Chapter 10 Dummy Variable Model

남준우·허인 (2018), 제 6장-3절

Gujarati/Porter (2018), Chapter 9

(1) Motivation

(2) Intercept Dummy Variable

(3) Slope Dummy Variable

(4) Intercept and Slope Dummy Variable

(5) Application: Seasonal Adjustment

(1) Motivation

► Independent variables(quantitative variable, qualitative variable) $\Rightarrow$ dependent variable

• Dummy variable: quantifying the qualitative variables.

(Example) Wage discrimination across gender.

• To investigate wage equation for each gender,

Male: $y_i = \alpha_1 + \beta_1 X_i + \varepsilon_i$

Female: $y_i = \alpha_2 + \beta_2 X_i + \varepsilon_i$ where y: wage rate, X: tenure,

► α: starting salary, β: wage promotion

• To compare (conduct hypothesis testing) α's and β's,

we need to merge two separate regression equations into one.

► Need for quantify gender variable.

(2) <u>Intercept Dummy Variable</u>($\alpha_1 \neq \alpha_2,\ \beta_1 = \beta_2$)

Male: $y_i = \alpha_1 + \beta X_i + \varepsilon_i$

Female: $y_i = \alpha_2 + \beta X_i + \varepsilon_i$

► Let $D_i = \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$.

Then,

$$\begin{aligned} y_i &= \alpha_1 \cdot D_i + \alpha_2 \cdot (1 - D_i) + \beta X_i + \varepsilon_i \\ &\equiv \alpha + \gamma D_i + \beta X_i + \varepsilon_i \end{aligned},$$

γ : wage discrimination against female worker.

► LS of y on (1, D, X).

① Testing $H_0: \gamma = 0$ or $H_0: \gamma \geq 0$ by t-test.

(Example) Wage discrimination across gender:

$$Male_i = \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$$

Model: $Income_i = \alpha + \gamma \cdot Male_i + \beta \cdot Age_i + \varepsilon_i$

Output:

$$Income_i = \underset{(3.133)}{17.564} + \underset{(2.412)}{6.776} \cdot Male_i + \underset{(3.181)}{0.494} \cdot Age_i + e_i \qquad R^2 = 0.260$$

► Male: $Income_i = 24.340 + 0.494 \cdot Age_i + e_i$

Female: $Income_i = 17.564 + 0.494 \cdot Age_i + e_i$

② Let $F_i = \begin{cases} 0 & \text{for male} \\ 1 & \text{for female} \end{cases}$. $y_i = \alpha + \delta \cdot F_i + \beta X_i + \varepsilon_i$

Then, $\delta = -\gamma$.

③ In time series data, the dummy variable can be used to examine policy effectiveness or structural change.

$$D_t = \begin{cases} 0 & \text{before policy} \\ 1 & \text{after policy} \end{cases}$$

$$y_t = \alpha + \gamma \cdot D_t + \beta X_t + \varepsilon_t$$

④ Suppose there are M categories, use only M-1 dummy variables.

(Example) Investigating the effect of education on income

Suppose there are four groups: (중졸이하, 고졸, 대졸, 대학원졸).

$$H_i = \begin{cases} 1 & \text{if 고졸} \\ 0 & \text{otherwise} \end{cases}, \qquad C_i = \begin{cases} 1 & \text{if 대졸} \\ 0 & \text{otherwise} \end{cases}, \qquad G_i = \begin{cases} 1 & \text{if 대학원졸} \\ 0 & \text{otherwise} \end{cases}$$

► $y_i = \alpha + \delta_1 H_i + \delta_2 C_i + \delta_3 G_i + \beta X_i + \varepsilon_i$

(Example) Effect of education on income

- Education group: (중졸 이하, 고졸, 대졸 이상)

$$H_i = \begin{cases} 1 & \text{if 고졸} \\ 0 & \text{otherwise} \end{cases}, \qquad C_i = \begin{cases} 1 & \text{if 대졸} \\ 0 & \text{otherwise} \end{cases}$$

Model: $Income_i = \alpha + \delta_1 H_i + \delta_2 C_i + \beta Age_i + \varepsilon_i$

Output:

$$Income_i = \underset{(0.551)}{4.217} + \underset{(0.904)}{3.691} H_i + \underset{(2.020)}{8.478} C_i + \underset{(5.680)}{0.816} Age_i + e_i \qquad R^2 = 0.272$$

► 중졸이하: $Income_i = 4.217 + 0.816 Age_i + e_i$

► 고졸: $Income_i = 4.217 + 0.816 Age_i + e_i$

► 대졸이상: $Income_i = 12.695 + 0.816 Age_i + e_i$

⑤ More than 2 qualitative variables

► gender(D) and education(H, C, G)

► $y_i = \alpha + \gamma D_i + \delta_1 H_i + \delta_2 C_i + \delta_3 G_i + \beta X_i + \varepsilon_i$

⑥ What if using M dummy variables for M categories?

► What if we use two dummies $F_i = \begin{cases} 0 & \text{for male} \\ 1 & \text{for female} \end{cases}$ and $D_i = \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$?

(3) Slope Dummy Variable($\alpha_1 = \alpha_2,\ \beta_1 \neq \beta_2$)

Male: $y_i = \alpha + \beta_1 X_i + \varepsilon_i$

Female: $y_i = \alpha + \beta_2 X_i + \varepsilon_i$

► Let $D_i = \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$.

Then,

$$\begin{aligned} y_i &= \alpha + \beta_1 D_i \cdot X_i + \beta_2 (1 - D_i) \cdot X_i + \varepsilon_i \\ &\equiv \alpha + \beta X_i + \gamma D_i \cdot X_i + \varepsilon_i \end{aligned},$$

γ: discrimination on wage promotion against female worker.

► LS of y on (1, X, D·X).

(Example) Discrimination on wage promotion against female worker:

$$Male_i = \begin{cases} 0 & \text{for} \quad \text{female} \\ 1 & \text{for} \quad \text{male} \end{cases}$$

Model: $Income_i = \alpha + \ \beta \cdot Age_i + \gamma \cdot Male_i \cdot Age_i + + \varepsilon_i$

Output:

$$Income_i = 22.902 + 0.334\,Age_i + \ 0.193\,Male_i \cdot Age_i + e_i$$

$$(3.626) \quad (1.711) \qquad\qquad (2.442) \qquad\qquad\qquad R^2 = 0.261$$

► Male: $Income_i = 22.902 + \ 0.527 \cdot Age_i + e_i$

Female: $Income_i = 22.902 + \ 0.334 \cdot Age_i + e_i$

(4) Intercept and Slope Dummy Variable($\alpha_1 \neq \alpha_2,\ \beta_1 \neq \beta_2$)

Male: $y_i = \alpha_1 + \beta_1 X_i + \varepsilon_i$

Female: $y_i = \alpha_2 + \beta_2 X_i + \varepsilon_i$

► Let $D_i = \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$.

Then,

$$y_i = \alpha + \gamma_1 \cdot D_i + \beta X_i + \gamma_2 D_i \cdot X_i + \varepsilon_i,$$

γ_1: wage discrimination against female worker,

γ_2: discrimination of wage promotion against female worker.

► LS of y on (1, D, X, D·X).

(a) $H_0: \gamma_1 = 0$ or $H_0: \gamma_1 \geq 0$ by t-test.

(b) $H_0: \gamma_2 = 0$ or $H_0: \gamma_2 \geq 0$ by t-test.

(c) $H_0: \gamma_1 = \gamma_2 = 0$ by F-test.

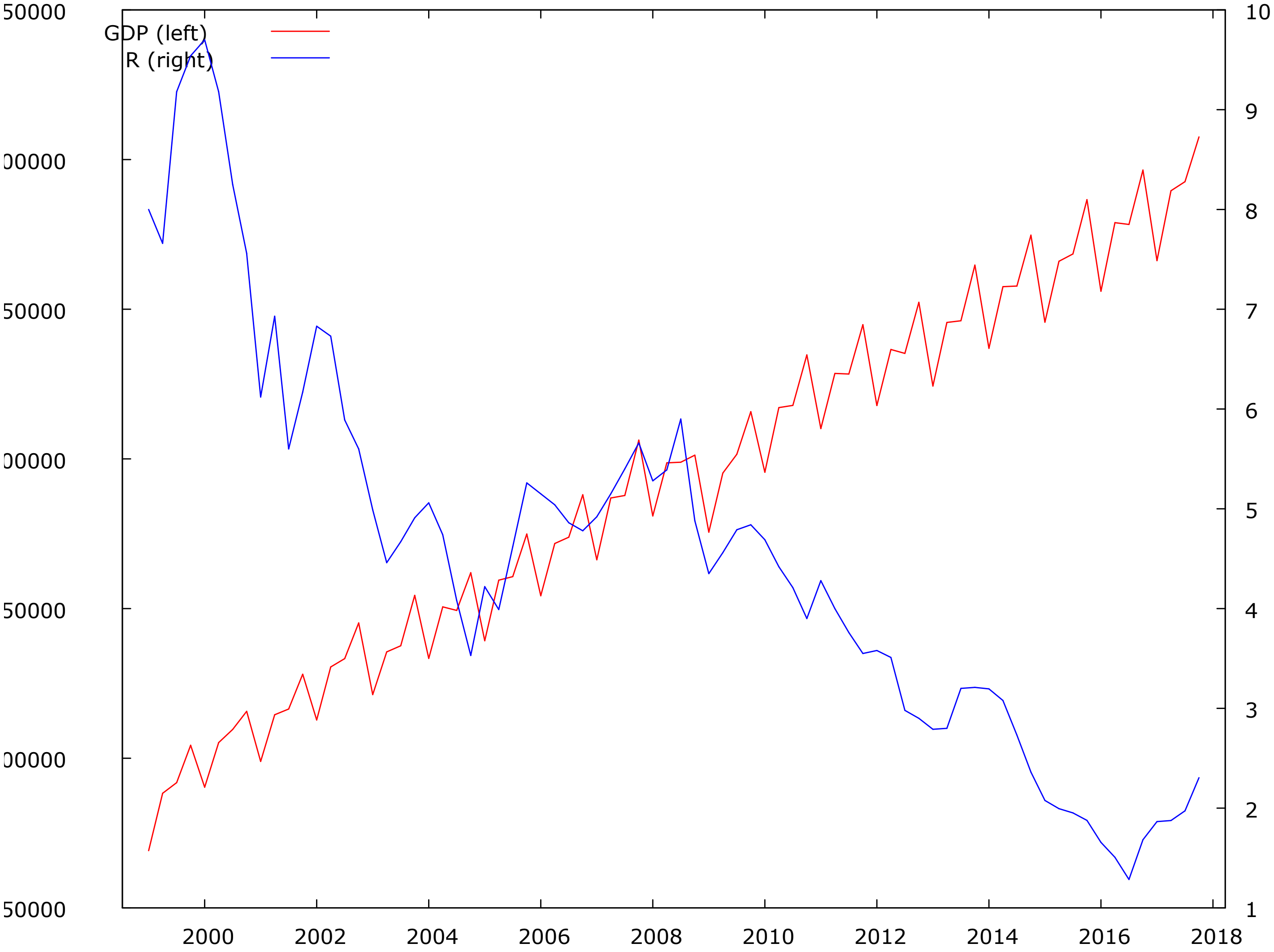
(5) Application: Seasonal Adjustment

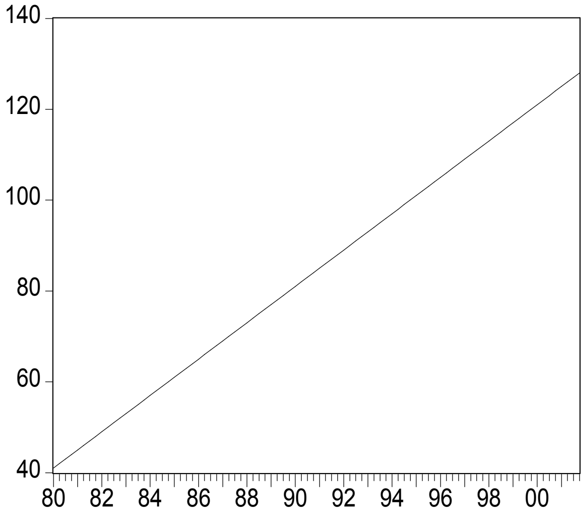
- Many economic (monthly or quarterly) time series data based on exhibit seasonal patterns.

(Example) sales on Christmas season, sales of ice-cream, unemployment rate on winter season and etc.

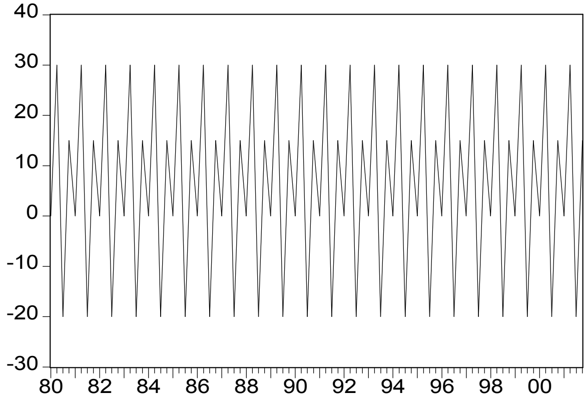
► Often, it is desirable to remove the seasonal factor from a time series, so that one may concentrate on the trend.

► Process of removing seasonal component: deseasonalization, seasonal adjustment.

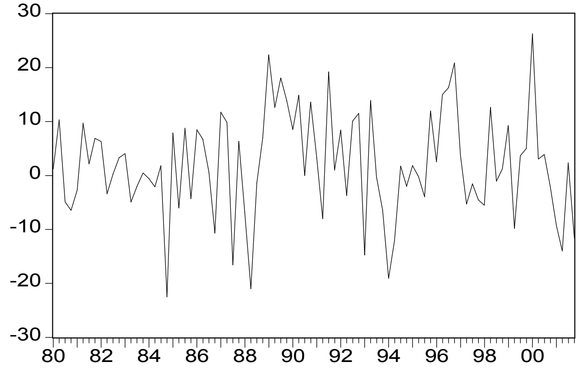

GDP (left)
R (right)
50000
00000
50000
00000
50000
00000
50000
10
9
8
7
6
5
4
3
2
1
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018


140
120
100
80
60
40
80 82 84 86 88 90 92 94 96 98 00

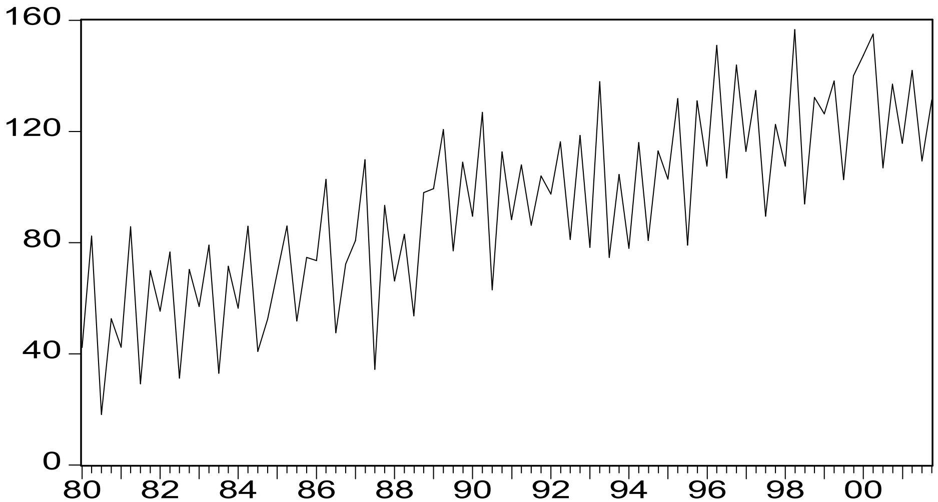
(a)


40
30
20
10
0
-10
-20
-30
80 82 84 86 88 90 92 94 96 98 00

(b)


30
20
10
0
-10
-20
-30
80 82 84 86 88 90 92 94 96 98 00

(c)


160
120
80
40
0
80 82 84 86 88 90 92 94 96 98 00

(d)

- Quarterly Dummy Variables:

► $D_{2t} = \begin{cases} 1 & \text{if 2nd quarter} \\ 0 & \text{otherwise} \end{cases}$, $D_{3t} = \begin{cases} 1 & \text{if 3rd quarter} \\ 0 & \text{otherwise} \end{cases}$, $D_{4t} = \begin{cases} 1 & \text{if 4th quarter} \\ 0 & \text{otherwise} \end{cases}$

- $y_t = \alpha + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta X_t + \varepsilon_t$

1st quarter: $y_t = \alpha + \beta X_t + \varepsilon_t$,

2nd quarter: $y_t = \alpha + \alpha_2 + \beta X_t + \varepsilon_t$,

3rd quarter: $y_t = \alpha + \alpha_3 + \beta X_t + \varepsilon_t$,

4th quarter: $y_t = \alpha + \alpha_4 + \beta X_t + \varepsilon_t$.

(Example) 분기별 자료를 이용한 소비함수의 추정

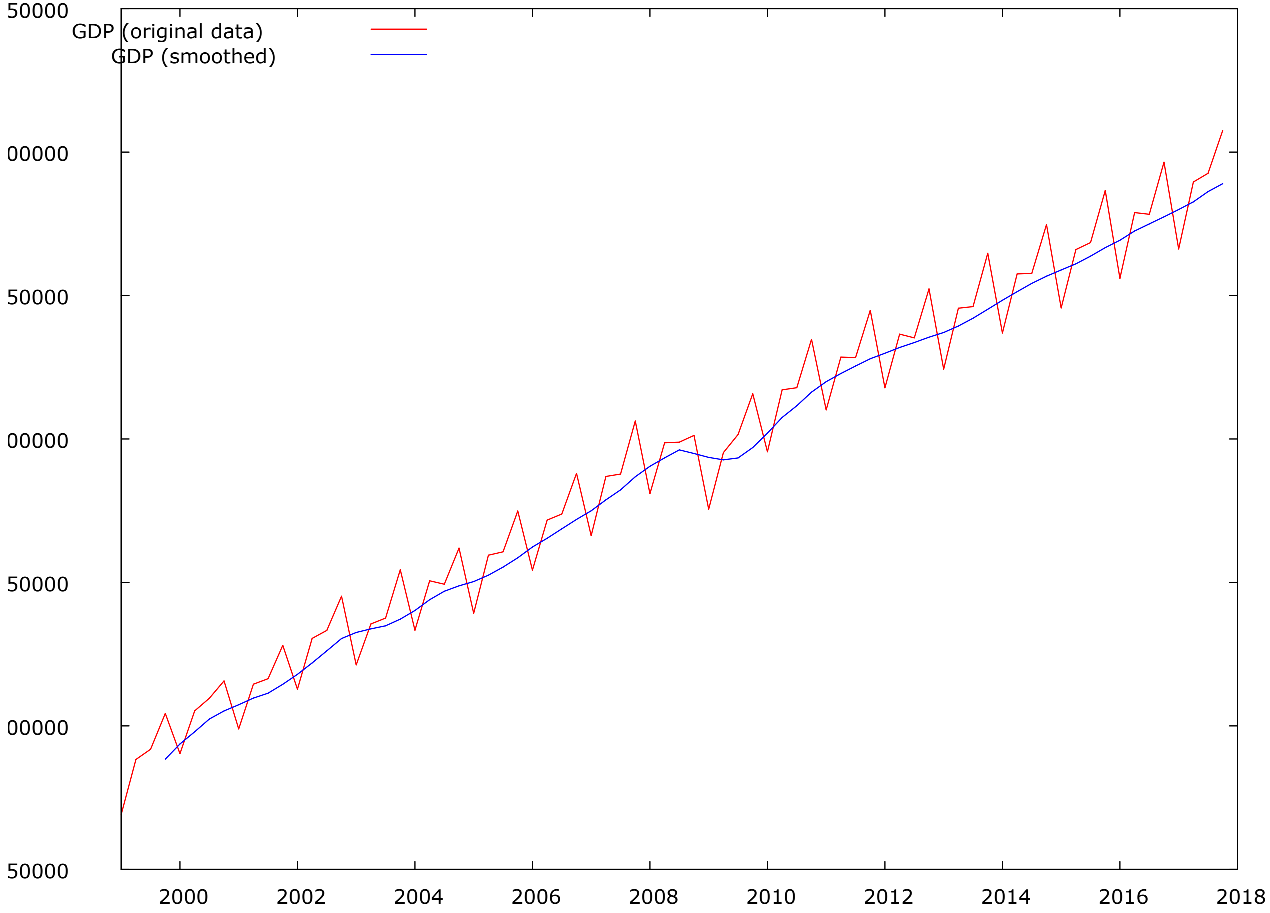
민간소비(CP)를 (상수항, 국민소득(GNI), 이자율(R), 계절더미)에 대해 회귀분석:

$$CPt = 39207.98 + 0.37GNI_t - 560.52R_t + 9070.33D1_t - 684.11D2_t + 1228.48D3_t + e_t$$

$$(4.40) \quad (18.34) \quad (-0.93) \quad (5.78) \quad (-0.46) \quad (0.83)$$

$$R^2 = 0.97$$

• Removing Seasonality(Seasonal Adjustment)



►In gretl, 변수를 선택 후 [variable]/[simple moving average]