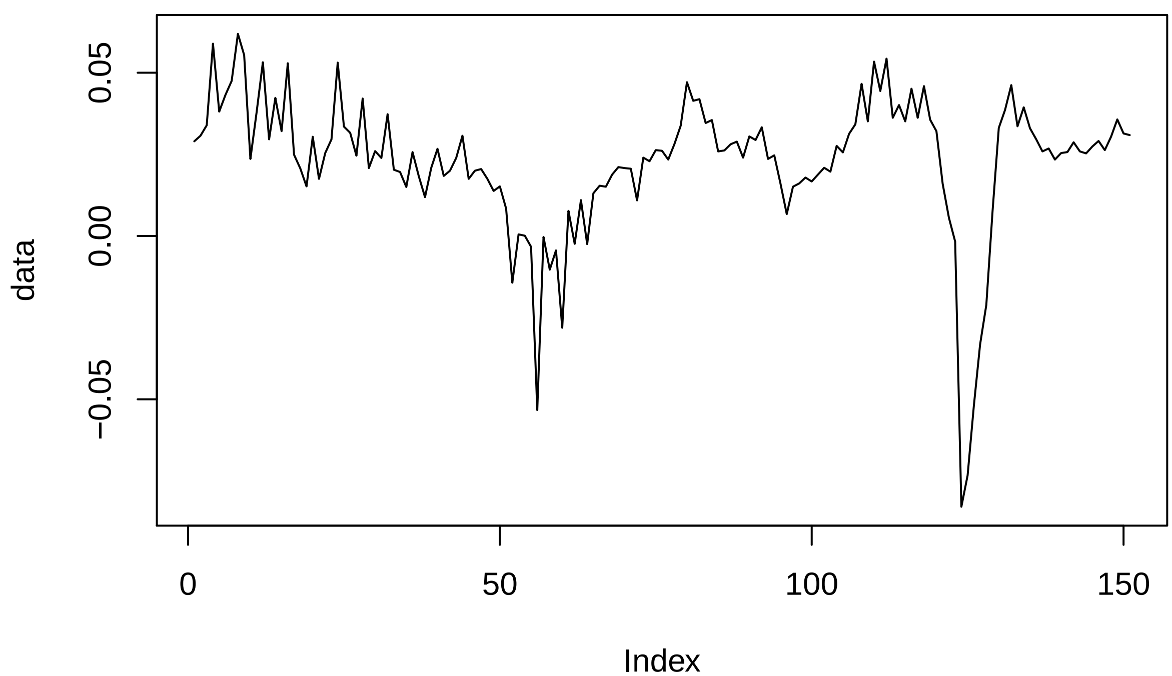
The data is acquired from NCREIF websiet via *http://www.ncreif.org/property-index-returns.aspx*. It contains the quarterly real estate appraisal returns time series from 1978 Q1 to 2015 Q3.

Part C. Data input and plot

```
setwd("/Users/cryan/Desktop/Assignment/STAT-2302/project")
library('itsmr')
library('forecast')
NPI <- read.csv("NPI.csv",fill = TRUE,header = TRUE)[,2:5]
data = c()
for (i in 1:length(NPI[,1])){
  data = c(data,as.numeric(NPI[i,]))
}
data =data[-length(data)] #the last elemet is NA corresponding to the Q4 of 2015
plota(data)
```




```r
plot(data,type = 'l')
```



The data plot doesn't show strong trend, and the retern is negative around 2008. Therefore, it not necessary or possible to take log to the data. In addition, the data doesn't show clear seasonal patterns.

Part D. Model Selection

```r
model1 = auto.arima(data,trace = T,d =0,ic = 'aicc',stepwise = T,allowmean=T,
                    allowdrift = F,seasonal = F,max.p =5,max.q=5)
```

```
##
##  ARIMA(2,0,2) with non-zero mean : -878.7927
##  ARIMA(0,0,0) with non-zero mean : -728.4797
##  ARIMA(1,0,0) with non-zero mean : -868.3238
##  ARIMA(0,0,1) with non-zero mean : -795.192
##  ARIMA(0,0,0) with zero mean     : -616.0297
##  ARIMA(1,0,2) with non-zero mean : -870.7688
##  ARIMA(3,0,2) with non-zero mean : -873.156
##  ARIMA(2,0,1) with non-zero mean : -880.0141
##  ARIMA(2,0,1) with zero mean     : -875.1583
##  ARIMA(1,0,1) with non-zero mean : -869.9902
##  ARIMA(3,0,1) with non-zero mean : -875.5481
##  ARIMA(2,0,0) with non-zero mean : -870.6764
##
##  Best model: ARIMA(2,0,1) with non-zero mean
```

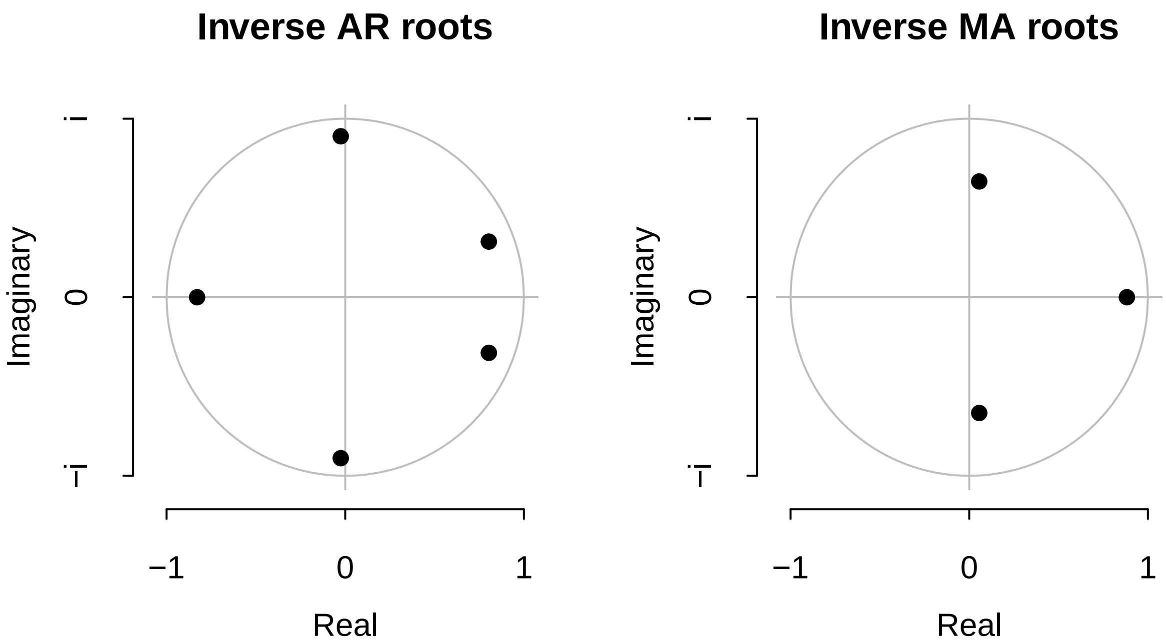
```r
model2 = auto.arima(data,trace = T,d =1,ic = 'aicc',stepwise = T,allowmean=T,
                    allowdrift = F,seasonal = F,max.p =5,max.q=5)
```

```
##
##  ARIMA(2,1,2)                    : -860.3413
##  ARIMA(0,1,0)                    : -841.6408
##  ARIMA(1,1,0)                    : -850.5195
##  ARIMA(0,1,1)                    : -849.6948
##  ARIMA(1,1,2)                    : -855.5401
##  ARIMA(3,1,2)                    : -863.4047
##  ARIMA(3,1,1)                    : -859.3243
##  ARIMA(3,1,3)                    : -858.881
##  ARIMA(2,1,1)                    : -853.1747
##  ARIMA(4,1,3)                    : -874.2867
##  ARIMA(5,1,3)                    : -882.9002
##  ARIMA(5,1,2)                    : -882.4005
##  ARIMA(5,1,4)                    : -881.5309
##  ARIMA(4,1,2)                    : -873.6499
```

```
## Warning in auto.arima(data, trace = T, d = 1, ic = "aicc", stepwise = T, :
## Unable to fit final model using maximum likelihood. AIC value approximated
```

```
##
##
##  Best model: ARIMA(5,1,3)
```

```r
plot.Arima(model2);model = model2
```




```r
model$coef
```

```
##         ar1         ar2         ar3         ar4         ar5         ma1
##  0.72845600 -0.18462933  0.04647569  0.44888090 -0.50122135 -0.99426414
##         ma2         ma3
##  0.52241882 -0.37384726
```

To fit an ARIMA model to the NPI data, I tried different p from 0 to 5, q from 0 to 5 and d from 0 to 1 for model with constant and without constant respectively. The best model with lowest AICc correspond to ARIMA(5,1,3) without constant. The characteristic roots of the model shows that it is invertible.

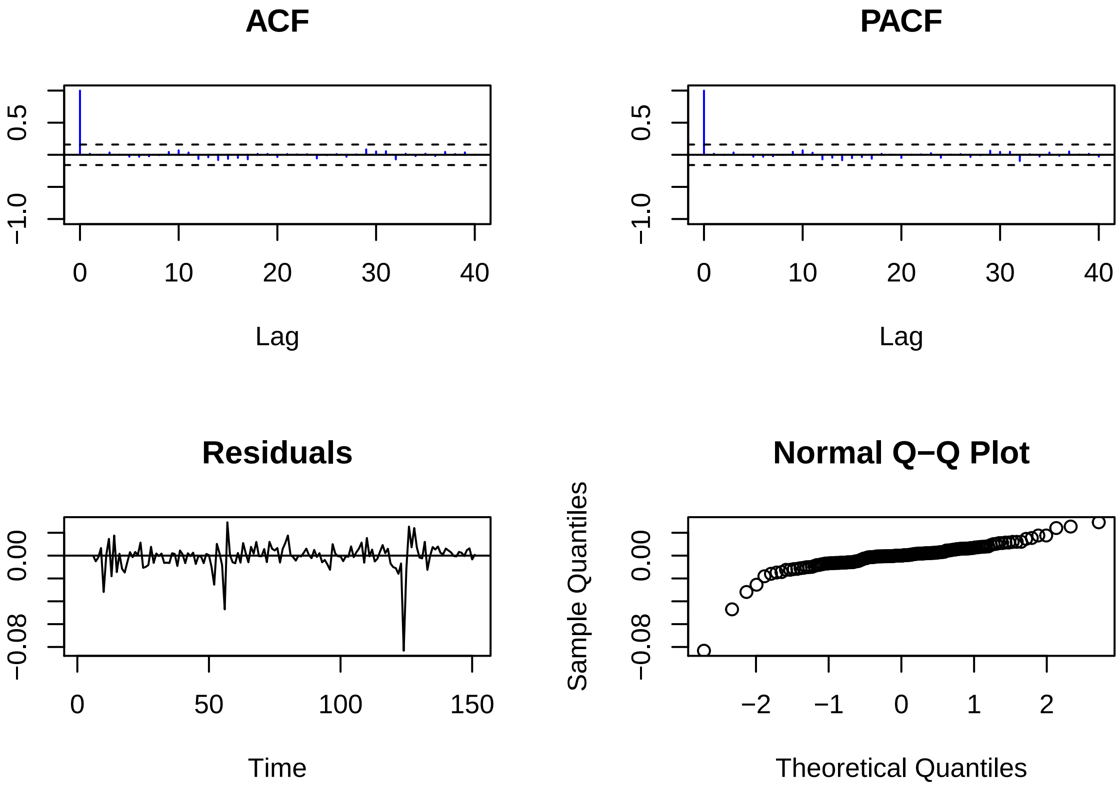
The complete model is:

$$X_t = 0.7285X_{t-1} - 0.1846X_{t-2} + 0.0465X_{t-3} + 0.4489X_{t-4} - 0.5012X_{t-5} + \epsilon_t - 0.9943\epsilon_{t-1} + 0.5224\epsilon_{t-2} - 0.3738\epsilon_{t-3}$$

Part E. Model Dignostic

```r
test(model$residuals)
```

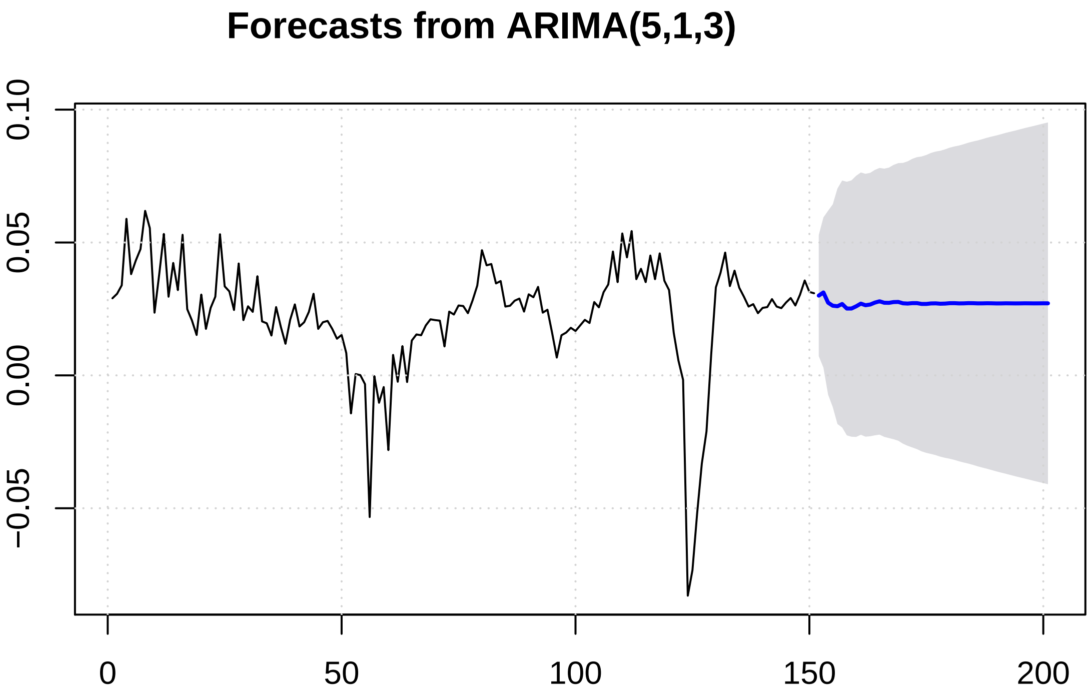
```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)       6.2    0.9986
## McLeod-Li Q                Q ~ chisq(20)      4.17    0.9999
## Turning points T   (T-99.3)/5.1 ~ N(0,1)        99    0.9484
## Diff signs S         (S-75)/3.6 ~ N(0,1)        65     0.005 *
## Rank P         (P-5662.5)/310.8 ~ N(0,1)      6292    0.0428 *
```



The Ljung-Box statistic for the residuals is 6.2 with p-value 0.9986>0.05. We do not reject the Null hypothesis that the Residuas are iid noise. This means the model is adequate. The ACF and PACF seems to be good since there is no lag outside the bar. The residual plot shows some outlier around time 125.In addition, the normal Q-Q plot shows some issues at the tails.

Part F.Forecast and Plot

```
fcast = forecast(model,h =50,level =95)
plot(fcast)
grid()
```



The forecast seems to be appropriate in the near future, but not the far future. The forecast converge to the mean of the series which is not reasonable from the data. The 0.95 forecast interval is not very good. it didn't even cover the variance of past returns.