Econometrics-Damodar N. Gujarati / Chapter 1

Furkan Zengin

16 08 2021

```
install.packages('remotes')


remotes::install_github("brunoruas2/gujarati",force = TRUE)

library(gujarati)

attach(Table1_1)

library(tidyverse)

plot(Table1_1$Y1,Table1_1$X1,xlab="Number of eggs produced (millions)",
     ylab="Price of eggs per dozen (in cent")


View(Table1_2)

library(ggplot2)
ggplot(Table1_2, aes(Table1_2$C.1, Table1_2$I)) + geom_point()
+facet_wrap(~Table1_2$FIRM)+ theme(legend.position = "none",
        panel.grid = element_blank(),
        axis.title = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.background = element_blank())

attach(Table1_3)


view(Table1_3)

ggplot(Table1_3, aes(YEAR, USA)) + geom_point()

attach(Table2_10)

view(Table2_10)

# using fix(), change the name and class of variables

MODEL1=lm(MATH~AVGI)
summary(MODEL1)

plot(AVGI,MATH)
abline(MODEL1,col="darkred",lwd=3)
```

$$Math = \beta_0 + \beta_1 AverageIncome + u_i$$

```
attach(Table2_8)

view(Table2_8)
fix(Table2_8)

MODEL2 = lm(Table2_8$TOTALEXP ~ Table2_8$FOODEXP)
summary(MODEL2)

plot(Table2_8$FOODEXP,Table2_8$TOTALEXP,xlab="Food Expenditure",
     ylab="Total Expenditure")
abline(MODEL2,col="darkred",lwd=3)
```

$$TotalExp = \beta_0 + \beta_1 FoodExp + u_i$$

```
                                        FORMULAS
```

$$\begin{aligned} \hat{\beta}_1 &= \\ &= \frac{\sum_i \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_i \left(x_i - \overline{x}\right)} \\ &= \frac{\sum_i \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)/(n-1)}{\sum_i \left(x_i - \overline{x}\right)/(n-1)} \\ &= \frac{\hat{\mathrm{Cov}}(x,\, y)}{\hat{\mathrm{Var}}(x)} \end{aligned}$$

$$\hat{\beta}_1 = \frac{\hat{\mathrm{Cov}}(x,\, y)}{\hat{\mathrm{Var}}(x)}$$

$$\hat{\beta}_1 = \frac{\hat{\mathrm{Cov}}(\tilde{x}_1,\, y)}{\hat{\mathrm{Var}}(\tilde{x}_1)}$$

$$R^2 = \frac{\sum_i (\hat{y}_i - \overline{y})^2}{\sum_i \left(y_i - \overline{y}\right)^2} = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \overline{y}\right)^2}$$

[1] Latex expressions are taken from :

---

1. https://github.com/tatanik501/EC421S20/blob/master/notes/03-review/03-review.Rmd (https://github.com/tatanik501/EC421S20/blob/master/notes/03-review/03-review.Rmd)↩

Econometrics-Damodar N. Gujarati Chapter 2/3

Furkan Zengin

16 08 2021

$$\mathrm{Y}_i = \beta_0 + \beta_1 \mathrm{X}_i + u_i$$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$$E[u \mid X] = 0$$

$$\hat{\mathrm{SE}}\left(\hat{\beta}_1\right) = \sqrt{\frac{s^2}{\sum_i (x_i - \overline{x})^2}}$$

```
library(gujarati)
library(tidyverse)
library(broom)
lm(y ~ x, data = pop_df) %>% tidy()

view(Table3_3)

fix(Table3_3)
MODEL1 = lm(Table3_3$Cellphone ~ Table3_3$Pcapincome)

summary(MODEL1)
plot(Table3_3$Pcapincome, Table3_3$Cellphone,xlab = "Income",
     ylab = "CellPhone")

abline(MODEL1)
```

$$Empirical Exercises$$

$$3.20$$

```
fix(Table3_6)
par(mfrow=c(2,2))
plot(Table3_6$OUT_BUS, Table3_6$COM_BUS,xlab = "Output per Hour of All Persons",
     ylab = "Real Compensation per Hour",main = "Business")
plot(Table3_6$OUT_NBUS, Table3_6$COM_NBUS,xlab = "Output per Hour of All Persons",
     ylab = "Real Compensation per Hour",main = "Non-Farm",col="red")
```

$$Output = \hat{\beta}_0 + \hat{\beta}_1 Compensation + e_i$$

$$3.22$$

```
fix(Table3_7)
MODEL2 = lm(Table3_7$Gold.Price ~ Table3_7$CPI)

summary(MODEL2)


MODEL3 = lm(Table3_7$NYSE ~ Table3_7$CPI)

summary(MODEL3)
```

$$GoldPrice = \hat{\beta}_0 + \hat{\beta}_1 CPI + e_i$$

$$GoldPrice = \underset{(54.4685)}{215.2856} + \underset{(0.4038)}{1.0384} CPI$$

$$Consistency$$

$$\lim_{n \to \infty} P(|B_n - \alpha| > \epsilon) = 0 \tag{1}$$

Econometrics-Damodar N. Gujarati / Chapter 5

Furkan Zengin

16 08 2021

$$H_o: \ \beta_1 = 0$$

$$t_{\text{stat}} = \frac{\hat{\beta}_1 - c}{\hat{\text{SE}}\left(\hat{\beta}_1\right)}$$

$$\text{SSE} = \sum_{i=1}^{n} e_i^2$$

$$\begin{aligned} e_i^2 &= (y_i - \hat{y}_i)^2 = \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 \\ &= y_i^2 - 2y_i\hat{\beta}_0 - 2y_i\hat{\beta}_1 x_i + \hat{\beta}_0^2 + 2\hat{\beta}_0\hat{\beta}_1 x_i + \hat{\beta}_1^2 x_i^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{SSE}}{\partial \hat{\beta}_0} &= \sum_i \left(2\hat{\beta}_0 + 2\hat{\beta}_1 x_i - 2y_i\right) = 2n\hat{\beta}_0 + 2\hat{\beta}_1 \sum_i x_i - 2\sum_i y_i \\ &= 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} - 2n\overline{y} \end{aligned}$$

$$\text{t} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum_{i=1}^{n} x_i}}{\hat{\sigma}}$$

```
$$Empirical Exercises$$

$$5.9$$

```{r}



options(scipen = 999)
fix(Table5_5)
par(mfrow=c(2,2))

MODEL2 = lm(Table5_5$SALARY ~ Table5_5$SPENDING)

summary(MODEL2)


plot(Table5_5$SPENDING, Table5_5$SALARY,xlab = "SPENDING",
     ylab = "SALARY")
abline(MODEL2)
predict1=predict(MODEL2,interval = "confidence")
predict1 # shows that fitted values and lower and upper intervals
fitted.values(MODEL2) # shows the same
```
```

$$\underset{(1197.3508)}{SALARY = 12129.3710} + \underset{(0.3117)}{3.3076}SPENDING$$

Using:

$$https://www.econometrics-with-r.org/index.html$$

```
t <- seq(-15, 15, 0.01)

plot(x = t,
     y = dnorm(t, 0, 1),
     type = "l",
     col = "steelblue",
     lwd = 2,
     yaxs = "i",
     axes = F,
     ylab = "",
     main = expression("Calculating the p-value of a Two-sided Test when" ~ t^act ~ "=10.61"),
     cex.lab = 0.7,
     cex.main = 1)

tact <- 10.61

axis(1, at = c(0, -1.96, 1.96, -tact, tact), cex.axis = 0.7)

# Shade the critical regions using polygon():

# critical region in left tail
polygon(x = c(-6, seq(-6, -1.96, 0.01), -1.96),
        y = c(0, dnorm(seq(-6, -1.96, 0.01)), 0),
        col = 'orange')

# critical region in right tail

polygon(x = c(1.96, seq(1.96, 6, 0.01), 6),
        y = c(0, dnorm(seq(1.96, 6, 0.01)), 0),
        col = 'orange')

# Add arrows and texts indicating critical regions and the p-value
arrows(-3.5, 0.2, -2.5, 0.02, length = 0.1)
arrows(3.5, 0.2, 2.5, 0.02, length = 0.1)

arrows(-5, 0.16, 10.61, 0, length = 0.1)
```

```
arrows(5, 0.16, -10.61, 0, length = 0.1)

text(-3.5, 0.22,
     labels = expression("0.025"~"="~over(alpha, 2)),
     cex = 0.7)
text(3.5, 0.22,
     labels = expression("0.025"~"="~over(alpha, 2)),
     cex = 0.7)

text(-5, 0.18,
     labels = expression(paste("-|",t[act],"|")),
     cex = 0.7)
text(5, 0.18,
     labels = expression(paste("|",t[act],"|")),
     cex = 0.7)

# Add ticks indicating critical values at the 0.05-level, t^act and -t^act
rug(c(-1.96, 1.96), ticksize  = 0.145, lwd = 2, col = "darkred")
rug(c(-tact, tact), ticksize  = -0.0451, lwd = 2, col = "darkgreen")
```

```
options(scipen = 999)
fix(Table5_6)


MODEL3 = lm(Table5_6$GNP ~ Table5_6$M1)
MODEL3_1 = lm(Table5_6$GNP ~ Table5_6$M2)
MODEL3_2 = lm(Table5_6$GNP ~ Table5_6$M3)
MODEL3_3 = lm(Table5_6$GNP ~ Table5_6$L)

library(stargazer)
s1 = stargazer(list(MODEL3, MODEL3_1, MODEL3_2,MODEL3_3), type = "text")
```

5.16

```
options(scipen = 999)
fix(Table5_9)


MODEL4 = lm(Table5_9$`Actual Exchange Rate` ~ Table5_9$`Implied PPP`)
a = summary(MODEL4)

plot(log(Table5_9$`Implied PPP`),log(Table5_9$`Actual Exchange Rate`)
     ,xlab = "Implied PPP",
     ylab = "Actual Exchange Rate")
```

$$ActualExchangeRate = \underset{(26.98784)}{-33.09170} + \underset{(0.02744)}{1.81472} ImpliedPPP$$

Econometrics-Damodar N. Gujarati / Chapter 7

Furkan Zengin

17 08 2021

$$\textit{Multiple Regression Analysis : The Problem of Estimation}$$

$$\textit{Assumptions}$$

$$Y_i = \beta_1 \ + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \tag{7.1.1}$$

$$\boldsymbol{E}[u \mid X_{2i}, X_{3i}] = 0 \tag{7.1.4}$$

$$\mathrm{Var}(u_i) = \sigma^2 \tag{7.1.5}$$

```
options(scipen = 999

library(gujarati)

library(ggplot2)

fix(Table6_4)

MODEL1 = lm(Table6_4$CM ~ Table6_4$FLR)

summary(MODEL1)

plot(Table6_4$FLR, Table6_4$CM,xlab = "Female Literacy",
 ylab = "Child Mortality")
abline(MODEL1)
```

$$C\hat{M}_i = \underset{(12.2250)}{263.8635} - \underset{(0.2133)}{2.3905} FLR_i \tag{7.3.1}$$

```
options(scipen = 999

MODEL2 = lm(Table6_4$CM ~ Table6_4$PGNP + Table6_4$FLR)
summary(MODEL2)
```

$$\hat{CM}_i = \underset{(11.593179)}{263.64} - \underset{(0.002003)}{0.005647} PGNPi - \underset{(0.209947)}{2.231586} FLR_i \tag{7.6.2}$$

$$R^2 = 1 - \frac{\sum_i \left(\hat{u}_i^2\right)}{\sum_i \left(\hat{y}_i^2\right)} \tag{7.8.1}$$

```
options(scipen = 999

fix(Table7_1)

MODEL3 = lm(Table7_1$Y ~ Table7_1$X)

summary(MODEL3)



MODEL4 = lm(log(Table7_1$Y) ~ log(Table7_1$X))

summary(MODEL4)
```

$$\hat{Y}_t = \underset{(0.1216)}{2.6911} - \underset{(0.1140)}{0.4795} X_t \quad R^2 = 0.6628 \tag{7.8.8}$$

$$\hat{Y}_t = \underset{(0.01524)}{0.77742} - \underset{(0.04937)}{0.25305} X_t \qquad R^2 = 0.7448 \tag{7.8.9}$$

$$The\ Cobb\text{–}\ Douglas\ Production\ Function : More\ on\ Functional\ Form$$

$$Y_i = \beta_1 \beta_2^{\beta_2} x_{2i} \beta_3^{\beta_3} x_{3i} e_i^u \tag{7.9.1}$$

```
options(scipen = 999

fix(Table7_3)

#Name them  as ValueAd, LaborIn and CapitalIn Respectively

MODEL5 = lm(log(Table7_3$ValueAd) ~ log(Table7_3$LaborIn) + log(Table7_3$CapitalIn))

summary(MODEL5)
```

$$In\hat{Y}_i = \underset{(0.39623)}{3.88760} + \underset{(0.04937)}{0.46833} InX_{2i} + + \underset{(0.09689)}{0.52128} InX_{3i} \qquad R^2 = 0.9642 \tag{7.9.4}$$

$$Polynomial\ Regression\ Models$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1\, X_1 + \hat{\beta}_2\, X_{2i}^2 + \cdots + \hat{\beta}_k\, X_{ki}^k + e_i \tag{7.10.3}$$

```
options(scipen = 999

fix(Table7_4)

x1 = Table7_4$X

x2 = (Table7_4$X)^2

x3 = (Table7_4$X)^3

MODEL6 = lm(Table7_4$Y ~ x1 + x2 +x3)

summary(MODEL6)

plot(x1, Table7_4$Y)
```

$$\hat{Y}_i = \underset{(6.37532)}{141.76667} + \underset{(4.77861)}{63.4776X_i} - \underset{(0.98566)}{12.96154X_i^2} + \underset{(0.05911)}{0.93959X_i^3} \qquad R^2 = 0.9983 \tag{7.10.6}$$

Econometrics-Damodar N. Gujarati / Chapter 8

Furkan Zengin

18 08 2021

$$Multiple\ Regression\ Analysis: The\ Problem\ \ of\ Inference$$

$$t = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\left(se(\hat{\beta}_1)\right)} \tag{8.1.1}$$

$$\vdots$$

```
options(scipen = 999)

library(gujarati)

library(ggplot2)

fix(Table6_4)

MODEL1 = lm(Table6_4$CM ~ Table6_4$PGNP + Table6_4$FLR)
summary(MODEL1)

# To find P-values, below formula can be used


library(AER)
2*(1 - pt(abs(coeftest(MODEL1, vcov. = vcovHC, type = "HC1")[2, 3]),
            df = MODEL1$df.residual))

#For Confidence Interval

confint(MODEL1,level = 0.95)
```

$$\hat{CM_i} = \underset{(11.593179)}{263.64} - \underset{(0.002003)}{0.005647} PGNPi - \underset{(0.209947)}{2.231586} FLR_i \tag{8.1.4}$$

$$t = (22.7411) \qquad (-2.8187) \qquad (-10.6293)$$

$$Hypothesis\ Testing\ in\ Multiple\ Regression$$

$$H_0:\ \beta_2 = 0$$
$$H_1:\ \beta_2 \neq 0$$

$$F = \frac{\left((R^2_{UR} - R^2_R)/m\right)}{\left(1 - R^2_{UR}\right)/(n-k)} \tag{8.6.10}$$

```
attach(Table6_4)
fix(Table6_4)


library(car)

MODEL1 = lm(CM ~ PGNP + FLR)
summary(MODEL1)

H0=c("PGNP","FLR")

linearHypothesis(MODEL1,H0)
```

The Chow Test

```
options(scipen = 999)

fix(Table8_9)

MODEL2 = lm(Table8_9$SAVINGS ~ Table8_9$INCOME)
summary(MODEL2)

library(strucchange)

sctest(Table8_9$SAVINGS ~ Table8_9$INCOME, type = "Chow", point = 12)

#Point is the year that the change occured


library(tidyverse)



new1 = Table8_9[Table8_9$YEAR >= "1970" & Table8_9$YEAR <= "1981",]

new2 = Table8_9[Table8_9$YEAR >= "1982" & Table8_9$YEAR <= "1995",]
par(mfrow=c(2,2))

plot(new1$INCOME,new1$SAVINGS,col = "steelblue",
     pch = 20,xlab = "INCOME",ylab = "SAVINGS")

plot(new2$INCOME,new2$SAVINGS,col = "steelblue",
     pch = 20,xlab = "INCOME",ylab = "SAVINGS")
```

Testing the Functional Form of Regression

```
options(scipen = 999)

fix(Table7_6)

MODEL3 = lm(Table7_6$Y ~ Table7_6$X2 + Table7_6$X3)
summary(MODEL3)

MODEL4 = lm(log(Table7_6$Y) ~ log(Table7_6$X2) + log(Table7_6$X3))
summary(MODEL4)

library(stargazer)
stargazer(list(MODEL3,MODEL4),type = "text")


library(lmtest)
petest(MODEL3, MODEL4, data = Table7_6)
```

Econometrics-Damodar N. Gujarati / Chapter 9

Furkan Zengin

19 08 2021

$$Dummy\,Variable\,Regression\,Models$$

$$Y_i = \beta_1 \ + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.1}$$

$$\boldsymbol{E}[(Y_i \mid D_{2i} = 1, D_{3i} = 0) = \beta_1 + \beta_2] \tag{9.2.2}$$

$$\ldots$$

```
options(scipen = 999)

library(gujarati)

library(ggplot2)

fix(Table9_1)

MODEL1 = lm(Table9_1$Salary ~ Table9_1$D2 + Table9_1$D3)

summary(MODEL1)

library(ggplot2)

labs =  as_labeller(c(`0` = "D2", `1` = "D3"))

ggplot(Table9_1, aes(Table9_1$Spending, Table9_1$Salary))+geom_point()+
  facet_wrap(~Table9_1$D2,labeller=labs)+xlab("Spending")+ylab("Salary")
```

$$\hat{Y}_t = \underset{(1857)}{48015} + \underset{(2363)}{1524} D_{2i} - \underset{(2467)}{1721} D_{3i} \quad R^2 = 0.04397 \tag{9.2.5}$$

```
options(scipen = 999)


MODEL2 = lm(Table9_1$Salary ~ Table9_1$D2 + Table9_1$D3 + Table9_1$Spending)

summary(MODEL2)

library(ggplot2)

labs =  as_labeller(c(`0` = "D2", `1` = "D3"))

ggplot(Table9_1, aes(Table9_1$Spending, Table9_1$Salary))+geom_point()+
  facet_wrap(~Table9_1$D2,labeller=labs)+xlab("Spending")+ylab("Salary")

ggplot(Table9_1, aes(Table9_1$Spending, Table9_1$Salary))+
  geom_point()+facet_wrap(~Table9_1$D2,labeller=labs)+geom_smooth(method = "lm")+xlab("Spending")+ylab("Salary")
```

$$\hat{Y}_t = \underset{(3262.5213)}{28694.9180} - \underset{(1862.575)}{2954.1268} D_{2i} - \underset{(1819.8725)}{3112.1948} D_{3i} + \underset{(0.3592)}{2.3404} X_i \qquad R^2 = 0.4977 \tag{9.4.2}$$

$$Interaction\ Effects\ Using\ Dummy\ Variables$$

$$Y_i = \alpha_1 \ + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \tag{9.6.2}$$

$$\boldsymbol{E}[(Y_i \mid D_{2i} = 1, D_{3i} = 1, X_i)] = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i) \tag{9.6.3}$$

$$Y = hourly\ wage\ in\ dollars$$
$$X = education$$
$$D_2 = 1\ if\ female,\ 0\ otherwise$$
$$D_3 = 1\ if\ NonWhite\ and\ NonHispanic,\ 0\ otherwise$$

$$There\ is\ no\ dataset\ for\ gender\ difference\ model,\ we\ can\ use\ Wooldrigde's\ data.$$

```
options(scipen = 999)
library(wooldridge)

attach(wage1)
MODEL3=lm(log(wage)~female+educ+female*educ+exper+expersq+tenure+tenursq)
summary(MODEL3)


labs =  as_labeller(c(`0` = "Male", `1` = "Female"))

ggplot(wage1, aes(educ, wage))+geom_point()+
  facet_wrap(~female,labeller=labs)+xlab("Education")+ylab("Wage")


library(sjPlot)
library(sjmisc)
library(ggplot2)
data(wage1)
theme_set(theme_sjplot())

wage1$female <- to_factor(wage1$female)

MODEL3 <- lm(log(wage)~educ+female+female*educ, data = wage1)

plot_model(MODEL3, type = "pred", terms = c("educ", "female"))
```

```
options(scipen = 999)

fix(Table9_4)


MODEL4=lm(Table9_4$FRIG ~  Table9_4$D2 + Table9_4$D3 +
            Table9_4$D4)
summary(MODEL4)


MODEL5=lm(Table9_4$FRIG ~  Table9_4$DUR +  Table9_4$D2 + Table9_4$D3 +
            Table9_4$D4)

summary(MODEL5)

library(stargazer)
stargazer(list(MODEL4,MODEL5) ,type = "text")



Fitted_Values = fitted(MODEL4)
Residuals = residuals(MODEL4)
Actual = Table9_4$FRIG
DF = data.frame(Actual,Fitted_Values,Residuals)

DF

plot(MODEL4$residuals)
```

```
options(scipen = 999)

fix(Table9_6)
plot(Table9_6$Output,Table9_6$TotalCost,type = "l")

attach(Table9_6)


#Threshold value is 5.500
X_star = 5500
D = ifelse(Output >= 5500,1,0)
subs = (Table9_6$Output - X_star)

New1 = data.frame(Table9_6,D,subs)
fix(New1)
MODEL6=lm(New1$TotalCost ~  New1$Output + New1$subs*D)
summary(MODEL6)
```

```
options(scipen = 999)

fix(Table9_7)
plot(Table9_7$WI,Table9_6$TotalCost,type = "l")

attach(Table9_6)


#Threshold value is 5.500
X_star = 5500
D = ifelse(Output >= 5500,1,0)
subs = (Table9_6$Output - X_star)

New1 = data.frame(Table9_6,D,subs)
fix(New1)
MODEL7=lm(log(Table9_7$WI) ~ Table9_7$AGE +Table9_7$DE2 + Table9_7$DE3 + Table9_7$DE4 +
            Table9_7$DPT + Table9_7$DSEX)
summary(MODEL7)
```

Empirical Exercises

9.24

```
options(scipen = 999)

fix(Table9_9)
attach(Table9_9)
MODEL8=lm(V ~ I + D + W +G*I + N +P)
summary(MODEL8)
```

$$\hat{Y}_t = \underset{(1857)}{0.499592} - \underset{(0.012934)}{0.00956}I_i - \underset{(0.016329)}{0.037411}D_i + \underset{(0.040128)}{0.007716}W_i + \underset{(0.002099)}{0.002621}G_i - \tag{9.24.c}$$

$$\underset{(0.003521)}{0.005109}N_i + \underset{(0.004331)}{0.001557}P_i + \underset{(0.001797)}{0.010298}IG_i \quad R^2 = 0.7958$$

Econometrics-Damodar N. Gujarati / Chapter 11

Furkan Zengin

20 08 2021

Heteroscedasticity

$$Homoscedasticity \tag{11.1.1}$$
$$E(u_i^2) = \sigma^2$$

```
library(pacman)
p_load(
  broom, latex2exp, ggplot2, ggthemes, viridis, extrafont,
  dplyr,
  magrittr, knitr, parallel
)
# Define pink color
red_pink <- "#e64173"
grey_light <- "grey70"
grey_mid <- "grey50"
grey_dark <- "grey20"
# Dark slate grey: #314f4f
# Notes directory
dir_slides <- "~/Dropbox/UO/Teaching/EC421W19/LectureNotes/02Review/"
# Knitr options
opts_chunk$set(
  comment = "#>",
  fig.align = "center",
  fig.height = 7,
  fig.width = 10.5,
  warning = F,
  message = F
)
# A blank theme for ggplot
theme_empty <- theme_bw() + theme(
  line = element_blank(),
  rect = element_blank(),
  strip.text = element_blank(),
  axis.text = element_blank(),
  plot.title = element_blank(),
  axis.title = element_blank(),
  plot.margin = structure(c(0, 0, -0.5, -1), unit = "lines", valid.unit = 3L, class = "unit"),
  legend.position = "none"
)
theme_simple <- theme_bw() + theme(
  line = element_blank(),
  panel.grid = element_blank(),
  rect = element_blank(),
  strip.text = element_blank(),
  axis.text.x = element_text(size = 18, family = "STIXGeneral"),
  axis.text.y = element_blank(),
  axis.ticks = element_blank(),
  plot.title = element_blank(),
  axis.title = element_blank(),
  # plot.margin = structure(c(0, 0, -1, -1), unit = "lines", valid.unit = 3L, class = "unit"),
  legend.position = "none"
)
theme_axes_math <- theme_void() + theme(
  text = element_text(family = "MathJax_Math"),
  axis.title = element_text(size = 22),
  axis.title.x = element_text(hjust = .95, margin = margin(0.15, 0, 0, 0, unit = "lines")),
  axis.title.y = element_text(vjust = .95, margin = margin(0, 0.15, 0, 0, unit = "lines")),
  axis.line = element_line(
    color = "grey70",
```

```
    size = 0.25,
    arrow = arrow(angle = 30, length = unit(0.15, "inches")
  )),
  plot.margin = structure(c(1, 0, 1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
  legend.position = "none"
)
theme_axes_serif <- theme_void() + theme(
  text = element_text(family = "MathJax_Main"),
  axis.title = element_text(size = 22),
  axis.title.x = element_text(hjust = .95, margin = margin(0.15, 0, 0, 0, unit = "lines")),
  axis.title.y = element_text(vjust = .95, margin = margin(0, 0.15, 0, 0, unit = "lines")),
  axis.line = element_line(
    color = "grey70",
    size = 0.25,
    arrow = arrow(angle = 30, length = unit(0.15, "inches")
  )),
  plot.margin = structure(c(1, 0, 1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
  legend.position = "none"
)
theme_axes <- theme_void() + theme(
  text = element_text(family = "Fira Sans Book"),
  axis.title = element_text(size = 18),
  axis.title.x = element_text(hjust = .95, margin = margin(0.15, 0, 0, 0, unit = "lines")),
  axis.title.y = element_text(vjust = .95, margin = margin(0, 0.15, 0, 0, unit = "lines")),
  axis.line = element_line(
    color = grey_light,
    size = 0.25,
    arrow = arrow(angle = 30, length = unit(0.15, "inches")
  )),
  plot.margin = structure(c(1, 0, 1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
  legend.position = "none"
)
```

This code Belongs to Edward Rubin:https://github.com/edrubin/EC421S20

```
set.seed(12345)
ggplot(data = tibble(
  x = runif(1e3, -3, 3),
  e = rnorm(1e3, 0, sd = 4 + 1.5 * x)
), aes(x = x, y = e)) +
geom_point(color = "darkslategrey", size = 2.75, alpha = 0.5) +
labs(x = "x", y = "u") +
theme_axes_math
```

```r
set.seed(12345)
ggplot(data = tibble(
  g = sample(c(F,T), 1e3, replace = T),
  x = runif(1e3, -3, 3),
  e = rnorm(1e3, 0, sd = 0.5 + 2 * g)
), aes(x = x, y = e, color = g, shape = g, alpha = g)) +
geom_point(size = 2.75) +
scale_color_manual(values = c("darkslategrey", red_pink)) +
scale_shape_manual(values = c(16, 1)) +
scale_alpha_manual(values = c(0.5, 0.8)) +
labs(x = "x", y = "u") +
theme_axes_math
```

```r
set.seed(12345)
# Data
gq_df <- tibble(
  x = runif(1e3, -3, 3),
  e = rnorm(1e3, 0, sd = 4 + 1.5 * x),
  y = 1 + 3 * x + e
)
# Quantiles
gq_x <- quantile(gq_df$x, probs = c(3/8, 5/8))
# Regressions
sse1 <- lm(y ~ x, data = gq_df %>% filter(x < gq_x[1])) %>%
  residuals() %>% raise_to_power(2) %>% sum()
sse2 <- lm(y ~ x, data = gq_df %>% filter(x > gq_x[2])) %>%
  residuals() %>% raise_to_power(2) %>% sum()
ggplot(data = gq_df, aes(x = x, y = e)) +
geom_point(color = "darkslategrey", size = 2.75, alpha = 0.5) +
labs(x = "x", y = "u") +
theme_axes_math
```

```
ggplot(data = gq_df, aes(
  x = x, y = e,
  color = cut(x, c(-Inf, gq_x, Inf)),
  alpha = cut(x, c(-Inf, gq_x, Inf)),
  shape = cut(x, c(-Inf, gq_x, Inf))
)) +
geom_vline(
  xintercept = gq_x,
  color = grey_mid,
  size = 0.25
) +
geom_point(size = 2.75) +
labs(x = "x", y = "u") +
scale_color_manual(values = c("darkslategrey", grey_mid, red_pink)) +
scale_shape_manual(values = c(19, 1, 19)) +
scale_alpha_manual(values = c(0.5, 0.8, 0.6)) +
theme_axes_math
```

OLS Estimation in the Presence of Heteroscedasticity

$$\hat{\beta}_2 = \frac{\sum_i (x_i y_i)}{\sum_i (x_i^2)} \tag{11.2.1}$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sum_i \left(x_i^2 \sigma_i^2\right)}{\sum_i (x_i^2)^2} \tag{11.2.2}$$

The Method of Generalized Least Squares

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \quad \text{and} \quad \frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \tag{11.3.3 and 11.3.4}$$

Detection of Heteroscedasticity

Goldfeld–Quandt Test

```
options(scipen = 999)

library(gujarati)

library(ggplot2)

library(lmtest)
fix(Table11_3)

attach(Table11_3)

MODEL1= lm(Table11_3$Y~Table11_3$X)
summary(MODEL1)

gqtest(MODEL1, order.by = ~Table11_3$X,fraction = 13)
```

Breusch–Pagan–Godfrey Test

$$p_i = \alpha_1 + \alpha_2 Z\_2i + \alpha_3 Z\_3i + \cdots + \alpha_m Z_{mi} + v_i \quad (11.5.15)$$

$$\text{LM} = n \times R_e^2$$

```
MODEL1= lm(Table11_3$Y~Table11_3$X)
summary(MODEL1)
bptest(MODEL1)

# For the White test

bptest(MODEL1,~fitted(MODEL1)+I(fitted(MODEL1)^2))
```

Remedial Measures
The Method of Weighted Least Squares

```
# Since the data in R are not compatible with the book,
# We can use wooldridge data.

library(wooldridge)

View(k401ksubs)

attach(k401ksubs)

#Using OLs

MODEL2=lm(nettfa~inc+I((age-25)^2)+male+e401k,subset=(fsize==1))

summary(MODEL2)

#Using WLS

MODEL3=lm(nettfa~inc+I((age-25)^2)+male+e401k,subset=(fsize==1),weights =
1/inc)

summary(MODEL3)

library(stargazer)

stargazer(MODEL2,MODEL3,type="text",column.labels = c("OLS","WLS"))

plot(inc,nettfa)

abline(MODEL2,lwd=3,lty=1,col="red")

abline(MODEL3,lwd=3,lty=3,col="green")

legend("topleft",c("OLS","WLS"),lty =c(1,3),bty = "n")


#When the weight is not known we can create by the process:

attach(smoke)

#Standard OLS

MODEL4=lm(cigs~lincome+lcigpric+educ+age+agesq+restaurn)

summary(MODEL4)

#Now we need residuals of previous regression

weight1=lm(log(residuals(MODEL4)^2)~lincome+lcigpric+educ+age+agesq+restaurn,
data = smoke)

MODEL5=lm(cigs~lincome+lcigpric+educ+age+agesq+restaurn,weights =
1/exp(fitted(weight1)),data = smoke)

summary(MODEL5)
```

```
stargazer(MODEL4,MODEL5,type="text",column.labels = c("OLS","WLS"))

#If we consider weight that we have chosen is wrong
#Then we can look at the Robust WLS.

library(lmtest)
library(sandwich)

robust3=vcovHC(MODEL7,type = "HC1")

MODEL3_Robust=coeftest(MODEL3,robust3)

stargazer(MODEL3,MODEL3_Robust,type = "text",column.labels = c("Non-Robust WLS","Robust WLS"))
```

Empirical Exercises

11.15

```
fix(Table11_7)


MODEL7 = lm(Table11_7$MPG ~ Table11_7$HP +Table11_7$WT+ Table11_7$SP)
summary(MODEL7)

bptest(MODEL7)

MODEL8 = lm(log(Table11_7$MPG) ~ log(Table11_7$HP) +log(Table11_7$WT)+ log(Table11_7$SP))
summary(MODEL8)

bptest(MODEL8)

#When we take log values, it becomes homoskedastic

stargazer(list(MODEL7,MODEL8),type = "text")
```

Econometrics-Damodar N. Gujarati / Chapter 12

Furkan Zengin

21 08 2021

Autocorrelation

$$E[u_i u_j] \neq 0 \tag{12.1.1}$$

$$\text{Cov}(u_i,\, u_j) \neq 0$$

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.1.8}$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \tag{12.1.9}$$

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \Delta u_t \tag{12.1.10}$$

OLS Estimation in the Presence of Autocorrelation

$$\text{Cov}(u_t,\, u_{t+s}) = E[u_t u_{t-s}] = \rho^s \frac{\sigma_\epsilon^2}{1-\rho^2} \tag{12.2.4}$$

The BLUE Estimator in the Presence of Autocorrelation

$$\hat{\beta}_2^{GLS} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + C \tag{12.3.1}$$

AR(1) Simulation

This code Belongs to Edward Rubin: https://github.com/edrubin/EC421S20

```
options(scipen = 999)

library(gujarati)

library(ggplot2)

library(tidyverse)

library(magrittr)


# Number of observations
T <- 1e2
# Rho
rho <- 0.95
# Set seed and starting point
set.seed(1234)
start <- rnorm(1)
# Generate the data
ar_df <- tibble(
  t = 1:T,
  x = runif(T, min = 0, max = 1),
  e = rnorm(T, mean = 0, sd = 2),
  u = NA
)
for (x in 1:T) {
  if (x == 1) {
    ar_df$u[x] <- rho * start + ar_df$e[x]
  } else {
    ar_df$u[x] <- rho * ar_df$u[x-1] + ar_df$e[x]
  }
}
ar_df %<>% mutate(y = 1 + 3 * x + u)
```

```
# Plot disturbances over time
ggplot(data = ar_df,
  aes(t, u)
) +
geom_line(color = "blue", size = 0.35) +
geom_point(color = "red", size = 2.25) +
ylab("u") +
xlab("t")
```

```
options(scipen = 999)

fix(Table12_4)

MODEL1 = lm(Table12_4$Y ~ Table12_4$X)
summary(MODEL1)

plot(Table12_4$X ,Table12_4$Y, col= "steelblue", xlab = "Productivity",
     ylab = "Wages" ,pch=20)

library(lmtest)

dwtest(MODEL1)
```

$$\hat{Y}_t = \underset{(1.39402)}{32.74190} + \underset{(0.01567)}{0.67041} X_t \tag{12.5.1}$$

$$R^2 = 0.9765 \qquad d = 0.17389$$

Detecting Autocorrelation

```
library(Hmisc)

RES1 = residuals(MODEL1)
LRES1 = Lag(RES1)


plot(RES1,type = "l",ylab = "Residuals and Lagged Residuals")
lines(LRES1, col = "red")
```

Durbin–Watson d Test

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{t=n} (\hat{u}_t)^2} \tag{12.6.5}$$

A General Test of Autocorrelation: The Breusch–Godfrey (BG) Test

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \epsilon_t \tag{12.6.15}$$

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_p = 0 \tag{12.6.16}$$

```
MODEL1 = lm(Table12_4$Y ~ Table12_4$X)
summary(MODEL1)

bgtest(MODEL1 ,type = c("Chisq", "F"),data = Table12_4)
```

Remedial Measures

```
MODEL1 = lm(Table12_4$Y ~ Table12_4$X)

library(prais)
library(orcutt)
prais_winsten(Table12_4$Y ~ Table12_4$X,data = Table12_4)

cochrane.orcutt(MODEL1)
```

The Newey–West Method

```
MODEL1 = lm(Table12_4$Y ~ Table12_4$X)


NW <- NeweyWest(lm(Table12_4$Y ~ Table12_4$X),
              lag = 4)

coeftest(MODEL1, vcov = NW_VCOV)
```

Empirical Exercises

12.26

```
fix(Table12_7)


MODEL2 = lm(log(Table12_7$C) ~ log(Table12_7$I) +log(Table12_7$L)+ log(Table12_7$H) +
              log(Table12_7$A))
summary(MODEL2)

RES2 = resid(MODEL2)


plot(RES2,type = "l",ylab = "Residuals")


dwtest(MODEL2)

bgtest(MODEL2) #can bu used to detect autocorrelation and can be decided.
```

Econometrics-Damodar N. Gujarati / Chapter 13

Furkan Zengin

22 08 2021

Econometric Modeling: Model Specification and Diagnostic Testing

Types of Specification Errors

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i} \tag{13.2.1}$$

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \tag{13.2.2}$$

$$u_{2i} = u_{1i} + \beta_4 X_i^3 \tag{13.2.3}$$

	$\text{Cov}(x_1, x_2) > 0$	$\text{Cov}(x_1, x_2) < 0$
$\beta_2 > 0$	Upward	Downward
$\beta_2 < 0$	Downward	Upward

This code Belongs to Edward Rubin: https://github.com/edrubin/EC421S20

```
library(tidyverse)
# Set seed
set.seed(12345)
# Sample size
n <- 1e3
# Parameters
beta0 <- 20; beta1 <- 0.5; beta2 <- 10
# Dataset
omit_df <- tibble(
  male = sample(x = c(F, T), size = n, replace = T),
  school = runif(n, 3, 9) - 3 * male,
  pay = beta0 + beta1 * school + beta2 * male + rnorm(n, sd = 7)
)
lm_bias <- lm(pay ~ school, data = omit_df)
bb0 <- lm_bias$coefficients[1] %>% round(1)
bb1 <- lm_bias$coefficients[2] %>% round(1)
lm_unbias <- lm(pay ~ school + male, data = omit_df)
bu0 <- lm_unbias$coefficients[1] %>% round(1)
bu1 <- lm_unbias$coefficients[2] %>% round(1)
bu2 <- lm_unbias$coefficients[3] %>% round(1)
```

```r
ggplot(data = omit_df, aes(x = school, y = pay)) +
geom_point(size = 2.5, color = "black", alpha = 0.4, shape = 16) +
geom_hline(yintercept = 0) +
geom_vline(xintercept = 0) +
xlab("Schooling") +
ylab("Pay") +
theme(
  axis.title = element_text(size = 18),
  plot.margin = structure(c(0, 0, 0.1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
)
```

```r
ggplot(data = omit_df, aes(x = school, y = pay)) +
geom_point(size = 2.5, alpha = 0.8, aes(color = male, shape = male)) +
geom_hline(yintercept = 0) +
geom_vline(xintercept = 0) +
geom_line(stat = "smooth", color = "orange", method = lm, alpha = 0.5, size = 1) +
xlab("Schooling") +
ylab("Pay") +
theme(
  axis.title = element_text(size = 18),
  plot.margin = structure(c(0, 0, 0.1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
) +
scale_color_manual("", values = c("red", "darkslategrey"), labels = c("Female", "Male")) +
scale_shape_manual("", values = c(16, 1), labels = c("Female", "Male"))
```

EXAMPLE 13.1

```
library(gujarati)

options(scipen = 999)

fix(Table6_4)

MODEL1 = lm(Table6_4$CM ~ Table6_4$FLR + Table6_4$PGNP) #Unbiased

summary(MODEL1)


MODEL1_1 = lm(Table6_4$CM ~ Table6_4$PGNP) # Biased

summary(MODEL1_1)




bb0 <- MODEL1_1$coefficients[1] %>% round(1)
bb1 <- MODEL1_1$coefficients[3] %>% round(1)

bu0 <- MODEL1$coefficients[1] %>% round(1)
bu1 <- MODEL1$coefficients[2] %>% round(1)
bu2 <- MODEL1$coefficients[3] %>% round(1)
```

```
ggplot(data = Table6_4, aes(x = Table6_4$PGNP, y = Table6_4$CM)) +
geom_point(size = 2.5, color = "red", alpha = 0.9, shape = 16) +
geom_hline(yintercept = 0) +
geom_vline(xintercept = 0) +
xlab("Income") +
ylab("Child Mortality") +
theme(
  axis.title = element_text(size = 18),
  plot.margin = structure(c(0, 0, 0.1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
)
```

```
ggplot(data = Table6_4, aes(x = log(Table6_4$PGNP), y = Table6_4$CM)) +
geom_point(size = 2.5, alpha = 0.8, aes(color = Table6_4$FLR, Table6_4$FLR)) +
geom_hline(yintercept = 0) +
geom_vline(xintercept = 0)  +
xlab("Income") +
ylab("Child Mortality") +
theme(
  axis.title = element_text(size = 18),
  plot.margin = structure(c(0, 0, 0.1, 0), unit = "lines", valid.unit = 3L, class = "unit"),
)
```

$$\mathrm{Var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \qquad \text{(13.3.3 and 13.3.4)}$$

$$\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2}\mathrm{VIF}$$

```
MODEL1 = lm(Table6_4$CM ~ Table6_4$FLR + Table6_4$PGNP) #Unbiased

summary(MODEL1)

library(car)

vif(MODEL1)
```

Tests of Specification Errors

Residuals and Durbin Watson

```
fix(Table7_4)

x1 = Table7_4$X
x2 = (Table7_4$X)^2
x3 = (Table7_4$X)^3


MODEL2 = lm(Table7_4$Y ~ x1 + x2 +x3)

MODEL2_1 = lm(Table7_4$Y ~ x1 + x2)

MODEL2_2 = lm(Table7_4$Y ~ x1)


RES1 = resid(MODEL2)
RES1_1 = resid(MODEL2_1)
RES1_2 = resid(MODEL2_2)

par(mfrow=c(2,2))
plot(RES1,type = "l")
plot(RES1_1,type = "l")
plot(RES1_2,type = "l")

library(lmtest)

dwtest(MODEL2)

dwtest(MODEL2_1)

dwtest(MODEL2_2)
```

Ramsey's RESET Test

```
library(fRegression)
lmTest(MODEL2,method = "reset")

resettest(MODEL2,order = 2:3,type = "fitted")
```

Errors of Measurement

$$Y_i = \alpha + \beta X_i^* + u_i \implies \text{Correct Model}$$

$$X_i = X_i^* + w_i \quad \text{where wi represents errors of measurement in}$$

$$\text{cov}(z_i, X_i) = -\beta\sigma_w^2$$

(z_i) is a compound of equation and measurement errors.

$$\text{plim}\hat{\beta} = \beta\left[\frac{1}{1+\frac{\sigma_w^2}{\sigma_X^2}}\right]$$

```
fix(Table13_2)

MODEL3 = lm(Table13_2$Y. ~Table13_2$X.)
summary(MODEL3)

MODEL3_1 = lm(Table13_2$Y ~Table13_2$X)
summary(MODEL3_1)
```

Davidson–MacKinnon J Test

$$

	Hypothesis : $\alpha_4 = 0$	
Hypothesis : $\beta_4 = 0$	**Do not Reject**	**Reject**
Do not Reject	Accept Both C and D	Accept D, Reject C
Reject	Accept C, Reject D	Reject Both C and D

$$

```
fix(Table13_3)

library(dynlm)

NEW1 = ts()

MODEL4 = dynlm(ts(Table13_3$PPCE) ~ ts(Table13_3$PDPI) + L(ts(Table13_3$PDPI))
               ,data = Table13_3)

summary(MODEL4)

MODEL5 = dynlm(ts(Table13_3$PPCE) ~ ts(Table13_3$PDPI) + L(ts(Table13_3$PPCE))
               ,data = Table13_3)

summary(MODEL5)

library(stargazer)

stargazer(list(MODEL4,MODEL5),type = "text")

Fit1 = fitted(MODEL4)

Fit2 = fitted(MODEL5)


MODEL4_1 = dynlm(ts(Table13_3$PPCE) ~ ts(Table13_3$PDPI) + L(ts(Table13_3$PDPI))+
                   Fit2)

summary(MODEL4_1)

model5_1 = dynlm(ts(Table13_3$PPCE) ~ ts(Table13_3$PDPI) + L(ts(Table13_3$PPCE))+
                   Fit1)
summary(model5_1)
```

```
fix(Table10_7)

MODEL6 = lm(log(Table10_7$C) ~ log(Table10_7$Yd) + log(Table10_7$W) + Table10_7$I)

summary(MODEL6)

dwtest(MODEL6)

MODEL7 = lm(log(Table10_7$C) ~ log(Table10_7$Yd) + log(Table10_7$W) + Table10_7$I +
            log(Table10_7$Yd)*log(Table10_7$W))

summary(MODEL7)

dwtest(MODEL7)

bgtest(MODEL7)

library(sandwich)

NW <- NeweyWest(MODEL7,
 lag = 4)
coeftest(MODEL7, vcov = NW)

library(strucchange)

sctest(MODEL6, data = Table10_7,
  type = "Chow", point = 44)
```

Econometrics-Damodar N. Gujarati / Chapter 16

Furkan Zengin

23 08 2021

Panel Data Regression Model

$$C_{it} = \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \tag{16.3.1}$$
$$i = 1, 2 \ldots, 6$$
$$t = 1, 2, \ldots, 15$$

Pooled OLS Regression

```
options(scipen = 999)

library(gujarati)

fix(Table16_1)

library(dynlm)

library(lmtest)

library(sandwich)

library(stargazer)

library(plm)

pdata.frame(Table16_1)

MODEL1 = plm(Table16_1$C ~ Table16_1$Q + Table16_1$PF + Table16_1$LF,data = Table16_1,
             index = c("T"),
                    model = "pooling")

summary(MODEL1)
```

```
library(gplots)

coplot(log(Table16_1$C)~Table16_1$Q|Table16_1$I ,data = Table16_1,xlab = "Output",
       ylab = "Total cost",type ="b")


plotmeans(log(Table16_1$C) ~ Table16_1$I, main="Heterogeineity across Airlines",
          xlab = "Output",
       ylab = "Total cost",data=Table16_1)
```

```
lin = lm(log(Table16_1$C) ~ Table16_1$Q)
quad =  lm(log(Table16_1$C) ~ Table16_1$Q+ I(Table16_1$Q^2))

plot(Table16_1$Q, log(Table16_1$C),
     col  = "steelblue",
     pch = 20,
     xlab = "Output",
     ylab = "Total Cost",
     )


abline(lin, col = "black", lwd = 2)

order_id <- order(Table16_1$Q)

lines(x = Table16_1$Q[order_id],
      y = fitted(quad)[order_id],
      col = "red",
      lwd = 2)
```

```
MODEL2 = plm(Table16_1$C ~ Table16_1$Q + Table16_1$PF + Table16_1$LF +
               factor(Table16_1$I) ,data = Table16_1,
               index = c("T"),
               model = "pooling")
summary(MODEL2)

library(car)

yhat = fitted(MODEL2)

scatterplot(yhat ~ Table16_1$Q|Table16_1$I, boxplots=FALSE, xlab="x1"
            , ylab="yhat",smooth=FALSE)

abline(lm(Table16_1$C ~ Table16_1$Q),lwd=3, col="red")
```

First-Difference Method

```
MODEL3 = plm(Table16_1$C ~ Table16_1$Q + Table16_1$PF + Table16_1$LF +
               factor(Table16_1$I) ,data = Table16_1,
               index = c("T"),
               model = "fd")
summary(MODEL3)
```

$$TC_{it} = \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + w_{it} \quad \text{(16.6.3 and 16.6.5)}$$

$$\text{where} \quad w_i = \epsilon_i + u_{it}$$

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

$$u_{it} \sim \mathcal{N}(0, \sigma_u^2)$$

The correlation coefficient is: (16.6.8)

$$\rho = \text{corr}(w_{it}, w_{is}) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2} \; ; t \neq s$$

```
MODEL4 = plm(Table16_1$C ~ Table16_1$Q + Table16_1$PF + Table16_1$LF,
                data = Table16_1,
                index = c("T","I"),
                model = "random")
summary(MODEL4)

phtest(MODEL2, MODEL4)

plmtest(MODEL2, c("time"), type=("bp"))
```

A useful website: https://www.princeton.edu/~otorres/Panel101R.pdf

Econometrics-Damodar N. Gujarati / Chapter 17

Furkan Zengin

24 08 2021

Dynamic Econometric Models: Autoregressive and Distributed-Lag Models

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \qquad \text{(Distributed-Lag Model)}$$

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t \qquad \text{(Autoregressive Model)}$$

Estimation of Distributed-Lag Models

```
#We can use the data in Table17_2

options(scipen = 999)

library(gujarati)

fix(Table17_2)

library(dynlm)

library(lmtest)

library(sandwich)

library(stargazer)

library(car)


MODEL1 = dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI) ,data = Table17_2)


MODEL1_1 = dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI)+ L(ts(Table17_2$PPDI))
                 ,data = Table17_2)


MODEL1_2 =dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI)+ L(ts(Table17_2$PPDI))+
                  L(ts(Table17_2$PPDI,2)) ,data = Table17_2)
```

```
MODEL1_3 = dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI)+ L(ts(Table17_2$PPDI))+
                  L(ts(Table17_2$PPDI,2))+ L(ts(Table17_2$PPDI,3))  ,data = Table17_2)


stargazer(list(MODEL1,MODEL1_1,MODEL1_2,MODEL1_3),type = "text")
```

The Koyck Approach to Distributed-Lag Models

$$\beta_k = \beta_0 \lambda^k \tag{17.4.1}$$

$$\sum_{k=0}^{\infty} \beta_k = \beta_0 (\frac{1}{1-\lambda}) \tag{17.4.2}$$

```
MODEL2 =  dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI) + L(ts(Table17_2$PPCE))
                ,data = Table17_2)

summary(MODEL2)

func = function(lambda,beta0) {beta0*(1/(1-lambda))}

func(0.797150,0.21389)
```

Koyck (17.4.7)

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

Adaptive expectations (17.5.5)

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + [u_t - (1-\gamma)u_{t-1}]$$

Partial adjustment (17.4.7)

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta u_t$$

Detecting Autocorrelation in Autoregressive Models: Durbin h Test

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[var(\hat{\alpha}_2)]}} \qquad (17.10.1)$$

```
MODEL2 =  dynlm(ts(Table17_2$PPCE)~  ts(Table17_2$PPDI) + L(ts(Table17_2$PPCE))
                ,data = Table17_2)

dwtest(MODEL2)

func2 = function(n,rho,var) {rho* sqrt(n/(1-n*var))}

1- (0.95862/2) # 1 - d/2 = rho

func2(47,0.52069,0.0053)




NW <- NeweyWest(MODEL2,
 lag = 3)
coeftest(MODEL2, vcov = NW)
```

17.11 The Demand for Money in Canada

$$M_t^* = \beta_0 R_t^{\beta_1} Y_t^{\beta_2} + e_t^{u_t} \quad (17.11.11)$$

```
options(scipen = 999)

fix(Table17_5)

MODEL3 = dynlm(ts(log(Table17_5$M1)) ~ log(ts(Table17_5$R)) + log(ts(Table17_5$GDP))+
                L(log(ts(Table17_5$M1))))

summary(MODEL3)
```

Causality in Economics: The Granger Causality Test

$$GDP_t = \sum_{i=1}^{n} \alpha_i M_{t-i} + \sum_{j=1}^{n} \beta_j GDP_{t-j} + u_{1t} \tag{17.14.1}$$

$$M_t = \sum_{i=1}^{n} \lambda_i M_{t-i} + \sum_{j=1}^{n} \delta_j GDP_{t-j} + u_{2t} \tag{17.14.2}$$

```
MODEL2_a =  dynlm(ts(Table17_5$M1) ~  ts(Table17_5$R), data = Table17_5)



gt1 = grangertest(ts(Table17_5$M1) ~  ts(Table17_5$R),order = 2)
gt2 = grangertest(ts(Table17_5$M1) ~  ts(Table17_5$R),order = 4)
gt3 = grangertest(ts(Table17_5$M1) ~  ts(Table17_5$R),order = 6)
gt4 = grangertest(ts(Table17_5$M1) ~  ts(Table17_5$R),order = 8)

rgt1 = grangertest(ts(Table17_5$R) ~  ts(Table17_5$M1),order = 2)
rgt2 = grangertest(ts(Table17_5$R) ~  ts(Table17_5$M1),order = 4)
rgt3 = grangertest(ts(Table17_5$R) ~  ts(Table17_5$M1),order = 6)
rgt4 = grangertest(ts(Table17_5$R) ~  ts(Table17_5$M1),order = 8)

Fgt1 = gt1$F
Fgt2 = gt2$F
Fgt3 = gt3$F
Fgt4 = gt4$F

Frgt1 = rgt1$F
Frgt2 = rgt1$F
Frgt3 = rgt1$F
Frgt4 = rgt1$F


df = data.frame(Fgt1,Frgt1,Fgt2,Frgt2,Fgt3,Frgt3,Fgt4,Frgt4)
df
```

Econometrics-Damodar N. Gujarati / Chapter 21

Furkan Zengin

25 08 2021

Time Series Econometrics

By using the database https://fred.stlouisfed.org/graph/?g=Gk2X

```
options(scipen = 999)

library(gujarati)

library(dynlm)

library(lmtest)

library(sandwich)

library(stargazer)

library(car)

attach(fredgraph)

library(latticeExtra)

fix(fredgraph)


xyplot(log(fredgraph$GDP) + log(fredgraph$DPI) + log(fredgraph$PCE)  ~ fredgraph$date, fredgraph, type = "l",
col=c("steelblue", "#69b3a2","red") , lwd=2,ylab = "LPCE,LGDP,LDPI",xlab = "Time")


plot(fredgraph$date , log(fredgraph$GDP),
     type = "l",
     col = 2,
```

```
     xlab = "Year",
     ylab = "Billion of Dollars Logged")
lines(fredgraph$date , log(fredgraph$DPI),
      type = "l",
      col = 3)
lines(fredgraph$date , log(fredgraph$PCE),
      type = "l",
      col = 4)
legend("topleft",
       c("LGDP", "LDPI", "LPCE"),
       lty = 1,
       col = 2:4)
```

```
#RUN THIS CODE AS THE LAST ONE !!!
library(AER)
library(forecast)
library(scales)
library(quantmod)
library(urca)

tsfred = ts(fredgraph)

fredgraph$date = as.Date(fredgraph$date)

Lgdp <- xts(log(fredgraph$GDP), fredgraph$date)["1959::2019"]

Lpce <- xts(log(fredgraph$PCE), fredgraph$date)["1959::2019"]




plot(merge(as.zoo(Lgdp), as.zoo(Lpce)),
     plot.type = "single",
     col = c("darkred", "steelblue"),
     lwd = 2,
     xlab = "Date",
     ylab = "GDP and PCE",
     main = "Logged GDP and PCE")


YToYQTR <- function(years) {
  return(
      sort(as.yearqtr(sapply(years, paste, c("Q1", "Q2", "Q3", "Q4"))))
  )
```

```
}


recessions <- YToYQTR(c(1961:1962, 1970, 1974:1975, 1980:1982, 1990:1991, 2001, 2007:2008,2019:2020))


plot(merge(as.zoo(Lgdp), as.zoo(Lpce)),
     plot.type = "single",
     col = c("darkred", "steelblue"),
     lwd = 2,
     xlab = "Date",
     ylab = "GDP and PCE",
     main = "Logged GDP and PCE")


xblocks(time(as.zoo(Lgdp)),
        c(time(Lgdp) %in% recessions),
        col = alpha("steelblue", alpha = 0.3))


legend("topleft",
       legend = c("LGDP", "LPCE"),
       col = c("darkred", "steelblue"),
       lwd = c(2, 2))
```

Stochastic Processes

$$\text{Mean} = \mathrm{E}(Y_t) = \mu \tag{21.3.1}$$

$$\text{Variance} = Var(Y_t) = \mathrm{E}(Y_t - \mu)^2 = \sigma^2 \tag{21.3.2}$$

$$\text{Covariance} = \gamma_k = \mathrm{E}[(Y_t - \mu)(Y_{t+k} - \mu)] \tag{21.3.2}$$

Nonstationary Stochastic Processes

Random Walk without Drift

$$Y_t = Y_0 + \sum u_t \tag{21.3.5}$$

$$Var(Y_t) = t\sigma^2 \tag{21.3.7}$$

Random Walk with Drift

$$Y_t = \delta + Y_{t-1} + u_t \tag{21.3.9}$$

$$
\begin{aligned}
\operatorname{Var}(Y_t) &= \operatorname{Var}(Y_{t-1} + \varepsilon_t) \\
&= \operatorname{Var}(Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\
&= \operatorname{Var}(Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\
&\cdots \\
&= \operatorname{Var}(Y_0 + \varepsilon_1 + \cdots + \varepsilon_{t_2} + \varepsilon_{t-1} + \varepsilon_t) \\
&= \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \\
&= t\sigma_\varepsilon^2
\end{aligned}
$$

Below code belongs to Edward Rubin : https://github.com/edrubin/EC421S20

```
library(tidyverse)

library(ggplot2)

set.seed(1246)
walk1 <- tibble(x = cumsum(rnorm(1e2)), t = 1:1e2, walk = "1")
walk2 <- tibble(x = cumsum(rnorm(1e2)), t = 1:1e2, walk = "2")
walk3 <- tibble(x = cumsum(rnorm(1e2)), t = 1:1e2, walk = "3")
walk4 <- tibble(x = cumsum(rnorm(1e2)), t = 1:1e2, walk = "4")
walk5 <- tibble(x = cumsum(rnorm(1e2)), t = 1:1e2, walk = "5")
ggplot(data = walk1, aes(x = t, y = x)) +
  geom_hline(yintercept = 0, color = "red", size = 1.25) +
  geom_path()
```

```r
library(viridis)

ggplot(data = bind_rows(walk1, walk2), aes(x = t, y = x, color = "blue")) +
  geom_hline(yintercept = 0, color = "red", size = 1.25) +
  geom_path() +
  scale_color_viridis(option = "magma", discrete = T, begin = 0.15, end = 0.85)
```

```r
ggplot(data = bind_rows(walk1, walk2, walk3, walk4, walk5), aes(x = t, y = x, color = walk)) +
  geom_hline(yintercept = 0, color = "grey85", size = 1.25) +
  geom_path() +
  scale_color_viridis(option = "magma", discrete = T, begin = 0.15, end = 0.85)
```

Unit Root Stochastic Process

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \tag{21.4.1}$$

Tests of Stationarity

2. Autocorrelation Function (ACF) and Correlogram

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{21.8.1}$$

$$\hat{\gamma}_k = \frac{\sum(Y_t - \bar{Y}_t)(Y_{t+k} - \bar{Y}_t)}{n} \tag{21.8.2}$$

$$\hat{\gamma}_0 = \frac{\sum(Y_t - \bar{Y}_t)^2}{n} \tag{21.8.3}$$

```
library(forecast)

acf(fredgraph$GDP, lag.max = 4, plot = F)

acf(log(fredgraph$GDP), lag.max = 4, plot = F)

ggAcf(fredgraph$GDP,24)
```

The Unit Root Test

```
#It is normal that we get different estimation from the book since period are not same

t = seq(1,61,1)

fredgraph$lgdp=log(fredgraph$GDP)

fredgraph1 = ts(fredgraph)

attach(fredgraph1)


MODEL1 = dynlm(diff(ts(log(fredgraph$GDP))) ~  L(ts(log(fredgraph$GDP))))

summary(MODEL1)

MODEL2 = dynlm(diff(ts(log(fredgraph$GDP))) ~ ts(t)+  L(ts(log(fredgraph$GDP))))

summary(MODEL2)
```

The Augmented Dickey–Fuller (ADF) Test

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-1} + \epsilon_t$$

```
adf.test(fredgraph$GDP, k = 3)

adf.test(log(fredgraph$GDP), k = 3)
```

Difference-Stationary Processes

```
MODEL3 = dynlm(diff(ts(log(fredgraph$GDP))) ~  L(diff(ts(log(fredgraph$GDP)))))

summary(MODEL3)

dlgdp = diff(log(fredgraph$GDP))

t2 = seq(2,61,1)

plot(t2, dlgdp,type = "l")
```

Cointegration

```
summary(ur.df(fredgraph$GDP , type = c("trend"), selectlags="AIC"))
```

EXAMPLE 21.3

By using the data: https://fred.stlouisfed.org/graph/?g=Gk7J

```
plot(CPIDATA$DATE,CPIDATA$CPI,type = "l" ,xlab= ("Time") , ylab=("CPI"))

t3 = seq(1,74,1)

library(dynlm)

MODEL4 = dynlm(diff(ts(CPIDATA$CPI)) ~ ts(t3) + L(ts(CPIDATA$CPI))
               + diff(L(ts(CPIDATA$CPI))))

summary(MODEL4)
```

$$\Delta C\hat{P}I_t = \underset{(0.49007)}{-0.51462} + \underset{(0.04360)}{0.14696}t - \underset{(0.01111)}{0.03176}CPI_{t-1} + \underset{(0.09792)}{0.51022}\Delta CPI_{t-1} \tag{21.12.2}$$

Econometrics-Damodar N. Gujarati / Chapter 22

Furkan Zengin

26 08 2021

Time Series Econometrics:Forecasting

AR, MA, and ARIMA Modeling of Time Series Data

$$\text{Autoregressive (AR) Process} \quad (22.2.1)$$
$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t$$

$$\text{Pth-order Autoregressive - AR(p)} \quad (22.2.3)$$
$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t$$

$$\text{Moving Average (MA) Process} \quad (22.2.4)$$
$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$$

Autoregressive and Moving Average (ARMA) Process (22.2.4)

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1}$$

The Box–Jenkins (BJ) Methodology

Step 1: Identification

```
options(scipen = 999)

library(gujarati)

library(dynlm)

library(vars)

library(fGarch)

library(quantmod)

library(scales)

library(forecast)

ggAcf(fredgraph$GDP,24,demean = TRUE)


ggtaperedacf(fredgraph$GDP,
  lag.max = 24,
  type = c("correlation", "partial"),
  plot = TRUE,
  calc.ci = TRUE,
  level = 95
)

ggAcf(fredgraph$GDP,24,type = "partial")
```

Simulation : FIGURE 22.4

This code belongs to Hüseyin Taştan:https://github.com/htastan

```
library(ggplot2)

n <- 1000
set.seed(123)
MD1 <- ts(rnorm(n,0,1))
autoplot(MD1)

ggAcf(MD1)




MDL <- stats::lag(MD1, -1)
x = MD1 + 0.5* MDL
autoplot(x)




set.seed(1234)
# define the lists for the ARIMA(p,d,q) models
# order = c(1, 0, 0) means ARIMA(1,0,0) = AR(1)
# ar is the AR coefficient and sd is the standard deviation
list1 <- list(order = c(1, 0, 0), ar = 0.5, sd = 1)
list2 <- list(order = c(1, 0, 0), ar = 0.8, sd = 1)
list3 <- list(order = c(1, 0, 0), ar = 0.9, sd = 1)
list4 <- list(order = c(1, 0, 0), ar = 0.95, sd = 1)
#
AR1_1 <- arima.sim(n = 500, model = list1)
AR1_2 <- arima.sim(n = 500, model = list2)
AR1_3 <- arima.sim(n = 500, model = list3)
```

```
AR1_4 <- arima.sim(n = 500, model = list4)
#autoplot(AR1_1)



plot1 <- autoplot(AR1_1) + xlab("") + ggtitle("AR(1) = 0.5")
plot2 <- autoplot(AR1_2) + xlab("") + ggtitle("AR(1) = 0.8")
plot3 <- autoplot(AR1_3) + xlab("") + ggtitle("AR(1) = 0.9")
plot4 <- autoplot(AR1_4) + xlab("") + ggtitle("AR(1) = 0.95")
library(grid)
library(gridExtra)
grid.arrange(grobs=list(plot1, plot2, plot3, plot4),
             ncol=2, top="Simulated AR(1) Processes")
```

Step 2 and Step 3 : Estimation of the ARIMA Model
Diagnostic Checking

```
MODEL1 = dynlm(log(ts(fredgraph$GDP)) ~  L(log(ts(fredgraph$GDP))))

summary(MODEL1)

MODEL1_1 = dynlm(diff(log(ts(fredgraph$GDP))) ~  L(MODEL1$residuals) +L(MODEL1$residuals,2))

summary(MODEL1_1)

RES1 = MODEL1_1$residuals

ggAcf(RES1,25)
```

Step 4 : Forecasting

```
library(stargazer)

tsdata = ts(fredgraph,start = 1959)

MODEL2 = dynlm(GDP ~  L(GDP),data = tsdata, end = 2008)

MODEL2_1 = dynlm(GDP ~  PCE + L(GDP),data = tsdata, end = 2008)


stargazer(MODEL2, MODEL2_1 ,type="text", keep.stat=c("n","adj.rsq","ser"))



PRED <- predict(MODEL2, newdata=window(tsdata,start=2009), interval="prediction")

PRED2 <- predict(MODEL2_1, newdata=window(tsdata,start=2009), interval="prediction")


gdp <- ts(fredgraph$GDP, start=1959)

AR1 <- ts(PRED, start=2009)

autoplot(gdp) + autolayer(AR1) +geom_point(aes(y=gdp)) +
  geom_vline(xintercept = 2009, linetype=2) +
  ggtitle("GDP Forecasts for 2009-2019 using AR(1) Model")


AR2 = ts(PRED2, start=2009)
autoplot(gdp) + autolayer(AR2) +geom_point(aes(y=gdp)) +
```

```
  geom_vline(xintercept = 2009, linetype=2) +
  ggtitle("GDP Forecasts for 2009-2019 using AR(1) Model")



gdpF <- forecast(fredgraph$GDP, h=30)

plot(gdpF)

y  <- window(tsdata,start=2009)[,"GDP"]
PRED <- predict( MODEL2, newdata=window(tsdata,start=2009) )
PRED1 <- predict( MODEL2_1, newdata=window(tsdata,start=2009) )


matplot(time(y), cbind(y,PRED,PRED1), type="l",  col="black",lwd=2,lty=1:3)
legend("topleft",c("GDP","Forecast 1","Forecast 2"),lwd=2,lty=1:3)
```

Vector Autoregression (VAR)

$$M_{1t} = \alpha + \sum_{j=1}^{k} \beta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{1t} \tag{22.9.1}$$

$$R_t = \alpha' + \sum_{j=1}^{k} \theta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{2t} \tag{22.9.2}$$

```
fix(Table17_5)

library(fpp2)

date1 = ts(data = 1979:1988, start = c(1979,1), end = c(1988,4), frequency = 4)

date11 <- as.yearqtr(date1, format = "%Y:0%q")

newdata = data.frame(Table17_5,date11)


TGDP <- ts(newdata$GDP,
          start = c(1980, 1),
          end = c(1987, 4),
          frequency = 4)

TM1 <- ts(newdata$M1,
          start = c(1980, 1),
          end = c(1987, 4),
          frequency = 4)

TR <- ts(newdata$R,
          start = c(1980, 1),
          end = c(1987, 4),
          frequency = 4)


VAR_data <- window(ts.union(TM1, TR), start = c(1980, 1), end = c(1987, 4))
```

```
VAR_est <- VAR(y = VAR_data, p = 4,type = "none",ic="AIC")

summary(VAR_est)



forecast(VAR_est) %>%
  autoplot() +
  xlab("year")



causality(VAR_est, cause = "TR")

causality(VAR_est, cause = "TM1")
```