

Lista 3 - Econometria I - 2017

Professor: Denisard Alves

Monitor: Adriano Teixeira

Data de entrega: 24 de abril (horário: início da monitoria)

Exercício 1: Considere as seguintes citações:

“Econometricians want too much from the data and hence tend to be disappointed by the answers, because the data are incomplete and imperfect. In part it is our fault, the appetite grows with eating. As we get larger samples, we keep adding variables and expanding our models, until on the margin, we come back to the same insignificance levels.” – Griliches, *American Economic Review*, 1985

“Controlling for covariates can increase the likelihood that regression estimates have a causal interpretation. But more control is not always better. Some variables are bad controls and should not be included in a regression model even when their inclusion might be expected to change the regression coefficients.” – Angrist & Pischke, *Mostly Harmless Econometrics*, 2009

Os autores alertam para importantes questões, em especial: a **omissão de variáveis relevantes** (subespecificação) e a **inclusão de irrelevantes** (sobreespecificação).

a) Defina cada conceito. Como você espera que cada problema mencionado afete os estimadores de MQO (viés, variância, R^2)? Explique intuitivamente.

b) Como você definiria controles ruins? E controles bons? Explique com exemplos.

Vamos agora formalizar cada um dos casos.

Caso 1: Suponha que o modelo verdadeiro é $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, mas o modelo estimado é $\hat{Y}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1i}$.

c) Derive o viés de variável omitida. Discuta em quais casos o viés é positivo, negativo e nulo.

d) Mostre que $Var(\hat{\gamma}_1|X) \leq Var(\hat{\beta}_1|X)$.

Caso 2: Suponha que o modelo verdadeiro é $Y_i = \beta_0 + \beta_1 x_{1i} + u_i$, mas o modelo estimado é $\hat{Y}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1i} + \hat{\gamma}_2 x_{2i}$.

e) Mostre, mencionando as hipóteses, que a inclusão de variável irrelevante não afeta viés: $E(\hat{\gamma}_1|X) = \gamma_1$ e $E(\hat{\gamma}_2|X) = \gamma_2$.

f) Mostre que $Var(\hat{\gamma}_1|X) \geq Var(\hat{\beta}_1|X)$.

Exercício 2: No trecho abaixo, Olivier Blanchard aborda outra questão comum em Econometria:

*“When students run their first regression, the first problem that they usually encounter is that of **multicollinearity**. Many of them conclude that there is something wrong with OLS; some resort to new and often creative techniques to get around that problem. But, we tell them, this is wrong. Multicollinearity is God’s will, not a problem with OLS or statistical techniques in general.”*

- a) Explique, com as suas palavras, o que é multicolinearidade.
- b) Explique detalhadamente quais as consequências de multicolinearidade.
- c) Explique, com as suas palavras, o que é (multi)colinearidade perfeita.
- d) Em notação matricial, vimos que $\hat{\beta} = (X'X)^{-1}X'Y$. O que acontece com $\hat{\beta}$ quando há perfeita colinearidade entre os X s? Como examinar sua existência?
- e) Seja a seguinte matriz X que contém as explicativas a serem usadas em uma regressão:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 7 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Examine se há perfeita colinearidade.

- f) Seja a seguinte matriz $X'X$ a ser usada em outra regressão:

$$\begin{pmatrix} 9 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 8 & 0 \\ 2 & 0 & 7 & 0 & 0 \\ 1 & 3 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}.$$

Examine se há perfeita colinearidade.

- g) Sob homoscedasticidade, a matriz de variância-covariância é expressa como:

$$\text{var} - \text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Explique o que acontece a esta matriz quando:

- (i) há perfeita colinearidade;
- (ii) a colinearidade é alta porém não perfeita.

Exercício 3: Um estudante faltou na última monitoria de Econometria mas acha que entendeu o **conceito de R^2** :

“ R^2 é tipo assim: mede o grau de acerto do modelo. R^2 alto significa que a reta estimada passa por muitos pontos da amostra. Se a reta estimada passa por todos os pontos, então o R^2 certamente é 1. Se houver dois modelos, prefiro o que tiver maior R^2 . Há dúvida sobre a inclusão de uma variável? Basta checar se o R^2 aumenta ou diminui.”

Há vários equívocos na explicação acima.

- a) Explique o que de fato R^2 representa. R^2 s de diferentes modelos são comparáveis?
- b) Mostre que a inclusão de uma nova variável nunca diminui o R^2 .
- c) Use *R2surprise.dta*. Regrida Y sobre X. Faça um gráfico e veja que apesar do R^2 alto, a reta estimada não necessariamente passa por um ponto da amostra.
- d) Use *R2miracle.dta*. Regrida W sobre Z. Faça um gráfico e veja que embora a reta estimada passa próxima a todos os pontos da amostra, $R^2 \approx 0$. Explique.
[Dica: `graph twoway (lfit W Z) (scatter W Z), yscale(r(0 10))`]

Exercício 4: Um econometrista quer avaliar o efeito da instalação de um incinerador de lixo no preço das casas. Os rumores de que o incinerador seria construído começou em 1979, e sua construção começou em 1981. A base *Lixo.dta*, da pesquisa de Kiel e McClain, tem dados sobre a localização e preços das casas vendidas em 1978 e 1981.

- a) Abra a base de dados. Identifique qual a estrutura da amostra: *cross-section*, painel, *pooled cross sections* ou séries temporais?

Sendo *price* o preço real do imóvel e *near* uma variável igual a 1 se o imóvel estiver próximo ao incinerador (raio de 5 km) e 0 caso contrário, o econometrista estima primeiro: $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 near$ com a amostra de dados posterior à construção do incinerador.

- b) Estime a regressão com a amostra de 1981 e interprete os resultados.
- c) As estimativas são causais? Em outras palavras, pode-se afirmar que o incinerador tenha causado preços menores dos imóveis próximos a ele? Explique.

Em seguida, o econometrista estima $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 near$ com a amostra de dados prévia ao anúncio.

- d) Estime a regressão com a amostra de 1978 e interprete os resultados.
- e) Como estes resultados corroboram com sua resposta no item c? Investigue.
- f) Após estas estimações, o econometrista escolheu uma estratégia que usa dados dos dois anos. Sugira uma especificação.

Exercício 5: Você foi contratado para estimar o efeito dos gastos de campanha nos resultados das eleições. A base de dados *Urna.dta* contém informações de 173 disputas eleitorais municipais entre dois partidos, o Partido Azul e o Partido Vermelho, ocorridas no ano corrente.

Inicialmente você pensou num modelo de regressão simples:

$$VotoAzul_i = \beta_0 + \beta_1 GastoAzul_i + u_i$$

onde *GastoAzul* e *VotoAzul* são respectivamente o montante de gastos e a porcentagem de votos recebida pelo candidato Azul.

a) Você suspeita que o Stata não está fazendo regressões adequadamente. Sem usar o comando *regress*, compute os coeficientes estimados usando comandos de somatório. Confirme seus resultados usando *regress*.

b) Interprete os resultados estimados. Será que $E[u|gastoAzul] = 0$? Você acha que há viés de variável omitida na estimação de $\hat{\beta}_1$? Se sim, cite uma variável omitida e elabore uma conjectura sobre o sinal do viés.

c) Verifique que as seguintes propriedades algébricas de MQO são válidas nesta regressão:

i) $\sum_{i=1}^n \hat{u}_i = 0$

ii) $\sum_{i=1}^n GastoAzul_i \cdot \hat{u}_i = 0$

iii) $(\overline{GastoAzul_i}, \overline{VotoAzul_i})$ pertence à reta de regressão estimada.

Posteriormente você pensou num modelo de regressão múltipla:

$$VotoAzul_i = \beta_0 + \beta_1 GastoAzul_i + \beta_2 GastoVermelho_i + \beta_3 ApoioAzul_i + u_i$$

onde *ApoioAzul* é a porcentagem de votos recebida pelo Partido Azul na última eleição presidencial (uma medida prévia de força do Partido Azul).

d) Sem usar o comando *regress*, compute os coeficientes estimados usando comandos de matrizes. Confirme seus resultados usando *regress*.

e) Interprete seus resultados.

f) Você incluiria *GastosTotais*, variável referente à soma dos gastos dos dois partidos, como variável explicativa do modelo? Que tal incluir a variável $\log(GastosTotais)$? Explique.

Ao longo do Capítulo 3, Wooldridge tem um subseção na qual aborda a interpretação de parcialidade (*partialled out*) de regressão múltipla.

g) Regrida *GastoAzul* sobre *GastoVermelho* e *ApoioAzul*. Salve os resíduos (\hat{r}_i). Faça a regressão de *VotoAzul* sobre \hat{r} . Explique seus resultados.