



Avaliação de Políticas Públicas usando o R

Introdução à avaliação de impacto

Vítor Wilher, Mestre em Economia e Cientista de Dados

```
library(tidyverse)
library(truncnorm)
library(broom)
library(kableExtra)
```

1 Introdução

Política pública refere-se às ações e normas formuladas pelos governos com o objetivo de solucionar problemas que impactam a sociedade. Assim, os propósitos podem ser os mais diversos, como aumentar a renda, melhorar a aprendizagem nas escolas ou aumentar a arrecadação de impostos. Entretanto, por diversos motivos, nem sempre o resultado da política é o mesmo que o esperado. Para avaliar isso, é importante que haja uma análise constante e rigorosa, com base em métricas que realmente respondam por um critério adequado ao objetivo. Por exemplo, no caso de uma política que vise melhorar a performance de aluno por meio de melhoria na infraestrutura, não cabe medir seu desempenho com base no número de computadores ou livros comprados. É necessário que a variável de interesse seja uma medida da aprendizagem dos estudantes. Esse tipo de avaliação faz parte de uma agenda global das últimas décadas de políticas públicas baseadas em evidências, que visa retirar o foco dos insumos e colocar nos resultados (Gertler et al. 2016).

Uma das motivações para se fazer políticas baseadas em evidência é simples, mas muito forte, e é a primeira lição da maioria dos manuais de introdução à economia: Os recursos são escassos. Isso é especialmente válido para países em desenvolvimento, como o Brasil. O governo tem uma dotação limitada de recursos, sejam eles financeiros, humanos ou de estrutura. Além disso, é necessário lembrar que o fim das políticas públicas é melhorar a vida das pessoas. Por isso é essencial que se façam políticas públicas que consigam atingir seu propósito, o que só pode ser averiguado se for feita uma avaliação. Além disso, a avaliação do impacto é importante para medir não somente se a ação governamental conseguiu ter efeito, mas também mensurar a magnitude do resultado, o que é necessário para estimar o custo benefício da política.

A utilização da avaliação de impacto de uma política pública tem dois tipos de usos principais: Interno e externo. No uso interno ela tem como objetivo subsidiar decisões acerca da operação do programa, a depender dos resultados encontrados. Se forem positivos, podem corroborar para manutenção ou ampliação da política.

Se forem negativos, uma diminuição ou cancelamento. Com o maior detalhamento das análises, elas podem auxiliar um processo de redesenho do programa, de forma a aprimorar e potencializar os resultados. Já o uso externo tem como principal objetivo subsidiar a aplicação ou alteração de programas semelhantes em outros lugares ou períodos. Além disso, avaliações servem para a identificação das melhores práticas na área da política em questão, o que permite que policy makers possam implementar políticas com resultados já consagrados em outros lugares.

2 Definição do problema e causalidade

A questão típica da avaliação de políticas públicas é centrada na definição do tratamento, T ou seja, a intervenção da política em uma pessoa, grupo de pessoas ou local. Assim:

$$T_i = \begin{cases} 1 & \text{Se o indivíduo recebeu o tratamento} \\ 0 & \text{Se não recebeu} \end{cases}$$

O interesse da avaliação tipicamente é analisar qual foi o efeito do tratamento em uma determinada variável de interesse, Y . O problema é que nós nunca conseguimos observar um mesmo indivíduo com e sem tratamento na mesma situação.

Por exemplo, suponha que alguém queira implementar uma política em determinado local. O impacto dessa política pode ser definido como a diferença entre duas realidades: o que aconteceu posteriormente à ação governamental e o que teria acontecido no lugar sem essa política. A maior dificuldade dessa formulação é justamente mostrar o que aconteceu na segunda realidade, pois ela não ocorreu de fato, ou seja, é hipotética. Assim, para definir o impacto, é preciso buscar meios de se aproximar desse contra-factual (Peixoto et al. 2012).

Uma das questões centrais desse tipo de avaliação é isolar o efeito causal da política sobre a variável de interesse. Caso fosse possível garantir que nada mais poderia afetar o comportamento do indicador, não seria nem necessária uma comparação com o a situação hipotética de não-intervenção. A variação da própria variável de interesse antes e depois da política seria o efeito puro da ação. Entretanto, não há caso em que isso seja factível. Assim, em uma definição mais estrita, definir uma relação de causa e efeito de uma política em um resultado seria impossível na prática. Por isso, a aproximação da situação ideal é feita com modelos estatísticos.

Na prática, a definição de causalidade é uma questão de convenção. Modelos são descrições do mundo hipotético obtidos pela variação dos fatores, também hipotéticos, que determinam os resultados. Assim, a aceitação de um modelo de contrafactuais depende da aceitação de seus componentes, que são as regras usadas para derivar um modelo, sejam suas características matemáticas ou suas correspondências com a teoria (Heckman 2008). Em outras palavras, para definir um impacto causal é preciso convencer o seu público que ele é de fato causal.

Nesse sentido, é preciso que a escolha da metodologia seja adequada para defender que o efeito da política foi de fato isolado. A simples análise estatística pode levar a conclusões erradas sobre os fenômenos. Um exemplo clássico e bastante ilustrativo é a correlação positiva entre vendas de sorvete e afogamentos no mar. Uma análise descuidada pode levar à tese de que tomar sorvete aumenta sua chance de se afogar, o que logicamente não é o caso. O que acontece de fato é que ambos os indicadores têm influência da temperatura. Ou seja, a temperatura aumenta a propensão das pessoas a consumir sorvetes e a entrar na água. Esse fenômeno é chamado de correlação espúria, que é quando há relação estatística, mas não causal.

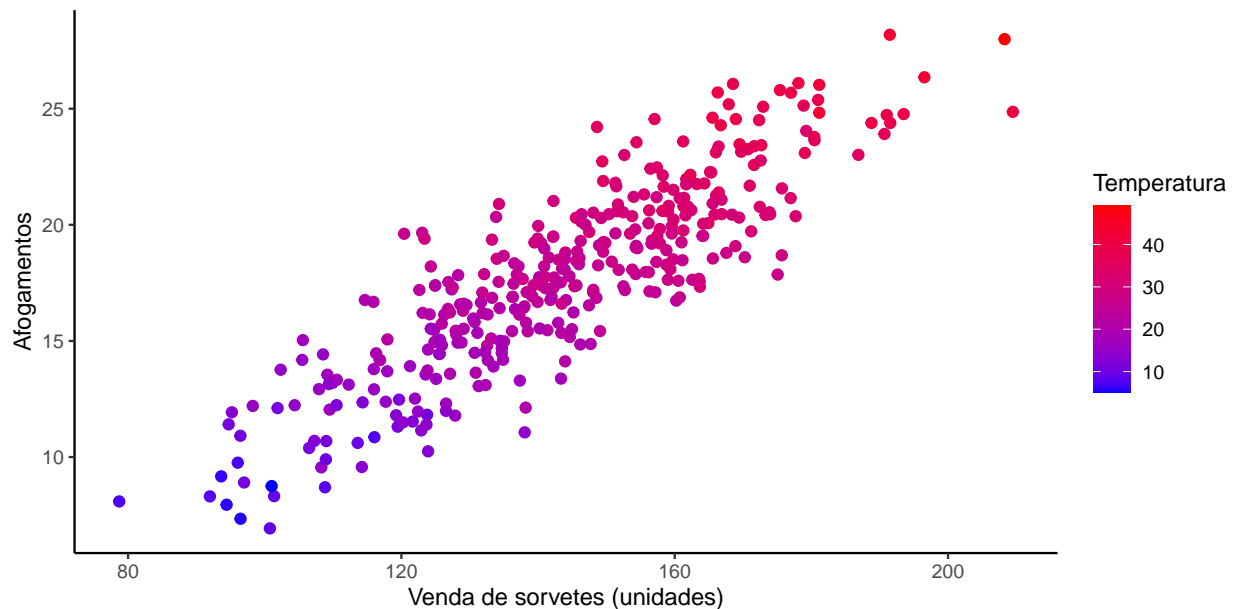
Podemos estender esse exemplo no R, assumindo dados hipotéticos para os 365 dias do ano. Ambas as variáveis são compostas de três termos, um intercepto, a influência da temperatura e um componente aleatório. Ou seja, elas não tem nenhum elemento de causa e efeito entre ambas, mas apresentam uma correlação muito forte de 0.87. Isso acontece pois ambas têm uma causa em comum, a temperatura.

```
set.seed(2000)

Temperatura <- rnorm(365, mean = 25, sd = 8)
venda_sorvetes <- 50 + Temperatura* 2.5 + rnorm(365, mean = 30, sd = 10)
afogamentos <- Temperatura* 0.5 + rnorm(365, mean = 5, sd = 1)

df <- data.frame(Temperatura, venda_sorvetes, afogamentos)

ggplot(df) +
  geom_point(aes(x = venda_sorvetes, y = afogamentos, color = Temperatura), size = 2) +
  theme_classic() +
  scale_color_gradient(low = "blue", high = "red", na.value = NA) +
  xlab("Venda de sorvetes (unidades)") +
  ylab("Afogamentos")
```



```
print(cor(df$venda_sorvetes, df$afogamentos))
```

```
[1] 0.8712027
```

2.1 Validação interna X Validação externa

Nem todos os desenhos de avaliação de impacto são iguais e há custos em optar por um tipo de análise ou outra. Dois componentes importantes para avaliarmos os resultados da nossa estimação do tratamento são sua validade interna e sua validade externa. Se uma análise tem validade interna isso significa que podemos estabelecer uma relação causal entre a política e a variável de interesse dentro do contexto do estudo. Já a validade externa significa que podemos generalizar os resultados do nosso estudo para outros contextos. No caso de políticas públicas é muito difícil fazer uma avaliação que tenha validação externa completa, pois alterações de local, grupos tratados, períodos e até pequena alterações na política podem alterar completamente o resultado final. Entretanto, a depender do caso, se bem executada, a análise pode ter um bom grau de validação externa. Assim, é importante estar atento a sensibilidade que os resultados

de uma política têm em relação ao contexto em que ela foi aplicada, principalmente para um gestor que a use como referência para um outro programa.

3 Heterogeneidade de impacto

Por definição, os objetos das políticas públicas são as pessoas pertencentes a um ou mais grupos. Como essas pessoas não são iguais entre si, o efeito que a ação governamental tem sobre elas vai ser naturalmente diferente. Por exemplo, um programa de transferência de renda hipotético que destina 200 reais para pessoas que recebem até 1000 reais vai atuar em grupo de pessoas bem heterogêneas. Por isso, independente do indicador que pretende-se avaliar, seja impacto na educação, mortalidade infantil ou até renda após o programa, é esperado que o efeito não seja homogêneo para todas as pessoas. A justificativa para isso pode ter várias causas, mas uma das principais é que se espera que o impacto dos 200 reais para quem já tinha renda perto do teto do programa seja muito menor do que quem não tinha renda. Outras diferenças prévias no grupo que sofreu a intervenção podem afetar o impacto do programa, como escolaridade, local onde vive, número de pessoas na família e assim por diante.

Normalmente, uma avaliação de impacto vai se preocupar com a média do impacto no grupo tratado. Porém, a depender do caso, pode ser mais importante fazer essa análise de efeitos heterogêneos. Com esse tipo de avaliação, é possível identificar fatores que influenciam na eficácia da política pública, o que pode permitir modificar o desenho do programa para atingir determinados sub-grupos onde o resultado esteja menos satisfatório (Peixoto et al. 2012).

4 Estratégias de identificação

Como dissemos, obviamente não é possível ter acesso à realidade alternativa em que não houve a política pública que queremos estimar o impacto. A solução mais usual nas ciências naturais, como a Física e a Química, é realizar experimentos. No caso de grande parte das questões dessas disciplinas é possível contruir um meio de controlar o efeito de fatores externos e testar uma hipótese de maneira isolada. Nas ultimas décadas, a popularidade dos estudos que utilizam experimentos para avaliar políticas públicas cresceu muito, pois são considerados o padrão ouro da ciência. Entretanto, a aplicação desse tipo de abordagem nem sempre é possível nessa área, pois experimentos as vezes não são exequíveis, ou são muito caros. Mesmo que não seja possível realizar experimentos, pensar no experimento ideal que responderia a pergunta em questão pode ajudar na hora de construir a solução para identificação causal e avaliação do impacto (Angrist and Pischke 2008).

O ponto mais importante para a robustez dos métodos experimentais é a atribuição aleatória entre grupos de controle e tratamento. Isto é, no primeiro momento os indivíduos não estão em nenhum grupo e são, posteriormente, colocados no grupos de forma aleatória e então, os que estiverem no grupo de tratamento, sofrem a intervenção. Assim, não há diferença entre os grupos no primeiro momento e, portanto, caso o experimento seja bem desenhado, as diferenças de trajetórias na variável de interesse são causadas exclusivamente pelo tratamento.

Para demonstrar o poder da atribuição aleatória, iremos utilizar o exemplo dado anteriormente do programa de transferência de renda de 200 reais. Suponha que, para avaliar o impacto da política na renda futura das pessoas, pudéssemos, antes do início da política, sortear parte do grupo focal para tratamento e parte para controle. Ou seja, não estaríamos de fato implementando a política em sua totalidade, mas sim, a testando. Para fins didáticos, vamos supor que a renda no momento anterior a política tenha uma distribuição normal. Também assumiremos que o impacto do tratamento é homogêneo e igual a 250 reais após a aplicação do programa, ou seja, afeta todos os indivíduos da mesma forma.

```
id <- seq(1:5000) #id de cada indivíduo
```

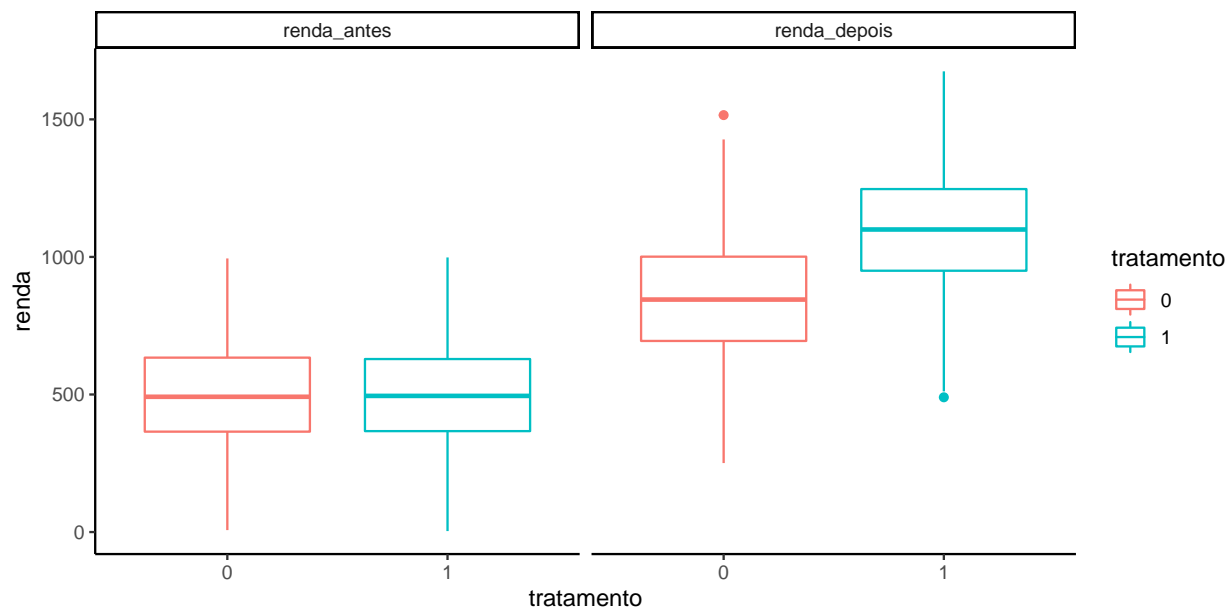
```
# distribuição normal truncada entre 0 e 1000
renda_antes <- rtruncnorm(n=5000, a=0, b=1000, mean=500, sd=200)

tratamento <- as.factor(rbinom(n=5000, size=1, prob=0.5))

df <- data.frame(id, renda_antes, tratamento)

df$renda_depois <- renda_antes*1.1 + as.numeric(tratamento)*250 + rnorm(5000, mean = 50, sd = 50)

df %>%
  pivot_longer(!c(id, tratamento), names_to = "tempo", values_to = "renda") %>%
  ggplot() +
  geom_boxplot(aes(x = tratamento, y = renda, color = tratamento)) +
  facet_wrap(vars(tempo)) +
  theme_classic()
```



Assim, nesse exemplo hipotético, poderíamos estimar o impacto do programa com algum procedimento estatístico simples, como uma regressão linear. Assim, podemos ver que o impacto encontrado, medido pelo parâmetro da regressão, é praticamente igual ao efeito definido.

```
#diferença entre antes e depois
df$diff <- df$renda_depois - df$renda_antes

lm(diff ~ tratamento, data = df) %>%
  tidy() %>%
  kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	349.7236	1.058283	330.4632	0
tratamento1	251.0386	1.504786	166.8268	0

No caso de políticas públicas, raramente temos atribuições aleatórias, o que gera um problema de viés de seleção. Primeiramente pois a própria definição de grupo de controle pode não ser óbvia. No caso do programa

de transferência de renda, a definição do grupo de controle seria especialmente difícil, pois se todas as pessoas com renda condizente com o programa estivessem dentro dele, não haveria indivíduos com renda baixa no local para fazer a comparação. Além disso, mesmo que uma política ainda pudesse dispor de um grupo de comparação, pouco provável que o tratamento seja 100% aleatório. Por exemplo, digamos que a tenhamos como objetivo avaliar o impacto da construção de uma linha do metrô na renda dos moradores próximos a ela. Não é factível colocar quem mora longe das estações como grupo de controle, pois, muito provavelmente, são grupos muito distintos. Isso acontece porque a decisão de onde colocar a linha do metrô não é aleatória, já que leva em consideração características do local, como renda média, densidade, demografia, entre outras.

Um exemplo clássico desse problema de viés de seleção é a comparação de pessoas que foram hospitalizadas com as não-hospitalizadas. Se formos comparar o quadro de saúde dos dois grupos, poderíamos concluir que hospital faz mal à saúde, o que é claramente errado, pois as pessoas já eram menos saudáveis antes de serem internadas.

Para tratar esse tipo de problema, existem alguns métodos estatísticos que podem ser aplicados. Esse métodos visam aproximar a análise a um experimento, são os chamados métodos quasi-experimentais ou experimentos naturais. A diferença entre essas duas definições é que nos Quase-experimentos o critério de atribuição de tratamento é definido pelo investigador, enquanto nos experimentos naturais a exposição ao tratamento é determinada pela natureza ou por fatores externos. Entretanto esses dois conceitos são semelhantes, dado que sempre há um papel do pesquisador em definir a pergunta a ser feita.

Existem diversos métodos e procedimentos que se encaixam nesse tipo de abordagem, os quais serão trabalhados ao longo do curso. Cada método tem um propósito e aplicação diferente e mais adequado. Boa parte do desafio da avaliação de políticas públicas é construir uma abordagem que utilize das ferramentas mais adequadas. Listamos alguns desses métodos, de maneira não exaustiva, que serão vistos e seus propósitos a seguir:

Pareamento - As técnicas de pareamento são utilizadas quando o tratamento não ocorre de maneira aleatória e os grupos de tratamento e controle acabam apresentando características muito diferentes. Esses métodos têm como objetivo balancear os grupos com base em covariáveis, ou seja, outros dados além da variável de interesse. É preciso que esses dados não sejam afetados pelo tratamento. Por exemplo, em um caso em que haja mais de um período de tempo é adequado parear pelos dados do período anterior ao tratamento. Na prática, essa técnica seleciona os indivíduos do grupo de controle que sejam os mais semelhantes possíveis com relação aos indivíduos que receberam o tratamento. Normalmente, esse procedimento faz com que algumas observações do grupo de controle sejam eliminadas da amostra.

O pareamento pode ser feito de maneira exata, isto é, você pareia cada indivíduo tratado com um do grupo de controle com exatamente os mesmos valores nas variáveis definidas para parear. Raramente isso é possível, principalmente quando buscamos parear por valores contínuos (ex: renda). Por isso, a estratégia mais comum é parear utilizando escores de propensão. Resumidamente, é estimada uma regressão logística, em que a variável dependente é o tratamento. Assim, com base nos parâmetros estimados, é calculada a probabilidade do indivíduo pertencer ao grupo de tratamento. Assim, cada indivíduo do grupo de tratamento é pareado com um, ou mais, do grupo de controle que tenham essa probabilidade, ou “escore,” semelhante.

Um exemplo de trabalho que utiliza matching para avaliar uma política pública é artigo de Andrade et al. (2012). Os autores investigaram os impactos do Programa Bolsa Família na imunização de crianças de 0 a 6 anos utilizando dados de 2005. Nele foram comparadas famílias que estavam no programa com famílias elegíveis, mas não presentes no programa. O ponto principal da utilização deste método é que esses grupos apresentam características diferentes, portanto é necessária uma readequação do grupo de controle. Os resultados encontrados mostraram que o programa não afetou a imunização das crianças.

Diferenças-em-diferenças - É um método de dados em painel muito popular na literatura econômica que calcula o efeito de um tratamento comparando a mudança média ao longo do tempo na variável de interesse para os tratados, em comparação com a mudança média para o grupo de controle. O principal requisito para esse método poder ser aplicado é o de que as tendências de ambos os grupos na variável de interesse sejam paralelas antes do tratamento. Ou seja, a hipótese é de que as diferenças não observáveis entre os grupos se mantenha constante ao longo do tempo. Assim, o efeito do tratamento é o responsável por alterações de trajetória.

Muitas vezes, a hipótese de tendência comum não é atendida, o que faz com que muitos estudos utilizem pareamento para atingir isso. (Mattos and Mazetto 2019) utilizam exatamente essa abordagem para estimar o efeitos de curto prazo do programa Mais Médicos, lançado em 2013. Assim, eles usam dados de municípios de 2010 até 2015 e mostram que o programa teve ao menos dois efeitos positivos: o aumento em consultas, agendamentos e visitas a domicílios, e o impacto negativo na hospitalização. Entretanto, não encontraram efeito significativo na mortalidade.

Controle sintético - Esse método é de certa forma semelhante a um modelo de diferenças-em-diferenças, mas utilizando normalmente quando há apenas uma unidade trata. É feita a construção de uma combinação linear ponderada da variável de interesse das unidades do grupo de controle antes do tratamento ocorrer, o que faz com que a trajetória dessa combinação se torna semelhante à trajetória da unidade tratada. Assim, se as duas trajetórias divergirem após à implementação de uma política, isso pode se dever ao tratamento.

Por sua característica, esse é um método interessante para aplicação de avaliação de políticas nacionais. Um exemplo disso é o paper de Chamon, Garcia, and Souza (2017), que avaliou um programa de intervenção cambial por meio de swaps do Banco Central do Brasil em 2013. Utilizando outros países emergentes como controle, eles estimaram que a política levou a uma apreciação do Real em um excedente de 10%.

Variáveis instrumentais - É um método que busca uma variável externa (instrumento) que é correlacionada com as variáveis explicativas, mas não com a variável de interesse. Assim, essa técnica permite estimar uma regressão quando as variáveis explicativas são correlacionadas com o termo de erro, ou seja, quando há endogeneidade. Basicamente, este método permite o instrumento dividir as variáveis explicativas em duas partes. Uma delas é correlacionada com o erro e a outra não. Assim, tornando possível separar a parte que não tem correlação para a estimação da regressão de forma não viesada. A grande dificuldade é encontrar instrumentos que sejam válidos para esse método. Alguns instrumentos se tornaram bastante populares, como é o caso de variáveis ligadas ao clima, principalmente em estudos sobre países com forte presença agrícola. A lógica por trás disso é que a chuva, por exemplo, afeta a produção agrícola, mas não é afetada pelas condições econômicas.

Um exemplo bastante comum do uso de variáveis instrumentais é a estimação do retorno da educação nos salários. Isso acontece pois existe um fator não observável, que é a habilidade pessoal. Pessoas com maior habilidade tem maior probabilidade de fazer ensino superior. Ao mesmo tempo, a habilidade também afeta os salários de forma independente à educação formal. Assim, um bom instrumento precisa ser correlacionado com a probabilidade de ir à universidade, mas não pode ser correlacionado com os salários. Card (1993), propoem estimar o efeito utilizando a distância em relação às universidades. Outros autores, como Blackburn and Neumark (1993) utilizam a educação dos pais como instrumento.

Regressão descontínua - O objetivo dessa abordagem é comparar observações que estejam perto de um limite que estabelece um limite entre o grupo de tratamento e o de controle. A lógica desse método é a de que indivíduos próximos a essa limite seriam extremamente semelhantes, sendo a designação para um ou para outro grupo completamente aleatória. A aplicação é possivelmente mais restritiva do que os outros métodos citados e muito dependente dos dados disponíveis e do critério de aplicação da política pública.

Chamon et al. (2019) utilizaram uma descontinuidade nas regras eleitorais municipais de 2000 para investigar se a competição política teria impacto nas escolhas das políticas fiscais. A descontinuidade em questão era a de que, para municípios menores de 200 mil habitantes, não havia segundo turno. Os resultados mostram que a competição faz com que se aumentem os gastos com infraestrutura em relação aos gastos correntes. Além disso, o impacto encontrado foi maior para os casos em que os candidatos poderiam concorrer à reeleição.

Referências

- Andrade, Mônica Viegas, Flávia Chein, Laeticia Rodrigues de Souza, and Jaume Puig-Junoy. 2012. “Income Transfer Policies and the Impacts on the Immunization of Children: The Bolsa Familia Program.” *Cadernos de Saude Publica* 28: 1347–58.
- Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

- Blackburn, McKinley L, and David Neumark. 1993. “Omitted-Ability Bias and the Increase in the Return to Schooling.” *Journal of Labor Economics* 11 (3): 521–44.
- Card, David. 1993. “Using Geographic Variation in College Proximity to Estimate the Return to Schooling.” *NBER Working Paper*, no. w4483.
- Chamon, Marcos, Sergio Firpo, João MP de Mello, and Renan Pieri. 2019. “Electoral Rules, Political Competition and Fiscal Expenditures: Regression Discontinuity Evidence from Brazilian Municipalities.” *The Journal of Development Studies* 55 (1): 19–38.
- Chamon, Marcos, Márcio Garcia, and Laura Souza. 2017. “FX Interventions in Brazil: A Synthetic Control Approach.” *Journal of International Economics* 108: 157–68.
- Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. 2016. *Impact Evaluation in Practice*. The World Bank.
- Heckman, James J. 2008. “Econometric Causality.” *International Statistical Review* 76 (1): 1–27.
- Mattos, Enlinson, and Debora Mazetto. 2019. “Assessing the Impact of More Doctors’ Program on Healthcare Indicators in Brazil.” *World Development* 123: 104617.
- Peixoto, Betânia, Cristine Campos de Xavier Pinto, Lycia Lima, Miguel Nathan FOGUEL, Ricardo Paes BARROS, and Naercio MENEZES FILHO. 2012. *Avaliação Econômica de Projetos Sociais*. Fundação Itaú Social.