BMECV1314D

COPENHAGEN BUSINESS SCHOOL

DEPARTMENT OF ECONOMICS

Winter Semester, 2024

Final Ordinary Exam, 10.12.2024

Econometrics

Time allowed: TWO hours

Students must answer all parts of the question.

General guidance:
You should not devote more than one hour to answer each of the questions.

DO NOT TURN THIS PAGE OVER UNTIL INSTRUCTED TO DO SO BY THE INVIGILATOR

QUESTION

This question considers the estimation of the determinants of regional crime rates. The data are a random sample of 714 regions in the US. The data comprise of the following variables:

lcrim	= natural logarithm of the crimes per 100,000 inhabitants
lpris	= natural logarithm of the number of prisoners per 100,000 inhabitants
po_lit_cy	= 1 if prison overcrowding litigation in the current year (0 otherwise)
po_lit_p2y	= 1 if prison overcrowding litigation in the previous 2 years (0 otherwise)
lincpc	= natural logarithm of the income (in $) per capita
lpolpc	= natural logarithm of the number of police officers per 100,000 inhabitants.

Note: Prison overcrowding litigation is a court verdict on whether a prison is overcowded. It puts the prison under court order and measures are enacted to ease the overcrowding.

The descriptive statistics are given in Table 1A.

TABLE 1A

Variable	Obs	Mean	Std. Dev.	Min	Max
lcrim	714	1.426112	.6546603	-.7431197	3.374785
lpris	714	5.129948	.5800487	3.036248	7.159943
po_lit_cy	714	.0084034	.0913479	0	1
po_lit_p2y	714	.0168067	.1286368	0	1
lincpc	714	9.563354	.2760826	8.836834	10.30426
lpolpc	714	5.556241	.2400228	5.082142	6.811186

The empirical analysis is conducted with R.

CONTINUED

A regression is estimated, and the results are shown in Table 1B.

TABLE 1B

```
> reg1 <- lm(lcrim ~ lpris+lpolpc, data=prison)
> summary(reg1)

Call:
lm(formula = lcrim ~ lpris + lpolpc, data = prison)

Residuals:
     Min       1Q   Median       3Q      Max
-1.22973 -0.24028  0.05792  0.28081  1.06923

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.31710    0.36032  -20.31   <2e-16 ***
lpris        0.57140    0.03126   18.28   <2e-16 ***
lpolpc       1.04603    0.07553   13.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4084 on 711 degrees of freedom
Multiple R-squared:  0.6119,    Adjusted R-squared:  0.6108
F-statistic: 560.4 on 2 and 711 DF,  p-value: < 2.2e-16
```

a) Interpret the coefficient on *lpris* in Table 1B. Define what it means that the variable *lpris* is endogenous in the model in Table 1B. Briefly list 4 different sources of endogeneity.

b) Additional code is executed which gives the following Output 1C.

CONTINUED

OUTPUT 1C

```
> yhatsq<-predict(reg1)^2
> yhatcub<-predict(reg1)^3
> reg2<-lm(lcrim ~ lpris+lpolpc+yhatsq+yhatcub,data=prison)
> linearHypothesis(reg2,c("yhatsq=0","yhatcub=0"))
Linear hypothesis test

Hypothesis:
yhatsq = 0
yhatcub = 0

Model 1: restricted model
Model 2: lcrim ~ lpris + lpolpc + yhatsq + yhatcub

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    711 118.61
2    709 111.77  2    6.8423 21.703 7.112e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Explain in detail what has been done in Output 1C. What do you conclude?

c) Another regression is estimated, and the results are shown in Table 1D.

Table 1D

```
> reg3 <- lm(lcrim ~ lpris+lpolpc+lincpc, data=prison)
> summary(reg3)

Call:
lm(formula = lcrim ~ lpris + lpolpc + lincpc, data = prison)

Residuals:
     Min       1Q   Median       3Q      Max
-1.20660 -0.25007  0.04623  0.27159  1.03628

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.42434    0.56883  -9.536  < 2e-16 ***
lpris        0.60920    0.03213  18.960  < 2e-16 ***
lpolpc       1.14090    0.07788  14.650  < 2e-16 ***
lincpc      -0.27332    0.06406  -4.267 2.25e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4036 on 710 degrees of freedom
Multiple R-squared:  0.6216,    Adjusted R-squared:   0.62
F-statistic: 388.7 on 3 and 710 DF,  p-value: < 2.2e-16
```

Explain in detail why the coefficients on *lpris* and *lpolpc* are different in Tables 1B and 1D under the assumption that the model in Table 1D satisfied Assumptions 1-4 of the course.

CONTINUED

d) Another regression is estimated, and the results are presented in Table 1E.

TABLE 1E

```
> reg4 <- lm(lpris ~ lcrim+lpolpc+lincpc+po_lit_cy+po_lit_p2y, data=prison)
> summary(reg4)

Call:
lm(formula = lpris ~ lcrim + lpolpc + lincpc + po_lit_cy + po_lit_p2y,
    data = prison)

Residuals:
     Min       1Q   Median       3Q      Max
-1.27116 -0.21621  0.01739  0.30338  0.80484

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.84376    0.57491  -1.468    0.143
lcrim        0.54641    0.02937  18.604   <2e-16 ***
lpolpc       0.04996    0.08491   0.588    0.556
lincpc       0.51381    0.05864   8.763   <2e-16 ***
po_lit_cy    0.20884    0.15800   1.322    0.187
po_lit_p2y   0.08197    0.11253   0.728    0.467
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.384 on 708 degrees of freedom
Multiple R-squared:  0.5648,    Adjusted R-squared:  0.5617
F-statistic: 183.7 on 5 and 708 DF,  p-value: < 2.2e-16
```

Using the regression in Table 1B as a starting point, how would you describe the regression in Table 1E? What do you conclude from the latter for the results in Table 1B. Explain in detail.

e) Another regression is estimated, and the results are shown in Table 1F.

CONTINUED

TABLE 1F

```
> reg5<-ivreg(lcrim~lpris+lpolpc+lincpc|po_lit_cy+po_lit_p2y+lpolpc+lincpc,data=prison)
> summary(reg5)

Call:
ivreg(formula = lcrim ~ lpris + lpolpc + lincpc | po_lit_cy +
    po_lit_p2y + lpolpc + lincpc, data = prison)

Residuals:
     Min       1Q   Median       3Q      Max
-1.20749 -0.33356 -0.03387  0.31979  1.55034

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.2459     1.9180  -1.171 0.242002
lpris         1.1641     0.3156   3.689 0.000242 ***
lpolpc        0.5817     0.3290   1.768 0.077485 .
lincpc       -0.5784     0.1884  -3.070 0.002220 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4809 on 710 degrees of freedom
Multiple R-Squared: 0.4626,     Adjusted R-squared: 0.4603
Wald test: 193.9 on 3 and 710 DF,  p-value: < 2.2e-16
```

Explain in detail the estimation approach that is applied in Table 1F. What do you think about the validity of the approach?

f) Additional code is executed, and the results are shown in Output 1G.

CONTINUED

OUTPUT 1G

```
> reg6<-lm(lpris~lpolpc+lincpc+po_lit_cy+po_lit_p2y,data=prison)
> u_hat<-reg6$residuals
> reg7<-lm(lcrim~lpris+lpolpc+lincpc+u_hat,data=prison)
> summary(reg7)

Call:
lm(formula = lcrim ~ lpris + lpolpc + lincpc + u_hat, data = prison)

Residuals:
     Min       1Q   Median       3Q      Max
-1.20769 -0.24584  0.04977  0.27061  1.03616

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.2459     1.6055  -1.399 0.162300
lpris         1.1641     0.2642   4.407 1.21e-05 ***
lpolpc        0.5817     0.2754   2.112 0.035029 *
lincpc       -0.5784     0.1577  -3.668 0.000263 ***
u_hat        -0.5632     0.2661  -2.116 0.034672 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4026 on 709 degrees of freedom
Multiple R-squared:  0.6239,    Adjusted R-squared:  0.6218
F-statistic: 294.1 on 4 and 709 DF,  p-value: < 2.2e-16
```

Explain in detail what has been done in Output 1G. What do you conclude?

END OF PAPER