

Econometria 1

Prof. Francis Petterini (PPGEco/UFSC)

f.petterini@ufsc.br

Notas de aula atualizadas em: Abril de 2026

Introdução

Estas notas de aula são baseadas essencialmente em [Wooldridge \(2010\)](#), separando o curso em duas partes, e uma avaliação para cada parte, denominadas de P1 e P2, respectivamente.

A P1 versa sobre os capítulos 1 ao 10 do livro.

A P2 versa sobre os capítulos 13 ao 19 do livro.

Os códigos dos exercícios práticos estão em formato de Stata.

Sumário por característica da variável dependente

1. P1 (alfabetização)

1.1 LGN e TLC

1.2 Modelos lineares: $y \in \mathbb{R}$

OLS, WLS, GLS

Endogeneidade e IV

Síntese para a P1 de 2026 Síntese para a P1 de 2026 (até o Cap. 7, não até o Cap. 10)

Painel sem dinâmica

Painel com dinâmica

2. P2 (métodos numéricos, máxima verossimilhança e modelos não lineares)

2.1 $y \in \{0, 1\}$

2.2 $y \geq$ mínimo

2.3 mínimo $\leq y \leq$ máximo

2.4 $y \in \mathbb{N}$

2.5 $y \in \{a, b, c\}$

2.6 $y_1 < y_2 < y_3$

2.7 y em fronteiras

P1

```
cls
clear all
set seed 54321
set obs 1000
g d1 = ceil(runiform(0,6))
g d2 = ceil(runiform(0,6))
g s = d1+d2
g n = _n
g cum_amount = sum(s)
g cum_mean = cum_amount/n
scatter cum_mean n
```

```
cls
clear all
set seed 54321
set obs 1000
g x = rpoisson(2)
*g x = rbeta(1/2,1/2)
g n = _n
g cum_amount = sum(x)
g cum_mean = cum_amount/n
scatter cum_mean n
```

```
cls
clear all
set seed 54321
set obs 10000
sca a = 2
sca b = 4
sca m = 500
forv i = 1/`=m' {
  g x`i' = a+(b-a)*runiform()
}
egen y = rowmean(x*)
g y_bar = (y-(a+b)/2)/(sqrt((b-a)^2/12)/sqrt(m))
kdensity y_bar
```

Modelos lineares

$$y \in \mathbb{R}$$

- $\{(y_{it}, y_{i0}, x_{it}, d_i)\}_{i=1, t=1}^{n, T}$, em que $y \in \mathbb{R}$, $x \in \mathbb{R}$ e $d \in \{0, 1\}$

$$\left\{ \begin{array}{lcl} y_{11} & = & \gamma y_{10} + \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + c_1 + \varepsilon_{11} \\ & \vdots & \\ y_{1T} & = & \gamma y_{1(T-1)} + \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + c_1 + \varepsilon_{1T} \\ & \vdots & \\ y_{n1} & = & \gamma y_{n0} + \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + c_n + \varepsilon_{n1} \\ & \vdots & \\ y_{nT} & = & \gamma y_{n(T-1)} + \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + c_n + \varepsilon_{nT} \end{array} \right.$$

- c_i é uma idiossincrasia constante em i entre $1 \leq t \leq T$
- ε_{it} é um ruído aleatório

$$\gamma = 0 \text{ e } T = 1$$

$$\varepsilon_{it} = c_i + \varepsilon_{it}$$

$$\begin{cases} y_1 &= \beta_1 + \beta_2 x_1 + \beta_3 d_1 + u_1 \\ y_2 &= \beta_1 + \beta_2 x_2 + \beta_3 d_2 + u_2 \\ &\vdots \\ y_n &= \beta_1 + \beta_2 x_n + \beta_3 d_n + u_n \end{cases} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & d_1 \\ 1 & x_2 & d_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & d_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$Y = X\beta + U \Rightarrow \hat{\beta}_{ols} = \operatorname{argmin}_{\beta} U'U = (X'X)^{-1}(X'Y)$$

- ▶ Se $\nexists$ correlação alguma entre quaisquer elementos de X e U , e a média do erro é zero, ocorre $E(X'U) = \mathbf{0}$

- ▶ O equivalente amostral de $E(X'U)$ é $n^{-1}X'U = \begin{bmatrix} \sum_{i=1}^n \frac{u_i}{n} \\ \sum_{i=1}^n \frac{x_i u_i}{n} \\ \sum_{i=1}^n \frac{d_i u_i}{n} \end{bmatrix}$

- ▶ $Y = X\beta + U \Rightarrow \textcolor{red}{X}'Y = \textcolor{red}{X}'X\beta + \textcolor{red}{X}'U \Rightarrow \textcolor{red}{X}'Y = \textcolor{red}{X}'X\beta + \textcolor{blue}{n}/\textcolor{blue}{n}\textcolor{red}{X}'U$

- ▶ $X'Y \stackrel{n \rightarrow \infty}{=} X'X\hat{\beta}_{mm} + \mathbf{0} \Rightarrow (X'X)^{-1}X'Y = (X'X)^{-1}X'X\hat{\beta}_{mm} \Rightarrow \hat{\beta}_{mm} = \hat{\beta}_{ols}$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n d_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i d_i \\ \sum_{i=1}^n d_i & \sum_{i=1}^n d_i x_i & \sum_{i=1}^n d_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n d_i y_i \end{bmatrix}$$

Base de datos hipotética

```
clear all
set seed 54321
set obs 100
sca b1 = -1
sca b2 = .5
sca b3 = .8
g x = rpoisson(9)
g c = rlogistic(.2,.9)
g z = rnormal(.9,.2)
g d = .4*c+.5*z+rnormal(0,1) > 0
g u = c+rnormal(0,1.2)
g y = round(b1+b2*x+b3*d+u,.01)
format y %9.2f
list y x d
```

Rotina padrão

```
g cte = 1
mkmat x d cte, matrix(X)
mkmat y, matrix(Y)

mat b_ols = inv(X'*X)*(X'*Y)
mat l b_ols

mat Y_hat = X*b_ols
mat u_hat = Y-Y_hat
mat SQR = u_hat'*u_hat
sca SQR = SQR[1,1]
sca s2 = SQR/(rowsof(X)-3)
mat V_beta = s2*inv(X'*X)
mat d_V_beta = vecdiag(V_beta)'
```

```
mat ep = J(3, 1, .)
mat t = J(3, 1, .)
forv i = 1/3 {
    mat ep[`i',1] = d_V_beta[`i', 1]^1.5
    mat t[`i',1] = b_ols[`i',1]/ep[`i',1]
}
sum y
sca y_m = r(mean)
mat y_c = Y-y_m*J(100,1,1)
mat SQT = y_c'*y_c
sca SQT = SQT[1,1]
sca R2 = 1-SQR/SQT

reg y x d
```

Bootstrap como alternativa não paramétrica para “ep”

```
reg y x d, vce(bootstrap, dots(1))
```

```
cap program drop ols_boot
pro de ols_boot, rclass
preserve
bsample
mkmat cte x d, matrix(Xb)
mkmat y, matrix(Yb)
mat b_boot = inv(Xb'*Xb)*(Xb'*Yb)
ret sca b_1 = b_boot[1,1]
ret sca b_2 = b_boot[2,1]
ret sca b_3 = b_boot[3,1]
restore
end
```

```
cap ols_boot
```

```
di b_1
```

```
loc sim 50
```

```
mat results = J(`sim',3,.)
```

```
forv i = 1/`sim' {
    cap ols_boot
    mat results[`i', 1] = r(b_1)
    mat results[`i', 2] = r(b_2)
    mat results[`i', 3] = r(b_3)
}
```

```
mat list results
```

```
forv j = 1/3 {
    g b_`j' = .
    forv i = 1/`sim' {
        replace b_`j' = results[`i', `j'] in `i'
    }
}
```

```
sum b_*
```

Endogeneidade

- ▶ Um estimador $\hat{\beta}$ é dito “consistente” para β se $\lim_{n \rightarrow \infty} \Pr(\|\hat{\beta} - \beta\| < \epsilon) = 1$. Nesse caso se escreve $\text{plim } \hat{\beta} = \beta$.
- ▶ Com um pouco de álgebra, temos que $\hat{\beta}_{ols} = \beta + (X'X)^{-1}(X'U) \Rightarrow \hat{\beta}_{ols} - \beta = (n^{-1}(X'X)^{-1})(n^{-1}(X'U))$.
- ▶ Logo, $\text{plim } \hat{\beta}_{ols} = \beta \iff E(X'U) = \mathbf{0}$.
- ▶ Se $E(X'U) \neq \mathbf{0}$, dizemos que há **endogeneidade** no modelo.
- ▶ As variáveis que causam $E(X'U) \neq \mathbf{0}$ são ditas “endógenas”.

Razões da endogeneidade

- ▶ **Variáveis Omitidas...** em um modelo que busca explicar os salários com base na educação, omitir a variável “habilidade inata” (que afeta tanto o salário quanto o nível de educação) causaria endogeneidade.
- ▶ **Causalidade Reversa...** em um modelo onde se tenta prever o consumo com base na renda, o consumo pode influenciar a renda, pois uma pessoa que consome mais bens de luxo pode atrair oportunidades de negócios.
- ▶ **Erros de Mensuração...** em um modelo de leilão onde se deseja entender a relação entre uma característica de um item (popularidade) e seu valor real. Suponha que temos dados observacionais sobre o preço final de itens vendidos em leilões, mas o preço não é uma medida perfeita do valor real. O preço, neste contexto, pode ser influenciado por fatores temporários ou subjetivos, que introduzem erros de mensuração na variável que representaria o valor.

Experimento para mostrar inconsistência

```
clear all

set seed 54321
set obs 1000

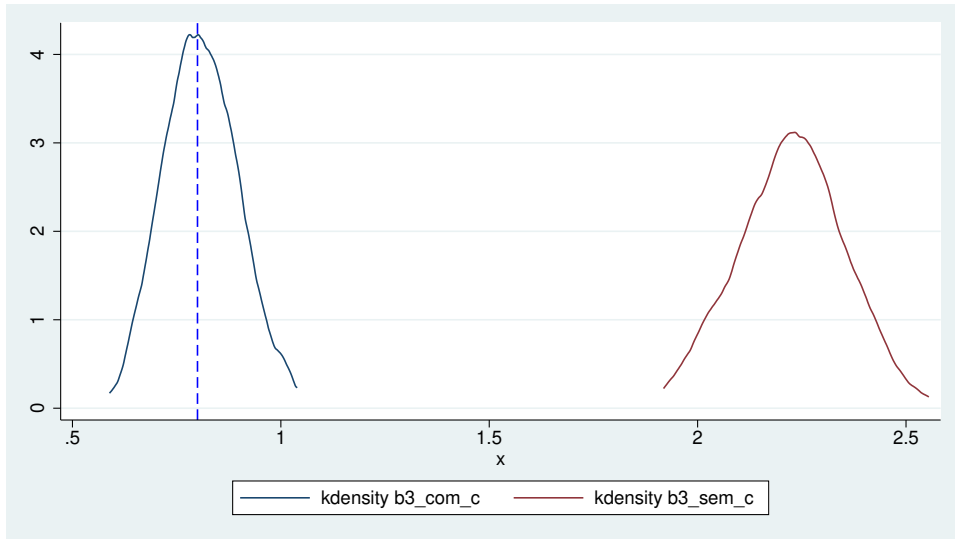
sca b1 = -1
sca b2 = .5
sca b3 = .8

g x = rpoisson(9)
g c = rlogistic(.2,.9)
g z = rnormal(.9,.2)
g d = .4*c+.5*z+rnormal(0,1)>0
g u = c+rnormal(0,1.2)
g y = b1+b2*x+b3*d+u

g b3_com_c = .
g b3_sem_c = .
```

```
forval j = 1/200 {
  set seed `=54321`j'
  replace x = rpoisson(9)
  replace c = rlogistic(.2,.9)
  replace z = rnormal(.9,.2)
  replace d = .4*c+.5*z+rnormal(0,1) > 0
  replace u = c+rnormal(0,1.2)
  replace y = b1+b2*x+b3*d+u
  reg y x d c
  mat b_com_c = e(b)
  replace b3_com_c = b_com_c[1,2] in `j'
  reg y x d
  mat b_sem_c = e(b)
  replace b3_sem_c = b_sem_c[1,2] in `j'
}

twoway (kdensity b3_com_c) (kdensity b3_sem_c), ///
       xline(.8, lpattern(dash) lcolor(blue)) xsize(16) ysize(9)
```



A solução clássica é uma variável instrumental (IV) z que atende duas condições (suponha d sendo a única variável endógena):

1. **Relevância:** $\text{Cov}(z, d) \neq 0$.
2. **Exogeneidade:** $\text{Cov}(z, u) = 0$.

$$Z = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix} \Rightarrow \hat{\beta}_{iv} = (Z'X)^{-1}(Z'U) = \beta + (Z'X)^{-1}(Z'U)$$

- ▶ Mais de um instrumento $\Rightarrow$ 2SLS.
- ▶ Desvantagens de IV:
 - ▶ É impossível checar $\text{Cov}(z, u) = 0$, exigindo suposições teóricas.
 - ▶ Um instrumento “fraco” pode causar mais viés que OLS.

Exemplos de instrumentos fracos

$$\text{plim}(\hat{\beta}_{ols}) = \beta_1 + \frac{\text{cov}(X, u)}{\text{var}(X)} \text{ e } \text{plim}(\hat{\beta}_{iv}) = \beta_1 + \frac{\text{cov}(Z, u)}{\text{cov}(Z, X)}$$

► Exemplo 1

- $\beta_1 = 2$, $\text{cov}(X, u) = 1$, $\text{var}(X) = 4$
- $\text{cov}(Z, u) = 0.5$, $\text{cov}(Z, X) = 0.2$
- **Viés OLS:** 0.25, **Viés IV:** 2.5

► Exemplo 2

- $\beta_1 = 1.5$, $\text{cov}(X, u) = 0.8$, $\text{var}(X) = 5$
- $\text{cov}(Z, u) = 2$, $\text{cov}(Z, X) = 1$
- **Viés OLS:** 0.16, **Viés IV:** 2.0

- **Conclusão:** IV pode ser pior que OLS quando o instrumento é fraco.

```
mkmat cte x d, matrix(X)
mat l X
mkmat y, matrix(Y)
mat b_ols = inv(X'*X)*(X'*Y)
mat l b_ols
mkmat cte x z, matrix(Z)
mat b_iv = inv(Z'*X)*(Z'*Y)
mat l b_iv

reg y x d
est sto ols

ivregress 2sls y x (d = z)
est sto iv

estout ols iv, cells(b(star fmt(3)) se(par fmt(3)))
```

Principais testes com IV

Teste de Hausman

- ▶ Avaliar se as variáveis suspeitas de serem endógenas realmente o são.
- ▶ O teste compara as estimativas 2SLS com OLS.

Teste de Sargan

- ▶ Para selecionar instrumentos, quando há muitas possibilidades.

Teste de Stock-Yogo

- ▶ Para verificar a fraqueza dos instrumentos.
- ▶ Explora uma estatística F modificada do 1º estágio.
 - ▶ Valores de teste surgem de um viés tolerável em relação ao viés do OLS.
 - ▶ Por exemplo, se o F do 1º estágio é 3.65 e o valor crítico para uma tolerância de 25% é 5.53, considera-se o instrumento fraco.

Síntese para a P1 de 2026
(até o Cap. 7, não até o Cap. 10)

Parte I: modelo de salários e escolaridade

A equação estrutural de interesse é:

$$\ln(w_i) = \beta_1 + \beta_2 \ln(x_i) + \beta_3 s_i + \beta_4 D_i + \beta_5 (D_i \ln x_i) + u_i.$$

- ▶ w_i : salário.
- ▶ x_i : experiência ou variável associada à experiência.
- ▶ s_i : escolaridade (anos de estudo).
- ▶ D_i : indicador de grupo.
- ▶ $D_i \ln(x_i)$: permite retorno diferente de experiência entre grupos.

Parâmetro central: β_3 , o retorno da escolaridade. Usaremos s_i para evitar confusão com a notação usual Z para matriz de instrumentos.

Por que s_i pode ser endógena?

Endogeneidade significa:

$$\text{Cov}(s_i, u_i) \neq 0 \quad \text{ou} \quad E(u_i \mid s_i, \text{controles}) \neq 0.$$

Possíveis fontes no problema:

- ▶ habilidade não observada;
- ▶ motivação;
- ▶ background familiar;
- ▶ qualidade da escola ou redes sociais não medidas.

Se essas variáveis afetam salários e também escolaridade, elas entram em u_i e ficam correlacionadas com s_i .

Consequência para MQO

No caso simples, o viés assintótico do MQO pode ser escrito como:

$$\text{plim } \hat{\beta}_3^{MQO} = \beta_3 + \frac{\text{Cov}(s_i, u_i)}{\text{Var}(s_i)}.$$

Com controles, a lógica é a mesma, mas usando a parte de s_i não explicada pelos demais regressores.

Questão de consistência lógica

Se escolaridade está positivamente correlacionada com habilidade omitida e habilidade aumenta salários, então $\text{Cov}(s_i, u_i) > 0$ e o MQO tende a **superestimar** β_3 .

Ideia de Proxy (da habilidade)

A proxy P_i pode reduzir viés de variável omitida se capturar parte da habilidade não observada. Se incluirmos P_i , a ideia é reduzir a correlação entre escolaridade e o erro:

$$\text{Cov}(s_i, u_i \mid P_i, \text{controles}) \approx 0.$$

Mas isso pode falhar porque a proxy pode ser fraca:

- ▶ a proxy mede habilidade com erro;
- ▶ habilidade é multidimensional;
- ▶ motivação e background familiar podem continuar omitidos;
- ▶ a proxy pode ser consequência de fatores já ligados à escolaridade e ao salário.

Ideia de Variáveis Instrumentais (da escolaridade)

Queremos uma fonte de variação em s_i que seja:

1. correlacionada com escolaridade;
2. não correlacionada com determinantes não observados dos salários.

Os instrumentos propostos são:

- ▶ Q_i : exposição a reforma educacional que elevou a escolaridade obrigatória;
- ▶ R_i : proximidade geográfica, na adolescência, de uma instituição pública de ensino médio.

Proxy versus IV/2SLS

Proxy P_i	IV/2SLS com Q_i, R_i
Habilidade	Escolaridade
Controla diretamente parte do fator omitido.	Usa variação exógena em escolaridade.
Depende de boa mensuração da habilidade.	Depende de relevância e exclusão dos instrumentos.
Pode reduzir viés, mas não necessariamente eliminar.	Pode gerar consistência mesmo sem observar habilidade.
Mantém interpretação mais próxima do MQO condicionado.	Pode identificar efeito local para os afetados pelos instrumentos.

Condições fundamentais de validade de instrumentos

Suponha dois candidatos a instrumentos... para Q_i e R_i serem instrumentos válidos:

1. Relevância

$$\text{Cov}(Q_i, s_i \mid X_i) \neq 0, \quad \text{Cov}(R_i, s_i \mid X_i) \neq 0.$$

Os instrumentos devem explicar escolaridade após controlar por $\ln(x_i)$, D_i e $D_i \ln(x_i)$.

2. Exogeneidade / restrição de exclusão

$$\text{Cov}(Q_i, u_i) = 0, \quad \text{Cov}(R_i, u_i) = 0.$$

Eles só podem afetar salários por meio da escolaridade, não diretamente nem por canais omitidos.

Notação matricial para IV/2SLS

Escreva o modelo estrutural como:

$$y = X\beta + u,$$

em que:

$$X = \begin{bmatrix} W & s \end{bmatrix}.$$

Aqui, W contém os regressores exógenos incluídos na equação estrutural:

$$W = \begin{bmatrix} \mathbf{1} & \ln(x) & D & D \ln(x) \end{bmatrix},$$

e s é a variável potencialmente endógena, escolaridade. A matriz de instrumentos é:

$$Z = \begin{bmatrix} W & Q & R \end{bmatrix}.$$

Ou seja, os instrumentos incluem os controles exógenos e os instrumentos excluídos Q e R .

Demonstração: 2SLS como MQO sobre variáveis projetadas

No primeiro estágio matricial, projetamos os regressores em Z :

$$\hat{X} = P_Z X.$$

No segundo estágio, rodamos MQO de y em $\hat{X}$:

$$\hat{\beta}_{2SLS} = (\hat{X}' \hat{X})^{-1} \hat{X}' y.$$

Substituindo $\hat{X} = P_Z X$:

$$\hat{\beta}_{2SLS} = (X' P_Z' P_Z X)^{-1} X' P_Z' y.$$

Como P_Z é simétrica e idempotente, isto é, $P_Z' = P_Z$ e $P_Z P_Z = P_Z$, então:

$$\hat{\beta}_{2SLS} = (X' P_Z X)^{-1} X' P_Z y.$$

Conclusão formal: IV e 2SLS coincidem

Da demonstração anterior:

$$\hat{\beta}_{2SLS} = (X'P_ZX)^{-1}X'P_Zy.$$

Essa é exatamente a forma matricial do estimador IV quando há mais instrumentos do que variáveis endógenas.

Caso exatamente identificado

Se o número de instrumentos é igual ao número de regressores, a expressão se reduz a:

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y.$$

Portanto, o 2SLS não é um estimador diferente do IV: ele é a forma operacional de calcular IV por projeções lineares.

Como funciona o 2SLS neste problema

1. Estimar o primeiro estágio:

$$s_i = \pi_1 + \pi_2 \ln(x_i) + \pi_3 D_i + \pi_4 (D_i \ln x_i) + \pi_5 Q_i + \pi_6 R_i + v_i.$$

2. Obter o componente previsto $\hat{s}_i$, isto é, a parte da escolaridade explicada pelos controles e pelos instrumentos.
3. Substituir s_i por $\hat{s}_i$ na equação de salários:

$$\ln(w_i) = \beta_1 + \beta_2 \ln(x_i) + \beta_3 \hat{s}_i + \beta_4 D_i + \beta_5 (D_i \ln x_i) + \varepsilon_i.$$

Primeiro estágio

O primeiro estágio estimado foi:

$$\hat{s}_i = 8,20 + 0,35 \ln(x_i) - 0,25D_i + 0,04(D_i \ln x_i) + 0,90Q_i + 0,15R_i.$$

Instrumento	Coeficiente	Erro-padrão	Estatística t
Q_i	0,90	0,28	3,21
R_i	0,15	0,06	2,50

Ambos têm associação positiva com escolaridade e são individualmente relevantes em níveis convencionais.

Interpretação dos instrumentos no primeiro estágio

- ▶ $\hat{\pi}_5 = 0,90$: mantidos os controles constantes, exposição à reforma está associada a 0,90 ano adicional de escolaridade, em média.
- ▶ $\hat{\pi}_6 = 0,15$: um aumento de uma unidade em proximidade geográfica está associado a 0,15 ano adicional de escolaridade, em média.

As estatísticas t são:

$$t_Q = \frac{0,90}{0,28} = 3,21, \quad t_R = \frac{0,15}{0,06} = 2,50.$$

Instrumentos fracos

Instrumentos fracos são instrumentos pouco correlacionados com a variável endógena. Consequências:

- ▶ estimador IV pode ter grande viés em amostras finitas;
- ▶ erros-padrão podem ser pouco informativos;
- ▶ testes de hipótese podem ter tamanho distorcido;
- ▶ intervalos de confiança podem ser enganosos.

Por isso, a relevância não deve ser avaliada apenas instrumento por instrumento: importa também a força conjunta dos instrumentos excluídos.

Teste F do primeiro estágio – Teste de Stock-Yogo

O teste conjunto dos instrumentos excluídos avalia:

$$H_0 : \pi_5 = \pi_6 = 0.$$

No enunciado:

$$F = 9,40.$$

Regra prática usual

Um critério informal muito usado é: $F > 10$ sugere que os instrumentos não são fracos. Aqui, $F = 9,40$ fica ligeiramente abaixo desse corte.

Stock-Yogo

Valores críticos de Stock-Yogo refinam essa análise, pois dependem do número de instrumentos, número de variáveis endógenas e tolerância ao viés/tamanho do teste.

Segundo estágio e semielasticidade

Segundo estágio:

$$\widehat{\ln(w_i)} = 0,95 + 0,44 \ln(x_i) + 0,11\hat{s}_i - 0,09D_i + 0,07(D_i \ln x_i).$$

O coeficiente de escolaridade é:

$$\hat{\beta}_3^{2SLS} = 0,11.$$

Leitura econômica

Como a variável dependente está em log e s_i está em nível, 0,11 é uma semielasticidade: um ano adicional de escolaridade aumenta o salário esperado em aproximadamente 11%, mantendo os demais controles constantes.

MQO versus 2SLS

Estimador	Coefficiente de escolaridade	Erro-padrão
MQO	0,08	0,02
2SLS	0,11	0,05

O coeficiente IV é maior que o coeficiente MQO.

Interpretação possível

Isso sugere que o MQO pode estar subestimando o retorno da escolaridade, talvez por erro de medida em escolaridade ou por heterogeneidade de retornos capturada pelos instrumentos.

A conclusão exige cautela porque IV estima um efeito local associado à variação induzida pelos instrumentos.

Teste de Hausman

O teste de Hausman compara MQO e IV:

H_0 : MQO é consistente.

Resultado:

$$H = 5,80, \quad p = 0,016.$$

Conclusão

Rejeita-se H_0 em nível de 5%. Há evidência contra a exogeneidade de s_i e a favor do uso de IV/2SLS.

Teste de Sargan

Como há dois instrumentos excluídos para uma variável endógena, o modelo é sobreidentificado. O teste de Sargan avalia a compatibilidade conjunta dos instrumentos com a restrição de exclusão:

H_0 : instrumentos são válidos, isto é, exógenos.

Resultado:

$$J = 1,25, \quad p = 0,264.$$

Conclusão

Não se rejeita H_0 . O teste não encontra evidência estatística contra a validade conjunta dos instrumentos.

Atenção: não rejeitar não prova validade; apenas não detecta violação.

Parte II: heterocedasticidade e WLS

Modelo:

$$y_i = \beta_1 + \beta_2 x_i + u_i, \quad i = 1, \dots, 4.$$

Amostra:

i	1	2	3	4
x_i	-1	1	-2	2
y_i	1	3	-1	7

Heterocedasticidade conhecida:

$$\text{Var}(u_i \mid x_i) = \sigma^2 h_i, \quad h_1 = h_2 = 1, \quad h_3 = h_4 = 4.$$

MQO: fórmulas

Com intercepto, os estimadores de MQO são:

$$\hat{\beta}_2 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \quad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}.$$

Neste exemplo:

$$\bar{x} = 0, \quad \bar{y} = \frac{1 + 3 - 1 + 7}{4} = 2,5.$$

Como $\bar{x} = 0$:

$$\hat{\beta}_2 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

MQO: cálculo

$$\sum_i x_i y_i = (-1)(1) + (1)(3) + (-2)(-1) + (2)(7) = 18.$$

$$\sum_i x_i^2 = 1 + 1 + 4 + 4 = 10.$$

Logo:

$$\hat{\beta}_2 = \frac{18}{10} = 1,8, \quad \hat{\beta}_1 = 2,5 - 1,8(0) = 2,5.$$

Reta ajustada por MQO

$$\hat{y}_i = 2,5 + 1,8x_i.$$

Por que usar WLS?

Sob heterocedasticidade conhecida:

$$\text{Var}(u_i \mid x_i) = \sigma^2 h_i.$$

Observações com maior variância são menos precisas. O WLS minimiza:

$$\sum_i w_i (y_i - \beta_1 - \beta_2 x_i)^2, \quad w_i = \frac{1}{h_i}.$$

Neste exemplo:

$$w_1 = w_2 = 1, \quad w_3 = w_4 = \frac{1}{4}.$$

WLS: médias ponderadas

A soma dos pesos é:

$$\sum_i w_i = 1 + 1 + \frac{1}{4} + \frac{1}{4} = 2,5.$$

A média ponderada de x é:

$$\bar{x}_w = \frac{\sum_i w_i x_i}{\sum_i w_i} = \frac{-1 + 1 - \frac{1}{2} + \frac{1}{2}}{2,5} = 0.$$

A média ponderada de y é:

$$\bar{y}_w = \frac{\sum_i w_i y_i}{\sum_i w_i} = \frac{1 + 3 - \frac{1}{4} + \frac{7}{4}}{2,5} = \frac{5,5}{2,5} = 2,2.$$

WLS: coeficientes

Como $\bar{x}_w = 0$:

$$\hat{\beta}_2^{WLS} = \frac{\sum_i w_i x_i y_i}{\sum_i w_i x_i^2}.$$

Calculando:

$$\sum_i w_i x_i y_i = -1 + 3 + \frac{1}{2} + \frac{14}{4} = 6,$$

$$\sum_i w_i x_i^2 = 1 + 1 + 1 + 1 = 4.$$

Logo:

$$\hat{\beta}_2^{WLS} = \frac{6}{4} = 1,5, \quad \hat{\beta}_1^{WLS} = 2,2.$$

Reta ajustada por WLS

$$\hat{y}_i^{WLS} = 2,2 + 1,5x_i.$$

Comparação numérica

Estimador	Intercepto	Inclinação
MQO	2,5	1,8
WLS	2,2	1,5

- ▶ MQO trata todas as observações como igualmente precisas.
- ▶ WLS reduz a influência das observações 3 e 4 porque elas têm variância quatro vezes maior.
- ▶ Como as observações 3 e 4 têm $|x|$ maior, a mudança de peso afeta especialmente a inclinação.

Intuição dos pesos

Observações 3 e 4 têm:

$$\text{Var}(u_i \mid x_i) = 4\sigma^2,$$

quatro vezes a variância das observações 1 e 2. Por isso, recebem peso:

$$w_3 = w_4 = \frac{1}{4}.$$

Ideia central

Se uma observação é mais ruidosa, ela contém menos informação sobre a reta populacional. O WLS dá menos influência a essa observação no ajuste.

Eficiência: MQO ou WLS?

Sob as hipóteses:

$$E(u_i | x_i) = 0, \quad \text{Var}(u_i | x_i) = \sigma^2 h_i$$

e com h_i corretamente especificado, MQO continua não viesado e consistente. Mas WLS é mais eficiente, porque usa a informação sobre a estrutura da heterocedasticidade.

MQO: resíduos

A reta ajustada por MQO foi:

$$\hat{y}_i = 2,5 + 1,8x_i.$$

Logo:

i	x_i	y_i	$\hat{y}_i$	$\hat{u}_i = y_i - \hat{y}_i$
1	-1	1	0,7	0,3
2	1	3	4,3	-1,3
3	-2	-1	-1,1	0,1
4	2	7	6,1	0,9

A soma dos quadrados dos resíduos é:

$$SQR_{MQO} = 0,3^2 + (-1,3)^2 + 0,1^2 + 0,9^2 = 2,6.$$

MQO: cálculo do R^2

A média de y é:

$$\bar{y} = 2,5.$$

A soma total dos quadrados é:

$$SQT = \sum_i (y_i - \bar{y})^2.$$

Logo:

$$SQT = (1 - 2,5)^2 + (3 - 2,5)^2 + (-1 - 2,5)^2 + (7 - 2,5)^2.$$

$$SQT = 2,25 + 0,25 + 12,25 + 20,25 = 35.$$

Portanto:

$$R_{MQO}^2 = 1 - \frac{SQR_{MQO}}{SQT} = 1 - \frac{2,6}{35} = 0,9257.$$

Interpretação

Neste exemplo, o MQO explica aproximadamente 92,6% da variação amostral de y .

MQO: variância residual

Como temos $n = 4$ observações e $k = 2$ parâmetros, os graus de liberdade são:

$$n - k = 4 - 2 = 2.$$

A estimativa da variância do erro no MQO é:

$$\hat{\sigma}_{MQO}^2 = \frac{SQR_{MQO}}{n - k} = \frac{2,6}{2} = 1,3.$$

Como:

$$X'X = \begin{bmatrix} 4 & 0 \\ 0 & 10 \end{bmatrix},$$

temos:

$$(X'X)^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Logo:

$$\widehat{\text{Var}}(\hat{\beta}_{MQO}) = \hat{\sigma}_{MQO}^2 (X'X)^{-1}.$$

MQO: erros-padrão e estatísticas t

Assim:

$$\widehat{\text{Var}}(\hat{\beta}_{MQO}) = 1,3 \begin{bmatrix} 1/4 & 0 \\ 0 & 1/10 \end{bmatrix} = \begin{bmatrix} 0,325 & 0 \\ 0 & 0,13 \end{bmatrix}.$$

Os erros-padrão são:

$$se(\hat{\beta}_1) = \sqrt{0,325} = 0,570,$$

$$se(\hat{\beta}_2) = \sqrt{0,13} = 0,361.$$

As estatísticas t , para testar $H_0 : \beta_j = 0$, são:

$$t_{\hat{\beta}_1} = \frac{2,5}{0,570} = 4,386,$$

$$t_{\hat{\beta}_2} = \frac{1,8}{0,361} = 4,992.$$

Observação

Essas são as estatísticas t convencionais de MQO. Sob heterocedasticidade conhecida, o WLS é o estimador mais adequado para inferência eficiente.

WLS: resíduos

A reta ajustada por WLS foi:

$$\hat{y}_i^{WLS} = 2,2 + 1,5x_i.$$

Logo:

i	x_i	y_i	w_i	$\hat{y}_i^{WLS}$	$\hat{u}_i^{WLS}$
1	-1	1	1	0,7	0,3
2	1	3	1	3,7	-0,7
3	-2	-1	1/4	-0,8	-0,2
4	2	7	1/4	5,2	1,8

No WLS, a soma relevante é a soma ponderada dos quadrados dos resíduos:

$$SQRP_{WLS} = \sum_i w_i (\hat{u}_i^{WLS})^2.$$

WLS: soma ponderada dos resíduos

Temos:

$$SQRP_{WLS} = 1(0,3)^2 + 1(-0,7)^2 + \frac{1}{4}(-0,2)^2 + \frac{1}{4}(1,8)^2.$$

Logo:

$$SQRP_{WLS} = 0,09 + 0,49 + 0,01 + 0,81 = 1,4.$$

A variância residual ponderada é:

$$\hat{\sigma}_{WLS}^2 = \frac{SQRP_{WLS}}{n - k} = \frac{1,4}{2} = 0,7.$$

Comparação

A soma ponderada dos resíduos no WLS é menor porque o critério de minimização agora leva em conta os pesos w_i .

WLS: cálculo do R^2 ponderado

A média ponderada de y foi:

$$\bar{y}_w = 2,2.$$

A soma total ponderada dos quadrados é:

$$SQTP = \sum_i w_i (y_i - \bar{y}_w)^2.$$

Logo:

$$SQTP = 1(1 - 2,2)^2 + 1(3 - 2,2)^2 + \frac{1}{4}(-1 - 2,2)^2 + \frac{1}{4}(7 - 2,2)^2.$$

$$SQTP = 1,44 + 0,64 + 2,56 + 5,76 = 10,4.$$

Portanto:

$$R_{WLS}^2 = 1 - \frac{SQRP_{WLS}}{SQTP} = 1 - \frac{1,4}{10,4} = 0,8654.$$

WLS: matriz de variância dos estimadores

No WLS:

$$\widehat{\text{Var}}(\hat{\beta}_{WLS}) = \hat{\sigma}_{WLS}^2 (X'WX)^{-1}.$$

Neste exemplo:

$$X'WX = \begin{bmatrix} \sum_i w_i & \sum_i w_i x_i \\ \sum_i w_i x_i & \sum_i w_i x_i^2 \end{bmatrix}.$$

Como:

$$\sum_i w_i = 2,5, \quad \sum_i w_i x_i = 0, \quad \sum_i w_i x_i^2 = 4,$$

temos:

$$X'WX = \begin{bmatrix} 2,5 & 0 \\ 0 & 4 \end{bmatrix}.$$

Logo:

$$(X'WX)^{-1} = \begin{bmatrix} 0,4 & 0 \\ 0 & 0,25 \end{bmatrix}.$$

WLS: erros-padrão e estatísticas t

Como:

$$\hat{\sigma}_{WLS}^2 = 0,7,$$

temos:

$$\widehat{\text{Var}}(\hat{\beta}_{WLS}) = 0,7 \begin{bmatrix} 0,4 & 0 \\ 0 & 0,25 \end{bmatrix} = \begin{bmatrix} 0,28 & 0 \\ 0 & 0,175 \end{bmatrix}.$$

Os erros-padrão são:

$$se(\hat{\beta}_1^{WLS}) = \sqrt{0,28} = 0,529,$$

$$se(\hat{\beta}_2^{WLS}) = \sqrt{0,175} = 0,418.$$

As estatísticas t , para testar $H_0 : \beta_j = 0$, são:

$$t_{\hat{\beta}_1^{WLS}} = \frac{2,2}{0,529} = 4,158,$$

$$t_{\hat{\beta}_2^{WLS}} = \frac{1,5}{0,418} = 3,586.$$

Comparação final: MQO e WLS

Estimador	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$t_{\hat{\beta}_1}$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$t_{\hat{\beta}_2}$
MQO	2,5	0,570	4,386	1,8	0,361	4,992
WLS	2,2	0,529	4,158	1,5	0,418	3,586

Estimador	Soma de quadrados residual	Variância residual	R^2
MQO	2,6	1,3	0,9257
WLS	1,4 ponderada	0,7	0,8654 ponderado

Mensagem

Os números de MQO e WLS não são diretamente comparáveis apenas pelo R^2 , pois o WLS usa outro critério: ele minimiza a soma ponderada dos resíduos ao quadrado.

Do WLS ao GLS

Até aqui, o problema era:

$$\text{Var}(u_i \mid x_i) = \sigma^2 h_i.$$

Ou seja, cada observação podia ter uma variância diferente, mas os erros eram tratados como não correlacionados entre si.

Em forma matricial, isso equivale a:

$$\text{Var}(u \mid X) = \sigma^2 \begin{bmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_3 & 0 \\ 0 & 0 & 0 & h_4 \end{bmatrix}.$$

Ideia

O WLS é suficiente quando a matriz de variância dos erros é diagonal. O GLS aparece quando queremos permitir uma estrutura mais geral.

Motivação para GLS

O MQO é construído para o caso em que:

$$\text{Var}(u \mid X) = \sigma^2 I.$$

Isto significa duas coisas:

$$\text{Var}(u_i \mid X) = \sigma^2$$

e

$$\text{Cov}(u_i, u_j \mid X) = 0, \quad i \neq j.$$

Mas, em muitos problemas empíricos, essas duas hipóteses podem falhar.

- ▶ As variâncias podem ser diferentes entre observações.
- ▶ Os erros podem ser correlacionados entre observações.

Motivação

O GLS ajusta a estimação quando a matriz de variância-covariância dos erros não é simplesmente $\sigma^2 I$.

Exemplos em que pode haver correlação dos erros

A correlação entre erros aparece naturalmente em vários contextos.

- ▶ **Séries temporais:** choques em um período podem persistir para períodos seguintes.
- ▶ **Dados em painel:** observações da mesma firma, pessoa ou município podem compartilhar componentes não observados.
- ▶ **Dados espaciais:** municípios vizinhos podem sofrer choques parecidos.
- ▶ **Dados agrupados:** estudantes da mesma escola ou trabalhadores da mesma firma podem ter fatores comuns.

Por exemplo, em uma série temporal:

$$\text{Cov}(u_t, u_{t-1}) \neq 0.$$

Problema

O MQO ignora essa dependência. O GLS usa essa informação para estimar de forma mais eficiente.

Modelo geral do GLS

Escreva o modelo em forma matricial:

$$y = X\beta + u.$$

Agora suponha:

$$E(u \mid X) = 0,$$

mas:

$$\text{Var}(u \mid X) = \sigma^2 \Omega.$$

A matriz Ω pode ter elementos diferentes na diagonal e fora da diagonal.

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{bmatrix}.$$

Estimador GLS

Se:

$$\text{Var}(u \mid X) = \sigma^2 \Omega,$$

e Ω é conhecida, o estimador GLS é:

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y.$$

Compare com o MQO:

$$\hat{\beta}_{MQO} = (X' X)^{-1} X' y.$$

Diferença

O GLS substitui a matriz identidade por Ω^{-1} . Assim, a estimação passa a levar em conta a estrutura de variância e covariância dos erros.

GLS como minimização ponderada

O MQO minimiza:

$$(y - X\beta)'(y - X\beta).$$

O GLS minimiza:

$$(y - X\beta)'\Omega^{-1}(y - X\beta).$$

Ou seja, o GLS não mede o tamanho dos resíduos apenas pela soma dos quadrados. Ele mede os resíduos levando em conta a matriz Ω^{-1} .

Intuição

O GLS penaliza mais os resíduos que vêm de observações mais precisas e penaliza menos os resíduos que vêm de observações mais ruidosas ou redundantes.

WLS como caso particular do GLS

No WLS, temos:

$$\Omega = \begin{bmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_3 & 0 \\ 0 & 0 & 0 & h_4 \end{bmatrix}.$$

Logo:

$$\Omega^{-1} = \begin{bmatrix} 1/h_1 & 0 & 0 & 0 \\ 0 & 1/h_2 & 0 & 0 \\ 0 & 0 & 1/h_3 & 0 \\ 0 & 0 & 0 & 1/h_4 \end{bmatrix}.$$

Portanto:

$$w_i = \frac{1}{h_i}.$$

Conclusão

Por que a correlação entre erros importa?

Suponha que duas observações tenham erros positivamente correlacionados:

$$\text{Cov}(u_i, u_j) > 0.$$

Isso significa que parte do erro de uma observação anda junto com o erro da outra.

Nesse caso, essas duas observações não trazem tanta informação independente quanto pareceria no MQO.

Intuição

Se duas observações carregam choques parecidos, tratá-las como totalmente independentes exagera a quantidade efetiva de informação na amostra.

O GLS corrige isso porque considera os termos de covariância dentro de Ω .

GLS como transformação do modelo

Se:

$$\text{Var}(u \mid X) = \sigma^2 \Omega,$$

podemos encontrar uma matriz P tal que:

$$P'P = \Omega^{-1}.$$

Multiplicando o modelo por P , temos:

$$Py = PX\beta + Pu.$$

Defina:

$$y^* = Py, \quad X^* = PX, \quad u^* = Pu.$$

Então:

$$y^* = X^*\beta + u^*.$$

Variância do erro transformado

Como:

$$u^* = Pu,$$

temos:

$$\text{Var}(u^* \mid X) = \text{Var}(Pu \mid X).$$

Logo:

$$\text{Var}(u^* \mid X) = P \text{Var}(u \mid X) P'.$$

Como:

$$\text{Var}(u \mid X) = \sigma^2 \Omega,$$

segue que:

$$\text{Var}(u^* \mid X) = \sigma^2 P \Omega P'.$$

Se a transformação é escolhida de modo que:

GLS factível

Na prática, raramente conhecemos exatamente Ω .

Por isso, muitas vezes usamos uma estimativa de Ω , denotada por $\hat{\Omega}$.

Nesse caso, o estimador é chamado de GLS factível:

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$

Exemplo

Se suspeitamos de autocorrelação AR(1), podemos primeiro estimar o parâmetro de autocorrelação e depois usar essa estimativa para construir $\hat{\Omega}$.

O mesmo raciocínio vale para heterocedasticidade modelada, dados agrupados ou estruturas específicas de painel.

MQO robusto ou GLS?

Há duas estratégias diferentes diante de heterocedasticidade ou correlação dos erros.

- ▶ **MQO com erros-padrão robustos:** mantém os coeficientes de MQO, mas corrige a inferência.
- ▶ **GLS ou FGLS:** muda os coeficientes estimados, usando uma estrutura para Ω .

Comparação

MQO robusto é mais defensivo: exige menos hipóteses sobre Ω . GLS é mais eficiente se a estrutura de Ω estiver bem especificada.

Mensagem

O ganho do GLS depende de termos uma boa descrição da matriz de variância-covariância dos erros.

Síntese: MQO, WLS e GLS

Estimador	Estrutura dos erros	Ideia central
MQO	$\sigma^2 I$	todos os erros têm mesma variância e não são correla
WLS	$\sigma^2 \text{diag}(h_i)$	erros com variâncias diferentes, sem correlação
GLS	$\sigma^2 \Omega$	variâncias diferentes e possíveis covariâncias entre err

Conclusão

O WLS corrige um problema específico: heterocedasticidade conhecida. O GLS é a versão geral: ele usa toda a matriz Ω para obter estimadores mais eficientes quando os erros têm uma estrutura conhecida de variância e covariância.

Painel sem dinâmica

► $\gamma = 0$ e $T > 1$

$$\left\{ \begin{array}{lcl} y_{11} & = & \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + \textcolor{red}{c}_1 + \varepsilon_{11} \\ y_{12} & = & \beta_1 + \beta_2 x_{12} + \beta_3 d_{12} + \textcolor{red}{c}_1 + \varepsilon_{12} \\ & \vdots & \\ y_{1T} & = & \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + \textcolor{red}{c}_1 + \varepsilon_{1T} \\ & \vdots & \\ y_{n1} & = & \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + \textcolor{red}{c}_n + \varepsilon_{n1} \\ y_{n2} & = & \beta_1 + \beta_2 x_{n2} + \beta_3 d_{n2} + \textcolor{red}{c}_n + \varepsilon_{n2} \\ & \vdots & \\ y_{nT} & = & \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + \textcolor{red}{c}_n + \varepsilon_{nT} \end{array} \right.$$

A abordagem da diferença da média (FE)

$$\left\{ \begin{array}{lcl} y_{11} - \bar{y}_1 & = & \beta_2(x_{11} - \bar{x}_1) + \beta_3(d_{11} - \bar{d}_1) + (\varepsilon_{11} - \bar{\varepsilon}_1) \\ y_{12} - \bar{y}_1 & = & \beta_2(x_{12} - \bar{x}_1) + \beta_3(d_{12} - \bar{d}_1) + (\varepsilon_{12} - \bar{\varepsilon}_1) \\ & \vdots & \\ y_{1T} - \bar{y}_1 & = & \beta_2(x_{1T} - \bar{x}_1) + \beta_3(d_{1T} - \bar{d}_1) + (\varepsilon_{1T} - \bar{\varepsilon}_1) \\ & \vdots & \\ y_{n1} - \bar{y}_n & = & \beta_2(x_{n1} - \bar{x}_n) + \beta_3(d_{n1} - \bar{d}_n) + (\varepsilon_{n1} - \bar{\varepsilon}_n) \\ y_{n2} - \bar{y}_n & = & \beta_2(x_{n2} - \bar{x}_n) + \beta_3(d_{n2} - \bar{d}_n) + (\varepsilon_{n2} - \bar{\varepsilon}_n) \\ & \vdots & \\ y_{nT} - \bar{y}_n & = & \beta_2(x_{nT} - \bar{x}_n) + \beta_3(d_{nT} - \bar{d}_n) + (\varepsilon_{nT} - \bar{\varepsilon}_n) \end{array} \right.$$

$$\hat{\beta}_{fe} = (\tilde{X}'\tilde{X})^{-1}(\tilde{X}'\tilde{Y}), \text{ sem constante}$$

A abordagem da primeira diferença (FD)

$$\left\{ \begin{array}{l} \Delta y_{11} = \beta_2 \Delta x_{11} + \beta_3 \Delta d_{11} + \Delta \varepsilon_{11} \\ \Delta y_{12} = \beta_2 \Delta x_{12} + \beta_3 \Delta d_{12} + \Delta \varepsilon_{12} \\ \vdots \\ \Delta y_{1T} = \beta_2 \Delta x_{1T} + \beta_3 \Delta d_{1T} + \Delta \varepsilon_{1T} \\ \vdots \\ \Delta y_{n1} = \beta_2 \Delta x_{n1} + \beta_3 \Delta d_{n1} + \Delta \varepsilon_{n1} \\ \Delta y_{n2} = \beta_2 \Delta x_{n2} + \beta_3 \Delta d_{n2} + \Delta \varepsilon_{n2} \\ \vdots \\ \Delta y_{nT} = \beta_2 \Delta x_{nT} + \beta_3 \Delta d_{nT} + \Delta \varepsilon_{nT} \end{array} \right. \Rightarrow \hat{\beta}_{fd} = (\Delta X' \Delta X)^{-1} (\Delta X' \Delta Y)$$

```

clear all
set seed 54321
loc n = 20
loc T = 40
set obs `n'
g i = _n
expand `T'
bys i: gen t = _n
sca b1 = -1
sca b2 = .5
sca b3 = .8
sort i t
g x = rpoisson(9)
bys i: g c = rlogistic(.2,.9) if _n == 1
bys i: replace c = c[_n-1] if mi(c)
g z = rnormal(.9,.2)
g d = .4*c+.5*z+rnormal(0,1)>0
g u = c+rnormal(0,1.2)
g y = .
bys i: replace y = b1+b2*x+b3*d+u

```

```

xtset i t

foreach var in y x d {
  bys i: egen `var' _mean = mean(`var')
  g `var' _til = `var' - `var' _mean
  g `var' _dif = D.`var'
}

mkmat x_til d_til, matrix(X_til)
mkmat y_til, matrix(Y_til)
mat beta_fe = inv(X_til'*X_til)*(X_til'*Y_til)
mat l beta_fe

drop if y_dif == .

mkmat x_dif d_dif, matrix(X_dif)
mkmat y_dif, matrix(Y_dif)
mat beta_fd = inv(X_dif'*X_dif)*(X_dif'*Y_dif)
mat l beta_fd

```

- ▶ Desvantagens da abordagem, independentemente de FE ou FD:
 - ▶ Só resolve a endogeneidade se o problema for relacionado com uma variável omitida que não muda com o tempo.
 - ▶ Essa ideia é estranha se T é grande.
 - ▶ Mesmo para T pequeno, a abordagem só resolve a endogeneidade se o problema que a gera é constante para todas as unidades de observação.
 - ▶ Não resolve problemas de causalidade reversa ou erro de medida.
 - ▶ O efeito de interesse não pode estar relacionado com uma constante, já que há perda de informação com variáveis constantes.
- ▶ FE/FD podem ser usados facilmente com IV.
- ▶ FD tem 2 desvantagens em relação a FE:
 - ▶ Perde n observações
 - ▶ Problemas em painel desbalanceado.

A abordagem de efeitos aleatórios (RE)

Se X não possui elementos endógenos, OLS é consistente e FE/FD são desnecessários, mas a matriz de variância-covariância do erro precisa ser ajustada:

$$\blacktriangleright u_{it} = c_i + \varepsilon_{it} \Rightarrow \sigma_u^2 = \sigma_c^2 + \sigma_\varepsilon^2 \Rightarrow UU'_{n=3, T=2} = \begin{bmatrix} 2.9 & 0.4 & 0 & 0 & 0 & 0 \\ 0.4 & 2.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.9 & 0.4 & 0 & 0 \\ 0 & 0 & 0.4 & 2.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.9 & 0.4 \\ 0 & 0 & 0 & 0 & 0.4 & 2.9 \end{bmatrix}$$

- ▶ É o caso de se usar $\hat{\beta}_{gls} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$.
- ▶ $\hat{\beta}_{re} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$.
 - ▶ Em que $\tilde{y}_{it} = y_{it} - \theta\bar{y}_i$ para $\theta = 1 - \sigma_\varepsilon/\sqrt{T\sigma_c^2 + \sigma_\varepsilon^2}$.
 - ▶ Se X não possui elementos endógenos, este GLS seria eficiente em detrimento ao OLS e ao FE, embora todos sejam consistentes.
 - ▶ O procedimento se dá em 2 estágios.
 - ▶ RE subsidia o teste de Hausman.

```
sort i t
xtset i t
reg y x d
est sto ols
xtreg y x d, pa corr(independent)
est sto pols
xtreg y x d, fe
est sto fe
reg D.y D.x D.d, nocon
est sto fd
xtivreg y x (d = z), fe
est sto ivfe
xtreg y x d, re
est sto re
estout ols pols fe fd re ivfe, cells(b(star fmt(3)) se(par fmt(3)))
hausman fe re, sigmamore
```

	ols b/se	pols b/se	fe b/se	fd b/se	re b/se	ivfe b/se
x	0.559*** (0.030)	0.559*** (0.030)	0.523*** (0.015)		0.524*** (0.016)	0.520*** (0.018)
d	3.548*** (0.191)	3.548*** (0.191)	0.808*** (0.117)		0.909*** (0.120)	0.101 (1.731)
D.x				0.514*** (0.015)		
D.d				0.735*** (0.112)		
_cons	-2.880*** (0.314)	-2.880*** (0.313)	-0.658*** (0.166)		-0.738* (0.333)	-0.136 (1.286)

---- Coefficients ----

	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	fe	re	Difference	Std. err.
-----+-----				
x	.5233016		.5244211	-.0011196.0004307
d	.8082739		.9092188	-.1009449.0134717

b = Consistent under H0 and Ha; obtained from xtreg.

B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

$$\begin{aligned}\text{chi2}(2) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 57.43\end{aligned}$$

Prob > chi2 = 0.0000

O p-valor 0 indica que se deve rejeitar a hipótese nula, o que significa que FE é preferível.

Acrescentando a dinâmica

Raíz unitária... só faz sentido painel dinâmico se $|\gamma| < 1$

$$y_{it} = \gamma y_{it-1} + c_i + \varepsilon_{it} \Rightarrow y_{it} = \gamma^t y_{i0} + \frac{1 - \gamma^t}{1 - \gamma} c_i + \sum_{j=0}^{t-1} \gamma^j \varepsilon_{it-j}$$

A lógica do ADF segue sendo usado (e aqui a questão do balanceamento do painel importa, por conta da FD): $\Delta y_{it} = \rho y_{it-1} + \sum_{j=1}^p \alpha_j \Delta y_{it-j} + \text{ruído}_{it}$

Testes populares	H0: $\rho = 0$	H1
LLC	Raiz unitária $\forall i$	Estacionariedade $\forall i$
IPS	Raiz unitária $\forall i$	Estacionariedade $\nexists i$
Fisher	Raiz unitária $\forall i$	Estacionariedade $\nexists i$

O p-valor pequeno rejeita H0 de que as séries contêm uma raiz unitária.

```

clear all
set seed 54321
loc n = 1000
loc T = 20
set obs `n'
g i = _n
expand `T'
bys i: gen t = _n
sca gamma = .6
sca b1 = -1
sca b2 = .5
sca b3 = .8
g x = rpoisson(9)

```

```

sort i t
bys i: g c = rlogistic(.2,.9) if _n == 1
bys i: replace c = c[_n-1] if mi(c)
g z = rnormal(.9,.2)
g d = .4*c+.5*z+rnormal(0,1)>0
g u = c+rnormal(0,1.2)
g y = .
bys i: replace y = b1+b2*x+b3*d+u if t == 1
bys i: replace y = gamma*y[_n-1]+b1+b2*x+b3*d+u if t > 1

xtset i t

* Teste Levin-Lin-Chu (LLC) com 1 defasagem
xtunitroot llc y, lags(1)

* Teste Im-Pesaran-Shin (IPS) com 2 defasagens
xtunitroot ips y, lags(2)

* Teste Fisher com 1 defasagem
xtunitroot fisher y, dfuller lags(1)

```

OLS, FE, FD etc. são inconsistentes em painel dinâmico

O procedimento de FE:

$$y_{it} - \bar{y}_i = \gamma(y_{it-1} - \bar{y}_{i,-1}) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Em que $\bar{y}_{i,-1} = (y_{i0} + y_{i1} + \dots + y_{it-1})/T$.

$$\hat{\gamma}_{fe} = \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(y_{it} - \bar{y}_i)}{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2} = \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(\gamma(y_{it-1} - \bar{y}_{i,-1}) + (\varepsilon_{it} - \bar{\varepsilon}_i))}{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2}$$

$$\text{plim } \hat{\gamma}_{fe} = \gamma + \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(\varepsilon_{it} - \bar{\varepsilon}_i)}{\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2} \neq \gamma$$

Mas $\sum_{i=1}^n \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(\varepsilon_{it} - \bar{\varepsilon}_i)/(nT)$ é a correlação entre o passado de y e o erro contemporâneo, que é parte do presente de y . Portanto, esse termo não é zero.

Anderson & Hsiao (1981)

A solução AH é aplicar FD com IV, porque (aqui a questão do balanceamento do painel importa, por conta da FD)...

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta_2 \Delta x_{it} + \beta_3 \Delta d_{it} + \Delta \varepsilon_{it}$$

- ▶ $\Delta y_{it-1} = y_{it-1} - y_{it-2}$ é correlacionado com o termo de erro $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$, por conta da contemporaneidade em $t - 1$.
- ▶ Mas y_{it-2} pode ser usado como instrumento para Δy_{it-1} , já que y_{it-2} não é correlacionado com $\Delta \varepsilon_{it}$, mas é correlacionado com Δy_{it-1} .
- ▶ Como há um único instrumento, basta aplicar apropriadamente IV.
- ▶ O problema é que y_{it-2} pode ser um instrumento fraco. – i.e., γ pode ser relativamente pequeno.

```
xtset i t
```

```
xtreg y L.y x d, fe
```

```
est sto fe
```

```
ivregress 2sls D.y D.x D.d (D.L.y = L2.y), nocon
```

```
est sto ah
```

```
estout fe ah, cells(b(star fmt(3)) se(par fmt(3)))
```

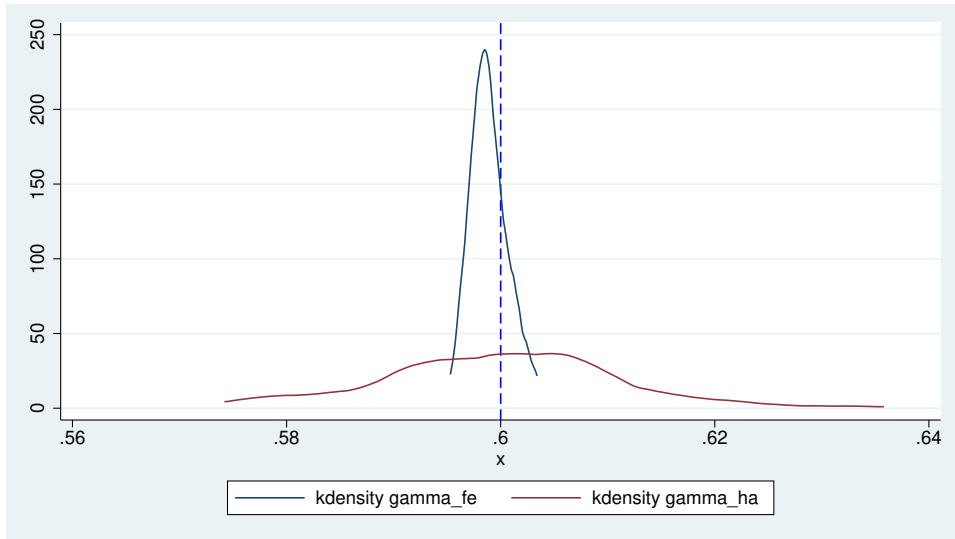
	fe	ah
	b/se	b/se
L.y	0.583*** (0.003)	
LD.y		0.666* (0.320)
x	0.502*** (0.003)	
d	0.771*** (0.021)	
D.x		0.520*** (0.081)
D.d		0.786*** (0.123)
_cons	-0.655*** (0.046)	

Exercício para mostrar que AH é consistente, mas não é eficiente

```
g gamma_fe = .
g gamma_ha = .

forval j = 1/100 {
  set seed `=54321`j'
  sort i t
  replace u = c+rnormal(0,.2)
  replace y = .
  bys i: replace y = b1+b2*x+b3*d+u if t == 1
  bys i: replace y = gamma*y[_n-1]+b1+b2*x+b3*d+u if t > 1
  xtset i t
  xtreg y L.y x d, fe
  mat fe = e(b)
  replace gamma_fe = fe[1,1] in `j'
  ivregress 2sls D.y D.x D.d (D.L.y = L2.y), nocon
  mat ha = e(b)
  replace gamma_ha = ha[1,1] in `j'
}

tw (kdensity gamma_fe) (kdensity gamma_ha), ///
xline(.6, lpattern(dash) lcolor(blue)) xsize(16) ysize(9)
```



Arellano & Bond (1991)

- ▶ Segue explorando $\Delta y_{it} = \gamma \Delta y_{it-1} + \beta_2 \Delta x_{it} + \beta_3 \Delta d_{it} + \Delta \varepsilon_{it}$, mas aplicando como instrumentos $y_{i,t-s}$ para $s \geq 2$.
- ▶ As principais vantagens: utiliza todas as defasagens válidas ($y_{i,t-2}, y_{i,t-3}, \dots$), então há menor risco de instrumentos fracos; e, permite o teste de Sargan.
- ▶ Se $E[y_{i,t-s} \Delta \varepsilon_{it}] = 0$, pode-se aplicar GMM.
- ▶ Mas agora há mais instrumentos que parâmetros a se identificar, então é preciso definir uma função objetivo $Q(\theta) = \hat{g}(\theta)' W \hat{g}(\theta)$, em que $\theta = [\gamma, \beta_2, \beta_3]$ e pode incluir outros eventuais parâmetros auxiliares.
 - ▶ $\hat{g}(\theta) = \frac{1}{n} \sum_{i=1}^n Z_i' \Delta \varepsilon_i(\theta)$ é o vetor de médias amostrais das condições de momento.
 - ▶ Z_i é um vetor contendo $y_{i,t-2}, y_{i,t-3}, \dots$ (e variáveis exógenas).
 - ▶ W é a matriz de pesos.

Exemplo GMM com $T = 4$

$$\begin{aligned}\|Q(\theta)\| &= \hat{g}(\theta)' W \hat{g}(\theta) \\&= \left\| \begin{bmatrix} \frac{\sum_{i=1}^n y_{i2} \Delta \varepsilon_{i4}}{n} \\ \frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i4}}{n} \\ \frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i3}}{n} \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix} \begin{bmatrix} \frac{\sum_{i=1}^n y_{i2} \Delta \varepsilon_{i4}}{n} \\ \frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i4}}{n} \\ \frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i3}}{n} \end{bmatrix} \right\| \\&= \sqrt{w_1 \left(\frac{\sum_{i=1}^n y_{i2} \Delta \varepsilon_{i4}}{n} \right)^2 + w_2 \left(\frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i4}}{n} \right)^2 + w_3 \left(\frac{\sum_{i=1}^n y_{i1} \Delta \varepsilon_{i3}}{n} \right)^2} \\ \Delta \varepsilon_{it} &= \Delta y_{it} - (\gamma \Delta y_{it-1} + \beta_2 \Delta x_{it} + \beta_3 \Delta d_{it}) \\ \hat{\theta} &= \operatorname{argmin}_{\theta} \|Q(\theta)\|\end{aligned}$$

- ▶ São empregados métodos numéricos para computar $\hat{\theta}$ sempre que há “superidentificação”.
- ▶ Poderíamos ter até 5 instrumentos ao considerar x e d como exógenas.

Arellano & Bover (1995)

A principal inovação é perceber que a equação pode ser tratada em nível:

$$y_{it} = \gamma y_{it-1} + \beta_1 + \beta_2 x_{it} + \beta_3 d_{it} + u_{it}$$

E explorar como instrumentos múltiplas defasagens das diferenças como instrumentos $\Delta y_{i,t-2}, \Delta y_{i,t-3} \dots$, já que essas diferenças não seriam correlacionadas com $u_{it} = c_i + \varepsilon_{it}$

A vantagem dessa abordagem é que ela não usa diferenças, então:

- ▶ se ajusta melhor para painéis desbalanceados
- ▶ permite estudar o efeito de variáveis constantes

```
clear all
set seed 54321
loc n = 100
loc T = 20
set obs `n'
g i = _n
expand `T'
bys i: gen t = _n
sca gamma = .5
sca beta = .5
g x = rpoisson(9)
sort i t
bys i: g c = rlogistic(.2,.9) if _n == 1
bys i: replace c = c[_n-1] if mi(c)
g u = c+rnormal(0,1.2)
g y = .
bys i: replace y = beta*x+u if t == 1
bys i: replace y = gamma*y[_n-1]+beta*x+u if t > 1
xtset i t
```

```
xtreg y L.y x, fe  
est sto fe
```

```
ivregress 2sls D.y D.x (D.L.y = L2.y), nocon  
est sto ah
```

```
xtabond y x, nocons lags(1) maxlags(1) pre(x)  
est sto ab
```

```
gmm (D.y - {gamma}*D.L.y - {beta}*D.x), instruments(L2.y L.x x D.x)  
est sto ab2
```

```
xtdpdsys y x, lags(1) maxlags(1)  
est sto bl
```

```
gmm (y - {gamma}*L.y - {beta}*x - {c}), instruments(L.D.y L2.D.y L2.y D.x)  
est sto bl2
```

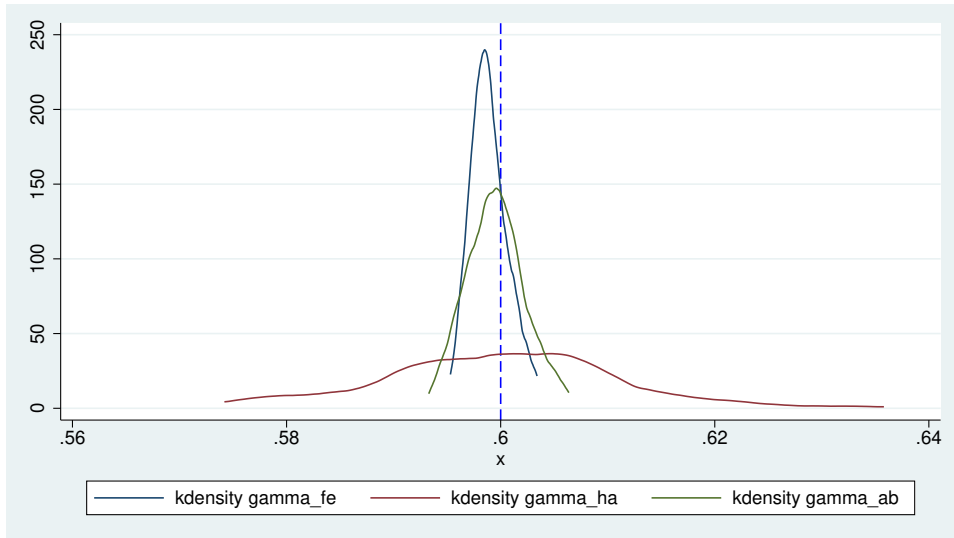
```
estout fe ah ab ab2 bl bl2, cells(b(star fmt(3)) se(par fmt(3)))
```

Ver pág. 40 de <https://www.stata.com/manuals13/rgmm.pdf>

Comparando FE, AH e AB

```
g gamma_fe = .
g gamma_ha = .
g gamma_ab = .
forval j = 1/100 {
  set seed `=54321`j''
  sort i t
  replace u = c+rnormal(0,.2)
  replace y = .
  bys i: replace y = b1+b2*x+b3*d+u if t == 1
  bys i: replace y = gamma*y[_n-1]+b1+b2*x+b3*d+u if t > 1
  xtset i t
  xtreg y L.y x d, fe
  mat fe = e(b)
  replace gamma_fe = fe[1,1] in `j'
  ivregress 2sls D.y D.x D.d (D.L.y = L2.y), nocon
  mat ha = e(b)
  replace gamma_ha = ha[1,1] in `j'
  xtabond y x d, nocons lags(1)
  mat ab = e(b)
  replace gamma_ab = ab[1,1] in `j'
}
tw (kdensity gamma_fe) (kdensity gamma_ha) (kdensity gamma_ab), ///
xline(.6, lpattern(dash) lcolor(blue)) xsize(16) ysize(9) legend(cols(3))
```

AB é consistente, e é eficiente



Escolhendo modelos de painel dinâmico: R^2 , AIC etc. não se aplicam

- ▶ Estimativas de GMM dependem substancialmente da escolha de Z e W .
 - ▶ Quanto maior a dimensão de Z , mais chance de instrumentos fracos.
 - ▶ 1 etapa (W determinístico) é consistente, mas menos eficiente.
 - ▶ 2 etapas (W estimado) é mais eficiente, mas pode ser inconsistente.
- ▶ Testes populares:
 - ▶ Sargan verifica se os instrumentos são exógenos.
 - ▶ Arellano-Bond verifica correlação nos resíduos: $AR(1)$ é esperado e $AR(2)$ não deve ser significativo.
- ▶ Como se faz na prática:
 - ▶ Evite usar muitos instrumentos, e compare robustez dos coeficientes entre diferentes especificações.
 - ▶ Use o conhecimento do fenômeno estudado para guiar a escolha de defasagens e instrumentos.

Exemplo para a equação de Mincer na PNAD-C

```
cls
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_fln_filtrada.xlsx", firstrow
g tri = qofd(t)
form tri %tq
replace t = tri
drop tri
*g lnw = ln(w/((30/7)*h))
g lnw = ln(w)
g idade2 = idade^2
global X edu idade idade2 homem branco
reg lnw $X if cond_dom == 1
est sto ols
xtset id t
xtreg lnw edu idade idade2 if cond_dom == 1, fe
est sto fe
ivregress 2sls D.lnw D.edu D.idade D.idade2 (D.L.lnw = L2.lnw) if cond_dom == 1, nocon
est sto ah
xtabond lnw edu idade idade2 if cond_dom == 1, lags(1)
est sto ab1
xtdpdsys lnw edu idade idade2 if cond_dom == 1, lags(1)
est sto ab2
estout ols fe ah ab1 ab2, cells(b(star fmt(3)) se(par fmt(3)))
```

Exemplo para projeção de taxa de homicídios

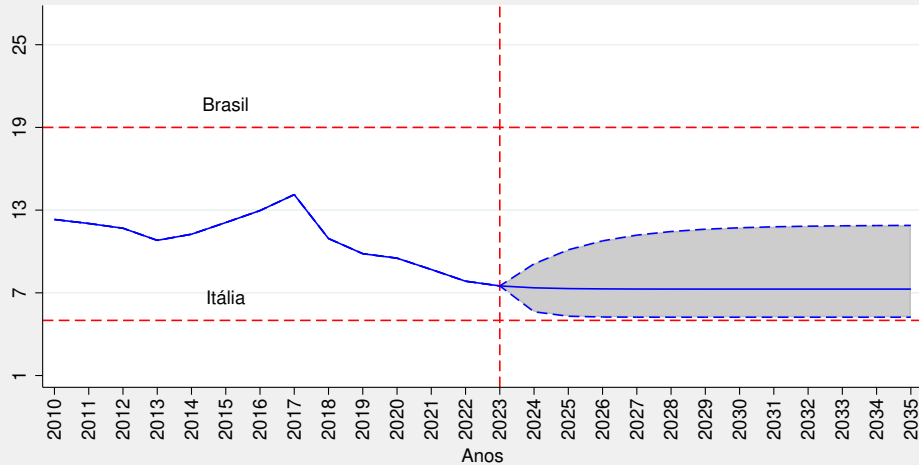
```
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_projeta_avanca_sc.xlsx", ///
sheet("data") firstrow
loc n = _N
loc N = _N+12
set obs `N'
forvalues i = `n'/'`N' {
    replace i = "SC" in `i'
    replace t = t[_n-1]+1 in `i'
}
fillin i t
drop _fillin
en i, g(id)
xtset id t
g y = tx_hom_total
gl nome_var "Taxa de homicídios"
*xtunitroot fisher y, dfuller lags(1)
ivregress 2sls D.y (D.L.y = L2.y) if i != "SC", nocon vce(cluster i)
sca gamma = e(b)[1,1]
sca sigma = sqrt(e(V)[1,1])
```

```
g y_trend = y
g y_minus = y
g y_plus = y
sort id t
bys id: g fe = y-gamma*L.y if t == 2023
bys id: egen fe2 = mean(fe)
bys id: replace fe = fe2
drop fe2
sca z = 1.64485
sum y
sca min = floor(r(min))
sca max = ceil(r(max))
sca interv = ceil((max-min)/5)
bys id: replace y_trend = gamma*y_trend[_n-1]+fe if y_trend == .
bys id: replace y_minus = (gamma-z*sigma)*y_minus[_n-1]+fe if y_minus == .
bys id: replace y_plus = (gamma+z*sigma)*y_plus[_n-1]+fe if y_plus == .
```

```
loc regioes "AMAI AMARP AMAUC AMAVI AMEOSC AMERIOS AMESC AMFRI AMMOC AMNOROESTE AMOSC AMPLANORTE
AMPLASC AMREC AMUNESC AMURC AMUREL AMURES AMVALI AMVE GRANFPOLIS SC"
```

```
foreach l of loc regioes {
  tw (rarea y_minus y_plus t, color(gs12)) ///
    (line y_trend t, lcolor(blue) lpattern(solid) lwidth(medium)) ///
    (line y_minus t, lcolor(blue) lpattern(dash) lwidth(medium)) ///
    (line y_plus t, lcolor(blue) lpattern(dash) lwidth(medium)) if i == "`l'", ///
    title("${nome_var} para `l'", justification(center)) ///
  xtitle("Anos", angle(90) lcolor(black)) ///
  xlabel(2010(1)2035, angle(90) format(%ty)) ///
  ylabel(`=min'(`=interv')`=max', grid) yscale(range( )) ///
  graphregion(color(gs15)) ///
  xsize(16) ysize(9) legend(off) ///
  xline(2023, lcolor(red) lpattern(dash) lwidth(medium)) ///
  text(20 2015 "Brasil", place(n)) ///
  yline(19, lcolor(red) lpattern(dash) lwidth(medium)) ///
  text(6 2015 "Itália", place(n)) ///
  yline(5 2015, lcolor(red) lpattern(dash) lwidth(medium))
  graph export "C:\Users\73771430020\Desktop\`l'.png", replace
}
```

Taxa de homicídios para SC



Exemplo para projeção de vendas de motos

```
cls
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_motos.xlsx", ///
sheet("data") firstrow
local n = _N
local N = _N + 12
set obs `N'
forvalues i = `='`n' + 1' / `N' {
    replace i = i[_n-1] in `i'
    replace t = t[_n-1]+1 in `i'
}
fillin i t
drop _fillin
sort i t
replace ano = 2013 if mi(ano)
bys i: replace mes = _n-84 if mi(mes)
replace modelo = modelo[_n-1] if mi(modelo)
bys i: replace p = p[_n-1] if mi(p)
```

```
gen cc = .
replace cc = 150 if modelo == "CG150"
replace cc = 125 if modelo == "BIZ"
replace cc = 125 if modelo == "CG125"
replace cc = 250 if modelo == "COMET"
replace cc = 600 if modelo == "HORNET"
replace cc = 250 if modelo == "MIRAGE"
replace cc = 115 if modelo == "NEO"
replace cc = 150 if modelo == "NXR"
replace cc = 750 if modelo == "SHADOW"
replace cc = 660 if modelo == "XT"
replace cc = 125 if modelo == "XTZ"
replace cc = 125 if modelo == "YBR"
g flex = 0
replace flex = 1 if modelo == "CG150" & ano >= 2009
replace flex = 1 if modelo == "BIZ" & ano >= 2011
replace flex = 1 if modelo == "NXR" & ano >= 2010
replace p = p/1000
sort t
bys t: egen qt = sum(q)
g s = q/qt if q != .
drop qt
g lns = ln(s)
sort i t
```

```
xtset i t
xtunitroot fisher lns, dfuller lags(1)
reg lns p i.cc flex i.mes i.ano
est sto ols
xtreg lns p flex i.mes i.ano, fe
est sto fe
xi: xtdpdsys lns p i.cc flex i.mes i.ano, lags(1)
est sto ab
xi: xtdpdsys lns L.lns p i.cc flex i.mes i.ano
est sto ab2
xi: xtdpdsys lns L12.lns p i.cc flex i.mes i.ano
est sto ab12
estat abond
estout ols fe ab ab2 ab12, cells(b(star fmt(3)) se(par fmt(3)))

forecast create projcoes
forecast estimates ab12
forecast solve, prefix(proj_) begin(85) end(96)
```

```

g proj_s = exp(proj_lns)

loc motos "CG150 BIZ CG125 COMET HORNET MIRAGE NEO NXR SHADOW XT XTZ YBR"
loc codes 1 2 3 4 5 6 7 8 9 10 11 12

forvalues j = 1/12 {
    local l : word `j' of `motos'
    local code : word `j' of `codes'
    line proj_s t if i == `code', lcolor(blue) lpattern(dash) lwidth(medium) ///
    title("Projeção para `l'", justification(center)) ///
    xtitle("t", angle(90) lcolor(black)) ///
    xlabel(1(3)96, angle(90) format(%ty)) ///
    graphregion(color(gs15)) ///
    xsize(16) ysize(9) legend(off) ///
    xline(85, lcolor(red) lpattern(dash) lwidth(medium))

    graph export "C:/Users/73771430020/Desktop/~l'.png", replace
}

```

Exemplo de projeção da variação do preço do querosene de aviação

```
clear all
import excel "https://petterini.ufsc.br/files/2024/11/querosene.xlsx", ///
sheet("data") firstrow
/*
loc regioes "NO NE CO SU SE BR"
foreach l of loc regioes {
    kdensity p if i == "`l'"
}
*/
loc n = _N
loc N = _N+10
set obs `N'
forvalues i = `n'/_N' {
    replace i = "BR" in `i'
    replace t = t[_n-1]+1 in `i'
}
fillin i t
drop _fillin
en i, g(id)
xtset id t
*xtline p
g y = ln(p/L.p)
xtunitroot fisher y, dfuller lags(1)
replace date1 = date1[_n-1]+7 if mi(date1)
ivregress 2sls D.y (D.L.y = L2.y) if i != "BR", nocon vce(cluster id)
g gamma = e(b)[1,1]
sort id
forv l = 1/6 {
    bys id: g sigma`l' = sqrt(abs(y[70]-y[69])*e(V)[1,1]) /// método delta
}
sort id t
```

```

bys id: g fe = y-gamma*L.y if t == 70
bys id: egen fe2 = mean(fe)
bys id: replace fe = fe2
drop fe2
sca z = 1.96
sum y
g y_trend = y
g y_minus = y
g y_plus = y
sort id
forv l = 1/6 {
replace y_trend = gamma*y_trend[_n-1]+fe if y_trend == . & id == `l'
replace y_minus = (gamma-z*sigma`l')*y_trend[_n-1]+fe if y_minus == . & id == `l'
replace y_plus = (gamma+z*sigma`l')*y_trend[_n-1]+fe if y_plus == . & id == `l'
}
loc regioes "NO NE CO SU SE BR"
foreach l of loc regioes {
  tw (rarea y_minus y_plus t, color(gs12)) ///
    (line y_trend t, lcolor(blue) lpattern(solid) lwidth(medium)) ///
    (line y_minus t, lcolor(blue) lpattern(dash) lwidth(medium)) ///
    (line y_plus t, lcolor(blue) lpattern(dash) lwidth(medium)) if i == "`l'", ///
    title("variação do preço do querosene de aviação no `l'", justification(center)) ///
    xtitle("semanas", angle(90) lcolor(black)) ///
    xlabel(0(5)80, angle(90)) ///
    graphregion(color(gs15)) ///
    xsize(16) ysize(9) legend(off) ///
    xline(70, lcolor(red) lpattern(dash) lwidth(medium)) ///
    yline(0 80, lcolor(red) lpattern(dash) lwidth(medium))
  graph export "C:\Users\73771430020\Desktop\g`l'.png", replace
}

```

P2

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

► $y = 1$ se $y^* \geq 0$ e $y = 0$ se $y^* < 0$

$$\left\{ \begin{array}{lcl} y_{11}^* & = & \gamma y_{10} + \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + c_1 + \varepsilon_{11} \\ y_{12}^* & = & \gamma y_{11} + \beta_1 + \beta_2 x_{12} + \beta_3 d_{12} + c_1 + \varepsilon_{12} \\ & \vdots & \\ y_{1T}^* & = & \gamma y_{1(T-1)} + \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + c_1 + \varepsilon_{1T} \\ & \vdots & \\ y_{n1}^* & = & \gamma y_{n0} + \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + c_n + \varepsilon_{n1} \\ y_{n2}^* & = & \gamma y_{n1} + \beta_1 + \beta_2 x_{n2} + \beta_3 d_{n2} + c_n + \varepsilon_{n2} \\ & \vdots & \\ y_{nT}^* & = & \gamma y_{n(T-1)} + \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + c_n + \varepsilon_{nT} \end{array} \right.$$

Breve revisão da alfabetização

Probit e Logit se $\gamma = 0$ e $T = 1 \dots$ e um único x

$$\Pr(y = 1 \mid \bar{x}, d = 1) - \Pr(y = 1 \mid \bar{x}, d = 0)$$

$$\Pr(y^* > 0 \mid \bar{x}, d = 1) - \Pr(y^* > 0 \mid \bar{x}, d = 0)$$

$$\Pr(\beta_1 + \beta_2 \bar{x} + \beta_3 + u > 0) - \Pr(\beta_1 + \beta_2 \bar{x} + u > 0) = .25$$

$$\beta_1 = 0 \text{ e } \bar{x} = 0 \text{ e } u \sim N(0, 1)$$

$$\Pr(u > -\beta_3) - \Pr(u > 0) = .25$$

$$\Phi(\beta_3) - \Phi(0) = .25 \Rightarrow \Phi(\beta_3) = .75 \Rightarrow \beta_3 = \Phi^{-1}(.75) \approx .67448975$$

```

cls
clear all
mat results = J(30,6,.)
forv i=1(1)30 {
    clear
    set obs `=1000*`i''
    set seed 54321`i'
    g x = runiform()-.5
    g d = runiform() > .5
    g y = x+invnormal(.75)*d+rnormal() > 0
    egen x_mean = mean(x)
    probit y x d
    sca delta_probit = normal(_b[_cons]+_b[x]*x_mean+_b[d])-normal(_b[_cons]+_b[x]*x_mean)
    mat results[`i',1] = delta_probit
    mat results[`i',3] = _b[x]
    mat results[`i',5] = _b[d]
    logit y x d
    sca delta_logit = invlogit(_b[_cons]+_b[x]*x_mean+_b[d])-invlogit(_b[_cons]+_b[x]*x_mean)
    mat results[`i',2] = delta_probit
    mat results[`i',4] = _b[x]
    mat results[`i',6] = _b[d]
}
mat l results

```

Principias elementos de ML

$$P(y_i = 1 \mid \mathbf{x}_i) = F(\beta' \mathbf{x}_i) \Rightarrow P(\{(y_i, \mathbf{x}_i)\}_{i=1}^n) \stackrel{\text{i.i.d.}}{=} \prod_{i=1}^n [F(\beta' \mathbf{x}_i)]^{y_i} [1 - F(\beta' \mathbf{x}_i)]^{1-y_i}$$

$$\ln L(\beta) = \sum_{i=1}^n [y_i \ln F(\beta' \mathbf{x}_i) + (1-y_i) \ln (1-F(\beta' \mathbf{x}_i))] \Rightarrow \begin{cases} \nabla(\beta) = \sum_{i=1}^n (y_i - F(\beta' \mathbf{x}_i)) \mathbf{x}_i \\ H(\beta) = -\sum_{i=1}^n f(\beta' \mathbf{x}_i) \mathbf{x}_i' \mathbf{x}_i \end{cases}$$

$$\hat{\beta}_{j+1} \leftarrow \hat{\beta}_j - [H(\hat{\beta}_j)]^{-1} \nabla(\hat{\beta}_j) \Rightarrow \hat{\beta} \xrightarrow{d} N\left(\mathbf{0}, \left(-H(\hat{\beta})\right)^{-1}\right)$$

$$\text{Vâr}(\hat{\beta}) = \left(-H(\hat{\beta})\right)^{-1} \Rightarrow \text{se}(\hat{\beta}_k) = \sqrt{\left[\text{Vâr}(\hat{\beta})\right]_{kk}} \Rightarrow \mathfrak{z}_k = \frac{\hat{\beta}_k}{\text{se}(\hat{\beta}_k)}$$

$$\text{Pseudo-}R^2 = 1 - \ln L(\hat{\beta}) / \ln L(\hat{\beta}_0)$$

```

cls
clear all

set obs 100
set seed 54321
g x = runiform()
g y = x+rnormal() > 0

probit y x
estat ml

program drop _all
program myprobit
args lnf xb
qui{
  replace `lnf' = $ML_y1*ln(normal(`xb'))+(1-$ML_y1)*ln(normal(-`xb'))
}
end
ml model lf myprobit (y = x)
ml check
ml maximize
estat ml

estout pro ml, cells(b(star fmt(3)) se(par fmt(3)))

```

```

mat b = (0 \ 0)
sca tol = 1e-5
sca diff = 1
sca iter = 0
sca max_iter = 100
g xb = .
g p = .
g score = .
g g0 = .
g g1 = .
g h = .
g h2 = .
g denom = .
g H00_part = .
g H01_part = .
g H11_part = .
g log_likelihood_contrib = .
while (diff > tol) & (iter < max_iter) {
replace xb = b[1,1]+b[2,1]*x
replace p = normal(xb)
replace score = y-p
replace g0 = score
replace g1 = score*x
replace h = normalden(xb)
replace h2 = h*h
replace denom = p*(1-p)
replace H00_part = h2/denom
replace H01_part = (h2*x)/denom
replace H11_part = (h2*(x^2))/denom

```

```

sum H00_part
sca H00 = -r(sum)
sum H01_part
sca H01 = -r(sum)
sum H11_part
sca H11 = -r(sum)
sum g0
sca G0 = r(sum)
sum g1
sca G1 = r(sum)
mat G = ( G0 \ G1 )
mat H = ( H00, H01 \ H01, H11 )
mat H_inv = inv(H)
mat delta = -H_inv*G
mat b = b+delta
sca diff = sqrt(delta[1,1]^2 + delta[2,1]^2)
replace log_likelihood_contrib = y*log(p)+(1-y)*log(1-p)
sum log_likelihood_contrib
sca log_likelihood = r(sum)
di "Iteration " iter
di ": b0 = " b[1,1] ", b1 = " b[2,1]
di ", diff = " diff
di ", log-likelihood = " log_likelihood
sca iter = iter + 1
}

```

C-Logit: Logit com efeitos fixos para $\gamma = 0$ e $T > 1$

$$P(y_{it} = 1 \mid \mathbf{x}_{it}, c_i) = \frac{\exp(\beta' \mathbf{x}_{it} + c_i)}{1 + \exp(\beta' \mathbf{x}_{it} + c_i)}$$

$T = 2$ temos $\mathbf{y}_i \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$

$$P(\mathbf{y}_i = (0, 0)) = \frac{1}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{1}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)}$$

$$P(\mathbf{y}_i = (1, 0)) = \frac{\exp(\beta' \mathbf{x}_{i1} + c_i)}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{1}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)} = \exp(c_i) \frac{\exp(\beta' \mathbf{x}_{i1})}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{1}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)}$$

$$P(\mathbf{y}_i = (0, 1)) = \frac{1}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{\exp(\beta' \mathbf{x}_{i2} + c_i)}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)} = \exp(c_i) \frac{1}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{\exp(\beta' \mathbf{x}_{i2})}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)}$$

$$P(\mathbf{y}_i = (1, 1)) = \frac{\exp(\beta' \mathbf{x}_{i1} + c_i)}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{\exp(\beta' \mathbf{x}_{i2} + c_i)}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)} = \exp(2c_i) \frac{\exp(\beta' \mathbf{x}_{i1})}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{\exp(\beta' \mathbf{x}_{i2})}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)}$$

$$P(\mathbf{y}_i \mid \mathbf{x}_{it}, c_i) = \frac{\exp(y_{i1}\beta' \mathbf{x}_{i1} + y_{i1}c_i)}{1 + \exp(\beta' \mathbf{x}_{i1} + c_i)} \times \frac{\exp(y_{i2}\beta' \mathbf{x}_{i2} + y_{i2}c_i)}{1 + \exp(\beta' \mathbf{x}_{i2} + c_i)} = \exp(c_i \sum_{t=1}^T y_{it}) \prod_{t=1}^T \frac{\exp(y_{it}\beta' \mathbf{x}_{it})}{1 + \exp(\beta' \mathbf{x}_{it} + c_i)}$$

$$P(y_{i1} + y_{i2} = 0 \mid \mathbf{x}_{it}, c_i) = P(\mathbf{y}_i = (0, 0) \mid \mathbf{x}_{it}, c_i)$$

$$P(y_{i1} + y_{i2} = 1 \mid \mathbf{x}_{it}, c_i) = P(\mathbf{y}_i = (1, 0) \mid \mathbf{x}_{it}, c_i) + P(\mathbf{y}_i = (0, 1) \mid \mathbf{x}_{it}, c_i)$$

$$P(y_{i1} + y_{i2} = 2 \mid \mathbf{x}_{it}, c_i) = P(\mathbf{y}_i = (1, 1) \mid \mathbf{x}_{it}, c_i)$$

$$P\left(\sum_{t=1}^T y_{it} = s \mid \mathbf{x}_i, c_i\right) = \sum_s \exp(c_i \sum_{t=1}^T y_{it}) \prod_{t=1}^T \frac{\exp(y_{it} \beta' \mathbf{x}_{it})}{1 + \exp(\beta' \mathbf{x}_{it} + c_i)}$$

$$P(\mathbf{y}_i \mid \sum_{t=1}^T y_{it} = s, \mathbf{x}_{it}) = \frac{\prod_{t=1}^T \exp(c_i \sum_{t=1}^T y_{it}) \frac{\exp(y_{it} \beta' \mathbf{x}_{it})}{1 + \exp(\beta' \mathbf{x}_{it} + c_i)}}{\sum_s \prod_{t=1}^T \exp(c_i \sum_{t=1}^T y_{it}) \frac{\exp(y_{it} \beta' \mathbf{x}_{it})}{1 + \exp(\beta' \mathbf{x}_{it} + c_i)}} = \frac{\exp(y_{it} \beta' \mathbf{x}_{it})}{\sum_{i \text{ é do tipo } s} \exp(y_{it} \beta' \mathbf{x}_{it})}$$

```

clear all
set seed 54321
loc n = 1000
loc T = 10
set obs `n'
g i = _n
expand `T'
bys i: g t = _n
g x = 5*(runiform()-.5)
g d = runiform()>.5
sort i t
bys i: g c = 5*(runiform()-.5) if _n == 1
bys i: replace c = c[_n-1] if mi(c)
g y = x+d+c+rlogistic() > 0
sum y
logit y x d c
est sto logit_com_c
logit y x d
est sto logit_sem_c
xtset i t
xtlogit y x d, fe
est sto xtlogit
clogit y x d, gr(i)
est sto clogit
estout logit_com_c logit_sem_c xtlogit clogit, cells(b(star fmt(3)) se(par fmt(3)))

```

```

clogit y x d, gr(i)
mat b = e(b)

range xval -2.5 2.5 11
g p_d1 = invlogit(b[1,1]*xval+b[1,2])
g p_d0 = invlogit(b[1,1]*xval)

tw (line p_d1 xval, lcolor(blue) lpattern(solid) lwidth(medium)) ///
    (line p_d0 xval, lcolor(red) lpattern(dash) lwidth(medium)), ///
xlabel(-2.5(.5)2.5, format(%2.1f)) ylabel(0(.2)1, format(%2.1f)) ///
legend(label(1 "P(y=1|x,d=1)" label(2 "P(y=1|x,d=0)" ///
placement(bottom) col(2) region(margin(0) lstyle(none))) graphregion(color(white))

```

Método Delta

- Deseja-se o erro-padrão de uma função diferenciável $g(\hat{\beta})$, talque $g : \mathbb{R} \rightarrow \mathbb{R}$.

$$g(\hat{\beta}) \approx g(\beta) + g'(\beta)(\hat{\beta} - \beta) \Rightarrow \text{Vâr}(g(\hat{\beta})) \approx [g'(\hat{\beta})]^2 \text{Vâr}(\hat{\beta})$$

- Se $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K)'$ e $g : \mathbb{R}^K \rightarrow \mathbb{R}$, o gradiente é:

$$\nabla g(\beta) = \left(\frac{\partial g}{\partial \beta_1}, \frac{\partial g}{\partial \beta_2}, \dots, \frac{\partial g}{\partial \beta_K} \right)'$$

$$\text{Vâr}(g(\hat{\beta})) \approx \nabla g(\hat{\beta})' \text{Vâr}(\hat{\beta}) \nabla g(\hat{\beta})$$

$$\Pr(y = 1 \mid x, d = 1) = g(\beta) = \Lambda(\beta_1 x + \beta_2) \Rightarrow \nabla g(\hat{\beta}) = \begin{bmatrix} \Lambda(\hat{\beta}_1 + \hat{\beta}_2)(1 - \Lambda(\hat{\beta}_1 + \hat{\beta}_2))x \\ \Lambda(\hat{\beta}_1 + \hat{\beta}_2)(1 - \Lambda(\hat{\beta}_1 + \hat{\beta}_2)) \end{bmatrix}$$

$$\Pr(y = 1 \mid x, d = 0) = g(\beta) = \Lambda(\beta_1 x) \Rightarrow \nabla g(\hat{\beta}) = \Lambda(\hat{\beta}_1)(1 - \Lambda(\hat{\beta}_1))x$$

$$\text{Var}(\hat{\beta}) = \left(-H(\hat{\beta})\right)^{-1}$$

```

drop p_d*
clogit y x d, group(i)
mat b = e(b)
mat V = e(V)
g p_d1 = .
g s_d1 = .
g l_d1 = .
g u_d1 = .
g p_d0 = .
g s_d0 = .
g l_d0 = .
g u_d0 = .

forv j = 1/11 {
  replace p_d1 = invlogit(b[1,1]*xval[`j']+b[1,2]) in `j'
  mat grad_d1 = (invlogit(b[1,1]*xval[`j']+b[1,2]))*(1-invlogit(b[1,1]*xval[`j']+b[1,2]))*xval[`j'] \ ///
    invlogit(b[1,1]*xval[`j']+b[1,2]))*(1-invlogit(b[1,1]*xval[`j']+b[1,2]))
  mat var_d1 = grad_d1'*V*grad_d1
  replace s_d1 = sqrt(var_d1[1,1]) in `j'
  replace l_d1 = p_d1-1.96*s_d1 in `j'
  replace u_d1 = p_d1+1.96*s_d1 in `j'
  replace p_d0 = invlogit(b[1,1]*xval[`j']) in `j'
  sca grad_d0 = invlogit(b[1,1]*xval[`j'])*(1-invlogit(b[1,1]*xval[`j']))*xval[`j']
  sca var_d0 = (grad_d0^2)*V[1,1]
  replace s_d0 = sqrt(var_d0) in `j'
  replace l_d0 = p_d0-1.96*s_d0 in `j'
  replace u_d0 = p_d0+1.96*s_d0 in `j'
}

tw (rarea l_d1 u_d1 xval, color(ltblue%50) lcolor(blue)) (line p_d1 xval, lcolor(blue) lpattern(solid) lwidth(medium)) ///
  (rarea l_d0 u_d0 xval, color(ltred%25) lcolor(red)) (line p_d0 xval, lcolor(red) lpattern(dash) lwidth(medium)), ///
  xlabel(-2.5(.5)2.5, format(%2.1f)) ylabel(0(.2)1, format(%2.1f)) ///
  legend(label(1 "P(y=1|x,d=1)") label(2 "P(y=1|x,d=0)") ///
  placement(bottom) col(2) region(margin(0) lstyle(none))) graphregion(color(white))

```

```

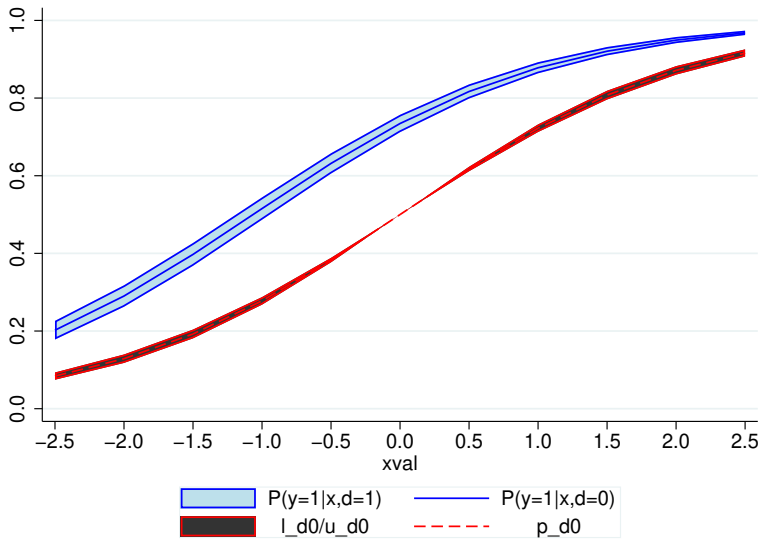
drop p_d*

g p_d1 = .
g l_d1 = .
g u_d1 = .
g p_d0 = .
g l_d0 = .
g u_d0 = .

forv j = 1/11 {
  clogit y x d, group(i)
  nlcom p1: invlogit(_b[x]*xval[`j']+_b[d]), post
  replace p_d1 = _b[p1] in `j'
  replace l_d1 = _b[p1]-1.96*_se[p1] in `j'
  replace u_d1 = _b[p1]+1.96*_se[p1] in `j'
  clogit y x d, group(i)
  nlcom p0: invlogit(_b[x]*xval[`j']), post
  replace p_d0 = _b[p0] in `j'
  replace l_d0 = _b[p0]-1.96*_se[p0] in `j'
  replace u_d0 = _b[p0]+1.96*_se[p0] in `j'
}

tw (rarea l_d1 u_d1 xval, color(ltblue%50) lcolor(blue)) (line p_d1 xval, lcolor(blue) lpattern(solid) lwidth(medium)) ///
(rarea l_d0 u_d0 xval, color(ltred%25) lcolor(red)) (line p_d0 xval, lcolor(red) lpattern(dash) lwidth(medium)), ///
xlabel(-2.5(.5)2.5, format(%2.1f)) ylabel(0(.2)1, format(%2.1f)) ///
legend(label(1 "P(y=1|x,d=1)") label(2 "P(y=1|x,d=0)") ///
placement(bottom) col(2) region(margin(0) lstyle(none))) graphregion(color(white))

```



union.dta

```
cls  
clear all  
webuse union  
describe  
sum
```

- ▶ **year** e **idcode**
- ▶ **age**: anos de idade do trabalhador
- ▶ **grade**: anos de escolaridade completados
- ▶ **not_smsa**: = 1 se reside fora de uma área metropolitana
- ▶ **south**: = 1 vive na região sul dos EUA
- ▶ **union**: = 1 afiliação sindical
- ▶ **black**: = 1 negro

```
gl X age black grade not_smsa south
```

```
logit union $X  
est sto logit  
xtset idcode year  
xtlogit union $X, fe  
est sto xtlogit  
estout logit xtlogit, cells(b(star fmt(3)) se(par fmt(3)))
```

```
tab grade
```

```
sum age  
sca age_m = r(mean)  
sum not_smsa  
sca not_smsa_m = r(mean)
```

```
range edu 0 18 19
```

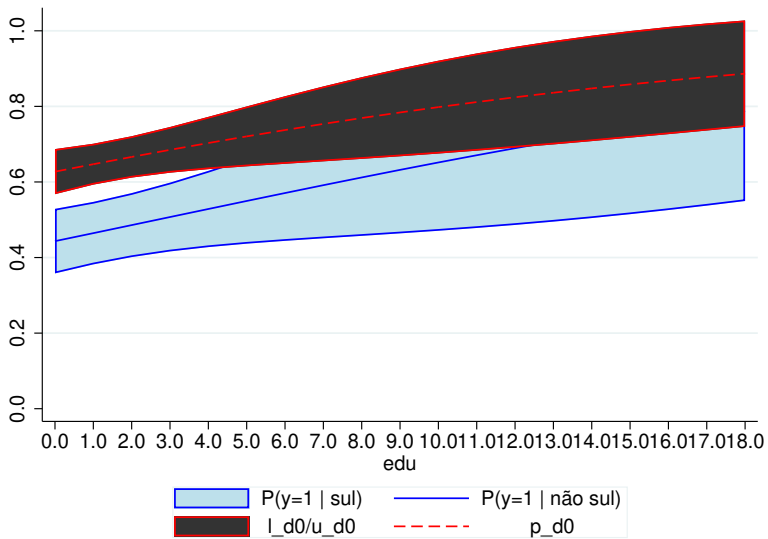
```
g p_d1 = .  
g l_d1 = .  
g u_d1 = .  
g p_d0 = .  
g l_d0 = .  
g u_d0 = .
```

```

forv j = 1/19 {
  clogit union age grade not_smsa south, group(idcode)
  nlcom p1: invlogit(_b[age]*age_m+_b[grade]*edu[`j']+_b[not_smsa]*not_smsa_m+_b[south]), post
  replace p_d1 = _b[p1] in `j'
  replace l_d1 = _b[p1]-1.96*_se[p1] in `j'
  replace u_d1 = _b[p1]+1.96*_se[p1] in `j'
  clogit union age black grade not_smsa south, group(idcode)
  nlcom p0: invlogit(_b[age]*age_m+_b[grade]*edu[`j']+_b[not_smsa]*not_smsa_m), post
  replace p_d0 = _b[p0] in `j'
  replace l_d0 = _b[p0]-1.96*_se[p0] in `j'
  replace u_d0 = _b[p0]+1.96*_se[p0] in `j'
}

tw (rarea l_d1 u_d1 edu, color(ltblue%50) lcolor(blue)) (line p_d1 edu, lcolor(blue) lpattern(solid)) ///
(rarea l_d0 u_d0 edu, color(ltred%25) lcolor(red)) (line p_d0 edu, lcolor(red) lpattern(dash)), ///
xlabel(0(1)18, format(%2.1f)) ylabel(0(.2)1, format(%2.1f)) ///
legend(label(1 "P(y=1 | sul)") label(2 "P(y=1 | não sul)") ///
placement(bottom) col(2) region(margin(0) lstyle(none))) graphregion(color(white))

```



Acrescentando a dinâmica

Probit com $|\gamma| < 1$ e o problema da condição inicial

$$\left\{ \begin{array}{l} y_{11}^* = \gamma y_{10} + \beta x_{11} + c_1 + \varepsilon_{11} \\ y_{12}^* = \gamma y_{11} + \beta x_{12} + c_1 + \varepsilon_{12} \\ y_{13}^* = \gamma y_{12} + \beta x_{13} + c_1 + \varepsilon_{13} \\ y_{21}^* = \gamma y_{20} + \beta x_{21} + c_2 + \varepsilon_{21} \\ y_{22}^* = \gamma y_{21} + \beta x_{22} + c_2 + \varepsilon_{22} \\ y_{23}^* = \gamma y_{22} + \beta x_{23} + c_2 + \varepsilon_{23} \end{array} \right. ; \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix} | (x_{it}, c_i) \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \right)$$

$$\varepsilon \sim \mathcal{N}(0, \Sigma) \therefore \Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \Rightarrow f(\varepsilon) = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} \varepsilon^T \Sigma^{-1} \varepsilon \right)$$

$$\begin{aligned}
\Pr(y_1 = 0, y_2 = 0, y_3 = 0) &= \Pr(\varepsilon_1 \leq 0, \varepsilon_2 \leq 0, \varepsilon_3 \leq 0) = \int_{-\infty}^0 \int_{-\infty}^0 \int_{-\infty}^0 f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 0, y_2 = 0, y_3 = 1) &= \Pr(\varepsilon_1 \leq 0, \varepsilon_2 \leq 0, \varepsilon_3 > 0) = \int_{-\infty}^0 \int_{-\infty}^0 \int_0^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 0, y_2 = 1, y_3 = 0) &= \Pr(\varepsilon_1 \leq 0, \varepsilon_2 > 0, \varepsilon_3 \leq 0) = \int_{-\infty}^0 \int_0^{\infty} \int_{-\infty}^0 f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 0, y_2 = 1, y_3 = 1) &= \Pr(\varepsilon_1 \leq 0, \varepsilon_2 > 0, \varepsilon_3 > 0) = \int_{-\infty}^0 \int_0^{\infty} \int_0^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 1, y_2 = 0, y_3 = 0) &= \Pr(\varepsilon_1 > 0, \varepsilon_2 \leq 0, \varepsilon_3 \leq 0) = \int_0^{\infty} \int_{-\infty}^0 \int_{-\infty}^0 f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 1, y_2 = 0, y_3 = 1) &= \Pr(\varepsilon_1 > 0, \varepsilon_2 \leq 0, \varepsilon_3 > 0) = \int_0^{\infty} \int_{-\infty}^0 \int_0^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 1, y_2 = 1, y_3 = 0) &= \Pr(\varepsilon_1 > 0, \varepsilon_2 > 0, \varepsilon_3 \leq 0) = \int_0^{\infty} \int_0^{\infty} \int_{-\infty}^0 f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3, \\
\Pr(y_1 = 1, y_2 = 1, y_3 = 1) &= \Pr(\varepsilon_1 > 0, \varepsilon_2 > 0, \varepsilon_3 > 0) = \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3.
\end{aligned}$$

Soluções mais populares

- ▶ FIML é complicado computacionalmente:

$$P(\{\mathbf{y}_i\}_{i=1}^T \mid \text{todos } y_{i0}, x_{it}) = \int \prod_{i=1}^n \left[\int_{\mathbf{y}_i} f(\boldsymbol{\varepsilon}_i(\mathbf{par\hat{a}metros})) d\boldsymbol{\varepsilon} \right] f(c) dc$$

- ▶ Heckman (1978) (Orme (2001) melhora) trata parametricamente y_0 , e mantém c como um efeito aleatório, mas ainda é complicado computacionalmente:

$$P(\{\mathbf{y}_i\}_{i=1}^T \mid \text{todos } y_{i0}, x_{it}) = \int \prod_{i=1}^n \Pr(y_{1i}, \dots, y_{iT} \mid \text{controles de } y_{0i}) f(c) dc$$

- ▶ Wooldridge (2005) modela c como função de y_0 (e outras variáveis iniciais), e monta uma função de verossimilhança condicionada em y_0 . Isso contorna problemas computacionais (e dá aplicabilidade para outros modelos).

$$c_i = \text{cte} + \rho y_0 + \text{controles} + \eta_i ; \eta_i \text{ é um componente aleatório}$$

$$P(\{\mathbf{y}_i\}_{i=1}^T \mid \text{todos } y_{i0}, x_{it}) = \int \prod_{i=1}^n \Pr(y_{1i}, \dots, y_{iT} \mid \text{controles de } c_i) f(\eta) d\eta$$

Simulando um Probit-Wooldridge

```
clear all
set seed 54321
loc n = 100
loc T = 10
loc reps = 50
mat m_probit = J(`reps', 2, .)
mat m_wooldr = J(`reps', 2, .)
tic
forv rep = 1/`reps' {
  clear
  set obs `n'
  g i = _n
  expand `T'
  bys i: g t = _n
  sca gamma = .25
  sca beta = 1
  sort i t
  bys i: gen c = rlogistic(0,2) if _n == 1  toc
  bys i: replace c = c[_n-1] if mi(c)
  g x = rpoisson(9)
  g u = c+rnormal()
  g y = runiform() > .5 if t == 1
  bys i: replace y = gamma*y[_n-1]+beta*x+u > 0 if t > 1

  xtset i t
  probit y L.y x, iter(30)
  mat m_probit[`rep',1] = _b[L.y]
  mat m_probit[`rep',2] = _b[x]
  // Wooldridge
  bys i: g y0 = y[1]
  bys i: egen x_mean = mean(x)
  xtprobit y L.y x y0 x_mean if t > 1, re i(i) iter(30)
  mat m_wooldr[`rep',1] = _b[L.y]
  mat m_wooldr[`rep',2] = _b[x]
}

svmat m_probit, names(p)
svmat m_wooldr, names(w)

sum p* w*
```

Exemplo com a PNAD-C

```
cls
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_fln_filtrada.xlsx", firstrow
g tri = qofd(t)
form tri %tq
replace t = tri
drop tri
keep if pea == 1 & cond_dom == 1
gl X idade homem branco edu
probit ocup $X
est sto probit_n
xtset id t
probit ocup L.ocup $X
est sto probit_l
sort id t
bys id: g ocup0 = ocup[1]
bys id: egen idade_mean = mean(idade)
bys id: egen edu_mean = mean(edu)
xtprobit ocup L.ocup $X ocup0 idade_mean edu_mean if t > 1, ///
re i(id) iter(30)
est sto probit_w
estout probit_n probit_l probit_w, cells(b(star fmt(3)) se(par fmt(3)))
disp normal(_b[L.ocup])
disp normal(_b[L.ocup])-.5
```

Keane & Wolpin (1997)

Resumo: 11 anos de observações de uma amostra de jovens do coorte de 1979 das National Longitudinal Surveys of Labor Market Experience (NLSY).

```
cls
clear all
*ssc install bcuse
*http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html
bcuse keane
keep if black == 1 & numyrs == 11
xtset id year
probit employ L.employ
est sto probit_normal
bys id: g employ0 = employ[1]
xtprobit employ L.employ employ0 if year > 81, re i(id) iter(30)
est sto probit_wooldridge
estout probit_normal probit_wooldridge, cells(b(star fmt(3)) se(par fmt(3)))
```

$$y \geq \text{mínimo}$$

- $y = y^*$ se $y^* > 0$ e $y = 0$ se $y^* \leq 0$ no caso contrário (abordagem de censura)

$$\left\{ \begin{array}{lcl} y_{11}^* & = & \gamma y_{10} + \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + c_1 + \varepsilon_{11} \\ y_{12}^* & = & \gamma y_{11} + \beta_1 + \beta_2 x_{12} + \beta_3 d_{12} + c_1 + \varepsilon_{12} \\ & \vdots & \\ y_{1T}^* & = & \gamma y_{1(T-1)} + \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + c_1 + \varepsilon_{1T} \\ & \vdots & \\ y_{n1}^* & = & \gamma y_{n0} + \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + c_n + \varepsilon_{n1} \\ y_{n2}^* & = & \gamma y_{n1} + \beta_1 + \beta_2 x_{n2} + \beta_3 d_{n2} + c_n + \varepsilon_{n2} \\ & \vdots & \\ y_{nT}^* & = & \gamma y_{n(T-1)} + \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + c_n + \varepsilon_{nT} \end{array} \right.$$

A verossimilhança padrão

$$\ln L(\beta) = \sum_{i=1}^n [I(y_i > 0) \ln f(y_i - \beta' \mathbf{x}_i) + I(y_i = 0) \ln F(-\beta' \mathbf{x}_i)]$$

$$\Pr(y_i = 0) = \Pr(y_i^* \leq 0) = \Pr(\beta' \mathbf{x}_i + \varepsilon_i \leq 0) = \Pr(\varepsilon_i \leq -\beta' \mathbf{x}_i) = F(-\beta' \mathbf{x}_i)$$

$$\beta_{j+1} \leftarrow \beta_j - [H(\beta_j)]^{-1} \nabla(\beta_j)$$

$$y = \begin{cases} y^* & \text{se } y^* > 0, \\ 0 & \text{se } y^* \leq 0, \end{cases} ; y^* = \beta' \mathbf{x}_i + \varepsilon ; \varepsilon \sim N(0, \sigma^2)$$

$$\mathcal{L}(\beta, \sigma) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_i - \beta' \mathbf{x}_i)^2}{2\sigma^2} \right) \right]^{I(y_i > 0)} \cdot \left[\Phi \left(-\frac{\beta' \mathbf{x}_i}{\sigma} \right) \right]^{1-I(y_i > 0)}$$

$$\begin{aligned} \mathcal{L}(\beta, \sigma, \sigma_c) = \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{t=1}^T \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_{it} - \beta' \mathbf{x}_i - c_i)^2}{2\sigma^2} \right) \right]^{I(y_{it} > 0)} \\ \cdot \left[\Phi \left(-\frac{\beta' \mathbf{x}_i + c_i}{\sigma} \right) \right]^{1-I(y_{it} > 0)} f(c_i) dc_i \end{aligned}$$

Tobit-Wooldridge

```
clear all
set seed 54321
loc n = 100
loc T = 10
loc reps = 50
mat m_tobit = J(`reps', 2, .)
mat m_woolldr = J(`reps', 2, .)
tic
forv rep = 1/`reps' {
  clear
  set obs `n'
  g i = _n
  expand `T'
  bys i: g t = _n
  sca gamma = .25
  sca beta = 1
  sort i t
  bys i: gen c = rlogistic(0,2) if _n == 1
  bys i: replace c = c[_n-1] if mi(c)
  g x = rpoisson(9)
  g u = c+rnormal()
  g y = max(0, beta*x+u) if t == 1
  g y_star = .
  forv j = 2/`T' {
    bys i: replace y_star = gamma*y[_n-1]+beta*x+u if t == `j'
    bys i: replace y = max(0,y_star) if t == `j'
  }

  xtset i t
  tobit y L.y x, ll(0) iter(30)
  mat m_tobit[`rep',1] = _b[L.y]
  mat m_tobit[`rep',2] = _b[x]
  // Wooldridge
  bys i: gen y0 = y[1]
  bys i: egen x_mean = mean(x)
  xttobit y L.y x y0 x_mean if t > 1, re ll(0) i(i) iter(30)
  mat m_woolldr[`rep',1] = _b[L.y]
  mat m_woolldr[`rep',2] = _b[x]
}
toc
svmat m_tobit, names(to)
svmat m_woolldr, names(w)
sum to* w*
```

Exemplo com taxa de homicídios municipal no Tobit

```
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_tx_hom_mun_sc.xlsx", sheet("data") firstrow
en municipio, g(i)
xtset i ano
reg tx L.tx, vce(cluster i)
est sto ols
reg tx L.tx if tx != 0, vce(cluster i)
est sto ols2
ivregress 2sls D.tx (D.L.tx = L2.tx), nocon vce(cluster i)
est sto ah
ivregress 2sls D.tx (D.L.tx = L2.tx) if tx != 0, nocon vce(cluster i)
est sto ah2
tobit tx L.tx, ll(0) vce(cluster i)
est sto to
bys i: gen tx0 = tx[1]
xttobit tx L.tx tx0 if ano > 2010, re ll(0) i(i) iter(30)
est sto wo
estout ols ols2 ah ah2 to wo, cells(b(star fmt(3)) se(par fmt(3)))
```

$$\text{mínimo} \leq y \leq \text{máximo}$$

Survival (censura na direita)

O principal caso envolve $y \geq 0$ representando duração de um evento.

$$y = \begin{cases} y^* & \text{se } y^* \leq \tau_i, \\ \tau_i & \text{se } y^* > \tau_i, \end{cases}$$

$$\mathcal{L}(\beta, \sigma) = \prod_{i=1}^n [f(y_i - \beta' \mathbf{x}_i)]^{I(y_i < \tau_i)} \cdot [1 - F(\tau_i - \beta' \mathbf{x}_i)]^{1 - I(y_i < \tau_i)}$$

$$\begin{aligned} \mathcal{L}(\beta, \sigma, \sigma_c) &= \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{t=1}^T [f(y_i - \beta' \mathbf{x}_i - c_i)]^{I(y_i < \tau_i)} \\ &\quad \cdot [1 - F(\tau_i - \beta' \mathbf{x}_i - c_i)]^{1 - I(y_i < \tau_i)} f(c_i) dc_i \end{aligned}$$

LogNormal

Sob a parametrização $\mu = \beta' \mathbf{x} + c$ e $\sigma > 0$ é o desvio padrão do logaritmo:

► fdp

$$f(t) = \frac{1}{t \cdot \sigma \sqrt{2\pi}} \cdot \exp \left(-\frac{(\ln t - (\beta' \mathbf{x} + c))^2}{2\sigma^2} \right)$$

► cdf

$$F(t) = \Phi \left(\frac{\ln t - (\beta' \mathbf{x} + c)}{\sigma} \right)$$

► esperança

$$E(t) = \exp \left(\beta' \mathbf{x} + c + \frac{\sigma^2}{2} \right)$$

► função quantílica

$$t = \exp \left(\mu + \sigma \cdot \Phi^{-1}(U) \right), \quad U \sim \text{Uniforme}(0, 1)$$

Weibull

Sob a parametrização $\lambda = \exp(-\alpha(\beta'\mathbf{x} + c))$, onde $\alpha > 0$ é o parâmetro de forma:

► fdp

$$f(t) = \exp(-\alpha(\beta'\mathbf{x} + c)) \cdot \alpha \cdot t^{\alpha-1} \cdot \exp(-\exp(-\alpha(\beta'\mathbf{x} + c)) \cdot t^\alpha)$$

► cdf

$$F(t) = 1 - \exp(-\exp(-\alpha(\beta'\mathbf{x} + c)) \cdot t^\alpha)$$

► esperança

$$E(t) = \exp(\beta'\mathbf{x} + c) \cdot \Gamma\left(1 + \frac{1}{\alpha}\right)$$

► função quantílica

$$t = \left(-\frac{\ln U}{\exp(-\alpha(\beta'\mathbf{x} + c))}\right)^{1/\alpha}$$

Exemplo “Table 22.1” do Wooldridge (2010)

```
cls
clear all
bcuse recid
* Definindo variáveis
generate f = 1 - cens
stset durat, fail(f)
global X workprg priors tserve felon alcohol drugs black married educ age
* Inicializando matriz para guardar os resultados
matrix AIC_BIC = J(4, 2, .)
matrix colnames AIC_BIC = AIC BIC
matrix rownames AIC_BIC = Weibull Lognormal Weibull_ML Lognormal_ML
* Modelo Weibull padrão
streg $X, dist(weibull) nohr difficult
est store weib
estat ic
matrix results = r(S)
matrix AIC_BIC[1, 1] = results[1, 5] // AIC
matrix AIC_BIC[1, 2] = results[1, 6] // BIC
stcurve, survival title("Survival Curve (Weibull)") ///
    xlabel(0(10)50) ylabel(0(0.2)1) ///
    saving(weibull_survival, replace)
* Modelo Lognormal padrão
streg $X, dist(lognormal) difficult
est store lognormal
estat ic
matrix results = r(S)
matrix AIC_BIC[2, 1] = results[1, 5] // AIC
matrix AIC_BIC[2, 2] = results[1, 6] // BIC
```

```

* Weibull ajustado manualmente com ML
capture program drop my_weibull
program define my_weibull
    args lnf xb a
    tempvar alph
    quietly {
        loc t "$ML_y1"
        gen double `alph' = exp(`a')
        replace `lnf' = `xb' + `a' + (`alph' - 1) * ln(`t') - exp(`xb') * (`t'`^`alph')
        replace `lnf' = -exp(`xb') * (`t'`^`alph') if cens == 1
    }
end
ml model lf my_weibull (durat = $X) ()
ml check
ml max, difficult iterate(50)
est store weib2
estat ic
matrix results = r(S)
matrix AIC_BIC[3, 1] = results[1, 5] // AIC
matrix AIC_BIC[3, 2] = results[1, 6] // BIC

```

```

* Lognormal ajustado manualmente com ML
capture program drop my_lognormal
program define my_lognormal
    args lnf xb sigma
    tempvar sigmasq
    quietly {
        loc t "$ML_y1"
        gen double `sigmasq' = exp(2 * `sigma')
        replace `lnf' = -ln(`t') - ln(sqrt(2*_pi)) - ln(`sigma') - (ln(`t') - `xb')^2 / (2 * `sigma'^2)
        replace `lnf' = ln(1 - normal((ln(`t') - `xb') / `sigma')) if cens == 1
    }
end
ml model lf my_lognormal (durat = $X) ()
ml check
ml max, difficult iterate(50)
est store lognormal2
estat ic
matrix results = r(S)
matrix AIC_BIC[4, 1] = results[1, 5] // AIC
matrix AIC_BIC[4, 2] = results[1, 6] // BI

```

```
* Exibindo a tabela com os resultados
display "Final AIC and BIC comparison table:"
matrix list AIC_BIC

estout weib weib2 lognormal lognormal2, cells(b(star fmt(3)) se(par fmt(3)))

* Ajustar modelo Weibull
streg $X, dist(weibull) nohr difficult

* Gerar dados para sobrevivência com black = 1
stcurve, survival at(black=1 black=0)

margins, at(priors=(0 1 2 3 4 5 6 7 8 9 10) black=(0 1)) plot
```

Simulando “duração até uma infecção” a partir da Weibull

```
cls
clear all
set seed 54321
sca n = 1000
sca T = 5
sca alfa = 1.5
sca beta = -.3
sca rho = .3
set obs `=n*T'
g i = int((_n-1)/T)+1
sort i
by i: g t = _n
g x = rnormal(0,1)
bys i: g c = rnormal(0,2) if t == 1
bys i: replace c = c[_n-1] if mi(c)
g lambda = .
by i: replace lambda = exp(-alfa*(beta*x+c)) if t == 1
g durat = .
by i: replace durat = (-ln(runiform())/lambda)^(1/alfa) if t == 1
forv j = 2/`=T' {
  replace lambda = exp(-alfa*(beta*x+c+rho*durat[_n-1])) if t == `j'
  replace durat = (-ln(runiform())/lambda)^(1/alfa) if t == `j'
}
```

```

by i: g durat_lag = durat[_n-1]
drop if t == 1
replace t = t-1
g cens = runiform() > 0.8
stset durat, fail(cens)
streg x c durat_lag, dist(weibull) time
est sto weib1
streg x durat_lag, dist(weibull) time
est sto weib2
sort i t
bys i: g durat0 = durat[1]
bys i: egen x_mean = mean(x)
xtset i
xtstreg x durat_lag durat0 x_mean if t>1, dist(weibull) time
est sto weib3
estout weib1 weib2 weib3, cells(b(star fmt(3)) se(par fmt(3)))

margins, at(x=(-2 -1 0 1 2) durat_lag=(0 1)) plot

```

$$y \in \mathbb{N}$$

- y é uma contagem Poisson, então $y^* = E(y \mid \text{controles})$

$$\left\{ \begin{array}{l} y_{11}^* = \exp(\gamma y_{10} + \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + c_1) \\ y_{12}^* = \exp(\gamma y_{11} + \beta_1 + \beta_2 x_{12} + \beta_3 d_{12} + c_1) \\ \vdots \\ y_{1T}^* = \exp(\gamma y_{1(T-1)} + \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + c_1) \\ \vdots \\ y_{n1}^* = \exp(\gamma y_{n0} + \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + c_n) \\ y_{n2}^* = \exp(\gamma y_{n1} + \beta_1 + \beta_2 x_{n2} + \beta_3 d_{n2} + c_n) \\ \vdots \\ y_{nT}^* = \exp(\gamma y_{n(T-1)} + \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + c_n) \end{array} \right.$$

$$y_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \exp(\beta' \mathbf{x}_i + c_i)$$

$$\mathcal{L}(\beta) = \prod_{i=1}^n \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} = \prod_{i=1}^n \frac{(\exp(\beta' \mathbf{x}_i + c_i))^{y_i} \exp(-\exp(\beta' \mathbf{x}_i + c_i))}{y_i!}$$

$$\ln \mathcal{L}(\beta) = \sum_{i=1}^n (y_i(\beta' \mathbf{x}_i + c_i) - \exp(\beta' \mathbf{x}_i + c_i) - \ln(y_i!))$$

$$\mathcal{L}(\beta, \sigma_c) = \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{t=1}^T \frac{\lambda_{it}^{y_{it}} \exp(-\lambda_{it})}{y_{it}!} f(c_i) dc_i$$

Há uma versão de FE ao estilo C-Logit

- ▶ Modelos de contagem são usados para variáveis dependentes representando números inteiros (ex.: número de eventos).
- ▶ Principais modelos:
 - ▶ Regressão de Poisson
 - ▶ Regressão Binomial Negativa (BN)
 - ▶ Regressão Zero-Inflated Poisson (ZIP)

Regressão de Poisson

Características principais:

- ▶ Modela contagens de eventos em um intervalo fixo.
- ▶ Assume que a variável dependente segue uma **distribuição de Poisson**.
- ▶ Fórmula do modelo:

$$\log(\lambda_i) = X_i\beta$$

Onde:

- ▶ λ_i : Média esperada do número de eventos.
- ▶ X_i : Vetor de covariáveis.
- ▶ β : Vetor de coeficientes.

Limitações:

- ▶ Sensível à **superdispersão** (variância maior que a média).
- ▶ Não lida bem com **excesso de zeros**.

Regressão Binomial Negativa (BN)

Características principais:

- ▶ Modela contagens de eventos com **superdispersão**.
- ▶ Relaxa a suposição de que média e variância são iguais.
- ▶ Fórmula da variância:

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

Onde:

- ▶ μ_i : Média esperada.
- ▶ θ : Parâmetro de dispersão.

Vantagens:

- ▶ Flexível para lidar com superdispersão.

Limitações:

- ▶ Não resolve problemas de **excesso de zeros**.

Regressão Zero-Inflated Poisson (ZIP)

Características principais:

- ▶ Lida com **excesso de zeros**.
- ▶ Assume dois processos geradores de zeros:
 - ▶ Um **processo estrutural** (sempre zero).
 - ▶ Um **processo Poisson**.
- ▶ Fórmula do modelo:

$$P(Y_i = 0) = \pi_i + (1 - \pi_i)e^{-\lambda_i}$$

Onde:

- ▶ π_i : Probabilidade do zero estrutural.
- ▶ $e^{-\lambda_i}$: Probabilidade do zero Poisson.

Vantagens:

- ▶ Modela zeros estruturais e incidentais separadamente.

Limitações:

- ▶ Interpretação mais complexa.

Comparação entre Modelos

Modelo	Situação Ideal	Limitação Principal
Poisson	Variância próxima à média	Sensível à superdispersão
BN	Superdispersão sem excesso de zeros	Não lida com excesso de zeros
ZIP	Excesso de zeros estruturais	Interpretação mais complexa

Escolha do Modelo

Cr terios para Escolher o Modelo:

- ▶ ****Superdispers o****:
 - ▶ Compare vari ncia e m dia da vari vel dependente.
- ▶ ****Excesso de Zeros****:
 - ▶ Compare a propor  o de zeros observados com o esperado.
- ▶ ****Testes Estat sticos****:
 - ▶ **Vuong Test**: ZIP vs Poisson.
 - ▶ **AIC/BIC**: Para comparar modelos alternativos.

- ▶ Cada modelo tem um uso específico:
 - ▶ Poisson: Dados simples sem superdispersão.
 - ▶ BN: Dados com superdispersão.
 - ▶ ZIP: Excesso de zeros estruturais.
- ▶ Use testes diagnósticos e critérios de seleção para escolher o modelo mais adequado.

Suplemento da PNAD-C

```
cls
clear all
tic
infix ///
uf 6-7 ///
capital 8-9 ///
rural 33-33 ///
sexo 95-95 ///
idade 104-106 ///
cor 107-107 ///
naointernet 505-505 ///
avaliapol 559-559 ///
chancevio 597-597 ///
foifurtdado 630-630 ///
vezesfurto 631-632 ///
foiroubado 785-785 ///
vezesroubo 786-787 ///
familia 576-576 ///
amigo 577-577 ///
colega 578-578 ///
```

```
vizinho 579-579 ///  
guarda 580-580 ///  
civil 581-581 ///  
pm 582-582 ///  
bombeiros 583-583 ///  
justica 584-584 ///  
forcas_arm 585-585 ///  
using "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.txt", clear  
compress  
sa "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta", replace  
toc
```

```

cls
clear all
use "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta"
keep if capital != .
tab vezesfurto vezesroubo, m
g y =.
replace y = 0 if vezesfurto == . & vezesroubo == .
replace y = 1 if (vezesfurto == 1 & vezesroubo == .) | (vezesfurto == . & vezesroubo == 1)
replace y = 2 if (vezesfurto == 2 & vezesroubo == .) | (vezesfurto == . & vezesroubo == 2) \\
    | (vezesfurto == 1 & vezesroubo == 1)
replace y = vezesfurto+vezesroubo if y == .
tab y
replace rural = rural-1
g homem = 2-sexo
g branco = cond(cor==1,1,0)
tab capital, gen(c_)
gl X rural homem branco c_2-c_27
sum y
poisson y $X
nbreg y $X, dispersion(mean)
zip y $X, inflate(_cons)

```

```
* Estime o modelo ZIP
zip y $X, inflate(_cons)
```

```
* Calcule o preditor linear para ambas as partes do modelo
```

```
gen xb_count_rural1 = _b[_cons] + _b[homem]*homem + _b[branco]*branco + _b[rural]*1 ///
                    + _b[c_2]*c_2 + _b[c_3]*c_3 + _b[c_4]*c_4 + _b[c_5]*c_5 + _b[c_6]*c_6 + ///
                    _b[c_7]*c_7 + _b[c_8]*c_8 + _b[c_9]*c_9 + _b[c_10]*c_10 + _b[c_11]*c_11 + ///
                    _b[c_12]*c_12 + _b[c_13]*c_13 + _b[c_14]*c_14 + _b[c_15]*c_15 + _b[c_16]*c_16 + ///
                    _b[c_17]*c_17 + _b[c_18]*c_18 + _b[c_19]*c_19 + _b[c_20]*c_20 + _b[c_21]*c_21 + ///
                    _b[c_22]*c_22 + _b[c_23]*c_23 + _b[c_24]*c_24 + _b[c_25]*c_25 + _b[c_26]*c_26 + ///
                    _b[c_27]*c_27
```

```
gen xb_count_rural0 = _b[_cons] + _b[homem]*homem + _b[branco]*branco + _b[rural]*0 ///
                    + _b[c_2]*c_2 + _b[c_3]*c_3 + _b[c_4]*c_4 + _b[c_5]*c_5 + _b[c_6]*c_6 + ///
                    _b[c_7]*c_7 + _b[c_8]*c_8 + _b[c_9]*c_9 + _b[c_10]*c_10 + _b[c_11]*c_11 + ///
                    _b[c_12]*c_12 + _b[c_13]*c_13 + _b[c_14]*c_14 + _b[c_15]*c_15 + _b[c_16]*c_16 + ///
                    _b[c_17]*c_17 + _b[c_18]*c_18 + _b[c_19]*c_19 + _b[c_20]*c_20 + _b[c_21]*c_21 + ///
                    _b[c_22]*c_22 + _b[c_23]*c_23 + _b[c_24]*c_24 + _b[c_25]*c_25 + _b[c_26]*c_26 + ///
                    _b[c_27]*c_27
```

```
gen xb_inflate_rural1 = _b[inflate:_cons]
gen xb_inflate_rural0 = _b[inflate:_cons]
```

```

* Calcule as probabilidades de zeros estruturais
gen p_zero_rural1 = invlogit(xb_inflate_rural1)
gen p_zero_rural0 = invlogit(xb_inflate_rural0)

* Calcule o valor esperado (mu) para a parte de contagem
gen mu_rural1 = exp(xb_count_rural1)
gen mu_rural0 = exp(xb_count_rural0)

* Calcule as probabilidades para Y = 0, 1, 2, 3
gen pr_y0_rural1 = p_zero_rural1 + (1 - p_zero_rural1) * exp(-mu_rural1)
gen pr_y1_rural1 = (1 - p_zero_rural1) * mu_rural1 * exp(-mu_rural1)
gen pr_y2_rural1 = (1 - p_zero_rural1) * (mu_rural1^2 / 2) * exp(-mu_rural1)
gen pr_y3_rural1 = (1 - p_zero_rural1) * (mu_rural1^3 / 6) * exp(-mu_rural1)

gen pr_y0_rural0 = p_zero_rural0 + (1 - p_zero_rural0) * exp(-mu_rural0)
gen pr_y1_rural0 = (1 - p_zero_rural0) * mu_rural0 * exp(-mu_rural0)
gen pr_y2_rural0 = (1 - p_zero_rural0) * (mu_rural0^2 / 2) * exp(-mu_rural0)
gen pr_y3_rural0 = (1 - p_zero_rural0) * (mu_rural0^3 / 6) * exp(-mu_rural0)

* Calcule médias das probabilidades
egen avg_pr_y0_rural1 = mean(pr_y0_rural1)
egen avg_pr_y1_rural1 = mean(pr_y1_rural1)
egen avg_pr_y2_rural1 = mean(pr_y2_rural1)
egen avg_pr_y3_rural1 = mean(pr_y3_rural1)

```

```
egen avg_pr_y0_rural0 = mean(pr_y0_rural0)
egen avg_pr_y1_rural0 = mean(pr_y1_rural0)
egen avg_pr_y2_rural0 = mean(pr_y2_rural0)
egen avg_pr_y3_rural0 = mean(pr_y3_rural0)
```

* Organize os dados para o gráfico

```
gen y_value = 0
replace y_value = 1 in 2
replace y_value = 2 in 3
replace y_value = 3 in 4
```

```
gen pr_rural1 = .
replace pr_rural1 = avg_pr_y0_rural1 in 1
replace pr_rural1 = avg_pr_y1_rural1 in 2
replace pr_rural1 = avg_pr_y2_rural1 in 3
replace pr_rural1 = avg_pr_y3_rural1 in 4
```

```
gen pr_rural0 = .
replace pr_rural0 = avg_pr_y0_rural0 in 1
replace pr_rural0 = avg_pr_y1_rural0 in 2
replace pr_rural0 = avg_pr_y2_rural0 in 3
replace pr_rural0 = avg_pr_y3_rural0 in 4
```

* Gráfico comparativo

```
twoway (line pr_rural1 y_value, sort) (line pr_rural0 y_value, sort), ///
       legend(label(1 "Rural = 1") label(2 "Rural = 0")) ///
       title("Comparação de Probabilidades Estimadas para Y (Modelo ZIP)") ///
       xlabel(0(1)3) ylabel(, angle(horizontal))
```

Base de dados de contagem simulada

```
clear all
set seed 54321
loc n = 100
loc T = 10
loc reps = 50
mat m_poisson = J(`reps', 2, .)
mat m_woolldr = J(`reps', 2, .)
tic
forv rep = 1/`reps' {
  clear
  set obs `n'
  g i = _n
  expand `T'
  bys i: gen t = _n
  sca gamma = .25
  sca beta = 1
  sort i t
  bys i: gen c = runiform()-.5 if _n == 1
  bys i: replace c = c[_n-1] if mi(c)
  g x = runiform()-.5
  g y = rpoisson(runiform()) if t == 1
  g lambda = .
  forv j = 2/`T' {
    bys i: replace lambda = exp(gamma*y[_n-1]+beta*x+c) if t == `j'
    bys i: replace y = min(rpoisson(lambda),20) if t == `j'
  }

  xtset i t
  poisson y L.y x
  mat m_poisson[`rep',1] = _b[L.y]
  mat m_poisson[`rep',2] = _b[x]
  // Wooldridge
  bys i: gen y0 = y[1]
  bys i: egen x_mean = mean(x)
  xtpoisson y L.y x y0 x_mean if t > 1, re iter(30)
  mat m_woolldr[`rep',1] = _b[L.y]
  mat m_woolldr[`rep',2] = _b[x]
}
toc
svmat m_poisson, names(po)
svmat m_woolldr, names(w)
sum po* w*
```

Modelo de Poisson com Efeitos Fixos

O modelo básico de Poisson é:

$$Y_{it} \sim \text{Poisson}(\mu_{it}),$$

com:

$$\mu_{it} = \exp(X_{it}\beta + \alpha_i + \ln(\text{pop}_{it})),$$

onde:

- ▶ Y_{it} : variável de contagem.
- ▶ μ_{it} : valor esperado de Y_{it} .
- ▶ $X_{it}\beta$: combinação linear das covariáveis.
- ▶ α_i : efeito fixo específico da unidade i .
- ▶ $\ln(\text{pop}_{it})$: offset (log da exposição pop_{it}).

Função de Verossimilhança Completa

A função de verossimilhança para todos os dados é:

$$\mathcal{L}(\beta, \alpha) = \prod_{i=1}^N \prod_{t=1}^T \frac{\mu_{it}^{Y_{it}} e^{-\mu_{it}}}{Y_{it}!}.$$

Substituindo μ_{it} :

$$\mathcal{L}(\beta, \alpha) = \prod_{i=1}^N \prod_{t=1}^T \frac{[\exp(X_{it}\beta + \alpha_i + \ln(\text{pop}_{it}))]^{Y_{it}} e^{-\exp(X_{it}\beta + \alpha_i + \ln(\text{pop}_{it}))}}{Y_{it}!}.$$

Problema: Não podemos estimar α_i diretamente devido ao problema dos parâmetros incidentais.

Condicionização pela Soma Total dos Eventos

A soma total dos eventos para cada unidade é:

$$S_i = \sum_{t=1}^T Y_{it}.$$

A probabilidade condicional de Y_{it} , dado S_i , elimina α_i :

$$P(Y_{i1}, Y_{i2}, \dots, Y_{iT} \mid S_i) = \frac{\prod_{t=1}^T \mu_{it}^{Y_{it}}}{\sum_{\{Y_{it}: \sum_t Y_{it} = S_i\}} \prod_{t=1}^T \mu_{it}^{Y_{it}}}.$$

Aqui, α_i desaparece porque afeta μ_{it} de forma multiplicativa.

Função de Verossimilhança Condicional

A função de verossimilhança condicional é:

$$\mathcal{L}_C(\beta) = \prod_{i=1}^N \frac{\prod_{t=1}^T [\exp(X_{it}\beta + \ln(\text{pop}_{it}))]^{Y_{it}}}{\sum_{\{Y_{it}: \sum_t Y_{it} = S_i\}} \prod_{t=1}^T [\exp(X_{it}\beta + \ln(\text{pop}_{it}))]^{Y_{it}}}.$$

Simplificando:

$$\mathcal{L}_C(\beta) = \prod_{i=1}^N \frac{\prod_{t=1}^T [\text{pop}_{it} \cdot \exp(X_{it}\beta)]^{Y_{it}}}{\sum_{\{Y_{it}: \sum_t Y_{it} = S_i\}} \prod_{t=1}^T [\text{pop}_{it} \cdot \exp(X_{it}\beta)]^{Y_{it}}}.$$

Interpretação

- ▶ A condicionalização elimina α_i , permitindo estimar β sem viés.
- ▶ O denominador, que depende de todas as combinações possíveis de Y_{it} que somam S_i , é calculado numericamente.
- ▶ Os coeficientes β podem ser interpretados como log-taxas relativas (log-rate ratios).

```

clear all
import excel "https://petterini.ufsc.br/files/2024/12/contagens.xlsx", sheet("fecam") firstrow
drop if i == "SC"
en i, g(id)
xtset id t
xtpoisson homM t, fe exposure(pop)
est sto pofe
sort id t
bys id: g y0 = homM[1]
bys id: egen t_mean = mean(t)
xtpoisson homM L.homM t y0 t_mean if t > 1, fe iter(30)
est sto wofe
xtpoisson homM L.homM t y0 t_mean if t > 1, re iter(30)
est sto wore
estout pofe wofe wore, cells(b(star fmt(3)) se(par fmt(3)))

```

$$y \in \{a, b, c\}$$

Logit Multinomial

1. Partindo da Função de Utilidade:

- ▶ O modelo assume que cada indivíduo i escolhe entre J alternativas.
- ▶ A utilidade da alternativa j para o indivíduo i é dada por:

$$U_{ij} = \beta_j' \mathbf{x}_{ij} + \xi_{ij},$$

onde:

- ▶ $\mathbf{x}_{ij}$: vetor de características observáveis da alternativa j e do indivíduo i ,
- ▶ β_j : vetor de coeficientes estimáveis específicos da alternativa j ,
- ▶ ξ_{ij} : componente de erro (não observado).

2. Assumindo Distribuição Gumbel para ξ_{ij} :

- ▶ Para simplificar a análise, assume-se que os erros ξ_{ij} são i.i.d. segundo a distribuição Gumbel (valor extremo):

$$F(\xi) = \exp(-\exp(-\xi)).$$

- ▶ Propriedades:
 - ▶ Esta distribuição tem diferenças entre erros bem comportadas, permitindo derivar probabilidades.
 - ▶ Produz uma função de densidade cumulativa (CDF) para escolhas discretas.

3. Derivando a Probabilidade Logística:

- ▶ A probabilidade de o indivíduo i escolher a alternativa j é:

$$\Pr(y_i = j) = \Pr(U_{ij} > U_{ik}, \forall k \neq j).$$

- ▶ Substituímos a utilidade:

$$\Pr(y_i = j) = \Pr(\beta'_j \mathbf{x}_{ij} + \xi_{ij} > \beta'_k \mathbf{x}_{ik} + \xi_{ik}, \forall k \neq j).$$

- ▶ Como $\xi_{ij} \sim \text{Gumbel}$, as diferenças entre ξ_{ij} têm distribuição logística, e podemos derivar:

$$\Pr(y_i = j) = \frac{\exp(\beta'_j \mathbf{x}_{ij})}{\sum_{k=1}^J \exp(\beta'_k \mathbf{x}_{ik})}.$$

4. Normalizando a Utilidade:

- ▶ A utilidade é identificável apenas até uma constante comum.
- ▶ Para resolver essa indeterminação, é comum normalizar a utilidade da categoria de referência ($j = 1$):

$$U_{i1} = 0 \quad \Rightarrow \quad \beta_1 = \mathbf{0}.$$

- ▶ As probabilidades ajustadas são:

$$\Pr(y_i = 1) = \frac{1}{1 + \sum_{k=2}^J \exp(\beta'_k \mathbf{x}_{ij})},$$
$$\Pr(y_i = j) = \frac{\exp(\beta'_j \mathbf{x}_{ij})}{1 + \sum_{k=2}^J \exp(\beta'_k \mathbf{x}_{ij})}, \quad j = 2, \dots, J.$$

Modelo Logit Multinomial com Efeitos Fixos

O modelo logit multinomial é usado para modelar uma variável de escolha discreta Y_{it} com J categorias ($j = 1, 2, \dots, J$).

$$P(Y_{it} = j) = \frac{\exp(V_{ijt})}{\sum_{k=1}^J \exp(V_{ikt})},$$

com:

$$V_{ijt} = X_{ijt}\beta + \alpha_{ij},$$

onde:

- ▶ X_{ijt} : covariáveis específicas da alternativa j no tempo t .
- ▶ β : coeficientes estimáveis.
- ▶ α_{ij} : efeito fixo específico da unidade i para a alternativa j .

Problema dos Efeitos Fixos

A função de verossimilhança completa seria:

$$\mathcal{L}(\beta, \alpha) = \prod_{i=1}^N \prod_{t=1}^T \prod_{j=1}^J P(Y_{it} = j)^{d_{ijt}},$$

com:

$$P(Y_{it} = j) = \frac{\exp(X_{ijt}\beta + \alpha_{ij})}{\sum_{k=1}^J \exp(X_{ikt}\beta + \alpha_{ik})}.$$

Problema:

- ▶ Muitos α_{ij} a estimar ($N \times J$).
- ▶ Viés dos parâmetros incidentais.

Condicionização para Eliminar α_{ij}

Condicionamos na soma total de escolhas por alternativa j para cada indivíduo i :

$$S_{ij} = \sum_{t=1}^T d_{ijt}, \quad \text{onde } d_{ijt} = 1 \text{ se } Y_{it} = j, \text{ e } 0 \text{ caso contrário.}$$

A probabilidade condicional das escolhas $\{Y_{it}\}$, dado S_{ij} , elimina α_{ij} :

$$P(\{Y_{it}\} \mid S_{i1}, \dots, S_{iJ}) = \frac{\prod_{t=1}^T \prod_{j=1}^J \exp(X_{ijt}\beta)^{d_{ijt}}}{\sum_{\{Y_{it}:S_{ij}\}} \prod_{t=1}^T \prod_{j=1}^J \exp(X_{ijt}\beta)^{d_{ijt}}}.$$

Função de Verossimilhança Condicional

A função de verossimilhança condicional é:

$$\mathcal{L}_C(\beta) = \prod_{i=1}^N \frac{\prod_{t=1}^T \prod_{j=1}^J \exp(X_{ijt}\beta)^{d_{ijt}}}{\sum_{\{Y_{it}: S_{ij}\}} \prod_{t=1}^T \prod_{j=1}^J \exp(X_{ijt}\beta)^{d_{ijt}}}.$$

Aqui:

- ▶ $\exp(X_{ijt}\beta)$: representa a utilidade estimada da alternativa j para a unidade i no tempo t .
- ▶ O denominador normaliza a probabilidade condicionada, eliminando α_{ij} .

Interpretação e Implementação

- ▶ Os efeitos fixos α_{ij} desaparecem devido à condicionalização.
- ▶ A verossimilhança condicional permite estimar β consistentemente.
- ▶ Aplicações incluem escolha de produtos, decisões políticas, transporte, etc.
- ▶ Estimável via pacotes estatísticos que suportam modelos logit multinomial de efeitos fixos.

```

cls
clear all
bcuse keane
xtset id year
mlogit status educ expwc expersq black, baseoutcome(1)
*est sto m1
femlogit status educ expwc expersq black, group(id) baseoutcome(1)
*est sto m2
*estout m1 m2, cells(b(star fmt(3)) se(par fmt(3)))
g d2 = cond(status==2,1,0)
g d3 = cond(status==3,1,0)
bys id: g d20 = d2[1]
bys id: g d30 = d3[1]
bys id: egen educ_mean = mean(educ)
bys id: egen expwc_mean = mean(expwc)
bys id: egen expersq_mean = mean(expersq)
femlogit status L.status educ expwc expersq black d20 d30 educ_mean expwc_mean expersq_mean if year > 81, group(id)

```

Em termos de dinâmica, a defasagem deve ser escrita com um jogo de dummies, mas precisa ter um comando de efeitos aleatórios

```

cls
clear all
use "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta"
keep if capital != . | naointernet == .
tab naointernet
replace rural = rural-1
g homem = 2-sexo
g branco = cond(cor==1,1,0)
tab capital, gen(c_)
global X i.rural homem branco idade c_2-c_27
mlogit naointernet $X, baseoutcome(7)
margins rural, at(homem=1 branco=1 idade=(20 30 40 50 60 70 80) ///
    c_2=1 c_3=0 c_4=0 c_5=0 c_6=0 c_7=0 c_8=0 c_9=0 c_10=0 ///
    c_11=0 c_12=0 c_13=0 c_14=0 c_15=0 c_16=0 c_17=0 c_18=0 ///
    c_19=0 c_20=0 c_21=0 c_22=0 c_23=0 c_24=0 c_25=0 c_26=0 c_27=0) ///
    predict(outcome(5))
marginsplot, ///
    title("Probabilidade de naointernet = 5 por Idade e Rural") ///
    xlabel(20(10)80) ylabel(, angle(horizontal)) ///
    ytitle("Probabilidade Pr(naointernet=5)") xtitle("Idade") ///
    legend(label(1 "Rural = 0") label(2 "Rural = 1"))

```

$$y_1 < y_2 < y_3$$

- ▶ Os modelos **Logit Ordenado** e **Probit Ordenado** são usados para variáveis dependentes ordinais.
- ▶ Baseiam-se na ideia de uma **utilidade latente** contínua (U_i^*), que determina as categorias observáveis (y_i).
- ▶ A estrutura subjacente permite inferir o impacto de variáveis independentes nas probabilidades das categorias ordinais.

Estrutura de Utilidade Latente

Modelo Subjacente:

$$U_i^* = X_i\beta + \epsilon_i,$$

onde:

- ▶ U_i^* : utilidade latente (não observada),
- ▶ X_i : vetor de covariáveis observáveis,
- ▶ β : vetor de coeficientes estimáveis,
- ▶ ϵ_i : termo de erro.

Categorias Observáveis:

$$y_i = j \quad \text{se} \quad \mu_{j-1} < U_i^* \leq \mu_j, \quad j = 1, 2, \dots, J.$$

- ▶ μ_j : pontos de corte estimados, separando as categorias de y_i .

Diferença entre Logit e Probit Ordenado

Distribuição do Termo de Erro ϵ_i :

► **Logit Ordenado:** $\epsilon_i \sim \text{Logística}(0, \pi^2/3)$

$$F(\epsilon_i) = \frac{\exp(\epsilon_i)}{1 + \exp(\epsilon_i)}.$$

► **Probit Ordenado:** $\epsilon_i \sim \mathcal{N}(0, 1)$

$$F(\epsilon_i) = \Phi(\epsilon_i),$$

onde $\Phi(\cdot)$ é a CDF da normal padrão.

Probabilidades:

$$\Pr(y_i = j) = F(\mu_j - X_i\beta) - F(\mu_{j-1} - X_i\beta).$$

- Para o Logit Ordenado: $F(\cdot)$ é a CDF logística. - Para o Probit Ordenado: $F(\cdot)$ é a CDF normal.

Pontos de Corte e Probabilidades

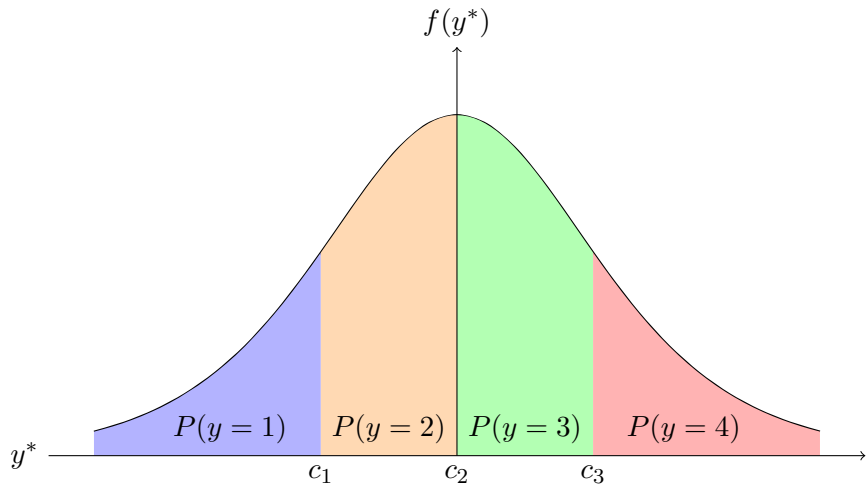
Pontos de Corte (μ_j):

- ▶ Os pontos de corte são estimados junto com os coeficientes β .
- ▶ Representam os limites entre as categorias observadas (y_i) na escala de U_i^* .

Probabilidades:

$$\Pr(y_i = j) = F(\mu_j - X_i\beta) - F(\mu_{j-1} - X_i\beta).$$

- $\Pr(y_i = j)$: probabilidade de y_i estar na categoria j . - Depende da posição relativa de $X_i\beta$ em relação aos pontos de corte μ_j e μ_{j-1} .



Interpretação de β

- ▶ Os coeficientes β capturam o efeito das covariáveis X_i na utilidade latente U_i^* .
- ▶ O impacto de β em $\Pr(y_i = j)$ é indireto, pois depende:
 - ▶ Dos valores de $X_i\beta$,
 - ▶ Da posição relativa dos pontos de corte μ_j .
- ▶ Em vez de mudanças absolutas, β representa mudanças relativas na propensão de y_i mudar de categoria.

```

cls
clear all
use "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta"
keep if capital != . | chancevio == .
tab chancevio
replace rural = rural-1
g homem = 2-sexo
g branco = cond(cor==1,1,0)
tab capital, gen(c_)
tab avaliapol chancevio, m
drop if avaliapol == . | avaliapol == 9 | chancevio == . | chancevio == 9
replace avaliapol = 6-avaliapol
replace chancevio = 5-chancevio
tab avaliapol chancevio, m
global X i.avaliapol rural homem branco idade c_2-c_27
ologit chancevio $X
margins avaliapol, at(rural=0 homem=1 branco=1 idade=30 ///
    c_2=0 c_3=0 c_4=0 c_5=0 c_6=0 c_7=0 c_8=0 c_9=0 c_10=0 ///
    c_11=0 c_12=0 c_13=0 c_14=0 c_15=0 c_16=0 c_17=0 c_18=0 ///
    c_19=0 c_20=0 c_21=0 c_22=0 c_23=0 c_24=0 c_25=0 c_26=0 c_27=0) ///
    predict(outcome(4)) predict(outcome(1))
marginsplot, ///
    title("Probabilidades de chancevio = 4 e chancevio = 1 por avaliapol") ///
    xlabel(1(1)5, format(%2.0f)) ylabel(, angle(horizontal)) ///
    ytitle("Probabilidade") xtitle("avaliapol") ///
    legend(label(1 "Pr(chancevio = 4)") label(2 "Pr(chancevio = 1)"))

```

```

cls
clear all
use "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta"
drop if capital == . | familia == . | familia == 9
tab familia, m
replace familia = 6-familia
replace rural = rural-1
g homem = 2-sexo
g branco = cond(cor==1,1,0)
tab capital, gen(c_)
global X rural homem branco i.idade c_2-c_27
ologit familia $X
margins idade, at(rural=0 homem=1 branco=1 ///
    c_2=0 c_3=0 c_4=0 c_5=0 c_6=0 c_7=0 c_8=0 c_9=0 c_10=0 ///
    c_11=0 c_12=0 c_13=0 c_14=0 c_15=0 c_16=0 c_17=0 c_18=0 ///
    c_19=0 c_20=0 c_21=0 c_22=0 c_23=0 c_24=0 c_25=0 c_26=0 c_27=0) ///
    predict(outcome(5))
marginsplot, ///
    title("Probabilidade de familia = 5 por idade") ///
    xlabel(, format(%2.0f)) ylabel(, angle(horizontal)) ///
    ytitle("Probabilidade Pr(familia = 5)") xtitle("Idade")

```

```

cls
clear all
use "G:/Meu Drive/pnad_c/PNADC_2021_trimestre4.dta"
drop if capital == . | bombeiros == . | bombeiros == 9
tab bombeiros, m
replace bombeiros = 6-bombeiros
replace rural = rural-1
g homem = 2-sexo
g branco = cond(cor==1,1,0)
tab capital, gen(c_)
global X rural homem branco i.idade c_2-c_27
ologit bombeiros $X
margins idade, at(rural=0 homem=1 branco=1 ///
    c_2=0 c_3=0 c_4=0 c_5=0 c_6=0 c_7=0 c_8=0 c_9=0 c_10=0 ///
    c_11=0 c_12=0 c_13=0 c_14=0 c_15=0 c_16=0 c_17=0 c_18=0 ///
    c_19=0 c_20=0 c_21=0 c_22=0 c_23=0 c_24=0 c_25=0 c_26=0 c_27=0) ///
    predict(outcome(5))
marginsplot, ///
    title("Probabilidade de bombeiros = 5 por idade") ///
    xlabel(, format(%2.0f)) ylabel(, angle(horizontal)) ///
    ytitle("Probabilidade Pr(bombeiros = 5)") xtitle("Idade")

```

Simulação de pesquisa de satisfação

```
cls
clear all
set seed 54321
set obs 1000
g renda = rnormal(3000,800)           // Renda média de 3000 com desvio padrão de 800
g idade = round(runiform()*47+18)     // Idade entre 18 e 65 anos
g frequencia_uso = ceil(runiform()*3) // Frequência de uso: 1 (baixa), 2 (média), 3 (alta)
* Gerar a variável latente z
g z = .5*(renda/1000)-.1*idade+.8*frequencia_uso+rnormal(0,.5)
* Definir os pontos de corte para gerar satisfação
g satisfacao = .
replace satisfacao = 1 if z <= -1.5
replace satisfacao = 2 if z > -1.5 & z <= -0.5
replace satisfacao = 3 if z > -0.5 & z <= 0.5
replace satisfacao = 4 if z > 0.5 & z <= 1.5
replace satisfacao = 5 if z > 1.5
* Limpar a variável latente
drop z

ologit satisfacao renda idade i.frequencia_uso

* Interpretação dos resultados
margins, at(renda=(2000(1000)5000) idade=(20(10)60) frequencia_uso=(1 2 3))
```

Em termos de dinâmica, a defasagem deve ser escrita pelo próprio y

y em fronteiras

Fronteira Estocástica

- ▶ A fronteira estocástica é usada para medir a eficiência técnica (ET) de unidades produtivas.
- ▶ O modelo considera uma fronteira de produção máxima e distingue:
 - ▶ **Ineficiência Técnica** (U_i): desvios sistemáticos da fronteira.
 - ▶ **Erro Aleatório** (V_i): choques imprevisíveis.
- ▶ Para dados em painel, a análise incorpora a dimensão temporal:

$$\ln Y_{it} = f(X_{it}; \beta) - U_{it} + V_{it}.$$

Componentes do Modelo

- ▶ Y_{it} : output observado para a unidade i no tempo t .
- ▶ $f(X_{it}; \beta)$: fronteira de produção máxima, especificada como:
 - ▶ **Cobb-Douglas:** $f(X_{it}; \beta) = \beta_0 + \sum_{j=1}^k \beta_j \ln X_{ijt}$,
 - ▶ **Translog:** Inclui termos quadráticos e de interação.
- ▶ U_{it} : termo de ineficiência técnica ($U_{it} \geq 0$).
- ▶ $V_{it} \sim N(0, \sigma_v^2)$: termo de erro aleatório.

Ineficiência Constante ou Variável no Tempo

Ineficiência Constante no Tempo:

$$U_{it} = U_i, \quad ET_{it} = \exp(-U_i).$$

Ineficiência Variável no Tempo:

- ▶ Modelo de decaimento exponencial (Battese e Coelli, 1992):

$$U_{it} = U_i \cdot \exp(-\eta t), \quad ET_{it} = \exp(-U_i \cdot \exp(-\eta t)).$$

- ▶ Modelo geral:

$$U_{it} \sim \text{Truncated-Normal}(\mu_{it}, \sigma_u^2).$$

Função de Verossimilhança

Ineficiência Constante no Tempo:

$$f(\epsilon_{i1}, \dots, \epsilon_{iT}) = \int_0^\infty \prod_{t=1}^T \phi(\epsilon_{it} + U_i; \sigma_v^2) \cdot \exp\left(-\frac{U_i}{\sigma_u}\right) dU_i,$$

onde:

- ▶ $\epsilon_{it} = V_{it} - U_i$: resíduo.
- ▶ $\phi(\cdot)$: densidade da normal padrão.

Ineficiência Variável no Tempo:

$$\mathcal{L} = \prod_{i=1}^N \prod_{t=1}^T \int_0^\infty f(\ln Y_{it} \mid X_{it}, U_{it}, \beta) \cdot g(U_{it}) dU_{it}.$$

Estimativa de Eficiência Técnica

- ▶ A eficiência técnica (ET_{it}) é dada por:

$$ET_{it} = \exp(-U_{it}).$$

- ▶ O termo U_{it} é estimado condicionalmente aos resíduos ϵ_{it} :

$$\mathbb{E}[U_{it} \mid \epsilon_{it}] = \frac{\sigma_u^2}{\sigma^2} \left(\epsilon_{it} + \frac{\phi\left(\frac{\epsilon_{it}}{\sigma}\right)}{\Phi\left(-\frac{\epsilon_{it}}{\sigma}\right)} \right),$$

com:

$$\sigma^2 = \sigma_u^2 + \sigma_v^2.$$

Vantagens e Desvantagens

Vantagens:

- ▶ Distingue ineficiência técnica (U_{it}) de erros aleatórios (V_{it}).
- ▶ Permite estimar mudanças na eficiência ao longo do tempo.
- ▶ Adequado para dados em painel com múltiplas observações por unidade.

Desvantagens:

- ▶ Requer supostos fortes sobre distribuições (U_{it} e V_{it}).
- ▶ Sensível à especificação da função de produção.

Eficiência em produção

```
clear all
set seed 54321
* Configuração: 20 firmas e 10 períodos de tempo
set obs 20
g firm_id = _n
expand 10
bys firm_id (firm_id): g year = _n
* Gerar inputs
g labor = rnormal(50, 10)
g capital = rnormal(100, 20)
* Gerar termo de ineficiência técnica ( $U_i \geq 0$ )
g inefficiency = abs(rnormal(1, 0.5))
bys firm_id (year): replace inefficiency = inefficiency[_n-1] if _n > 1
* Gerar termo de erro aleatório ( $V_{it}$ )
g error = rnormal(0, 0.2) // Erro aleatório
* Gerar output ( $Y_{it}$ ) com uma fronteira Cobb-Douglas
g ln_output = 2+.7*ln(labor)+.3*ln(capital)-inefficiency+error
g output = exp(ln_output)
* Criar dummies de tempo para capturar efeitos fixos no tempo
tab year, gen(time_)
keep firm_id year labor capital output
* Aplicar o comando xtfreier no Stata
g lnY = ln(output)
g lnL = ln(labor)
g lnK = ln(capital)
xtset firm_id year
xtfreier lnY lnL lnK, ti
xtfreier lnY lnL lnK, tvd
predict efficiency, te
kdensity efficiency
graph box efficiency, over(year, sort(ascending))
```

Eficiência em Custos

- ▶ A eficiência em custos mede a capacidade de uma unidade produtiva de minimizar seus custos para um dado nível de output.
- ▶ Compara os custos observados (C_{it}) com os custos mínimos (C_{it}^*).
- ▶ Essencial para:
 - ▶ Diagnosticar ineficiências.
 - ▶ Identificar economias de custo.
 - ▶ Melhorar a competitividade.

Função de Custo

A função de custo representa o custo mínimo necessário para produzir um dado nível de output:

$$C_{it}^* = f(Y_{it}, P_{it}; \beta),$$

onde:

- ▶ C_{it}^* : custo mínimo,
- ▶ Y_{it} : output produzido,
- ▶ P_{it} : vetor de preços dos insumos,
- ▶ β : parâmetros da função de custo.

O custo observado (C_{it}) pode ser maior que C_{it}^* devido a:

- ▶ **Ineficiência Técnica** (U_{it}): desvio sistemático,
- ▶ **Erro Aleatório** (V_{it}): choques não sistemáticos.

Modelo de Fronteira de Custo

A fronteira estocástica para custos é modelada como:

$$\ln C_{it} = f(\ln Y_{it}, \ln P_{it}; \beta) + U_{it} + V_{it},$$

onde:

- ▶ $f(\ln Y_{it}, \ln P_{it}; \beta)$: função de custo mínima (log-linearizada),
- ▶ $U_{it} \geq 0$: termo de ineficiência de custo,
- ▶ $V_{it} \sim N(0, \sigma_v^2)$: erro aleatório.

A eficiência de custo é calculada como:

$$EC_{it} = \exp(-U_{it}),$$

onde:

- ▶ $EC_{it} = 1$: eficiência perfeita (custo mínimo),
- ▶ $EC_{it} < 1$: presença de ineficiência de custo.

Distribuição de Ineficiência (U_{it})

- ▶ O termo U_{it} é não-negativo ($U_{it} \geq 0$) e segue uma distribuição:
 - ▶ Truncada normal: $U_{it} \sim \text{Truncated-Normal}(\mu, \sigma_u^2)$,
 - ▶ Exponencial: $U_{it} \sim \text{Exponential}(\lambda)$.
- ▶ V_{it} é simétrico, enquanto U_{it} induz assimetria em $\ln C_{it}$.
- ▶ A separação de U_{it} e V_{it} é feita pela maximização da função de verossimilhança.

Estimativa de Eficiência em Custos

A eficiência técnica condicional (U_{it}) é estimada usando a expectativa condicional:

$$\mathbb{E}[U_{it} \mid \epsilon_{it}] = \frac{\sigma_u^2}{\sigma^2} \left(\epsilon_{it} + \frac{\phi\left(\frac{\epsilon_{it}}{\sigma}\right)}{\Phi\left(-\frac{\epsilon_{it}}{\sigma}\right)} \right),$$

onde:

- ▶ $\epsilon_{it} = \ln C_{it} - f(\ln Y_{it}, \ln P_{it}; \beta),$
- ▶ $\sigma^2 = \sigma_u^2 + \sigma_v^2.$

A eficiência de custo é dada por:

$$EC_{it} = \exp(-U_{it}).$$

Aplicações e Interpretação

Aplicações Práticas:

- ▶ Análise de eficiência em setores industriais.
- ▶ Benchmarking entre empresas.
- ▶ Identificação de economias de escala e escopo.

Interpretação:

- ▶ $EC_{it} = 1$: Unidade opera no custo mínimo.
- ▶ $EC_{it} < 1$: Unidade apresenta custos excessivos em relação ao mínimo.
- ▶ Comparação entre EC_{it} ao longo do tempo identifica mudanças na eficiência.

Eficiência em custos

```
clear all
set seed 54321
* Configuração: 20 firmas e 10 períodos de tempo
set obs 20
gen firm_id = _n
expand 10
bys firm_id (firm_id): gen year = _n
* Gerar inputs
gen labor = rnormal(50, 10) // Trabalho
gen capital = rnormal(100, 20) // Capital
gen output = rnormal(200, 50) // Output produzido
* Gerar termo de ineficiência técnica (U_i >= 0)
gen inefficiency = abs(rnormal(0.1, 0.05)) // Ineficiência nos custos
bys firm_id (year): replace inefficiency = inefficiency[_n-1] if _n > 1
* Gerar termo de erro aleatório (V_it)
gen error = rnormal(0, 0.02) // Erro aleatório nos custos
* Preço dos inputs
gen price_labor = rnormal(20, 5) // Preço do trabalho
gen price_capital = rnormal(10, 2) // Preço do capital
* Função de custo Cobb-Douglas (C_it = f(Y, P_L, P_K) + U_i + V_it)
gen ln_cost = 2 + 0.6 * ln(output) + 0.3 * ln(price_labor) + 0.1 * ln(price_capital) + inefficiency + error
gen cost = exp(ln_cost)
```

```

* Criar dummies de tempo
tab year, gen(time_)
* Variáveis para o modelo
gen lnY = ln(output)
gen lnPL = ln(price_labor)
gen lnPK = ln(price_capital)
gen lnC = ln(cost)
* Manter variáveis relevantes
keep firm_id year lnC lnY lnPL lnPK
* Configurar painel
xtset firm_id year
* Estimar o modelo com ineficiência constante no tempo
xtfrontier lnC lnY lnPL lnPK, ti
* Estimar o modelo com ineficiência variável no tempo
xtfrontier lnC lnY lnPL lnPK, tvd
* Predizer eficiência de custo
predict efficiency_cost, te
graph box efficiency_cost, over(year, sort(ascending)) ///
    title("Eficiência de Custo por Ano") ytitle("Eficiência de Custo")

```

Acrescentando dinâmica

```
clear all
set seed 54321
set obs 20
gen firm_id = _n
expand 10
bys firm_id (firm_id): gen year = _n
gen labor = rnormal(50, 10)
gen capital = rnormal(100, 20)
gen output = rnormal(200, 50)
gen price_labor = rnormal(20, 5)
gen price_capital = rnormal(10, 2)
gen cost = exp(2+.6*ln(output)+.3*ln(price_labor)+.1*ln(price_capital)+abs(rnormal(0, 0.2)))
bysort firm_id (year): replace cost = exp(0.2 * ln(cost[_n-1]) + ///
    (2+.6*ln(output)+.3*ln(price_labor)+.1*ln(price_capital)+abs(rnormal(0,.2)))) ///
    if _n > 1
* Gerar log-transformações
gen lnC = ln(cost)
gen lnY = ln(output)
gen lnPL = ln(price_labor)
gen lnPK = ln(price_capital)
bysort firm_id (year): gen lnC_lag = lnC[_n-1] if _n > 1
* Configurar painel
xtset firm_id year
* Estimar modelo de fronteira de custo com lag
xtfrontier lnC lnY lnPL lnPK lnC_lag, ti
* Estimar modelo com ineficiência variável no tempo
xtfrontier lnC lnY lnPL lnPK lnC_lag, tvd
* Predizer eficiência de custo
predict efficiency_cost, te
```

Referências

- ANDERSON, THEODORE WILBUR, & HSIAO, CHENG. 1981. Estimation of dynamic models with error components. *Journal of the American statistical Association*, **76**(375), 598–606.
- ARELLANO, MANUEL, & BOND, STEPHEN. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, **58**(2), 277–297.
- ARELLANO, MANUEL, & BOVER, OLYMPIA. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, **68**(1), 29–51.
- HECKMAN, JAMES J. 1978. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, **5**(4), 475–492.
- KEANE, MICHAEL P, & WOLPIN, KENNETH I. 1997. The career decisions of young men. *Journal of political Economy*, **105**(3), 473–522.

- ORME, CHRIS. 2001. Two-Stage Estimation of Panel Data Models with Sample Selection. *Journal of Econometrics*, **114**(1), 121–139.
- WOOLDRIDGE, JEFFREY M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, **20**(1), 39–54.
- WOOLDRIDGE, JEFFREY M. 2010. *Econometric analysis of cross section and panel data*. MIT press.