

# Trabalho Individual de Análise de Dados Econômicos II

Fellipe Silva

08 de dezembro de 2020

## Conteúdo

<b>Introdução</b>	<b>1</b>
<b>Artigo</b>	<b>2</b>
Principais conclusões . . . . .	2
Método utilizado . . . . .	3
<b>Reprodutividade</b>	<b>4</b>
Replicação do artigo e de seus resultados . . . . .	4
Documentação . . . . .	6
Código e seus obstáculos . . . . .	7
<b>Sugestões de reprodutividade</b>	<b>7</b>
<b>Importância do artigo</b>	<b>8</b>
<b>Sugestões para a extensão da análise</b>	<b>9</b>
<b>Conclusão</b>	<b>10</b>

## Introdução

A disponibilização de grandes bancos de dados, muitos deles de acesso aberto à comunidade científica e a sociedade como um todo, serviu como um belo incremento para a produção e extensão de pesquisas dentro e fora das universidades. Ao mesmo tempo, exige-se cada vez mais domínio “informático” dos cientistas em exercício, e dos cientistas em potencial, para que suas pesquisas tenham as qualidades necessárias para serem reproduzidas nos melhores jornais do campo acadêmico.

Além da qualidade das pesquisas e da transparência nos métodos utilizados, dá-se cada vez mais importância à abertura na divulgação de códigos e dados

utilizados na produção de “papers”. Com estes, torna-se possível reproduzir e averiguar mais a fundo os métodos supramencionados, assim como verificar se as conclusões destas pesquisas condizem com o que os dados têm para “dizer”.

À luz destes desenvolvimentos, verifica-se a grande importância de se aprender como reproduzir artigos científicos usando os métodos informáticos mais populares do campo das Ciências Econômicas, e também o aprendizado das linguagens de programação voltadas para a academia. Para além de oferecerem dicas preciosas nos processos individuais vindouros de desenvolvimento acadêmico, estes ensinamentos ajudam a preparar de forma mais rica a ciência como um todo, com futuros economistas prontos para contribuir na checagem constante dos padrões de qualidade das pesquisas do ramo.

## Artigo

O relatório presente delinea a reprodução do artigo “Voters’ response to in-kind transfers: Quasi-experimental evidence from prescription drug cost-sharing in Brazil”. O mesmo pode ser encontrado no jornal científico *Economic Letters*, cujo foco é a divulgação rápida e concisa dos mais novos achados das Ciências Econômicas em termos de métodos, modelos e resultados em pesquisas nos vários ramos da ciência (Lehmann & Matarazzo, 2019).

O artigo em questão foi escrito por Michael Christian Lehmann e Hellen Matarazzo, pesquisadores do Departamento de Economia da Universidade de Brasília, no Brasil. O artigo foi recebido em 31 de agosto de 2018 pela revista, e aceito quase um ano depois em 12 de agosto de 2019, sendo publicado na internet em 20 de agosto de 2019.

A classificação JEL do artigo é D72 - Processos Políticos e H72 - Orçamento e Despesas do Estado e do Poder Local (American Economic Association, 2020). As palavras-chave são “Eleições”, “Transferências por Bens e Serviços (‘in-kind’)” e “Políticas de Saúde”.

## Principais conclusões

O objetivo do artigo é contribuir ao ramo de Economia Política, avaliando os efeitos eleitorais da implementação de um programa de transferência de bens em áreas metropolitanas do Brasil chamado “Farmácia Popular” nas eleições municipais de 2008. O programa do governo federal foi lançado em 2005, distribuindo remédios de controle de hipertensão e diabetes de forma gratuita, e também oferecendo remédios para outras doenças a preços 90% abaixo do valor de mercado (Blog da Saúde, 2016).

No Brasil, cerca de 24% dos adultos do país tem hipertensão, e 8% tem diabetes (Correio Braziliense, 2020). Para a cidade ser elegível ao programa, é necessário que ela tenha pelo menos 70 mil habitantes. Além disso, é necessário que a

prefeitura requisite o programa junto ao governo federal, que usa as instalações de farmácias privadas para oferecer os remédios sob as condições descritas acima.

A pesquisa de Lehmann e Matarazzo encontrou um retorno positivo entre 11 a 17% em prefeitos em exercício que trouxeram para suas cidades o programa, na comparação entre as eleições municipais de 2004 e 2008. Este é um retorno acima dos encontrados em transferências monetárias (De La O, 2015), suportando os achados prévios da literatura científica que transferências de bens e serviços tem maior suporte por parte de eleitores em comparação a transferências monetárias, apesar de terem uma relação custo-benefício pior (Cunha, 2014).

## Método utilizado

O método econométrico utilizado para a avaliação dos resultados eleitorais foi o “design de regressão descontínua”. Este é um método utilizado comumente na avaliação de resultados estatísticos antes e depois de um teste ou experimento, onde se encontra uma “quebra” na tendência linear de um modelo a partir de um limiar que pode ser um piso ou teto em várias formas - desde o coeficiente de rendimento para seleção em um programa de bolsas acadêmicas, até o quociente populacional necessário para fazer parte de um programa federal. Este último é o caso do estudo a ser reproduzido. Pelo método, é possível identificar os efeitos causais das intervenções que motivaram a “quebra” a partir de valores em torno do limiar, que estão um pouco abaixo ou um pouco acima deste valor determinado (CAF, 2019).

Uma vez que a “quebra” no limiar não é em um valor absoluto entre 0 e 1 e sim em degraus variáveis, uma vez que cerca de 30% dos prefeitos com cidades elegíveis para se integrar ao “Farmácia Popular” não o fizeram - presumidamente por motivos de rivalidade política/ideológica e receio de “alimentar” votos para a oposição, segundo os autores do artigo - usa-se o design de regressão descontínua em formato difuso. Com isso, as chances dos estimadores serem enviesados é reduzida (Development Impact Evaluation, 2020).

O design de regressão descontínua usa o método de estimação de variáveis instrumentais, com o artigo lançando mão da regressão de mínimos quadrados em dois estágios. A regressão do primeiro estágio, para a variável independente  $FPB_{c,s}$ , é:

$$FPB_{c,s} = \alpha_1 + \alpha_2(\widetilde{pop}_{c,s}^{04} \geq 0) + f(\widetilde{pop}_{c,s}^{04}) + \varepsilon_{c,s}$$

- $FPB_{c,s}$  indica se uma cidade  $c$  em um estado  $s$  possui qualquer unidade de uma “Farmácia Popular”
- $(\widetilde{pop}_{c,s}^{04})$  é a distância entre a população de uma cidade e o limiar de 70 mil habitantes para entrada no Farmácia Popular em 2004
- $f(\widetilde{pop}_{c,s}^{04})$  é uma função polinomial que combina a variável acima com uma variável dummy com valores iguais a 1 se a população da cidade está acima

do limiar do programa, e 0 se a população da cidade está abaixo do limiar do programa

- $\varepsilon_{c,s}$  são os resíduos da regressão de primeiro estágio

E a regressão do segundo estágio, para a variável dependente  $VoteShare_{c,s}^{08}$ , é:

$$VoteShare_{c,s}^{08} = \beta_1 + \beta_2 \widehat{FPB}_{c,s} + f(\widetilde{pop}_{c,s}^{04}) + \mu_{c,s}$$

- $VoteShare_{c,s}^{08}$  é a percentagem de votos obtida pelo prefeito em exercício da cidade  $c$  no estado  $s$  nas eleições municipais de 2008
- $\widehat{FPB}_{c,s}$  são os valores estimados de  $FPB_{c,s}$  na regressão de primeiro estágio
- $\mu_{c,s}$  são os resíduos da regressão de segundo estágio
- $\beta_2$  indica o efeito da adesão ao programa Farmácia Popular na percentagem de votos para os prefeitos em exercício, na comparação entre as eleições de 2004 e 2008

Para determinar quais cidades seriam analisadas via design de regressão descontínua, os pesquisadores usaram larguras de banda para determinar as observações a serem avaliadas nas equações a partir da relação  $-h < \widetilde{pop}_{c,s}^{04} < h$ , com  $-h$  e  $h$  sendo os limites inferiores e superiores da banda, respectivamente. Uma vez que não existe consenso sobre qual a largura de banda ótima para o parâmetro de suavização dentro do design de regressão descontínua, os autores usaram os três métodos proeminentes: Ludwig e Miller, Imbens e Kalyanaraman, e Calonico et al.

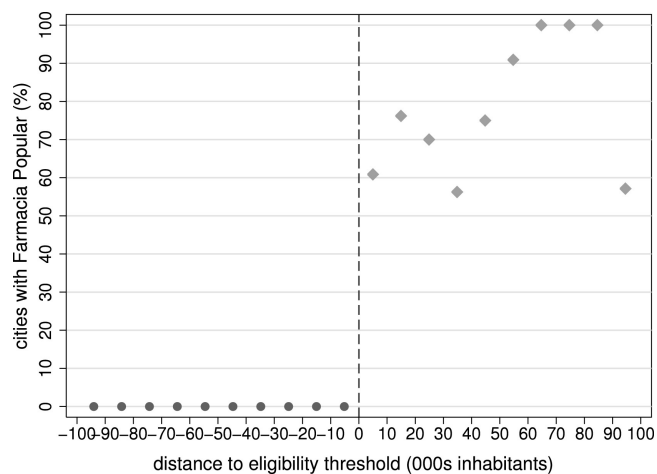
## Reprodutividade

Os arquivos para reprodução do artigo podem ser encontrados no Mendeley Data, um repositório aberto para dados de pesquisa que podem ser compartilhados entre pesquisadores e também para o público. Os “datasets” publicados no repositório contam com um DOI, ou “identificador de objeto digital” que é único.

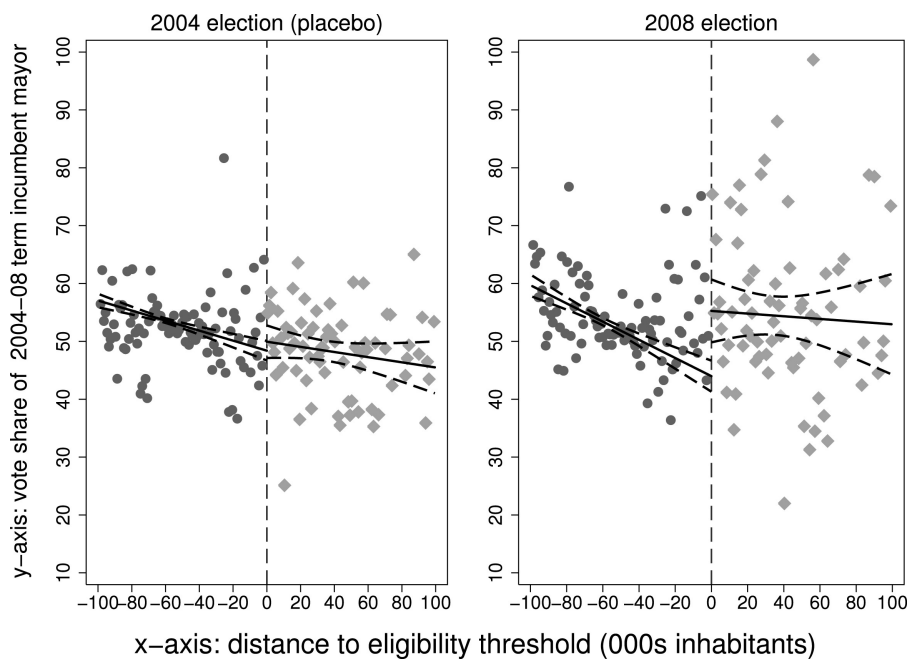
O artigo possui o DOI 10.1016/j.econlet.2019.108614. E o dataset possui o DOI 10.17632/3t9wshp34p.1.

## Replicação do artigo e de seus resultados

O código e o dataset tem a intenção de reproduzirem através do Stata as figuras abaixo (estas foram retiradas do próprio Economic Letters, no portal Science Direct):



Adesão ao Farmácia Popular conforme proximidade com o limítar populacional do programa



Farmácia Popular e percentagem de votos em prefeitos em exercício, nas eleições municipais de 2004 e 2008

Table 1

FPB and mayor vote shares.

Dep. var. in panels A,B,C:		Vote share of 2004-2008 term incumbent mayor (%)					
Election		2004 (placebo)	2008	2004 (placebo)	2008	2004 (placebo)	2008
<i>Panel A. OLS estimates</i>							
$FPB_{c,t}$		0.3 [2.56]	7.0** [3.10]	1.1 [2.63]	5.7* [3.20]	-5.0* [2.40]	-6.9 [6.83]
<i>Panel B. Reduced-form estimates (sharp regression discontinuity design)</i>							
$\widehat{pop}_{c,t}^{04} \geq 0$		1.7 [3.82]	11.7*** [3.26]	3.9 [3.82]	10.4*** [3.19]	0.5 [5.86]	8.1 [6.10]
<i>Panel C. 2SLS estimates (fuzzy regression discontinuity design)</i>							
$FPB_{c,t}$		2.6 [5.74]	18.1*** [5.48]	6.6 [6.31]	17.6*** [6.02]	1.1 [11.65]	17.1 [11.19]
<i>First-stage estimates</i>					Dep. var.: $FPB_{c,t}$		
$\widehat{pop}_{c,t}^{04} \geq 0$		0.65*** [0.07]				0.47*** [0.11]	
Kleibergen-Paap F-statistic		77.5				50.7	17.1
Bandwidth method		LM				CCT	
Bandwidth $h$		99,000				16,000	
Observations (cities)		2,856				62	
Sample mean of dep. var.:		53.37	53.45	53.32	52.98	49.99	55.05

Notes: Square brackets contain robust standard errors, clustered at the State level.

\*Statistically significant at 10%.

\*\*Statistically significant at 5%.

\*\*\*Statistically significant at 1%.

Farmácia Popular e percentagem de votos em prefeitos em exercício, nas eleições municipais de 2004 e 2008

## Documentação

Os autores do artigo mencionaram no mesmo a existência de um apêndice online que adiciona informações que corroboram a validade do design de regressão descontínua a partir de quatro pontos:

1. Cidades em torno do limiar tem as mesmas características socioeconômicas entre o lançamento do programa em 2005, e as eleições municipais em 2008;
2. Não há um incremento notável no número de cidades com população um pouco acima do limiar;
  - Logo, prefeitos não manipularam dados para aumentar artificialmente o número de habitantes de suas cidades.
3. Não há diferença no alinhamento do prefeito com o presidente em exercício em torno do limiar;
  - Logo, não houve escolha a partir de ideologias (a partir do governo federal) no processo de implementação do programa “Farmácia Popular”
4. Outras políticas públicas que utilizam “pisos” populacionais têm poucas chances de terem influenciado os resultados encontrados na pesquisa

Entretanto, o apêndice não foi encontrado nem no Mendeley Data, nem em outras partes da internet durante pesquisas no Google. Por isso foi enviado no dia 02 de dezembro de 2020 um e-mail aos pesquisadores solicitando o envio deste arquivo, caso ele esteja disponível.

## Código e seus obstáculos

De início, o do-file apresentou um problema: a dependência RDROBUST não era instalada a partir de uma linha de código já presente no arquivo, e sim no cabeçalho do mesmo. E uma vez que se entrava no website, descobria-se que o mesmo havia sido descontinuado. Por sorte, o que houve na verdade foi uma atualização do código, que era agora encontrado no website RD Packages.

Outro problema é a utilização do comando `cd` para definir a pasta onde o arquivo “zip” com o dataset foi salvo. Esta pasta também teria em si todo o output do processo, desde o log até as figuras e a tabela descritas acima. Além disso, o dataset original que era alterado para a realização do processo de regressão, não era salvo em nenhuma das etapas.

Infelizmente o código original não funcionou, e foi possível apurar que o erro foi causado pelo comando `rdbwselect_2014`. Este é um comando da versão antiga do pacote RDROBUST, que aparentemente não se encontra na sua versão atual. Além disso, nem a documentação do pacote nem o website que hospeda o mesmo contam com algum item explicando como o antigo código poderia ser adaptado para funcionar em sua nova versão.

No processo de tentativa e erro, a eliminação da tentativa de reproduzir a tabela de regressão do artigo foi o que aparentemente deu certo em primeiro momento. Mas logo depois encontrou-se outro erro: a falta de instalação do pacote CMOGRAM para a reprodução das figuras 1 e 2.

Uma vez que o pacote foi instalado por meio do `ssc install`, a reprodução foi bem sucedida. Entretanto a figura 2, que é a junção de dois gráficos, teve como “output” apenas a sua versão final - o que pode ser prejudicial em um processo de extensão da pesquisa presente no artigo.

Depois houve a tentativa de se adaptar o código problemático usando o código mais novo, que era usado para reproduzir as colunas 5 e 6 da tabela de regressão. Mas a tentativa não foi bem sucedida.

No processo de tentativa de correção do erro, não foram encontradas referências ao comando `rdbwselect_2014` para o Stata. Este foi encontrado somente para o RStudio.

## Sugestões de reprodutividade

São várias as sugestões de reprodutividade. As mais importantes seriam:

1. Em termos de organização, o ideal seria utilizar uma “árvore” de pastas, com uma pasta para o dataset original, uma pasta para os datasets alterados, uma pasta para os do-files, uma pasta para os logs e uma pasta final para o “output” em figuras e tabelas

- Logo, lançaria mão do comando `capture mkdir` para a execução deste processo

```
capture mkdir O_ECOLET-FS
capture mkdir O_ECOLET-FS\1_source
capture mkdir O_ECOLET-FS\2_data
capture mkdir O_ECOLET-FS\2_data\1_auxiliary_data
capture mkdir O_ECOLET-FS\3_progs
capture mkdir O_ECOLET-FS\4_logs
capture mkdir O_ECOLET-FS\5_output
```

2. Adicionar a cada uma destas pastas um arquivo `.txt` explicando o que cada arquivo significa
3. Para evitar a necessidade de alterar o `do-file` com a pasta-destino do dataset, seria melhor ter o dataset em um repositório online com acesso livre ao mesmo, permitindo que fosse feito um download das variáveis necessárias para o programa
  - Semelhante ao que ocorre no SDMX
4. Ao invés de ter a dependência principal do `do-file` no cabeçalho do mesmo, o ideal seria ter uma linha de código que já instalasse a mesma no Stata
  - Assim, uma das primeiras linhas de comando do `do-file` seria

```
net install rdrobust, from(https://raw.githubusercontent.com/rdpackages/rdrobust/master)
```
5. O mesmo precisa ser feito para o pacote CMOGRAM, necessário para a reprodução das figuras do artigo, que não foi mencionado na linha de código do `do-file`
  - `ssc install cmogram`
6. É mais do que necessário atualizar o código de reprodução das tabelas, checando se existe alguma documentação que explica se existe compatibilidade entre o código antigo e o código novo
7. E seria interessante manter as figuras originais que montaram a figura 2 do artigo, para referências futuras.

## Importância do artigo

A principal contribuição do artigo para o campo de Economia Política é a verificação dos efeitos diretos de uma política pública específica nos resultados eleitorais de várias cidades em um país de dimensões continentais e com grande diversidade socioeconômica como o Brasil. De acordo com os autores, a maior parte das pesquisas que tentam encontrar estas relações geralmente lançam mão de conceitos como as despesas públicas agregadas, com resultados mistos entre



si. O uso das despesas públicas agregadas impede que se encontre quais políticas públicas geram os efeitos encontrados, sejam eles positivos, negativos ou neutros.

Além disso, a pesquisa realizada é uma das poucas do campo que verificam os efeitos das transferências de bens e serviços nos resultados eleitorais, uma vez que pesquisas anteriores cobriram o efeito de transferências monetárias (De La O, 2015). E na comparação com uma das pesquisas mais proeminentes desse âmbito, os retornos eleitorais encontrados foram superiores.

Os resultados do artigo validam algumas das teorias vigentes do campo de Economia Política listados pelos autores, como o fato das transferências em bens e serviços terem maior suporte por parte de eleitores por manterem incentivos ao trabalho, melhorarem a definição dos públicos-alvo que são beneficiados por estes programas, e satisfazer preferências paternalistas. Além disso, os ganhos de utilidade dos beneficiários do programa (Da Motta et al, 2015), do governo federal através das reduções de mortalidade e hospitalizações por conta de problemas de hipertensão e diabetes (Américo & Rocha, 2017), e até do público não beneficiado diretamente são significativos.

Estes retornos de utilidade podem explicar o retorno eleitoral do programa, apesar dos programas de transferência monetária apresentarem vantagens de custo-benefício (através de custos administrativos mais baixos, e maior flexibilidade dada aos seus beneficiários) sobre transferências de bens e serviços. Por isso, é bem provável que os programas de transferência em bens e serviços continuarão a ser escolhidos por governantes, apesar dos protestos de economistas.

## Sugestões para a extensão da análise

O método utilizado no artigo para encontrar os retornos eleitorais sobre um programa de política pública específico, é bastante promissor. Por isso, seria de bom grado ver o mesmo método sendo utilizado para analisar os retornos eleitorais de outros programas de política pública em outras cidades, e checando desta forma quais destes tem efeitos positivos, negativos e/ou neutros nas eleições municipais do Brasil. E a pesquisa poderia também tentar fazer uma análise “geográfica”, observando variações populacionais e/ou eleitorais em cidades vizinhas às que foram beneficiadas por estes programas, mas que não os receberam.

Seria também interessante verificar, se possível, o impacto de grandes obras públicas nos resultados de zonas eleitorais específicas em grandes cidades brasileiras. Algumas das grandes cidades do país como São Paulo e Rio de Janeiro possuem populações na casa das dezenas de milhões de pessoas, com bairros que funcionam como pequenos municípios dentro destas localidades. E verifica-se à primeira vista que eleitores destes bairros agraciam prefeitos em exercício que executam grandes obras na região, algo que pode ser vital nos esforços de reeleição à prefeitura ou no lançamento de candidaturas em esferas acima (estadual e federal).

Para além do Brasil, a pesquisa poderia verificar os impactos das políticas públicas que envolvem tanto transferências monetárias, quanto transferências de bens e serviços, nos resultados eleitorais e fazer comparações diretas entre os mesmos. Isso poderia verificar a diferença nas preferências de populações que envolvem níveis educacionais e socioeconômicos diferentes entre si.

Outra extensão é verificar potenciais diferenças de preferência entre programas de transferência em dinheiro, e de transferência em bens e serviços, nos diferentes níveis de renda e capital de uma mesma localidade. É uma pesquisa que poderia trazer luz às “preferências paternalistas” mencionadas anteriormente.

Uma última sugestão é ver a duração do efeito destes programas nos retornos eleitorais dos governantes. Pode-se presumir que existe um grande impacto inicial que vai se dissipando ao longo do tempo, mas seria interessante verificar se isso realmente ocorre, e qual a dimensão desse efeito caso ele aconteça.

## Conclusão

Por décadas, a pesquisa acadêmica do campo econômico foi predominantemente teórica. Muito disso se explica pelas limitações tecnológicas que não permitiam o acesso a ferramentas com poder de processamento necessário para esmiuçar dados que muitas vezes já se encontravam disponíveis a partir de pesquisas censitárias e também em pesquisas de campo. Mas este panorama mudou radicalmente nos últimos tempos, com máquinas domésticas podendo realizar cálculos a velocidades antes inimagináveis pelos antigos cientistas de dados.

Mas nem toda pesquisa acadêmica precisa de um grande arcabouço de dados, algo que ocorre no artigo que acaba de ser analisado. Com um corte temporal de dois momentos, e dados das eleições das mais de 5 mil cidades brasileiras, foi possível analisar o impacto de um programa do governo federal que segundo os resultados da análise presente, impulsionou os votos dos prefeitos em exercício em até 17% em comparação a seus pares e ao resultado do ano anterior. E enquanto este impacto poderia ser identificado de forma individual e qualitativa, fazendo comparações cidade a cidade, os métodos econômicos e econométricos se destacam por permitirem a mensuração deste efeito de forma quantitativa e massiva, economizando principalmente o tempo dos pesquisadores.

Infelizmente o processo de reprodução dos resultados não foi totalmente eficaz, o que mostra que mesmo artigos que poderiam ser considerados diminutos também podem se mostrar problemáticos em suas reproduções. Mas isso serve como um excelente aprendizado aos cientistas em potencial que aprendem com os seus próprios erros, e também com os erros de seus pares, adotando para si as melhores práticas organizacionais do campo para permitir que pesquisas sejam reproduzidas e também apuradas.