Home assignment

Alessandro Bramucci

Contents

1) Data description and data manipulation **1**

2) Data visualization **2**

3) Regression **3**

1) Data description and data manipulation

The raw data downloaded from the FRED database cover the gross domestic product and unemployment rate of the United States from 1950 to 2008. Data for GDP are annual while data for the unemployment rate are monthly. We will therefore have to transform the latter into an annual value. First let's load the raw data into R using the *read.csv* function.[1]

```
GDPCA <- read.csv("GDPCA.csv",
                  header = TRUE,
                  sep = ",")

UNRATE <- read.csv("UNRATE.csv",
                   header = TRUE,
                   sep = ",")
```

We now need to create the variables for the regression. To do this we will use the functions of two widely used R packages, *tidyr* and *dplyr*. We begin by calculating the average annual unemployment rate. If you are new to the packages or you keep forgetting stuff like me, it is good to take a look at the cheat sheet of the packages.

```
# We create annual averages of unemployment
UNRATE_ready <- UNRATE %>%
  separate(DATE, c("Year", "Month", "Day")) %>%
  group_by(Year) %>%
  summarise(UNRATE_avg = mean(UNRATE))

# We remove month and day from DATE
GDPCA_ready <- GDPCA %>%
  separate(DATE, c("Year", "Month", "Day")) %>%
  select(GDPCA, Year)
```

We are now ready to join the two data sets; we will use the variable Year as key. All that remains to do is to calculate the variables for the regression.

```
# Join the two data sets
Okun <- inner_join(GDPCA_ready, UNRATE_ready, by = "Year")
```

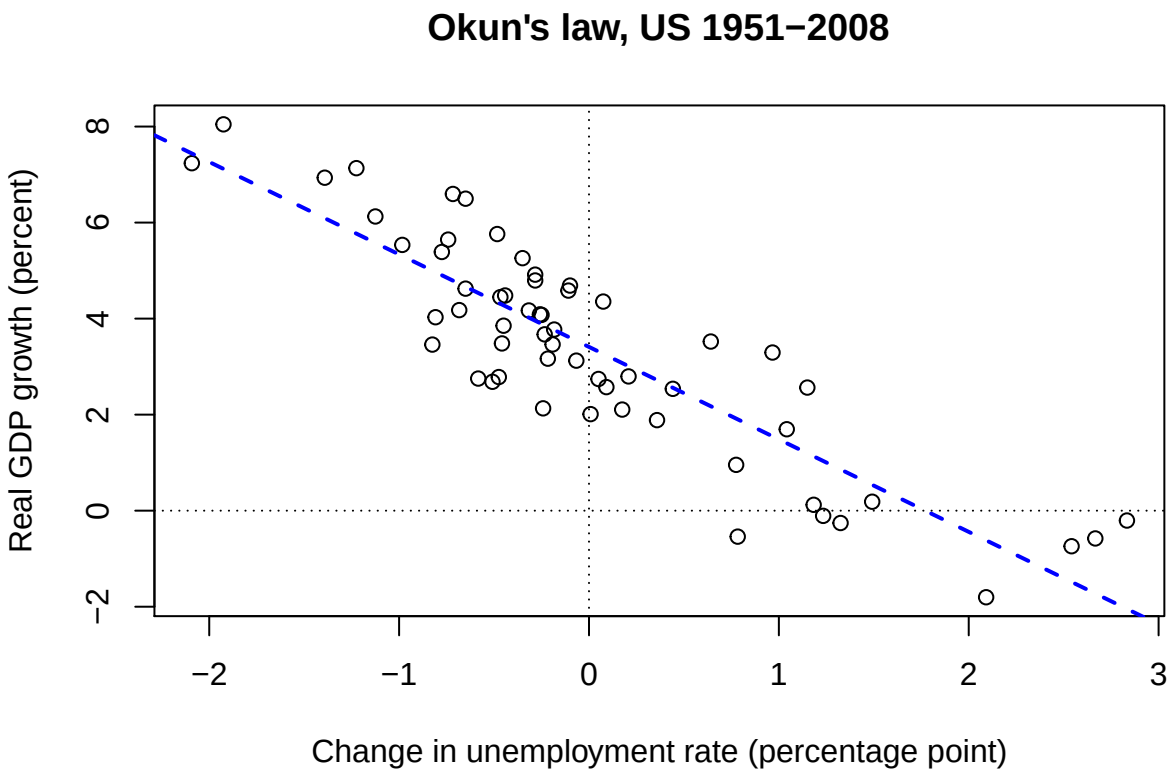
[1] This function is not the only one for importing csv into R. It is not even the most up-to-date but it's good enough for now.

```r
# First difference of the unemployment rate
Okun <- Okun %>%
  mutate(dUNRATE_avg = c(NA, diff(UNRATE_avg)))

# Real GDP growth rate
Okun <- Okun %>%
  mutate(GDPCA_gr = c(NA, diff(GDPCA)) / lag(GDPCA, shift = 1) * 100)

# Remove the first row because of NAs
Okun <- Okun[-1,]

# Let's take a look at the data
head(Okun, 2L)
##      GDPCA Year UNRATE_avg dUNRATE_avg GDPCA_gr
## 2 2475.449 1951   3.283333  -1.9250000 8.045791
## 3 2576.658 1952   3.025000  -0.2583333 4.088511
tail(Okun, 2L)
##       GDPCA Year UNRATE_avg dUNRATE_avg  GDPCA_gr
## 58 15623.87 2007   4.616667 0.008333333 2.0105063
## 59 15642.96 2008   5.800000 1.183333333 0.1221912
```

2) Data visualization

In this section, we are going to visualize our data with a scatter plot. In the horizontal axis we report the change in unemployment while in the vertical axis we report the growth rate in real GDP. We also introduce the regression line. In the next section, we will report the estimates of intercept and slope.

```r
plot(Okun$dUNRATE_avg, Okun$GDPCA_gr,
     main = "Okun's law, US 1951-2008",
     xlab = "Change in unemployment rate (percentage point)",
     ylab = "Real GDP growth (percent)")
abline(lm(Okun$GDPCA_gr ~ Okun$dUNRATE_avg), col = "blue", lty = 2, lwd = 2)
abline(h = 0, lty = 3)
abline(v = 0, lty = 3)
```



3) Regression

We have reached almost the end of our exercise. As you can see from the table, the estimates obtained are similar (though not identical) to those reported in the book.

```
# We regress real GDP growth on the change in unemployment rate
reg <- lm(GDPCA_gr ~ dUNRATE_avg, data = Okun)

# Regression table
stargazer(reg, type = "text", keep.stat = c("n", "rsq"))
##
## ========================================
##                   Dependent variable:
##               ---------------------------
##                          GDPCA_gr
## ----------------------------------------
## dUNRATE_avg             -1.926***
##                          (0.140)
##
## Constant                3.411***
##                          (0.141)
##
## ----------------------------------------
## Observations               58
## R2                        0.771
## ========================================
```

```
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

From the table, we observe that Okun's coefficient for the US from 1951 to 2008 is equal -1.926 (2 in the book). What does this mean? It means that if unemployment decreases by **one percentage point**, GDP increases on average by **1.926 percent**. The intercept of the regression is 3.411. How can we interpret this coefficient? If unemployment remains constant, GDP will grow by 3.411 percent on average.

$$\beta_1 = -1.926$$