# Regression Analysis on Unemployment Rate

## Data

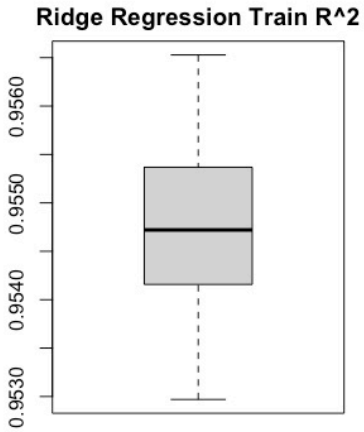
1. data: USA national economic monthly data. xls (06/1960-01/2020)

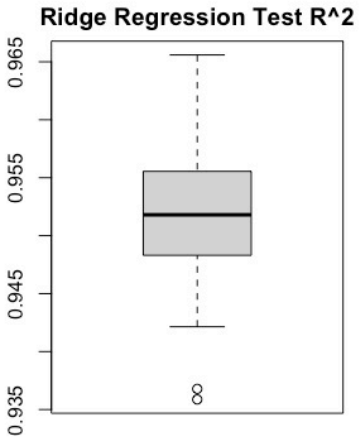
- Observation : n=721
- Variables: p=42 , numerical

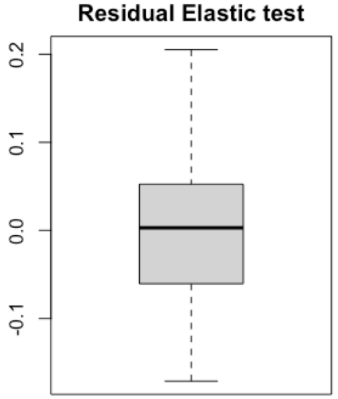
```
df<-read_excel('USA_national_economic_data.xls')
df<-as.data.frame.data.frame(df)

str(df)

## 'data.frame':    721 obs. of  42 variables:
##  $ DATE            : POSIXct, format: "1960-01-01" "1960-02-01" ...
##  $ AAA             : num  4.61 4.56 4.49 4.45 4.46 4.45 4.41 4.28 4.25 4.3
...
##  $ AWHMAN          : num  40.6 40.3 40 40 40.1 39.9 39.9 39.7 39.4 39.7
...
##  $ BAA             : num  5.34 5.34 5.25 5.2 5.28 5.26 5.22 5.08 5.01 5.11
...
##  $ CIVPART         : num  59.1 59.1 58.5 59.5 59.5 59.7 59.5 59.5 59.7
59.4 ...
##  $ CPIAUCSL        : num  29.4 29.4 29.4 29.5 29.6 ...
```

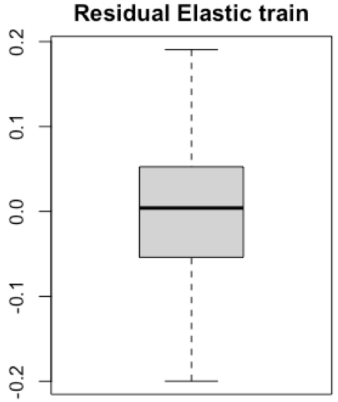
- 2. data source: https://fred.stlouisfed.org (FRED Economic Data)

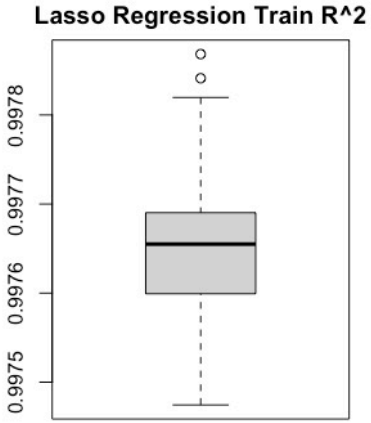
# Boxplots of $R^2_{test}$ & $R^2_{train}$.

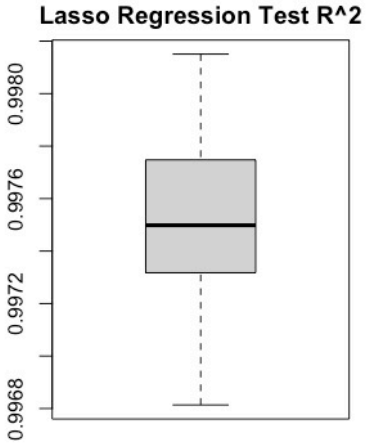


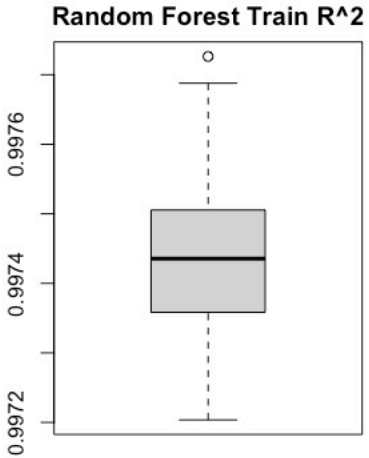



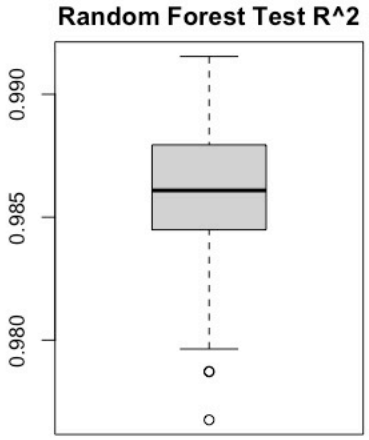



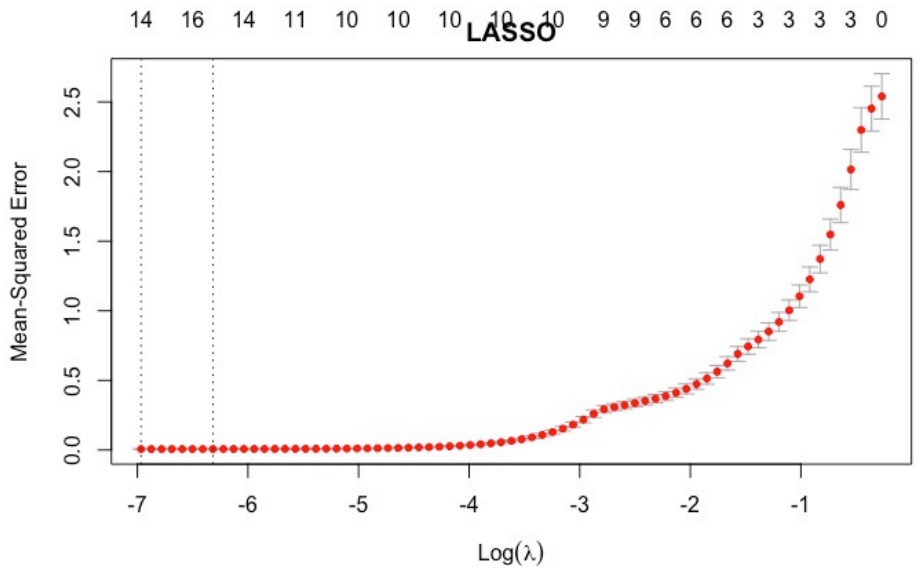



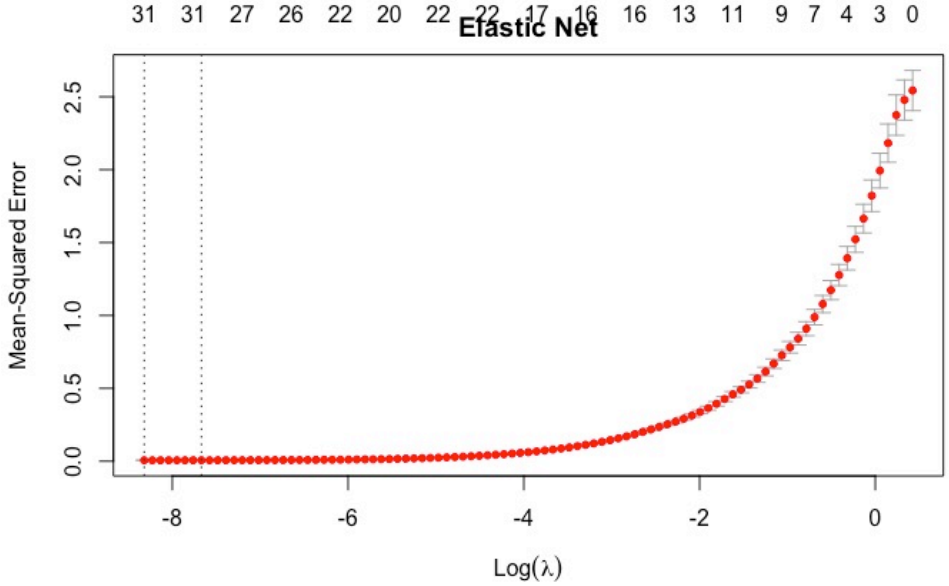



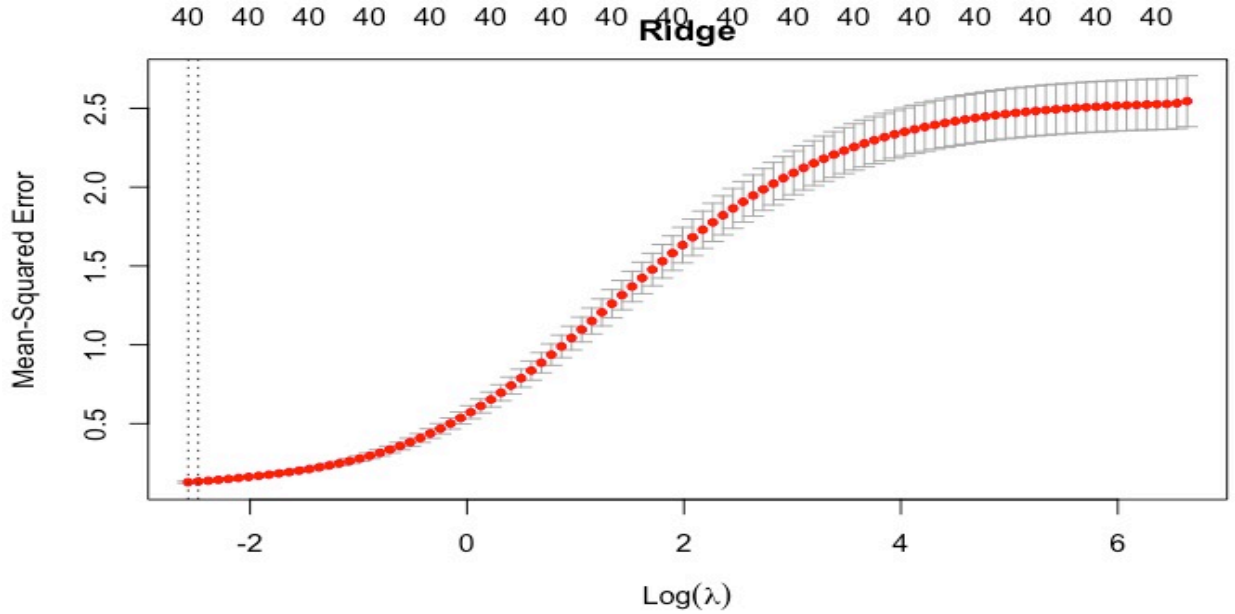



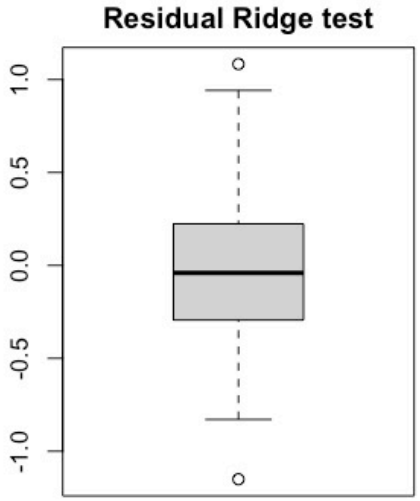
# 10-fold CV curves

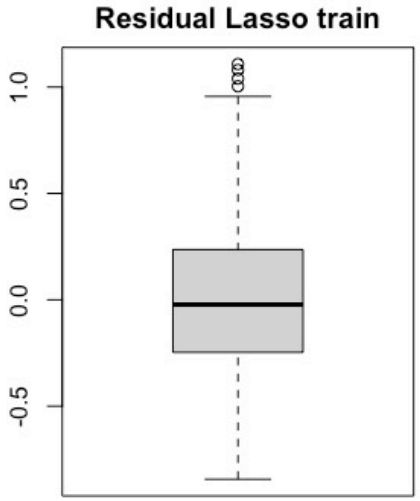


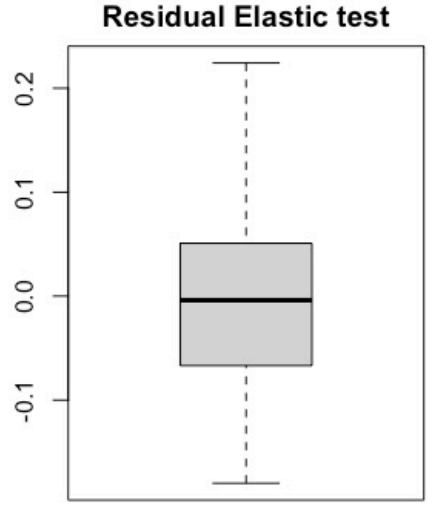



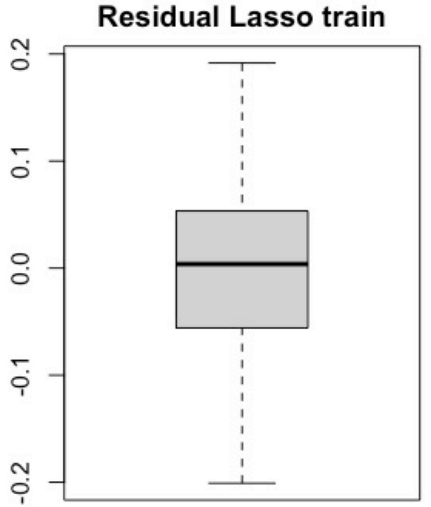



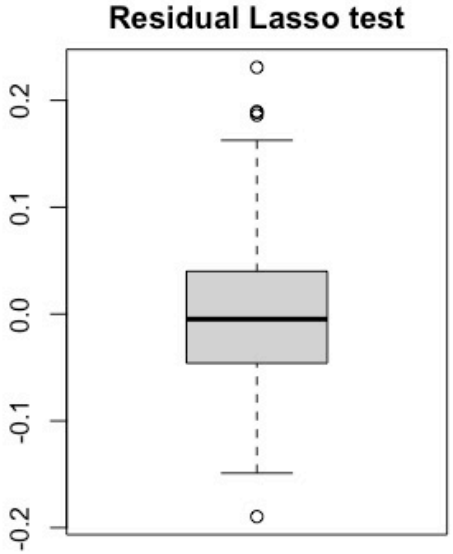
# Residuals For Test Model and Train Model

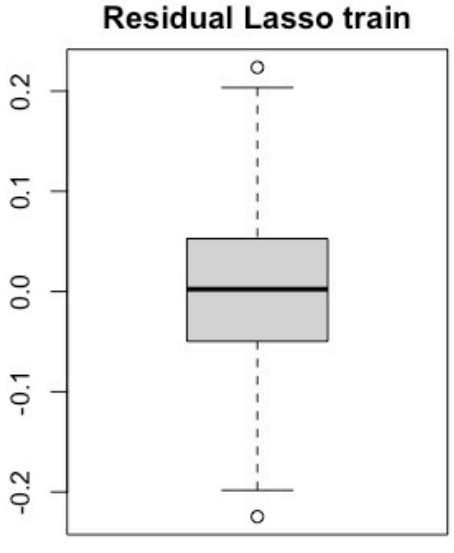


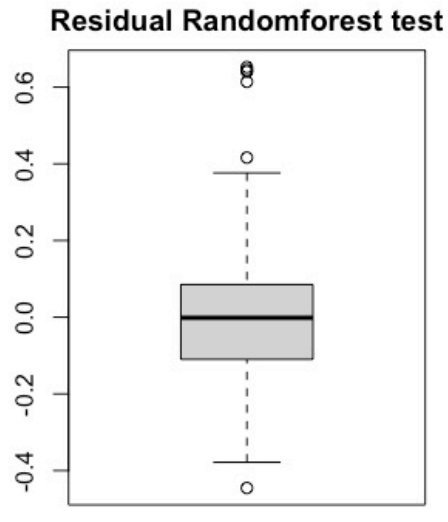



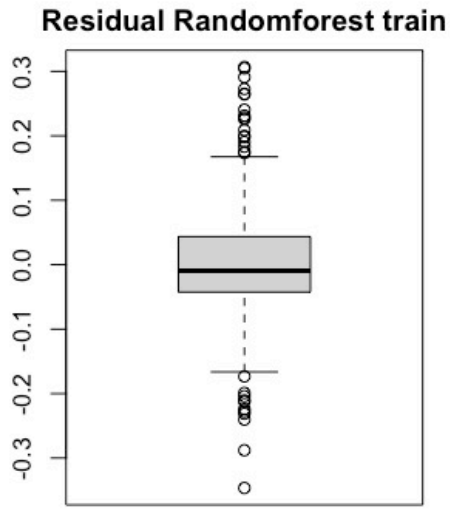



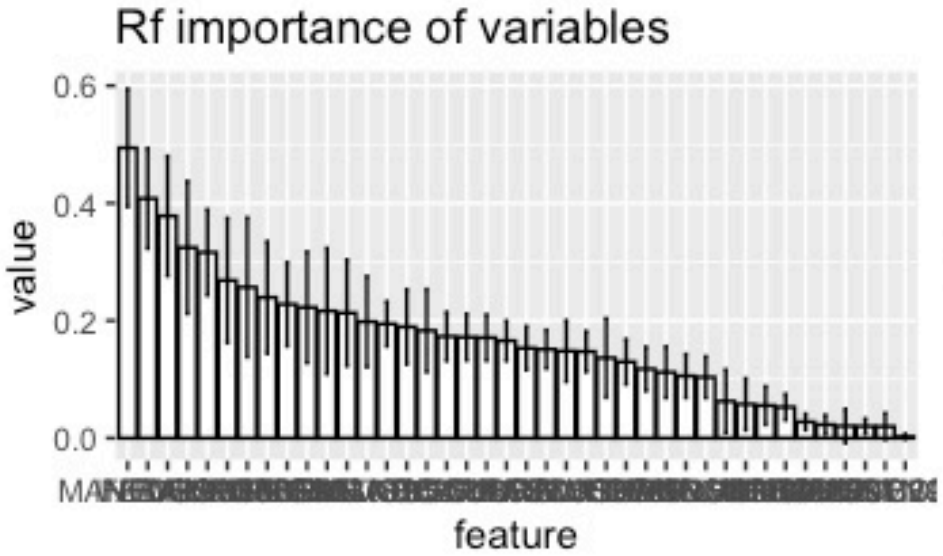



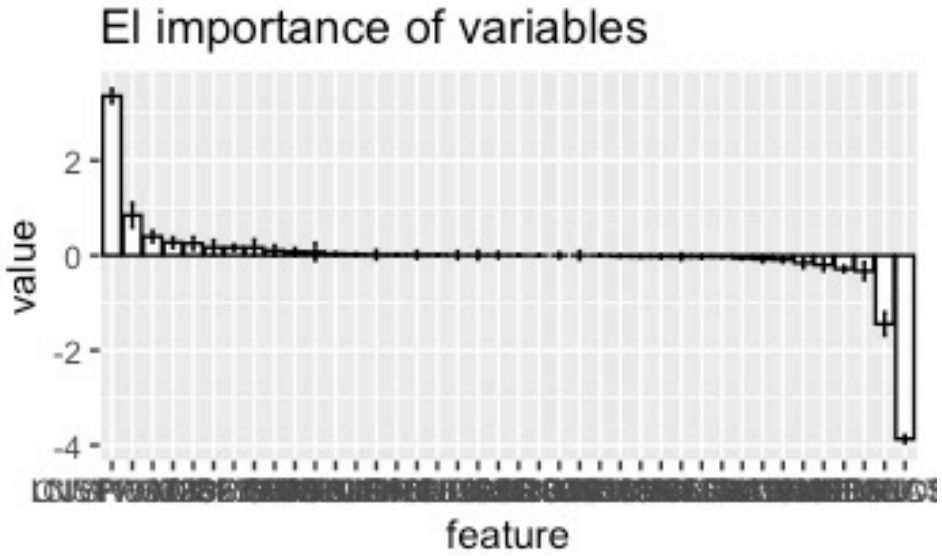



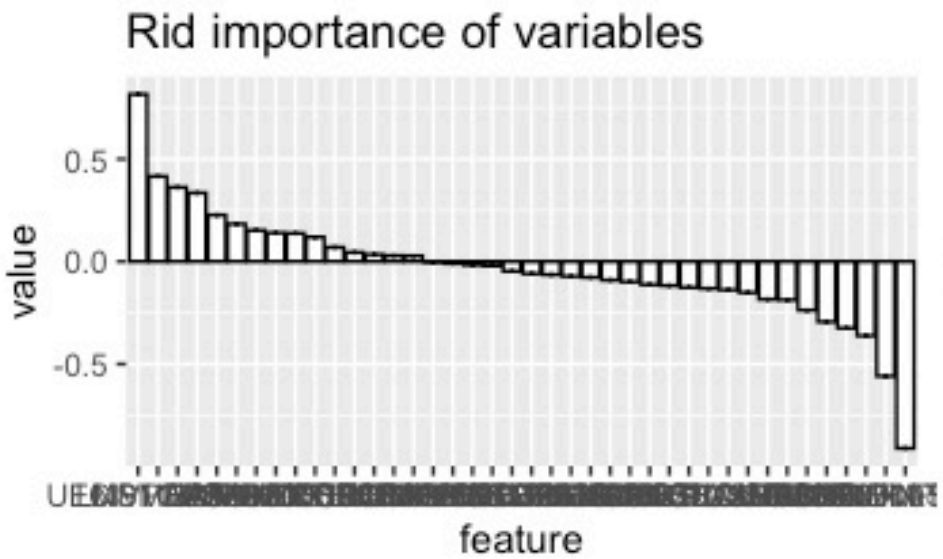



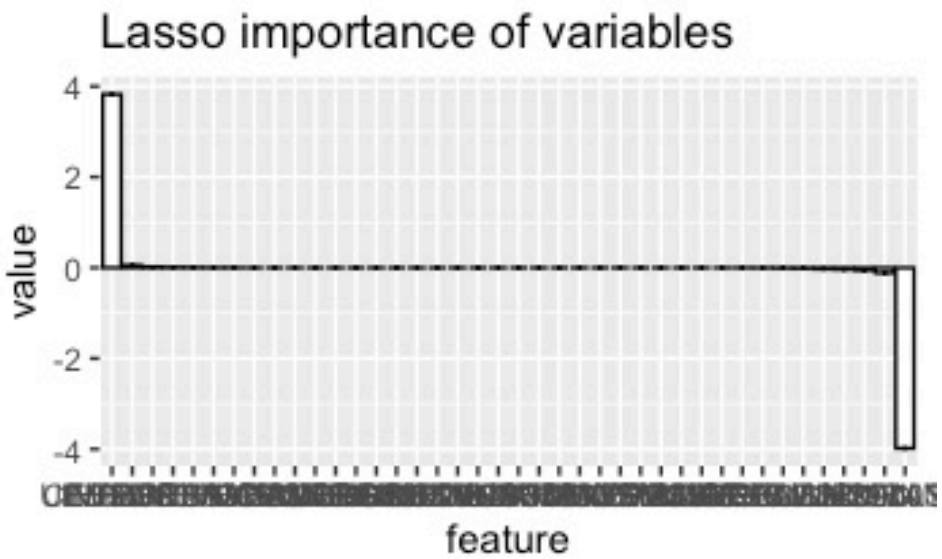



# Boxplot With Bootstrapped Error Bars



# Performance and Time

- **Performance:**
- Both Lasso and Elastic model have relatively smaller cross-validation error(k=10)
- Elastic model does better job based on residual residual boxplot
- **Running Time:**
- Ridge regression:20
- Lasso regression:10
- El regression:12
- Random forest:214
- **Comment:**
- Random forest is more time consuming compared with other models.