Παράρτημα ΙΙΙ

Κώδικας R για την ανάλυση πολλαπλής γραμμικής παλινδρόμησης

Εισαγωγή δεδομένων:

```
data = read.csv(file.choose(), row.names=1)

names(data) = c("n2o", "ctl", "pig", "sheep", "goat", "crop", "chem", "chc", "cl", "cng", "cpp",
"cco");

x0<-data$n2o;
x1<-data$ctl;
x2<-data$pig;
x3<-data$sheep;
x4<-data$goat;
x5<-data$crop;
x6<-data$chem;
x7<-data$chc;
x8<-data$cl;
x9<-data$cng;
x10<-data$cpp;
x11<-data$cco;
```

Έλεγχος συσχέτισης και πίνακας τιμών VIF:

```
 install.packages("corrplot");
library(corrplot);
dedomena <- data.frame(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11)
M <- cor(dedomena);
M;
corrplot(M, method = "circle")

install.packages("usdm");
library(usdm);
vif(dedomena);
```

Προσαρμογή μοντέλου:

```
full<-lm(x0~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11);
 summary(full);
```

Τιμές των κριτηρίων AIC και BIC:

```
AIC(full);
BIC(full);
```

Μέθοδοι Penalized:

Lasso:

```
dedomena2<-data.matrix(dedomena, rownames.force = NA)
library(glmnet)
lasso2=glmnet(dedomena2,x0)
plot(lasso2,label=T)
lasso3<-cv.glmnet(dedomena2,x0)
lasso3$lambda.min
plot(lasso3$lambda[40:70], lasso3$cvm[40:70], xlab=’lambda’, ylab=’MSE’)
blasso3<-coef(lasso3,s="lambda.min")
blasso3
```

Ridge Regression:

```
library(MASS)
ridge2<-
lm.ridge(x0~x1+x2+x3+x4+x5+x6+x7,data=dedomena,lambda=seq(0,500,length.out=1500))
plot(ridge2)
legend('topright',c("x1","x2","x3","x4","x5","x6","x7"),col=1:7,lty=1:7)
library(parcor)
ridgecv<-
ridge.cv(as.matrix(dedomena2),x0,plot.it=T,lambda=seq(0.001,0.25,length.out=10000),k=10)
ridgecv
cridge<-coef(ridgecv,s=0.03068068)
```

Βέλτιστο μοντέλο με βάση τα AIC και BIC:

```
integer.base.b<-function(x, b=2){
xi <- as.integer(x)
if(any(is.na(xi) | ((x-xi)!=0)))
print(list(ERROR="x not integer", x=x))
N <- length(x)
xMax <- max(x)
ndigits <- (floor(logb(xMax, base=2))+1)
Base.b <- array(NA, dim=c(N, ndigits))
for(i in 1:ndigits){#i <- 1
Base.b[, ndigits-i+1] <- (x %% b)
x <- (x %/% b)
}
if(N ==1) Base.b[1, ] else Base.b
}

full_enumeration.linear_BIC_AIC<-function(x,y){
 p<-ncol(x) # number of explanatotory variables
 n<-nrow(x) # sample size
 namesg<-colnames(x)
 X<-cbind(1, x)
```

```
nmodels<-2^p
 result<-matrix(0, nrow=nmodels, ncol=p+2)
 gamma<-integer.base.b(0:(nmodels-1))
 gamma<-cbind(1,gamma)
 bic<-BIC(lm(y~1))
aic<-AIC(lm(y~1))
 result[1,]<-c(gamma[1,-1],bic,aic)

 for (i in 2:nmodels){
   if(floor(i/100)==i/100) print(i)
   Xg<-as.matrix(X[,gamma[i,]==1], nrow=n)
           xg<-Xg[,-1]
           bic<-BIC(lm(y~xg))
           aic<-AIC(lm(y~xg))
   result[i,]<-c(gamma[i,-1],bic,aic)
 }
 # sort results according to BIC
 index <- order(result[,p+1], decreasing=FALSE)
 result<-result[index,]
 colnames(result)<-c(namesg,"BIC","AIC")
 print(result)

# sort results according to AIC
index <- order(result[,p+2], decreasing=FALSE)
 result<-result[index,]
 colnames(result)<-c(namesg,"BIC","AIC")
  print(result)
}
full_enumeration.linear_BIC_AIC(dedomena,x0)
```

Προσαρμογή τελικού μοντέλου:

```
results_teliko<-lm(x0~x2+x3+x4+x5+x8+x9+x11);
summary(results_teliko);
```

Διαστήματα Εμπιστοσύνης:

```
confint(results_teliko);
```

Τιμές των κριτηρίων AIC και BIC:

```
AIC(results_teliko);
BIC(results_teliko);
```

Έλεγχος Γραμμικότητας:

```
 par(mfrow=c(2,4));
 plot(residuals(results_teliko,"partial")[,1],x2);
 plot(residuals(results_teliko,"partial")[,2],x3);
 plot(residuals(results_teliko,"partial")[,3],x4);
plot(residuals(results_teliko,"partial")[,4],x5);
plot(residuals(results_teliko,"partial")[,5],x8);
plot(residuals(results_teliko,"partial")[,6],x9);
plot(residuals(results_teliko,"partial")[,7],x11);
```

Έλεγχος Κανονικότητας:

```
 qqnorm(residuals(results_teliko));
 qqline(residuals(results_teliko));

library(moments)
jarque.test(results_teliko$residuals)
```

Έλεγχος Ομοσκεδαστικότητας:

```
plot(results_teliko$res,results_teliko$fitted);

install.packages("lmtest")
library(lmtest)
bptest(results_teliko)
```

Έλεγχος Ανεξαρτησίας Υπολοίπων:

```
 plot(results_teliko$res);
abline(0,0);
plot(results_teliko$res,type='l');

dwtest(results_teliko)
```

Πρόβλεψη:

```
predict(results_teliko,list(x2<-12212300,x3<-1033570,x4<-412550,x5<-20027572,x8<-31 ,
x9<-1391534,x11<-52570),int="p");
```

Κώδικας για την ανάλυση με χρήση Panel Data

Γραφήματα ερμηνείας δεδομένων:

```
library(plm)
library(car)
scatterplot(y~Year|Country, boxplots=FALSE, smooth=TRUE, reg.line=FALSE, data=data)

library(gplots)
plotmeans(y ~ Country, main="Heterogeneity across countries", data=data)

library(gplots)
plotmeans(y ~ Year, main="Heterogeneity across time", data=data)
```

Μέθοδος Ελαχίστων Τετραγώνων:

```
ols <-lm(y ~ x1+x2+x3+x4, data=data)
summary(ols)
```

Μέθοδος Ελαχίστων Τετραγώνων με Ψευδομεταβλητές:

```
fixed.dum <-lm(y~ x1+x2+x3+x4+ factor(Country) - 1, data=data)
summary(fixed.dum)
```

Μοντέλο Σταθερών Επιδράσεων:

```
library(plm)
fixed <- plm(y ~x1+x2+x3+x4, data=data, index=c("Country", "Year"), model="within")
summary(fixed)
```

Έλεγχος pFtest για επιλογή καλύτερου μοντέλου ανάμεσα στα fixed και ols:

```
pFtest(fixed, ols)
```

Μοντέλο Τυχαίων επιδράσεων:

```
random <- plm(y ~x1+x2+x3+x4, data=data, index=c("Country", "Year"), model="random")
summary(random)
```

Έλεγχος Hausman Test:

```
phtest(fixed,random)
```

Εναλλακτικός έλεγχος για την επιλογή καλύτερου μοντέλου ανάμεσα στα ols και random:

```
random<- plm(y ~ x4+x7+x8+x9+x10+x12, data=data, index=c("Country", "Year"),
model="pooling")
summary(pool)
plmtest(pool, type=c("bp"))
```

Επιλογή καταλληλότερου μοντέλου με βάση τα κριτήρια AIC και BIC:

```
integer.base.b<-function(x, b=2){
xi <- as.integer(x)
if(any(is.na(xi) | ((x-xi)!=0)))
print(list(ERROR="x not integer", x=x))
N <- length(x)
xMax <- max(x)
ndigits <- (floor(logb(xMax, base=2))+1)
Base.b <- array(NA, dim=c(N, ndigits))
for(i in 1:ndigits){#i <- 1
Base.b[, ndigits-i+1] <- (x %% b)
x <- (x %/% b)
}
if(N ==1) Base.b[1, ] else Base.b
}

full_enumeration.linear_BIC_AIC<-function(x,y){
 p<-ncol(x) # number of explanatotory variables
 n<-nrow(x) # sample size
 namesg<-colnames(x)
 X<-cbind(1, x)

nmodels<-2^p
 result<-matrix(0, nrow=nmodels, ncol=p+2)
 gamma<-integer.base.b(0:(nmodels-1))
 gamma<-cbind(1,gamma)
 bic<-BIC(lm(y~1))
aic<-AIC(lm(y~1))
 result[1,]<-c(gamma[1,-1],bic,aic)

 for (i in 2:nmodels){
   if(floor(i/100)==i/100) print(i)
   Xg<-as.matrix(X[,gamma[i,]==1], nrow=n)
           xg<-Xg[,-1]
           bic<-BIC(lm(y~xg))
           aic<-AIC(lm(y~xg))
   result[i,]<-c(gamma[i,-1],bic,aic)
 }
 # sort results according to BIC
 index <- order(result[,p+1], decreasing=FALSE)
 result<-result[index,]
 colnames(result)<-c(namesg,"BIC","AIC")
 print(result)
```

```
# sort results according to AIC
index <- order(result[,p+2], decreasing=FALSE)
 result<-result[index,]
 colnames(result)<-c(namesg,"BIC","AIC")
  print(result)
}
full_enumeration.linear_BIC_AIC(dedomena,y)
```

Έλεγχος Ομοσκεδαστικότητας:

```
library(lmtest)
bptest(y ~x1 + x2 + x4 + x7 + x11 + x12 + factor(Country), data = data, studentize=F)
```

Συντελεστές robust για αντιμετώπιση ετεροσκεδαστικότητας:

```
coeftest(fixed, vcovHC(fixed, method = "arellano"))
coeftest(random, vcovHC(random, method = "white1"))
```