Esta aula

- Plano
  - Panel Data
  - Serial Correlation

- Bibliografia
  - Wooldridge, J. M. Introductory Econometrics: A modern Approach, 6th Ed.

Panel Data

Panel Data

- *Panel Data* ou Dados em Painel ocorre quando temos dados das mesmas unidades cross-section em diferentes momentos do tempo
- Diferentemente, quando há dados para distintas unidades cross-sections em diferentes momentos do tempo, temos o que se denomina *Pooled Cross-Sections*.

Panel Data: Difference-in-Differences (DiD)

- Imagine que pudessemos fazer um experimento, atribuindo aleatoriamente uma unidade cross-section ao grupo de tratamento B e a outra ao grupo de controle A. Suponha também que pudessemos observar as duas unidades antes (t=1) e após o tratamento (t=2).
- O efeito do tratamento poderia ser calculado como:

$$(y_{B,2} - y_{A,2}) - (y_{B,1} - y_{A,1})$$

esse é o estimador DiD.

Panel Data: Difference-in-Differences (DiD)

- Podemos encontrar o estimador DiD através de uma regressão com duas dummies: a primeira que toma o valor unitário após o tratamento e a segunda que toma o valor unitário para o grupo de tratamento
- Considere o model: $y_{it} = \beta_0 + \beta_1 treatment_{it} + \beta_2 after_{it} + \beta_3 treatment_{it} * after_{it} + u_{it}$
- β_3 será o DiD para os grupos
- *Vantagem da regressão: permite adicionar variáveis de controle relacionadas às diferenças entre os grupos*

Panel Data

- Dados em painel nos permitem corrigir viés devido a omissão de variáveis que caracterizam cada unidade cross-section.
- Suponha que o modelo populacional seja:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}$$

em que temos um componente de erro constante no tempo a_i , $v_{it} = a_i + u_{it}$

- Se a_i for correlacionado com as variáveis explicativas, OLS será enviesado.

Panel Data

- Nesse caso, podemos corrigir o problema e subtraindo o valor de cada variável no período t pelo seu valor no período anterior t-1, obtendo-se:

$$\Delta y_i = \beta_1 \Delta x_{i1} + \ldots + \beta_k \Delta x_{ik} + \Delta u_i$$

- Se u_{it} não for correlacionado com as variáveis explicativas, OLS não implicará em viés na estimação. Esse é o estimador das primeiras diferenças ou *First-Differences Estimator* (FD)

Panel Data

- Ao invés do estimador FD, podemos utilizar o estimador de efeitos fixos ou *Fixed-Effects Estimator* (FE). Se tomarmos a média temporal de cada variável para cada unidade cross section:

$$y_i = \frac{\sum_t y_{it}}{T}; \quad x_i = \frac{\sum_t x_{it}}{T};$$

Note que a média de a_i é equivalente ao próprio a_i.

- Dessa forma, obtemos a equação para a média: $y_i = \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + a_i + u_i$

Panel Data

- Equação para a média:

$$y_i = \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + a_i + u_i$$

- Retirando-se a média do modelo original, obtém-se:

$$y_{it} - y_i = \beta_1 x_{it1} - x_{i1} + \ldots + \beta_k x_{itk} - x_{ik} + u_{it} - u_i$$

- Se u_{it} não for correlacionado com as variáveis explicaticas, OLS não implicará em viés na estimação. Esse é o estimador de efeitos fixos ou *Fixed-Effects Estimator* (FE)
-

Panel Data

- Novamente, considere o modelo:

  $$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots \beta_k x_{itk} + a_i + u_{it}$$

-
- Anteriormente, assuminos que a_i era correlacionado com as variáveis explicativas.
- Se não for correlacionado, não precisamos eliminá-lo por FD ou FE.
- Nesse caso, OLS na equação original ainda seria consistente para os betas, porém o termo de erro seria serialmente correlacionado e as estimativas dos erros padrões enviesadas.

Panel Data

- Há uma transformação possível para corrigir as inferências e melhorar a eficiência na estimação:

$$\lambda = 1 - \left[\sigma_u^2 \Big/ \left(\sigma_u^2 + T\sigma_a^2\right)\right]^{1/2}$$

$$y_{it} - \lambda \bar{y}_i = \beta_0\left(1 - \lambda\right) + \beta_1\left(x_{it1} - \lambda \bar{x}_{i1}\right) + \ldots + \beta_k\left(x_{itk} - \lambda \bar{x}_{ik}\right) + \left(v_{it} - \lambda \bar{v}_i\right)$$

esse é o estimador de efeitos aleatórios ou Random-Effects (RE)

Panel Data

- Se $\lambda = 1$, esse é o estimador FE
- Se $\lambda = 0$, esse é o estimador OLS
- RE é uma média ponderada de OLS e FE.
- Quanto maior a variância de a_i, mais próximo RE é de FE.
- Quanto menor a variância de a_i, mais próximo RE é de OLS.

Panel Data

- Podemos ainda testar pela presença de heterocedastidade e correlação serial, corrigindo por ambas.

Serial Correlation

Serial Correlation

- Considere o modelo: $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$
- Queremos testar a hipótese nula de ausência de correlação serial $\rho = 0$ em $u_t = \rho u_{t-1} + e_t$,
- *Se as variáveis explicativas forem estritamente exógenas, faça a regressão dos resíduos contra os seus valores defasados e aplique um teste t.*
- *Se não forem estritamente exógenas, faça a regressão:*
- $u_t = \rho u_{t-1} + \alpha_1 x_{t1} + \ldots + \alpha_k x_{tk} + e_t$ e *aplique um teste t em* ρ

Serial Correlation

- Pode-se testar pela presença de correlação serial de ordens maiores também, incluindo-se as demais defasagens do resíduo:

$$u_t = \rho_1 u_{t-1} + \cdots . + \rho_5 u_{t-5} + e_t$$

E aplicando-se um teste F em todos os ρ.

Serial Correlation

- Assuma que todos as variáveis explicativas sejam estritamente exógenas e que os erros sigam um processo AR(1) $u_t = \rho u_{t-1} + e_t$, $t = 2, \ldots, n$
- $\mathrm{Var}(u_t) = \sigma^2_e / (1 - \rho^2)$

Serial Correlation

- Considere, por exemplo, o modelo $y_t = \beta_0 + \beta_1 x_t + u_t$, então $y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1}$
- Multiplicando-se a segunda equação por ρ, e subtraindo-se da primeira:
- $y_t - \rho\, y_{t-1} = (1 - \rho)\beta_0 + \beta_1(x_t - \rho\, x_{t-1}) + e_t$
- Em que e_t não apresenta correlação serial
- Estime ρ através da regressão: $u_t = \rho u_{t-1} + e_t$
- E aplique OLS em $y_t - \rho\, y_{t-1} = (1 - \rho)\beta_0 + \beta_1(x_t - \rho\, x_{t-1}) + e_t$
- Esse seria um FGLS! (Cochrane-Orcutt ou Prais-Winsten)

Serial Correlation

- FGLS é apenas aplicavel com variáveis explicativas estritamente exógenas.
- Ao invés de FGLS, para todos os casos, podemos calcular erros padrões robustos a serial correlation. (Newey-West Standard Errors (comando *newey* no Stata)

Obrigada!