

Thiago Moreira da Silva
Sergio Simoni Jr.
Argelina Figueiredo

Introdução

As eleições presidenciais recentes na América Latina são caracterizadas, em maior ou menor grau, pela polarização político-ideológica e pelo conflito distributivo. Em diversos países da região, o período recente foi caracterizado pela redução da desigualdade de renda do trabalho (Cepal, 2018) e pela implementação de políticas de transferência de renda condicionada, atingindo, em alguns casos, amplas parcelas da população mais pobre (Fizbein et al., 2009).

O Brasil é um caso importante para a literatura. Entre 2002 a 2014, liderando a coalizão de centro-esquerda, o PT venceu quatro eleições presidenciais em disputas de segundo turno contra o PSDB, que capitaneou a coalizão de centro-direita. As análises correntes ressaltam que a partir de 2006 ocorreu uma clara distribuição social e regional do voto: pessoas de renda mais baixa nos estados mais pobres dariam maior apoio aos candidatos do PT (Hunter e Power, 2007; Nicolau e Peixoto, 2007; Zucco, 2008, 2010; Soares e Terron, 2008; Canêdo-Pinheiro, 2015; Singer, 2012; Zucco e Power, 2013; Marzagão, 2013; Corrêia, 2015; Limongi e Guarnieri, 2014).

Essa correlação social e regional também se fez presente na acirrada disputa de 2014, quando a candidata do PT se reelegeu com margem de menos de 4% dos votos. A campanha se caracterizou pelo forte embate ideológico, o que levou alguns analistas a ressaltar que estaria ocorrendo um fenômeno de polarização política (Couto, 2014; Reis, 2014). Nas palavras de Nicolau (2015: 40) “Na última disputa presidencial, o Brasil ficou mais dividido, com cada partido ampliando a votação no seu reduto e diminuindo no território adversário”.

Neste artigo procuramos avançar no entendimento da relação entre renda e as disputas presidenciais no Brasil. Iniciamos com uma constatação básica: as explicações vigentes fazem referência a duas dimensões de análise, a individual e a agregada. No entanto, ao analisar exclusivamente dados geográficos (em nível estadual ou municipal) ou individuais (por meio de *survey*), os pesquisadores pecam em tomar como equivalentes os votos de

pessoas pobres (ricas) e os votos de regiões pobres (ricas). Para Singer (2012), por exemplo, “o realinhamento eleitoral de 2006 significa a mudança de um padrão histórico de comportamento eleitoral das camadas populares no Brasil, em particular no Nordeste”.

Como mostra o trabalho de Gelman et al. (2007; 2009) para o caso norte-americano, essa associação não deve ser imediata. Um partido pode ter sua base eleitoral entre os ricos, mas ganhar em estados pobres, como o Partido Republicano, ou ter base eleitoral entre os pobres, mas ganhar em estados ricos, como o Partido Democrata. Para eles, é equivocado identificar o eleitor republicano típico como o rico dos estados pobres e o democrata como os pobres dos estados ricos. Mostram que, no país como um todo, os republicanos tem votos entre os mais ricos e os democratas entre os mais pobres. No entanto, nos estados mais ricos a relação entre renda e voto é menor, ou seja, os ricos tem maior probabilidade de votar nos democratas do que nos estados mais pobres. O contexto, fatores sociais e culturais, seria a explicação para essa diferença.

Seguindo nessa linha, analisamos neste artigo a eleição de 2014, testando uma metodologia ainda não aplicada para o caso nacional: a *regressão multinível com pós-estratificação* (MRP, de agora em diante) (Gelman et al., 2007, 2009). Utilizamos estatística bayesiana com equações hierárquicas para estimar as preferências de subgrupos em nível estadual, tendo por base os dados do ESEB (Estudo Eleitoral Brasileiro), do Censo Demográfico Brasileiro, do TSE (Tribunal Superior Eleitoral).

Este trabalho tem duplo objetivo: do ponto de vista metodológico, argumentamos que o MRP permite prever com razoável grau de certeza os resultados de 2014 no nível estadual com um modelo simples, com apenas algumas variáveis demográficas. Do ponto de vista substantivo, avançamos nas explicações para o caso brasileiro sublinhando um aspecto não notado pela literatura: o nível de riqueza estadual importa para o comportamento eleitoral tanto de eleitores pobres, quanto de eleitores não-pobres. Especificamente, quanto mais pobre o estado, maior o apoio do PT em ambos os grupos de eleitores. Logo, ao contrário da ideia de que o eleitor típico do PT é o eleitor pobre nordestino, mostramos como o contexto estadual impacta o comportamento das diferentes categorias de renda.

Indivíduo e região nas eleições presidenciais no Brasil

Os estudos recentes pautam-se no papel da renda e da região para explicar o voto nas eleições presidenciais brasileiras. Sejam as teses que defendem a ocorrência de um realinhamento eleitoral, sejam as teses que minimizam esse fenômeno, é amplamente aceito que a partir de 2006 existe uma clara clivagem do voto em termos de renda individual e em termos geográficos. Os candidatos petistas tendem a receber maior apoio entre eleitores pobres e nas regiões mais carentes.

O pleito de 2014 parece confirmar esse diagnóstico. Em pesquisa Datafolha conduzida a seis dias do segundo turno, Dilma Rousseff, candidata do PT, tinha 52% das intenções de votos válidos. Entre os eleitores com até dois salários mínimos, seu patamar chegava a 62%. Já entre moradores da região Nordeste, a mais pobre do país, o apoio à candidata mandatária alcança 70% das intenções de voto válido. Os resultados efetivos caminham na mesma direção: no Piauí, estado no qual a porcentagem de pobres de acordo com o Censo de 2010 era de 34%, o PT conseguiu 78% dos votos válidos no dia 26 de outubro de 2014. Em Santa Catarina, cujo nível de pobreza era dez vezes menor, o candidato do PSDB logrou 64% das preferências.

Para alguns autores esse fenômeno seria explicado pela distribuição das políticas sociais dos governos petistas, notadamente o Programa Bolsa Família (PBF), focado nas famílias mais pobres e altamente concentrado no Nordeste (Nicolau e Peixoto; 2007; Soares e Terron, 2008; Lício et al., 2009; Zucco e Power, 2013; Magalhães et. al., 2015; Zucco, 2013). Para outros, se trataria de uma manifestação particular do fenômeno geral do governismo, no qual pobres e localidades pobres tendem a apoiar o mandatário, independente de sua performance ou plataforma (Zucco, 2008, 2010; Canêdo-Pinheiro, 2015). Singer (2012) acrescenta que a correlação renda-voto seria, em parte, expressão da congruência entre o lulismo e elementos ideológicos que permeiam as camadas populares. Por fim, Carraro et al (2009) ressaltam que o desempenho econômico heterogêneo das diferentes regiões do país explicaria as disparidades eleitorais no nível geográfico.

Observa-se que parte das explicações diz respeito às características individuais do eleitor, como a renda, ser beneficiário do PBF, sua identidade ideológica, enquanto outra faz referência às características agregadas, como PIB, renda média, região, estado, localidade de residência do eleitor. Metodologias e fontes de dados diferentes foram utilizados para

embasar essas teses. Alguns trabalhos utilizam dados agregados no nível municipal ou estadual. Alguns lançam mão de *surveys* eleitorais.

O debate sobre explicações de nível micro e macro perpassa a história das Ciências Sociais (Dogan e Rokkan, 1969; Przeworski, 1974; Coleman, 1994; Achen e Shively; 1995; King, 1997; Sprinz, 2000; Cho e Manski, 2010)¹. Referência central é a demonstração de Robinson (1950) de que correlações entre fenômenos medidos em nível agregado não podem ser imputadas às unidades que compõe essa agregação, proposição que ficou conhecida como falácia ecológica (Selvin, 1958).

A área de comportamento eleitoral, por outro lado, foi fortemente influenciada pela escola de Michigan, que estabeleceu o *survey* como instrumento primordial de análise eleitoral (Campbel et. al., 1980; Bartels, 2010). A explicação do voto deveria ser buscada, segundo essa visão, unicamente em variáveis individuais, de cunho cognitivo. Dados agregados, logo, seriam necessariamente enviesados.

Parte da literatura no debate sobre as eleições presidenciais no Brasil se alinha com essa perspectiva. Por exemplo, Bohn (2011) critica o uso de dados agregados para mensurar o efeito eleitoral do PBF. Correlações nesse nível padeceriam, nas palavras de Lício et al. (2009), de um “elo perdido” explicativo, que necessariamente deveria ser buscado em informações de *surveys*. 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 nosso), 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”. 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 (...). Desse modo, *não há como escapar do uso de dados agregados*”.

Na contramão dessa visão, uma série de argumentos metodológicos foram construídos para 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 (Subramanian *et al.* (2009); Schwartz (1994) Gelman et.

¹ Os próximos parágrafos se baseiam, em parte, em Simoni Jr (2017).

al., 2007; 2009; Kramer, 1983; Wright, 1989; Huckfelt e Sprague, 1986). Parte da literatura nacional também argumenta a favor de hipóteses contextualistas, lançando mão de métodos de econometria espacial (Soares e Terron, 2008; Carraro *et al.*, 2009; Marzagão, 2013). No entanto, ao focarem apenas em dados agregados, não conseguem captar como eles se relacionam com informações individuais.

Como notam Przeworski e Sprague (1986), embora o voto seja um ato individual, é ao mesmo tempo um fenômeno social e político. Wright (1989) argumenta que os resultados eleitorais não são uma mera agregação de comportamentos individuais. O impacto de certas variáveis micro pode ser diferente dado fatores contextuais. Nas palavras de Katz e King (1999: 15, tradução nossa) “pesquisas de survey tem vantagens enormes por estudar preferências no nível individual, mas na medida em que estão baseadas em seleção aleatória de indivíduos isolados de localidades geográficas desconhecidas, elas necessariamente não captam importantes elementos da política eleitoral. Como tal, essas pesquisas são frequentemente melhor complementadas com estudos de retorno eleitoral agregado”.

Nosso argumento é que explicações adequadas para o comportamento eleitoral devem cobrir tanto variáveis de nível individual, quanto de nível agregado. Perspectiva semelhante já foi utilizada por alguns trabalhos no caso brasileiro. Limongi e Guarnieri (2015) analisam a variação de votos entre 2002 e 2006 por seção eleitoral, agrupadas por estados por meio de um modelo multinível. Ainda que as seções eleitorais sejam, *stricto sensu*, uma agregação de indivíduos, são o menor nível possível que se pode observar resultados eleitorais efetivos. Além disso, tendem a ser mais homogêneas que outras unidades de análise agregada. No entanto, a natureza do dado limita o número de variáveis disponíveis, impedindo que análises mais complexas sejam realizadas.

Zucco e Power (2013) utilizam análise de *survey* agregando os dados por tipo de localidade, estratégia compartilhada por Zucco (2010) que também lança mão de métodos de inferência ecológica de dados no plano municipal. No entanto, como o próprio autor reconhece, seus resultados não são conclusivos: os métodos de inferência ecológica são controversos (Cho, 1998), e as amostras dos *surveys* utilizados não são representativas de subgrupos no nível subnacional, não permitindo inferir com segurança o comportamento de eleitores pobres e ricos em diferentes localidades.

Em Figueiredo et al. (2018), avançamos nesse sentido utilizando *surveys* com grande número de observações e ajustando interações para verificar se o comportamento do

eleitorado pobre varia conforme a localidade, entendida como macro-região e porte dos municípios. No entanto, a metodologia utilizada não permite verificar variações de comportamento eleitoral de subgrupos em um nível geográfico menor, como os estados.

Os estados são tradicionalmente vistos como importantes para a política brasileira. No debate sobre as eleições recentes, Hunter e Power (2007) e Limongi e Guarnieri (2015) focam nessa unidade. Limongi e Guarnieri (2015), especificamente, levantam argumentos importantes para essa escolha: a lógica das disputas estaduais é importante para a coordenação de campanhas dos partidos competitivos nacionalmente.

Neste trabalho, utilizamos a metodologia MRP, mais adequada para lidar com a inclusão de variáveis contextuais no nível dos estados da federação, por combinar análise multinível com pós-estratificação. A análise multinível avança em relação às interações ao modelar diretamente as relações entre variáveis de diferentes níveis. A pós-estratificação permite a estimação do comportamento de subgrupos ao nível estadual com maior segurança, pois incorpora informações dos Censos Demográficos. Falaremos mais sobre o método na próxima seção.

Utilizamos as variáveis individuais tal como coletadas pelo ESEB 2014, realizado pelo Centro de Estudos de Opinião Pública (Cesop) da Universidade Estadual de Campinas (Unicamp), com amostra nacional de 2.506 entrevistas e com trabalho de campo entre os dias 1 e 19 de novembro de 2014. Os dados das unidades da federação foram coletados no Censo demográfico de 2010 e no TSE.

Dados e metodologia

Embora as pesquisas de opinião tenham se tornado cada vez mais comuns no Brasil em tempos recentes, com institutos nacionais e estrangeiros fazendo a coleta dos dados, esses bancos de informações não fornecem diretrizes adequadas para a estimação das preferências e opiniões de subgrupos da população. Isso porque as amostras são representativas no nível estritamente nacional.

Tendo esse aspecto em vista, normalmente os pesquisadores se debruçam sobre preditores sociodemográficos de *surveys*, especialmente as regiões, ou se utilizam de técnicas de desagregação – método que consiste na combinação e estimação de médias de preferências nos estados em diferentes bancos individuais (Erikson, Wright e McIver, 1993). O modelo de regressão múltipla com pós-estratificação (MRP), no entanto, figura como uma forma mais

consistente de análise do comportamento dos subgrupos que compõem a população dos países investigados. Com essa abordagem utilizamos *surveys* nacionais e estatística bayesiana com equações hierárquicas para estimar as preferências de subgrupos em nível estadual (Park, Gelman e Bafumi, 2004). Dessa forma, pretendemos identificar possíveis diferenças contextuais na atuação eleitoral de segmentos de renda específicos.

O MRP é composto de duas etapas. A primeira compreende no cálculo de um modelo multinível, com as inclinações dos eleitores ponderadas de acordo com as características demográficas e geográficas dos respondentes. Assim, consideramos explicitamente o agrupamento das preferências conforme padrões estaduais, de modo que todos os entrevistados da pesquisa produzam indicações sobre os padrões demográficos de uma dada localidade. Enquanto isso, preditores de nível geográfico capturam diferenças residuais entre as áreas destacadas. Posteriormente, pesamos os coeficientes obtidos no modelo multinível referentes aos subgrupos nas porcentagens desses tipos populacionais em cada estado. Nesse segundo passo, conhecido como pós-estratificação, usamos as estatísticas fornecidas pelo censo nacional para analisar a porcentagem correta das categorias examinadas.

O MRP apresenta vantagens potenciais em relação aos métodos até agora utilizados pelos estudos eleitorais no Brasil (Lax e Phillips, 2009). Primeiro porque permite o escrutínio das inclinações dos votantes de forma localizada, mesmo com quantidades reduzidas de informação. Ademais, a MRP corrige eventuais incongruências entre as amostras e a população real, minimizando os problemas decorrentes de não-respostas e de desenhos amostrais pouco sofisticados. Por fim, a metodologia facilita a elaboração de estimativas de segmentos não contemplados pelas bases de dados existentes, haja vista que poucas pesquisas contam com dados relativos a todas as unidades federativas.

Neste artigo faremos uso do *Estudo Eleitoral Brasileiro* (ESEB) de 2014, a fim de mapear os condicionantes do voto da eleição presidencial entre Dilma Rousseff (PT) e Aécio Neves (PSDB). A preferência pelo ESEB se deve à representatividade nacional da amostra e pelo fato de a pesquisa se tratar de uma coleta pós-eleitoral, o que minimiza o impacto da volatilidade eleitoral e afiança maior congruência entre as opiniões expostas pelos entrevistados. Além disso, o uso desse banco permite a comparação com outros estudos que, via de regra, tomaram o ESEB como base privilegiada, embora com técnicas distintas de análise. Para a pós-estratificação adotamos o Censo Demográfico de 2010, o mais próximo do período analisado, elaborado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Na tabela 1, apresentamos as variáveis utilizadas, junto com as devidas codificações.

Tabela 1 – Codificação das variáveis utilizadas

Variável	Descrição	Tipo	Número de Grupos
Voto no 2o turno % voto PT (2010)	Binária. Dilma = 1, Aécio = 0 Contínua. Normalizada, valores entre 0 e 1	Variável dependente	-
Renda domiciliar estadual	Contínua. Normalizada, valores entre 0 e 1	Preditor Linear	-
Renda	Binária. Até 2 SM = 1, + 2 SM = 0	Preditor Linear	-
Estado	Categórica. Todos os estados + DF	Interceptos e inclinações aleatórias	2
Raça	Binária. Brancos = 1, Outros = 0	Inclinações aleatórias	27
Idade	Categórica. 18-25, 26-46, 46+	Inclinações aleatórias	2
Sexo	Binária. Mulher = 1, Homem = 0	Inclinações aleatórias	3
		Inclinações aleatórias	2

A variável dependente do modelo possui uma configuração binária, levando em consideração o segundo turno no pleito de 2014. Excluímos as abstenções, os votos em branco e nulos e os que não responderam. Assinalamos os concorrentes petistas como categorias principais. Seguimos nossa proposta em Figueiredo et al (2018) e criamos dois grupos de renda: pobres e não pobres, para separar a população de menor renda do restante dos cidadãos. Escolhemos o critério de 2 salários mínimos domiciliar do ESEB como delimitador desses grupos, o que, assumindo um domicílio típico de 4 pessoas, corresponde a um quarto de salário per capita, tal como o critério do Censo Demográfico para definição de pobreza. Nossa justificativa em criar uma categoria dicotômica de renda se deve ao fato de que, dessa forma, concentramos a análise no comportamento dos pobres. Além disso, minimizamos o problema de se delimitar categorias de renda acima desse patamar (2 a 5 SM, 5 a 10 SM, etc.), normalmente estabelecidas, e logo, interpretadas, sem critérios substantivos. Ademais, essa estrutura favorece a estimação em um modelo MRP, posto que condensa informações em poucos níveis - diminuindo, portanto, os casos faltantes. Como controle, adicionamos as variáveis *sexo*, *idade* (18-25, 26-46, 46+) e *raça* (brancos e outros). Assim, formalmente temos:

$$Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{s[i]} + \beta x_{s[i]k}), \text{ cada } i = 1, \dots, n,$$

em que o subscrito $s[i]$ indexa o estado (de 1 a 27) no qual reside o respondente i ; $x_{s[i]k}$ são os preditores considerados na equação por distrito subnacional e n é o número de observações em cada base. O modelo com variância nos interceptos e nos parâmetros autoriza a avaliação do apoio à candidata petista. Com efeito, os coeficientes de renda, sexo, idade e raça representam a medida pela qual Dilma foi votada nas 27 unidades federativas.

Além disso, incorporamos medidas atinentes à votação no PT em 2010 e a renda domiciliar média por estado como preditores de nível estadual. O modelo pode ser explicitado como:

$$\alpha_s = U_s \gamma + \varepsilon_s, \text{ cada } s = 1, \dots, 27$$

em que U é a matriz dos preditores em nível estadual e os erros ε tem média 0 e variância σ^2_a , que pode ser estimada com a combinação do nível individual dos modelos.

Resultados

Na tabela 2 estimamos 3 modelos, do mais simples ao mais complexo, para enfatizar as diferenças do efeito da renda sobre o voto, quando levamos em consideração as unidades subnacionais. No primeiro, o modelo nulo, omitimos os preditores e autorizamos somente o intercepto a variar pelos estados, a fim de investigar a estrutura dos dados e prover informações acerca da variância entre indivíduos e *clusters*. A ideia é mostrar o quanto o intercepto e o termo residual se distanciam do intercepto médio. Fora isso, computamos a correlação intraclasse (ICC), com o objetivo de quantificar a proporção da variância total do sufrágio no PT concernente às diferenças estaduais. Valores próximos a zero sugerem a inadequação de um arcabouço hierárquico; enquanto coeficientes maiores do que zero, mesmo inferiores 0,10, indicam a conveniência de uma estrutura multinível. Conforme os coeficientes expostos, o ICC do modelo nulo é de 0,11, revelando que as diferenças estaduais são responsáveis por 11% da variação do voto.

Tabela 2 – Coeficientes estimados dos modelos multinível

	Modelo 1	Modelo 2	Modelo 3
(Intercept0)	0.51*** (0.14)	0.32* (0.13)	0.27* (0.12)
Renda		0.53*** (0.09)	0.60*** (0.11)
AIC	3459.05	3428.00	3429.55
BIC	3470.79	3445.60	3458.90
Log Likelihood	-1727.53	-1711.00	-1709.77
Num. obs.	2617	2617	2617
Num. grupos: estado	27	27	27
Var: estado (Intercept)	0.40	0.31	0.23
Var: estado renda			0.04
Cov: estado (Intercepto) renda			0.09

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

Fonte: elaboração própria

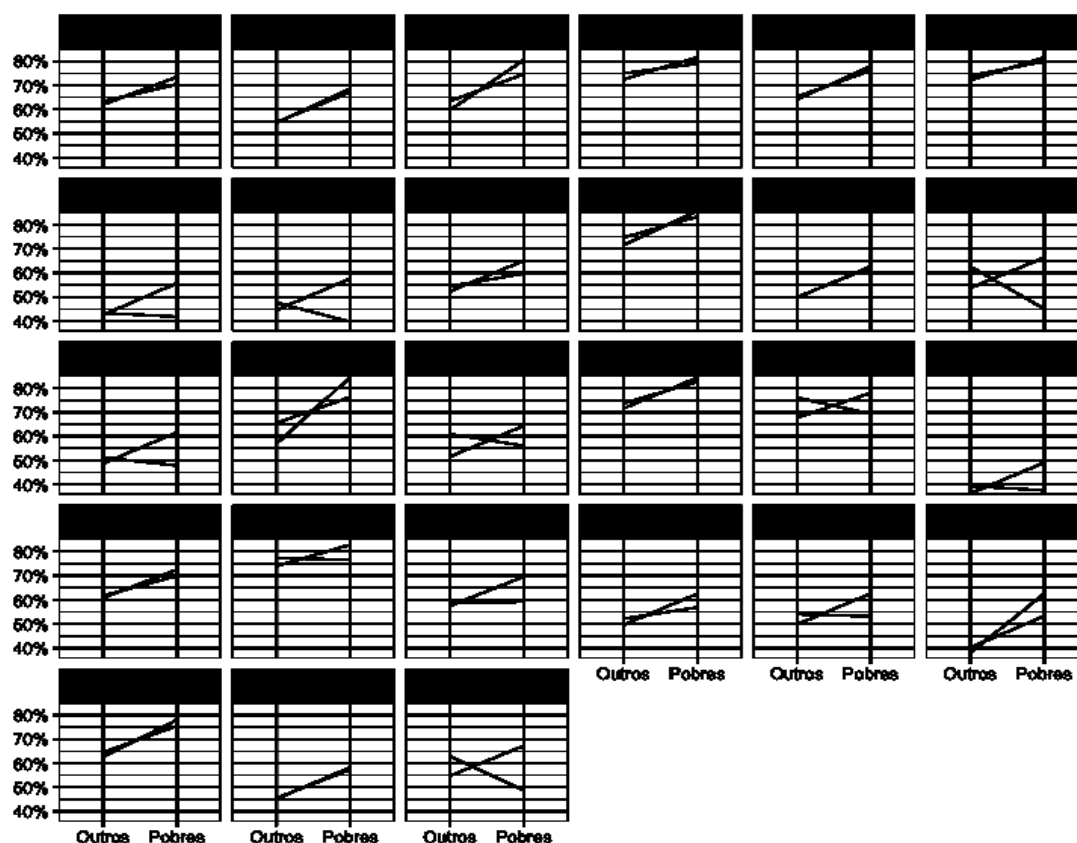
No modelo 2, incluímos a variável renda individual como preditor com efeitos fixos e autorizamos a variação do intercepto por estados. No resultado podemos ver que nossa principal variável independente tem um impacto significativo no voto em Dilma Rousseff e que os pobres têm maior probabilidade de sufragar a candidata petista que os não-pobres. Ademais, observamos uma variação entre os estados de 0,55 e entre indivíduos de 0,99.

O modelo 3, mais sintonizado com a proposta deste artigo, autoriza a variação dos interceptos e a inclinação da variável renda. Ou seja, nessa conformação percebemos as diferenças de comportamento de pobres e não-pobres, conforme as unidades estaduais. A variação de 0,19 no coeficiente de renda, ainda que menor do que a variação dos interceptos entre os estados, demonstra um efeito não desprezível, confirmando a hipótese de que a conduta dos eleitores difere de acordo com contextos específicos.

As estimações multinível acima mencionadas oferecem indícios de diferenças comportamentais segundo padrões geográficos, mas não nos autorizam a fazer inferências confiáveis, em virtude do número exíguo de observações e por ser uma amostra com representatividade nacional, e não subnacional. Para superar esse obstáculo, ponderamos as médias das probabilidades estimadas nas proporções desses subgrupos contidas no Censo Demográfico de 2010. Na figura 1 apresentamos as diferenças de predição dos modelos 2 e 3, por estado.

As linhas pretas enfatizam a variação dos coeficientes dos estados, com renda mantida constante; enquanto as linhas vermelhas informam as disparidades do comportamento eleitoral em níveis subnacionais. A consideração desse contraste revela dois quadros distintos. No primeiro caso, o modelo realça a preferência dos pobres pelo PT em todas as circunscrições examinadas. No segundo, por outro lado, atentamos para um paradigma mais matizado, com larga variância do coeficiente de renda, em que nem sempre os pobres optam em maioria por Dilma Rousseff. Senão vejamos: se, por exemplo, no Pará, Maranhão, Sergipe, São Paulo e Santa Catarina há vantagem considerável dos estratos mais carentes pelo PT; no Distrito Federal, no Rio Grande do Sul em Roraima e Rondônia esse efeito simplesmente não se manifesta e; mais interessante, no Espírito Santo, Tocantins, Mato Grosso do Sul o favoritismo muda de lado, com os não-pobres apoiando de maneira mais intensa os petistas.

Figura 1: Probabilidades estimadas dos modelos 2 e 3, com pós-estratificação



Fonte: Elaboração própria.

Nota: As linhas pretas correspondem às probabilidades estimadas dos coeficientes do modelo 2 e as linhas vermelhas correspondem às probabilidades estimadas dos coeficientes do modelo 3.

Como foi dito, o MRP estima respostas individuais em função de preditores demográficos e geográficos. Isto é, o MRP assume que os efeitos em um grupo de variáveis estão relacionados em razão de uma estrutura hierárquica. Ao contrário da regressão clássica, os modelos multinível agrupam os parâmetros em relação à média em um determinado nível. O grau de agrupamento surge endogenamente a partir dos dados, “com os pesos relativos determinados pelo tamanho da amostra do grupo, com variação interna e entre os segmentos avaliados” (Gelman e Hill, 2007:254, *tradução nossa*).

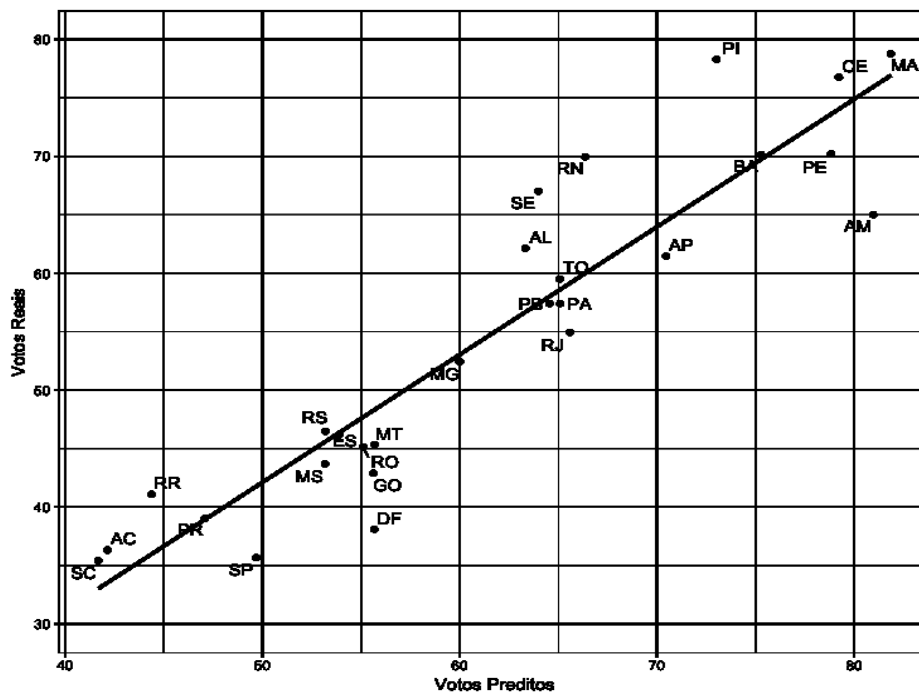
Assim, vale tornar nossas estimações um tanto mais complexas. No modelo final, avaliamos os efeitos de variáveis controle de nível estadual e individual. Para tanto, examinamos a conformação acima descrita, com *sexo, idade e raça* como controles de nível individual; e votação no PT em 2010 e renda domiciliar média como controles de nível estadual².

A figura 2 expõe a comparação entre os votos reais obtidos pelo PT nos estados e as predições do modelo final. De modo geral, nota-se a proximidade dos pontos em relação à curva de ajuste dos dados, o que fornece uma boa indicação de ajuste das estimações com o desempenho da candidatura petista no país. O ponto merece ser destacado: um modelo simples, apenas com renda individual, sexo, idade e raça como variáveis individuais, e renda estadual e votação em 2010 como variáveis de segundo nível, encontramos votos preditos que se aproxima sobremaneira dos votos observados, em cada estado, em 2014.

Os casos que mais se afastam da linha – Piauí, Amazonas e Acre – são também os que contam com menos observações na base de dados trabalhada, o ESEB 2014. A correlação das variáveis aponta um coeficiente de 0,94 e o erro médio absoluto das estimações é de 6,46 pontos percentuais, valores que certificam ainda mais a adaptação do modelo aos resultados da última eleição presidencial.

² Não utilizamos interações entre os preditores para não fugir do escopo do artigo – focado na disseminação subnacional do comportamento político baseado em renda domiciliar – e pelo fato das interações não melhorarem a capacidade preditiva dos parâmetros calculados (Lax e Phillips, 2013).

Figura 2 – Comparação dos votos reais e dos votos preditos no PT (2014)



Fonte: TSE

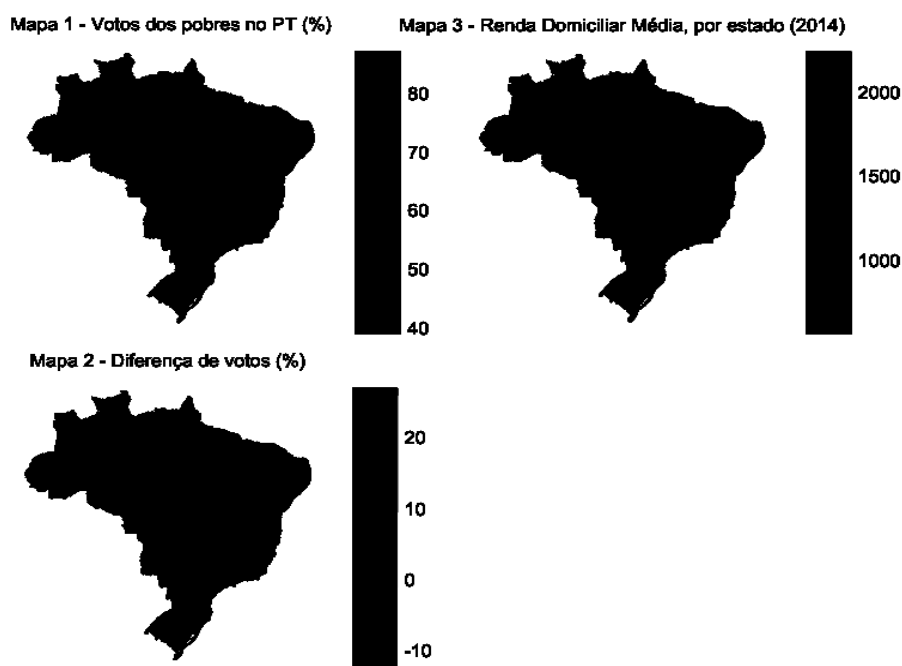
Se a principal vantagem do MRP é a capacidade de predição do comportamento político de subgrupos da população, algo impeditivo para outros modelos empregados até o momento pela literatura, os mapas abaixo exibem resultados analiticamente interessantes. No mapa 1, indicamos a porcentagem predita de voto dos eleitores pobres no PT e no mapa 2, a renda domiciliar média, ambas em nível estadual. A correlação negativa entre ambos é alta, de 0,74. Ou seja, quanto mais pobre é o estado, maior apoio dos pobres ao PT. No Maranhão, por exemplo, a preferência supera 80%.

À primeira vista, isso parece apenas corroborar a visão de o eleitorado petista são basicamente pobres nordestinos. No entanto, o modelo aplicado permite matizar essa interpretação. O mapa 3 indica o *gap* do efeito renda sobre o voto no PT. Ou seja, a diferença de voto entre pobres e não-pobres em direção ao PT. A comparação dessa distribuição com o mapa 2, de renda média estadual, indica que não existe correlação entre esses fenômenos, de 0,05.

Ou seja, se os indivíduos pobres apoiam mais o PT nos estados pobres que nos estados ricos, o mesmo acontece com os eleitores não-pobres. A coloração mais clara no Nordeste mostra que os não-pobres também manifestam simpatia pelo PT. No Piauí, Maranhão, Ceará, Pernambuco, Rio Grande do Norte e Paraíba, as taxas de votação dos não-pobres excedem os 70%. Os maiores *gaps* de efeito de renda ocorrem em localidades com

características econômicas diversas: as principais desproporções entre pobres e não-pobres aparecem no Pará (26%) e em Santa Catarina (22%)³.

Figura 4 - Voto no PT e PIB Estadual nas Eleições Presidenciais de 2014



A variação regional é, portanto, uma tônica de nossos achados. Os resultados estaduais destoam sobremaneira em um exame agregado. No entanto, diferentemente do que encontram Gelman et al. (2005; 2009) para o caso norte-americano, qual seja, que o comportamento dos eleitores ricos varia conforme contexto estadual, no caso brasileiro encontramos que tanto pobres quanto não-pobres são afetados, na sua direção de voto, pelo nível de riqueza de seus estados.

³ Nessa perspectiva, o presente trabalho reforça o esforço anterior de Argelina et al. (2018). No artigo, os autores enfatizam que o impacto do termo interativo entre renda e região não apresenta significância estatística em 2014. Ou seja, de acordo com a interpretação dos coeficientes, não existiam diferenças comportamentais entre pobres e não-pobres nas regiões investigadas.

Considerações Finais

Ao longo do artigo testamos um novo método para estimar a escolha o voto de um subgrupo importante nas eleições nacionais: os pobres. Desde 2006, quando a literatura de comportamento político dedicou suas atenções à ideia de realinhamento, as condições econômicas dos eleitores ganharam ainda mais proeminência nas teses de comportamento político. Explicações atinentes à ascensão de classe (Peixoto e Rennó, 2011), à identificação simbólico-afetiva com a figura de Lula (Singer, 2012) ou mesmo uma suposta tendência governista das camadas populares (Zucco, 2008) exprimem algumas das justificativas levantadas para a predileção recente dos pobres com o PT. Se as teorias anunciam desacordos, nem sempre pontuais, a noção de que a desigualdade de renda pode se transformar em desigualdade política permanece algo bastante consensual.

O modelo utilizado, baseado em uma estrutura hierárquica e em pós-estratificação, revela, a um só tempo, predições próximas aos resultados eleitorais e nuances não exploradas pelos autores do tema. Embora os pobres, em média, tenham mesmo preferido Dilma Rousseff, essa situação muda um tanto conforme o contexto estadual. Ademais, os níveis de riqueza dos estados influenciaram o comportamento de ambos os estratos sociais examinados. As razões para tal, no entanto, serão objeto de nossos estudos futuros.

Referências Bibliográficas

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

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.

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.

Cepal (2018) Panorama Social de América Latina, 2017. (LC/PUB.2018/1-P), Santiago

Cesop (2018). Estudo Eleitoral Brasileiro 2014 (ESEB) In.: Banco de Dados do Centro de Estudos de Opinião Pública – CESOP-UNICAMP.

Cho, Wendy (1998) “If the Assumption Fits...: A Comment on the King Ecological Inference Solution”. *Political Analysis*, 7, 143-163.

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

Coleman, James (1994) Foundations of Social Theory. Belknap Press of Harvard University Press.

Corrêa, Diego (2015) “Os custos eleitorais do Bolsa Família: reavaliando seu impacto sobre a eleição presidencial de 2006”. In.: *Opinião Pública*, vol. 21, n.3.

Couto, Cláudio (2014) “2014 - Novas eleições críticas?”. Em Debate, Belo Horizonte, v.6, n.6, p.17-24.

Dogan, Mattei e Rokkan, Stein. (1969). Quantitative ecological analysis in the social science. Massachusetts Institute of Technology.

Figueiredo, Argelina; Maciel, Natália; Simoni Jr, Sergio; Silva, Thiago Moreira da (2018) “O Voto do Eleitor Pobre nas Eleições Presidenciais (1989-2014)”. No prelo.

Fizbein *et al.*, (2009). Conditional cash transfer – A World Bank Policy Research Report.

Gelman *et al.* (2007) “Rich State, Poor State, Red State, Blue State: What’s the Matter with Connecticut?”. *Quarterly Journal of Political Science*, 2: 345–367.

Gelman et al. (2009). Red State, Blue State, Rich State, Poor State Why Americans Vote the Way They Do. Princeton University Press.

Hunter, Wendy e Power, Timothy. (2007). “Rewarding Lula: Executive power, social policy and the Brazilian elections of 2006” In.: *Latin American politics and society*. Vol., 49, n. 1.

Kastellec, J. P., Lax, J. R., & Phillips, J. (2010). Estimating state public opinion with multi-level regression and poststratification using R. *Unpublished manuscript, Princeton University*.

Katz, Jonathan e King, Gary. (1999). “A statistical model for multiparty electoral data”. In.: *American Political Science Review*, vol. 93, n.1.

King, Gary. (1997). A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton: Princeton University Press.

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.

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. (2014). “A base e os partidos”. In.: *Novos Estudos Cebrap*, n. 99.

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.

Marzagão, Thiago. (2013). “A dimensão geográfica das eleições brasileiras”. In.: *Opinião Pública*, vol. 19, n2.

Nicolau, Jairo (2015) “Eleições presidenciais nos municípios brasileiros: uma breve comparação dos resultados do primeiro turno de 2010 e 2014”, Em Debate, Belo Horizonte, v.7, n.2, p.29-41.

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.

Peixoto, Vitor e Rennó, Lúcio (2011) “Mobilidade social ascendente e voto: as eleições presidenciais de 2010 no Brasil”. *Opinião Pública*, Campinas , v. 17, n. 2, p. 304-332.

Przeworski, Adam (1974) “Contextual models of political behavior”, *Political Methodology*, Vol. 1, No. 1.

Przeworski, Adam e Sprague, John. (1986). *Paper stones*. The University of Chicago Press.

Reis, Fábio Wanderley (2014) “Eleição de 2014: “País dividido” e a questão social”. Em Debate, Belo Horizonte, v.6, n.6, p.8-16

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.

Simoni Jr, Sergio (2017) *Política distributiva e competição presidencial no Brasil: Programa Bolsa-Família e a tese do realinhamento eleitoral*. Tese de doutorado, DCP-USP.

Singer, André. (2012). *Os sentidos do lulismo*. Ed. Companhia das Letras.

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.

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.

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. (2010). "Poor places, poor voters: Persisting patterns in presidential elections in Brazil", Working paper.

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).

Zucco, Cesar e Power, Timothy. (2013). "Bolsa-Família and the Shift in Lula's Electoral Base, 2002-2006". In.: *Latin American Research Review*, 48(2).