

MATERIAL SUPLEMENTAR

Carreiras políticas na Câmara dos Deputados:
Uma análise quase-experimental

Setembro, 2016

Este documento apresenta resultados e detalhes adicionais do artigo *Carreiras políticas na Câmara dos Deputados: Uma análise quase-experimental*.

Anexo A

Conforme discutido na seção sobre o desenho de pesquisa, as variáveis dependentes analisadas são *dummies* que indicam candidatura e vitória em eleições futuras. A tabela abaixo apresenta algumas estatísticas descritivas destas variáveis.

Tabela 1 – Estatísticas descritivas das variáveis utilizadas na análise

Variável	N	Média	Desvio	Min	Max
Posição na lista	16,609	-1875	2044	-115	24
Margem Bruta	16,609	-43,61947	58,16702	-493,768	1,503,156
Margem (%)	16,609	-0.10	0.15	-0.97	0.96
Concorrer t+4	16,609	0.17	0.37	0	1
Vencer t+4	16,609	0.06	0.24	0	1
Concorrer dep. est. t+4	16,609	0.06	0.24	0	1
Vencer dep. estadual t+4	16,609	0.01	0.07	0	1
Concorrer vereador t+2	16,609	0.18	0.38	0	1
Vencer vereador t+2	16,609	0.03	0.16	0	1
Concorrer executivo t+2	16,609	0.06	0.24	0	1
Vencer executivo t+2	16,609	0.01	0.09	0	1
Concorrer t+2	16,609	0.24	0.43	0	1
Vencer t+2	16,609	0.03	0.18	0	1
Concorrer t+8	16,609	0.12	0.32	0	1
Concorrer t+12	16,609	0.05	0.21	0	1
Vencer t+8	16,609	0.04	0.19	0	1
Vencer t+12	16,609	0.02	0.12	0	1

Nota – A tabela apresenta a estatística descritiva das variáveis utilizadas na análise. As entradas são estatísticas numéricas, indicadas pelo título das colunas, das variáveis indicadas pela primeira coluna. As três primeiras variáveis são contínuas e, as demais, variáveis binárias, onde 1 indica o evento de interesse.

Anexo B

Ao longo do artigo, são discutidos alguns testes de validação, que servem para verificar se o efeito do tratamento ocorre independentemente da sua atribuição. Para o desenho de regressão descontínua utilizado no artigo, estes testes foram feitos de duas formas: primeiro, por meio da análise do efeito do tratamento em variáveis que sabemos ser anteriores ao tratamento, i.e. pré-tratamento (i.e., a literatura chama estes procedimentos de testes de placebo); segundo, utilizando falsas descontinuidades. Os resultados dos dois conjuntos de teste são reportados abaixo. Utilizando o bens declarados dos candidatos, os gastos de campanha e a filiação partidária como *outcomes*, a Tabela 2 mostra que os grupos de tratamento e de controle não são significativamente diferentes em seus perfis e campanhas. A Tabela 3, por sua vez, mostra que o efeito causal encontrado no artigo desaparece nestas especificações, corroborando a validade do desenho.

Tabela 2 – Efeito do tratamento em *outcomes* pré-tratamento

Variável	Estimador	IC (95%)	N
Bens declarados (em milhares)	-3254.76	[-7379.13, 869.6]	2505
Total gasto em campanha (em milhares)	-307.28	[-2518.18, 1903.61]	1842
Total gasto com pessoal (em milhares)	-61.57	[-628.32, 505.19]	2782
Total gasto em comícios (em milhares)	1.55	[-2.12, 5.22]	3274
Concorrer pelo PT	0.03	[-0.04, 0.09]	2942
Concorrer pelo PMDB	0.04	[-0.02, 0.09]	3983
Concorrer pelo PSDB	-0.02	[-0.07, 0.03]	3377

Nota – * p-valor < 0.05. Os intervalos de confiança, entre colchetes, foram calculados a partir de erros-padrão robustos. As janelas usadas foram selecionadas através do método proposto por Imbens e Kalyanamaram (2011). As estimativas mostram o efeito causal de ser incumbente sobre *outcomes* pré-tratamento, como gastos de campanha e filiação partidária.

Tabela 3 – Efeito do tratamento utilizando descontinuidades falsas

Variável	Estimador	IC (95%)	N
Concorrer t+4 (<i>cutoff</i> = 0.02)	-0.06	[-0.15, 0.03]	2062
Concorrer t+4 (<i>cutoff</i> = 0.05)	0.03	[-0.13, 0.19]	684
Concorrer t+4 (<i>cutoff</i> = -0.02)	-0.13*	[-0.2, -0.07]	2429
Concorrer t+4 (<i>cutoff</i> = -0.05)	-0.02	[-0.06, 0.02]	5273
Vencer t+4 (<i>cutoff</i> = 0.02)	0.05	[-0.06, 0.16]	1666
Vencer t+4 (<i>cutoff</i> = 0.05)	0.12	[-0.06, 0.3]	714
Vencer t+4 (<i>cutoff</i> = -0.02)	-0.08*	[-0.1, -0.05]	3462
Vencer t+4 (<i>cutoff</i> = -0.05)	-0.01	[-0.02, 0.01]	2992

Nota – * p-valor < 0.05. Os intervalos de confiança, entre colchetes, foram calculados a partir de erros-padrão robustos. As janelas usadas foram selecionadas através do método proposto por Imbens e Kalyanamaram (2011). As estimativas mostram o efeito causal de ser incumbente sobre concorrer e sobre concorrer e vencer eleições para diversos cargos dois e/ou quatro anos depois utilizando falsas descontinuidades.

Anexo C

A operacionalização da *forcing variable* pode influenciar a seleção de observações usadas nos modelos estimados. Ao usar a margem bruta de votos, distritos maiores são penalizados, já que as diferenças totais de votos entre dois candidatos dentro de uma lista tendem a ser maiores; por outro lado, diferenças percentuais de votos são maiores onde existem menos eleitores. O mapa a seguir ilustra esse ponto. A Tabela 4, abaixo, contudo, mostra que o uso da margem bruta de votos não altera os principais resultados reportados, sobre a reeleição.

Gráfico 1 – Observações selecionadas por estado usando diferentes forcing variables

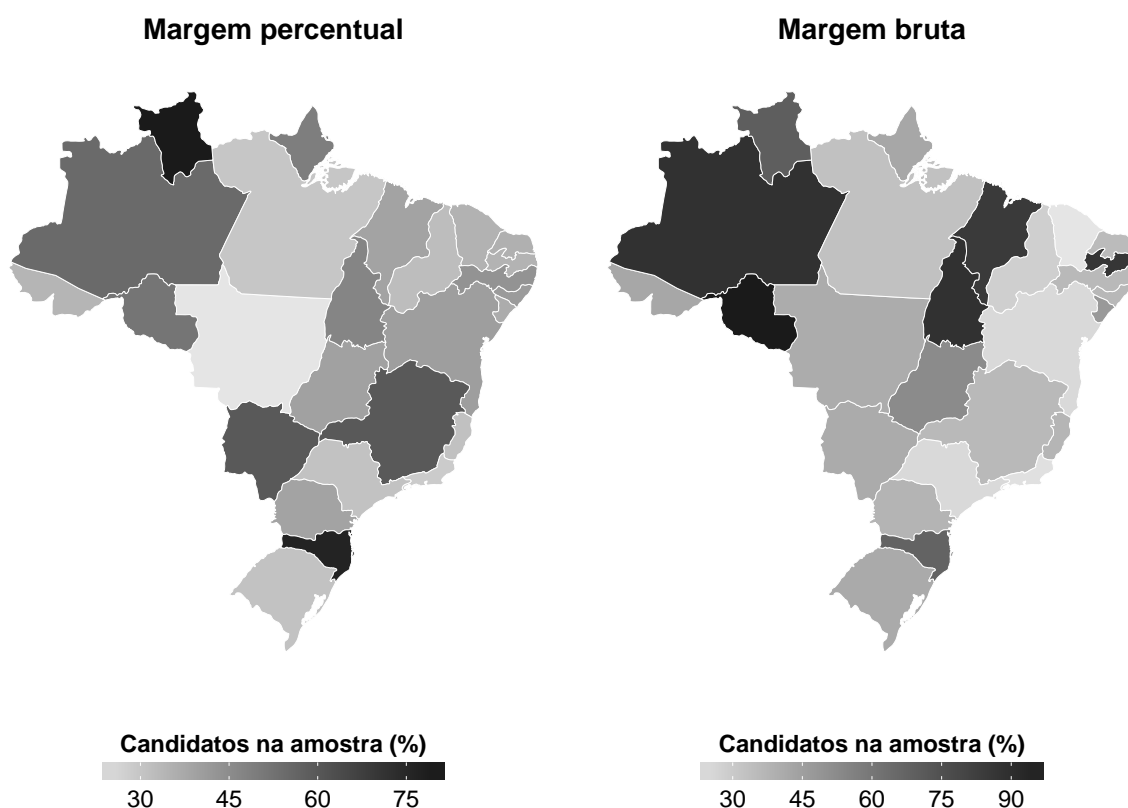


Tabela 4 – Estimativas do efeito do tratamento local médio de vencer em t na eleição para o mesmo cargo em t_{+4} utilizando as margens brutas de votos como *forcing variable*

		Polinômio	1	2	3	4
Concorrer em t_{+4}	Estimador		0.24*	0.21*	0.22*	0.14*
	Erro		(0.04)	(0.05)	(0.05)	(0.06)
	BW		19639.3	34261.71	60838.43	48198.19
	N		2293	4256	7623	5953
Vencer em t_{+4}	Estimador		0.3*	0.23*	0.22*	0.19*
	Erro		(0.03)	(0.04)	(0.05)	(0.05)
	BW		48045.5	38148.72	52791.64	56939.16
	N		5934	4712	6454	6974

Nota – * p-valor < 0.05. Erros-padrão robustos entre parêntesis. As janelas foram selecionadas através do método proposto por Imbens e Kalyanaramam (2011). Os modelos estão ordenados segundo a ordem dos polinômios usados para estimá-los. As estimativas mostram o efeito causal de ser incumbente sobre concorrer e sobre concorrer e vencer a eleição para o mesmo cargo 4 anos depois.

Anexo E

As Tabelas 5 e 6 reanalisam os principais achados do artigo com outras amostras. Isto é feito para garantir que o desequilíbrio na amostra original – existem mais candidatos no grupo de controle (não-eleitos) do que no de tratamento (eleitos) – não afeta os resultados. Como pode ser visto, mesmo usando apenas os últimos eleitos e os primeiros suplentes, ou restringindo a amostra para ter um mesmo número de observações nos dois grupos, os resultados não se alteram.

Tabela 5 – Estimativas do efeito do tratamento local médio de vencer em t na eleição para o mesmo cargo em t_{+4} (amostra apenas com último eleito e primeiro suplente)

		Polinômio	1	2	3	4
Concorrer em t_{+4}	Estimador		0.14*	0.15	0.16	0.16
	Erro		(0.07)	(0.09)	(0.11)	(0.13)
	BW		0.02	0.03	0.03	0.03
	N		468	488	478	480
Vencer em t_{+4}	Estimador1		0.2*	0.23*	0.23*	0.28*
	Erro		(0.07)	(0.08)	(0.09)	(0.11)
	BW		0.03	0.03	0.03	0.03
	N		484	492	522	508

Nota – * p-valor < 0.05. Erros-padrão robustos entre parêntesis. As janelas foram selecionadas através do método proposto por Imbens e Kalyanaraman (2011). Os modelos estão ordenados segundo a ordem dos polinômios usados para estimá-los. As estimativas mostram o efeito causal de ser incumbente sobre concorrer e sobre concorrer e vencer a eleição para o mesmo cargo 4 anos depois utilizando uma amostra com apenas últimos eleitos e primeiros suplentes.

Tabela 6 – Estimativas do efeito do tratamento local médio de vencer em t na eleição para o mesmo cargo em t_{+4} (amostra com número igual de eleitos e não-eleitos)

		Polinômio	1	2	3	4
Concorrer em t_{+4}	Estimador		0.25*	0.17*	0.15*	0.11
	Erro		(0.04)	(0.06)	(0.07)	(0.09)
	BW		0.03	0.02	0.03	0.03
	N		2073	1786	2121	2131
Vencer em t_{+4}	Estimador		0.29*	0.27*	0.25*	0.19*
	Erro		(0.04)	(0.06)	(0.07)	(0.08)
	BW		0.03	0.03	0.03	0.03
	N		1818	1951	2115	2078

Nota – * p-valor < 0.05 . Erros-padrão robustos entre parêntesis. As janelas foram selecionadas através do método proposto por Imbens e Kalyanaraman (2011). Os modelos estão ordenados segundo a ordem dos polinômios usados para estimá-los. As estimativas mostram o efeito causal de ser incumbente sobre concorrer e sobre concorrer e vencer a eleição para o mesmo cargo 4 anos depois com uma amostra balanceada.

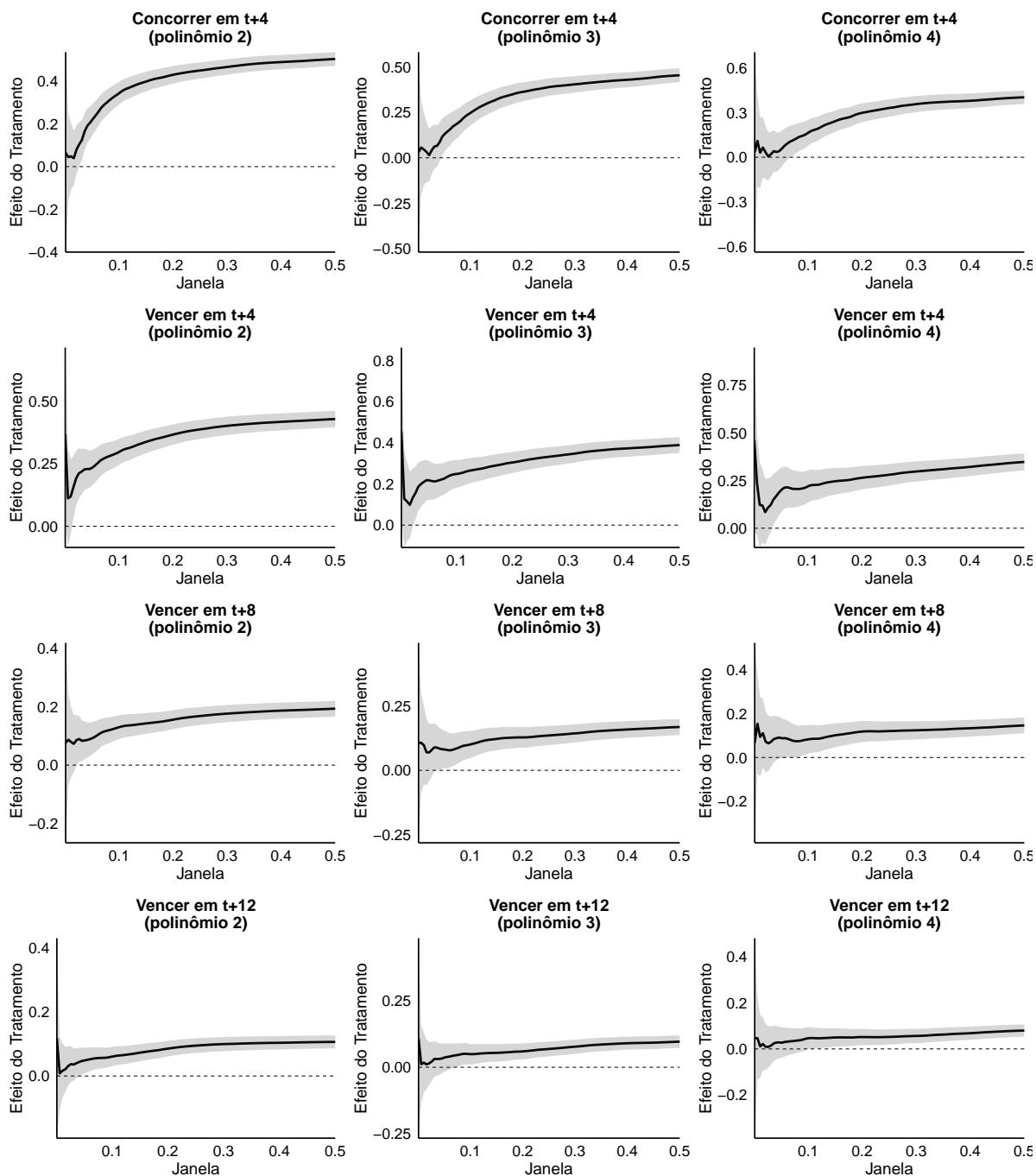
Anexo F

Também reestimei alguns resultados com polinômios nos termos lineares (i.e., *forcing variable* ao quadrado e ao cubo). Estes modelos seguem a seguinte forma:

$$Y_{ijt} = \alpha + \beta T_{ijt} + \gamma_1 M_{ipt} + \delta_1 T_{ijt} M_{ipt} + \gamma_2 M_{ipt}^2 + \delta_2 T_{ijt} M_{ipt}^2 + \dots \gamma_n M_{ipt}^n + \delta_n T_{ijt} M_{ipt}^n + \epsilon_{ipt}$$

Como os gráficos abaixo mostram, apesar de variações, os resultados se mantêm.

Gráfico 2 – Uso de polinômios nos modelos



Anexo G

Na subseção do texto intitulada *Resultados adicionais*, são reportados, em forma de gráficos, os efeitos marginais do tratamento variando a magnitude dos distritos e o *status* dos candidatos, isto é, se eram ou não incumbentes no momento da eleição. Aqui, segue a tabela com os resultados desses modelos, usados para criar os gráficos.

Tabela 7 – Resultados completos do gráfico 6

	Disputar a reeleição (dummy)	
	(1)	(2)
Tratamento	0.099 (0.064)	0.393*** (0.033)
Magnitude do distrito	-0.004*** (0.001)	
Tratamento * Magnitude	0.005*** (0.001)	
Incumbente		0.124*** (0.044)
Tratamento * incumbente		-0.119* (0.061)
Constante	0.593*** (0.044)	0.340*** (0.019)
Janela	1%	1%
N	1,255	1,255

*p < .1; **p < .05; ***p < .01

Anexo H

Por fim, cabe uma nota quanto à preparação dos dados. No artigo, menciono um procedimento de *fuzzy matching*, usado para identificar candidatos ao longo do tempo. Como os dados do TSE são alimentados pelos próprios candidatos, erros de grafia e digitação são comuns, o que faz com que seja difícil cruzar ocorrências de uma mesma pessoa em eleições diferentes. Para piorar, dados como data de aniversário possuem formatações distintas em diferentes períodos (e.g., 22/10/1950 e 22 de out. 1950, ou ainda 22-10-50) e, até recentemente, dados como número de CPF e número do título de eleitor não eram integrados à base (mesmo que com erros de digitação).

Para lidar com isto, implementei um algoritmo chamado *Levenshtein distance*¹, que mensura quão diferentes são duas palavras ou expressões de acordo com o número mínimo de caracteres que teriam que ser movidos ou acrescentados para compatibilizar as palavras. Por exemplo, para igualar João a Jãoo, precisaríamos apenas deslocar ã de lugar na segunda palavra, contabilizando uma operação. Posteriormente, este número é ponderado pelo total de letras nas palavras e retorna um *score* de 0 a 1, onde 1 indica que dois termos são idênticos. Com esse algoritmo, implementado numa combinação de

¹ Disponível em: https://en.wikipedia.org/wiki/Edit_distance. Acesso em: 13/08/2016.

nome inteiro, UF e data de aniversário, identifiquei cerca de 2% dos candidatos cujas informações não batiam – ainda que o número seja irrisório e não provoque qualquer alteração nos resultados (porque, provavelmente, a distribuição deste tipo de erro está distribuída aleatoriamente entre os candidatos).

Referências

IMBENS, G.; KALYANARAMAN, K. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, Oxford University Press, p. rdr043, 2011.