# Modelos paramétricos lineares e não lineares de 1 equação com potencial dinâmica

Prof. Francis Petterini (PPGEco/UFSC)

f.petterini@ufsc.br

Dezembro de 2024

# Sumário por característica da variável dependente

1. Moldelos lineares

1.1 $y \in \mathbb{R}$

2. 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

# 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 $d \in \{0,1\}$

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

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

# Motivações para a dinâmica

1. Mitigar a endogeneidade sob a hipótese de que a variável defasada traz consigo um componente omitido que muda com o tempo, e então estimar consistentemente efeitos marginais $\partial E(y \mid x, d)/\partial x = \beta_2$ (elasticidades ou equivalente) ou $E(y \mid x, d = 1) - E(y \mid x, d = 0) = \beta_3$ (avaliação de políticas ou equivalente)

2. Projetar $y$ quando $T$ é pequeno e técnicas regulares de séries temporais não podem ser aplicadas, fazendo com que o pesquisador explore $n$.

# Um pouco mais sobre a função “projeção”

Makridakis & Hibon (2000), Diebold & Mariano (2002), Stock & Watson (2007), Campbell & Thompson (2008), Welch & Goyal (2008), entre outros, mostram que o AR(1) muitas vezes supera modelos mais sofisticados em termos de projeção.

$$\begin{cases} y_{11} & = \gamma y_{10} + c_1 + u_{11} \\ & \vdots \\ y_{1T} & = \gamma y_{1(T-1)} + c_1 + u_{1T} \\ & \vdots \\ y_{n1} & = \gamma y_{n0} + c_n + u_{n1} \\ & \vdots \\ y_{nT} & = \gamma y_{n(T-1)} + c_n + u_{nT} \end{cases}$$

# Breve revisão da alfabetização

## $\gamma = 0$ e $T = 1$

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

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

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

# MM

- 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 $n^{-1}X'U = \begin{bmatrix} \sum_{i=1}^{n} \frac{u_i}{n} \\ \sum_{i=1}^{n} \frac{x_i u_i}{n} \\ \sum_{i=1}^{n} \frac{d_i u_i}{n} \end{bmatrix}$
- $Y = X\beta + U \Rightarrow X'Y = X'X\beta + X'U \Rightarrow X'Y = X'X\beta + n/nX'U$
- $X'Y \overset{n\rightarrow\infty}{=} X'X\hat{\beta}_{mm} + \mathbf{0} \Rightarrow (X'X)^{-1}X'Y = (X'X)^{-1}X'X\hat{\beta}_{mm} \Rightarrow \hat{\beta}_{mm} = \hat{\beta}_{ols}$

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

# Base de dados 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]^.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 $\beta$ se $\lim_{n\to\infty} \Pr(||\hat{\beta} - \beta|| < \epsilon) = 1$. Nesse caso se escreve 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, 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".

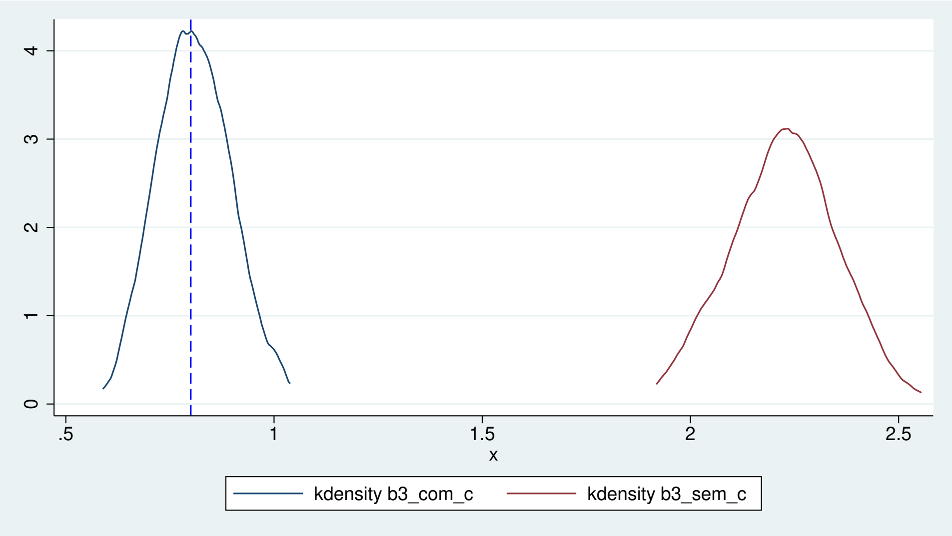
# Razões da endogeneidade

- **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)
```



0
1
2
3
4
.5
1
1.5
2
2.5
x
kdensity b3_com_c
kdensity b3_sem_c

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.

- $\gamma = 0$ e $T > 1$

$$\begin{cases} y_{11} & = \beta_1 + \beta_2 x_{11} + \beta_3 d_{11} + c_1 + \varepsilon_{11} \\ y_{12} & = \beta_1 + \beta_2 x_{12} + \beta_3 d_{12} + c_1 + \varepsilon_{12} \\ & \vdots \\ y_{1T} & = \beta_1 + \beta_2 x_{1T} + \beta_3 d_{1T} + c_1 + \varepsilon_{1T} \\ & \vdots \\ y_{n1} & = \beta_1 + \beta_2 x_{n1} + \beta_3 d_{n1} + c_n + \varepsilon_{n1} \\ y_{n2} & = \beta_1 + \beta_2 x_{n2} + \beta_3 d_{n2} + c_n + \varepsilon_{n2} \\ & \vdots \\ y_{nT} & = \beta_1 + \beta_2 x_{nT} + \beta_3 d_{nT} + c_n + \varepsilon_{nT} \end{cases}$$

# A abordagem da diferença da média (FE)

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

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

# A abordagem da primeira diferença (FD)

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

```
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:

- $u_{it} = c_i + \varepsilon_{it} \Rightarrow \sigma_u^2 = \sigma_c^2 + \sigma_\varepsilon^2 \Rightarrow UU'_{n=3,T=2} = \begin{bmatrix} 2.9 & 0.4 & 0 & 0 & 0 & 0 \\ 0.4 & 2.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.9 & 0.4 & 0 & 0 \\ 0 & 0 & 0.4 & 2.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.9 & 0.4 \\ 0 & 0 & 0 & 0 & 0.4 & 2.9 \end{bmatrix}$
- 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 | pols | fe | fd | re | ivfe |
|---|---|---|---|---|---|---|
| | b/se | b/se | b/se | b/se | b/se | b/se |
| x | 0.559*** | 0.559*** | 0.523*** | | 0.524*** | 0.520*** |
| | (0.030) | (0.030) | (0.015) | | (0.016) | (0.018) |
| d | 3.548*** | 3.548*** | 0.808*** | | 0.909*** | 0.101 |
| | (0.191) | (0.191) | (0.117) | | (0.120) | (1.731) |
| D.x | | | | 0.514*** | | |
| | | | | (0.015) | | |
| D.d | | | | 0.735*** | | |
| | | | | (0.112) | | |
| _cons | -2.880*** | -2.880*** | -0.658*** | | -0.738* | -0.136 |
| | (0.314) | (0.313) | (0.166) | | (0.333) | (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

    chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
    =  57.43
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 $\not\forall i$ |
| Fisher | Raiz unitária $\forall i$ | Estacionariedade $\not\forall i$ |

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

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

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

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)...

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

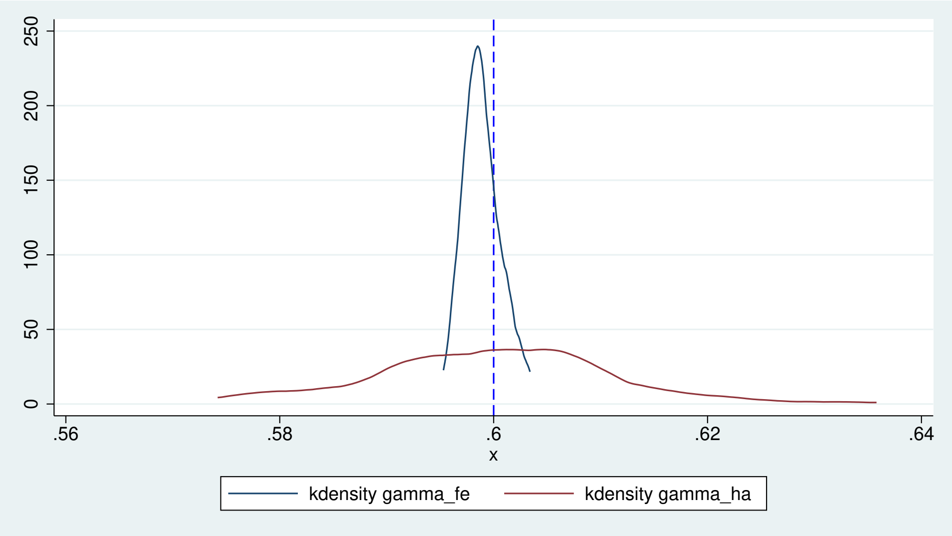
- $\Delta y_{it-1} = y_{it-1} - y_{it-2}$ é correlacionado com o termo de erro $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$, por conta da contemporaneidade em $t-1$.
- Mas $y_{it-2}$ pode ser usado como instrumento para $\Delta y_{it-1}$, já que $y_{it-2}$ não é correlacionado com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



250
200
150
100
50
0
.56
.58
.6
.62
.64
x
kdensity gamma_fe
kdensity gamma_ha

# 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 $T = 4$

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

- São empreendidos 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 $\Delta y_{i,t-2}, \Delta y_{i,t-3} \ldots$, já que essas difrenças 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
```

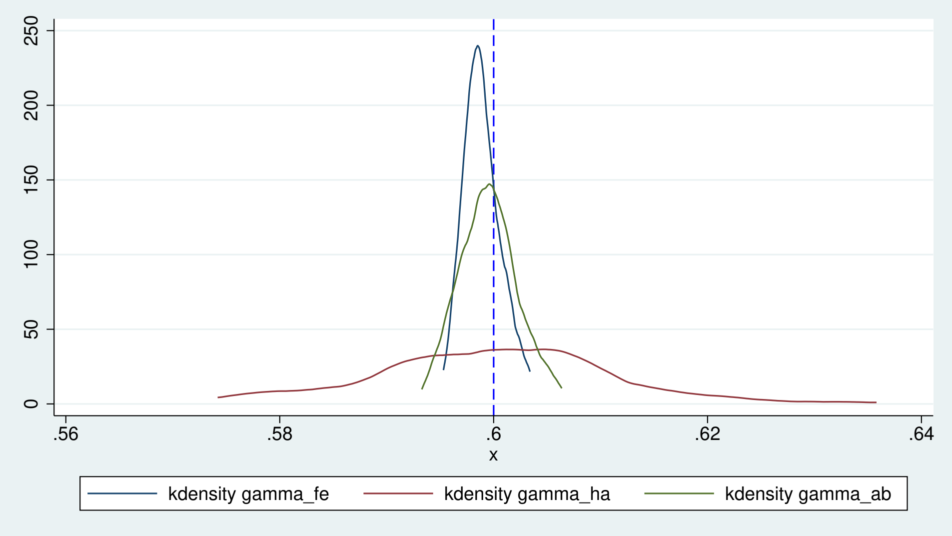
```
xtreg y L.y x, fe
est sto fe

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: R2, 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

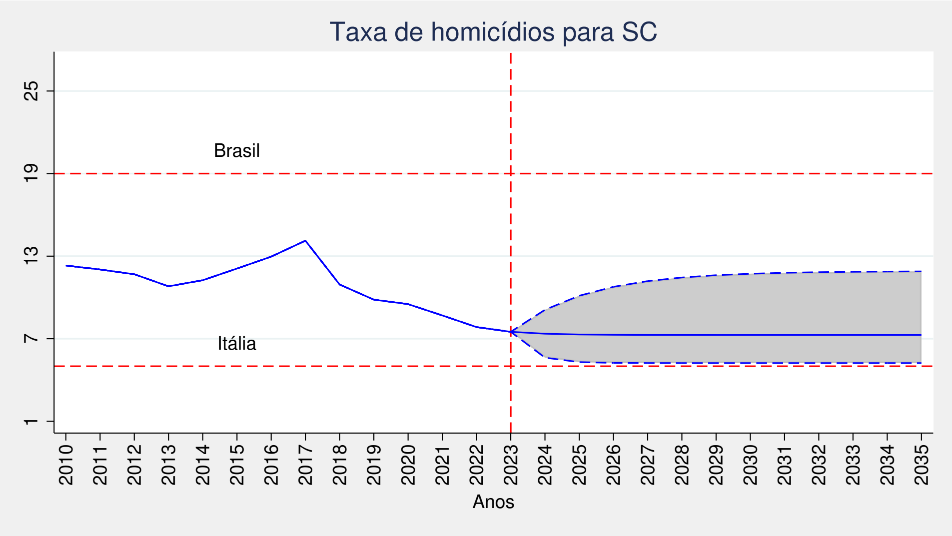
```
cls
clear all
import excel "https://petterini.ufsc.br/files/2024/11/data_fln_filtrada.xlsx", firstrow
g tri = qofd(t)
form tri %tq
replace t = tri
drop tri
*g lnw = ln(w/((30/7)*h))
g lnw = ln(w)
g idade2 = idade^2
global X edu idade idade2 homem branco
reg lnw $X if cond_dom == 1
est sto ols
xtset 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
25
19
13
7
1
Brasil
Itália
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
Anos

# 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 projecoes
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
}
```

# Modelos não lineares

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

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

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

# Breve revisão da alfabetização

# Probit e Logit se $\gamma = 0$ e $T = 1$... 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

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

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

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

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

$$\text{Pseudo-}R^2 = 1 - \ln L(\hat\beta)/\ln L(\hat\beta_0)$$

```
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 $T > 1$

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

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

$$P(\boldsymbol{y}_i = (0,0)) = \frac{1}{1 + \exp(\beta' \boldsymbol{x}_{1t} + c_i)} \times \frac{1}{1 + \exp(\beta' \boldsymbol{x}_{2t} + c_i)}$$

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

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

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

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

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

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

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

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

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

```
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 $g : \mathbb{R} \to \mathbb{R}$.

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

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

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

$$\hat{\mathrm{Var}}(g(\hat{\beta})) \approx \nabla g(\hat{\beta})' \hat{\mathrm{Var}}(\hat{\beta}) \nabla g(\hat{\beta})$$

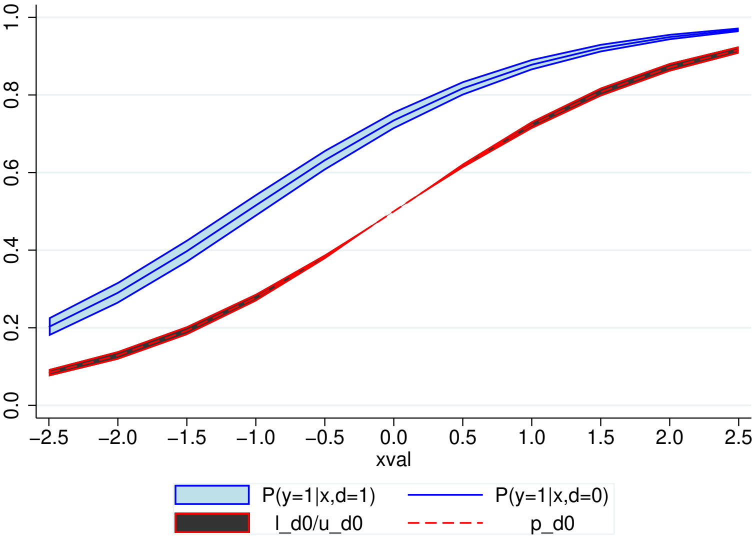
$$\Pr(y = 1 \mid x, d = 1) = g(\beta) = \Lambda(\beta_1 x + \beta_2) \Rightarrow \nabla g(\hat{\beta}) = \begin{bmatrix} \Lambda(\hat{\beta}_1 + \hat{\beta}_2)(1 - \Lambda(\hat{\beta}_1 + \hat{\beta}_2))x \\ \Lambda(\hat{\beta}_1 + \hat{\beta}_2)(1 - \Lambda(\hat{\beta}_1 + \hat{\beta}_2)) \end{bmatrix}$$

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

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

```
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))
```



0.0
0.2
0.4
0.6
0.8
1.0
–2.5
–2.0
–1.5
–1.0
–0.5
0.0
0.5
1.0
1.5
2.0
2.5
xval
P(y=1|x,d=1)
P(y=1|x,d=0)
l_d0/u_d0
p_d0

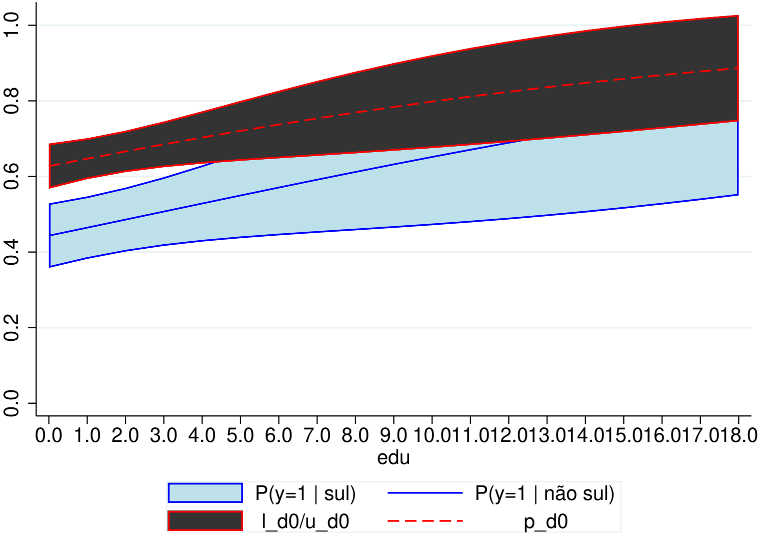
# union.dta

```
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))
```


1.0
0.8
0.6
0.4
0.2
0.0
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0
edu
P(y=1 | sul)
P(y=1 | não sul)
l_d0/u_d0
p_d0

# Acrescentando a dinâmica

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

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

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

$$\Pr(y_1 = 0, y_2 = 0, y_3 = 0) = \Pr(\varepsilon_1 \le 0, \varepsilon_2 \le 0, \varepsilon_3 \le 0) = \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{-\infty}^{0} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 0, y_2 = 0, y_3 = 1) = \Pr(\varepsilon_1 \le 0, \varepsilon_2 \le 0, \varepsilon_3 > 0) = \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{0}^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 0, y_2 = 1, y_3 = 0) = \Pr(\varepsilon_1 \le 0, \varepsilon_2 > 0, \varepsilon_3 \le 0) = \int_{-\infty}^{0}\int_{0}^{\infty}\int_{-\infty}^{0} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 0, y_2 = 1, y_3 = 1) = \Pr(\varepsilon_1 \le 0, \varepsilon_2 > 0, \varepsilon_3 > 0) = \int_{-\infty}^{0}\int_{0}^{\infty}\int_{0}^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 1, y_2 = 0, y_3 = 0) = \Pr(\varepsilon_1 > 0, \varepsilon_2 \le 0, \varepsilon_3 \le 0) = \int_{0}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{0} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 1, y_2 = 0, y_3 = 1) = \Pr(\varepsilon_1 > 0, \varepsilon_2 \le 0, \varepsilon_3 > 0) = \int_{0}^{\infty}\int_{-\infty}^{0}\int_{0}^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 1, y_2 = 1, y_3 = 0) = \Pr(\varepsilon_1 > 0, \varepsilon_2 > 0, \varepsilon_3 \le 0) = \int_{0}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{0} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3,$$

$$\Pr(y_1 = 1, y_2 = 1, y_3 = 1) = \Pr(\varepsilon_1 > 0, \varepsilon_2 > 0, \varepsilon_3 > 0) = \int_{0}^{\infty}\int_{0}^{\infty}\int_{0}^{\infty} f(\varepsilon_1, \varepsilon_2, \varepsilon_3)\, d\varepsilon_1\, d\varepsilon_2\, d\varepsilon_3.$$

# Soluções mais populares

- FIML é complicado computacionalmente:
  $P(\{\boldsymbol{y}_i\}_{i=1}^{T} \mid \text{todos } y_{i0}, x_{it}) = \int \prod_{i=1}^{n} \left[ \int_{\boldsymbol{y}_i} f(\boldsymbol{\varepsilon}_i(\textbf{parâmetros}))d\boldsymbol{\varepsilon} \right] f(c)dc$
- Heckman (1978) (Orme (2001) melhora) trata parametricamente $y_0$, e mantém $c$ como um efeito aleatório, mas ainda é complicado computacionalmente:
  $P(\{\boldsymbol{y}_i\}_{i=1}^{T} \mid \text{todos } y_{i0}, x_{it}) = \int \prod_{i=1}^{n} \Pr(y_{1i}, ..., y_{iT} \mid \text{controles de } y_{0i}) f(c)dc$
- 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).

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

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

# 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
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*

toc
```

# 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)

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

# A verossimilhança padrão

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

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

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

# Tobit

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

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

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

$$\cdot \left[ \Phi\left( -\frac{\beta' \boldsymbol{x}_i + c_i}{\sigma} \right) \right]^{1 - I(y_{it} > 0)} f(c_i) \, dc_i$$

# Tobit-Wooldridge

```
clear all
set seed 54321
loc n = 100
loc T = 10
loc reps = 50
mat m_tobit = 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
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_wooldr[`rep',1] = _b[L.y]
mat m_wooldr[`rep',2] = _b[x]
}
toc
svmat m_tobit, names(to)
svmat m_wooldr, 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.

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

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

$$\mathcal{L}(\beta, \sigma, \sigma_c) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[ f(y_i - \beta' \boldsymbol{x}_i - c_i) \right]^{I(y_i < \tau_i)}$$

$$\cdot \left[ 1 - F\left( \tau_i - \beta' \boldsymbol{x}_i - c_i \right) \right]^{1 - I(y_i < \tau_i)} f(c_i)\, dc_i$$

# LogNormal

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

- fdp

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

- cdf

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

- esperança

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

- função quantílica

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

# Weibull

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

- fdp

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

- cdf

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

- esperança

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

- função quantílica

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

# 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 tserved 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 $y^* = E(y \mid \text{controles})$

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

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

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

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

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

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:
  - $\lambda_i$: Média esperada do número de eventos.
  - $X_i$: Vetor de covariáveis.
  - $\beta$: 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:

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

  Onde:
  - $\mu_i$: Média esperada.
  - $\theta$: 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:
  - $\pi_i$: 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

**Critérios para Escolher o Modelo:**

- **Superdispersão**:
  - Compare variância e média da variável dependente.
- **Excesso de Zeros**:
  - Compare a proporção de zeros observados com o esperado.
- **Testes Estatísticos**:
  - **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_wooldr = 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_wooldr[`rep',1] = _b[L.y]
mat m_wooldr[`rep',2] = _b[x]
}
toc
svmat m_poisson, names(po)
svmat m_wooldr, 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.
- $\mu_{it}$: valor esperado de $Y_{it}$.
- $X_{it}\beta$: combinação linear das covariáveis.
- $\alpha_i$: efeito fixo específico da unidade $i$.
- $\ln(\text{pop}_{it})$: offset (log da exposição $\text{pop}_{it}$).

# 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 $\mu_{it}$:

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

Problema: Não podemos estimar $\alpha_i$ diretamente devido ao problema dos parâmetros incidentais.

# Condicionalização 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 $\alpha_i$:

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

Aqui, $\alpha_i$ desaparece porque afeta $\mu_{it}$ de forma multiplicativa.

# Função de Verossimilhança Condicional

A função de verossimilhança condicional é:

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

Simplificando:

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

# Interpretação

- A condicionalização elimina $\alpha_i$, permitindo estimar $\beta$ sem viés.
- O denominador, que depende de todas as combinações possíveis de $Y_{it}$ que somam $S_i$, é calculado numericamente.
- Os coeficientes $\beta$ 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:

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

   onde:
   - $\boldsymbol{x}_{ij}$: vetor de características observáveis da alternativa $j$ e do indivíduo $i$,
   - $\beta_j$: vetor de coeficientes estimáveis específicos da alternativa $j$,
   - $\xi_{ij}$: componente de erro (não observado).

2. **Assumindo Distribuição Gumbel para $\xi_{ij}$:**
   - Para simplificar a análise, assume-se que os erros $\xi_{ij}$ 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:

$$\Pr(y_i = j) = \Pr(\beta_j' \boldsymbol{x}_{ij} + \xi_{ij} > \beta_k' \boldsymbol{x}_{ik} + \xi_{ik}, \forall k \neq j).$$

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

$$\Pr(y_i = j) = \frac{\exp(\beta_j' \boldsymbol{x}_{ij})}{\sum_{k=1}^{J} \exp(\beta_k' \boldsymbol{x}_{ij})}.$$

**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:

$$\Pr(y_i = 1) = \frac{1}{1 + \sum_{k=2}^{J} \exp(\beta_k' \boldsymbol{x}_{ij})},$$

$$\Pr(y_i = j) = \frac{\exp(\beta_j' \boldsymbol{x}_{ij})}{1 + \sum_{k=2}^{J} \exp(\beta_k' \boldsymbol{x}_{ij})}, \quad j = 2, \ldots, J.$$

# Modelo Logit Multinomial com Efeitos Fixos

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

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

com:

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

onde:

- $X_{ijt}$: covariáveis específicas da alternativa $j$ no tempo $t$.
- $\beta$: coeficientes estimáveis.
- $\alpha_{ij}$: 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 $\alpha_{ij}$ a estimar ($N \times J$).
- Viés dos parâmetros incidentais.

# Condicionalização para Eliminar $\alpha_{ij}$

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

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

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

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

# 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 $\alpha_{ij}$.

# Interpretação e Implementação

- Os efeitos fixos $\alpha_{ij}$ desaparecem devido à condicionalização.
- A verossimilhança condicional permite estimar $\beta$ 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,
- $\beta$: vetor de coeficientes estimáveis,
- $\epsilon_i$: termo de erro.

**Categorias Observáveis:**

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

- $\mu_j$: pontos de corte estimados, separando as categorias de $y_i$.

# Diferença entre Logit e Probit Ordenado

**Distribuição do Termo de Erro $\epsilon_i$:**

- **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

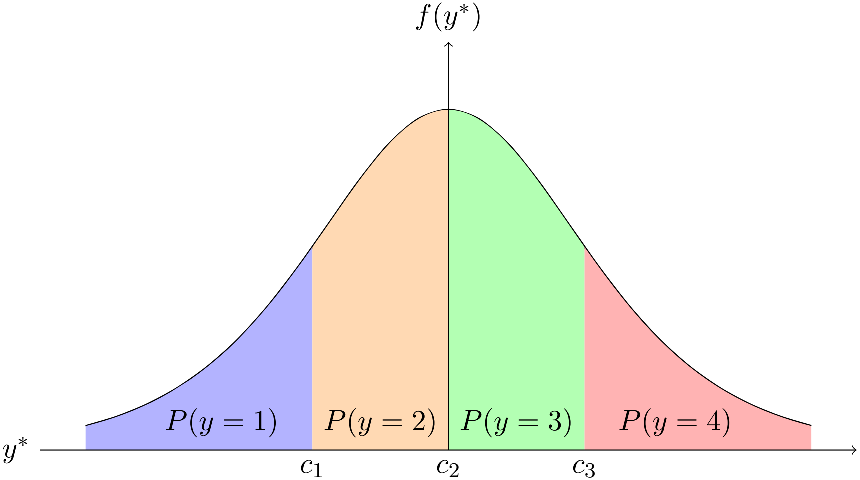
**Pontos de Corte ($\mu_j$):**

- Os pontos de corte são estimados junto com os coeficientes $\beta$.
- 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 $\mu_j$ e $\mu_{j-1}$.


$f(y^*)$
$P(y=1)$
$P(y=2)$
$P(y=3)$
$P(y=4)$
$y^*$
$c_1$
$c_2$
$c_3$

# Interpretação de $\beta$

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

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

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 $\epsilon_{it}$:

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

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 >= 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 xtfrontier no Stata
g lnY = ln(output)
g lnL = ln(labor)
g lnK = ln(capital)
xtset firm_id year
xtfrontier lnY lnL lnK, ti
xtfrontier 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,
- $\beta$: 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.

Campbell, John Y, & Thompson, Samuel B. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, **21**(4), 1509–1531.

Diebold, Francis X, & Mariano, Robert S. 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, **20**(1), 134–144.

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.

Makridakis, Spyros, & Hibon, Michele. 2000. The M3-Competition: results, conclusions and implications. *International journal of forecasting*, **16**(4), 451–476.

Orme, Chris. 2001. Two-Stage Estimation of Panel Data Models with Sample Selection. *Journal of Econometrics*, **114**(1), 121–139.

Stock, James H, & Watson, Mark W. 2007. Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, **39**, 3–33.

Welch, Ivo, & Goyal, Amit. 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, **21**(4), 1455–1508.

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.