

Microeconometria

Prof. Francis Petterini



Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Economia

`f.petterini@ufsc.br`

Atualizado em 29 de julho de 2023

1 Introdução

2 Revisão de conceitos

3 Microdados

4 Modelo linear

5 Violações das hipóteses

6 Monte Carlo

7 Instrumentos e painel

8 Simultaneidade

9 Resposta binária

10 Resposta de contagem

11 Resposta multinomial

12 Censura e truncagem

13 Análise de sobrevivência

14 Avaliação econômica

15 Pareamento

16 Regressão na descontinuidade

17 Dif-in-Dif

18 Controle sintético

19 Econometria de cópulas

20 Econometria de leilões

21 Econometria espacial

Introdução

- “[M]icroeconometric [...] the analysis of individual-level data on the economic behavior of individuals or firms” (Cameron and Trivedi, 2005) [p. 3]; “The goal [...] is to determine whether a change in one variable, say w , causes a change in another variable, say y . For example, does having another year of education cause an increase in monthly salary?” (Wooldridge, 2010) [p. 3]
- A pedra angular da microeconometria é $y = E(y | X) + u$ tal que $E(u | X) = 0$
- Se modela $E(y | X)$ [variáveis dependente (resposta ou regressando) e explicativas (independentes, covariadas, regressores ou controles)], para inferir $\partial E(y | X) / \partial x$, para $x \in X$
- Variáveis contínuas (irrestritas, censuradas ou truncadas) e categóricas (contagem, ordenadas e não ordenadas)
- “The notion of **ceteris paribus** – that is, holding all other (relevant) factors fixed – is at the crux of establishing a **causal relationship**”
- “Deciding on the list of proper controls is not always straightforward, and using different controls can lead to different conclusions about a causal relationship between y and w ” [**endogeneidade**]
- dados: experimentais vs. não experimentais; e, transversais (cross section), longitudinais (painel) e empilhados (pooled) ... $\{\{y_{it}, X_{it}\}_{i=1}^{n_t}\}_{t=1}^T$
- só consideraremos: **random sampling assumption**; modelos paramétricos e estimadores MQ, ML e GMM (não consideraremos modelos semi-paramétricos e não-paramétricos, nem métodos Bayesianos); e, usaremos as propriedades assintóticas dos estimadores, e eventualmente usaremos bootstrapping

Revisão de conceitos

- **Processo gerador de dados de um experimento** (e.g., lançamento simultâneo de duas moedas)
- **Espaço amostral** é um conjunto dos resultados possíveis (e.g., $S = \{\text{duas caras, duas coroas, uma cara e uma coroa}\}$) e **evento** é um conjunto de resultados (e.g., $A_i = \{\text{duas caras, duas coroas}\}$)
- **Distribuição de probabilidade** é uma função $P : S \rightarrow \mathbb{R}$ tal que:
 - $P(A_i) \in [0, 1]$
 - $P(S) = 1$
 - $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, $\forall i, j = 1, \dots, n : A_i \cap A_j = \emptyset$
- **Variável aleatória** é uma função $X : S \rightarrow \mathbb{R}$ (e.g., $X = 1$ para faces iguais e $X = 0$ caso contrário)
- X é **discreta** se $X \subseteq \mathbb{Z}$, c.c. é **contínua**
- **Função de probabilidade** (CDF) é $F(x) = P(X \leq x)$
- **Função densidade de probabilidade** (PDF) é:
 - $f(x) = P(X = x)$ se X é discreta
 - $f(x) = F'(x)$ se X é contínua
- **Suporte** é o conjunto $S(X) = \{X = x : f(x) > 0\}$

- X discretas:

- Bernoulli: $f(x) = p^x(1-p)^{(1-x)}$ para $S(X) = \{0, 1\}$
- Binomial: $f(x) = \frac{n!}{x!(n-x)!} p^x(1-p)^{(1-x)}$ para $S(X) = \{0, 1, 2, \dots, n\}$
- Poisson: $f(x) = \frac{\exp(-\lambda)\lambda^x}{x!}$ para $S(X) = \mathbb{N}$

- X contínuas:

- Uniforme: $f(x) = \frac{1}{b-a}$ para $S(X) = [a, b]$
- Normal padrão: $f(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$ para $S(X) = \mathbb{R}$
- Logística padrão: $f(x) = \frac{\exp(x)}{(1 + \exp(x))^2}$ para $S(X) = \mathbb{R}$
- Exponencial: $f(x) = \lambda \exp(-\lambda x)$ para $S(X) = \mathbb{R}^{++}$
- Potência: $f(x) = \theta x^{\theta-1}$ para $S(X) = [0, 1]$

- Poisson:

- $P(S) = \sum_{x=0}^{+\infty} \frac{\exp(-\lambda)\lambda^x}{x!} = \exp(-\lambda) \left[\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} \right]$
- $\exp(\lambda) = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}$ é a expansão de Taylor para $\lambda^* = 0$
- então $P(S) = 1$ ocorre para todo $\lambda \neq 0$
- mas se $\lambda < 0$ ocorre $P(A_i = \{X = 1\}) < 0$
- então a restrição é $\lambda > 0$

- Potência:

- $P(S) = \int_0^1 \theta x^{\theta-1} dx = \theta \left[\frac{x^\theta}{\theta} \right]_0^1 = 1$
- então $P(S) = 1$ ocorre para $\theta \neq 0$
- mas se $\theta < 0$ ocorre $P(A_i = \{0, 1 < X < 0, 9\}) < 0$
- então a restrição é $\theta > 0$

https://en.wikipedia.org/wiki/List_of_probability_distributions

- **Esperança** (matemática) é a função $E : X \rightarrow \mathbb{R}$ escrita

- $E(X) = \sum_{x \in S(X)} xf(x)$ no caso discreto
- $E(X) = \int_{x \in S(X)} xf(x)dx$ no caso contínuo

- Se $h : X \rightarrow \mathbb{R}$ ocorre

- $E(h(X)) = \sum_{x \in S(X)} h(x)f(x)$ no caso discreto
- $E(h(X)) = \int_{x \in S(X)} h(x)f(x)dx$ no caso contínuo

- Exemplos:

- Poisson: $E(X) = \sum_{x=0}^{+\infty} \frac{x \exp(-\lambda) \lambda^x}{x!} = \lambda \exp(-\lambda) \sum_{x=1}^{+\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda$

- Potência: $E(X) = \int_0^1 \theta x x^{\theta-1} dx = \frac{\theta}{\theta+1}$

- Se $r \in \mathbb{N} \setminus \{0\}$: $E(X^r)$ é o **momento r** ; e, $E((X - E(X))^r)$ é o **momento r centrado**

- A variância é o 2º momento centrado: $V(X) = E(X^2) - (E(X))^2$

- Exemplo... Potência: $E(X^2) = \int_0^1 \theta x^2 x^{\theta-1} dx = \theta \int_0^1 x^{\theta+1} dx = \frac{\theta}{\theta+2}$, logo $V(X) = \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2}$

1 Vetor aleatório

2 Distribuição conjunta: BiNormal, BiPoisson etc.

3 **Marginais:**

- Caso discreto: $f_1(x_1) = \sum_{x_2 \in S(X)} f(x_1, x_2)$
- Caso contínuo: $f_1(x_1) = \int_{x_2 \in S(X)} f(x_1, x_2) dx_2$

4 **Condicionais:** $P(X_1 | X_2) = \frac{P(X_1 \cap X_2)}{P(X_2)} \Rightarrow g_1(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$

5 Covariância $Cov(x_1, x_2) = E((x_1 - E(x_1))(x_2 - E(x_2))) = E(x_1 x_2) - E(x_1)E(x_2)$

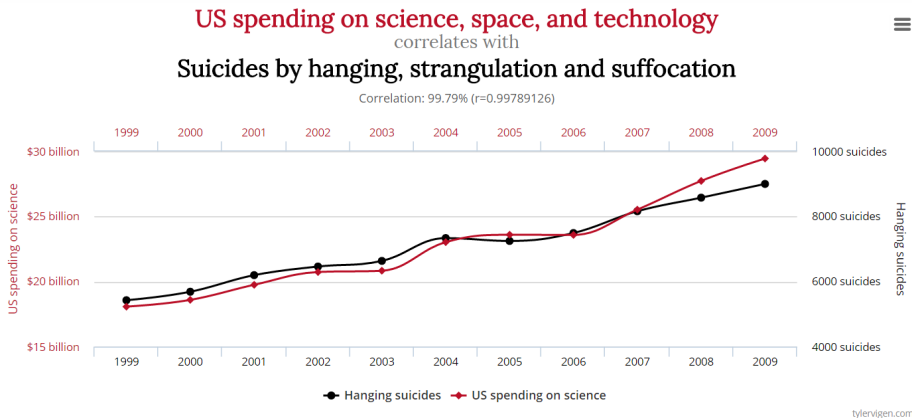
6 $Cov(x_1, x_2) = 0 \Rightarrow E(x_1 x_2) = E(x_1)E(x_2) \Rightarrow E(x_1 x_2) = 0$ se $E(x_k) = 0$

7 LEI: $E(x_1) = \sum_j E(x_1 | x_{2j}) \Pr(x_{2j})$, então $E(x_1 | x_2) = 0 \Rightarrow E(x_1) = 0$

8 Correlação de Pearson $\rho = \frac{Cov(x_1, x_2)}{\sigma_1 \sigma_2} \in [-1, 1]$

9 Independência $\Rightarrow$ Correlação zero (mas o inverso não é verdade)

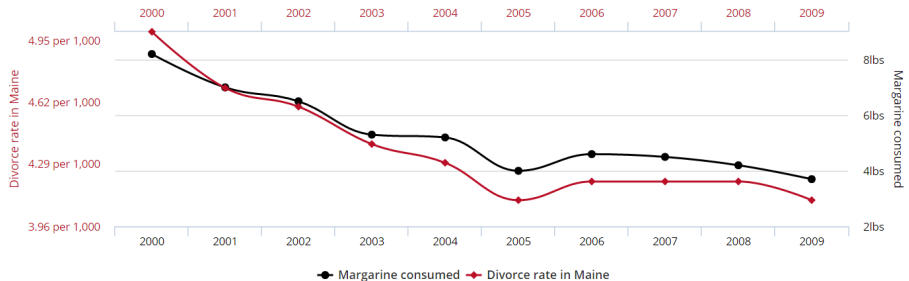
10 Causalidade $\Rightarrow$ Correlação diferente de zero (mas o inverso não é verdade)



Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

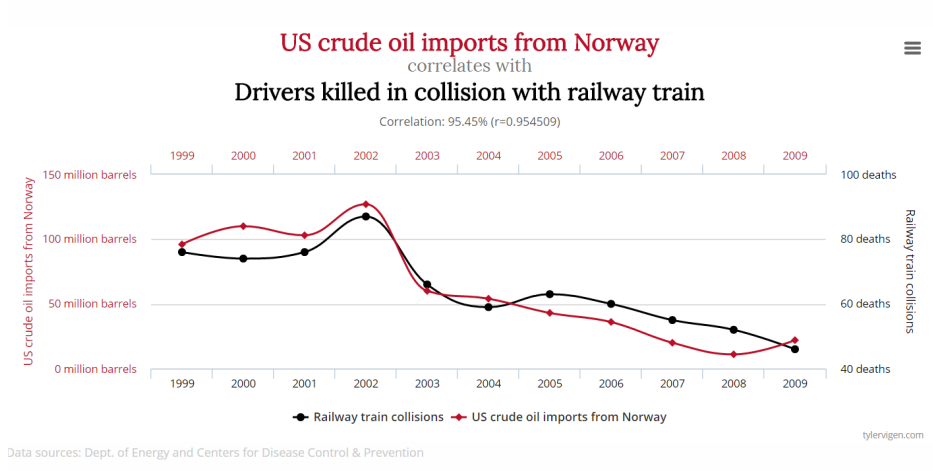
Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% ($r=0.992558$)



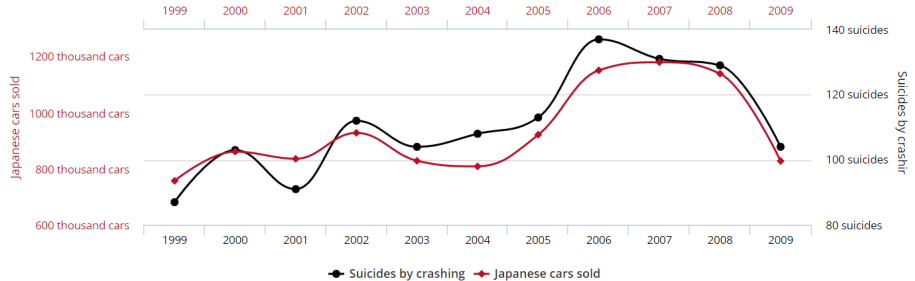
Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

tylervigen.com



Japanese passenger cars sold in the US correlates with Suicides by crashing of motor vehicle

Correlation: 93.57% ($r=0.935701$)



Data sources: U.S. Bureau of Transportation Statistics and Centers for Disease Control & Prevention

tylervigen.com

- **Amostra** é um subconjunto (ou vetor) $\{x_i\}_{i=1}^n$ de realizações observadas de um experimento (e o que é população?)
- A densidade de probabilidade *pró veritate* é $f(x_1, \dots, x_n)$
- **Amostra aleatória** (ou independente) $\Rightarrow f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$
- **Amostra aleatória e identicamente distribuída (i.i.d.)** $\Rightarrow f(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)$

O termo “estatística” também é usado para representar qualquer função dos valores de uma amostra... as principais estatísticas são divididas em classes de:

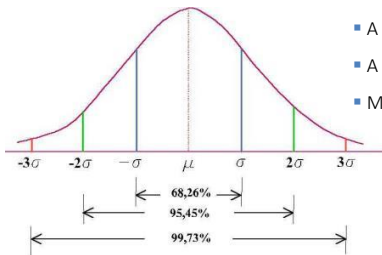
- **Centralização** de $f(x_1, \dots, x_n)$: médias aritmética, ponderada, harmônica e geométrica; moda e mediana
- **Dispersão** de $f(x_1, \dots, x_n)$: variância, desvio padrão e amplitude
- Indicadores relativos de **Assimetria e Curtose** de $f(x_1, \dots, x_n)$
- Indicadores de **Correlação** no caso de amostras $\{(x_i, y_i)\}_{i=1}^n$ etc. (fórmulas de Spearman e Pearson)
- **Estimadores** são estatísticas usadas para se estimar os parâmetros de $f(x_1, \dots, x_n)$, e seus valores calculados são chamados de **estimativas**

- Na estatística clássica os momentos são determinados por parâmetros, portanto, estimar um momento é equivalente a estimar um parâmetro (ou um vetor deles)
- θ é um parâmetro e $\hat{\theta}$ é o seu estimador (e *per si* é uma variável aleatória porque não é determinística... é uma função da realização de uma amostra em particular)
- $\hat{\theta}$ é **viesado** se $E(\hat{\theta} - \theta) \neq 0$
- $\hat{\theta}$ é **consistente** se $P(|\hat{\theta} - \theta| < u) = 1$ (*plim*) para todo $u > 0$ quando $n \rightarrow \infty$ (isso caracteriza uma propriedade **assintótica**)

- $\hat{\theta}$ é **eficiente** em relação a $\tilde{\theta}$ se possui menor variância
- Inevitavelmente, para se encontrar $\hat{\theta}$ é preciso **modelar** θ
- LGN: $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) \Rightarrow \bar{X}_{n \rightarrow \infty} \rightarrow \mu$
- TLC: $\sqrt{n} \left[\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - \mu \right] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$
- IC: $P \left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$
- EP: $t = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n-1}} = \frac{\bar{X}_n - \mu}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}}$
- Formulação de hipóteses nula e alternativa
- Não rejeição de H_0 e tipos de erros

TEOREMA DO LIMITE CENTRAL E A DISTRIBUIÇÃO NORMAL

- A distribuição normal é uma das mais importantes distribuições de probabilidade
- O Teorema do Limite Central (TLC) postula que, numa população de tamanho N , se você retirar A amostras de tamanho n , você verificará que $\frac{\bar{y}_a - \mu}{S_a} \sim N(0, 1)$, onde $a = 1, 2, \dots, A$ indica a amostragem. $S_a = \sigma / \sqrt{n}$ é o desvio padrão de $\bar{x}_a$



- A representação formal é $N(\mu, \sigma^2)$
- A distribuição é simétrica
- Média = Mediana = Moda

Exercício #1 (<https://www.scilab.org/>)

Verificando a validade do TLC...

```
clc
clear

rand('seed',12345)

b = 4
a = 2
n = 100
s = 1000
x = a+(b-a).*rand(n,s,'uniform')

y = (mean(x,'r')-(a+b)/2)./(sqrt((b-a)^2/12)/sqrt(n))
histplot(1000,y)
```

Seja $q = \frac{p^{-.2} + p^{-.5}}{2}$: se $q = q^*$, quem é p^* ?

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots$$

$$f(x) \approx f(a) + f'(a)(x-a)$$

$$f(p^*) = \frac{(p^*)^{-.2} + (p^*)^{-.5}}{2} - q^* = 0$$

$$f(x) = 0 \Rightarrow f'(a)(x-a) = -f(a) \Rightarrow x = a - \frac{f(a)}{f'(a)}$$

$$x_{j+1} = x_j - \frac{f(x_j)}{f'(x_j)}$$

Exercício #2

```
clc
clear

x = 0:0.01:1
y = .5*x^-.2+.5*x^-.5

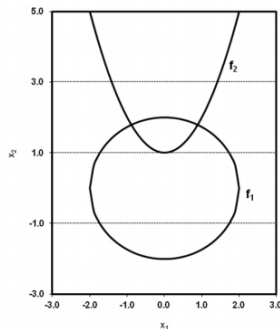
plot(x,y)
plot(x,2,'r')

p = .1
q = 2
for i = 1:100
    delta = (.5*p^-.2+.5*p^-.5-q) / (.1*p^-1.2+.25*p^-1.5)
    p = p + delta
    if abs(delta) < 10^-5
        break
    end
end
```

$$f_1(x_1, x_2) = (x_1)^2 + (x_2)^2 - 4 = 0$$

$$f_2(x_1, x_2) = (x_1)^2 - (x_2) + 1 = 0$$

The first equation is that of a circle with radius 2 centered at the origin. The second equation is that of a parabola. In Figure 5.1. We plot the solution to the two equations independently. Since there are two variables in each equation, there are an infinite number of solutions for the equations treated independently. The solution to the system of nonlinear algebraic equations corresponds to ordered pairs of (x_1, x_2) that satisfy both equations. In Figure 5.1., this corresponds to the intersection between the two curves.



$$J_{1,1} = \left(\frac{\partial f_1}{\partial x_1} \right) = 2x_1 \quad J_{1,2} = \left(\frac{\partial f_1}{\partial x_2} \right) = 2x_2$$

$$J_{2,1} = \left(\frac{\partial f_2}{\partial x_1} \right) = 2x_1 \quad J_{2,2} = \left(\frac{\partial f_2}{\partial x_2} \right) = -1$$

$$\begin{bmatrix} 0.8895 \\ 1.7913 \end{bmatrix}$$

$$\begin{bmatrix} 1.7791 & 3.5826 \\ 1.7791 & -1.0000 \end{bmatrix}$$

$$\begin{bmatrix} 0.5159e-8 \\ 0.4887e-8 \end{bmatrix}$$

$$\begin{bmatrix} -0.2780e-8 \\ -0.0059e-8 \end{bmatrix}$$

So one of the two roots is located at $(x_1, x_2) = (0.8895, 1.7913)$. (The other is located at $(x_1, x_2) = (-0.8895, 1.7913)$ by symmetry. From an examination of the final value of $\frac{\partial x}{\partial x}$ both solutions have converged to an absolute error on x of less than 10^{-8} .

Exercício #2 (continuação)

```
x = [.5;.5]
for i = 1:100
    g1 = x(1)^2+x(2)^2-4
    g2 = x(1)^2-x(2)+1
    g = [g1;g2]
    h11 = 2*x(1)
    h12 = 2*x(2)
    h21 = 2*x(1)
    h22 = -1
    h = [h11 h12; h21 h22]
    if abs(inv(h)*g) < 10^-5
        break
    end
    x = x - inv(h)*g
end
```

Microdados

- IBGE (Censo, PNAD, PME, POF, INPC/IPCA ...)
- INEP (Censo Escolar, Enem, Enad, SAEB ...)
- Ministério do Trabalho (RAIS, CAGED e MEI)
- Ministério da Justiça (Infopen e Sisnarm)
- Ministério do Desenvolvimento (ComexStat)
- Ministério da Saúde (DataSUS)
- TSE (Eleições)
- Defesa Civil (S2ID)
- Tesouro Nacional (FINBra)
- ANP (SLP)
- BCB (DadosAbertos)
- Registros Civil e Comercial (nascimentos, casamentos, óbitos, abertura e fechamento de empresas etc.)

<https://www.ibge.gov.br/estatisticas/sociais/trabalho/27946-divulgacao-semanal-pnadcovid1.html>

 Instituto Brasileiro de Geografia e Estatística

Buscar

 Estatísticas Experimentais

Pesquisa Nacional por Amostra de Domicílios - PNAD COVID19

O que é

Microdados

Divulgação semanal

Setembro 2020 - semana 4

Principais resultados

Tabelas

Conceitos e métodos

Divulgação mensal

Downloads

Informações técnicas

Downloads

> Mensal

> Microdados

> Dados

PNAD_COVID_052020.zip

PNAD_COVID_062020.zip

PNAD_COVID_072020.zip

PNAD_COVID_082020.zip

PNAD_COVID_092020.zip

PNAD_COVID_102020.zip

PNAD_COVID_112020.zip

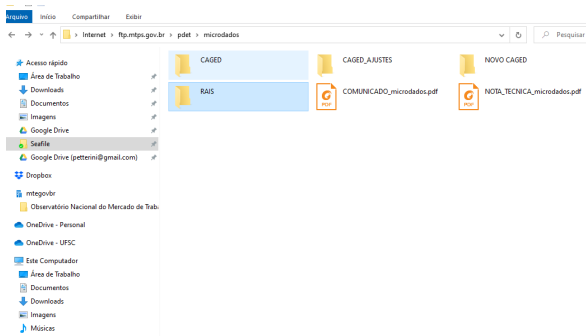
Documentacao

> Notas_Tecnicas

> Semanal

A	B	C	D	E	
Dicionário das variáveis da PNAD COVID					
Tamanho	Código da variável	Quesito		Categorias	
		nº	Descrição	Tipo	Descrição
Parte 1 - Identificação e Controle					
4	Ano		Ano de referência		
2	UF		Unidade da Federação	11	Rondônia
				12	Acre
				13	Amazonas
				14	Roraima
				15	Pará
				16	Amapá
				17	Tocantins
				21	Maranhão
				22	Piauí
				23	Ceará
				24	Rio Grande do Norte
				25	Paraíba
				26	Pernambuco
				27	Alagoas
				28	Sergipe
				29	Bahia
				31	Minas Gerais
				32	Espírito Santo
				33	Rio de Janeiro
				35	São Paulo
				41	Paraná
				42	Santa Catarina
				43	Rio Grande do Sul
				50	Mato Grosso do Sul
				51	Mato Grosso
				52	Goiás
				53	Distrito Federal
				11	Município de Porto Velho (RO)

- 1 Inspeccionar o dicionário, identificando mudanças entre períodos
- 2 Tabular algum estado
- 3 Explicar os diferentes tipos de variáveis (contínuas, categóricas, truncadas etc.)
- 4 Renomear variáveis de interesse
- 5 Fazer exercícios de estatística descritiva (gráficos, testes-t, kernel, lfit etc.)



RAIS: Vínculos + Estabelecimentos



[\[início \]](#)
[\[guias e tutoriais \]](#)
[\[perguntas frequentes \]](#)
[\[microdados \]](#)
[\[equipe \]](#)
[\[contato \]](#)

[\[english version \]](#)

Data Zoom: simplificando o acesso aos microdados do Brasil



Bem-vindo ao Data Zoom

O Data Zoom disponibiliza gratuitamente pacotes em Stata para a leitura dos microdados das pesquisas domiciliares do IBGE. Para gerar os dados, basta ter os dados originais e utilizar o programa. Além do instrumental para extração dos dados, o Data Zoom oferece ferramentas para compatibilização de pesquisas entre diferentes anos, construção de bases em painel e deflacionamento de valores monetários.

O Data Zoom foi desenvolvido pelo Departamento de Economia da PUC-Rio com financiamento da FINEP. O acesso é aberto. Envie uma cópia de seu trabalho para datazoom@econ.puc-rio.br. Favor nos citar em seus agradecimentos.

Novidade: O programa para a leitura dos microdados da Pesquisa de Orçamentos Familiares de 2017-18 já está disponível. Orientações para download e mais informações sobre a pesquisa, clique aqui.

Modelo linear

O modelo linear de uma equação é a pedra angular da Econometria...

$$y_i = E(y_i | X_i) + u_i \text{ tal que } \begin{cases} E(y_i | X_i) = X_i\beta \\ E(u_i | X_i) = 0 \Rightarrow E(u_i) = 0 \text{ [LEI]} \end{cases}$$

- Francis Galton e Karl Pearson: Biometrika (1901)
- Ragnar Frisch e Jan Tinbergen: Econometric Society (1933)
- Os termos “causalidade” e “provar” devem ser evitados ou, no mínimo, usados com muita cautela... o exemplo do Daniel Kahneman com a experiência dos instrutores de pilotos de caça

Considere que o experimento que gerou pares (x, y) ... sob a hipótese de linearidade temos que:

$$\begin{cases} y_1 = \beta_1 + \beta_2 x_1 + u_1 \\ y_2 = \beta_1 + \beta_2 x_2 + u_2 \\ y_3 = \beta_1 + \beta_2 x_3 + u_3 \\ y_4 = \beta_1 + \beta_2 x_4 + u_4 \\ y_5 = \beta_1 + \beta_2 x_5 + u_5 \\ y_6 = \beta_1 + \beta_2 x_6 + u_6 \end{cases}$$

Na prática pesquisador observa uma amostra (e.g., sem as observações $i = 1$ e $i = 6$ da população)

$$\begin{cases} y_2 = ? + ?x_2 + ? \\ y_3 = ? + ?x_3 + ? \\ y_4 = ? + ?x_4 + ? \\ y_5 = ? + ?x_5 + ? \end{cases}$$

- Como você sabe que o processo gerador é, de fato, linear?
- Como você sabe que a amostra é i.i.d.?

Sob a hipótese de linearidade, qual é o âmago do problema?

R.: trata-se da resolução de um sistema com n equações e $n + K$ incógnitas ... que possui infinitas respostas, mas um único e desconhecido processo gerador

Considere a representação matricial do sistema ...

- $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + u_i$

- $$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} = X\beta + u$$

Sob as hipóteses clássicas, β são parâmetros (**constantes**) que você deseja estimar, e u é uma **variável aleatória não observada** (erro, distúrbio etc.)

Note que:

- $y = E(y | X) + u \Rightarrow E(y | X) = X\beta + E(u | X) \dots E(u | X) = 0$ é o ponto!
- Você tem em mãos $\{y_i, X_i\}_{i=1}^n$, $n < N$, qual seria um critério para estimar β ?
- O análogo amostral de $E(u | X) = E(u) = 0$ é $\sum_{i=1}^n u_i/n^* = 0 \Rightarrow \sum_{i=1}^n (y_i - X_i\beta)/n^* = 0 \Rightarrow \bar{y} - \bar{X}\beta = 0$

[1 equação e K incógnitas, não ajuda muito]

- E se minimizar a variância amostral condicionada na média zero do erro?

$$\begin{cases} \text{Min.}_{\beta} f(\beta) = \sum_{i=1}^n w_i \frac{(y_i - X_i\beta - \hat{E}(u | X))^2}{n^*} \\ \text{s.r. : } \sum_{i=1}^n w_i (u_i/n^*) = 0 \\ w_i \geq 0 \forall i \end{cases} \Rightarrow \begin{cases} \text{Min.}_{\beta} f(\beta) = \sum_{i=1}^n w_i (y_i - X_i\beta)^2 \\ \text{s.r. : } x_{1i} = 1 \end{cases}$$

$$\bullet \begin{cases} -2 \sum_{i=1}^n w_i (y_i - X_i\hat{\beta}) = 0 \\ -2 \sum_{i=1}^n w_i (y_i - X_i\hat{\beta}) x_{2i} = 0 \\ \vdots \\ -2 \sum_{i=1}^n w_i (y_i - X_i\hat{\beta}) x_{Ki} = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^n w_i u_i = 0 \\ \sum_{i=1}^n w_i (u_i x_{2i}) = 0 \\ \vdots \\ \sum_{i=1}^n w_i (u_i x_{Ki}) = 0 \end{cases} \Rightarrow X'Wu = 0$$

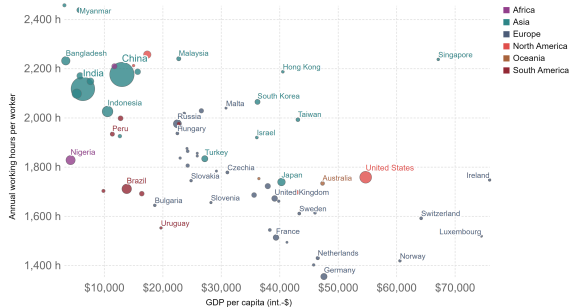
- onde W é uma matriz diagonal $n \times n$ de pesos
- intrinsecamente, para haver identificação de $\hat{\beta}$ se requer que a covariância entre o erro e **todos** os x_k seja 0
- $X'Wu = 0 \Rightarrow X'W(y - X\hat{\beta}) = 0 \Rightarrow X'Wy - X'WX\hat{\beta} = 0$
- Se existe $(X'WX)^{-1}$ (i.e., inexistente multicolinearidade perfeita), então $\hat{\beta}_{OLS/WLS} = (X'WX)^{-1}(X'Wy)$

Do workers in richer countries work longer hours?

- <https://ourworldindata.org/rich-poor-working-hours>
- http://petterini.paginas.ufsc.br/files/2021/08/data_hours_gdp.xls

Annual working hours vs. GDP per capita

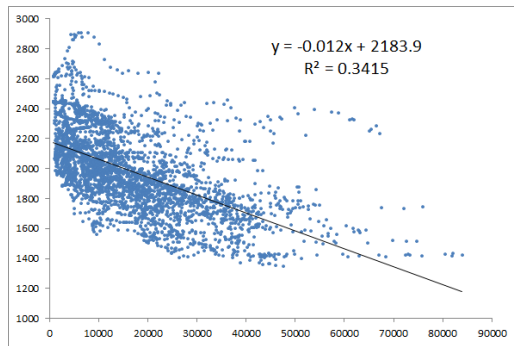
Working hours are the annual average per worker. GDP per capita is measured in constant 2011 international \$, which means it is adjusted for price differences between countries (PPP adjustment) and for inflation to allow comparisons between countries and over time.



Source: Feenstra et al. (2015) Penn World Table 9.1

OurWorldInData.org/working-hours • CC BY

- Erro (u) $\neq$ resíduo ($\hat{u}$)
- $\hat{Y} = X\hat{\beta} \Rightarrow Y = \hat{Y} + \hat{u}$
- $SQT = \sum_{i=1}^n (y_i - \bar{y})^2$
- $SQT = (Y - \bar{y})'(Y - \bar{y}) = ([\hat{Y} - \bar{y}] + \hat{u})'([\hat{Y} - \bar{y}] + \hat{u})$
- $SQT = (\hat{Y} - \bar{y})'(\hat{Y} - \bar{y}) + 2(\hat{Y} - \bar{y})'\hat{u} + \hat{u}'\hat{u}$
- $\hat{Y}'\hat{u} = (X\hat{\beta})'\hat{u} = \hat{\beta}'(X'\hat{u}) = 0$
- $\bar{y}'\hat{u} = 0$
- $SQT = SQE + SQR$
- $R^2 = 1 - SQR/SQT$



RESUMO DOS RESULTADOS								
<i>Estatística de regressão</i>								
R múltiplo	0.58436							
R-Quadrado	0.341477							
R-quadrado ajustado	0.341281							
Erro padrão	221.4691							
Observações	3373							
ANOVA								
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>			
Regressão	1	85738272	85738272	1748.028	3.7E-308			
Resíduo	3371	1.65E+08	49048.57					
Total	3372	2.51E+08						
	<i>Coefficiente</i>	<i>erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>% inferior</i>	<i>% superior</i>	<i>perior 95.0</i>	<i>perior 95.0</i>
Interseção	2183.903	6.113787	357.2095	0	2171.916	2195.89	2171.916	2195.89
gdp	-0.012	0.000286	-41.8094	3.7E-308	-0.01253	-0.01141	-0.01253	-0.01141

Matriz de variância-covariância clássica

- $\begin{cases} H_0 : \hat{\beta}_k = 0 \\ H_1 : \hat{\beta}_k \neq 0 \end{cases}$

- $\frac{\hat{\beta}_k}{\text{ep}_k} \sim t_{n-1}$ onde ep_k é uma estimativa de $\sqrt{E(\hat{\beta}_k - \beta_k)^2}$

- $\hat{\beta} - \beta = (X'X)^{-1}(X'u)$

- $(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (X'X)^{-1}(X'u)((X'X)^{-1}(X'u))'$

- $\underbrace{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'}_{K \times K} = (X'X)^{-1}X' \underbrace{uu'}_{n \times n} X(X'X)^{-1}$

$$\underbrace{\begin{bmatrix} (\hat{\beta}_1 - \beta_1)^2 & (\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) & \cdots & (\hat{\beta}_1 - \beta_1)(\hat{\beta}_K - \beta_K) \\ (\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1) & (\hat{\beta}_2 - \beta_2)^2 & \cdots & (\hat{\beta}_2 - \beta_2)(\hat{\beta}_K - \beta_K) \\ \vdots & \vdots & \ddots & \vdots \\ (\hat{\beta}_K - \beta_K)(\hat{\beta}_1 - \beta_1) & (\hat{\beta}_K - \beta_K)(\hat{\beta}_2 - \beta_2) & \cdots & (\hat{\beta}_K - \beta_K)^2 \end{bmatrix}}_{\text{a diagonal tem estrutura } (X - E(X))^2}$$

Portanto, a estrutura de variância-covariância de $\hat{\beta}$ depende de... $uu' = \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_n \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n u_1 & u_n u_2 & \cdots & u_n^2 \end{bmatrix} \dots E(uu' | X) = \sigma^2 I_n \Rightarrow \hat{E}(uu' | X) = \frac{SQR}{n - K} (X'X)^{-1}$

Exercício #3 (Stata)

```
clear all
cls

import excel "C:\...\data_hours_gdp.xls",
             sheet("data_hours_gdp") firstrow

*tw (scatter gdp hours) (lfit gdp hours)

gen ones = 1
mkmat ones, matrix(ones)
matrix list ones

mkmat ones gdp, matrix(X)
matrix list X

mkmat hours, matrix(Y)
matrix list Y

matrix b = inv(X' * X) * (X' * Y)
matrix list b

matrix SQR = (Y - X * b)' * (Y - X * b)
matrix list SQR

matrix SQT = (Y - ((ones' * Y) / rowsof(Y)) * ones)' *
```

```
(Y - ((ones' * Y) / rowsof(Y)) * ones)
matrix list SQT

matrix R2 = 1 - SQR * inv(SQT)
matrix list R2

matrix VarCov = SQR / (rowsof(X) - colsof(X)) * inv(X' * X)
matrix list VarCov

matrix ep = VarCov
forval i = 1 / '=' rowsof(VarCov)' {
    forval j = 1 / '=' colsof(VarCov)' {
        matrix ep['i', 'j'] = sqrt(VarCov['i', 'j'])
    }
}
matrix ep = vecdiag(ep)'
matrix list ep

matrix t = ep
forval i = 1 / '=' rowsof(ep)' {
    matrix t['i', 1] = b['i', 1] / ep['i', 1]
}
matrix list t

reg hours gdp
```

- $$\begin{cases} H_0 : \underbrace{R}_{s \times K} \beta - \underbrace{r}_{s \times 1} = [0] \\ H_1 : \underbrace{R}_{s \times K} \beta - \underbrace{r}_{s \times 1} \neq [0] \end{cases}$$

- Sob a hipótese de que $u \sim N([0], \sigma^2 I_n) \Rightarrow \hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$

- Teste de Wald:**

- $$\underbrace{W}_{s \times 1} = (R\hat{\beta} - r)' [\sigma^2 R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r) \sim \underbrace{\chi^2}_{s \times 1}$$

- $$\underbrace{W}_{s \times 1} = (R\hat{\beta} - r)' \left[s \frac{SQR}{n - K} R(X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \sim \underbrace{F(s, n - K)}_{s \times 1}$$

Há vários outros tipos de testes

APPENDIX 4A

4A.1 MAXIMUM LIKELIHOOD ESTIMATION OF TWO-VARIABLE REGRESSION MODEL

Assume that in the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$ the Y_i are normally and independently distributed with mean $= \beta_1 + \beta_2 X_i$ and variance $= \sigma^2$. [See Eq. (4.3.9).] As a result, the joint probability density function of $Y_1, Y_2, \dots, Y_n$, given the preceding mean and variance, can be written as

$$f(Y_1, Y_2, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2)$$

But in view of the independence of the Y 's, this joint probability density function can be written as a product of n individual density functions as

$$\begin{aligned} f(Y_1, Y_2, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2) \\ = f(Y_1 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 | \beta_1 + \beta_2 X_i, \sigma^2) \cdots f(Y_n | \beta_1 + \beta_2 X_i, \sigma^2) \end{aligned} \quad (1)$$

where

$$f(Y_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \right\} \quad (2)$$

which is the density function of a normally distributed variable with the given mean and variance.

(Note: exp means e to the power of the expression indicated by [].)

Substituting (2) for each Y_i into (1) gives

$$f(Y_1, Y_2, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \right\} \quad (3)$$

If $Y_1, Y_2, \dots, Y_n$ are known or given, but β_1, β_2 , and σ^2 are not known, the function in (3) is called a **likelihood function**, denoted by $LF(\beta_1, \beta_2, \sigma^2)$,

$$\begin{aligned}\ln \text{LF} &= -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \\ &= -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\end{aligned}\quad (5)$$

Differentiating (5) partially with respect to β_1 , β_2 , and σ^2 , we obtain

$$\frac{\partial \ln \text{LF}}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum (Y_i - \beta_1 - \beta_2 X_i)(-1) \quad (6)$$

$$\frac{\partial \ln \text{LF}}{\partial \beta_2} = -\frac{1}{\sigma^2} \sum (Y_i - \beta_1 - \beta_2 X_i)(-X_i) \quad (7)$$

$$\frac{\partial \ln \text{LF}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_1 - \beta_2 X_i)^2 \quad (8)$$

Setting these equations equal to zero (the first-order condition for optimization) and letting $\tilde{\beta}_1$, $\tilde{\beta}_2$, and $\tilde{\sigma}^2$ denote the ML estimators, we obtain³

$$\frac{1}{\tilde{\sigma}^2} \sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) = 0 \quad (9)$$

$$\frac{1}{\tilde{\sigma}^2} \sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) X_i = 0 \quad (10)$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = 0 \quad (11)$$

$u_i = y_i - X_i\beta$ e é conveniente escrever $\sigma = \exp(\gamma)$

$$L = \prod_{i=1}^n \frac{\exp\left(-\frac{1}{2} \frac{u_i^2}{\sigma^2}\right)}{\sigma \sqrt{2\pi}} = \prod_{i=1}^n \frac{\exp\left(-\frac{1}{2} \frac{u_i^2}{\exp(2\gamma)}\right)}{\exp(\gamma) \sqrt{2\pi}}$$

$$\ln L = \sum_{i=1}^n \left(-\gamma - \frac{\ln(2\pi)}{2} - \frac{1}{2} \frac{u_i^2}{\exp(2\gamma)} \right)$$

$$\ln L \propto - \sum_{i=1}^n \left(2\gamma + \frac{u_i^2}{\exp(2\gamma)} \right)$$

Exercício #4

```
clear
set obs 100
set seed 12345
gen x = invnormal(uniform())
gen y = 2*x + invnormal(uniform())
reg y x

program drop _all
program mlols
    version 16.0
    args lnf xb gamma
    tempvar u
    quietly{
        gen `u' = $ML_y1-`xb'
        replace `lnf' = -2*`gamma' - (`u'^2)/exp(2*`gamma')
    }
end

ml model lf mlols (y = x) ()
ml check

ml maximize
```

Violações das hipóteses

∃ 3 hipóteses básicas (além da linearidade)

- 1 posto($X'X$) = $K \Rightarrow Cov(x_k, x_l) \neq 0 \forall k \neq l$; se $Cov(x_k, x_l) \approx 0$ há multicolinearidade (às vezes micronumerosidade)
- 2 $E(uu') = \sigma^2 I_n$, cuja violação (mesmo que parcial) é chamada de heterocedasticidade e/ou autocorrelação residual
- 3 $E(X'u) = [0]$, cuja violação (mesmo que parcial) é chamada de endogeneidade (às vezes omissão)

O PROBLEMA MAIOR ESTÁ NESSA ÚLTIMA...

- $y_i - \bar{y} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i - \bar{y} \Rightarrow y_i^* = \beta_2 x_{i2}^* + \beta_3 x_{i3}^* + u_i$
- $$\begin{cases} \hat{\beta}_2 = \frac{(\sum y_i^* x_{i2}^*)(\sum x_{i3}^{*2}) - (\sum y_i^* x_{i3}^*)(\sum x_{i2}^* x_{i3}^*)}{(\sum x_{i2}^{*2})(\sum x_{i3}^{*2}) - (\sum x_{i2}^* x_{i3}^*)^2} \\ \hat{\beta}_3 = \frac{(\sum y_i^* x_{i3}^*)(\sum x_{i2}^{*2}) - (\sum y_i^* x_{i2}^*)(\sum x_{i2}^* x_{i3}^*)}{(\sum x_{i2}^{*2})(\sum x_{i3}^{*2}) - (\sum x_{i2}^* x_{i3}^*)^2} \end{cases}$$
- Sob multicolinearidade perfeita (i.e., $x_3 = \lambda x_2$) o numerador do estimador de OLS é zero
- Sob multicolinearidade menos perfeita (i.e., $x_3 = \lambda x_2 + \nu_i$) o numerador do estimador de OLS não é zero
- $$\hat{\beta}_2 = \frac{(\sum y_i^* x_{i2}^*)(\lambda^2 \sum x_{i2}^{*2} + \sum \nu_i^2) - (\lambda \sum y_i^* x_{i2}^* + \sum y_i \nu_i)(\lambda^2 \sum x_{i2}^{*2})}{(\sum x_{i2}^{*2})(\lambda^2 \sum x_{i2}^{*2} + \sum \nu_i^2) - (\lambda^2 \sum x_{i2}^{*2})^2}$$
 [adiante se vê que isso não é viesado]

$$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{i2}^{*2}(1 - R_{23}^2)} \text{ e } V(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{i3}^{*2}(1 - R_{23}^2)}, \text{ em que } R_{23}^2 \text{ diz respeito a } x_3 = \lambda x_2 + \nu_i$$

Portanto, multicolinearidade é um problema para inferência! ...

https://en.wikipedia.org/wiki/Principal_component_analysis

$$E(uu') \neq \sigma^2 I_n$$

- Heterocedasticidade

- GLS/FGLS

- $y = X\beta + u$ com $E(uu' | X) = \sigma^2 \Omega$
 - $\hat{\beta}_{GLS} = (X' \Omega X)^{-1} (X' \Omega y)$
 - $\hat{\beta}_{FGLS} = (X' \hat{\Omega} X)^{-1} (X' \hat{\Omega} y)$

- Matriz de variância-covariância robusta:

$$\begin{aligned} V_{\hat{\beta}_{OLS}} &= \frac{1}{n} \left(\frac{1}{n} \sum_i X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_i X_i X_i' \hat{u}_i^2 \right) \left(\frac{1}{n} \sum_i X_i X_i' \right)^{-1} \\ &= (X' X)^{-1} (X' \text{diag}(\hat{u}_1^2, \dots, \hat{u}_n^2) X) (X' X)^{-1} \end{aligned}$$

- Bootstrap

- Autocorrelação é um tema mais de séries temporais, e tem importância menor na microeconometria

- $\hat{\beta}_{OLS} = (X'X)^{-1}(X'y) \Rightarrow \hat{\beta} = (X'X)^{-1}(X'(X\beta + u))$
- $\hat{\beta} = (X'X)^{-1}(X'X)\beta + (X'X)^{-1}(X'u) \Rightarrow \hat{\beta} = \beta + (X'X)^{-1}(X'u)$
- $\hat{\beta} = \beta + ((1/n)X'X)^{-1}((1/n)X'u) \Rightarrow \lim_{n \rightarrow \infty} \hat{\beta} = \beta + E(X'X)^{-1}E(X'u)$
- $\hat{\beta}$ é não viesado e consistente se $E(X'u) = 0$
- Há três razões para que falhe esta hipótese, e quando isso acontece, dizemos que há “endogeneidade”
 - 1 X está medido de forma muito errada (isso não costuma ser um problema para n grande)
 - 2 u contém variáveis omitidas bastante correlacionadas com pelo menos um dos X
 - 3 X e y são definidos simultaneamente

$$y_i = \beta_1 + \beta_2 x_i + \underbrace{\beta_3 O_i + \xi_i}_{u_i}$$

$$\hat{\beta} = \beta + ((1/n)X'X)^{-1}((1/n)X'u)$$

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(1/n)X'X = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{\sum_{i=1}^n x_i^2}{n} \end{bmatrix}$$

$$X'u = \begin{bmatrix} \beta_3 \sum_{i=1}^n O_i + \sum_{i=1}^n \xi_i \\ \beta_3 \sum_{i=1}^n O_i x_i + \sum_{i=1}^n \xi_i x_i \end{bmatrix}$$

Supondo $E(X'\xi) = 0$,

$$(1/n)X'u = \beta_3 \begin{bmatrix} \bar{O} \\ \frac{\sum_{i=1}^n O_i x_i}{n} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \beta_3 \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{\sum_{i=1}^n x_i^2}{n} \end{bmatrix}^{-1} \begin{bmatrix} \bar{O} \\ \frac{\sum_{i=1}^n O_i x_i}{n} \end{bmatrix}$$

Note que quanto maior a correlação entre a covariada x e a variável omitida (maior β_3 em valor absoluto), maior será o viés

Note também que, às vezes, podemos não nos preocupar com a magnitude do viés, mas apenas com o sinal

$$y_i = \beta_1 + \beta_2(x_i + O_i) + \xi_i$$

$$y_i = \beta_1 + \beta_2 x_i + \underbrace{\quad}_{\beta_2 O_i + \xi_i} u_i$$

(...) cai no resultado anterior

Note que se $\bar{O} = 0$ e $C(X, O) = 0$ os erros de medida não são problema

A Lei dos Grandes Números mitiga o primeiro, desde que se use um intercepto

Mas os questionários autodeclarados podem ser um problema no segundo – e.g., quando se pergunta sobre a renda, a maioria das pessoas respondem valores menores do que de fato ganham

Monte Carlo

Exercício #5

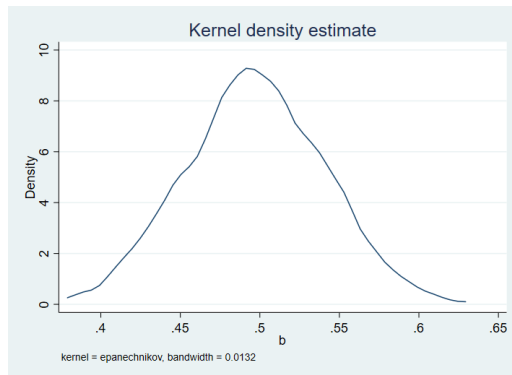
```
clear all
cls
set obs 100
set seed 12345

scalar constant = .5
scalar slope = .5
scalar sigma2 = 1
scalar alpha = .05
gen x = exp(rnormal())
gen Ey = constant + slope*x /* mean function */
gen u = . /* initiate u -- all missing values */
gen y = . /* initiate y -- all missing values */
tempname sim

postfile `sim' b se coverage using results, replace
quietly {
  forvalues i = 1/200 {
    replace u = rnormal(0,sqrt(sigma2))
    replace y = Ey + u
    reg y x
    scalar b = _b[x]
    scalar se = _se[x]
    scalar lb = b - se*invttail(e(df_r),alpha/2)
    scalar ub = b + se*invttail(e(df_r),alpha/2)
    scalar pv = slope<ub & slope>lb
    post `sim' (b) (se) (pv)
  }
}
postclose `sim'
```

```
use results, clear
summarize
```

```
kdensity b
```




```

clear all
cls
set obs 100
set seed 12345

scalar constant = .5
scalar slope = .5
scalar sigma2 = 1
scalar alpha = .05
gen x = exp(rnormal())
gen Ey = constant + slope*x /* mean function */
gen u =. /* initiate u -- all missing values */
gen y =. /* initiate y -- all missing values */
tempname sim

postfile `sim' b se coverage using results, replace
quietly {
forvalues i = 1/200 {
replace u = rnormal(
-X
, sqrt(sigma2))
replace y = Ey + u
reg y x
scalar b = _b[x]
scalar se = _se[x]
scalar lb = b - se*invttail(e(df_r), alpha/2)
scalar ub = b + se*invttail(e(df_r), alpha/2)
scalar pv = slope<ub & slope>lb
post `sim' (b) (se) (pv)
}
}

```

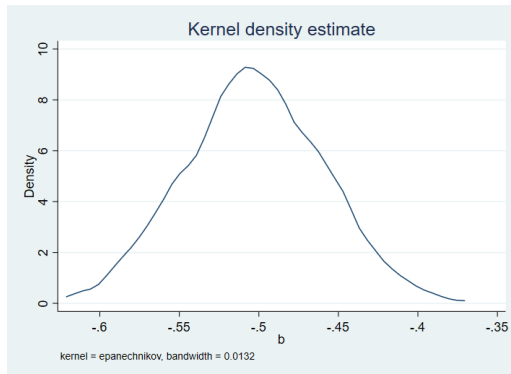
```

postclose `sim'

use results, clear
summarize

kdensity b

```



omite x_2

```
clear all
cls

set obs 100
set seed 12345

* Set the values of the parameters
scalar constant = .5
scalar slope1 = .5
scalar slope2 = .5
scalar sigma2 = 1
scalar alpha = .05

* Define the exogenous variable
gen x1 = exp(rnormal())
gen x2 = cond(rnormal()>1,1,0)

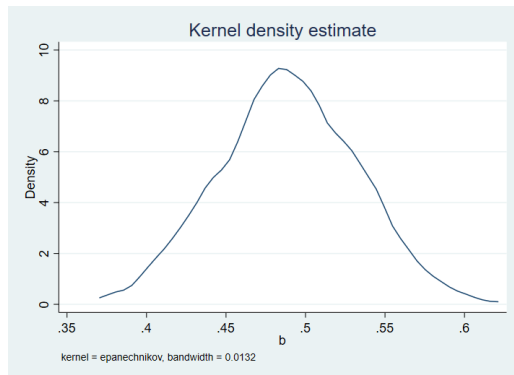
* Take the experiment
gen Ey = constant + slope1*x1 + slope2*x2 /* mean function */
gen u =. /* initiate u -- all missing values */
gen y =. /* initiate y -- all missing values */
tempname sim

postfile `sim' b se coverage using results, replace
quietly {
  forvalues i = 1/200 {
    replace u = rnormal(0,sqrt(sigma2))
    replace y = Ey + u
    reg y x1
    scalar b = _b[x1]
    scalar se = _se[x1]
    scalar lb = b - se*invttail(e(df_r),alpha/2)
    scalar ub = b + se*invttail(e(df_r),alpha/2)
    scalar pv = slope1<ub & slope1>lb
```

```
post `sim' (b) (se) (pv)
}
}
postclose `sim'
```

```
use results, clear
summarize
```

```
kdensity b
```



mais endogeneidade

```
clear all
cls

set obs 100
set seed 12345

* Set the values of the parameters
scalar constant = .5
scalar slope1 = .5
scalar slope2 = .5
scalar sigma2 = 1
scalar alpha = .05

* Define the exogenous variable
gen x1 = exp(rnormal())
gen x2 = x1*cond(rnormal()>0,1,0)

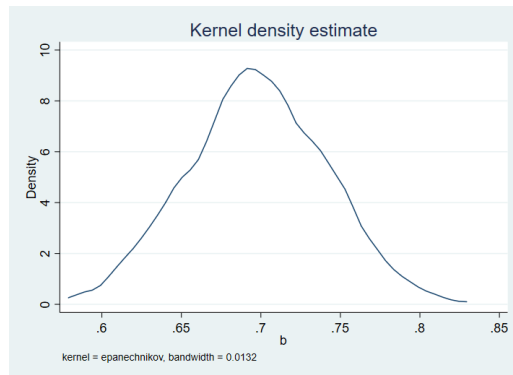
* Take the experiment
gen Ey = constant + slope1*x1 + slope2*x2 /* mean function */
gen u =. /* initiate u -- all missing values */
gen y =. /* initiate y -- all missing values */
tempname sim

postfile `sim' b se coverage using results, replace
quietly {
  forvalues i = 1/200 {
    replace u = rnormal(0,sqrt(sigma2))
    replace y = Ey + u
    reg y x1
    scalar b = _b[x1]
    scalar se = _se[x1]
    scalar lb = b - se*invttail(e(df_r),alpha/2)
    scalar ub = b + se*invttail(e(df_r),alpha/2)
    scalar pv = slope1<ub & slope1>lb
```

```
post `sim' (b) (se) (pv)
}
}
postclose `sim'
```

```
use results, clear
summarize
```

```
kdensity b
```



a hipótese de linearidade falha

```
clear all
cls

set obs 100
set seed 12345

* Set the values of the parameters
scalar constant = .5
scalar slope1 = .5
scalar slope2 = .5
scalar sigma2 = 1
scalar alpha = .05

* Define the exogenous variable
gen x1 = exp(rnormal())
gen x2 = cond(rnormal()>0,1,0)

* Take the experiment
gen Ey = constant*(x1^slope1)*exp(slope2*x2) /* mean function */
gen u =. /* initiate u -- all missing values */
gen y =. /* initiate y -- all missing values */
tempname sim

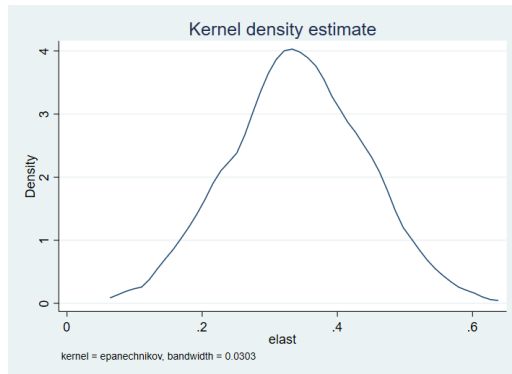
sum x1 Ey

postfile `sim' b elast using results, replace
quietly {
  forvalues i = 1/200 {
    replace u = rnormal(0,sqrt(sigma2))
    replace y = Ey + u
    reg y x1 x2
    scalar b = _b[x1]
    scalar elast = b*1.81012/.7898326
    post `sim' (b) (elast)
```

```
}
}
postclose `sim'

use results, clear
summarize

kdensity elast
```



Então, o recado é...

- A teoria é essencial para se definir a especificação
- Estatística t não significativa não é necessariamente um problema, pode ser simplesmente multicolinearidade ou erro na definição da matriz de variância-covariância
- Não linearidade não é um problema, porque as elasticidades costumam apontar na direção correta a despeito de formas funcionais
- Variável omitida *per si* pode não ser um problema
- **Endogeneidade** é o problema!

Instrumentos e painel

5.1 Instrumental Variables and Two-Stage Least Squares

5.1.1 Motivation for Instrumental Variables Estimation

To motivate the need for the method of instrumental variables, consider a linear population model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u, \quad (5.1)$$

$$E(u) = 0, \quad \text{Cov}(x_j, u) = 0, \quad j = 1, 2, \dots, K - 1, \quad (5.2)$$

$$\hat{\boldsymbol{\beta}} = \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' y_i \right) = \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' u_i \right)$$

$$y = \mathbf{x}\boldsymbol{\beta} + u, \quad (5.7)$$

where the constant is absorbed into $\mathbf{x}$ so that $\mathbf{x} = (1, x_2, \dots, x_K)$. Write the $1 \times K$ vector of all exogenous variables as

$$\mathbf{z} \equiv (1, x_2, \dots, x_{K-1}, z_1).$$

Assumptions (5.2) and (5.3) imply the K population orthogonality conditions

$$E(\mathbf{z}'u) = \mathbf{0}. \quad (5.8)$$

Multiplying equation (5.7) through by $\mathbf{z}'$, taking expectations, and using equation (5.8) gives

$$[E(\mathbf{z}'\mathbf{x})]\boldsymbol{\beta} = E(\mathbf{z}'y), \quad (5.9)$$

where $E(\mathbf{z}'\mathbf{x})$ is $K \times K$ and $E(\mathbf{z}'y)$ is $K \times 1$. Equation (5.9) represents a system of K linear equations in the K unknowns $\beta_1, \beta_2, \dots, \beta_K$. This system has a unique solution if and only if the $K \times K$ matrix $E(\mathbf{z}'\mathbf{x})$ has full rank; that is,

$$\text{rank } E(\mathbf{z}'\mathbf{x}) = K, \quad (5.10)$$

in which case the solution is

$$\boldsymbol{\beta} = [E(\mathbf{z}'\mathbf{x})]^{-1}E(\mathbf{z}'y). \quad (5.11)$$

Example 5.1 (Instrumental Variables for Education in a Wage Equation): Consider a wage equation for the U.S. working population:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + u, \quad (5.12)$$

where u is thought to be correlated with educ because of omitted ability, as well as other factors, such as quality of education and family background. Suppose that we can collect data on mother's education, motheduc . For this to be a valid instrument for educ we must assume that motheduc is uncorrelated with u and that $\theta_1 \neq 0$ in the reduced-form equation

$$\text{educ} = \delta_0 + \delta_1 \text{exper} + \delta_2 \text{exper}^2 + \theta_1 \text{motheduc} + r.$$

There is little doubt that educ and motheduc are partially correlated, and this assertion is easily tested given a random sample from the population. The potential problem with motheduc as an instrument for educ is that motheduc might be correlated with the omitted factors in u : mother's education is likely to be correlated with child's ability and other family background characteristics that might be in u .

Preview

Lifetime Earnings and the Vietnam Era Draft Lottery:
Evidence from Social Security
Administrative Records

By JOSHUA D. ANGRIST*

The randomly assigned risk of induction generated by the draft lottery is used to construct estimates of the effect of veteran status on civilian earnings. These estimates are not biased by the fact that certain types of men are more likely than others to serve in the military. Social Security administrative records indicate that in the early 1980s, long after their service in Vietnam was ended, the earnings of white veterans were approximately 15 percent less than the earnings of comparable nonveterans. (JEL 824)

A central question in the debate over military manpower policy is whether veterans are adequately compensated for their service. The political process clearly reflects the desire to compensate veterans: since World War II, millions of veterans have received

schooling. Regarding the general position of veterans, a member of the Twentieth Century Fund's Task Force on Policies Toward Veterans concludes that "Within any age group, veterans have higher incomes, more education, and lower unemployment rates

JOURNAL ARTICLE

Lifetime Earnings and the
Vietnam Era Draft
Lottery: Evidence from
Social Security
Administrative Records

Joshua D. Angrist

The American
Economic Review
Vol. 80, No. 3 (Jun.,
1990), pp. 313-336 (24
pages)
Published By:
American Economic
Association



<https://www.jstor.org/stable/2006669>

5.1.2 Multiple Instruments: Two-Stage Least Squares

Consider again the model (5.1) and (5.2), where x_K can be correlated with u . Now, however, assume that we have more than one instrumental variable for x_K . Let $z_1, z_2, \dots, z_M$ be variables such that

$$\text{Cov}(z_h, u) = 0, \quad h = 1, 2, \dots, M, \quad (5.13)$$

so that each z_h is exogenous in equation (5.1). If each of these has some partial correlation with x_K , we could have M different IV estimators. Actually, there are many more than this—more than we can count—since any linear combination of $x_1, x_2, \dots, x_{K-1}, z_1, z_2, \dots, z_M$ is uncorrelated with u . So which IV estimator should we use?

In Section 5.2.3 we show that, under certain assumptions, the **two-stage least squares (2SLS) estimator** is the most efficient IV estimator. For now, we rely on intuition.

$$\hat{x}_{iK} = \hat{\delta}_0 + \hat{\delta}_1 x_{i1} + \dots + \hat{\delta}_{K-1} x_{i,K-1} + \hat{\theta}_1 z_{i1} + \dots + \hat{\theta}_M z_{iM}. \quad (5.16)$$

Now, for each observation i , define the vector $\hat{\mathbf{x}}_i \equiv (1, x_{i1}, \dots, x_{i,K-1}, \hat{x}_{iK})$, $i = 1, 2, \dots, N$. Using $\hat{\mathbf{x}}_i$ as the instruments for $\mathbf{x}_i$ gives the IV estimator

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \hat{\mathbf{x}}_i' \mathbf{x}_i \right)^{-1} \left(\sum_{i=1}^N \hat{\mathbf{x}}_i' y_i \right) = (\hat{\mathbf{X}}' \mathbf{X})^{-1} \hat{\mathbf{X}}' \mathbf{Y}, \quad (5.17)$$

THEOREM 5.1 (Consistency of 2SLS): Under Assumptions 2SLS.1 and 2SLS.2, the 2SLS estimator obtained from a random sample is consistent for β .

Proof: Write

$$\hat{\beta} = \beta + \left[\left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{z}_i \right) \left(N^{-1} \sum_{i=1}^N \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{z}_i' \mathbf{x}_i \right) \right]^{-1} \\ \cdot \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{z}_i \right) \left(N^{-1} \sum_{i=1}^N \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{z}_i' \mathbf{u}_i \right)$$

Even in large samples, IV methods can be ill-behaved if we have **weak instruments**. Consider the simple model $y = \beta_0 + \beta_1 x_1 + u$, where we use z_1 as an instrument for x_1 . Assuming that $\text{Cov}(z_1, x_1) \neq 0$, the plim of the IV estimator is easily shown to be

$$\text{plim } \hat{\beta}_1 = \beta_1 + \text{Cov}(z_1, u) / \text{Cov}(z_1, x_1),$$

which can be written as

$$\text{plim } \hat{\beta}_1 = \beta_1 + (\sigma_u / \sigma_{x_1}) [\text{Corr}(z_1, u) / \text{Corr}(z_1, x_1)], \quad (5.36)$$

where $\text{Corr}(\cdot, \cdot)$ denotes correlation. From this equation we see that if z_1 and u are correlated, the inconsistency in the IV estimator gets arbitrarily large as $\text{Corr}(z_1, x_1)$ gets close to zero. Thus, seemingly small correlations between z_1 and u can cause severe inconsistency—and therefore severe finite sample bias—if z_1 is only weakly correlated with x_1 . In fact, it may be better to just use OLS, even if we look only at the inconsistencies in the estimators. To see why, let $\tilde{\beta}_1$ denote the OLS estimator and write its plim as

$$\text{plim}(\tilde{\beta}_1) = \beta_1 + (\sigma_u / \sigma_{x_1}) \text{Corr}(x_1, u). \quad (5.37)$$

- $y_i = \beta_1 + \beta_2 x_i + u_i$; $E(X'u) \neq [0]$; $E(Z'u) = [0]$; $E(X'Z) \neq [0]$
- $\hat{\beta}_{IV} = (Z'X)^{-1}(Z'Y) \Rightarrow \hat{\beta}_2 = \frac{\sum(z_i - \bar{z})(y_i - \bar{y})}{\sum(z_i - \bar{z})(x_i - \bar{x})}$
- 2SLS
 - $X = Z\delta + \xi$
 - $\hat{\delta} = (Z'Z)^{-1}Z'X$
 - $\hat{X} = Z\hat{\delta} = Z(Z'Z)^{-1}Z'X$
 - $Y = \hat{X}\beta + u$
 - $\hat{\beta}_{2SLS} = ((Z(Z'Z)^{-1}Z'X)'(Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'Y$

File
Edit
History
Help

help ivregress X

+

[R] **ivregress** — Single-equation instrumental-variables regression
[\(View complete PDF manual entry\)](#)

Syntax

ivregress estimator *depvar* [*varlist1*] (*varlist2* = *varlist_iv*) [*if*] [*in*] [*weight*] [, options]

varlist1 is the list of exogenous variables.
varlist2 is the list of endogenous variables.
varlist_iv is the list of exogenous variables used with varlist1 as instruments for varlist2.

estimator	Description
2sls	two-stage least squares (2SLS)
liml	limited-information maximum likelihood (LIML)
gmm	generalized method of moments (GMM)

options	Description
Model	
<code>noconstant</code>	suppress constant term
<code>hasconst</code>	has user-supplied constant
# GMM	
<code>wmatrix(wmtype)</code>	wmtype may be <code>robust</code> , <code>cluster</code> <i>clustvar</i> , <code>hac</code> <i>kernel</i> , or <code>unadjusted</code>
<code>center</code>	center moments in weight matrix computation
<code>lgmm</code>	use iterative instead of two-step GMM estimator
<code>* eps(#)</code>	specify # for parameter convergence criterion; default is <code>eps(1e-6)</code>
<code>* weps(#)</code>	specify # for weight matrix convergence criterion; default is <code>weps(1e-6)</code>
<code>* optimization_options</code>	control the optimization process; seldom used
SE/Robust	
<code>vce(vcetype)</code>	vcetype may be <code>unadjusted</code> , <code>robust</code> , <code>cluster</code> <i>clustvar</i> , <code>bootstrap</code> , <code>jackknife</code> , or <code>hac</code> <i>kernel</i>
Reporting	
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
<code>first</code>	report first-stage regression
<code>small</code>	make degrees-of-freedom adjustments and report small-sample statistics
<code>noheader</code>	display only the coefficient table
<code>deprname(deprname)</code>	substitute dependent variable name
<code>iform(string)</code>	report exponentiated coefficients and use string to label them
<code>display_options</code>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
<code>perfect</code>	do not check for collinearity between endogenous regressors and excluded instruments
<code>coeflegend</code>	display legend instead of statistics

These options may be specified only when `gmm` is specified.
* These options may be specified only when `lgmm` is specified.
unadjusted *unadjusted* and *unadjusted* do not contain further variables: see *fourvarlist*

Exercício #6

```
clear

webuse hsng2

ivregress 2sls rent pcturban (hsngval = faminc i.region)
estimates store reg1

reg hsngval faminc i.region pcturban
estimates store reg2
predict x_hat, xb

reg rent pcturban x_hat
estimates store reg3

esttab reg1 reg2 reg3
```

```
. esttab reg1 reg2 reg3
```

	(1) rent	(2) hsngval	(3) rent
hsngval	0.00224*** (6.82)		
pcturban	0.0815 (0.27)	182.2 (1.58)	0.0815 (0.32)
faminc		2.731*** (4.01)	
1.region		0 (.)	
2.region		-5095.0 (-1.24)	
3.region		-1778.1 (-0.44)	
4.region		13413.8** (3.31)	
x_hat			0.00224*** (8.03)
_cons	120.7*** (7.93)	-18671.9 (-1.56)	120.7*** (9.33)
N	50	50	50

t statistics in parentheses

* p<0.05, ** p<0.01, *** p<0.001

- **Hausman Test for Exogeneity:** We calculate the squared difference of the estimators, normalised by the covariance of the difference.
- Let $\mathbf{d} = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})$, and write:

$$\begin{aligned} H &= \mathbf{d}' \widehat{\text{Var}}[\mathbf{d}]^{-1} \mathbf{d} \\ &= \mathbf{d}' \left[\hat{\sigma}^2 \left((\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1} - (\mathbf{X}' \mathbf{X})^{-1} \right) \right]^{-1} \mathbf{d} \stackrel{a}{\sim} \chi^2(k^+), \end{aligned}$$

where k^+ is the number of potentially endogenous variables and $\hat{\sigma}^2$ is either the OLS or IV estimator of σ^2 (both are consistent).

... continuação do Exercício #6

```
ivregress 2sls rent pcturban (hsngval = faminc i.region)
estimates store iv
```

```
reg hsngval faminc i.region pcturban
hausman iv ., sigmamore
```

---- Coefficients ----				
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	iv	.	Difference	S.E.
pcturban	.081516	182.2201	-182.1386	48.24537

b = consistent under H_0 and H_a ; obtained from ivregress
B = inconsistent under H_a , efficient under H_0 ; obtained from regress

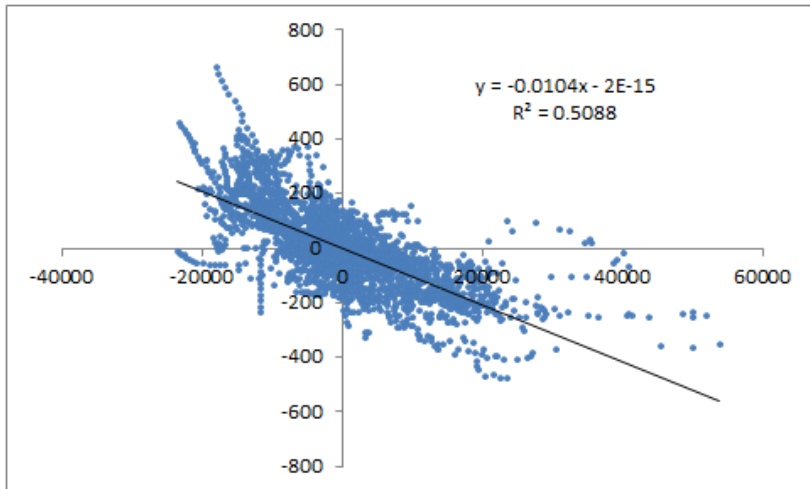
Test: H_0 : difference in coefficients not systematic

$$\begin{aligned}\chi^2(1) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\ &= 14.25 \\ \text{Prob}>\chi^2 &= 0.0002\end{aligned}$$

- $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \underbrace{\beta_3 O_i + \xi_i}_{u_i}$
- Se o pesquisador observa i em $t = 1, \dots, T$ a endogeneidade pode ser parcialmente mitigada pela introdução de efeitos fixos
- $y_{it} = \beta_1 + \beta_2 x_{it2} + \beta_3 x_{it3} + \underbrace{\beta_3 O_i + \xi_{it}}_{u_{it}}$
- $y_{it} = \beta_1 + \beta_2 x_{it2} + \beta_3 x_{it3} + \sum_{i=1}^n \alpha_i \mathbf{1}_i + \xi_{it}$
- $y_{it} = \beta_1 + \beta_2 x_{it2} + \beta_3 x_{it3} + \alpha_i + \xi_{it}$
- $y_{it} - \bar{y}_i = \beta_1 + \beta_2 x_{it2} + \beta_3 x_{it3} + \alpha_i + \xi_{it} - \bar{y}_i$
- $\bar{y}_i^* = \beta_2 x_{i2}^* + \beta_3 x_{i3}^* + \xi_i^*$

Do workers in richer countries work longer hours?

Pincel de Formatação		N Z S		T A		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M		M M			
----------------------	--	-------	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	--	--



Assuming joint normality of the estimators.^{[3][6]}

$$\sqrt{N} \begin{bmatrix} b_1 - b \\ b_0 - b \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \text{Var}(b_1) & \text{Cov}(b_1, b_0) \\ \text{Cov}(b_1, b_0) & \text{Var}(b_0) \end{bmatrix} \right)$$

Consider the function : $q = b_0 - b_1 \Rightarrow \text{plim } q = 0$

By the [delta method](#)

$$\sqrt{N}(q - 0) \xrightarrow{d} \mathcal{N} \left(0, [1 \quad -1] \begin{bmatrix} \text{Var}(b_1) & \text{Cov}(b_1, b_0) \\ \text{Cov}(b_1, b_0) & \text{Var}(b_0) \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$$

$$\text{Var}(q) = \text{Var}(b_1) + \text{Var}(b_0) - 2 \text{Cov}(b_1, b_0)$$

Using the commonly used result, showed by Hausman, that the covariance of an efficient estimator with its difference from an inefficient estimator is zero yields

$$\text{Var}(q) = \text{Var}(b_0) - \text{Var}(b_1)$$

The chi-squared test is based on the Wald criterion

$$H = \chi^2[K - 1] = (b_1 - b_0)' (\text{Var}(b_0) - \text{Var}(b_1))^{\dagger} (b_1 - b_0),$$

where † denotes the [Moore–Penrose pseudoinverse](#)

Panel data [\[edit \]](#)

The Hausman test can be used to differentiate between [fixed effects model](#) and [random effects model](#) in [panel analysis](#). In this case, Random effects (RE) is preferred under the null hypothesis due to higher efficiency, while under the alternative Fixed effects (FE) is at least as consistent and thus preferred.

	H_0 is true	H_1 is true
b_1 (RE estimator)	Consistent Efficient	Inconsistent
b_0 (FE estimator)	Consistent Inefficient	Consistent

Exercício #7

```
import excel "C:\ ... \data_hours_gdp.xls", sheet("data_hours_gdp") firstrow clear
xtset i t, yearly
xtreg hours gdp, fe
estimates store fixo
xtreg hours gdp, re
hausman fixo ., sigmamore
```

---- Coefficients ----				
	(b)	(B)	(b-B)	$\sqrt{\text{diag}(V_b - V_B)}$
	fixo	.	Difference	S.E.
gdp	-.0104265	-.0104356	9.14e-06	.0000118

b = consistent under H_0 and H_a ; obtained from xtreg
B = inconsistent under H_a , efficient under H_0 ; obtained from xtreg

Test: H_0 : difference in coefficients not systematic

```
chi2(1) = (b-B)' [(V_b-V_B)^(-1)] (b-B)
        = 0.60
Prob>chi2 = 0.4370
```

Simultaneidade

The essential form of a model for equilibrium in a market is

$$Q_{Demand} = \alpha_1 + \alpha_2 Price + \alpha_3 Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand},$$

$$Q_{Supply} = \beta_1 + \beta_2 Price + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply},$$

$$Q_{Equilibrium} = Q_{Demand} = Q_{Supply},$$

where $\mathbf{d}$ and $\mathbf{s}$ are other variables that influence the equilibrium through their impact on the demand and supply curves, respectively. This model differs from those suggested thus far because the implication of the third equation is that *Price* is not exogenous in the equation system. The equations of this model fit more appropriately in the instrumental variables framework developed in Chapter 8 than in the regression models developed in Chapters 1 to 7. The multiple equations framework developed in this chapter provides additional results for estimating “simultaneous equations models” such as this.

All the examples suggested in the chapter introduction have a common multiple equation structure, which we may write as

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\ \mathbf{y}_2 &= \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\ &\dots \\ \mathbf{y}_M &= \mathbf{X}_M \boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M. \end{aligned} \tag{10-1}$$

There are M equations and T observations in the sample of data used to estimate them.¹ The second and third examples embody different types of constraints across equations and different structures of the disturbances. A basic set of principles will apply to them all, however.² **The seemingly unrelated regressions (SUR)** model in (10-1) is

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, M. \tag{10-2}$$

Define the $MT \times 1$ vector of disturbances,

$$\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_M]'$$

We assume strict exogeneity of $\mathbf{X}_i$,

$$E[\boldsymbol{\varepsilon} | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \mathbf{0},$$

and homoscedasticity

$$E[\boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}'_m | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{mm} \mathbf{I}_T.$$

Example 10.5 Identification

Consider a market in which q is quantity of Q , p is price, and z is the price of Z , a related good. We assume that z enters both the supply and demand equations. For example, Z might be a crop that is purchased by consumers and that will be grown by farmers instead of Q if its price rises enough relative to p . Thus, we would expect $\alpha_2 > 0$ and $\beta_2 < 0$. So,

$$q_d = \alpha_0 + \alpha_1 p + \alpha_2 Z + \varepsilon_d \quad (\text{demand}),$$

$$q_s = \beta_0 + \beta_1 p + \beta_2 Z + \varepsilon_s \quad (\text{supply}),$$

$$q_d = q_s = q \quad (\text{equilibrium}).$$

The reduced form is

$$q = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2 - \alpha_2 \beta_1}{\alpha_1 - \beta_1} Z + \frac{\alpha_1 \varepsilon_s - \alpha_2 \varepsilon_d}{\alpha_1 - \beta_1} = \pi_{11} + \pi_{21} Z + v_q,$$

$$p = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 - \alpha_2}{\alpha_1 - \beta_1} Z + \frac{\varepsilon_s - \varepsilon_d}{\alpha_1 - \beta_1} = \pi_{12} + \pi_{22} Z + v_p.$$

With only four reduced-form coefficients and six structural parameters, it is obvious that there will not be a complete solution for all six structural parameters in terms of the four reduced parameters. Suppose, though, that it is known that $\beta_2 = 0$ (farmers do not substitute the alternative crop for this one). Then the solution for β_1 is π_{21} / π_{22} . After a bit of manipulation, we also obtain $\beta_0 = \pi_{11} - \pi_{12} \pi_{21} / \pi_{22}$. The restriction identifies the supply parameters, but this step is as far as we can go.

The structural form of the model is³⁸

There are M equations and M endogenous variables, denoted $y_1, \dots, y_M$. There are K exogenous variables, $x_1, \dots, x_K$, that may include predetermined values of $y_1, \dots, y_M$ as well. The first element of $\mathbf{x}_t$ will usually be the constant, 1. Finally, $\varepsilon_{t1}, \dots, \varepsilon_{tM}$ are the **structural disturbances**. The subscript t will be used to index observations, $t = 1, \dots, T$.

In matrix terms, the system may be written

$$[y_1 \ y_2 \ \cdots \ y_M]_t \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ & & \ddots & \\ \gamma_{M1} & \gamma_{M2} & \cdots & \gamma_{MM} \end{bmatrix} \\ + [x_1 \ x_2 \ \cdots \ x_K]_t \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ & & \ddots & \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KM} \end{bmatrix} = [\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_M]_t,$$

$$\begin{cases} y_1 = \gamma_{12}y_2 + \delta_{11} + \delta_{12}x_2 + \delta_{13}x_3 + u_1 \\ y_2 = \gamma_{21}y_1 + \delta_{21} + \delta_{22}x_2 + \delta_{23}x_3 + u_2 \end{cases}$$

$$\begin{aligned} \Gamma_{2 \times 2} Y_{2 \times 1} &= \Delta_{2 \times 3} X_{3 \times 2} + U_{2 \times 1} \\ \text{if } \gamma_{12}\gamma_{21} &\neq 1 \\ Y &= \Gamma^{-1} \Delta X + \Gamma^{-1} U \\ Y &= \Pi_{2 \times 3} X + V \end{aligned}$$

Identification condition: $\Gamma\Pi = \Delta$.

Suppose $\delta_{13} = \delta_{22} = 0$.

$$\begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix} \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} = \begin{bmatrix} \delta_{11} & \delta_{12} & 0 \\ \delta_{21} & 0 & \delta_{23} \end{bmatrix}$$

Then, $\pi_{13} - \gamma_{12}\pi_{23} = 0$ and $-\gamma_{21}\pi_{12} + \pi_{22} = 0$. Or, $\gamma_{12} = \frac{\pi_{13}}{\pi_{23}}$ and $\gamma_{21} = \frac{\pi_{22}}{\pi_{12}}$.

O recado aqui é que você tem 3 caminhos:

- 1 Se for estimar uma equação do sistema por vez, ache instrumentos para a variável endógena e aplique 2SLS
- 2 Se for estimar todo o sistema por SUR, se preocupe com restrições de identificação e com a matriz de variância-covariância
- 3 Se quiser estimar todo o sistema de uma vez, sem ter instrumentos evidentes, vá para a econometria de cópulas

Resposta binária

- Lembre que nosso interesse está na análise da estrutura $y_i = E(y_i | X_i) + u_i$.
- Se a variável de estudo é binária, sem perda de generalidade, é melhor definir $y_i \in \{0, 1\}$, porque assim $E(y_i | X_i) = 0 \times \Pr(y_i = 0 | X_i) + 1 \times \Pr(y_i = 1 | X_i) = \Pr(y_i = 1 | X_i) \in [0, 1]$ (i.e., você só precisa se preocupar em modelar uma probabilidade).
- Quais os problemas de modelar essa probabilidade como uma projeção linear? (i.e., $\Pr(y_i = 1 | X_i) = X_i\beta$) R.: os mesmos de se modelar uma variável dependente não binária.
- Mas tem lógica considerar essa esperança projeção linear?
- A ideia da variável latente: $y_i^* = X_i\beta + u_i^*$, em que $y_i = \mathbb{1}(y_i^* > 0)$.
- $\Pr(y_i = 1 | X_i) = \Pr(y_i^* > 0 | X_i)$.
- $\Pr(X_i\beta + u_i^* > 0 | X_i) = \Pr(u_i^* > -X_i\beta | X_i) = 1 - \Pr(u_i^* < -X_i\beta | X_i) = 1 - F(-X_i\beta | X_i)$, em que F é a c.d.f. de u_i^* .
- Se F for simétrica (qual o problema com isso?), $\Pr(y_i = 1 | X_i) = F(X_i\beta) \Rightarrow E(y_i | X_i) = F(X_i\beta) \in [0, 1]$
- Se F for Normal/Logística padrão, qual o problema com isso?
- $$\frac{\partial E(y_i | X_i)}{\partial x_{ki}} = \frac{\partial F(X_i\beta)}{\partial x_{ki}} = \frac{\partial F(X_i\beta)}{\partial (X_i\beta)} \times \frac{\partial (X_i\beta)}{\partial x_{ki}} = f(X_i\beta) \times \beta_k$$

$$y_i \in \{0, 1\}$$

$$\Pr(y_i = 1) = \Lambda_i$$

$$\Lambda_i = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

$$\ln L = \sum_{i=1}^n [y_i \ln \Lambda_i + (1 - y_i) \ln(1 - \Lambda_i)]$$

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i=1}^n x_{ki}(y_i - \Lambda_i)$$

$$\frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_{k'}} = - \sum_{i=1}^n x_{ki} \Lambda_i (1 - \Lambda_i) x_{k'i}$$

Exercício #8

```
clear
clc
rand('seed',12345)
n = 30
x = rand(n,1,'uniform')
u = rand(n,1,'normal')
b1 = 0.5
b2 = 0.5
y = zeros(n,1)
for j = 1:n
    if b1+b2*x(j,1)+u(j,1) > 0 then
        y(j,1) = 1
    end
end
XX = [ones(n,1) x]
beta_ols = ((XX'*XX)^-1)*(XX'*y)

beta_logit = beta_ols
for k=1:20
    a = beta_logit(1,1)
    b = beta_logit(2,1)
```

```
p = exp(a+b*x)./(1+exp(a+b*x))
f1 = sum(y-p)
f2 = sum((y-p).*x)
grad = [f1;f2]
g11 = -sum(p.*(1-p))
g12 = -sum(p.*(1-p).*x)
g21 = g12
g22 = -sum(p.*(1-p).*(x.^2))
hessiano = [g11 g12; g21 g22]
delta = (hessiano^-1)*grad
    if abs(delta)<10^-5
        break
    end
    beta_logit = beta_logit - delta
end

lnL = sum(log(y.*p+(1-y).*(1-p)))

dp = diag((-hessiano).^(.5))
z = beta_logit./dp
```

```
logit var2 var1, technique(nr)
```

```
Iteration 0:   log likelihood = -13.516836
Iteration 1:   log likelihood = -12.593299
Iteration 2:   log likelihood =  -12.5464
Iteration 3:   log likelihood = -12.546231
Iteration 4:   log likelihood = -12.546231
```

Logistic regression	Number of obs	=	30
	LR chi2(1)	=	1.94
	Prob > chi2	=	0.1635
Log likelihood = -12.546231	Pseudo R2	=	0.0718

var2	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
var1	2.352477	1.783962	1.32	0.187	-1.144023	5.848978
_cons	.5918586	.8265458	0.72	0.474	-1.028142	2.211859

Exercício #9

```
clear
cls

set obs 30
*set obs 100
*set obs 100000
set seed 12345

gen x = runiform()
gen u = rnormal()
gen y = cond(0.5+0.5*x+u>0,1,0)

logit y x
estat classification
margins, dydx(*)

probit y x
estat classification
margins, dydx(*)
```

Exercício #10

```
clear
set obs 100
set seed 12345
gen x = invnormal(uniform())
gen y = cond(0.5+0.5*x+invnormal(uniform())>0,1,0)
probit y x

program drop _all
program myprobit
    version 16.0
    args lnf xb
    quietly{
        replace `lnf' = $ML_y1 * ln(normal(`xb')) + (1-$ML_y1) * ln(1-normal(`xb'))
    }
end

ml model lf myprobit (y = x)
ml check

ml maximize
```

No Probit/Logit, onde está o problema da endogeneidade?

$$y_i^* = X_i\beta + u_i \text{ tal que } \begin{cases} y_i = 1 \text{ se } y_i^* > 0 \\ y_i = 0 \text{ se } y_i^* \leq 0 \end{cases}$$

$$E(y_i | X_i) = \Pr(y_i = 1 | X_i) = \underbrace{\Pr(u_i < X_i\beta | X_i)}_{\text{sob a hip. } E(u_i | X_i)=0} = F(X_i\beta)$$

Mas se $y_i^* = \gamma w_i + X_i\beta + u_i$ e $E(u_i | w_i) \neq 0$?

Como no modelo linear, é possível usar instrumentos ou painel.

$$\begin{cases} y_i^* = \gamma w_i + X_i \beta + u_i \\ w_i = Z_i \alpha + v_i \\ \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \sigma_v \\ \rho \sigma_v & \sigma_v^2 \end{bmatrix} \right) \end{cases}$$

$$\ln L = \sum_{i=1}^n \ln \Phi \left[(2y_i - 1) \left(\frac{\mathbf{x}_i' \beta + \gamma w_i + (\rho / \sigma_u)(w_i - \mathbf{z}_i' \alpha)}{\sqrt{1 - \rho^2}} \right) \right] + \ln \left[\frac{1}{\sigma_u} \phi \left(\frac{w_i - \mathbf{z}_i' \alpha}{\sigma_u} \right) \right]$$

<https://www.stata.com/manuals/rivprobit.pdf>

Arminger et al. (2013) página 413 ... $\mathbf{y}_i = \{y_{it}\}_{t=1}^T$

$$\Pr(y_{it} = 1) = \Lambda_{it} = \frac{\exp(c_i + X_{it}\beta)}{1 + \exp(c_i + X_{it}\beta)}$$

$$\Pr(\mathbf{y}_i) = \prod_{t=1}^T \Lambda_{it}^{y_{it}} (1 - \Lambda_{it})^{(1-y_{it})} = \exp \left(\sum_{t=1}^T y_{it} \ln \left(\frac{\Lambda_{it}}{1 - \Lambda_{it}} \right) + \sum_{t=1}^T \ln(1 - \Lambda_{it}) \right)$$

$$\Pr(\mathbf{y}_i) = \exp \left(\sum_{t=1}^T y_{it} (c_i + X_{it}\beta) + \sum_{t=1}^T \ln \left(\frac{1}{1 + \exp(c_i + X_{it}\beta)} \right) \right)$$

$$\Pr(\mathbf{y}_i) = \exp \left([c_i \sum_{t=1}^T y_{it}] + [\sum_{t=1}^T y_{it} X_{it}\beta] + \ln \left(\prod_{t=1}^T \frac{1}{1 + \exp(c_i + X_{it}\beta)} \right) \right)$$

$$\Pr(\mathbf{y}_i) = \frac{\exp \left(c_i [\sum_{t=1}^T y_{it}] + [\sum_{t=1}^T y_{it} X_{it}] \beta \right)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it}\beta)]}$$

Agora defina $\sum_{t=1}^T y_{it} = \tau_i = 0, 1, \dots, T \dots$ e, para simplificar, considere $T = 2$

$$\Pr(\tau_i = 0) = \frac{\exp(c_i[0] + [0]\beta)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it}\beta)]} = \frac{1}{\prod_{t=1}^T [1 + \exp(c_i + X_{it}\beta)]}$$

$$\Pr(\tau_i = 1) = \frac{\exp(c_i[1] + X_{i1}\beta) + \exp(c_i[1] + X_{i2}\beta)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it}\beta)]}$$

$$\Pr(\tau_i = 2) = \frac{\exp(2c_i + (X_{i1} + X_{i2})\beta)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it}\beta)]}$$

Agora note que...

$$[1 + \exp(c_i + X_{i1}\beta)] \times [1 + \exp(c_i + X_{i2}\beta)] = 1 + \exp(c_i + X_{i1}\beta) + \exp(c_i + X_{i2}\beta) + \exp(2c_i + (X_{i1} + X_{i2})\beta)$$

$$\Pr(\tau_i) = \frac{\sum_{\text{combinações que chegam em } \tau_i} \exp \left(c_i \tau_i + [\sum_{t=1}^T y_{it} X_{it}] \beta \right)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it} \beta)]}$$

$$\Pr(\mathbf{y}_i \mid \tau_i) = \frac{\Pr(\mathbf{y}_i)}{\Pr(\tau_i)} = \frac{\frac{\exp \left(c_i [\sum_{t=1}^T y_{it}] + [\sum_{t=1}^T y_{it} X_{it}] \beta \right)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it} \beta)]}}{\frac{\sum_{\text{combinações que chegam em } \tau_i} \exp \left(c_i \tau_i + [\sum_{t=1}^T y_{it} X_{it}] \beta \right)}{\prod_{t=1}^T [1 + \exp(c_i + X_{it} \beta)]}}$$

$$\Pr(\mathbf{y}_i \mid \tau_i) = \frac{\exp \left([\sum_{t=1}^T y_{it} X_{it}] \beta \right)}{\sum_{\text{combinações que chegam em } \tau_i} \exp \left([\sum_{t=1}^T y_{it} X_{it}] \beta \right)}$$

```
clear
cls
use https://www.stata-press.com/data/r17/lowbirth2
describe
list low lwt smoke pairid in 1/10
logit low lwt smoke ptd ht ui i.race
estimates store logit
clogit low lwt smoke ptd ht ui i.race, group(pairid)
estimates store clogit
esttab logit clogit
logit low lwt smoke ptd ht ui i.race, or
clogit low lwt smoke ptd ht ui i.race, group(pairid) or
```

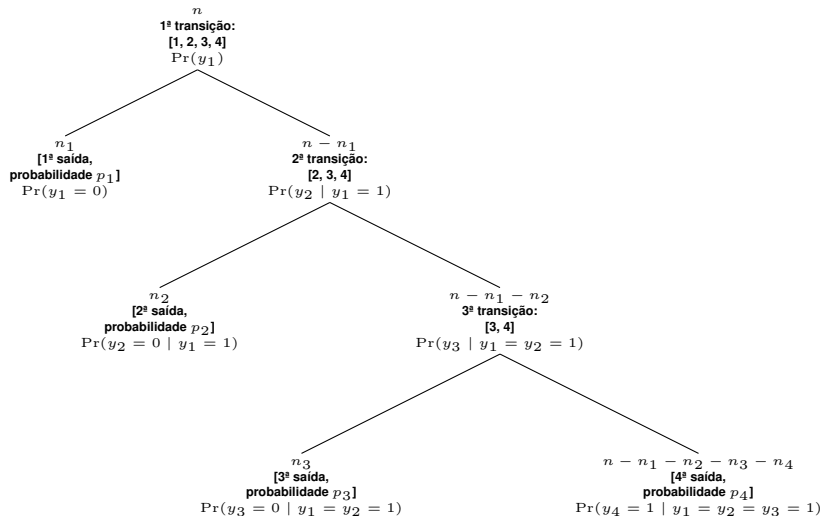
Ver também os efeitos marginais <https://www.stata.com/manuals13/rlogitpostestimation.pdf>

$$\begin{aligned}y_1^* &= \mathbf{x}_1' \boldsymbol{\beta}_1 + \varepsilon_1, & y_1 &= 1 \text{ if } y_1^* > 0, 0 \text{ otherwise,} \\y_2^* &= \mathbf{x}_2' \boldsymbol{\beta}_2 + \varepsilon_2, & y_2 &= 1 \text{ if } y_2^* > 0, 0 \text{ otherwise,} \\& & \left(\begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \end{array} \middle| \mathbf{x}_1, \mathbf{x}_2 \right) &\sim N \left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right].\end{aligned}$$

<https://www.stata.com/manuals/rbiprobit.pdf>

Fazer o Exercício #12 do paper da PPP

Triprobit: <https://ideas.repec.org/c/boc/bocode/s424302.html>



$$p_1 = \Pr(y_1 = 0)$$

$$p_2 = \Pr(y_2 = 0, y_1 = 1)$$

$$= \Pr(y_1 = 1) \times \Pr(y_2 = 0 \mid y_1 = 1)$$

$$p_3 = \Pr(y_1 = 1) \times \Pr(y_2 = 1 \mid y_1 = 1) \times \Pr(y_3 = 0 \mid y_1 = y_2 = 1)$$

$$p_4 = \Pr(y_1 = 1) \times \Pr(y_2 = 1 \mid y_1 = 1) \times \Pr(y_3 = 1 \mid y_1 = y_2 = 1) \times \Pr(y_4 = 0 \mid y_1 = y_2 = y_3 = 1)$$

$$\Pr(y_{it} \mid \cdot) = \Lambda(X_{it}\beta)^{y_{it}}(1 - \Lambda(X_{it}\beta))^{1-y_{it}}$$

$$p_{i1} = 1 - \Lambda(X_{i1}\beta)$$

$$p_{i2} = \Lambda(X_{i1}\beta)(1 - \Lambda(X_{i2}\beta))$$

$$p_{i3} = \Lambda(X_{i1}\beta)\Lambda(X_{i2}\beta)(1 - \Lambda(X_{i3}\beta))$$

$$p_{i4} = \Lambda(X_{i1}\beta)\Lambda(X_{i2}\beta)\Lambda(X_{i3}\beta)$$

$$L(\beta \mid \{(X_{it}, y_{it})\}_{i=1, t=1}^{n, T}) = \prod_i (1 - p_{iT})^{1-I_{iT}} \prod_t p_{it}^{I_{it}}$$

Exercício #13

```
set lstretch off
*ssc install seqlogit

use "C:\Users\pett7\Google Drive\disciplina_microeconometria\dados\seqlogit_data.dta", clear

recode degree 4=3

label define degree 0 "lt high school" ///
                  1 "high school"      ///
                  2 "junior college"   ///
                  3 "college", modify

label value degree degree

gen byte res16 = 1*country + 2*town + 3*suburb + 4*city

label define res16 1 "country" ///
                  2 "town"      ///
                  3 "suburb"    ///
                  4 "city"

label value res16 res16

replace coh = coh - 5

label var coh "(year of birth-1950)/10"

seqlogit degree south i.res16 c.sibs##c.sibs ///
        c.coh##c.coh if black == 0 , or      ///
        tree(0 : 1 2 3 , 1 : 2 3 , 2 : 3 )   ///
        ofinterest(paeduc)                  ///
        over(c.coh##c.coh)                  ///
        levels(0=9, 1=12, 2=14, 3=16)
```

Resposta de contagem

- Lembre que nosso interesse está na análise de $y_i = E(y_i | X_i) + u_i$.
- Se a variável de estudo é $y_i \in \{0, 1, 2, 3, \dots\}$, com densidade em valores baixos, uma opção é modelar $E(y_i | X_i) = \exp(X_i\beta)$.
- Uma distribuição de probabilidade que gera esse tipo de esperança é a Poisson, de média $\lambda_i = \exp(X_i\beta)$.
- $E(y_i | X_i) = \sum_{j=0}^{\infty} j \times \Pr(y_i = j | X_i)$, então a questão é modelar $\Pr(y_i = j | X_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$
- $L = \prod_{i=1}^n \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \Rightarrow \ln L = \sum_{i=1}^n (-\lambda_i + y_i \times X_i\beta - \ln y_i!)$
- $J = [\sum_{i=1}^n (y_i - \lambda_i)x_{ik}]$
- $H = [\sum_{i=1}^n (y_i - \lambda_i)x_{ik}x_{ik'}]$
- $\frac{\partial E(y_i | X_i)}{\partial x_{ki}} = \exp(X_i\beta)\beta_k$
- A questão da variância igual a média, e os modelos Binomial Negativo e Poisson Generalizado
- Variável *offset*

Exercício #14

```
n = 100
rand('seed',12345)

x = grand(n,1,"gam",2,5)
mu = exp(1-x)
y = zeros(n,1)
for i = 1:n
    y(i) = grand(1,1,"poi", mu(i))
end
```

```
b = [1;1]
for i = 1:50
    lambda = exp(b(1)+b(2)*x)
    g1 = sum(y-lambda)
    g2 = sum((y-lambda).*x)
    G = [g1; g2]
    H11 = -sum(lambda)
    H12 = -sum(lambda.*x)
    H21 = H12
    H22 = -sum(lambda.*(x.^2))
    H = [H11 H12; H21 H22]
```

```
b = b-inv(H)*G
if abs(inv(H)*G) < 10^-5
end
end
```

```
z = b./(diag(-inv(H)).^.5)
```

```
b0 = 1
for i = 1:50
    g = sum(y-exp(b0))
    b0 = b0-g/(-n*exp(b0))
    if abs(g/(-n*exp(b0))) < 10^-5
end
end
```

```
lnL = sum(-lambda+y.*log(lambda)
          -log(factorial(y)))
lnL0 = sum(-exp(b0)+y.*b0-log(factorial(y)))

pseudoR2 = 1-lnL/lnL0
```

ZIP e o caso dos homicídios

Pode existir duas naturezas de zero na observação dos homicídios (independente de terem ou não uma motivação econômica). Na primeira, ocorre $H = 0$ porque o local é “pacato”, ou seja, no período observado não existiram situações suficientemente violentas capazes de gerar homicídios. Ou, na segunda, existiram situações suficientemente violentas capazes de gerar homicídios, mas por algum outro motivo observou-se $H = 0$.

$$\begin{aligned}\Pr(H = y_i) &= \begin{cases} \pi_i + (1 - \pi_i) \exp(-\lambda_i) & \text{se } H = 0 \\ (1 - \pi_i) \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{y_i!} & \text{se } H > 0 \end{cases} \\ \pi_i &= \frac{\exp(Z_i \gamma)}{1 + \exp(Z_i \gamma)} \\ \lambda_i &= \exp(\ln P_i + X_i \theta + \beta \ln A_i) \\ E(H|X_i, P_i, A_i; \beta, \theta) &= (1 - \pi_i) \lambda_i \\ \text{Var}(H|X_i, P_i, A_i; \alpha, \beta, \theta) &= (1 - \pi_i) \lambda_i (1 + \pi_i \lambda_i)\end{aligned}$$

No modelo *zero inflated* a probabilidade de $H = 0$ em decorrência do município ser “pacato” é dada por π_i . Ou seja, uma forma Logit condicionada a uma vetor Z_i de covariadas. Note também que, sendo $E(H|\cdot) = (1 - \pi_i) \lambda_i$, se $\pi_i \rightarrow 0$ o modelo abarca a “equidispersão” se aproximando forma canônica do Processo de Poisson.

O Teste de Vuong, o AIC e o SIC... e usar a base de dados dos homicídios

$$Income = \mathbf{z}_i' \boldsymbol{\gamma} + u_i,$$

$$\text{Prob}(DocVis_i = j | \mathbf{x}_i, Income_i) = \exp(-\lambda_i) \lambda_i^j / j!, \lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta} + \delta Income_i).$$

Endogeneity as we have analyzed it, for example, in Chapter 8 and Sections 17.3.5 and 17.5.5, arises through correlation between the endogenous variable and the unobserved omitted factors in the main equation. **But, the Poisson model does not contain any unobservables.** This is a major shortcoming of the specification as a “regression” model; all of the regression variation of the dependent variable arises through variation of the observables. There is no accommodation for unobserved heterogeneity or omitted factors. This is the compelling motivation for the negative binomial model or, in RWM’s case, the Poisson-normal mixture model. [See Terza (2010, pp. 555–556) for discussion of this issue.] If the model is reformulated to accommodate heterogeneity, as in

$$\lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta} + \delta Income_i + \varepsilon_i),$$

then $Income_i$ will be endogenous if u_i and ε_i are correlated.

A bivariate normal model for (u_i, ε_i) with zero means, variances σ_u^2 and σ_ε^2 and correlation ρ provides a convenient (and the usual) platform to operationalize this idea. By projecting ε_i on u_i , we have

$$\varepsilon_i = (\rho \sigma_\varepsilon / \sigma_u) u_i + v_i,$$

$$\begin{aligned}\varepsilon_i &= \rho\sigma_\varepsilon(\text{Income}_i - \mathbf{z}_i'\boldsymbol{\gamma})/\sigma_u + v_i, \\ &= \tau[(\text{Income}_i - \mathbf{z}_i'\boldsymbol{\gamma})/\sigma_u] + \theta w_i.\end{aligned}$$

where w_i will be normally distributed with mean zero and variance one while $\tau = \rho\sigma_\varepsilon$ and $\theta^2 = \sigma_\varepsilon^2(1 - \rho^2)$. Then, combining terms,

$$\varepsilon_i = \tau u_i^* + \theta w_i.$$

With this parameterization, the conditional mean function in the Poisson regression model is

$$\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \delta \text{Income}_i + \tau u_i^* + \theta w_i).$$

The parameters to be estimated are $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, δ , σ_ε , σ_u , and ρ . There are two ways to proceed. A two-step method can be based on the fact that $\boldsymbol{\gamma}$ and σ_u can consistently be estimated by linear regression of *Income* on $\mathbf{z}$. After this first step, we can compute values of u_i^* and formulate the Poisson regression model in terms of

$$\hat{\lambda}_i(w_i) = \exp[\mathbf{x}_i'\boldsymbol{\beta} + \delta \text{Income}_i + \tau \hat{u}_i + \theta w_i].$$

The log-likelihood to be maximized at the second step is

$$\ln L(\boldsymbol{\beta}, \delta, \tau, \theta | \mathbf{w}) = \sum_{i=1}^n -\hat{\lambda}_i(w_i) + y_i \ln \hat{\lambda}_i(w_i) - \ln y_i!.$$

Consider first a fixed effects approach. The Poisson distribution is assumed to have conditional mean

$$\log \lambda_{it} = \beta' \mathbf{x}_{it} + \alpha_i, \quad (18-24)$$

where now, $\mathbf{x}_{it}$ has been redefined to exclude the constant term. The approach used in the linear model of transforming y_{it} to group mean deviations does not remove the heterogeneity, nor does it leave a Poisson distribution for the transformed variable. However, the Poisson model with fixed effects can be fit using the methods described for the probit model in Section 17.4.3. The extension to the Poisson model requires only the minor modifications, $g_{it} = (y_{it} - \lambda_{it})$ and $h_{it} = -\lambda_{it}$. Everything else in that derivation applies with only a simple change in the notation. The first-order conditions for maximizing the log-likelihood function for the Poisson model will include

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{t=1}^{T_i} (y_{it} - e^{\alpha_i} \mu_{it}) = 0 \quad \text{where } \mu_{it} = e^{\mathbf{x}_{it}' \beta}.$$

This implies an explicit solution for α_i in terms of β in this model,

$$\hat{\alpha}_i = \ln \left(\frac{(1/T_i) \sum_{t=1}^{T_i} y_{it}}{(1/T_i) \sum_{t=1}^{T_i} \hat{\mu}_{it}} \right) = \ln \left(\frac{\bar{y}_i}{\bar{\hat{\mu}}_i} \right). \quad (18-25)$$

The Statistician (2003)
52, Part 3, pp. 381–393

Analysis of sports data by using bivariate Poisson models

Dimitris Karlis

Athens University of Economics and Business, Greece

and Ioannis Ntzoufras

University of the Aegean, Chios, Greece

[Received November 2001. Final revision April 2003]

Summary. Models based on the bivariate Poisson distribution are used for modelling sports data. Independent Poisson distributions are usually adopted to model the number of goals of two competing teams. We replace the independence assumption by considering a bivariate Poisson model and its extensions. The models proposed allow for correlation between the two scores, which is a plausible assumption in sports with two opposing teams competing against each other. The effect of introducing even slight correlation is discussed. Using just a bivariate Poisson distribution can improve model fit and prediction of the number of draws in football games. The model is extended by considering an inflation factor for diagonal terms in the bivariate joint distribution. This inflation improves in precision the estimation of draws and, at the same time, allows for overdispersed, relative to the simple Poisson distribution, marginal distributions. The properties of the models proposed as well as interpretation and estimation procedures are provided. An illustration of the models is presented by using data sets from football and water-polo.

Keywords: Bivariate Poisson regression; Difference of Poisson variates; Inflated distributions; Soccer

Resposta multinomial

- A variável de resposta é, por exemplo, “insatisfeito”, “satisfeito” e “muito satisfeito”.
- $y_i \in \{0, 1, 2\} \Rightarrow E(y_i | X_i) = \Pr(y_i = 1 | X_i) + 2 \times \Pr(y_i = 2 | X_i) \in [0, 1]$.
- Quais os problemas de modelar essa probabilidade como uma projeção linear? (i.e., $E(y_i | X_i) = X_i\beta$) R.: Nesse caso, você perde a oportunidade de modelar cada $\Pr(y_i = j | X_i)$, para $j = 0, 1, 2$
- A ideia da variável latente: $y_i^* = X_i\beta + u_i^*$, em que $y_i = \begin{cases} 0 & \text{se } y_i^* \leq 0 \\ 1 & \text{se } 0 < y_i^* \leq c \\ 2 & \text{se } y_i^* > c \end{cases}$
- Se F é a c.d.f. de u_i^* , ocorre que $\begin{cases} \Pr(y_i = 0 | X_i) = F(-X_i\beta) \\ \Pr(y_i = 1 | X_i) = F(c - X_i\beta) - F(-X_i\beta) \\ \Pr(y_i = 2 | X_i) = 1 - F(c - X_i\beta) \end{cases}$
- $\frac{\partial \Pr(y_i = j | X_i)}{\partial x_{ki}} = \begin{cases} -f(-X_i\beta) \times \beta_k \\ -[f(c - X_i\beta) - f(-X_i\beta)] \times \beta_k \\ f(c - X_i\beta) \times \beta_k \end{cases}$

```
cls
clear
use https://www.stata-press.com/data/r16/fullauto
tabulate rep77

ologit rep77 foreign length mpg
predict poor fair avg good exc
margins, dydx(*)
tw (kdensity poor) (kdensity fair)
```

Como nos casos anteriores, aqui também existe uma discussão sobre endogeneidade e simultaneidade.

Ordered logistic regression

Number of obs = 66

LR chi2(3) = 23.29

Prob > chi2 = 0.0000

Pseudo R2 = 0.1295

Log likelihood = -78.250719

rep77	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
foreign	2.896807	.7906411	3.66	0.000	1.347179	4.446435
length	.0828275	.02272	3.65	0.000	.0382972	.1273579
mpg	.2307677	.0704548	3.28	0.001	.0926788	.3688566
/cut1	17.92748	5.551191			7.047344	28.80761
/cut2	19.86506	5.59648			8.896161	30.83396
/cut3	22.10331	5.708936			10.914	33.29262
/cut4	24.69213	5.890754			13.14647	36.2378

Ex.: a variável dependente é variante de Covid com a qual a pessoa foi contaminada... $y_i = \{1, 2, 3\}$

Nesse caso $\mathbb{1}(y_i = j) = E(\mathbb{1}(y_i = j) \mid X_i) + u_i$, então $\Rightarrow E(\mathbb{1}(y_i = j) \mid X_i) = \Pr(y_i = j \mid X_i)$

$$\begin{cases} y_{1i}^* = X_i\beta_1 + u_{1i}^* \\ y_{2i}^* = X_i\beta_2 + u_{2i}^* \\ y_{3i}^* = X_i\beta_3 + u_{3i}^* \end{cases}$$

$$\begin{cases} \Pr(y_i = 1 \mid X_i) = \Pr(y_{1i}^* > y_{2i}^*; y_{1i}^* > y_{3i}^*) \\ \Pr(y_i = 2 \mid X_i) = \Pr(y_{2i}^* > y_{1i}^*; y_{2i}^* > y_{3i}^*) \\ \Pr(y_i = 3 \mid X_i) = \Pr(y_{3i}^* > y_{1i}^*; y_{3i}^* > y_{2i}^*) \end{cases} \Rightarrow \begin{cases} \Pr(y_i = 1 \mid X_i) = \Pr(y_{1i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \epsilon_i) \\ \Pr(y_i = 2 \mid X_i) = \Pr(y_{2i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \epsilon_i) \\ \Pr(y_i = 3 \mid X_i) = \Pr(y_{3i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \epsilon_i) \end{cases}$$

https://en.wikipedia.org/wiki/Extreme_value_theory

$$\begin{cases} \Pr(y_i = 1 \mid X_i) = \Pr(y_{1i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \varepsilon_i) = \frac{\exp(X_i\beta_1)}{\sum_{j=1}^3 \exp(X_i\beta_j)} = \frac{\exp(X_i\beta_1)}{1 + \sum_{j=1}^2 \exp(X_i\beta_j)} \\ \Pr(y_i = 2 \mid X_i) = \Pr(y_{2i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \varepsilon_i) = \frac{\exp(X_i\beta_2)}{\sum_{j=1}^3 \exp(X_i\beta_j)} = \frac{\exp(X_i\beta_2)}{1 + \sum_{j=1}^2 \exp(X_i\beta_j)} \\ \Pr(y_i = 3 \mid X_i) = \Pr(y_{3i}^* = \max\{y_{1i}^*, y_{2i}^*, y_{3i}^*\} = \varepsilon_i) = \frac{\exp(X_i\beta_3)}{\sum_{j=1}^3 \exp(X_i\beta_j)} = \frac{1}{1 + \sum_{j=1}^2 \exp(X_i\beta_j)} \end{cases}$$

<https://www.stata.com/manuals/rmlogit.pdf>

$$\begin{cases} \Pr(y_i = 1 \mid X_i) = \Pr((\beta_1 - \beta_{j'})x_i \geq u_{j',i}^* - u_{1i}^* ; \forall j' = 1, 2, 3) \\ \Pr(y_i = 2 \mid X_i) = \Pr((\beta_2 - \beta_{j'})x_i \geq u_{j',i}^* - u_{2i}^* ; \forall j' = 1, 2, 3) \\ \Pr(y_i = 3 \mid X_i) = \Pr((\beta_3 - \beta_{j'})x_i \geq u_{j',i}^* - u_{3i}^* ; \forall j' = 1, 2, 3) \end{cases}$$

Unordered choice models can be motivated by a random utility model. For the i th consumer faced with J choices, suppose that the utility of choice j is

$$U_{ij} = \mathbf{z}_{ij}'\boldsymbol{\theta} + \varepsilon_{ij}.$$

If the consumer makes choice j in particular, then we assume that U_{ij} is the maximum among the J utilities. Hence, the statistical model is driven by the probability that choice j is made, which is

$$\text{Prob}(U_{ij} > U_{ik}) \quad \text{for all other } k \neq j.$$

identically distributed with Gumbel (type 1 extreme value) distributions,

$$F(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij})), \quad (18-1)$$

$$\text{Prob}(Y_i = j) = \frac{\exp(\mathbf{z}'_{ij}\boldsymbol{\theta})}{\sum_{j=1}^J \exp(\mathbf{z}'_{ij}\boldsymbol{\theta})},$$

Utility depends on $\mathbf{z}_{ij}$, which includes aspects specific to the individual as well as to the choices. It is useful to distinguish them. Let $\mathbf{z}_{ij} = [\mathbf{x}_{ij}, \mathbf{w}_i]$ and partition $\boldsymbol{\theta}$ conformably into $[\boldsymbol{\beta}', \boldsymbol{\alpha}']'$. Then $\mathbf{x}_{ij}$ varies across the choices and possibly across the individuals as well. The components of $\mathbf{x}_{ij}$ are typically called the **attributes** of the choices. But, $\mathbf{w}_i$ contains the **characteristics** of the individual and is, therefore, the same for all choices. If we incorporate this fact in the model, then (18-2) becomes

$$\text{Prob}(Y_i = j) = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_i\boldsymbol{\alpha})}{\sum_{j=1}^J \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_i\boldsymbol{\alpha})} = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) \exp(\mathbf{w}'_i\boldsymbol{\alpha})}{\left[\sum_{j=1}^J \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) \right] \exp(\mathbf{w}'_i\boldsymbol{\alpha})}. \quad (18-3)$$

$$\Pr(Y_i = j) = \Pr(U_{ij} > U_{ic}, \quad \forall \quad j \neq c)$$

$$\Pr(Y_i = j) = \Pr(V_{ij} + \varepsilon_{ij} > V_{ic} + \varepsilon_{ic}, \quad \forall j \neq c) \Rightarrow \Pr(\varepsilon_{ic} - \varepsilon_{ij} < V_{ij} - V_{ic}, \quad \forall j \neq c)$$

$$\Pr(Y_i = j) = \frac{e^{V_{ij}}}{1 + \sum_{j=1}^C e^{V_{ij}}}$$

Assumindo que a opção externa, $V_{i0} = 0$...

$$s_{0t} = \frac{1}{1 + \sum_{j=1}^C e^{\beta x_{jt} + \xi_j}}$$

Com alguma manipulação algébrica...

$$\ln(s_j) - \ln(s_0) = X_j \beta + \xi_j$$

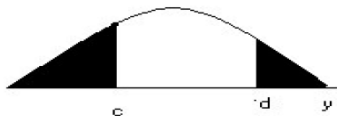
Base de dados dos carros e Exercício #16 [aqui falaremos de GMM

<https://www.jstor.org/stable/2171802>]

Censura e truncagem

- Se y não passa de zero (e.g., renda), então ela é truncada em zero
- Se y não é observada acima de $y^{\max}$, então ela é censurada em $y^{\max}$

Seja Y uma variável aleatória e y um valor particular



$$\Pr(Y < y \mid c < Y < d) = \frac{F(y) - F(c)}{F(d) - F(c)}$$

A densidade de probabilidade no ponto y da variável aleatória truncada Y é

$$\lim_{\delta \rightarrow 0} \frac{\Pr(Y < y \mid c < Y < d) - \Pr(Y < y + \delta \mid c < Y < d)}{\delta} = \frac{f(y)}{F(d) - F(c)}$$

Note também que

$$\begin{aligned} E(Y \mid c < Y < d) &= \int_c^d \frac{yf(y)}{F(d) - F(c)} dy \\ &= \frac{\int_c^d ydF(y)}{F(d) - F(c)} = \frac{E(Y)}{F(d) - F(c)} \end{aligned}$$

$y_i^* = X_i\beta + u_i$, com $u_i \sim N(0, \sigma^2)$ i.i.d.

$$y_i = \begin{cases} 0 & , \quad y_i^* \leq 0 \\ y_i^* & , \quad y_i^* > 0 \end{cases}$$

Então.

$$\begin{aligned} \Pr(y_i = 0 \mid X_i, y_i^* \leq 0) &= \\ \Pr(y_i^* \leq 0 \mid \cdot) &= \Pr\left(\frac{u_i}{\sigma} \leq -\frac{X_i\beta}{\sigma}\right) = \\ \Phi\left(-\frac{X_i\beta}{\sigma}\right) &= 1 - \Phi\left(\frac{X_i\beta}{\sigma}\right) \end{aligned}$$

$$f(y) = \left[\frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(y - \mathbf{x}'\beta)^2\right\} \right]^d \left[1 - \Phi\left(\frac{\mathbf{x}'\beta}{\sigma}\right) \right]^{1-d}, \quad (16.14)$$

where the binary indicator d is defined in (16.6) with $L = 0$.

The Tobit MLE $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$ maximizes the censored log-likelihood function (16.8). Given (16.14) this becomes

$$\begin{aligned} \ln L_N(\beta, \sigma^2) &= \sum_{i=1}^N \left\{ d_i \left(-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y_i - \mathbf{x}_i'\beta)^2 \right) \right. \\ &\quad \left. + (1 - d_i) \ln \left(1 - \Phi\left(\frac{\mathbf{x}_i'\beta}{\sigma}\right) \right) \right\}, \end{aligned} \quad (16.15)$$

a mixture of discrete and continuous densities. The first-order conditions are

$$\begin{aligned} \frac{\partial \ln L_N}{\partial \beta} &= \sum_{i=1}^N \frac{1}{\sigma^2} \left(d_i (y_i - \mathbf{x}_i'\beta) - (1 - d_i) \frac{\sigma \phi_i}{(1 - \Phi_i)} \right) \mathbf{x}_i = \mathbf{0} \\ \frac{\partial \ln L_N}{\partial \sigma^2} &= \sum_{i=1}^N \left\{ d_i \left(-\frac{1}{2\sigma^2} + \frac{(y_i - \mathbf{x}_i'\beta)^2}{2\sigma^4} \right) + (1 - d_i) \frac{\phi_i \mathbf{x}_i' \beta}{(1 - \Phi_i) 2\sigma^3} \right\} = 0, \end{aligned} \quad (16.16)$$

16.3.5. Marginal Effects in the Tobit Model

The marginal effect is the effect on the conditional mean of the dependent variable of changes in the regressors. This effect varies according to whether interest lies in the latent variable mean $\mathbf{x}'\beta$ or the truncated or censored means given in (16.23).

541

TOBIT AND SELECTION MODELS

Differentiating each with respect to $\mathbf{x}$ yields

$$\begin{aligned} \text{latent variable:} \quad & \partial E[y^*|\mathbf{x}]/\partial \mathbf{x} = \beta, \\ \text{left-truncated (at 0):} \quad & \partial E[y, y > 0|\mathbf{x}]/\partial \mathbf{x} = \{1 - w\lambda(w) - \lambda(w)^2\}\beta, \\ \text{left-censored (at 0):} \quad & \partial E[y|\mathbf{x}]/\partial \mathbf{x} = \Phi(w)\beta, \end{aligned} \tag{16.25}$$

where $w = \mathbf{x}'\beta/\sigma$ and we use $\partial \Phi(z)/\partial z = \phi(z)$ and $\partial \phi(z)/\partial z = -z\phi(z)$. The simple expression for the censored mean is obtained after some manipulation. It can be decomposed into two effects, one for $y = 0$ and one for $y > 0$ (see McDonald and Moffitt, 1980).

Exercício #17

```
clear all
set seed 54321

set obs 1000

gen x = rpoisson(1)
gen u = rnormal()
gen y_star = 1+2*x+u
gen y = y_star
replace y = 0 if y_star < 0

program tobit1
version 16
args lnf xb s
quietly replace `lnf' = ln(1-normal(`xb'/exp(`s')))) if $ML_y1 == 0
quietly replace `lnf' = -`s'+ln(normalden(($ML_y1-`xb')/exp(`s')))) if $ML_y1 > 0
end
ml model lf tobit1 (y = x) ()
ml check
ml maximize
disp exp(2*_b[eq2:_cons])

tobit y x
truncreg y x, ll(0)
```

The truncated *joint* density of y and z is

$$f(y, z | z > a) = \frac{f(y, z)}{\text{Prob}(z > a)}.$$

To obtain the incidentally truncated marginal density for y , we would then integrate z out of this expression. The moments of the incidentally truncated normal distribution are given in Theorem 19.5.²⁵

THEOREM 19.5 Moments of the Incidentally Truncated Bivariate Normal Distribution

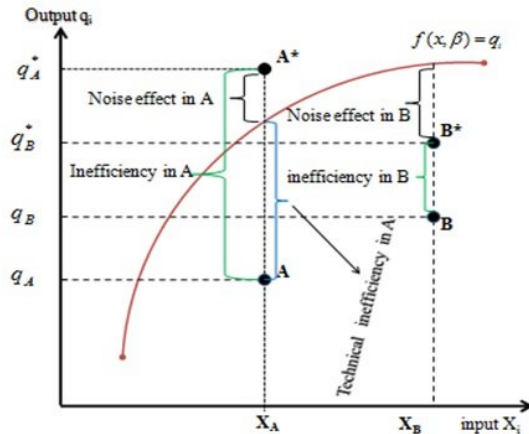
If y and z have a bivariate normal distribution with means μ_y and μ_z , standard deviations σ_y and σ_z , and correlation ρ , then

$$\begin{aligned} E[y | z > a] &= \mu_y + \rho\sigma_y\lambda(\alpha_z), \\ \text{Var}[y | z > a] &= \sigma_y^2[1 - \rho^2\delta(\alpha_z)], \end{aligned}$$

where

$$\alpha_z = (a - \mu_z)/\sigma_z, \lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)], \text{ and } \delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z].$$

<https://www.stata.com/manuals13/rheckman.pdf>



<https://www.stata.com/manuals/xtxtfrontier.pdf>

Análise de sobrevivência

Na abordagem de sobrevivência, uma vez que o crédito foi concedido, a duração da inadimplência é uma variável aleatória contínua não negativa T^* . No instante t , a PD é $P(T^* \leq t) = F(t)$, onde $F(\cdot)$ é uma função de distribuição acumulada. Logo, a probabilidade de sobrevivência da inadimplência é $S(t) = 1 - F(t)$, e consequentemente $f(t) = -S'(t)$, onde $f(\cdot)$ é a densidade de probabilidade.

Analogamente, o risco instantâneo de inadimplência é $h(t) = \lim_{\xi \rightarrow 0} \frac{P(t < T^* \leq t + \xi \mid T^* > t)}{\xi} = -\frac{S'(t)}{S(t)}$.

Ou seja, as funções de risco e sobrevivência são inversamente proporcionais em um ponto fixo de $f(\cdot)$.

$$\lim_{\xi \rightarrow 0} \frac{P(t < T^* \leq t + \xi \mid T^* > t)}{\xi} = \lim_{\xi \rightarrow 0} \frac{P(t < T^* \leq t + \xi)}{\xi} \frac{1}{P(T^* > t)} = \lim_{\xi \rightarrow 0} \frac{F(t + \xi) - F(t)}{\xi} \frac{1}{1 - F(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

Com essas definições, a modelagem começa com alguma conjectura acerca do processo gerador do tempo de inadimplência, através da suposição de uma forma funcional para $f(\cdot)$ ou $h(\cdot)$. Ao se modelar $f(\cdot)$, costumam ser usadas as distribuições Weibull e Gompertz. A primeira tende a se ajustar melhor aos dados quando é mais frequente observar a inadimplência nas primeiras parcelas dos contratos, e no caso contrário a segunda tende a se ajustar melhor.

A Tabela [...] apresenta as funções densidade, sobrevivência e risco para essas duas distribuições. Considerando n mutuários indexados por $i = 1, \dots, n$, o risco condicionado é fundamentalmente determinado por $\gamma_i = \exp(\sum_{k=1}^K \beta_k x_{ik})$, onde β_k é um parâmetro e $x_{ik} \in X_i$ é uma covariada de um conjunto contendo características do cliente (e.g., setor de atividade), do contrato (e.g., taxa de juros) ou da situação macroeconômica (e.g., desemprego). Se ocorre $\beta_k > 0$, $k \neq 0$, significa que aumentos em x_{ik} aumentam o risco de inadimplência. Complementarmente, β_0 é o parâmetro que regula a relação do risco com o tempo – e.g., no caso da Weibull, se $\beta_0 = 1$ a probabilidade de sobrevivência diminui com o tempo, embora o risco dependa exclusivamente do perfil representado por X_i . Ambas as distribuições geram análises denominadas de “risco relativo”, já que para todo perfil $i \neq j$, a razão de risco é $\frac{\gamma_i}{\gamma_j}$, que independe do tempo.

Tabela: Funções das distribuições Weibull e Gompertz.

Função	Weibull	Gompertz
Densidade, $f(t \mid \beta, X_i)$:	$\gamma_i \beta_0 t^{\beta_0 - 1} \exp(-\gamma_i t^{\beta_0})$	$\gamma_i \exp(\beta_0 t - \frac{\gamma_i}{\beta_0} [\exp(\beta_0 t) - 1])$
Sobrevivência, $S(t \mid \beta, X_i)$:	$\exp(-\gamma_i t^{\beta_0})$	$\exp(-\frac{\gamma_i}{\beta_0} [\exp(\beta_0 t) - 1])$
Risco, $h(t \mid \beta, X_i)$:	$\gamma_i \beta_0 t^{\beta_0 - 1}$	$\gamma_i \exp(\beta_0 t)$
$\beta_0 > 0$; $\gamma_i = \exp(\sum_{k=1}^K \beta_k x_{ik})$; no caso da Weibull, se $\beta_0 = 1$ ela é a Exponencial, se $\beta_0 > 1$ o risco é acelerativo e se $\beta_0 < 1$ risco é não acelerativo; a Gompertz é risco acelerativo $\forall \beta_0$.		

Para a estimação do vetor de parâmetros β , primeiro deve-se notar que a duração da inadimplência é uma variável latente. Isto é, ela é observável apenas no caso do cliente indicar saldo vencido de alguma fatura, e poderá ainda ser observada caso o contrato não tenha findado. Assim, se o sistema do banco registra a data de início do contrato, d_i , e os dias de atraso de pagamento, A_i , em um determinado dia D da coleta dos dados, o pesquisador pode computar a seguinte variável – cujo menor valor é zero, e o limite superior é móvel:

$$T_i = \begin{cases} \frac{D - d_i - A_i}{30} := T_i^*, & \text{se } A_i > 0 \\ \frac{D - d_i}{30}, & \text{se } A_i = 0 \end{cases}$$

A equação [...] define uma fração de tempo de 30 dias que indica o ponto de revelação da inadimplência caso ocorra saldo vencido, ou o ponto onde se encontra o contrato do mutuário ainda adimplente. Por exemplo, suponha um contrato assinado em 1º de outubro de 2013, se a data de coleta das informações foi 15 de abril 2015, então

$$T_i = \frac{D - d_i}{30} = \frac{561}{30} = 18,7. \text{ Se há saldo vencido desde 1º de outubro de 2014, então } A_i = 196, \text{ logo}$$

$T_i^* = \frac{561 - 196}{30} = 12,2$. Neste caso a inadimplência falhou 12,2 meses após a assinatura do contrato que perdura a 18,7 meses. Por consequência, sendo $I(\cdot)$ uma dummy de censura de observação da inadimplência, a densidade de probabilidade condicionada de T_i será:

$$\begin{aligned}\tilde{f}(T_i | \beta, X_i) &= \begin{cases} f(T_i | \beta, X_i), & \text{se } A_i > 0 \\ 1 - F(T_i | \beta, X_i) := S(T_i | \beta, X_i), & \text{se } A_i = 0 \end{cases} \\ &= [f(T_i | \beta, X_i)]^{1-I(A_i=0)} [S(T_i | \beta, X_i)]^{I(A_i=0)}\end{aligned}$$

Logo, na hipótese de T_i ser independente e identicamente distribuída, a função de log-verosimilhança será:

$$\ln L(\beta) = \sum_{i=1}^n [[1 - I(A_i = 0)] \ln f(T_i | \beta, X_i) + I(A_i = 0) \ln S(T_i | \beta, X_i)]$$

Caso o pesquisador defina uma forma funcional para $h(\cdot)$ ao invés de $f(\cdot)$, a mais comum é $h_i(t) = \tilde{h}(t)\gamma_i$, denominada de “abordagem semi-paramétrica de Cox”. Fazendo uma comparação com os casos das distribuições Weibull e Gompertz, as funções de risco seriam semelhantes na abordagem de Cox se $\tilde{h}(t) = \beta_0 t^{\beta_0 - 1}$ e $\tilde{h}(t) = \exp(\beta_0 t)$, respectivamente.

A base de dados foi construída a partir da carteira de crédito de um banco de desenvolvimento que atua no sul do país, tabulando-se os contratos ativos iniciados entre outubro de 2013 e março de 2015, assinados por empresas ou por agropecuaristas como pessoas físicas, tendo obrigações de pagamentos de qualquer natureza mesmo estando em prazo de carência, totalizando 7.461 observações. Devido a política de confidencialidade de informações do banco, seu nome não pode ser explicitamente citado, e só foi autorizada a apresentação de um recorte restrito das características dos contratos.

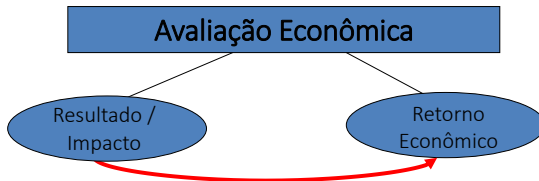
Em 15 de abril de 2015, data da coleta dos dados, 88 mutuários (1,2% do total observado) apresentavam saldo vencido há mais de 30 dias. Segundo os relatórios do Banco Central do Brasil (BCB) que tratam da inadimplência bancária no período em tela [...], dentre os bancos públicos, sendo ou não de desenvolvimento, os atrasos acima de 30 dias ocorreram em cerca de 1,5% dos empréstimos – o que indica que o nível de inadimplência da instituição aqui analisada estava próximo da média. Complementarmente, 391 mutuários apresentavam saldo vencido em até 30 dias, totalizando 479 contratos com pagamentos atrasados (6,4% do total observado).

Fazer o Exercício # 18, dados do BRDE

Avaliação econômica

INTRODUÇÃO

- A avaliação econômica é um componente importante de qualquer projeto social: “Apesar de bilhões de dólares serem gastos todos os anos com projetos sociais, muito pouco se sabe do seu impacto sobre beneficiários e sua viabilidade econômica” **(Banco Mundial... Pq esses caras entraram nessa?!)**
- A avaliação econômica pode ser dividida em duas partes: **avaliação de resultado/impacto e avaliação de retorno econômico**



PROJETOS SOCIAIS E POPULAÇÃO-ALVO

- Um projeto (programa ou política) social sempre parte do pressuposto de que se trará benefícios para uma **população-alvo** (ou público-alvo)
- Assim, para realizarmos uma avaliação econômica, primeiro é preciso definir claramente o grupo ao qual o projeto se destina, isto é, o grupo de pessoas que o projeto pretende beneficiar (o **grupo de tratamento**)
- Em geral, para a identificação clara da população-alvo, analisam-se os chamados **critérios de elegibilidade** ao projeto
 - ✓ Ex.: o Programa Bolsa Família atende todas as famílias que tenham renda familiar per capita abaixo do nível de pobreza

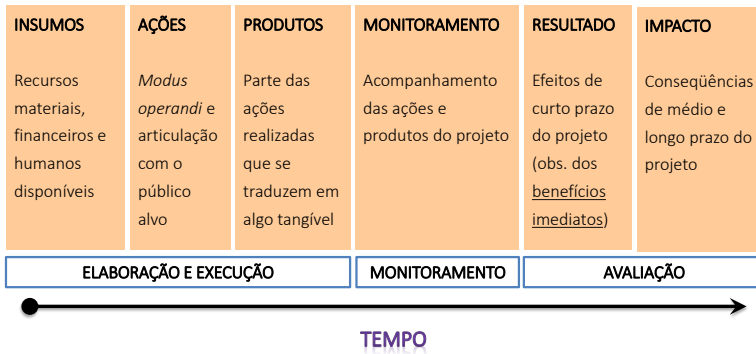
MARCO LÓGICO E AVALIAÇÕES EX-ANTE E EX-POST

- A decisão pela implementação de um programa é geralmente precedida pela apresentação de argumentos que justificam sua existência. Esta argumentação é chamada de **“Marco Lógico”**
- Quando a argumentação é feita com algum tipo de simulação quantitativa, diz-se que foi feita uma avaliação *ex-ante*
- A metodologia que será discutida no decorrer deste curso é a *ex-post*
- As avaliações *ex-ante* e *ex-post* se baseiam cada qual em um conjunto distinto de informações
 - ✓ Na avaliação *ex-ante*, o papel da teoria é primordial, uma vez que as estimativas de impacto são obtidas a partir de simulações sobre o comportamento hipotético dos futuros beneficiários do programa
 - ✓ Na avaliação *ex-post*, quando uma grande quantidade de informações foi coletada, a teoria será testada pela evidência empírica

MARCO LÓGICO E AVALIAÇÃO

- O Marco Lógico é uma ferramenta que deve facilitar o processo de justificativa, elaboração, execução e avaliação de projetos em gerais (sociais em particular)
- O objetivo do Marco Lógico é dar estrutura ao processo de planejamento e informação essencial relativa ao projeto
- Pode ser utilizado em todas as etapas de preparação, execução e avaliação do projeto
- Na perspectiva da avaliação, especificamente, é no Marco Lógico onde se deve encontrar os elementos de análise das relações causais e a forma de construção do grupo de controle **(ver o exemplo de aplicação do Ceará no Moodle)**

DESCRIÇÃO DO MARCO LÓGICO



INDICADORES

- Como dito, para a avaliação econômica é preciso indicadores objetivos e mensuráveis
 - ✓ Indicadores de **oferta** estão associados aos recursos necessários para a realização do projeto
 - ✓ Indicadores de **monitoramento** estão associados aos recursos utilizados para a realização do projeto
 - ✓ Indicadores de **resultado** estão associados aos objetivos de curto prazo do projeto (coletados para os grupos de tratamento e controle)
 - ✓ Indicadores de **impacto** estão associados aos objetivos de médio e longo prazo do projeto (coletados para os grupos de tratamento e controle)

INDICADORES – EXEMPLO: PROGRAMA SAÚDE DA FAMÍLIA



A RELAÇÃO ENTRE IMPACTO, AMBIENTE SOCIOECONÔMICO E CARACTERÍSTICAS DA POPULAÇÃO BENEFICIADA

- Projetos sociais são desenhados para uma população-alvo em um ambiente pré-estabelecido
- É de se esperar que uma mesma ação dirigida a outra população, outro local e outro momento do tempo não beneficie da mesma forma os que dela participam
- Uma das utilidades da avaliação econômica é, precisamente, determinar em que medida a eficiência, a eficácia e a efetividade de um programa depende das características da população beneficiada, do momento no tempo e da natureza do ambiente socioeconômico em que ocorre

EFICIÊNCIA, EFICÁCIA E EFETIVIDADE

- Eficiência é uma medida relativa, usada para comparar duas ações de um mesmo projeto ou dois projetos diferentes. Refere-se à relação entre os resultados obtidos (indicadores de resultado e/ou impacto) e os recursos empregados (indicadores de oferta)
- Eficácia é uma medida relacionada a uma meta. Diz-se que um projeto social foi eficaz se ele alcançou um objetivo mensurável por um indicador de impacto e/ou resultado
- Efetividade é uma medida de duração de um impacto. Isto é, se o projeto social for efetivo, se ele deixar de existir os resultados alcançados não serão desfeitos

O “BRILHO NOS OLHOS”

- Alguns projetos sociais objetivam causar impactos psicológicos na população-alvo. Por exemplo, melhorar a autoestima ou reduzir o nível de tensão em ambientes extremos (como é o caso do projeto “Doutores da Alegria”)
- Nestes casos, os avaliadores precisam medir o que popularmente ficou conhecido por “brilho nos olhos” das pessoas
- Para tanto, é possível usar escalas psicométricas para medir algo que parece tão intangível como o “brilho nos olhos”. Dentre estas escalas, é possível citar:
 - ✓ A escala Likert que explora o nível de concordância com alguma afirmação. Por exemplo, em relação a afirmação “você está feliz” é possível dizer que: (1) concordo; ou, (0) discordo
 - ✓ A escala Rosenberg, que é mais complexa que a Likert mas é muito popular em estudos psicométricos

CONTRAFACTUAL

- O problema central da área de avaliação de programas é construir o **contrafactual** do grupo tratado pelo programa
- Na forma mais simples de apresentar o problema, podemos pensar que qualquer indivíduo está sempre em uma de duas situações mutuamente excludentes: ter sido ou não ter sido tratado
- Idealmente, o melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem tratados. Contudo, claramente não é possível observar os mesmos indivíduos na condição de tratados e de não tratados ao mesmo tempo

CONTRAFACTUAL, GRUPO DE CONTROLE E RESULTADO POTENCIAL

- O desafio do avaliador é encontrar um grupo de indivíduos que represente adequadamente a situação de não tratamento, ou seja, um grupo que funcione como um bom contrafactual do grupo tratado, chamado de **grupo de controle**
- Um elemento crucial para um bom entendimento desse arcabouço é ter sempre em mente a **potencial existência** de uma situação contrafactual
- Por isso usamos o nome de **“Modelo de Resultados Potenciais”** para o que veremos daqui a pouco

MÉTODOS INGÊNUOS DE COMPARAÇÃO

- Antes de apresentarmos formalmente o **Modelo de Resultados Potenciais**, vale a pena analisar dois métodos “ingênuos”, porém amplamente utilizados por não especialistas para construir o grupo de controle
 - ✓ O primeiro é baseado na comparação do que ocorre com o grupo de tratamento antes e depois da intervenção
 - ✓ O segundo consiste em comparar o grupo que escolhe passar pelo programa com um grupo que decide não participar da intervenção
- Como veremos, esses métodos dificilmente são capazes de fornecer o impacto do programa

COMPARANDO OS TRATADOS ANTES E DEPOIS

- Considere uma situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade numa certa localidade
- Esse programa consistia de um curso ministrado por pessoal especializado e tinha como objetivo aumentar o rendimento do trabalho dos participantes
- Um pouco antes do começo do programa, todos os participantes responderam um questionário que levantou informações sobre suas características socioeconômicas, incluindo a situação de emprego e rendimento laboral naquele momento

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

- Suponha que o rendimento médio do trabalho para todos os indivíduos que entraram no programa de treinamento foi de R\$ 1.000
- Um ano depois, os mesmos tipos de informações foram novamente coletadas com todos os participantes do programa. O rendimento médio encontrado nesse segundo momento foi igual a R\$ 1.100
- Alguns diriam que o efeito do programa foi de R\$ 100, o que equivale a um aumento de 10% no rendimento médio inicial. Mas será que podemos afirmar que esse foi o verdadeiro efeito do programa?
- Para responder essa pergunta, temos que nos questionar se o grupo de tratamento antes do programa representa o contrafactual adequado para a situação de não tratamento depois do programa

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

- Existe uma série de fatores que podem ter afetado o rendimento médio do grupo tratado além do programa
- Por exemplo, é possível que a situação do mercado de trabalho da localidade onde residem os participantes do programa tenha melhorado ao longo do tempo entre as coletas das informações
- A menos que possamos controlar a influência do mercado de trabalho e de todos os outros fatores que afetam o rendimento médio do trabalho, a simples comparação entre o que ocorre com o grupo de tratados antes e depois do programa não identifica corretamente o efeito causal de uma intervenção

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

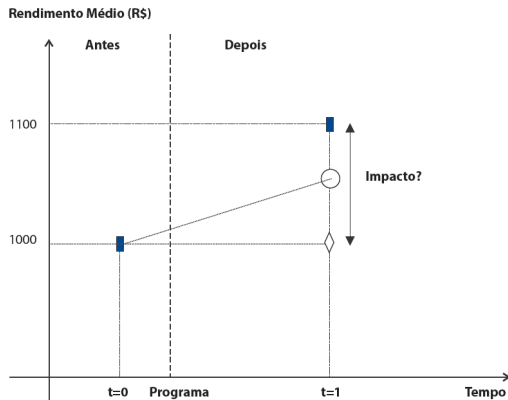


Figura da p. 37.

O verdadeiro impacto é ilustrado pela distância entre o retângulo e o círculo

COMPARANDO OS TRATADOS COM OS NÃO TRATADOS

- Outra prática “ingênua” é a que compara o grupo de tratados diretamente com um grupo de não tratados. Apenas por não ter passado pela intervenção não significa que o grupo de não tratados representa bem o que ocorreria com o grupo de tratamento caso este não tivesse sido tratado
- As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis (pelo avaliador). É possível que, mesmo que os grupos sejam parecidos em atributos observáveis (gênero, idade, escolaridade, renda familiar etc.), eles difiram em atributos não observáveis (habilidades, motivação, esforço etc.)
- Na medida em que muitas das características podem afetar a variável de resultado sobre a qual se quer medir o impacto, o uso da comparação direta entre os grupos não necessariamente mede o efeito causal do programa. Afinal, as diferenças nas características não observáveis poderiam estar misturadas ao efeito do programa

O ARCABOUÇO DE RESULTADOS POTENCIAIS

- Para apresentar o arcabouço de modo formal, utilizaremos o algarismo “1” para denotar a situação em que o indivíduo é tratado e o algarismo “0” para a situação contrafactual na qual esse mesmo indivíduo não é tratado
- Estamos interessados em avaliar o impacto de um programa (e.g., um programa de qualificação de mão de obra) sobre uma variável de resultado (e.g., o rendimento do trabalho)
- Denotando por y essa variável, podemos pensar que seu valor para indivíduo i será $y_i(1)$ caso ele esteja na situação de tratado, mas poderia assumir, pelo menos potencialmente, o valor $y_i(0)$ caso ele não seja tratado
- A vantagem da formalização é explicitar que cada uma das situações está associada a um resultado potencialmente distinto para um mesmo indivíduo. Assim, se fosse possível observar o indivíduo nas duas situações, a diferença $\beta_i = y_i(1) - y_i(0)$ seria o impacto do programa para i

O ARCABOUÇO DE RESULTADOS POTENCIAIS (CONTINUAÇÃO)

- Seja T_i uma variável discreta que assume valor igual a **1** se o indivíduo participa do programa e valor **0** caso ele não participe. Trata-se, portanto, do que chamamos de uma variável binária (ou *dummy*, em inglês), cujos valores unitário ou nulo podem depender da decisão do indivíduo de participar ou não da intervenção, respectivamente
- Dado o apresentado até aqui, podemos expressar a equação de resultados potenciais como:

$$\begin{aligned} y_i &= T_i y_i(1) + (1 - T_i) y_i(0) \\ &= y_i(0) + T_i (y_i(1) - y_i(0)) \\ &= y_i(0) + \beta_i T_i \\ &= \alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_K x_{Ki} + \varepsilon_i + \beta_i T_i \end{aligned}$$

Eq. (5) da p. 41

Cada “ x ” é uma covariada (um atributo observável que pode explicar parte do indicador de resultado), e “epsilon” representa os fatores não observáveis

ATT

- Em geral, estamos interessados em computar o que se costuma denominar de efeito médio do programa sobre os tratados (EMPT ou **ATT**, do inglês *average treatment effect on the treated*)

$$\mathbf{ATT} = E[\mathbf{y}_i(\mathbf{1})|\mathbf{T}_i = \mathbf{1}] - E[\mathbf{y}_i(\mathbf{0})|\mathbf{T}_i = \mathbf{1}] = E[\boldsymbol{\beta}_i]$$

✓ Onde $E[\dots]$ é o operador “esperança matemática”

- O problema enfrentado pelo avaliador na estimação do **ATT** é que não se observa $\mathbf{y}_i(\mathbf{0})|\mathbf{T}_i = \mathbf{1}$, logo não se observa $E[\mathbf{y}_i(\mathbf{0})|\mathbf{T}_i = \mathbf{1}]$

VIÉS DE AUTO SELEÇÃO

- Por outro lado, é possível observar $y_i(0)|T_i = 0$
- Notando isso, ingenuamente o avaliador pode tentar substituir $E[y_i(0)|T_i = 1]$ por $E[y_i(0)|T_i = 0]$ na estimação do ATT, viesando a estimativa do resultado

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 1]$$

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 0] + E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]$$

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 0] + \underbrace{E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]}_{\text{viés}}$$

- Este viés também é chamado de “viés de auto seleção”

VIÉS DE AUTO SELEÇÃO: EXEMPLO

- Considere uma situação em que um programa de treinamento profissional
- y_i é o salário observado após o treinamento
- Existe uma variável não observada, x_i , que representa “motivação”. Se $x_i = 1$ o indivíduo é motivado, e se $x_i = 0$ o indivíduo não é motivado
- Considere que apenas indivíduos motivados busquem o treinamento.

Então ocorre que

$$\underbrace{E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]}_{\text{viés}} = \underbrace{E[y_i(0)|x_i = 0] - E[y_i(0)|x_i = 1]}_{\text{viés}}$$

- Portanto, se os indivíduos motivados tiverem salários maiores (na média), existirá um viés de auto seleção

MÉTODO DE ALEATORIZAÇÃO

- O principal problema da avaliação é encontrar um grupo de controle que represente adequadamente o contrafactual do grupo tratado
- A situação contrafactual pode ser **bastante bem aproximada** pelo uso do método de aleatorização. Este método, muitas vezes chamado de método experimental, é baseado na seleção aleatória dos que farão parte dos grupos de tratamento e controle
- Quando bem implementado, o mecanismo da aleatorização fornece um balanceamento satisfatório tanto das características observadas quanto das não observadas das unidades que compõem os dois grupos
- Por construção, o método permite criar uma situação na qual não há correlação entre ser ou não tratado e os atributos das unidades de observação
- Assim, o problema do viés de auto seleção é mitigado

ALEATORIZAÇÃO TENDE A ELIMINAR O VIÉS DE AUTO SELEÇÃO

- Pela própria natureza da aleatorização, as esperanças condicionais serão idênticas as esperanças não condicionais
- Portanto ocorre:

$$\underbrace{E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]}_{\text{viés}} = E[y_i(0)] - E[y_i(0)] = 0$$

CUSTOS DA ALEATORIZAÇÃO

- Por conta de fatores éticos, financeiros e técnicos o método de aleatorização está longe de ser o predominantemente utilizado
- Sem dúvida, não se cogita utilizar aleatorização para casos de intervenções de ampla escala (e.g., programas de vacinação), ou para públicos-alvo muito específicos (e.g., pessoas com algum tipo de deficiência). No entanto, há pelo menos duas situações importantes em que esse problema ético não se coloca:
 - ✓ Quando existe escassez de vagas para atender toda a demanda do programa. Nessa situação, pode-se inclusive justificar o uso da aleatorização como o critério mais justo de alocação das vagas existentes entre os candidatos ao programa
 - ✓ Quando os recursos do programa são limitados e, apesar de se planejar atender toda a população-alvo ao final da implementação, só é possível incluir os elegíveis ao longo do tempo. Em casos como esse, ninguém será privado de participar do programa; o que se aleatoriza é somente a ordem com que os elegíveis entram no programa

OS RESULTADOS POTENCIAIS E A ANÁLISE DE REGRESSÃO

- No modelo de resultados potenciais temos y_i e T_i como um indicador de resultado e uma *dummy* que indica “tratamento”, respectivamente. Lembre-se da aula passada que o modelo é baseado na seguinte equação:

$$y_i = T_i y_i(1) + (1 - T_i) y_i(0) = y_i(0) + T_i (y_i(1) - y_i(0)) = y_i(0) + \beta_i T_i$$

- Definimos que o valor de y para indivíduo i será $y_i(1)$ caso ele esteja na situação de tratado, mas poderia assumir, potencialmente, o valor $y_i(0)$ caso ele não seja tratado. Assim explicitamos que cada uma das situações está associada a um resultado potencialmente distinto para um mesmo indivíduo e, se fosse possível observar o indivíduo nas duas situações, a diferença β_i seria o impacto do programa para i

EXEMPLO DE ESTRATÉGIA DE IDENTIFICAÇÃO

- Como a variável estudada assume o valor $y_i(0)$ caso o indivíduo não seja tratado, ela será função apenas de características específicas dele. Por exemplo, suponha um contexto de treinamento profissional, onde $y_i(0)$ seria o salário do indivíduo. O salário deve variar em função atributos observáveis (gênero, idade, escolaridade, renda familiar etc.) e atributos não observáveis (habilidades, motivação, esforço etc.)
- Para simplificar, suponha que exista um salário mínimo dado por α_0 (R\$). E que o salário em média aumente em α_1 (R\$) para cada ano de escolaridade x_{1i} (que assume os valores $x_{1i} = 0, 1, 2, \dots$)
- Suponha também que a escolaridade é o único atributo observado pelo pesquisador, e que escrevemos um índice ε_i representando todo o mais que não observamos
- Nestas condições, o pesquisador pode definir o seguinte modelo: $y_i(0) = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i$

EXEMPLO DE ESTRATÉGIA DE IDENTIFICAÇÃO (CONTINUAÇÃO)

- Sob a hipótese de que $y_i(0) = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i$, o salário dos indivíduos (tratados ou não tratados pelo programa) deve ser representado pela seguinte equação:

$$y_i = y_i(0) + \beta_i T_i = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i + \beta_i T_i$$

- Para simplificar, faz-se a hipótese adicional de que o impacto do programa é igual para todos os indivíduos, ou seja, $\beta_i = \beta$ para qualquer indivíduo da população. Assim, a equação acima se transformará na equação (5) da p. 41 do livro. Onde β é o **ATT** do programa

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \beta T_i + \varepsilon_i$$

- Na sequência da aula nos dedicaremos a entender a técnica dos “mínimos quadrados”, que usaremos para calcular de α_0 , α_1 e β e verificar se nossas hipóteses de fato mimetizam a realidade, ou seja, se construímos uma boa “estratégia de identificação”

CONSIDERAÇÕES SOBRE AS LIMITAÇÕES DO MÉTODO

- O uso do método da regressão linear que veremos a seguir possui duas grandes limitações na avaliação de impacto:
 - ✓ A primeira, discutida em seguida, é que ele mede efeitos de correlação entre a variável de tratamento (T_i) e a variável de impacto (y_i). Então a conclusão sobre a causalidade dependerá da razoabilidade das hipóteses feitas na estratégia de identificação (que nem sempre são testáveis)
 - ✓ E a segunda, que é vista em cursos avançados de econometria, é que a análise de regressão pressupõe a ausência de correlação entre ε_i e todos os outros regressores (x_i e T_i). Se essa hipótese for violada, as estimativas dos parâmetros α_0 , α_1 e β serão viesadas
- Quanto a esta segunda limitação, as técnicas de “diferenças em diferenças” e “pareamento”, que serão vistas nas próximas aulas, tem o propósito de resolver a questão

Pareamento

MÉTODO DE AVALIAÇÃO POR PAREAMENTO

- A técnica do “pareamento” é um método não-experimental que busca construir um grupo de controle semelhante ao grupo de tratamento em termos de determinadas características observáveis pelo pesquisador
- De acordo com as hipóteses do método, cada membro do grupo de tratamento teria um par no grupo de controle que representa o resultado que ele teria obtido caso não fosse tratado
- De outra forma, as hipóteses do pareamento postulam que, ao comparar dois indivíduos, um no grupo de controle e outro no grupo de tratamento, com as mesmas características observáveis, o único fator que diferencia os resultados destes indivíduos é a participação ou não no programa

1ª HIPÓTESE: SELEÇÃO NOS OBSERVÁVEIS

- A hipótese principal deste método, que tem como objetivo estimar o efeito médio do tratamento sobre os tratados (ATT), é que o vetor de variáveis observáveis X contém todas as informações que descrevem o resultado potencial na ausência do tratamento, $Y(0)$
- Ao levar em consideração estas características observáveis, o avaliador controla por todas as variáveis que estão relacionadas ao resultado potencial na ausência de tratamento e que também afetam a decisão do indivíduo em participar ou não
- Assim, ao controlar pelo vetor X , a variável $Y(0)$ torna-se independente de T
- Esta hipótese é conhecida como seleção nos observáveis, ou ignorabilidade, ou ainda não-confundimento

2ª HIPÓTESE: SUPORTE COMUM

- Precisamos que a região do vetor X que engloba as características dos indivíduos tratados também represente as características dos indivíduos que estão no grupo de não-tratados
- Então é preciso que, ao observar apenas o vetor X , o pesquisador não saiba ao certo se o indivíduo participou ou não do tratamento
- Formalmente, isto significa que

$$0 < Pr[T_i = 1|X_i] < 1$$

PROBLEMAS COMUNS COM O MÉTODO DE PAREAMENTO

Este método deve lidar com dois tipos de problemas:

1. O problema da dimensão:
 - Com K variáveis **binárias**, o número de potenciais pares é 2^K e cresce exponencialmente com K
 - Com K variáveis **contínuas**, a dimensionalidade é intratável
2. Falta de suporte comum: isto ocorre quando não é possível encontrar uma unidade no grupo de controle que tenha os mesmos valores de X para uma unidade do grupo de tratados

RESOLVENDO O PROBLEMA DA DIMENSIONALIDADE: ESCORE DE PROPENSÃO

- O “escore de propensão” transforma o problema multidimensional em um problema unidimensional
- O “escore de propensão” é a probabilidade condicional de receber o tratamento dadas as variáveis observadas X antes do tratamento:

$$p(X) = \Pr\{T = 1 | X\} = E\{T | X\}$$

ESTRATÉGIA DE ESTIMAÇÃO DO ATT USANDO PAREAMENTO

✓ Passos a seguir:

1. Estimação do “escore de propensão”
2. Identificação do suporte comum
3. Identificação da estimativa de efeito individual
 $\beta_i = y_i(1) - y_i(0)$ dado o “escore de propensão”
4. Estimação do efeito médio de tratamento (ATT)

PAREAMENTO POR VIZINHO MAIS PRÓXIMO

Pela observação anterior, precisamos agora de um critério para parear as unidades de tratamento e controle. Usaremos aqui o critério do “vizinho mais próximo” (detalhes nas pág. 90 a 92 do livro).

- **Vantagens:**
 - Para cada unidade tratada sempre é encontrado um par, não tratado, evitando a exclusão de observações tratadas
 - É facilmente implementado usando as fórmulas do Excel
- **Principal Problema:** em algumas bases de dados pode gerar pareamento de indivíduos com escore de propensão muito diferentes, pois o vizinho mais próximo pode não ser tão próximo. O que não será o caso deste exercício em particular, já que para toda combinação **X** no grupo de tratamento encontra-se uma igual combinação no grupo de controle.

PAREAMENTO E REGRESSÃO LINEAR

- Primeiro, porque o método de regressão linear usou todas as informações da base de dados, e não apenas aquelas que se encontravam no suporte comum. Isso fez com que se ponderasse uma parte do grupo de controle que não mimetiza satisfatoriamente o contrafactual do tratamento, gerando um viés.
- Segundo, porque o estimador de mínimos quadrados (regressão linear) parte do pressuposto de que não existe correlação entre os regressores e o termo de erro. Mas a escolaridade da mãe (observada) pode estar correlacionada com a habilidade (não observada) do aluno, gerando outra fonte de viés.
- Terceiro, se tivéssemos usado uma outra regra de pareamento, diferente do “vizinho mais próximo”, teríamos encontrado outro ATT



ATUAL ARQUIVOS SOBRE ▾

INÍCIO / ARQUIVOS / V. 20 N. 3 (2016) / Artigos

Uma avaliação econômica do Plano Setorial de Qualificação (PLANSEQ)

Francis Carlo Petterini
Universidade Federal de Santa Catarina. Departamento de Economia e Relações Internacionais

DOI: <https://doi.org/10.11606/1413-8050/ea117882>

Palavras-chave: PLANSEQ, PLANFOR, qualificação profissional, avaliação econômica

 **PDF**

PUBLICADO

2016-09-01

COMO CITAR

Regressão na descontinuidade

Sharp RD design

- **Notation:** $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \dots, n$, X_i continuous

- **Treatment:** $T_i \in \{0, 1\}$, $T_i = \mathbf{1}(X_i \geq \bar{x})$.

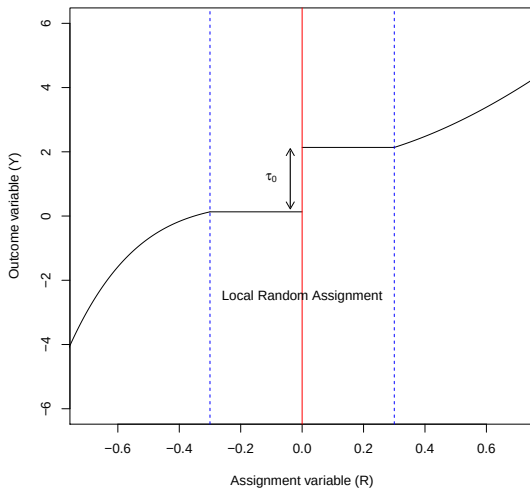
- **Data:** (Y_i, T_i, X_i) , $i = 1, 2, \dots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- **Average Treatment Effect at the cutoff:**

$$\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = \bar{x}] = \lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i | X_i = x]$$

- **Quasi-Experimental Design:** “local randomization” (more later)



Artigos • Rev. adm. contemp. 24 (1) • Jan-Feb 2020 • <https://doi.org/10.1590/1982-7849rac2020190230> [COPIAR](#)

Política de Expansão das Universidades Federais: É Possível Potencializar os Impactos Econômicos?

Expansion of Brazilian Federal Universities: Is It Possible to Raise Economic Impacts?

Marcelo Ponte Barbosa Francis Carlo Petterini Roberto Tatiwa Ferreira [SOBRE OS AUTORES](#)

- » Resumo
- » Abstract
- » Text
 - Introdução
 - Resultados
 - Conclusão
- » ANEXO APÊNDICES
- » Referências
- » Dados de Publicação
- » Histórico



Resumo

A investigação dos efeitos de um campus universitário sobre a economia local é ainda incipiente no Brasil. Mesmo em nível internacional, são raros estudos que utilizem métodos robustos de estimação desses efeitos e, muito menos, que busquem diferenciá-los ao longo do tempo. Neste artigo avaliam-se os impactos iniciais da política de expansão das universidades federais ocorrida na década de 2000 sobre as economias municipais. Para tanto, partindo de um amplo conjunto de dados socioeconômicos, aplica-se a técnica de pareamento por escore de propensão e, em seguida, estima-se o Efeito Tratamento sobre Tratados pela política. Os resultados indicam que a expansão da rede federal elevou a renda per capita dos municípios beneficiados entre 3,5% e 5,8%, em média, mas que não houve ganho adicional relacionado ao tempo de implantação dos campi. Este padrão sugere que os novos campi foram capazes de gerar um impulso inicial sobre o nível de demanda local, fruto dos gastos envolvidos na implantação e funcionamento, mas sinaliza para uma possível dificuldade em gerar saltos de produtividade normalmente esperados com a implantação dessas estruturas. Neste caso, seria possível buscar um maior retorno desses investimentos a partir da reavaliação dos critérios para criação de novos campi e da melhor focalização das ações daqueles já existentes.

Palavras-chave:

Dif-in-Dif

MÉTODOS DE AVALIAÇÃO

- Os métodos de avaliação de impacto são geralmente divididos em dois grupos: **experimental** e **não-experimental**
 - ✓ O primeiro é baseado na seleção aleatória dos participantes e não participantes do programa e é considerado o método de referência na área de avaliação
 - ✓ O segundo grupo é composto por uma variedade de métodos distintos que, embora façam uso de hipóteses específicas para identificar o efeito causal do programa, procuram mimetizar o método experimental
- A técnica das diferenças em diferenças (DD) é um método não-experimental que requer dados amostrais para os tratados e não tratados em pelo menos dois períodos no tempo: **antes e depois da intervenção**

HIPÓTESE

- A principal hipótese do método de DD é que a trajetória temporal da variável de resultado para o grupo de controle representa o que ocorreria com o grupo tratado caso não houvesse a intervenção
- Naturalmente, essa hipótese não pode ser testada diretamente nos dados, mas uma indicação de sua validade aparece quando as trajetórias dos dois grupos são parecidas pré-programa
- A ideia básica é que, se as trajetórias se assemelham durante o período antes do programa, então parece razoável supor que a evolução do grupo de controle após o programa represente com fidedignidade o que ocorreria com o grupo de tratados

A ilustração gráfica do exemplo da p. 77

OUTRA VANTAGEM PARA O CASO DE DADOS EM PAINEL

- Outra importante vantagem do método de DD é que ele permite controlar para características não observáveis dos indivíduos que sejam invariantes no tempo (por exemplo, habilidades inatas)
- Embora o método de DD só consiga levar em conta esse tipo de atributo que é fixo no tempo, esse controle é capaz de resolver o problema de auto seleção para um conjunto amplo de aplicações
- Essa capacidade confere ao DD uma vantagem relativa a outros métodos não-experimentais que não conseguem controlar para vieses de auto seleção decorrentes de atributos não observáveis, fixos ou não no tempo

O MODELO DE DIFERENÇAS EM DIFERENÇAS

- A forma mais simples de expressar o estimador do método de DD é calculando uma dupla diferença de médias da variável de resultado. Se denotarmos por $T = \{1,0\}$ a participação ou não no programa e por $t = \{1,0\}$ os períodos posterior e anterior à intervenção, respectivamente, o estimador de DD será dado por:

$$\beta_{DD} = \{E[Y|T = 1, t = 1] - E[Y|T = 1, t = 0]\} \\ - \{E[Y|T = 0, t = 1] - E[Y|T = 0, t = 0]\}$$

- Embutida nesse estimador está a hipótese de que a variação temporal na variável de resultado para o grupo de controle representa a variação contrafactual do grupo tratado. Em decorrência dessa hipótese, a diferença entre a variação efetivamente observada para o grupo de tratamento e a variação contrafactual fornecida pelo grupo de controle vai capturar o efeito causal da intervenção

UMA IMPORTANTE OBSERVAÇÃO

- A média da variável de resultado para o grupo de controle no período anterior ao programa não precisa coincidir com a média correspondente para o grupo de tratamento, ou seja, os grupos podem ou não partir de um mesmo ponto
- Na realidade, na maior parte das aplicações do procedimento de DD, os dois grupos têm essas médias distintas, um fato que tipicamente reflete as diferentes influências dos atributos observáveis e não observáveis dos indivíduos sobre a variável de resultado
- O que o método de fato requer é que a variação temporal do que ocorre com o grupo de controle antes e depois do programa reflita corretamente a variação temporal do grupo de tratados na situação contrafactual de não tratamento

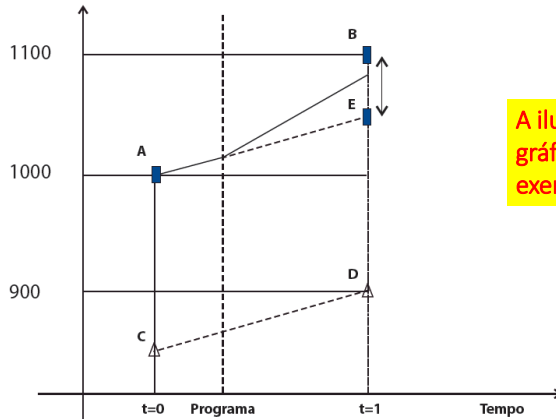
EXEMPLO DO LIVRO

- Considere uma situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade numa certa localidade
- Esse programa consistia de um curso ministrado por pessoal especializado e tinha como objetivo aumentar o rendimento do trabalho dos participantes
- O interesse recaí sobre o impacto do programa sobre o salário médio dos tratados, que era de R\$ 1.000 um pouco antes do início do curso e de R\$ 1.100 alguns meses após o programa
- A tabela a seguir contém os valores dos salários médios dos grupos de tratamento e controle – respectivamente, (T_1, T_0) – para os períodos anterior e posterior ao programa – (t_0, t_1) , respectivamente

	Tratados	Controle	Diferença
Antes	$\bar{Y}_{T_1, t_0} = 1000$	$\bar{Y}_{T_0, t_0} = 850$	$\Delta \bar{Y}_{T_1 - T_0, t_0} = 150$
Depois	$\bar{Y}_{T_1, t_1} = 1100$	$\bar{Y}_{T_0, t_1} = 900$	$\Delta \bar{Y}_{T_1 - T_0, t_1} = 200$
Diferença	$\Delta \bar{Y}_{T_1, t_1 - t_0} = 100$	$\Delta \bar{Y}_{T_0, t_1 - t_0} = 50$	$\begin{aligned} \Delta \bar{Y}_{T_1, t_1 - t_0} - \Delta \bar{Y}_{T_0, t_1 - t_0} \\ &= \Delta \bar{Y}_{T_1 - T_0, t_1} \\ &\quad - \Delta \bar{Y}_{T_1 - T_0, t_0} \\ &= 50 \end{aligned}$

- A notação $\bar{Y}$ representa a média amostral do salário e Δ significa o operador de diferença

Rendimento Médio (R\$)



A ilustração gráfica do exemplo

UTILIZANDO A REGRESSÃO LINEAR

- ✓ $\text{Renda} = \alpha + \beta_1 \text{'programa'} + \beta_2 \text{'depois'} + \beta_3 \text{'programa*depois'} + \text{erro}$
- ✓ β_3 = a diferença das diferenças, ou seja, é o coeficiente que mede o impacto do programa.
- ✓ β_1 captura se os grupos são diferentes, independentemente do programa.
- ✓ β_2 captura se o indicador muda no tempo, independentemente do programa.

PORQUE β_3 É O ESTIMADOR DE DIFERENÇAS EM DIFERENÇAS?

Médias por grupo	Antes	Depois	Variação
Tratamento	TA	TD	TD - TA
Controle	CA	CD	CD - CA
Variação das variações	(TD - TA) - (CD - CA)		

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

$$\checkmark \text{ TA} = \alpha + \beta_1 \times \text{'1'} + \beta_2 \times \text{'0'} + \beta_3 \times 0$$

$$\checkmark \text{ TA} = \alpha + \beta_1$$

$$\checkmark \text{ TD} = \alpha + \beta_1 \times \text{'1'} + \beta_2 \times \text{'1'} + \beta_3 \times 1$$

$$\checkmark \text{ TD} = \alpha + \beta_1 + \beta_2 + \beta_3$$

$$\checkmark \text{ TD} - \text{TA} = \beta_2 + \beta_3$$

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

$$\checkmark \text{ CA} = \alpha + \beta_1 \times \text{'0'} + \beta_2 \times \text{'0'} + \beta_3 \times 0$$

$$\checkmark \text{ CA} = \alpha$$

$$\checkmark \text{ CD} = \alpha + \beta_1 \times \text{'0'} + \beta_2 \times \text{'1'} + \beta_3 \times 0$$

$$\checkmark \text{ CD} = \alpha + \beta_2$$

$$\checkmark \text{ CD} - \text{CA} = \beta_2$$

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

$$\checkmark (TD - TA) = \beta_2 + \beta_3$$

$$\checkmark (CD - CA) = \beta_2$$

$$\checkmark (TD - TA) - (CD - CA) = \text{dif em dif} = (\beta_2 + \beta_3 - \beta_2) = \beta_3$$

Controle sintético



ATUAL ARQUIVOS SOBRE ▾

INÍCIO / ARQUIVOS / V. 48 N. 1 (2018) / Artigo

Maiores navios no mundo, mais um desafio no Brasil: uma análise do programa nacional de dragagem

Luciano Menegazzo
Universidade Federal de Santa Catarina
<http://orcid.org/0000-0002-3838-1318>

Francis Petterini
Universidade Federal de Santa Catarina
<http://orcid.org/0000-0003-4410-0970>



DOI: <https://doi.org/10.1590/0101-416148161751mf>

Econometria de cópulas



Economics Bulletin

General Information

- Editorial Board
- Mission Statement
- Directory of Economists

Submissions

- New Submission
- Submit a Revision
- Submit a Final Draft
- Submit Referee Report

Guidelines

- Referee Guidelines
- Author Guidelines

Contents

- Search
- Content by years

Profile and Directory

- Add me to Directory
- Modify my Profile
- Lost Password

All Rights Reserved
2008, 2008.
Powered by JSC
www.journal.org

Francis Petterini and Akauã Flores

"Copula econometrics to simulate effects of private policing on crime"
(2021, Vol. 0 No.0)

Clotfelter's model indicates that an increase in private policing has an ambiguous effect on crime across a city. Although this is often cited in the literature, we have found few attempts to test this hypothesis, perhaps because it involves several econometric challenges. This research presents a framework to solve this using city-level samples, applying copulas in Tobit marginals over indicators of crime and private policing. Using a sample of Brazilian cities, we find evidence in favor of Clotfelter's hypothesis.

Keywords: copula econometrics, private policing, economics of crime

JEL: C1 - Econometric and Statistical Methods: General

K4 - Legal Procedure, the Legal System, and Illegal Behavior: General

Description of Supplemental Material:

Data and codes for Stata.

[EB-20-01192-Supplemental-Material.zip](#)

Manuscript Received : Dec 08 2020

Manuscript Accepted : Jul 18 2021

This abstract has been downloaded 161 times

The Full PDF of this paper has been downloaded 135942 times

Econometria de leilões

- Estimar preços de reserva
- Aplicações
- Custo computacional x precisão
- Mais referências: <https://core.ac.uk/download/pdf/41339395.pdf>.
- Comportamento da relação não é claro.
- $p = X\beta + \varepsilon$. Viés de especificação? (Preferências)
- $V_i = X\beta + \varepsilon_i$. V_i é observável?

- 1) $V_{il} = \mu_l + \sigma_l \varepsilon_{il}$, aonde $\varepsilon_{il} = (V_{il} - \mu_l)/\sigma_l$ e ε_{il} segue uma distribuição F qualquer em comum para todos os i licitantes e l leilões.
- 2) $\mu_l = X_l \beta$ e $\sigma_l = Z_l \alpha$
- 3) n_l é o número de licitantes, este é conhecido por todos e é exógenamente determinado. Cada licitante maximiza seu *payoff* para cada leilão separadamente.
- 4) Regras do leilão são tais que o licitante que mais valore o bem ganha o leilão, e o licitante que menos valore o bem espera pagar nada. Os leilões Inglês; o de “maior lance selado”; segundo preço e *all-pay* respeitam esta hipótese.
- Ex: <https://receita.economia.gov.br/sobre/acoes-e-programas/mercadorias-apreendidas/leilao/leilao-eletronico/manual-sle/manual-do-licitante-setembro-2020.pdf> pg8

- *expected revenue equivalence*: sob as hipóteses anteriores (exceto a 2) tem-se que o pagamento esperado pelo item do leilão l é:
$$E(V_{(2:n_l)l}) \rightarrow E(p_l|X_l, Z_l, n_l) = E(V_{(2:n_l)}|X_l, Z_l, n_l)$$
- Suponha que conheçamos F .
- $E(p_l|X_l, Z_l, n_l) = \mu_l + \sigma_l E(\varepsilon_{(2:n)})$
- $a(n) = E(\varepsilon_{(2:n)}) = n(n-1) \int tF(t)^{n-1}(1-F(t))dF(t)$
- tem-se: $E(p_l|X_l, Z_l, n_l) = X_l\beta + a(n)Z_l\alpha$
- F comuns de se trabalhar são: Uniforme, Normal, Logística, Laplace, Gumbel.
- $a(n)$ existe para $n \geq 2$.
- Estimador MQO não é eficiente.

- Dados: 654 leilões de ipod mini realizados no ebay entre o período de 27/06/2006 a 18/07/2006.
- Originalmente foram realizados 1225 leilões, mas excluiu-se leilões com altos preços de reserva (maiores que U\$ 17)
- tem-se que $POS = \text{Log}_2(P + 1)$ e $NEG = \text{Log}_2(N + 1)$ aonde P é o total de reviews positivos do vendedor e N é o total de reviews negativos do vendedor

Table III. Estimated coefficients, F Gumbel

Regressor	OLS				WLS		
	I	II	III	IV	I	II	III
<i>On μ</i>							
Constant	65.9728 (20.2600)	67.8343 (61.6944)	72.6487 (48.1675)	68.6269 (222.9034)	65.8326 (14.5745)	67.6936 (55.0027)	72.5584 (42.9749)
BROKEN		-39.3721 (15.1692)	-40.9110 (13.9647)	-45.8312 (43.4235)		-39.2975 (18.9874)	-40.8381 (18.5294)
POS		1.9572 (0.7003)	1.9924 (0.7031)	2.1551 (4.6552)		1.9562 (0.6954)	1.9924 (0.6761)
NEG		-3.2092 (0.7145)	-3.1656 (0.7192)	-2.1885 (4.3013)		-3.2107 (0.7354)	-3.1688 (0.7455)
New		59.8633 (36.4966)	61.5885 (27.7271)	74.0562 (167.2685)		59.8172 (34.1179)	61.5575 (29.0913)
Used		-1.0429 (10.3430)				-1.0479 (11.7372)	
Refurbished		8.9332 (19.7835)				8.9111 (30.8950)	
Silver		5.9684 (21.7494)				6.0242 (17.2327)	
Blue		-4.0321 (20.4925)				-4.0017 (19.5332)	
Pink		6.6170 (27.6143)				6.6749 (21.9380)	
Green		15.7220 (28.2819)				15.7465 (26.5944)	
# GB		0.9592 (0.1522)	0.8072 (0.1738)	0.7739 (0.1695)		0.9547 (0.2544)	0.8027 (0.2595)
<i>On σ</i>							
Constant	55.0320 (27.4887)	34.2180 (24.0819)	33.8169 (24.1540)	39.7211 (337.1709)	55.1977 (20.9938)	34.3841 (17.0459)	33.9580 (17.4974)
BROKEN				12.4026 (99.6003)			
POS				-0.1737 (7.2029)			
NEG				-1.4749 (6.7277)			
New				-14.0496 (174.1814)			

Econometria espacial

Caso simples de **simultaneidade espacial** com duas regiões:

$n = 2$ equações

$$y_1 = \alpha_{12}y_2 + X_1\beta + \varepsilon_1 \quad \varepsilon_1 \sim N(0, \sigma^2)$$

$$y_2 = \alpha_{21}y_1 + X_2\beta + \varepsilon_2 \quad \varepsilon_2 \sim N(0, \sigma^2)$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & \alpha_{12} \\ \alpha_{21} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \rho \begin{pmatrix} 0 & \alpha_{12}/\rho \\ \alpha_{21}/\rho & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \rho \begin{pmatrix} 0 & w_{12} \\ w_{21} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

- Caso geral - n regiões (equações):

$$y = \rho W y + X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I)$$

em que ρ é o coeficiente de dependência espacial (defasagem auto-regressiva) e W é a matriz de peso espacial.

- Observação:

$$w_{ij} = 0 \text{ se } i = j$$

$w_{ij} \neq 0$ se $i \neq j$: existe relação de vizinhança entre i e j .

- Arminger, G., Clogg, C. C., and Sobel, M. E. (2013). *Handbook of statistical modeling for the social and behavioral sciences*. Springer Science & Business Media.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.