

3ª Lista de Exercícios

Professor: Rafael Terra
Inferência Causal (Econometria SP)

19 de janeiro de 2025

Responda as questões de 1 a 4 abaixo.

1. Considere que o efeito do tratamento de um programa de distribuição de material didático à escolas públicas realizado em 2006 sobre a proficiência média das escolas em leitura e matemática se baseie na seguinte relação

$$\begin{aligned} \text{proficiência}_{it} = & \alpha + \beta_1 \cdot \text{grupo_tratado}_i \cdot \text{d.2007}_t + \beta_2 \cdot \text{grupo_tratado}_i + \\ & + \beta_3 \cdot \text{d.2007}_t + \varepsilon_i \end{aligned} \quad (1)$$

em que a variável “grupo_tratado” é uma dummy de escolas tratadas independentemente do ano. A variável “d.2007” é a variável dummy que indica o ano de 2007 (igual a 0 caso contrário), e a variável “tratamento” é uma dummy de interação entre as duas variáveis anteriores. Use o arquivo “banco_diff_diff.dta”.

- (a) Estime por OLS e interprete os coeficientes dos efeitos do tratamento sobre as proficiências de leitura e matemática. O programa foi efetivo?
- (b) Informe a inclinação dos grupos de escolas tratadas e de controle e seus interceptos.
- (c) Estime o seguinte modelo por efeitos fixos

$$\begin{aligned} \text{proficiência}_{it} = & \alpha + \beta_1 \cdot \text{tratamento}_{it} + \\ & + \beta_3 \cdot \text{d.2007}_t + \beta_{2i} + \varepsilon_{it} \end{aligned} \quad (2)$$

Use a função `plm[, type=“within”]`, e compare os coeficientes com o modelo estimado na equação 1. Existe alguma diferença entre os dois procedimentos?

- (d) Se utilizássemos o método de primeiras diferenças obteríamos resultados diferentes?
- (e) Suponha que você tivesse dados para o ano de 2003. Se adicionássemos os dados

desse ano ao nosso banco e na equação 1 inseríssemos uma dummy para o ano de 2005 e a interação desta com “grupo_tratado_i”, qual seria a interpretação de um coeficiente $\hat{\beta}_0 = 10$ com erro-padrão $s.e.(\hat{\beta}) = 3$?

2. Considere a existência de Gileade do Sul, um país que passou por uma pandemia de Covid19 no último ano. Após um ano, e o fim da pandemia, você foi contratado para avaliar o impacto do “Distanciamento Social” (social_distancing) sobre a taxa de mortalidade do vírus por milhão de habitantes (covid_fatality_pm). O distanciamento social é endógeno, pois se correlaciona com uma variável omitida que afeta a mortalidade e o distanciamento social (percentual de pobreza no município). Você dispõe de duas candidatas a variáveis instrumentais: 1) uma variável binária referente a uma campanha de conscientização sobre o distanciamento realizada em municípios sorteados (campaign) e , 2) outra variável referente a implantação do lockdown durante a pandemia (lockdown). Responda.
 - (a) Analise a variável “campaign” e responda qual é o percentual aproximado de compliers?
 - (b) Estime um modelo por 2SLS com dois instrumentos. Interprete os coeficientes de 1º e 2º estágios e os resultados dos testes.
 - (c) Considerando os resultados, proponha um modelo com instrumentos válidos. Qual a direção do viés do modelo em (b)? Explique a lógica por trás desse resultado.
 - (d) Desenhe um DAG que represente o modelo estrutural desse problema com as setas apontando na direção da causalidade.
3. Um programa de treinamento realizado pelo governo federal visa qualificar pessoas de baixa renda para reduzir desigualdes socioeconômicas. O programa distribui *vouchers* para trabalhadores classificados como pobres (variável “poverty_index”>50). O programa de qualificação não impede que outras pessoas que não receberam o voucher participem do treinamento, e nem obriga os recipientes do voucher a participarem do programa. Após um ano da distribuição dos vouchers foram coletadas informações sobre renda do trabalho. Usando a base de dados “Training_voucher.dta” e as libraries rdrobust, rdplot, rddensity, rdplotdensity e AER no R, responda:
 - (a) Qual é o desenho mais adequado para avaliar o efeito do programa sobre o rendimento?
 - (b) Usando a função rdplot estime e interprete o gráfico que mostre o salto na probabilidade de participação no programa ao redor do threshold de pobreza.

- (c) Usando a função `rdplot`, estime e interprete o gráfico que mostre se há salto na variável de renda pós-tratamento ao redor do threshold de pobreza. O que representa a diferença de salários no threshold de pobreza?
 - (d) Usando a função `rdrobust` estime e interprete o coeficiente que mostre esse salto na probabilidade de participação no programa ao redor do threshold de pobreza. O que o teste informa?
 - (e) No último gráfico, qual o polinômio que melhor se ajusta a relação entre a forcing variable e a renda pós-tratamento?
 - (f) Usando a função `rdrobust` estime e interprete o coeficiente que mostra se há diferença no valor esperado da variável dependente ao redor do threshold de pobreza.
 - (g) Faça o teste de diferença de densidade ao redor do threshold usando a função `rddensity`. Qual o propósito do teste? O que o resultado obtido significa?
 - (h) Estime e interprete o gráfico do teste de diferença de densidade ao redor do threshold usando a função `rdplotdensity`.
 - (i) Usando a função `rdplot` estime e interprete o gráfico que mostre se há salto na variável de renda pré-tratamento e na habilidade ao redor do threshold de pobreza. O que podemos concluir com esses testes de falsificação?
 - (j) Usando a função `rdrobust` estime e interprete o coeficiente que mostre se há diferença no valor esperado da variável de renda pré-tratamento ao redor do threshold de pobreza.
 - (k) Reestime o item (f) e interprete usando variáveis instrumentais e os polinômios na forcing variable (e interações) que ache adequados. Use todas as observações ao redor do threshold.
 - (l) Reestime o item (f) e interprete usando variáveis instrumentais via regressão local para tamanhos de banda $h=1, 2.5$ e 5 . Use polinômios somente de primeira ordem na forcing variable. Por que não necessitamos de ajustar polinômios de ordem mais alta quando trabalhamos com bandas pequenas ao redor do threshold?
4. Usando o banco de dados `JTRAIN3.dta`, referente a um programa de treinamento
- (a) Supondo ignorabilidade forte do tratamento, usando a função `rq` no R estime e interprete o efeito do tratamento nos quantis de rendimento em 78

$(\tau = 0.1, 0.25, 0.5, 0.75, 0.9)$, dado pelo seguinte modelo populacional.

$$Q_\tau(y|X) = \alpha(\tau) + X_i\beta(\tau) \quad (3)$$

- (b) Desenhe e interprete a distribuição dos resultados potenciais estimados com base nos efeitos estimados nos quantis. Use a função `rq` no R junto com a função `plot`, como no script da aula prática.
- (c) Relaxe a hipótese de ignorabilidade forte do tratamento. Use o banco de dados `lista3ex4itc.csv`. Supondo que a variável “Z_voucher” corresponda ao sorteio do voucher de treinamento, e “treat_itemC” a variável de treinamento propriamente dita, estime e interprete o efeito nos quantis $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ de “re78_c” usando o peso “kappa de Abadie” - especificamente a fórmula (7.2.5, rodapés 7 e 8 no livro).