Analysis of cross section and panel data - J.Wooldridge Ch4 - Ex13

Thalles Quinaglia Liduares

30/03/2023

Use the data in CORNWELL.RAW (from Cornwell and Trumball, 1994) to estimate a model of county level crime rates, using the year 1987 only.

a. Using logarithms of all variables, estimate a model relating the crime rate to the deterrent variables prbarr, prbconv, prbpris, and avgsen.

Upload package

```
library(wooldridge)
```

Upload data

```
data<-wooldridge::crime4
```

Restrict the sample to 1987

```
data_1987 <- subset(data, year == 87)
```

Take the natural logarithm of the variables

```
data_1987$log_crmrte <- log(data_1987$crmrte)
data_1987$log_prbarr <- log(data_1987$prbarr)
data_1987$log_prbconv <- log(data_1987$prbconv)
data_1987$log_prbpris <- log(data_1987$prbpris)
data_1987$log_avgsen <- log(data_1987$avgsen)
```

Estimate the regression model

```
model <- lm(log_crmrte ~ log_prbarr + log_prbconv + log_prbpris + log_avgsen, data = data_198
7)
```

Results

```
summary(model)
```

```
##
## Call:
## lm(formula = log_crmrte ~ log_prbarr + log_prbconv + log_prbpris +
##     log_avgsen, data = data_1987)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36105 -0.19129  0.07939  0.27754  0.86843
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.86792    0.43153 -11.281  < 2e-16 ***
## log_prbarr  -0.72397    0.11532  -6.278 1.39e-08 ***
## log_prbconv -0.47251    0.08311  -5.686 1.80e-07 ***
## log_prbpris  0.15967    0.20644   0.773    0.441
## log_avgsen   0.07642    0.16347   0.467    0.641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.429 on 85 degrees of freedom
## Multiple R-squared:  0.4162, Adjusted R-squared:  0.3888
## F-statistic: 15.15 on 4 and 85 DF,  p-value: 2.171e-09
```

b. Add log(crmrte) for 1986 as an additional explanatory variable, and comment on how the estimated elasticities di¤er from part a.

Define variable

```
data_1987$log_crmrte86 <- log(subset(data, year == 86)$crmrte)
```

Estimate model

```
model2 <- lm(log_crmrte ~ log_prbarr + log_prbconv + log_prbpris + log_avgsen + log_crmrte86,
data = data_1987)
```

Results

```
summary(model2)
```

```
##
## Call:
## lm(formula = log_crmrte ~ log_prbarr + log_prbconv + log_prbpris +
##     log_avgsen + log_crmrte86, data = data_1987)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28023 -0.08931  0.03055  0.11019  0.32422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.76663    0.31310  -2.449  0.01643 *
## log_prbarr   -0.18504    0.06276  -2.948  0.00414 **
## log_prbconv  -0.03868    0.04660  -0.830  0.40890
## log_prbpris  -0.12669    0.09885  -1.282  0.20351
## log_avgsen   -0.15202    0.07829  -1.942  0.05552 .
## log_crmrte86  0.77981    0.04521  17.248  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2025 on 84 degrees of freedom
## Multiple R-squared:  0.8715, Adjusted R-squared:  0.8638
## F-statistic: 113.9 on 5 and 84 DF,  p-value: < 2.2e-16
```

We can see from the regression results that the estimated elasticities in part b are generally smaller in magnitude than those in part a, except for the elasticity of the crime rate with respect to the probability of imprisonment, which is slightly larger in part b. This suggests that the inclusion of the lagged crime rate as an additional explanatory variable partially explains the relationship between the deterrent variables and the crime rate in 1987, and reduces the estimated elasticities. The coefficient of the lagged crime rate β_5 is positive and statistically significant, indicating that a higher crime rate in 1986 is associated with a higher crime rate in 1987, holding constant the other variables in the model.

```
library(car)
```

```
## Carregando pacotes exigidos: carData
```

F statistic for joint significance of the wage variables

```
linearHypothesis(model2, c("log_prbarr = 0", "log_prbconv = 0", "log_prbpris = 0", "log_avgse
n = 0"), test = "F")
```

```
## Linear hypothesis test
##
## Hypothesis:
## log_prbarr = 0
## log_prbconv = 0
## log_prbpris = 0
## log_avgsen = 0
##
## Model 1: restricted model
## Model 2: log_crmrte ~ log_prbarr + log_prbconv + log_prbpris + log_avgsen +
##     log_crmrte86
##
##   Res.Df    RSS Df Sum of Sq     F   Pr(>F)
## 1     88 4.0576
## 2     84 3.4447  4   0.61284 3.736 0.007573 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F statistic is 3.73 and its associated p-value is 0.007, which is less than the conventional significance level of 0.05. Therefore, we can reject the null hypothesis and conclude that at least one of the wage variables is significantly related to the crime rate in 1987.

```
library(sandwich)
```

Compute robust standard errors

```
robust_se <- sqrt(diag(vcovHC(model2)))
```

Compute the F statistic using robust standard errors

```
linearHypothesis(model2, c("log_prbarr = 0", "log_prbconv = 0", "log_prbpris = 0", "log_avgse
n = 0"),
                 vcov = vcovHC(model2), test = "F")
```

```
## Linear hypothesis test
##
## Hypothesis:
## log_prbarr = 0
## log_prbconv = 0
## log_prbpris = 0
## log_avgsen = 0
##
## Model 1: restricted model
## Model 2: log_crmrte ~ log_prbarr + log_prbconv + log_prbpris + log_avgsen +
##     log_crmrte86
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F  Pr(>F)
## 1     88
## 2     84  4 2.0149 0.09973 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F statistic is 2.01 and its associated p-value is 0.09, which is higher than the conventional significance level of 0.05. Therefore, we cannot reject the null hypothesis that at least one of the wage variables is significantly related to the crime rate in 1987.