Ridge Regression

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

library(lmridge)

## Warning: package 'lmridge' was built under R version 4.0.5

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.0.5

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.3

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(regclass)

## Warning: package 'regclass' was built under R version 4.0.5

## Loading required package: bestglm

## Warning: package 'bestglm' was built under R version 4.0.5

## Loading required package: leaps

## Warning: package 'leaps' was built under R version 4.0.3

## Loading required package: VGAM

## Warning: package 'VGAM' was built under R version 4.0.5

## Loading required package: stats4

## Loading required package: splines

##   
## Attaching package: 'VGAM'

## The following object is masked from 'package:lmtest':  
##   
## lrtest

## Loading required package: rpart

## Warning: package 'rpart' was built under R version 4.0.3

## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 4.0.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

## Important regclass change from 1.3:  
## All functions that had a . in the name now have an \_  
## all.correlations -> all\_correlations, cor.demo -> cor\_demo, etc.

#loading data  
strfile="E:\\xampp\\htdocs\\ANN COVID-19\\ALL\_FINAL.csv"  
data\_covid\_IHSG=read.csv(file = strfile)

RidgeFM<-Close~.-Date-total\_boosters-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality

#iterative of K's  
Kopt=seq(0.0,1.0,0.005)

RidgeFit<-lmridge(RidgeFM,data\_covid\_IHSG,K= Kopt)

lmridge::kest(RidgeFit)

## Ridge k from different Authors  
##   
## k values  
## Thisted (1976): 0.00114  
## Dwividi & Srivastava (1978): 0.00013  
## LW (lm.ridge) 0.60215  
## LW (1976) 0.00151  
## HKB (1975) 0.00140  
## Kibria (2003) (AM) 0.00712  
## Minimum GCV at 0.00000  
## Minimum CV at 0.00000  
## Kibria 2003 (GM): 0.00306  
## Kibria 2003 (MED): 0.00251  
## Muniz et al. 2009 (KM2): 59.75445  
## Muniz et al. 2009 (KM3): 0.16733  
## Muniz et al. 2009 (KM4): 18.06683  
## Muniz et al. 2009 (KM5): 0.05535  
## Muniz et al. 2009 (KM6): 19.96884  
## Mansson et al. 2012 (KMN8): 60.95859  
## Mansson et al. 2012 (KMN9): 0.07431  
## Mansson et al. 2012 (KMN10): 23.77035  
## Mansson et al. 2012 (KMN11): 0.04207  
## Mansson et al. 2012 (KMN12): 23.32618  
## Dorugade et al. 2010: 0.00000  
## Dorugade et al. 2014: 0.53746

RidgeFit<-lmridge(RidgeFM,data\_covid\_IHSG,K= 0.16733)

summary(RidgeFit)

##   
## Call:  
## lmridge.default(formula = RidgeFM, data = data\_covid\_IHSG, K = 0.16733)  
##   
##   
## Coefficients: for Ridge parameter K= 0.16733   
## Estimate Estimate (Sc) StdErr (Sc) t-value (Sc)  
## Intercept 7.3921e+03 -1.7705e+11 1.0630e+10 -16.6554  
## new\_cases 5.6000e-03 1.5278e+03 1.9559e+02 7.8114  
## new\_deaths 1.8750e-01 1.4579e+03 2.0256e+02 7.1973  
## pneumonia\_mean -2.5802e+03 -4.1326e+03 1.7992e+02 -22.9688  
## reproduction\_rate -7.9693e+01 -6.2402e+02 1.8628e+02 -3.3500  
## new\_tests 1.3000e-03 2.5002e+03 1.8419e+02 13.5741  
## mask\_use\_mean 5.9576e+02 2.0276e+03 1.7652e+02 11.4866  
## mobility\_mean 1.6864e+01 3.5895e+03 2.0055e+02 17.8986  
## total\_vaccinations 0.0000e+00 2.7219e+02 1.9563e+02 1.3914  
## cumulative\_all\_vaccinated 0.0000e+00 3.9593e+03 1.5883e+02 24.9275  
## infection\_detection 2.7998e+04 5.5735e+03 1.7912e+02 31.1161  
## infection\_hospitalization 1.0392e+04 7.8605e+02 1.8848e+02 4.1704  
## Pr(>|t|)   
## Intercept <2e-16 \*\*\*  
## new\_cases <2e-16 \*\*\*  
## new\_deaths <2e-16 \*\*\*  
## pneumonia\_mean <2e-16 \*\*\*  
## reproduction\_rate 0.0009 \*\*\*  
## new\_tests <2e-16 \*\*\*  
## mask\_use\_mean <2e-16 \*\*\*  
## mobility\_mean <2e-16 \*\*\*  
## total\_vaccinations 0.1648   
## cumulative\_all\_vaccinated <2e-16 \*\*\*  
## infection\_detection <2e-16 \*\*\*  
## infection\_hospitalization <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Ridge Summary  
## R2 adj-R2 DF ridge F AIC BIC   
## 0.7889 0.7844 7.3200 503.7673 5181.4800 8233.0270   
## Ridge minimum MSE= 93219589 at K= 0.16733   
## P-value for F-test ( 7.32 , 479.094 ) = 3.754492e-220   
## -------------------------------------------------------------------

#assumption check  
  
lmtest::bptest(RidgeFit)

##   
## studentized Breusch-Pagan test  
##   
## data: RidgeFit  
## BP = 71.853, df = 11, p-value = 5.421e-11

lmtest::gqtest(RidgeFit)

##   
## Goldfeld-Quandt test  
##   
## data: RidgeFit  
## GQ = 0.23907, df1 = 232, df2 = 232, p-value = 1  
## alternative hypothesis: variance increases from segment 1 to 2

lmtest::hmctest(RidgeFit)

##   
## Harrison-McCabe test  
##   
## data: RidgeFit  
## HMC = 0.73464, p-value = 1

lmtest::dwtest(RidgeFit)

##   
## Durbin-Watson test  
##   
## data: RidgeFit  
## DW = 0.2268, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0

lmtest::bgtest(RidgeFit)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: RidgeFit  
## LM test = 385.9, df = 1, p-value < 2.2e-16

vif(RidgeFit)

## new\_cases new\_deaths pneumonia\_mean reproduction\_rate new\_tests  
## k=0.16733 0.94729 1.01603 0.80157 0.85921 0.84005  
## mask\_use\_mean mobility\_mean total\_vaccinations  
## k=0.16733 0.77153 0.99589 0.94764  
## cumulative\_all\_vaccinated infection\_detection  
## k=0.16733 0.6247 0.79444  
## infection\_hospitalization  
## k=0.16733 0.87969

Ridgeauto=dwtest(RidgeFit)  
pRidge=(1-(Ridgeauto$statistic/2))  
bptest(RidgeFit)

##   
## studentized Breusch-Pagan test  
##   
## data: RidgeFit  
## BP = 71.853, df = 11, p-value = 5.421e-11

Y\_var<-data\_covid\_IHSG$Close  
X\_var<-subset(data\_covid\_IHSG,select=-c(Date,total\_boosters,Close))

RidgeFMIt<-Close~.-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality

#remove auto-correlation  
#iter=1  
  
Y\_var<-Y\_var[2:length(Y\_var)]-(pRidge\*Y\_var[1:length(Y\_var)-1])  
X\_var<-X\_var[2:nrow(X\_var),]-(pRidge\*X\_var[1:nrow(X\_var)-1,])  
data\_transform=data.frame("Close"=Y\_var,X\_var)  
  
RidgeFitIt<-lmridge(RidgeFMIt,data\_transform,K= 0.16733)  
RidgeautoIt=dwtest(RidgeFitIt)  
pRidge=(1-(RidgeautoIt$statistic/2))  
RidgeautoIt

##   
## Durbin-Watson test  
##   
## data: RidgeFitIt  
## DW = 1.7568, p-value = 0.00187  
## alternative hypothesis: true autocorrelation is greater than 0

#iter=2  
  
Y\_var<-Y\_var[2:length(Y\_var)]-(pRidge\*Y\_var[1:length(Y\_var)-1])  
X\_var<-X\_var[2:nrow(X\_var),]-(pRidge\*X\_var[1:nrow(X\_var)-1,])  
data\_transform=data.frame("Close"=Y\_var,X\_var)  
  
RidgeFitIt<-lmridge(RidgeFMIt,data\_transform,K= 0.16733)  
RidgeautoIt=dwtest(RidgeFitIt)  
pRidge=(1-(RidgeautoIt$statistic/2))  
RidgeautoIt

##   
## Durbin-Watson test  
##   
## data: RidgeFitIt  
## DW = 2.0022, p-value = 0.4425  
## alternative hypothesis: true autocorrelation is greater than 0

vif.lmridge(RidgeFitIt)

## new\_cases new\_deaths pneumonia\_mean reproduction\_rate new\_tests  
## k=0.16733 0.89432 0.82552 0.86363 0.76855 0.83778  
## mask\_use\_mean mobility\_mean total\_vaccinations  
## k=0.16733 0.82362 0.89237 0.79545  
## cumulative\_all\_vaccinated infection\_detection  
## k=0.16733 0.98606 0.87998  
## infection\_hospitalization  
## k=0.16733 0.9651

summary(RidgeFitIt)

##   
## Call:  
## lmridge.default(formula = RidgeFMIt, data = data\_transform, K = 0.16733)  
##   
##   
## Coefficients: for Ridge parameter K= 0.16733   
## Estimate Estimate (Sc) StdErr (Sc) t-value (Sc)  
## Intercept 7.1492e+02 -3.4765e+09 4.0892e+08 -8.5018  
## new\_cases 2.0000e-04 1.4890e+01 6.6121e+01 0.2252  
## new\_deaths 3.9800e-02 9.4985e+01 6.3527e+01 1.4952  
## pneumonia\_mean -2.7044e+03 -4.5093e+02 6.4977e+01 -6.9399  
## reproduction\_rate -8.4662e+01 -1.9256e+02 6.1296e+01 -3.1415  
## new\_tests 1.0000e-04 1.0390e+02 6.3997e+01 1.6236  
## mask\_use\_mean 1.2669e+03 3.8488e+02 6.3454e+01 6.0655  
## mobility\_mean 1.0593e+01 2.8369e+02 6.6049e+01 4.2951  
## total\_vaccinations 0.0000e+00 8.8980e-01 6.2359e+01 0.0143  
## cumulative\_all\_vaccinated 0.0000e+00 7.7381e+02 6.9430e+01 11.1452  
## infection\_detection 1.5820e+04 3.8329e+02 6.5589e+01 5.8438  
## infection\_hospitalization 1.9440e+04 2.0943e+02 6.8688e+01 3.0491  
## Pr(>|t|)   
## Intercept <2e-16 \*\*\*  
## new\_cases 0.8219   
## new\_deaths 0.1355   
## pneumonia\_mean <2e-16 \*\*\*  
## reproduction\_rate 0.0018 \*\*   
## new\_tests 0.1051   
## mask\_use\_mean <2e-16 \*\*\*  
## mobility\_mean <2e-16 \*\*\*  
## total\_vaccinations 0.9886   
## cumulative\_all\_vaccinated <2e-16 \*\*\*  
## infection\_detection <2e-16 \*\*\*  
## infection\_hospitalization 0.0024 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Ridge Summary  
## R2 adj-R2 DF ridge F AIC BIC   
## 0.38830 0.37540 8.88787 41.26012 4135.58643 7179.29030   
## Ridge minimum MSE= 146458.1 at K= 0.16733   
## P-value for F-test ( 8.88787 , 475.5171 ) = 7.98858e-54   
## -------------------------------------------------------------------

sqrt(mean(residuals.lmridge(RidgeFitIt)^2))

## [1] 69.16056

#assumption check  
  
lmtest::bptest(RidgeFitIt)

##   
## studentized Breusch-Pagan test  
##   
## data: RidgeFitIt  
## BP = 100.24, df = 11, p-value < 2.2e-16

lmtest::gqtest(RidgeFitIt)

##   
## Goldfeld-Quandt test  
##   
## data: RidgeFitIt  
## GQ = 0.35162, df1 = 231, df2 = 231, p-value = 1  
## alternative hypothesis: variance increases from segment 1 to 2

lmtest::hmctest(RidgeFitIt)

##   
## Harrison-McCabe test  
##   
## data: RidgeFitIt  
## HMC = 0.70054, p-value = 1

lmtest::dwtest(RidgeFitIt)

##   
## Durbin-Watson test  
##   
## data: RidgeFitIt  
## DW = 2.0022, p-value = 0.4425  
## alternative hypothesis: true autocorrelation is greater than 0

lmtest::bgtest(RidgeFitIt)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: RidgeFitIt  
## LM test = 0.049681, df = 1, p-value = 0.8236

RidgeFMIt<-Close~.-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality-new\_cases-new\_deaths-new\_tests-total\_vaccinations

RidgeFitIt<-lmridge(RidgeFMIt,data\_transform,K= 0.16733)

summary(RidgeFitIt)

##   
## Call:  
## lmridge.default(formula = RidgeFMIt, data = data\_transform, K = 0.16733)  
##   
##   
## Coefficients: for Ridge parameter K= 0.16733   
## Estimate Estimate (Sc) StdErr (Sc) t-value (Sc)  
## Intercept 7.0427e+02 -3.6286e+09 3.0418e+08 -11.9291  
## pneumonia\_mean -2.6181e+03 -4.3654e+02 6.4760e+01 -6.7408  
## reproduction\_rate -8.2085e+01 -1.8670e+02 6.0746e+01 -3.0735  
## mask\_use\_mean 1.2751e+03 3.8737e+02 6.3482e+01 6.1020  
## mobility\_mean 1.0181e+01 2.7264e+02 6.5698e+01 4.1499  
## cumulative\_all\_vaccinated 0.0000e+00 8.0871e+02 6.7793e+01 11.9291  
## infection\_detection 1.6186e+04 3.9217e+02 6.5501e+01 5.9873  
## infection\_hospitalization 1.9558e+04 2.1071e+02 6.8372e+01 3.0818  
## Pr(>|t|)   
## Intercept <2e-16 \*\*\*  
## pneumonia\_mean <2e-16 \*\*\*  
## reproduction\_rate 0.0022 \*\*   
## mask\_use\_mean <2e-16 \*\*\*  
## mobility\_mean <2e-16 \*\*\*  
## cumulative\_all\_vaccinated <2e-16 \*\*\*  
## infection\_detection <2e-16 \*\*\*  
## infection\_hospitalization 0.0022 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Ridge Summary  
## R2 adj-R2 DF ridge F AIC BIC   
## 0.38120 0.37340 5.64877 63.87687 4134.14491 7164.28924   
## Ridge minimum MSE= 106295.7 at K= 0.16733   
## P-value for F-test ( 5.64877 , 479.3339 ) = 1.08346e-55   
## -------------------------------------------------------------------

sqrt(mean(residuals.lmridge(RidgeFitIt)^2))

## [1] 69.51986

#assumption check  
  
lmtest::bptest(RidgeFitIt)

##   
## studentized Breusch-Pagan test  
##   
## data: RidgeFitIt  
## BP = 99.427, df = 7, p-value < 2.2e-16

lmtest::gqtest(RidgeFitIt)

##   
## Goldfeld-Quandt test  
##   
## data: RidgeFitIt  
## GQ = 0.35779, df1 = 235, df2 = 235, p-value = 1  
## alternative hypothesis: variance increases from segment 1 to 2

lmtest::hmctest(RidgeFitIt)

##   
## Harrison-McCabe test  
##   
## data: RidgeFitIt  
## HMC = 0.69778, p-value = 1

lmtest::dwtest(RidgeFitIt)

##   
## Durbin-Watson test  
##   
## data: RidgeFitIt  
## DW = 1.9751, p-value = 0.3123  
## alternative hypothesis: true autocorrelation is greater than 0

lmtest::bgtest(RidgeFitIt)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: RidgeFitIt  
## LM test = 0.0059227, df = 1, p-value = 0.9387

vif.lmridge(RidgeFitIt)

## pneumonia\_mean reproduction\_rate mask\_use\_mean mobility\_mean  
## k=0.16733 0.85586 0.75303 0.82241 0.88082  
## cumulative\_all\_vaccinated infection\_detection  
## k=0.16733 0.9379 0.87554  
## infection\_hospitalization  
## k=0.16733 0.95397