

A falácia distrital nos estudos eleitorais

Sergio Simoni Junior¹

sergiojr_ssj@yahoo.com.br

Rafael Nunes Magalhães²

rnunesmagalhaes@gmail.com

Trabalho preparado para apresentação no VIII Seminário Discente da Pós-Graduação em Ciência Política da USP, de 7 a 11 de maio de 2018

¹ Doutor em Ciência Política pela USP. Realiza pós-doutorado no CESOP/UNICAMP

² Doutorando em Ciência Política pela USP.

Introdução

O objetivo desse artigo é demonstrar que a devida incorporação da dimensão do distrito eleitoral nas análises sobre determinantes das eleições em sistemas majoritários, como as eleições presidenciais e para governador no Brasil, tem consequências importantes nos resultados encontrados e, particularmente, para algumas teses correntes na literatura. A competição eleitoral, como muitos objetos de estudo da Ciência Política, consiste em resultado agregado de comportamentos individuais. O voto é um ato individual, mas também um fenômeno social e político, nos seus determinantes e nas suas consequências (Przeworski e Sprague, 1986). A relação entre as dimensões macro e micro dos fenômenos sociais perpassa grande parte do desenvolvimento teórico e metodológico das Ciências Sociais (Cho e Manski, 2010; Sprinz, 2000; King, 1997; Achen e Shively, 1995; Dogan e Rokkan, 1969).

Os estudos eleitorais, entretanto, foram fortemente influenciados pela escola de Michigan do comportamento eleitoral, que estabeleceu o survey como instrumento primordial de análise eleitoral (Campbel et. al., 1980; Bartels, 2010). O comportamento individual se tornou a unidade a ser explicada. Dados agregados, segundo essa visão, seriam necessariamente enviesados pois sofreriam de falácia ecológica, ou seja, seriam incapazes de explicar o comportamento individual devido a problemas de viés de agregação.

Um conjunto de literatura levanta argumentos que matizam essa crítica. Esses trabalhos procuram demonstrar que as dimensões contextuais, agregadas e/ou geográficas, seriam importantes para explicar o próprio comportamento individual. Para essa visão, a ênfase exclusiva em variáveis individuais cairia em uma falácia individualista, ao desconsiderar a importância dos efeitos agregados (Katz e King, 1999; Kramer, 1983; Wright, 1989; Huckfelt e Sprague, 1986).

Esse debate, no entanto, ignora uma dimensão central. Os votos de cada eleitor são somados e convertidos em cadeiras de acordo com regras determinadas, que definem, além dos mecanismos matemáticos de conversão, os limites geográficos onde os sufrágios serão agregados, ou em outros termos, os distritos eleitorais. Assim, para além do debate acerca da importância dos diferentes níveis e mecanismos de explicação do comportamento eleitoral individual, o ponto que chamamos atenção diz respeito à extrapolação desses efeitos para determinantes dos resultados eleitorais.

O problema ocorre quando as análises estão baseadas em porcentagens como variável dependente e/ou independente calculadas em subunidades de tamanho desigual, como os municípios no caso brasileiro, por exemplo. Ao proceder dessa forma, os modelos estatísticos estão estimando efeitos médios de porcentagens, que, no entanto, são desiguais em tamanho absoluto. Como o resultado eleitoral não consiste na média de porcentagens das subunidades do distrito eleitoral, mas sim na sua soma absoluta, a extrapolação direta dos efeitos substantivos a partir de coeficientes estimados em análises desse tipo cairia em um tipo de viés que chamamos de *falácia distrital*. Nosso objetivo neste artigo é discutir as propriedades dessa falácia, suas consequências empíricas e propor soluções.

Para tanto, adotamos duas estratégias argumentativas. Primeiro, procedemos a simulações de um cenário de competição eleitoral e de sua análise em níveis subdistritais. Nesses exercícios, mostraremos o viés de agregação que ocorre quando se extrapola, para explicação do resultado eleitoral, efeitos substantivos a partir de porcentagens de votos e porcentagens de uma variável de interesse calculadas em unidades do distrito.

Segundo, discutimos como caso empírico particular o debate recente sobre eleições majoritárias no Brasil. Existe praticamente um consenso, ainda que com exceções e refinamentos, de que a distribuição do Programa Bolsa-Família (daqui em diante, chamado PBF), programa de transferência de renda iniciado em 2003 implementado pelo PT, foi importante para a conformação dos resultados eleitorais presidenciais de 2006 a 2014, ajudando a explicar as eleições e reeleições de candidatos desse partido (p. ex., Hunter e Power, 2007; Nicolau e Peixoto, 2007; Zucco, 2008; Soares e Terron, 2008; Zucco, 2013; Limongi e Guarnieri, 2015). Um conjunto associado de literatura estuda os impactos recíprocos desse fenômeno nas eleições para governador, particularmente em estados nordestinos, historicamente associados à existência de “oligarquias” locais e clientelismo e que nos últimos pleitos assistem à emergência de partidos de esquerda (Montero, 2010, 2012; Borges, 2010; 2011).

Dada a importância para o debate, nos concentramos na eleição de 2006. Nossa análise se baseia em dados e modelos próprios e em replicação de dois estudos importantes do debate: Zucco (2013), para o pleito presidencial, e Montero (2010), que

analisa eleições para governadorias estaduais³. Nosso objetivo é discutir o impacto substantivo das variáveis independentes de interesse desses estudos. Mostraremos como esse impacto substantivo se altera quando estimado com falácia distrital, ou seja, atribuindo, de forma equivocada, o mesmo peso aos municípios, em comparação com estimativa adequada, que incorpora de fato o distrito eleitoral.

Nossa proposta de adequação do modelo estatístico ao distrito eleitoral é simples e direta: consiste em recuperar em votos absolutos, município a município, os efeitos percentuais preditos da variável de interesse. O efeito total estimado da variável de interesse é obtido pela soma dos valores absolutos de todos os municípios. Dessa forma, mantemos o município como unidade de análise, concordando com a crítica da falácia individualista, mas evitamos a falácia distrital.

Os resultados tanto das simulações, quanto das análises empíricas própria e replicadas mostra a importância desse exercício. No caso particular, mostramos que os efeitos eleitorais do PBF são menos pronunciados do que se obtém a partir da extrapolação direta dos coeficientes, e que, ademais, *coattail effects* de Lula em 2006 não se mostram importantes para compreender a variação da performance de partidos de esquerda no nível estadual em alguns estados do Nordeste. Como argumento geral, ressaltamos a importância de se levar a sério o distrito nas análises de determinantes dos resultados eleitorais.

Falácia distrital e falácia individualista

As ciências sociais e particularmente os estudos eleitorais foram fortemente impactados pelo artigo de Robinson (1950)⁴. O autor demonstra lógica e empiricamente que correlações entre fenômenos medidos ao nível agregado não podem ser imputadas ao nível das unidades que compõe essa agregação.

O exemplo utilizado pelo autor é didático para a exposição desse problema. Suponha-se que se quer estudar a relação entre grau de escolaridade e marcadores sociais da diferença entre os indivíduos. Mais especificamente, se quer saber se analfabetismo está relacionado à raça e à origem nacional.

³ Agradecemos a Cesar Zucco e Alfred Montero o compartilhamento de suas bases de dados, e ressaltamos seu espírito público e colaborativo para a comunidade acadêmica.

⁴ Em maio de 2017, o google acadêmico contabilizava mais de 5000 citações do artigo.

Utilizando dados censitários dos anos 30 dos EUA, Robinson (1950) encontra que a correlação entre raça e nacionalidade com índices de analfabetismo estimada ao nível dos estados norte-americanos (logo, nível agregado) apresenta valores consideravelmente diferentes (e no caso da origem nacional, até em sentido inverso) aos oriundos de uma correlação ao nível dos indivíduos.

Dito de outra forma, as conclusões que uma analista poderia atingir acerca das relações entre escolaridade e características sociais seriam diferentes, e até mesmo opostas, a se depender do nível de análise considerado. Mais do que isso, e aqui o autor generaliza sua crítica, dado que a unidade de explicação última é o indivíduo, proposições baseadas ao nível agregado correriam o risco de encontrar relações espúrias, incorrendo em diagnósticos explicativos equivocados. Selvin (1958), alguns anos depois, cunhou o termo “falácia ecológica” para designar a influente crítica proposta por Robinson (1950)⁵.

Robinson (1950: 357) encerra seu texto com uma prescrição:

“the purpose of this paper will have been accomplished (...) if it prevents the future computation of meaningless correlation and stimulates the study of similar problems with the use of meaningful correlations between properties of individuals”.

Em certa medida sua intenção foi atingida. Nas palavras de Firebaugh (2001: 4024) *apud* Subramanian *et al.*, 2009) “it would be difficult to overstate the impact Robinson’s article has had on social science research during the second half of the twentieth century”. No campo específico dos estudos eleitorais, a valorização de informações e análises focadas no comportamento individual foi notável.

Desde os anos 50, assistiu-se a uma proliferação de aplicação e análise de *surveys* eleitorais, produzindo talvez a mais extensa bibliografia na Ciência Política. A tradição da “escola de Michigan” do comportamento eleitoral impactou fortemente a disciplina, com ênfase em explicações de cunho cognitivo, como a identificação partidária, coletadas em *surveys* nacionais (Bartels, 2010).

Evidentemente, fatores sociais não eram descartados, mas o foco da explicação estava nas características dos indivíduos. Abordagens vigorosas tais como os estudos de Key Jr (1955; 1959), sobre as dinâmicas de longo-prazo nos resultados eleitorais em

⁵ Cabe notar que os termos “unidade” e “indivíduo” não designa apenas pessoas, mas qualquer unidade de análise que é passível de agregação de alguma forma.

unidades agregadas do território norte-americano acabaram sendo relegadas para um segundo plano. Análises ecológicas eram, segundo essa visão, necessariamente falaciosas.

As proposições de Robinson (1950), no entanto, não ficaram inunes à críticas. Já a partir dos anos 60 (Alker, 1969), alguns trabalhos chamam atenção para o que Subramanian *et al.* (2009) iriam cunhar de “falácia individualista”. Basicamente, são dois tipos de argumentos. Primeiro, a ausência de incorporação de variáveis contextuais acabaria por prejudicar a explicação ao nível dos indivíduos⁶; segundo, alerta-se que as prescrições formuladas por Robinson (1950) acabaram por levar, muitas vezes, a se incorrer em equívoco semelhante ao identificado pelo autor, mas em sentido inverso, a saber, extrapolação de associações encontradas ao nível das unidades para os fenômenos coletivos.

A primeira crítica foi também formulada por Schwartz (1994) como um problema de validade interna e por Hanushek *et al.* (1974) como um problema de especificação do modelo, notadamente viés de variável omitida. De fato, ao replicar os dados de Robinson (1950) incluindo variáveis sobre políticas educacionais estaduais, e portanto, uma dimensão agregada, Subramanian *et al.* (2009) explicam melhor a escolaridade individual, fenômeno de interesse do autor.

A segunda crítica feita à ênfase em dados individuais é importante para esse trabalho. Como nota Schwartz (1994), as perguntas de pesquisa e os fenômenos de interesse muitas vezes não são, e nem devem ser, relativos ao nível micro ou individual, como supõe Robinson (1950). Assim, conclusões sobre o nível agregado a partir de associações analisadas no nível micro são também um tipo de erro de falácia ecológica.

Grande parte dos estudos eleitorais lança mão de unidades agregadas subdistritais como nível empírico de análise. A maioria dos estudos recentes sobre as eleições presidenciais e para governador no Brasil, em particular, está baseada em análises de dados municipais. Tendo em vista o debate apresentado acima, pode-se dizer que esse procedimento permite captar efeitos indiretos, de caráter local, sobre o comportamento

⁶ Sem mencionar ainda outros problemas inerentes ao *survey*, como erro de medida causado por viés de *desejabilidade social* (respostas não sinceras ditadas por constrangimentos sociais), esquecimento por parte do respondente, interpretação incorreta da pergunta, não resposta, etc.

eleitoral. No entanto, pouco se atenta para a forma adequada de extrapolação de efeitos substantivos de análises desse tipo para o resultado eleitoral.

Na próxima seção, apresentaremos o caso particular que embasa nosso argumento, a saber, debate da literatura brasileira sobre eleições presidenciais e para governador.

Estudos recentes sobre eleição presidencial e para governador no Brasil

O debate recente sobre o caso brasileiro das eleições presidenciais e para governador ilustra como essas questões são tratadas de forma geral nos estudos eleitorais. Utilizaremos esse conjunto de literatura como exemplo de nossa proposição metodológica, que, no entanto, tem caráter geral.

Uma das teses centrais nos estudos presidenciais recentes, ainda que não sem importantes exceções e resultados matizados⁷, é de que o PBF tem importante efeito eleitoral. Grande parte das análises é realizada com dados agregados, que encontram correlação entre porcentagem de beneficiários do PBF e porcentagem de votos do PT⁸ de 2006 a 2014 (Hunter e Power, 2007; Zucco, 2008, 2013, 2015; Nicolau e Peixoto, 2007; Soares e Terron, 2008; Canêdo-Pinheiro, 2015; Marzagão, 2013; Magalhães *et al.*, 2015; Pereira *et al.*, 2015; Zucco, 2015; Martins *et al.*, 2016).

Esse fenômeno empírico é correlato de uma tese maior acerca de um possível realinhamento eleitoral que ocorreria desde 2006 (Singer, 2012). Segundo essa interpretação, locais e eleitores pobres, que outrora votavam em partidos conservadores, passaram a apoiar majoritariamente o PT.

Parte da literatura estuda esses fenômenos no nível estadual. Montero (2010; 2012) e Borges (2010; 2011) chamam atenção para o fato de que partidos de esquerda conseguiram, em 2006, pela primeira vez eleger governadores em alguns estados nordestinos historicamente associados à existência de “oligarquias” locais e clientelismo, como a Bahia e o Maranhão. No entanto, os resultados empíricos desses autores levam-nos a minimizar a importância do PBF e ressaltar antes o efeito da votação de Lula

⁷ Para argumentos mais matizados, ver Carraro *et al.* (2009); Limongi e Guarnieri (2015); Corrêa (2015).

⁸ Zucco (2013) encontra um efeito também do Programa Bolsa-Escola e Bolsa-Alimentação nos votos para o então incumbente PSDB em 2002.

(*coattail effect*) e o fato de que a votação das esquerdas está associada com o grau de urbanização dos municípios.

No debate sobre as eleições presidenciais, alguns autores levantam cautela para as inferências encontradas em dados agregados, como os municípios. Bohn (2011; 2013), por exemplo, argumenta que os resultados advindos de dados agregados por municípios mostram apenas que existe uma correlação positiva entre porcentagem de famílias beneficiárias do PBF e os votos do PT no município. Por si só, esses resultados não permitem afirmar que são os indivíduos que recebem o PBF que estão votando no partido. Afirmer isso seria incorrer em falácia ecológica.

Para essa visão, o único mecanismo plausível de explicação dos resultados eleitorais deve ser de ordem individual. Os dados agregados, nas palavras de Lício *et al.* (2009), conteriam um “elo perdido” explicativo. Mesmo estudos baseados em dados agregados, muitas vezes, consideram esse nível um *second-best* em comparação com informações sobre o comportamento individual. Canêdo-Pinheiro (2015: 34, grifos nossos), por exemplo, nota que: “por conta do mecanismo do voto secreto, é necessário o uso de dados agregados para investigar o impacto do programa Bolsa Família”. Em nota de rodapé, afirma que “alguns autores tentam contornar esse problema usando dados individuais extraídos de pesquisas de opinião. Entretanto, muitas dessas pesquisas sequer conseguem reproduzir o percentual de votação obtido pelos candidatos nas eleições de 2006. Desse modo, *não há como escapar do uso de dados agregados*”.

Minimizando essa posição, alguns autores levanta argumentos que ressaltam a importância de efeitos indiretos e contextuais ao nível local, em consonância com a crítica da falácia individualista. Zucco (2015), por exemplo, defende que o número de beneficiários do PBF ao nível municipal impactou o comportamento eleitoral local de não-beneficiários. Diversos estudos quantitativos e qualitativos mostram que a distribuição do programa social é importante para o funcionamento da economia nacional (Mostafa *et al.*, 2010; Neri *et al.*, 2013) e principalmente para a economia local dos pequenos municípios, por seu efeito multiplicativo no comércio de alimentos, medicamentos e vestuário, afetando, portanto, as condições econômicas de não-beneficiários (Paes-Sousa e Vaitsman, 2007; Landim Jr. 2007; Cohn, 2012; Rego e Pinzani, 2013; Marques, 2005). No caso das análises para eleição de governador, o nível de urbanização dos municípios é ressaltado como importante para compreender a diferença na força dos partidos.

A importância de dados agregados não é alheio a teorias de comportamento eleitoral. Por exemplo, Kramer (1983) ressalta sua importância na análise do voto econômico, Huckfelt e Sprague (1986) e Zuckerman (2005) chamam atenção para o impacto das redes de contato; Agnew (1996) e Cox (2005) ressaltam que, em si mesma, a localidade é um elemento chave da competição eleitoral⁹; e Wright (1989) defende a importância de mecanismos de mediação entre comportamento micro e resultados macro, inclusive das regras dos distritos eleitorais. Nas palavras de Katz e King (1999: 15) “survey research has enormous advantages for studying individual-level preferences, but as analyses of random selections of isolated individuals from unknown geographical locations, they necessarily miss much of electoral politics. As such, they are often best complemented with studies of aggregate electoral returns”.

No caso brasileiro, que estamos discutindo, municípios parecem ser uma boa unidade de análise agregada. Além de informações municipais servirem de base para a distribuição do PBF (ver Fried, 2009; Paes de Barros *et al.*, 2008), e de outras políticas, as cidades são também a última instância de poder político do federalismo brasileiro, e prefeitos podem ter importante poder de *brokers* na intermediação na eleição de outros cargos, como para governador, por exemplo (Avelino *et al.* 2012; Ames, 1994).

No entanto, a ênfase no debate sobre as possibilidades e limites dos níveis de análise, com foco nos problemas de inferência a partir dos dados agregados, tem como consequência a desatenção a outro ponto tão importante, a saber, a extrapolação dos dados a partir das cidades para a competição presidencial ou pra governador.

Os modelos estatísticos empregados estão estimando *o efeito médio da distribuição do PBF (ou de qualquer outra variável) na votação percentual municipal do PT*¹⁰. A extrapolação direta dos impactos encontrados para o resultado eleitoral implica em assumir que o sistema eleitoral da eleição presidencial ocorre como um colégio eleitoral com cerca de 5.500 unidades de igual importância. Dentro dessa lógica, 60% de votos em um município de 1.000 eleitores teriam a mesma utilidade para um partido que busca a vitória eleitoral que 60% em um município de 5.000.000 de eleitores. Em outras

⁹ Para mais detalhes consultar revisão elaborada por Ethington e McDaniel (2007) e o volume editado por Cox (2005).

¹⁰ Utilizamos aqui o exemplo de estudos presidenciais, mas o mesmo raciocínio, com as devidas adaptações, pode ser usado para o caso das eleições para governador.

palavras, ao se concentrar no valor dos coeficientes, implicitamente, assume-se que as unidades de agregação são homogêneas.

No entanto, como é sabido, as eleições presidenciais brasileiras têm como distrito todo o território nacional. Um voto ganho (ou perdido) em uma pequena localidade como Borá (SP) tem o mesmo peso que um voto ganho (ou perdido) em São Paulo. Entretanto, 1% de votos ganhos (ou perdidos) em cada um destes municípios representam valores absolutos completamente diferente. Em Borá, 1% dos eleitores significam menos de 10 votos em 2006, em São Paulo, cerca de 63.000.

O resultado eleitoral é obtido no distrito eleitoral. Não se trata de médias de municípios, mas antes da sua agregação. Isso significa que o efeito do PBF, ou de qualquer outra variável, no resultado eleitoral, não pode ser obtido a partir de resultados médios obtidos a partir dos municípios, pois esses têm peso muito desigual na composição do eleitorado nacional.

Para demonstrar com propriedade esse argumento, primeiramente elaboramos simulações que apresentamos abaixo, para posteriormente discutir empiricamente.

Falácia distrital – simulações

Nesta seção, procedemos a duas simulações para mostrar a importância da falácia distrital. Na primeira, mais simples e descritiva, mostraremos como a média de porcentagens de votos dos partidos com base em subunidades de tamanho heterogêneo de um distrito é um mecanismo equivocado de agregação de votos, que não condiz com a regra eleitoral vigente nos sistemas majoritários.

Na segunda simulação, mostraremos como esse procedimento tem consequências para inferências de impacto substantivo realizadas a partir de regressões.

Simulação I

Imagine um país hipotético composto por quatro cidades, como no quadro abaixo. A cidade A, a maior de todas, teve 1000 votos válidos em uma eleição presidencial. A cidade B apresentou 100, e as C e D, tiveram 10 votos válidos cada, totalizando 1120 votos válidos nacionais.

Quadro I - Eleição presidencial país hipotético

Cidade	Votos válidos	Votos partido A	Votos partido B	% partido A	% partido B
A	1000	600	400	60	40
B	100	40	60	40	60
C	10	2	8	20	80
D	10	4	6	40	60

Fonte: Elaboração própria.

Existem dois competidores. O partido A obteve 600, 40, 2 e 4 votos, em cada cidade respectivamente. Isso significa 60%, 40%, 20% e 40%. O partido B, de forma complementar, teve 40%, 60%, 80% e 60% de votos em cada cidade.

Vamos supor que neste país hipotético os juízes eleitorais podem escolher entre dois sistemas eleitorais para transformar os votos na cadeira presidencial. No primeiro, trata-se de fazer uma média da proporção de votos obtidos por cada partido nos municípios. As cidades, nesse caso, atuam como colégios eleitorais de igual tamanho. Nesse caso, como mostra o quadro II, o partido B sairia vencedor, com 60% da média de votos por município.

Quadro II Resultado eleitoral sob 2 sistemas eleitorais

	Partido A	Partido B
1- Média de votos por cidade	40	60
2- % Votos válidos	57,68	42,32

Fonte: Elaboração própria.

No segundo sistema eleitoral em exame, somam-se os votos absolutos recebidos por cada partido em todas as cidades, e depois se dividem pelo total de votos válidos no país. Assim, o partido A obteve $(646/1120) = 58\%$ dos votos e o partido B, $(474/1120) = 42\%$, e, dessa feita, o partido A seria eleito.

Esse simples exemplo hipotético serve para mostrar a importância da crítica que estamos elaborando. Um sistema eleitoral majoritário agrega os votos em um distrito determinado, como no caso do segundo sistema eleitoral examinado. Uma agregação diferente, com base na média de porcentagem de votos, por exemplo, é equivocada.

Na simulação abaixo verificaremos as consequências da falácia distrital nas inferências obtidas em regressões.

Simulação II

Na segunda simulação, criamos aleatoriamente com uma distribuição beta, 5000 observações (equivalentes aos municípios) de tamanho variável entre 1000 e 10000 de votos válidos. Estabelecemos uma variável independente de interesse (que pode ser entendida, em casos particulares, como a distribuição do PBF, o grau de urbanização, *coattail effects*, etc) de distribuição normal, com média 33% e desvio-padrão de 1%. A distribuição absoluta dessa variável em cada unidade é obtida pela multiplicação de seu percentual com o número de votos válidos em cada município.

Nossa variável dependente são os votos absolutos do partido A. Em cada unidade de observação, essa variável é definida da seguinte forma:

$$Voto A_{absoluto,i} = 0,8 * V.I._{absoluto,i} + \varepsilon_i$$

Ou seja, a cada aumento de 1 ponto absoluto na variável independente aumenta em 0,8 o número predito de votos absolutos do partido A. O termo de erro (ε) tem distribuição normal com média zero e 50 de desvio-padrão.

Suponhamos que uma pesquisadora está interessada em estudar esse cenário eleitoral ressaltando a importância dos efeitos locais que ocorrem nas unidades subdistritais. Para tanto, ela trabalha com porcentagens de votos e da variável independente em cada município. Qual estratégia de análise essa pesquisadora deve adotar para obter impactos substantivos da variável independente ao nível do distrito eleitoral?

Existem duas possibilidades. A primeira consiste em, após estimação da regressão em porcentagens de votos, aplicar o coeficiente em cada município, de acordo a distribuição porcentual da variável independente em cada unidade. Esse valor porcentual, então, é recuperado em votos absolutos em cada município, de acordo com seu tamanho. Os resultados de todos os municípios são depois somados para se obter o impacto total da variável explicativa no resultado eleitoral. Assim, se mantém o nível local como

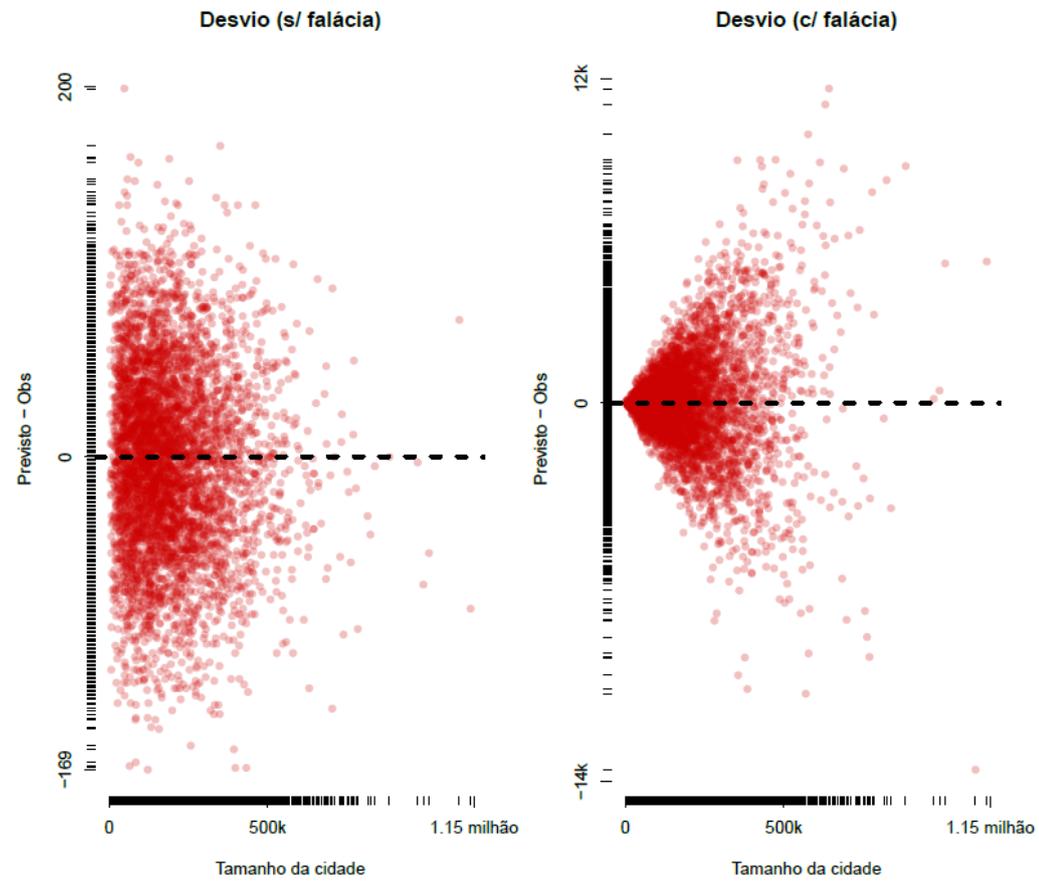
importante para a análise, mas se respeita a regra do distrito eleitoral nas inferências sobre os determinantes do resultado eleitoral.

A segunda possibilidade comete a falácia distrital. Após a estimação da regressão em porcentagens de votos, calcula-se efeitos substantivos da variável de interesse multiplicando-se o coeficiente por sua distribuição, por exemplo, por sua média. De modo geral, esta é a forma mais intuitiva e direta de se obter efeitos substantivos a partir de regressões. O impacto em votos absolutos por município, nesse cenário, pode ser obtido por multiplicação desse efeito, que está em termos percentuais, com o número de votos válidos em cada município. A posterior soma desses valores resultaria no impacto geral.

No caso de eleições majoritárias cujos distritos são compostos por subunidades de tamanho e distribuição desigual das variáveis de interesse, esse segundo raciocínio comete o equívoco da falácia distrital, pois assume equivalência entre as unidades subdistritais.

Para demonstrar as consequências e as diferenças de ambos procedimentos, procedemos da seguinte forma. Verificamos, em cada município, como as estimativas obtidas em cada método se desvia do observado estabelecido acima. Os resultados são apresentados no gráfico abaixo:

Gráfico 1 – Diferença entre observado e predito na agregação sem e com falácia distrital



No lado esquerdo temos os resultados advindos do método que respeita as regras do distrito eleitoral, ou seja, que não comete a falácia distrital. Observando-se a escala do eixo y, vê-se que os desvios são pequenos e devem-se ao termo de erro da regressão. No lado direito temos a comparação dos efeitos substantivos estimados sob a falácia distrital com os valores observados. O cenário é completamente diferente: os desvios são elevados e aumentam conforme o tamanho das cidades.

Os resultados advindos dessa simulação mostram a importância do ponto que estamos levantando. A extrapolação direta dos coeficientes estimados a partir de porcentagens de variáveis calculadas a nível sudistrital de tamanho heterogêneo é falaciosa.

Para verificar como isso ocorre em casos empíricos, vejamos o caso brasileiro

Análise empírica

Nossa estratégia analítica será a seguinte. Inicialmente iremos trabalhar com dados e modelos próprios, mostrando o impacto substantivo de se calcular equivocadamente os efeitos substantivos de análises baseadas em porcentagens municipais. Depois, replicaremos estudos de Zucco (2013), para o nível presidencial, e Montero (2010), para nível estadual.

Nossa análise com dados próprios diz respeito ao nível presidencial e na discussão sobre os efeitos do PBF. Nos concentramos na eleição de 2006, considerada pela literatura eleição crítica, balizadora dos pleitos posteriores. Utilizamos o modelo básico OLS. A variável dependente é o percentual de votos válidos no primeiro turno do PT em 2006 por cidade¹¹. A variável explicativas de interesse é a cobertura do Programa Bolsa-Família, mensurada como a proporção de domicílios que receberam o programa social em outubro de 2006¹². Para escolha das variáveis de controle, adotamos indicadores comuns à literatura. O IDH permite dimensionar em que medida o impacto da cobertura do PBF no voto está relacionado com o desenvolvimento social do município; o tamanho da população (adotado aqui em forma logarítmica) e a % de população urbana funciona como *proxies* de “grotões”, e buscam verificar se o PT de fato teve mais votos em cidades menores e rurais; a proporção de habitantes de minorias raciais¹³ e de evangélicos pentecostais busca controlar por características sociais importantes da sociedade com possíveis impactos políticos-eleitorais¹⁴.

As variáveis econômicas que utilizo são o crescimento do PIB municipal *per capita* no ano eleitoral, ou seja de 2006 em relação a 2005; e a proporção do PIB formada pelo setor público, e que é utilizada pela demonstrar a hipótese do governismo (Zucco, 2008; Canêdo-Pinheiro, 2015; Ribeiro, 2014). São incluídas medidas de poder local do

¹¹ Mesmo não sendo um município, incluí Brasília devido ao seu tamanho e sua importância política. Toda minha análise utiliza dados do primeiro turno. Utilizar segundo turno, seja de 2002, seja de 2006, não altera os resultados.

¹² Na medida em que um dos critérios de elegibilidade ao PBF é a renda domiciliar *per capita*, é mais adequado ponderar os beneficiários do programa social pelo número de domicílios antes que por famílias, como usualmente faz a literatura.

¹³ Me refiro às categorias censitárias de pretos, pardos, amarelos e indígenas.

¹⁴ Alguns estudos incluem medidas como porcentagem de pobres, analfabetos, PIB per capita, no entanto essas variáveis são altamente correlacionadas com o IDH e a população urbana, de modo que, para evitar multicolinearidade, não são incluídas aqui. A substituição entre essas medidas não afeta os resultados. Todas as variáveis percentuais estão na escala 0-100.

PT: uma *dummy* que indica se o prefeito pertence a este partido e outra que considera se o governador incumbente era do PT¹⁵. Por fim, são incluídas *dummies* regionais.

Abaixo se encontra uma tabela com a distribuição dos valores dessas variáveis:

Tabela 1 – Estatísticas descritivas

Variáveis contínuas					
Variável	Média	Desvio-Padrão	Mínimo	Máximo	Observações
% PT 2006	51,61	17,85	11,40	93,36	5565
% PBF	33,59	19,23	0,00	91,50	5565
% PT 2002	42,42	12,05	5,77	79,79	5563
(Ln) população	9,3859	1,1554	6,7190	16,2149	5564
IDH	0,59	0,09	0,339	0,841	5565
% População Urbana	61,17	22,61	4,42	100	5565
% Minorias raciais	49,98	24,45	0,50	96,45	5564
% Pentecostais	9,61	5,87	0,00	49,10	5564
% PIB setor público	31,49	16,65	1,65	93,67	5564
% Δ PIB 2006-2005	8,77	15,99	-60,35	387,44	5564

Variáveis categóricas			
Variável binária	0	1	Observações
Governador PT	5242	323	5564
Prefeitura PT	5160	404	5564

Fonte: elaboração própria.

A literatura já diagnosticou em diversas ocasiões que as observações ao nível dos municípios são relacionadas entre si. A solução convencionalmente utilizada, a partir de testes apropriados, é o modelo de erro espacial (Soares e Terron, 2008; Carraro *et al.*, 2009; Marzagão, 2013).

Ainda que necessário para corrigir por viés e eficiência, os modelos de erro espacial são os menos interessantes teoricamente pois não permitem analisar por quais

15 Os dados que utilizo são de fonte primária, coletados no IBGE, IPEADATA e TSE. IDH, população urbana, minorias raciais e pentecostais são coletadas ao nível municipal apenas em Censo, ou seja, em 2000 e 2010. De modo a buscar se aproximar do valor dessas variáveis em 2006, adotei uma média das medidas censitárias. Os dados de prefeitura se referem ao partido do prefeito em 2005, coletados na Munic/IBGE.

mecanismos ocorrem as relações espaciais. Em outras palavras, corrigem o problema da autocorrelação tratando-a apenas como “ruído” (Ward e Gleditsch, 2008). Além disso, e importante ressaltar, os resultados apresentados pela literatura que utiliza modelos espaciais não alteram as conclusões substantivas do modelo OLS.

Dado isso, e de modo a permitir maior flexibilidade às análises aqui empreendidas, utilizamos o modelo OLS com erro-padrão *clusterizado* ao nível dos estados. Isso significa que assumimos que municípios da mesma unidade de federação têm relação entre si em variáveis não observáveis, ou seja, seus erros são correlacionados¹⁶. Além de corrigir por heterocedasticidade, esse procedimento também se justifica substantivamente, pois é de esperar cidades de mesmo estado tenham características políticas, sociais e culturais semelhantes. A aplicação de modelos de regressão espacial (resultados não mostrados aqui) não alteram as conclusões.

A tabela abaixo, mostra os resultados da regressão:

Tabela 2 – Regressão OLS - % PT presidente

VD % PT 2006	Coefficiente	Erro-padrão
% PT 2002	0.360***	(0.0519)
% PBF	0.332***	(0.0464)
Sul	0.378	(2.082)
Centro-Oeste	-5.715***	(1.672)
Norte	-3.801	(3.140)
Nordeste	5.310**	(2.275)
Prefeitura PT	-1.713**	(0.797)
Governador PT	-4.835**	(2.003)
% Δ PIB 2006-2005	0.0550***	(0.0137)
% População Urbana	0.0231	(0.0210)
IDH	-17.29	(12.85)
(Ln) população	0.888**	(0.391)
% Minorias raciais	0.240***	(0.0486)
% Pentecostais	0.109	(0.0644)
% PIB setor público	0.135***	(0.0426)
Constante	7.295	(7.377)
<i>N</i>	5563	
adj. <i>R</i> ²	0.726	

Os erros-padrões são *clusterizados* ao nível das UF.

p*< 0.1, *p*< 0.05, ****p*< 0.01

¹⁶ Snijders e Bosker (2012: 329) dizem que, assim como modelos de erro espacial, a utilização de erro-padrão *clusterizado* é uma forma de tratar a dependência entre as observações.

Os resultados de ambos os modelos mostram que o coeficiente do PBF é significativo e a magnitude do efeito encontrado é semelhante a outros da literatura (Zucco, 2008; Corrêa, 2015).

De modo geral, é a partir da visualização desse coeficiente que se baseia o diagnóstico de que o PBF tem efeitos importantes no resultado eleitoral presidencial, principalmente para o pleito de 2006. Soares e Terron (2008: 270), por exemplo, afirmam que sua “hipótese (...) é que as políticas públicas contaram nas eleições de 2006¹⁷”. Hunter e Power (2007: 24) sentenciam que “Lula’s victory in 2006 may well go down in history for its association with the Bolsa-Família”. Corrêa (2015: 531), em síntese de sua argumentação, diz que “segue a abordagem corrente da literatura de elevar o programa [Bolsa-Família] ao patamar de principal variável explicativa para o padrão eleitoral de 2006”. Lício *et al.* (2009: 37), em balanço da literatura, afirmam que “esse debate deixa claro que o Programa Bolsa Família teve implicações significativas para o resultado da eleição de 2006”.

Esses efeitos teriam ocorrido também nas eleições posteriores. Analisando o pleito de 2010, Magalhães *et al.* (2015: 535) defendem que “os resultados indicam um forte efeito conjunto do Bolsa Família e do ex-presidente Lula para as votações recebidas por Dilma”. Em trabalhos sobre 2014, Zucco (2015) e Martins *et al.* (2016) defendem que o programa social manteve ou mesmo aumentou sua importância eleitoral.

No entanto, como ressaltam King *et al.* (2000), uma simples visualização dos coeficientes e significância estatística não é suficiente para inferir os impactos substantivos das variáveis de um modelo de regressão. O mais adequado é extrair valores esperados ou preditos do fenômeno de estudo de acordo com certas distribuições das variáveis independentes.

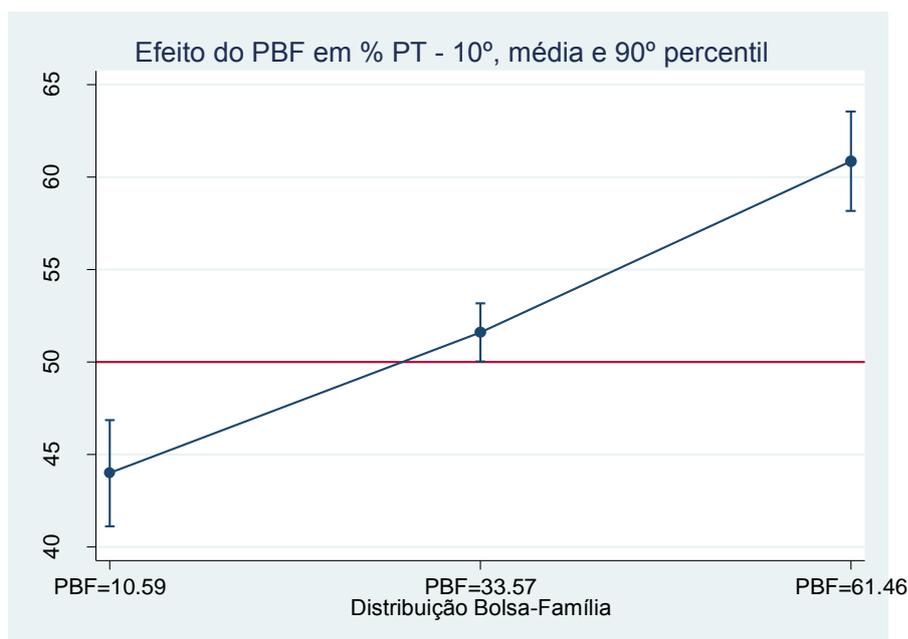
Ainda que essa prática seja mais comum nos trabalhos que utilizam modelos não-lineares, - tal como procedem, por exemplo, Canêdo-Pinheiro (2015) e Souza e Cribari Neto (2013), que utilizam regressões beta, -, mesmo em modelos OLS, nos quais o efeito marginal é constante, faz-se necessário exercícios desse tipo, pois os impactos substantivos das variáveis depende da distribuição da amostra.

¹⁷ Apesar do plural, analisam apenas os PBF.

Nesse sentido, baseado em raciocínio de Montero (2010: 126), adotamos o seguinte critério: comparamos o efeito estimado do PBF nos votos do PT quando esta variável assume valor de seu último decil (no caso, 10,59%), quando assume valor médio (33,57%), e quando assume valor do primeiro decil (61,46%).

No gráfico abaixo, apresentamos os resultados a partir de um “*average predictive margins*”, onde alteramos apenas os valores do PBF, seguindo essas especificações, e mantivemos as demais variáveis em suas médias, estimando valores preditos da variável dependente.

Gráfico 2 – % PT 2006 predito a partir de modelo OLS



Os dados parecem de fato indicar que a cobertura do PBF tem impacto importante na *performance* eleitoral do PT. O partido obtém porcentagem predita de votos da ordem de 44% quando o PBF assume valores baixos de sua distribuição, passa para 51% quando todas as variáveis, PBF inclusive, são ajustadas na média, e chega a cerca de 60% com valores elevados do programa social.

Outra forma de se analisar impacto substantivo das variáveis é por meio de simulações (King *at al.*, 2000)¹⁸. A tabela abaixo mostra os valores esperados da porcentagem de votos válidos do PT por meio desse exercício, alterando os valores do PBF na mesma escala utilizada anteriormente, mantendo as demais variáveis na média:

Tabela 3

Efeito do PBF no % PT - 2006			
	% PT 2006 predito	IC - 95%	
PBF - Primeiro decil	44	41,1	46,84
PBF - Média	51,64	50,18	53,22
PBF - Último decil	60,9	58,22	63,54

Fonte: Elaboração própria. Demais variáveis foram ajustadas em seus valores médios

Os resultados das simulações são muito semelhantes às estimativas do gráfico acima, indicando, à primeira vista, que os resultados substantivos extraídos a partir de regressões com base nos municípios parecem confirmar que o Programa Bolsa-Família tem um efeito importante e fundamental no resultado eleitoral, aparentemente justificando os diagnósticos da literatura.

No entanto, uma simples inspeção nos resultados preditos mostra dúvidas quanto à esse procedimento. A se manter todas as variáveis na média, o PT teria mais de 50% dos votos em 2006 (patamar indicado pela linha vermelha no gráfico) e, logo, ganharia as eleições no primeiro turno.

No entanto, se sabe que isso de fato não ocorreu. Essa contradição nos resultados não é explicada por erros na amostra utilizada. Antes, ela revela um problema analítico maior, qual seja, a inadequação de extrapolar diretamente resultados da eleição presidencial com base em percentagens municípios. Em outras palavras, essa estimativa está baseada em proporções municipais como unidade de análise, e não os votos válidos em cada localidade, o que seria adequado dada a regra eleitoral. Ou seja, comete o erro de falácia distrital.

Nossa proposta pra contornar esse problema é a seguinte: aplicamos o coeficiente do PBF (0,332) e recuperamos depois em votos absolutos cidade a cidade, de acordo com sua cobertura e o número de votos válidos em cada município. Usando esse procedimento, o impacto do PBF se revela da ordem de 6,9 milhões de votos. Em termos percentuais,

¹⁸ Utilizo aqui o Clarify (King *et al.* 2000) que realiza 1000 simulações estocásticas dos parâmetros do modelo a partir de sua distribuição amostral assintótica.

significa 7% do total de votos válidos em 2006, um efeito menor que o apontado no gráfico 2 e na tabela 3 acima.

Replicação I

De forma a reforçar esse ponto, procedemos aqui a uma replicação do modelo de Zucco (2013). O autor defende a tese que programas de transferência de renda no Brasil (o autor analisa não apenas o PBF, mas também os programas do PSDB Bolsa-Escola e Bolsa-Alimentação) garantem retorno eleitoral para o incumbente, ainda que esse efeito seja de curto-prazo e sem impacto em eleições legislativas e na formação de identificações partidárias¹⁹.

Uma das principais contribuições do estudo é de natureza metodológica. Como o PBF não foi distribuído de forma aleatória, - ao contrário, por exemplo, dos programas de transferência de renda mexicano e uruguaio (De La O, 2013; Ima *et al.*, 2016; Manacorda *et al.*, 2011), - Zucco (2013) argumenta que o impacto estimado do PBF no resultado eleitoral pode estar correlacionado com outras variáveis, observáveis e não-observáveis, o que poderia enviesar os resultados advindos de modelos tradicionais utilizados pela literatura.

Como forma de minimizar esses problemas, o autor emprega o *propensity score matching* (PSM, daqui em diante). Este procedimento metodológico procura aproximar a análise de dados observacionais de situações experimentais, criando um pareamento de covariáveis em indivíduos beneficiários e indivíduos não-beneficiários (Ho *et al.*, 2007; Morgan e Wiship, 2007).

No caso de análise de dados de *survey*, o pareamento implica em restringir a amostra de não-beneficiários do PBF apenas aos indivíduos que são semelhantes aos beneficiários. No caso dos dados agregados, como estamos tratando aqui, o PSM utilizado por Zucco (2013) divide os municípios em estratos de acordo com a cobertura estimada do Bolsa-Família a partir das seguintes variáveis: a meta de cobertura estipulada pelo governo, o IDH municipal (em forma linear e polinomial), os votos do PT em 1998, a distância à capital de estado, o (log) do PIB *per capita*, a proporção de minorias raciais e de evangélicos pentecostais, *dummies* sobre a presença de governadores e prefeitos de

¹⁹ Trataremos aqui apenas dos modelos sobre a cobertura do programa social em 2006.

PT e PSDB, e efeitos fixos por unidade da federação. Em outros termos, o autor utiliza, para a *estimação da cobertura do PBF*, que é o primeiro passo da análise de *matching*, grande parte das variáveis que são de modo geral tratadas pela literatura como controle nos modelos explicativos para os votos do PT²⁰.

Zucco (2013) divide os municípios em agrupamentos de acordo com sua cobertura estimada do PBF, com pontos de corte a cada 10% de famílias beneficiárias, conformando então 9 estratos diferentes. Para estimar o efeito causal do programa no voto, o autor emprega regressões OLS da proporção de votos do PT por cidade sobre a proporção de beneficiários do PBF efetivamente observada, controlando pela cobertura estimada e pela taxa de crescimento econômico.

A cada estrato o autor roda uma regressão. A suposição é que, dentro de cada grupo, a proporção de beneficiários do PBF efetivamente observada varia aleatoriamente em relação à proporção estimada. Dessa forma, seria possível minimizar problemas de inferência causal. O efeito médio do Bolsa-Família (ATE – *average treatment effect*) seria obtido pela média dos efeitos em cada estrato ponderada pelo número de municípios em cada agrupamento²¹.

Para 2006, eleição central para o debate, o autor encontra efeito de 0,145, com erro padrão de 0,041. A partir destas estimativas, Zucco (2013: 816) conclui que “results are statistically significant and substantively meaningfully when plausible changes in coverage are considered”. Ou seja, como grande parte da literatura, o autor defende que o impacto dos programas de transferência de renda é substantivo.

Analisemos seu impacto efetivo nos dois cenários que estamos propondo. No primeiro, cometendo o erro da falácia distrital, considerando os municípios não ponderados como unidade de análise. Ao se considerar o valor do 10º percentil de distribuição do PBF, seu impacto é de 1,5% dos votos preditos do PT em 2006. Com seu valor médio, o efeito é de 4,9%, e no 90º percentil, é de 8,9%. Ainda que o patamar seja menor que o encontrado no exercício com OLS, as mudanças são proporcionalmente elevadas.

²⁰ Diferentemente de outros trabalhos na literatura, no entanto, o autor inclui a meta de cobertura estipulada pelo governo. Esta variável tem correlação elevada com a cobertura observada, o que pode explicar a pequena magnitude do efeito do PBF encontrado por Zucco (2013) em comparação com resultados de outros estudos.

²¹ Esse é o número de cidades por estrato: 402; 1219; 1143; 664; 672; 830; 458; 64; 20.

Entretanto, esse resultado se encontra superestimado, porque baseado na suposição de que todos os municípios contam igualmente no resultado eleitoral. Quando se calcula o efeito em votos absolutos, conforme nossa proposta, o resultado é de pouco mais de 3,1% do total de votos válidos²².

Em suma, ao se ajustar o modelo seguindo as regras de funcionamento da competição eleitoral presidencial brasileira, resolvendo, logo, a falácia distrital, os efeitos do programa social diminuem, tanto no nosso modelo original, quanto no de Zucco (2013)²³.

Frisamos que não estamos argumentando que o efeito marginal do PBF varia com o tamanho das cidades. Tampouco o argumento não consiste em dizer que PBF não teve efeito eleitoral. Os resultados mostram que seu impacto é estatisticamente significativo. Antes, nosso objetivo é ressaltar as distorções decorrentes da falta de atenção à falácia distrital, com impactos nas inferências substantivas dos efeitos encontrados.

Replicação II

Montero (2010) tem como objeto de estudos a eleição inédita de partidos de esquerda em alguns estados nordestinos em 2006, particularmente, Bahia, Maranhão e Ceará²⁴. Essa questão é importante pois, em alguns deles, notadamente Bahia e Maranhão, são tidos como dominados por máquinas políticas e coronéis, perpassados por clientelismo. O pleito de 2006, assim, pode ter assumido caráter importante não apenas para as configurações da competição eleitoral no nível nacional, mas também no nível estadual.

²² Cabe ressaltar que utilizamos, neste exercício, o coeficiente de efeito de cada cidade de acordo com seu estrato e não o ATE.

²³ Considero a discussão sobre inferência causal que Zucco (2013) elabora importante, contribuindo para o avanço dessa área de pesquisa. No entanto, como o próprio autor nota, a utilização do *propensity score matching* não resolve as dificuldades inerentes da análise observacional. Afinal, a constituição dos pareamentos e dos estratos continua sendo formatada por variáveis observáveis, como notam Morgan e Wiship (2007: 122). Angrist e Pischke (2009) argumentam que a regressão OLS se constitui em uma forma particular de *matching*. Além disso, afirmam que “the differences between regression and matching estimates are unlikely to be of major empirical importance” (Angrist e Pischke, 2009: 69-70). De fato, não existem diferenças significativas nos resultados apresentados por Zucco (2010) entre a estimação por OLS e a estimação via PSM.

²⁴ O PT ganhou a eleição na Bahia, o PDT no Maranhão e o PSB no Ceará.

O autor discute algumas possíveis explicações para esse fenômeno. A distribuição do PBF, o crescimento econômico e o *coattail effect* de Lula aparecem como candidatos óbvios. O ponto do autor, no entanto, consiste em ressaltar a importância da construção partidária local. Montero (2010) chama atenção para o fato de que os partidos de esquerda se estabelecem inicialmente em áreas urbanas e grandes centros, para depois tentar se espalhar, com maior ou menor dificuldade, para o interior. A *proxy* que o autor utiliza para mensurar a presença e força local de partidos de esquerda é o grau de urbanização dos municípios. Sua tese é de que essa variável é importante para explicar os resultados da eleição para governador.

A análise empírica realizada por Montero (2010) tem como unidade de análise os municípios dos estados da Bahia, do Maranhão e do Ceará. Seus resultados demonstram que urbanização de fato tem impacto nos resultados eleitorais, ao passo que PBF e crescimento econômico não apresentam significância estatística. A votação de Lula no nível municipal, por sua vez, tem efeito positivo.

Nas tabelas abaixo apresentamos a replicação da tabela 2 de Montero (2010: 125), com as regressões OLS básicas que embasam seu argumento²⁵. Tratam-se de 6 modelos, 3 cuja variável dependente são os votos de partidos de esquerda para governador, e outros 3 com os votos para partidos de direita. As alterações entre os modelos devem-se a exclusões e inclusões de variáveis de controle.

²⁵ Como Montero (2010: 124) adverte, as dummies estaduais são incluídas no modelo mas não apresentadas no artigo.

Tabela 4 – Replicação tabela 2 Montero (2010: 125)

	Left Vote			Right Vote		
	(1)	(2)	(3)	(4)	(5)	(6)
Urbanization	5.695* (2.457)	7.782** (2.632)	5.791* (2.252)	-6.768** (2.488)	-9.325*** (2.635)	-7.340** (2.240)
BF_Scope	-1.076 (3.371)	3.174 (3.488)	-1.264 (3.169)	1.554 (3.645)	-3.111 (3.869)	2.343 (3.475)
Economic growth	2.433 (2.097)	3.373 (1.982)	2.469 (2.091)	-3.246 (2.008)	-4.453* (2.137)	-3.239 (2.007)
Left support in 2002	0.398*** (0.0416)	0.440*** (0.0450)	0.402*** (0.0386)			
Right support in 2002				0.329*** (0.0358)	0.353*** (0.0382)	0.338*** (0.0327)
Lula vote 2006	0.409*** (0.0438)		0.409*** (0.0435)	-0.463*** (0.0422)		-0.461*** (0.0419)
HDI (2000)	2.877 (12.31)	-7.918 (12.83)		-8.808 (12.62)	1.472 (13.30)	
BA	-14.99*** (1.161)	-17.57*** (1.235)	-15.08*** (1.127)	7.970*** (0.920)	10.22*** (0.974)	8.030*** (0.920)
MA	-42.89*** (1.657)	-42.10*** (1.790)	-43.12*** (1.529)	20.30*** (1.370)	18.38*** (1.378)	20.68*** (1.290)
Constant	21.03* (8.708)	53.23*** (8.427)	22.84*** (3.734)	59.56*** (10.13)	22.21* (10.11)	53.18*** (4.660)
Observations	816	816	818	816	816	818
Adjusted R ²	0.625	0.578	0.625	0.429	0.342	0.430

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Como interpretar efeitos substantivos a partir dessa regressão? Vejamos o raciocínio de Montero (2010: 126) nas suas próprias palavras:

“The results demonstrate that urban localization and Lula’s coattails are the most consistent predictors of the gubernatorial results in the three states in 2006. The latter proves a weightier factor since a move from the 10th percentile to the 90th on Lula’s municipal share of the vote while holding all other predictors at their mean produces a -12 percentage point change in the rightist incumbent’s share and a

+10.6 percentage point change in the leftist opposition's share. Urbanization has a less powerful, but still appreciable, effect. Moving from the lowest decile to the highest in urban population rates produces a shift of 3 percentage points in favor of the left and a decline of 4 percentage points in the right's share of the gubernatorial vote.”

O procedimento adotado pelo autor para inferior efeitos substantivos de sua análise cai na falácia distrital. Os efeitos preditos a partir de valores das variáveis de interesse estão equivocados pois baseados em porcentagens calculadas em unidades de tamanho e atributos heterogêneos.

De modo a contornar esse problema, adotamos o procedimento exposto acima. Recuperamos em valores absolutos, em cada município, o impacto das variáveis de urbanização e *coattail effect* de Lula. Utilizamos os resultados dos modelos 1 e 4, os mesmos usamos pelo autor para embasar seu raciocínio exposto acima.

Os resultados demonstram a importância da nossa proposta. Os efeitos da variável urbanização pouco se alteram: no modelo 1, cuja variável dependente são os votos de partidos de esquerda, urbanização tem efeito positivo de 3,89% dos votos. No modelo 4, cuja variável dependente são os votos de partidos de direita, urbanização tem efeito negativo de 4,62%.

No entanto, diferenças gritantes aparecem no impacto da variável de *coattail effect* de Lula. Se na análise de Montero (2010) essa variável apresentou impacto substantivo maior, a correção da falácia distrital mostra, na verdade que os ganhos para os partidos de esquerda e as perdas para os de direita são muito pequenos: 0,28% e -0,32%, respectivamente.

Ou seja, a depender da heterogeneidade do tamanho e da distribuição das variáveis, a falácia distrital tem consequências importantes para as inferências substantivas dos determinantes dos resultados eleitorais.

Conclusão

O objetivo desse artigo foi demonstrar que a devida incorporação da dimensão do distrito eleitoral nas análises sobre determinantes das eleições em sistemas majoritários, como as eleições presidenciais e para governador no Brasil, tem consequências

importantes nos resultados encontrados e, particularmente, para algumas teses correntes na literatura.

Os estudos eleitorais utilizam como unidade de análise tanto dados individuais quanto agregados. Parte do debate metodológico consiste em ressaltar as possibilidades e os limites explicativos da cada nível de análise. O termo falácia ecológica foi cunhado para designar as inferências enviesadas sobre comportamento individual a partir de dados agregados. Já o termo falácia individualista consiste na crítica de que análises focadas unicamente nos indivíduos são também equivocadas, pois perdem de vista efeitos contextuais.

Neste artigo chamamos atenção para uma terceira falácia ainda não escrutinada pela literatura. Chamamos de falácia distrital o procedimento de inferir determinantes do resultado eleitoral a partir de extrapolações de efeitos substantivos estimados em percentagens das distribuições das variáveis de interesse em unidades subdistritais. O problema é premente quando as unidades tem tamanho e distribuição heterogênea das variáveis.

Mostramos a importância desse ponto por meio de simulações e de análise própria e replicações para o caso brasileiro, no debate sobre efeito do PBF na eleição presidencial e sobre os determinantes das eleições de partidos de esquerda nos governos estaduais do Nordeste. Nossa proposta para superar o problema da falácia distrital é simples e direta: consiste em recuperar em votos absolutos, município a município, os efeitos percentuais preditos da variável de interesse. O efeito total estimado da variável de interesse é obtido pela soma dos valores absolutos de todos os municípios. Dessa forma, mantemos o município como unidade de análise, concordando com a crítica da falácia individualista e permitindo captar efeitos locais, mas evitamos a falácia distrital.

Os resultados da comparação de efeitos substantivos em cenários com e sem falácia distrital mitigam alguns dos efeitos encontrados na literatura, mostrando a importância da nossa proposta, que vai além do caso brasileiro e tem implicações gerais.

De modo geral, nossa crítica e proposta chama atenção para importância de algo básico, mas negligenciado: de se considerar as regras eleitorais nos estudos sobre determinantes de seus resultados.

Bibliografia:

Achen, Christopher; Shively, Phillips. (1995). **Cross-level inference**. University of Chicago Press.

Agnew, John. (1996). "Mapping politics: how context counts in electoral geography". In.: *Political Geography*, 15.

Alker, Hayward. (1969) "A typology of ecological fallacies". In.: Dogan, Mattei e Rokkan, Stein (ed.). **Quantitative ecological analysis in the social science**. Massachusetts Institute of Technology.

Ames, Barry (1994) "The Reverse Coattails Effect: Local Party Organization in the 1989 Brazilian Presidential Election". In.: *American Political Science Review*, vol. 88, no 1.

Angrist, Joshua e Pischke, Jorn-Steffen (2008). **Mostly Harmless Econometrics: An Empiricist's Companion**. Princeton University Press

Avelino, George; Biderman, Ciro; Barone, Leonardo. (2012) "Articulações intrapartidárias e desempenho eleitoral no Brasil". In.: *Dados*, vol. 55, n.4.

Bartels, Larry. (2010). "The study of electoral behavior". In.: Jan E. Leighley (ed.). **The Oxford Handbook of American Politics and Political Behavior**.

Bohn, Simone. (2011) "Social policy and vote in Brazil". In.: *Latin American Research Review*, vol. 46, n.1

Bohn, Simone. (2013) "The electoral behavior of the poor in Brazil" In.: *Latin America Research Review*, vol. 48, n.2.

Borges, André (2010) "Já não se fazem mais máquinas políticas como antigamente: Competição vertical e mudança eleitoral nos Estados brasileiros", In: *Revista de Sociologia e Política*, v. 18, n. 35.

Borges, André (2011). "The political consequences of Center-led redistribution in Brazilian federalism", In: *Latin American Research Review*, Vol. 46, No. 3

Campbell, Angus; Converse, Philip; Miller, Warren; Stokes, Donald. (1980) *The American Voter*. University of Chicago Press, Chicago.

Canêdo-Pinheiro, Maurício. (2015). "Bolsa família ou desempenho da economia? Determinantes da reeleição de lula em 2006". In.: *Economia Aplicada*, v. 19, n.1.

Carraro, André *et al.* (2009) "It is the economy, companheiro!": an empirical analysis of Lula's re-election based on municipal data". In.: *Economics Bulletin*, 01/2009; 29(2):976-991.

Cho, Wendy e Manski, Charles (2010) "Cross-Level/Ecological Inference" In.: **Oxford Handbook of Political Methodology**. Oxford University Press.

- Cohn, Amélia (2012). **Cartas ao Presidente Lula – Bolsa Família e Direitos Sociais**. 1.ed. Ed. Pensamento Brasileiro.
- Cox, Kevin (2005). **Political Geography: critical concepts in the Social Sciences**. Volume III. Ed. Routledge.
- De La O, Ana. (2013) “Do conditional cash transfers affect electoral behavior? Evidence from a randomized experiment in Mexico”. In.: *American Journal of Political Science*, 57 (1): 1-14.
- Dogan, Mattei e Rokkan, Stein. (1969). **Quantitative ecological analysis in the social science**. Massachusetts Institute of Technology.
- Ethington, Phillip e McDaniel, Jason. (2007). “Political places and institutional spaces”. In.: *Annual Review of Political Science*, 10.
- Firebaugh, G. (2001). “Ecological fallacy, statistics of. In: Smelser NJ, Baltes PB (ed.) **International Encyclopedia of the Social and Behavioral Sciences**. Oxford: Elsevier Science.
- Fried, Brian. (2009). “Distributive politics and conditional cash transfers: the case of Brazil’s Bolsa Família”. In.: *World Development*, vol. 40, n. 5.
- Hanushek, E; Jackson, J.; Kain, J. (1974). “Model specification, use of aggregate data, and the ecological correlation fallacy”. In.: *Political Methodology*, n.1, vol.1
- Ho, Daniel *et al.* (2007). “Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference”. In.: *Political Analysis*, 15.
- Huckfeldt, Robert e Sprague, John. (1987). “Networks in context: The social flow of political information”. In.: *The American Political Science Review*, vol. 81, issue 4.
- Hunter, Wendy e Power, Timothy. (2005). “Lula’s Brazil at midterm”. In.: *Journal of Democracy*, Volume 16, Number 3.
- Imai, Kosuke; King, Gary e Rivera, Carlos Velasco. (2016). “Do nonpartisan programmatic policies have partisan electoral effects? Evidence from two large scale randomized experiments”. Working paper.
- Katz, Jonathan e King, Gary. (1999). “A statistical model for multiparty electoral data”. In.: *American Political Science Review*, vol. 93, n.1.
- Key Jr., V. O. (1955). “A Theory of Critical Elections”. In.: *Journal of Politics*, Vol. 17, No. 1.
- Key Jr., V.O. (1959). “Secular Realignment and the party system”. In.: *Journal of Politics*, Vol. 21, No. 2.
- King, Gary. (1997). **A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data**. Princeton: Princeton University Press.
- King, Gary; Tomz, Michael; Wittenberg, Jason (2000). “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”. In.: *American Journal of Political Science*, Vol. 44, No. 2.

- Kramer, Gerald. (1983). "The ecological fallacy revisited: aggregate versus individual level findings on economics and elections". In.: *American Political Science Review*, vol. 77, n. 1.
- Landim Jr., Henrique. (2007). "Os Efeitos do Programa Bolsa Família sobre a Economia dos Municípios Brasileiros". Disponível em: <http://www.ipc-undp.org/publications/mds/33P.pdf>.
- Lício, Elaine; Rennó, Lúcio; Castro, Henrique Carlos. (2009). "Bolsa Família e Voto na Eleição Presidencial de 2006: em busca do elo perdido". In.: *Opinião Pública*, vol. 15, nº 1.
- Limongi, Fernando e Guarnieri, Fernando. (2015). "Competição partidária e voto nas eleições presidenciais no Brasil"; In.: *Opinião pública*, vol. 21, nº 1.
- Magalhães, André; Silva, Marcelo Eduardo Alves; Dias, Fernando de Mendonça. (2015). "Eleição de Dilma ou segunda reeleição de Lula? Uma análise espacial do pleito de 2010". In.: *Opinião Pública*, vol. 21, n.3.
- Manacorda, Marco, E. Miguel, and A. Vigorito. (2011). "Government Transfers and Political Support". In.: *American Economic Journal: Applied Economics*, 3(3): 1–28.
- Marques, Rosa. (2005). "A importância do Bolsa-Família nos municípios brasileiros", In.: *Cadernos de Estudo Desenvolvimento Social em Debate*, n.1.
- Martins, Diogo; Mansano, Fernanda; Parré, José; Plassa, Wander. (2016). "Fatores que contribuíram para a reeleição da presidente Dilma Rousseff". In.: *Política e Sociedade*, vo. 15, n. 32.
- Marzagão, Thiago. (2013). "A dimensão geográfica das eleições brasileiras". In.: *Opinião Pública*, vol. 19, n2.
- Montero, Alfred (2010) "No Country for Leftists? Clientelist Continuity and the 2006 Vote in the Brazilian Northeast", In: *Journal of Politics in Latin America* 2:2
- Montero, Alfred (2012) "A Reversal of Political Fortune: The Transitional Dynamics of Conservative Rule in the Brazilian Northeast", In: *Latin American Politics and Society*
- Morgan, Stephen e Wiship, Christopher. (2007). **Counterfactuals and Causal Inference: Methods and Principles for Social Research. Analytical Methods for Social Research**. Cambridge: Cambridge University Press.
- Mostafa, Joana *et al.*, (2010). "Efeitos econômicos do gasto social no Brasil". In: Castro, Jorge Abrahao *et al.* (Ed.). **Perspectivas da política social no Brasil**. Brasília: Ipea, 2010.
- Neri, Marcelo *et al.*, (2013). "Efeitos Macroeconômicos do Programa Bolsa Família". In Campello e Cortez (org.) **Programa Bolsa Família: uma década de inclusão e cidadania**. Brasília: IPEA.
- Nicolau, Jairo e Peixoto, Vitor. (2007). "Uma disputa em três tempos: uma análise das bases municipais das eleições presidenciais de 2006". XXXI Encontro da Anpocs.

- Paes de Barros, Ricardo *et al.*, (2008). “A importância das cotas para a focalização do programa bolsa-família”, Texto para discussão, n. 1349.
- Paes-Sousa, Rômulo e Vaitsman, Jeni. (2007). “Cadernos de Estudos Desenvolvimento em Debate – Síntese das Pesquisas de Avaliação de Programas Sociais do MDS 2004-2006”, Ministério do Desenvolvimento Social e Combate à Fome.
- Pereira, Ana *et al.* (2015). “A eleição de Dilma em 2010 e seus determinantes: evidências empíricas do Programa Bolsa Família”. In.: *Análise Econômica*, n. 64.
- Przeworski, Adam e Sprague, John. (1986). **Paper stones**. The University of Chicago Press.
- Rego, Walquíria Leão e Pinzani, Alessandro. (2013). **Vozes do Bolsa-Família – autonomia dinheiro e cidadania**. São Paulo: Editora Unesp.
- Ribeiro, Ricardo Mendes. (2014). “Decadência longe do poder: refundação e crise do PFL”. In.: *Revista de Sociologia e Política*. Vol. 22, n. 49.
- Robinson, W. S. (1950). “Ecological correlations and the behavioral of individuals”. In: *American Sociological Review*, vol. 15, n.3.
- Schwartz, Sharon. (1994). “The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences”. In.: *American Journal of Public Health*. 84(5).
- Selvin, Kevin. (1958). “Durkheim’s Suicide and Problems of Empirical Research”, In.: *American Journal of Sociology*, vol. 63, n. 6.
- Singer, André. (2012). **Os sentidos do lulismo**. Ed. Companhia das Letras.
- Snijders, Tom e Bosker, Roel. (2012). **Multilevel analysis**. SAGE Publications.
- Soares, Gláucio e Terron, Sonia. (2008). "Dois Lulas: a geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise geoespacial)". In.: *Opinião Pública*, vol. 14, nº 2
- Souza, Tatiene e Cribari Neto, Francisco. (2013). “Uma estimativa do impacto eleitoral do Programa Bolsa-Família”. In.: *Rev. Bras. Biom.*, vol. 31, n.1.
- Sprinz, Detlef. (2000). “Cross-level inference in political Science”. In.: *Climatic Change*, 44.
- Subramanian *et al.* (2009). “Revisiting Robinson: the perils of individualistic and ecologic fallacy”. In.: *International Journal of Epidemiology*, n.38.
- Ward, Michael D. and Gleditsch, Kristian Skrede. (2008). **Spatial Regression Models**. Sage Publications.
- Wright, Gerald. (1989). “Level of analysis effects on explanation of voting: the case the 1982 US Senate Elections”. In.: *British Journal of Political Science*, vol. 19, n.3.
- Zucco, Cesar. (2008). "The President's `New' Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Election". In.: *Journal of Latin American Studies*, vol. 40(1).

Zucco, Cesar. (2013). "When pay outs pay off: conditional cash-transfer and voting behavior in Brazil 2002-2010". In.: *American Journal of Political Science*, 47(3).

Zucco, Cesar. (2015). "The Impacts of Conditional Cash Transfers in Four Presidential Elections (2002-2014)". In.: *Brazilian Political Science Review*, 9(1).

Zuckerman, Alan. (2005). **The social logic of politics: personal networks as contexts for political behavior**. Temple University Press