Lista 4
Econometria
Diogo Wolff Surdi

June 20, 2020

Questão 1

Temos por propriedade do somatório que:

$$\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x}) = \sum_{i=1}^{n} z_i(x_i - \bar{x})$$

Sendo que vale a mesma transformação para y. Temos então a fórmula:

$$\sum_{i=1}^{n} z_i(y_i - \bar{y}) = \sum_{i=1}^{n} z_i y_i - \left(\sum_{i=1}^{n} z_i\right)\bar{y}$$

Note que o primeiro termo será a média de observações de y tais que $z_i = 1$ multiplicada pelo número de vezes que isso ocorre, enquanto o segundo será a média de y multiplicada por esse mesmo número. Chamando tal média condicional de $\bar{y}_1$ e o número de ocorrências de n_1 temos:

$$\sum_{i=1}^{n} z_i(y_i - \bar{y}) = n_1\bar{y}_1 - n_1\bar{y}$$

Seja n_0 o número de observações de $z = 0$; então, devemos ter $n_0 + n_1 = n$, e podemos escrever $\bar{y}$ como:

$$\bar{y} = \frac{n_0}{n}\bar{y}_0 + \frac{n_1}{n}\bar{y}_1$$

Com isso:

$$\bar{y}_1 - \bar{y} = \left[\frac{n - n_1}{n}\right]\bar{y}_1 - \frac{n_0}{n}\bar{y}_0 = \frac{n_0}{n}(\bar{y}_1 - \bar{y}_0)$$

E encontramos então que:

$$n_1\bar{y}_1 - n_1\bar{y} = n_1(\bar{y}_1 - n_1\bar{y}) = \frac{n_1 n_0}{n}(\bar{y}_1 - \bar{y}_0)$$

O mesmo vale para x, logo temos que:

$$\hat{\beta}_1 = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

Questão 2

(a)

Existem diversos fatores que não estão presentes na equação mas que podem afetar o salário mínimo, como previsões de atividade econômica, logo há correlação entre ele e o erro da regressão.

(b)

O salário mínimo federal deve ser mais robusto do que o estadual, logo a performance nacional (contabilizada através do PIB) já deve tomar boa parte da possível correlação com o erro. Com isso, acredito que ele não estaria correlacionado.

(c)

Devido à regra, aumentos no salário mínimo estadual devem seguir de algum modo os aumentos do salário mínimo federal. Como deduzimos que esse último não é correlacionado com o erro, então ele será uma boa variável instrumental.

Questão 3

(a)

Temos a equação do livro:

$$plim\hat{\beta}_1 = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)}\frac{\sigma_u}{\sigma_x}$$

Com isso, um simples cálculo encontra que o viés será de 0.5.

(b)

Da segunda equação dada na questão, temos que a correlação entre x e u precisaria ser maior do que $\frac{1}{2}$ para o viés de MQO superar o viés de MQ2E.

Questão 4

(a)

A equação se torna então:

$$y_t = \beta_0 + \beta_1(x_t - e_t) + u_t = \beta_0 + \beta_1 x_t + u_t - \beta_1 e_t = \beta_0 + \beta_1 x_t + v_t$$

Como e_t e x_t^* não são correlacionados, temos $\mathbb{E}[x_t e_t] = \mathbb{E}[(x_t^* + e_t)e_t] = \mathbb{E}[(e_t^2)] = \sigma_e^2$. Temos então que $Cov(x_t, v_t) = Cov(x_t, u_t) - \beta_1 Cov(x_t, e_t) = -\beta_1\sigma_e^2 < 0$ para $\beta_1 > 0$.

(b)

Por hipótese, temos que $\mathbb{E}[x_{t-1}^* u_t] = \mathbb{E}[e_{t-1}u_t] = \mathbb{E}[e_{t-1}e_t] = 0$, logo $\mathbb{E}[x_{t-1}u_t] = \mathbb{E}[x_{t-1}e_t] = 0$. Com isso, temos $\mathbb{E}[x_{t-1}v_t] = \mathbb{E}[x_{t-1}u_t] - \beta_1\mathbb{E}[x_{t-1}e_t] = 0$.

(c)

É possível (e provável) que as variáveis sejam correlacionadas, dado que variáveis econômicas tendem a crescer.

(d)

Deveríamos tomar x_{t-1} como variável instrumental para x_t.

Questão 5

(a)

```
library(foreign)
dados <- read.dta('(...)/wage2.dta')

> lm(log(wage) ~ sibs, dados)

Call:
lm(formula = log(wage) ~ sibs, data = dados)

Coefficients:
(Intercept)          sibs
6.8611          -0.0279
```

A regressão diz que um irmão a mais causa uma queda de 2.8% no salário.

(b)

Como são as mais velhas, crianças que vieram primeiro podem acabar tendo maior prioridade para ensino superior, além de que com mais filhos a restrição orçamentária das famílias fica mais apertada.

```
> lm(educ ~ brthord, dados)

Call:
lm(formula = educ ~ brthord, data = dados)

Coefficients:
(Intercept)       brthord
14.1494         -0.2826
```

(c)

```
> summary(lm(educ ~ sibs + brthord, dados))

Call:
lm(formula = educ ~ sibs + brthord, data = dados)

Residuals:
Min      1Q  Median      3Q     Max
-5.1438 -1.6854 -0.6852  2.0090  5.9950

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.29650    0.13329 107.260  < 2e-16 ***
sibs        -0.15287    0.03987  -3.834 0.000135 ***
brthord     -0.15267    0.05708  -2.675 0.007619 **
```

A estatística t é de -2.675, logo rejeitamos a hipótese de $\pi_2 = 0$ e as hipóteses de identificação são válidas.

(d)

```
> summary(ivreg(log(wage) ~ educ + sibs | brthord + sibs, data = dados))

Call:
ivreg(formula = log(wage) ~ educ + sibs | brthord + sibs, data = dados)

Residuals:
Min       1Q   Median      3Q     Max
-1.84808 -0.26227  0.03841  0.29901  1.30836

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.938528   1.055690   4.678 3.37e-06 ***
educ        0.136994   0.074681   1.834   0.0669 .
sibs        0.002111   0.017372   0.122   0.9033
```

(e)

```
educhat <- fitted(lm(educ ~ sibs + brthord, dados))

completeFun <- function(data, desiredCols) {
completeVec <- complete.cases(data[, desiredCols])
return(data[completeVec, ])
}

validos <- completeFun(dados, "brthord")
```

```
> cor(teste, validos$sibs)
[1] -0.9294818
```

A correlação negativa é bem forte, indicando um problema de multicolinearidade.

Questão 6

(a)

```
 lm(i3 ~ inf, dados[-1, ])

Call:
lm(formula = i3 ~ inf, data = dados[-1, ])

Coefficients:
(Intercept)          inf
2.3208       0.6981
```

(b)

```
library(AER)

> ivreg(i3 ~ inf | inf_1, data = dados)

Call:
ivreg(formula = i3 ~ inf | inf_1, data = dados)

Coefficients:
(Intercept)          inf
1.5426       0.9025
```

Com a nova regressão, o coeficiente em relação à inflação não é estatisticamente diferente de 1, logo um aumento na inflação poderia levar a um aumento de igual magnitude na taxa de juros.

(c)

```
> lm(ci3 ~ cinf, dados)

Call:
lm(formula = ci3 ~ cinf, data = dados)

Coefficients:
(Intercept)         cinf
0.02296      0.22118
```

A estimativa é muito menor.

(d)

```
atual <- dados$cinf[3:56]
passado <- dados$cinf[2:55]

> lm(atual ~ passado)

Call:
lm(formula = atual ~ passado)

Coefficients:
(Intercept)      passado
0.06358       -0.01028
```

O coeficiente é quase nulo, logo há pouca correlação entre os dois e não podemos utilizar como variável instrumental.