
LISTA 4

Curso: Economia 4ECO

Disciplina: Econometria

Período letivo: 2016/1

Professor: Hedibert Freitas Lopes - www.hedibert.org

Monitora: Paloma Vaissman Uribe - PalomaVU@insper.edu.br

Questão 1

Considere o seguinte modelo para estimar o efeito de possuir um computador pessoal (PC) na média de notas de estudantes universitários no seu último ano de graduação (GPA):

$$GPA_i = \beta_0 + \beta_1 PC_i + \epsilon_i$$

em que, PC_i é uma variável dummy que vale 1 se o estudante possui um computador.

- Por que você esperaria que PC_i seja correlacionada com ϵ_i ?
- Explique porque PC_i deve ser correlacionada com a renda anual dos pais do estudante. Isso significa que renda anual é um bom instrumento para PC_i ?
- Suponha que, quatro anos atrás, a universidade deu bolsas para metade dos estudantes para que eles pudessem comprar seus computadores. Além disso, assuma que os estudantes que receberam a bolsa foram escolhidos de forma aleatória. Explique como você poderia usar essa informação para construir uma variável instrumental para PC_i .

Questão 2

Suponha que você seja um pesquisador interessado em entender a dinâmica salarial de uma determinada população. Considere, por exemplo, a seguinte equação para os salários:

$$\log(\text{salario}_i) = \beta_0 + \beta_1 \text{exper}_i + \beta_2 \text{exper}_i^2 + \beta_3 \text{educ}_i + \beta_4 \text{negro}_i + \epsilon_i \quad (1)$$

em que:

salario – é o salário pago mensalmente aos indivíduos;

educ – educação formal, em anos;

exper – experiência profissional, em anos;

negro – variável dummy que assume o valor 1 para as pessoas negras.

Utilizando a base de dados disponível em <http://hedibert.org/wp-content/uploads/2014/02/wage2-wooldridge.txt>, responda o que se pede.

- Explique quais variáveis explicativas podem ser endógenas e por que motivos pode haver endogeneidade no modelo (1) proposto.
- Usando as variáveis $feduc$ (escolaridade do pai) e $meduc$ (escolaridade da mãe) como instrumentos para a variável $educ$, estime os parâmetros do modelo proposto em (1). Escreva os resultados na forma usual.
- Utilizando um nível de significância de 10%, teste se a possível endogeneidade da variável educação é um problema para o estimador de mínimos quadrados.
- Teste a validade dos dois instrumentos propostos em (b). Utilize um nível de significância de 10%.

Questão 3

Um economista tem o objetivo de estimar uma curva de oferta. Para tanto, ele tem a sua disposição a quantidade negociada (y) e o preço praticado (p) em 150 lojas em um determinado mês, além de um conjunto de variáveis que podem ser usadas como instrumentos.

Observação: assumamos que o processo gerador dos dados é tal que $\log(y_i) = \beta_0 + \beta_1 \log(p_i) + \varepsilon_i$.

- a) Durante o processo de formulação do modelo, o economista afirma: “a estimação da oferta por MQO é incorreta, pois sabemos que neste caso sempre há endogeneidade. Assim, usarei como instrumento o preço dos três insumos principais na produção do bem”. Esta afirmação é correta? Justifique sua resposta com palavras.
- b) O economista estimou o modelo, conforme a afirmação feita em (a), obtendo:

$$\log(y_i) = 1,2 + 0,5 \log(p_i) + \varepsilon_{i,TLS}$$

(0,2) (0,1)

em que TLS significa Mínimos quadrados em Dois Estágios (entre parênteses apresentamos o erro-padrão de cada estimativa). Além disso, obteve-se por MQO:

$$\varepsilon_{i,TLS} = 0,01 + 0,31Z_1 + 0,44Z_2 + 0,23Z_3 + \hat{\varepsilon}_i$$

(0,02) (0,12) (0,18) (0,08)

$$R^2 = 0,09$$

em que Z_1 , Z_2 e Z_3 são os três instrumentos mencionados anteriormente (entre parênteses apresentamos o erro-padrão de cada estimativa). Os resultados da equação fazem sentido? Para auxiliar a discussão, faça um teste de hipótese, a 5% de significância, para verificar se os instrumentos são válidos.

Questão 4

Os dados em <http://fmwww.bc.edu/ec-p/data/wooldridge/fertil2.dta> incluem, para as mulheres em Botswana em 1988, informações sobre o número de filhos, anos de escolaridade, idade e variáveis de status econômico e religioso. Para usar os dados, instale o pacote *foreign* no R e use o comando: `read.dta`. Estime por MQO, o seguinte modelo:

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + \varepsilon,$$

em que:

children é o número de filhos de determinada mulher na amostra;

educ é o número de anos de estudo da mulher;

age é a idade da mulher.

- a) Mantendo *age* constante, qual é o efeito estimado de um ano adicional de educação na fertilidade da mulher?
- b) A variável *frsthalf* é uma variável dummy que vale 1 se a mulher (mãe) nasceu nos primeiros seis meses do ano. Quais são as condições que *frsthalf* deve atender para que ela possa ser usada como instrumento para *educ*? Você acha que essas condições são razoáveis, nesse exemplo?
- c) Faça uma regressão de primeiro estágio e comente se *frsthalf* parece ser um bom instrumento para *educ*. Qual das duas condições você está "testando" nesse caso?
- d) Estime o modelo usando *frsthalf* como variável instrumental para *educ*. Compare o coeficiente estimado por MQO com o coeficiente estimado por IV. Você acha que os resultados estão mais coerentes?

Questão 5

A partir dos modelos (1), (2) e (3), respectivamente, pretendemos estudar a demanda por frangos nos Estados Unidos. Mais especificamente, pretendemos investigar o consumo per capita de frango nos Estados Unidos, durante os anos de 1960 a 1982.

$$C_t = \beta_1 + \beta_2 \log(R_t) + \beta_3 F_t + \epsilon_t$$

$$C_t = \beta_1 + \beta_2 \log(R_t) + \beta_3 F_t \beta_4 S_t + \beta_5 B_t + \epsilon_t$$

$$C_t = \beta_1 + \beta_2 \log(R_t) + \beta_3 F_t \beta_4 S_t + \beta_5 B_t + \beta_6 B_t S_t + \epsilon_t$$

em que:

C_t – consumo per capita de frango em libras-peso;

R_t – renda real disponível per capita, em US\$;

F_t – preço real do frango, no varejo, em centavos de dólar por libra-peso;

S_t – preço real da carne suína, no varejo, em centavos de dólar por libra-peso;

B_t – preço real da carne bovina, no varejo, em centavos de dólar por libra-peso;

O Quadro 1 apresenta, entre outras coisas, os resultados das estimativas dos parâmetros dos modelos propostos em (1), (2) e (3), respectivamente, por MQO.

- a) Qual é a interpretação de β_2 no modelo proposto em (1)?
- b) Analisando o modelo (3), qual é o efeito sobre C_t quando se aumenta S_t em uma unidade? Esse efeito é sempre constante? Como podemos interpretar β_6 ?
- c) Utilizando-se de algum critério adequado, qual modelo você escolheria: (1), (2) ou (3)?
- d) Considerando o modelo escolhido no item (c), faça o teste de heterocedasticidade de White (sem termos cruzados). Adote um nível de significância de 10%.
- e) Considerando o modelo escolhido no item (c), conduza um teste para verificar se os erros apresentam correlação serial de primeira ordem. Adote um nível de significância de 10%. (Ler o Capítulo 12 de Wooldridge)
- f) Considerando os resultados obtidos nos itens (d) e (e), é necessário usar algum estimador robusto à heterocedasticidade ou à correlação serial? Em caso afirmativo, qual?
- g) Um economista acredita que a variável explicativa F_t é endógena. Considerando o modelo escolhido no item (c), responda:
 - (i) Indique como você conduziria um teste para verificar se, realmente, F_t é uma variável explicativa endógena. Qual instrumento você proporia para a variável endógena F_t ?
 - (ii) Supondo que os resultados disponibilizados na Tabela 1 sejam referentes ao instrumento(s) proposto(s) no item anterior, é possível inferir, com 90% de confiança, que a variável F_t é endógena?
 - (iii) Considerando o resultado obtido no item (ii), qual método de estimação deveria ser utilizado para estimar os parâmetros do modelo de interesse? Justifique a sua resposta levando em consideração, inclusive, as propriedades dos estimadores dos parâmetros do modelo de interesse.

Quadro1

Variável Dependente: C_t

Número de observações: 23

Variáveis Explicativas	Modelos		
	(1)	(2)	(3)
Constante	-61.27514 (4.91804)	-47.97135 (12.87808)	-52.37624 (15.57921)
$\ln(R_t)$	16.33948 (15.48051)	14.11194 (2.14885)	15.16099 (2.96261)
F_t	-0.20613 (-3.80854)	-0.26935 (0.10559)	-0.25081 (0.11337)
S_t		0.00127 (0.04647)	-0.02496 (0.06876)
B_t		0.03799 (0.01796)	0.01468 (0.04789)
$S_t B_t$			0.00013 (0.00025)
R^2	0.97732	0.98211	0.98239
R^2 ajustado	0.97505	0.97813	0.97722
Teste de Heterocedasticidade de White sem termos cruzados	2.91491	4.40195	3.33572
Teste LM de Correlação Serial de ordem 1 de Breusch-Godfrey	2.43201	0.40835	0.22945
Teste de Endogeneidade	0.342498	1.11700	1.067404

Nota: O teste de Endogeneidade apresenta o resultado da estatística t do teste para a variável F_t , considerando as demais variáveis exógenas.