

DIFERENÇAS EM DIFERENÇAS

Luiz Guilherme Scorzafave



FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
UNIVERSIDADE DE SÃO PAULO

BIBLIOGRAFIA

- ABADIE, Alberto; DIAMOND, Alexis; HAINMUELLER, Jens. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. **Journal of the American Statistical Association**, 2010.
- ANGRIST, Joshua D.; PISCHKE, Jörn-Steffen. **Mostly harmless econometrics: An empiricist's companion**. Princeton university press, 2008, cap. 5 e cap. 8
- BERTRAND, Marianne; DUFLO, Esther; MULLAINATHAN, Sendhil. How Much Should We Trust Differences-in-Differences Estimates?. **Quarterly journal of economics**, v. 119, n. 1, p. 249-275, 2004.
- BESLEY, Timothy; BURGESS, Robin. Can Labor Regulation Hinder Economic Performance? Evidence from India. **The Quarterly Journal of Economics**, v. 119, n. 1, p. 91-134, 2004.



BIBLIOGRAFIA

- BESLEY, Timothy; BURGESS, Robin. Can Labor Regulation Hinder Economic Performance? Evidence from India. **The Quarterly Journal of Economics**, v. 119, n. 1, p. 91-134, 2004.
- CARD, David. Using regional variation in wages to measure the effects of the federal minimum wage. **Industrial & Labor Relations Review**, v. 46, n. 1, p. 22-37, 1992.
- CARD, David; KRUEGER, Alan B. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: reply. **The American Economic Review**, v. 90, n. 5, p. 1397-1420, 2000.
- CARD, David; KRUEGER, Alan B. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. **The American Economic Review**, v. 84, n. 4, p. 772-793, 1994.



BIBLIOGRAFIA

- HIRANO, Keisuke; IMBENS, Guido W.; RIDDER, Geert. Efficient estimation of average treatment effects using the estimated propensity score. **Econometrica**, v. 71, n. 4, p. 1161-1189, 2003.
- PISCHKE, Jörn-Steffen. The impact of length of the school year on student performance and earnings: Evidence from the German short school years. **The Economic Journal**, v. 117, n. 523, p. 1216-1242, 2007.
- SCORZAFAVE, Luiz; SOARES, Milena; DORIGAN, Tulio. Vale a pena pagar para desarmar? Uma avaliação do impacto da campanha de entrega voluntária de armas sobre as mortes com armas de fogo. **Estudos Econômicos (São Paulo)**, v. 45, n. 3, p. 475-497, 2015.



INTRODUÇÃO

- **D-D**: painel ou *cross-section* repetidas (se a composição dos grupos de tratamento e controle for estável no tempo).
- **D-D**: compara mudança dos resultados dos grupos de T e C antes e depois da intervenção.
- Seja $t=0$ o período anterior ao programa e $t=1$ o período posterior:

Y_t^T e $Y_t^C \rightarrow$ resultados para os grupos no tempo t .



EX: TRABALHO PIONEIRO

- Snow (1855) comparou a incidência de cólera entre habitantes de 2 áreas de Londres. Inicialmente, cada área era abastecida por uma companhia diferente e, inicialmente, ambas coletavam água suja.
- Em 1952, uma delas passou a coletar e distribuir água limpa. As taxas de mortalidade despencaram nas regiões abastecidas por essa companhia!!!



EXEMPLO

- Efeito de salário mínimo sobre emprego (Card e Krueger, 1994).
- Se mercado de trabalho é competitivo, **↑ SM**, **↓ demanda por trabalho** e **↓ emprego** justamente dos que eu queria beneficiar com tal política.
- Card e Krueger (1994) usam variação estadual do SM para verificar isso.



CARD E KRUEGER (1994)

- 01/04/92: NJ eleva SM de \$4,25 para \$5,05;
- Card e Krueger (1994) coletam dados de emprego em lojas *fast-food* em NJ em fev/92 e nov/92, pois *fast-food* é grande empregador que paga salário mínimo.
- Coletam dado semelhante na Pensilvânia, na divisa com NJ. Lá, o SM continuou sendo \$4,25.
- Comparam a mudança de emprego de fev. a nov. em NJ com a mudança na Pensilvânia no mesmo período.



CARD E KRUEGER (1994)

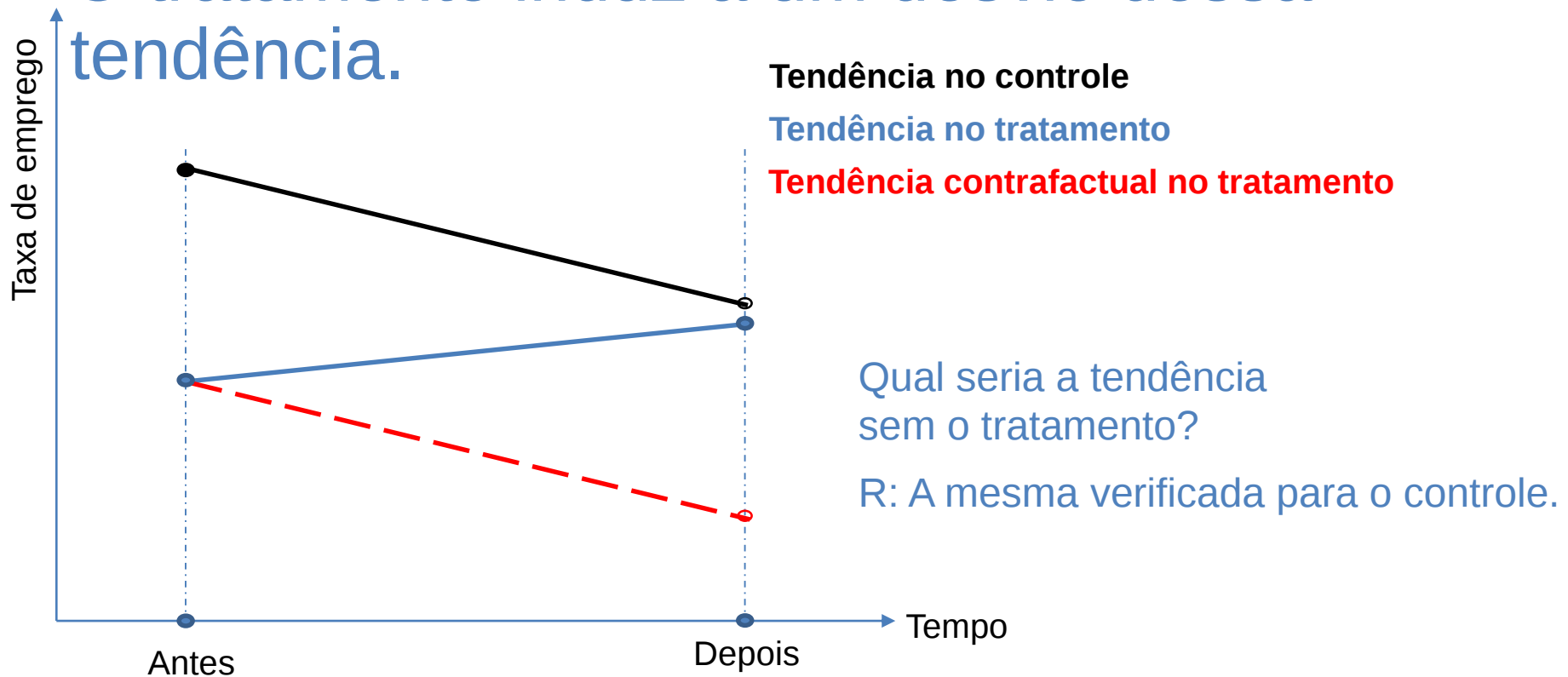
	PA	NJ	Δ NJ-PA
ANTES	23,33 (1,35)	20,44 (0,51)	-2,89 (1,44)
DEPOIS	21,17 (0,94)	21,03 (0,52)	-0,14 (1,07)
ΔTEMPO	-2,16 (1,25)	0,59 (0,54)	2,76 (1,36)

- O emprego cresceu no estado que $\uparrow$ SM;
- Hipótese chave de identificação: tendência do emprego seria a mesma nos dois estados na ausência do tratamento.



CARD E KRUEGER (1994)

- O tratamento induz a um desvio dessa tendência.



HIPÓTESE DE TENDÊNCIA COMUM

- Assim, se escrevermos D-D como:

$$DD = E(Y_1^T - Y_0^T | T=1) - E(Y_1^C - Y_0^C | T=1),$$

não observável

A hipótese de tendência comum implica que:

$$E(Y_1^C - Y_0^C | T=1) = E(Y_1^C - Y_0^C | T=0).$$

observável

De modo que:

$$DD = E(Y_1^T - Y_0^T | T=1) - E(Y_1^C - Y_0^C | T=0)$$

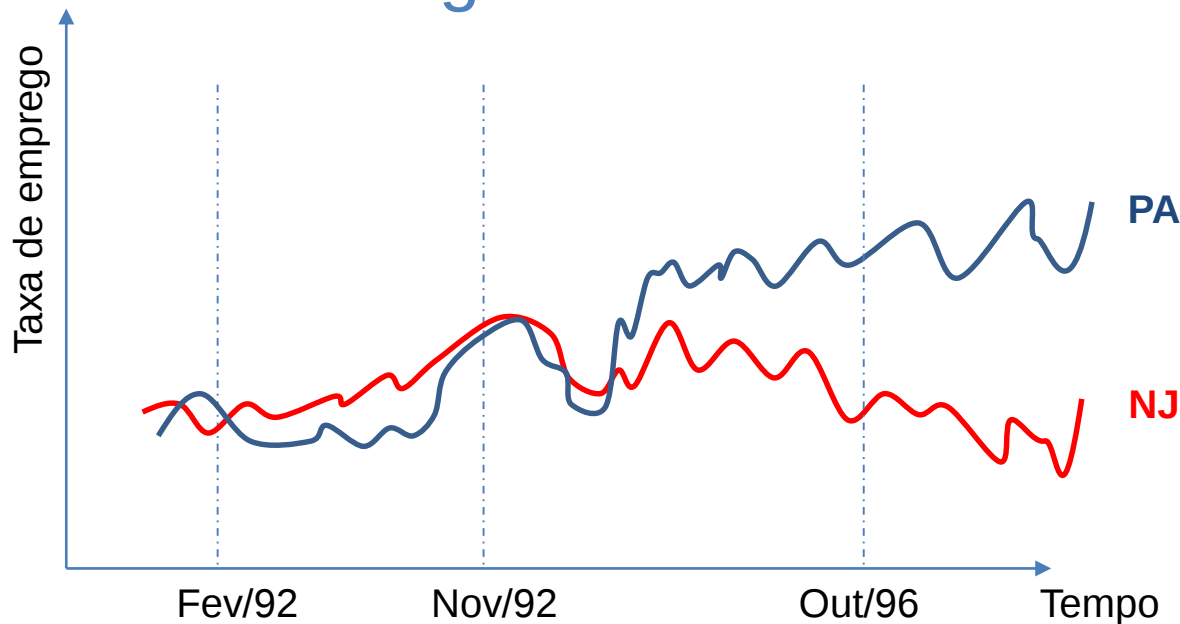
é estimável!!! Mas será que a hipótese é razoável?

R: Se tivermos dados de múltiplos períodos, é possível avaliar!!!



CARD E KRUEGER (2000)

- Dados de folha de pagamento dos restaurantes em NJ e algumas cidades de PA.



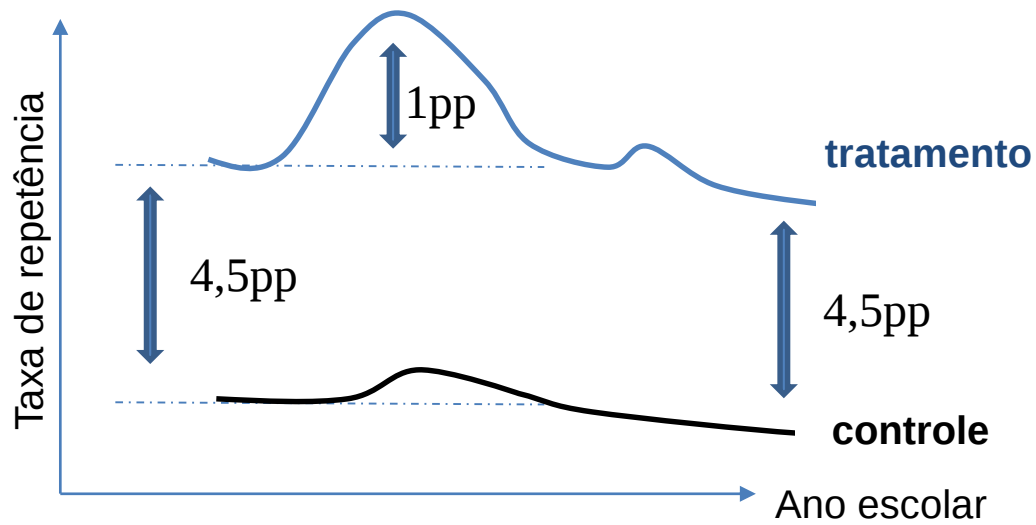
Não parece muito convincente a hipótese nesse caso!!!

Aumento de SM para \$4,75 que afetou PA, mas não afetou NJ.



PISCHKE (2007)

- Até 1960, todos os estados alemães, exceto a Bavária começavam as aulas na primavera. Em 1966-67, todos passaram a começar no outono. Assim, duas gerações tiveram menos dias letivos na transição (24 ao invés de 37 semanas).



Hipótese mais plausível



DIF-DIF EM EQUAÇÕES

$$Y_{it} = \alpha + \beta T_i t + \rho T_i + \gamma t + \varepsilon_{it}$$

onde t = tempo (0,1) e T_i = tratamento (0,1)

β = efeito causal D-D!!!

- Modelo Saturado: $E(Y_{it}|T_i, t)$ só pode assumir 4 valores e há quatro parâmetros.



MODELO SATURADO

- **CONTROLE**

Antes: $E(Y_{it}|T=0, t=0) = \alpha + E(\varepsilon_{it}|T=0, t=0)$

Depois: $E(Y_{it}|T=0, t=1) = \alpha + \gamma + E(\varepsilon_{it}|T=0, t=1)$

- **TRATAMENTO**

Antes: $E(Y_{it}|T=1, t=0) = \alpha + \rho + E(\varepsilon_{it}|T=1, t=0)$

Depois: $E(Y_{it}|T=1, t=1) = \alpha + \beta + \rho + \gamma + E(\varepsilon_{it}|T=1, t=1)$

- **Hipótese chave:**

$$E(\varepsilon_{it}|T=0, t=1) - E(\varepsilon_{it}|T=0, t=0) = E(\varepsilon_{it}|T=1, t=1) - E(\varepsilon_{it}|T=1, t=0)$$



VANTAGENS E DESVANTAGENS DO D-D

- **Vantagens**

1. Relaxa a hipótese de seleção somente em observáveis.
2. É facilmente generalizável para mais períodos de tempo.
3. Facilita estudo de políticas diferentes da binária.



EXEMPLO VANTAGEM (2)

$$Y_{it} = \alpha + \sum \lambda_t + \rho T_i + \beta_1 \sum I_{it} + \varepsilon_{it}$$

- Onde I_{it} assume valor 1 para aqueles indivíduos tratados apenas após passarem a receber o tratamento
- Ex: Bônus em SP: $t=2007$ (antes) e 2009 e 2011(depois)

$$Nota_{it} = \beta_0 + \beta_1 EE_SP_i + \beta_2 I_{i2009} + \beta_3 I_{i2011} + \beta_4 d_{2009} + \beta_5 d_{2011} + \varepsilon_{it}$$



EXEMPLO VANTAGEM (3)

- Efeito de SM sobre emprego olhando todos os estados ao invés de olharmos apenas NJ e PA e $t=2$. Se houver variação entre estados e/ou tempos no SM, posso estimar.
- Card (92): variação regional do SM entre estados

$$Y_{ist} = \gamma_s + \lambda_t + \delta.(Fa_s.d_t) + \varepsilon_{ist}$$

{ 1 → depois 1990 }

$$\Delta \bar{Y}_s = \lambda^* + \delta.FA_s + \Delta \varepsilon_s$$

← Fração de adolescentes (afetados por SM)

Variação média do
emprego no estado
s entre t e t-1

→ $(\lambda_t - \lambda_{t-1})$



VANTAGENS E DESVANTAGENS DO D-D

- **Vantagens**

4. Facilita o uso de controles adicionais X_{ist} para ajudar a garantir o paralelismo

5. Fornece maneira fácil de checar a estratégia de identificação: adicionando tendências temporais estado-específicas (X_{st}) como controles. Talvez o tratamento tenha mudado tendência estado prévia.



VANTAGEM (5)

$$Y_{ist} = \underbrace{\gamma_{0s}}_{\text{Efeito específico de estados (como antes) de intercepto}} + \underbrace{\gamma_{1s} \cdot t}_{\text{Efeito específico de estados de tendência (dummy de interação s.t)}} + \lambda_t + \delta \cdot D_{st} + X_{ist} \cdot \beta + \varepsilon_{ist}$$

Efeito específico de estados (como antes) de intercepto.

- São necessários 3 períodos de tempo pelo menos.
- Não deve mudar o efeito estimado (δ) para garantir robustez do resultado.



BESLEY E BURGUESS(2004)

TABLE 5.2.3

Estimated effects of labor regulation on the performance of firms in Indian states

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	-.186 (.064)	-.185 (.051)	-.104 (.039)	.0002 (.020)
Log development expenditure per capita		.240 (.128)	.184 (.119)	.241 (.106)
Log installed electricity capacity per capita		.089 (.061)	.082 (.054)	.023 (.033)
Log state population		.720 (.96)	0.310 (1.192)	-1.419 (2.326)
Congress majority			-.0009 (.01)	.020 (.010)
Hard left majority			-.050 (.017)	-.007 (.009)
Janata majority			.008 (.026)	-.020 (.033)
Regional majority			.006 (.009)	.026 (.023)
State-specific trends	No	No	No	Yes
Adjusted R ²	.93	.93	.94	.95

- Efeito de regulação trabalhista na produtividade na Índia. Diferentes estados colocam regulação em diferentes pontos do tempo.
- Quando adiciona, some o efeito. Ou seja, a regulação do trabalho aumentou em estados onde o produto declinaria de qualquer forma.

Notes: Adapted from Besley and Burgess (2004), table IV. The table reports regression DD estimates of the effects of labor regulation on productivity. The dependent variable is log manufacturing output per capita. All models include state and year effects. Robust standard errors clustered at the state level.

LIMITAÇÕES

1. A composição do grupo de tratamento e controle pode mudar como resultado do tratamento.

Ex: Efeito da generosidade da assistência pública sobre oferta de trabalho. Nos EUA, essa generosidade varia entre estados. Mas será que Δ 's que de qualquer modo ofertariam pouco trabalho migram para estados mais generosos? No D-D tradicional, isso faria o programa parecer pior do que ele é.



LIMITAÇÕES

2. Difícil garantir que o tratamento é a única coisa que mudou entre o tratamento e controle, sendo responsável pelos resultados.

3. A escolha do período do tempo depois do tratamento:

- Se escolhe muito perto do tratamento, chance do problema 2 ocorrer é menor, mas pode ser que não tenha dado tempo do efeito aparecer
- Se escolhe muito longe do tratamento, aumenta a chance do problema 2 ocorrer (outras coisas vão ocorrendo nos dois grupos), mas dá mais tempo do efeito aparecer.



INFERÊNCIA

- O fato de que muitas vezes a variável de interesse é agregada além do nível do indivíduo dificulta a inferência estatística
- Ex: Card e Krueger: Sal. Mínimo
- Solução: Cluster por Estado e Ano
- No entanto, como pode haver dependência temporal (autocorrelação serial), Angrist sugere clusterizar apenas por Estado
- Problema: pode gerar poucos clusters.



EXTENSÕES

A. Modelo de efeitos fixos em painel (períodos múltiplos)

Imagine que tenho $t > 2$ e que quero também corrigir por X variando no tempo e estado.

$$Y_{it} = \Phi \cdot T_{it} + \delta X_{it} + \eta_i + \varepsilon_{it}$$

$$\underbrace{Y_{it} - Y_{it-1}}_{\Delta Y_{it}} = \Phi \cdot \underbrace{(T_{it} - T_{it-1})}_{\Delta T_{it}} + \delta \underbrace{(X_{it} - X_{it-1})}_{\Delta X_{it}} + \underbrace{\varepsilon_{it} - \varepsilon_{it-1}}_{\Delta \varepsilon_{it}}$$

Que pode ser estimado por MQO se $\text{cov}(\Delta \varepsilon_{it}, \Delta T_{it}) = 0$ e $\text{cov}(\Delta \varepsilon_{it}, \Delta X_{it}) = 0$. Bertrand, Duflo et. al (2004) mostram que a estimação do erro-padrão tem que ser corrigido pela correlação serial do erro.



EXTENSÕES

B. Mais cuidado com a hipótese de identificação.
“Ashenfelter’s dip” → efeito do treinamento sobre salário.

- Fazer o curso é o mais provável se uma diminuição temporária de renda ocorreu logo antes da introdução do programa.
- Assim, os tratados teriam crescimento de renda, mesmo sem o treinamento (reversão a média).
- D-D superestima o impacto do programa nesse caso (diferença entre controle e contrafactual muda no tempo)



EXTENSÕES

C. PSM com D-D

Antes de fazer D-D, usa PSM para casar melhor T e C, pré-tratamento. Depois, aplico D-D na amostra “casada”.

$DD_i = (Y_1^T - Y_0^T) - \sum_{j \in C} w(ij) \cdot (Y_{j1}^C - Y_{j0}^C)$, onde w_{ij} é o peso definido no PSM dado ao Δj a ser casado com o Δ tratamento i .

- Hirano, Imbens e Ridder (2003) : Mínimos Quadrados Ponderados, em que os pesos dos controles são dados pelo seu $p(x)$, gera estimador eficiente:

$$\Delta Y_{it} = \alpha + \beta \cdot T_i + \gamma \Delta X_{it} + \varepsilon_{it}$$

β = DD effect.

Para o tratamento: peso = 1

Para o controle: peso = $\hat{p} (1 - \hat{p})$



EXTENSÕES

D. Diferenças Triplas

Ex: Imagine que faça treinamento para trabalhadores e alguns desistem e outros ficam até o fim. Se fizer D-D entre esses 2 grupos, só será não viesado se as oportunidades de renda fora do programa forem iguais para os grupos (o que não deve ser verdade, já que quem fica no programa deve ter menor chance), subestimando assim, o efeito do programa.

Solução: usar informações de Δ 's não tratados mesmo!!! (não participantes)



(D) DIFERENÇAS TRIPLAS

- I. D-D entre os que ficaram e os não participantes.
- II. D-D entre os que saíram e os não participantes.
- III. D-D entre os que saíram e ficaram.

$$\left. \begin{array}{l} D_{it}=1 \text{ participantes} \\ D_{it}=0 \text{ não participantes} \\ T=1,2. \end{array} \right\} \begin{array}{l} I = (Y_2^T - Y_1^T) - (Y_2^C - Y_1^C) | \underline{D_{i2}=1} \longrightarrow \text{Ficaram até o fim} \\ II = (Y_2^T - Y_1^T) - (Y_2^C - Y_1^C) | \underline{D_{i2}=0} \longrightarrow \text{Saíram no meio} \end{array}$$

$$III = I - II$$



EXTENSÕES

D. Diferenças Triplas

- Outra aplicação de diferenças triplas envolve em calcular a evolução entre dois períodos do tempo antes do tratamento da evolução entre tratamento e controle.
- Serve como robustez para garantir a validade da hipótese de paralelismo.

$$DD = [E(Y_1^T - Y_0^T | T_1 = 1) - E(Y_1^C - Y_0^C | T_1 = 0)] \\ - [E(Y_t^T - Y_{t-1}^T | T_1 = 1) - E(Y_t^C - Y_{t-1}^C | T_1 = 0)]$$



EXTENSÕES

D. Diferenças Triplas: Exemplo

$$\text{NotaMédia}_{it} = \beta_0 + \beta_1 \text{SP}_i + \beta_2 \text{EE}_i + \beta_3 d_{2009} + \beta_4 d_{2011} + \beta_5 \text{SP}_i d_{2009} + \beta_6 \text{SP}_i d_{2011} + \beta_7 \text{EE}_i d_{2009} + \beta_8 \text{EE}_i d_{2011} + \beta_9 \text{EE}_i \text{SP}_i + \beta_{10} \text{EE}_i \text{SP}_i d_{2009} + \beta_{11} \text{EE}_i \text{SP}_i d_{2011} + \varepsilon_{it}, \quad (7)$$

em que $\text{SP}_i = 1$, se a escola pertence ao estado de São Paulo e 0, caso contrário; $\text{EE}_i = 1$, se a escola é estadual. Neste caso, os coeficientes que medem o impacto do programa em 2009 e 2011 são, respectivamente, β_{10} e β_{11} , e podem ser escritos como

$$\beta_{10} = \left[\left(\text{Nota}_{2009}^{\text{E,SP}} - \text{Nota}_{2007}^{\text{E,SP}} \right) - \left(\text{Nota}_{2009}^{\text{M,SP}} - \text{Nota}_{2007}^{\text{M,SP}} \right) \right] - \left[\left(\text{Nota}_{2009}^{\text{E,F}} - \text{Nota}_{2007}^{\text{E,F}} \right) - \left(\text{Nota}_{2009}^{\text{M,F}} - \text{Nota}_{2007}^{\text{M,F}} \right) \right] \quad (8)$$

$$\beta_{11} = \left[\left(\text{Nota}_{2011}^{\text{E,SP}} - \text{Nota}_{2007}^{\text{E,SP}} \right) - \left(\text{Nota}_{2011}^{\text{M,SP}} - \text{Nota}_{2007}^{\text{M,SP}} \right) \right] - \left[\left(\text{Nota}_{2011}^{\text{E,F}} - \text{Nota}_{2007}^{\text{E,F}} \right) - \left(\text{Nota}_{2011}^{\text{M,F}} - \text{Nota}_{2007}^{\text{M,F}} \right) \right] \quad (9)$$

em que E = Estadual, M = Municipal, SP = Estado de São Paulo e F = Fora do Estado de São Paulo.

CONTROLE SINTÉTICO

Abadie, Diamond e Hainmueller (2010), J. American Stat. Association

- Efeito de intervenções em nível mais agregado

Ex: Efeito de lei estadual, política estadual específica.

Ao invés de escolher 1 UF como controle, ou estados fronteiriços, que poderia ser “contaminado” pela política, construo média ponderada dos estados no período anterior ao tratamento que melhor represente o comportamento passado do estado tratado.

- Suponha $T+1$ períodos, tratamento é período final
- Suponha $G+1$ estados, estado G é o tratado.



CONTROLE SINTÉTICO

Seja $G_i = G \Rightarrow$ tratamento

Seja $G_i = 0 \Rightarrow$ controle

- O resultado do grupo de tratamento do período final, na ausência do tratamento é estimado como a média ponderada dos resultados dos G estados nos T períodos de tempo iniciais:

$$\hat{E}(Y_i(0)|T_i=T, G_i=G) = \sum_{g=0}^{G-1} \lambda_g \cdot \bar{Y}_{gT}$$

Com os pesos λ_g tq $\sum_{g=0}^{G-1} \lambda_g = 1$ e $\lambda_g \geq 0$



CONTROLE SINTÉTICO

- Os pesos são escolhidos tq minimizem a distância entre o grupo de tratamento e o grupo de controle sintético antes do tratamento.

$$\| \bar{y}_{g_0} - \sum_{g=0}^{G-1} \lambda_g \cdot \bar{y}_{g_0} \|$$

...



$$\sqrt{(Y_1 - Y_0 \lambda)' V (Y_1 - Y_0 \lambda)} = \| Y_1 - Y_0 \lambda \|$$

$$\| \bar{y}_{g,T-1} - \sum_{g=0}^{G-1} \lambda_g \cdot \bar{y}_{g,T-1} \|$$

Onde V é positiva semidefinida e pode ser escolhida para minimizar o EQM de previsão produzida.

- Para inferência, os autores propõe teste de permutação de Fisher, que requer estimar um efeito tratamento placebo para cada UF com o mesmo método usado inicialmente.



CONTROLE SINTÉTICO

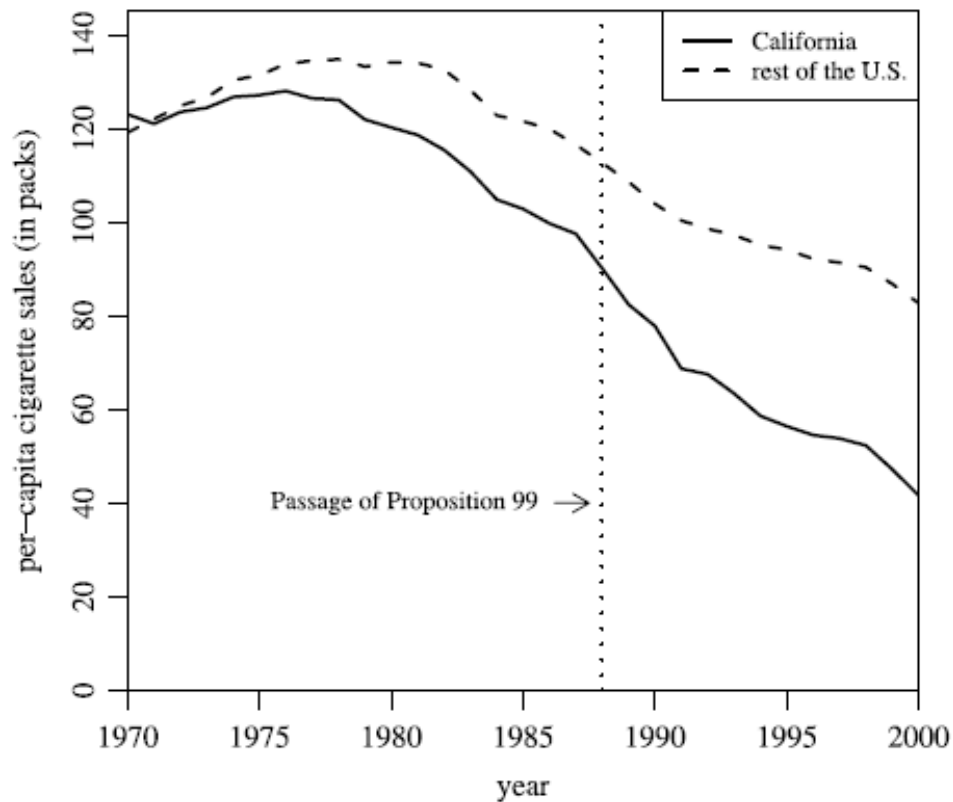


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.



CONTROLE SINTÉTICO

Table 1. Cigarette sales predictor means

Variables	California		Average of 38 control states
	Real	Synthetic	
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

NOTE: All variables except lagged cigarette sales are averaged for the 1980–1988 period (beer consumption is averaged 1984–1988). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.



CONTROLE SINTÉTICO

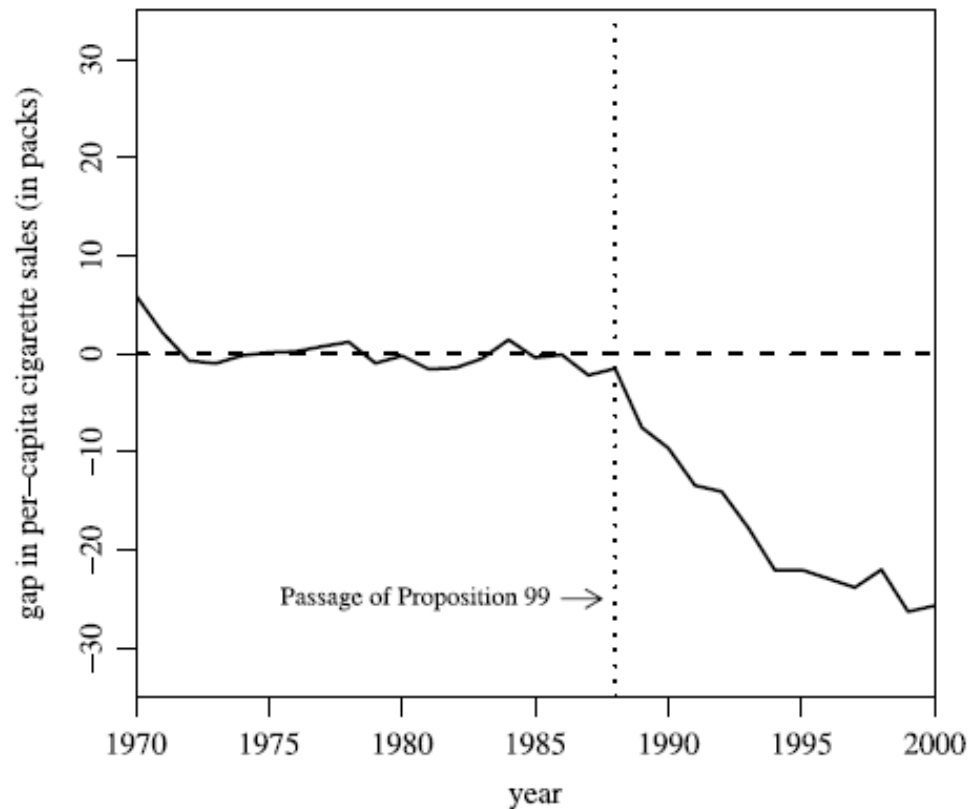


Figure 3. Per-capita cigarette sales gap between California and synthetic California.



CONTROLE SINTÉTICO

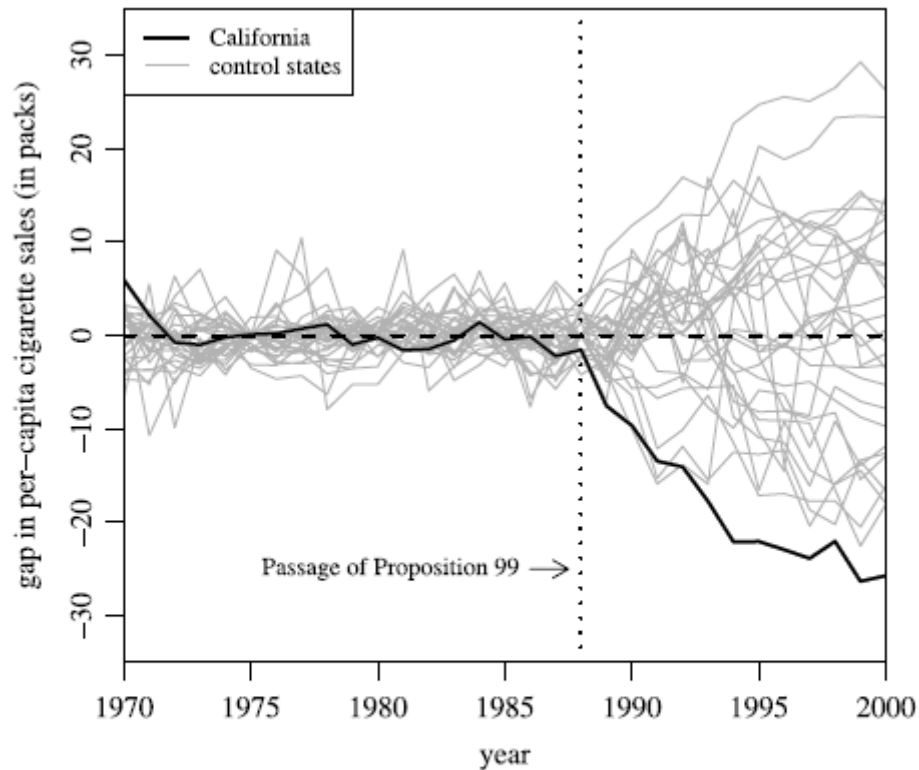


Figure 6. Per-capita cigarette sales gaps in California and placebo gaps in 29 control states (discards states with pre-Proposition 99 MSPE five times higher than California's).



CONTROLE SINTÉTICO

- Força pesquisador a demonstrar afinidade entre unidades afetadas e não afetadas pela intervenção
- A inferência é válida independentemente do número de unidades disponíveis para comparação.
- Permite avaliar intervenções em nível mais agregado

