

O BOLSA FAMÍLIA E AS ELEIÇÕES PRESIDENCIAIS NO BRASIL: UM MODELO DE PREDIÇÃO ELEITORAL¹

Rodrigo Lins²
Dalson Figueiredo Filho³
Lucas Silva⁴
Enivaldo Rocha⁵

Resumo: Este artigo apresenta um modelo de predição para as eleições presidenciais no Brasil. A hipótese sustenta que quanto maior a dependência do programa Bolsa Família, maior é a quantidade de votos auferidos pelo candidato mandatário. O desenho de pesquisa combina estatística descritiva, multivariada e espacial para analisar um banco de dados original elaborado a partir de dados secundários. Os resultados sugerem que: (1) em 2006, o modelo previu 50,82% versus um observado de 48,61% (erro de 2,21%); (2) em 2010, o modelo previu 47,49% versus um observado de 46,91% (erro de 0,58%) e (3) em 2014, o modelo estimou 45,25% versus um observado de 41,25% (erro de 3,66%). Comparativamente, nosso modelo apresentou um erro médio de 2,15%, sendo mais preciso do que outros presentes na literatura. Com esse artigo, esperamos difundir a utilização de modelos de predição na Ciência Política brasileira.

Palavras-chave: modelos de predição; eleições presidenciais; métodos quantitativos.

Abstract: This paper develops a prediction model for presidential elections in Brazil. The hypothesis states that the higher the level of dependence of *Bolsa Família* resources, higher it is the incumbent share of the votes. The research design combines descriptive, multivariate and spatial statistics to analyze an original dataset based on secondary information. The results suggest that: (1) in 2006, the model predicted 50,82% versus an observed value of 48,61% (error of 2,21%); (2) in 2010, the model predicted 47,49% versus an observed value of 46,91% (error of 0,58%) and (3) in 2014, the model predicted 45,25% versus an observed of 41,25% (error of 3,66%). Comparatively, our model has a lower mean error than other models in the literature. With this paper we hope to diffuse the use of prediction models in Brazilian Political Science.

Keywords: prediction models; presidential elections; quantitative methods.

Resumen: Este artículo presenta un modelo de pronóstico para las elecciones presidenciales en Brasil. La hipótesis sostiene que cuanto mayor es la dependencia del programa *Bolsa Família*, mayor será el número de votos recibidos por el presidente beneficiado. El diseño de investigación mezcla análisis estadístico descriptivo, multivariado y espacial para analizar una base de datos única a partir de datos secundarios. Los resultados sugieren que: (1) en 2006, el modelo predijo 50,82% frente al observado de 48,61% (error de 2,21%); (2) en 2010, el modelo predijo 47,49% frente a un observado de 46,91% (0,58% de error) y (3) en 2014, el modelo estimó 45,25% frente a un observado de 41,25% (error de 3,66%). Comparativamente, el modelo mostró un error promedio de 2,15%, siendo más preciso que otros en la literatura. Con este artículo, esperamos contribuir a la utilización de modelos predictivos en la ciencia política brasileña.

Palabras-clave: modelos de predicción; elecciones presidenciales; métodos cuantitativos.

All models are wrong, but some are useful
George Box

The only function of economic forecasting is to make astrology look respectable
John Kenneth Galbraith

¹ Agradecemos ao *Berkeley Institute for Transparency in the Social Sciences* (BITSS) pelo treinamento recebido em 2014 e 2016, ao *Teaching Integrity in Empirical Research* (TIER) workshop pelo apoio logístico durante 2016-2017 e ao CNPq/CAPES pelo apoio financeiro. Este trabalho também se beneficiou dos comentários dos pareceristas anônimos da Revista Eletrônica de Ciência Política (RECP) e dos membros do grupo de Métodos de Pesquisa em Ciência Política da Universidade Federal de Pernambuco (DCP – UFPE). Erros remanescentes são monopólio dos autores.

² Doutorando em Ciência Política pela UFPE

³ Doutor em Ciência Política pela UFPE e professor na UFPE

⁴ Graduando em Ciência Política pela UFPE

⁵ Professor titular da UFPE

INTRODUÇÃO

O volume 47 de Abril de 2014 da *Political Science and Politics* traz uma coleção de 16 artigos sobre previsão das eleições presidenciais norte-americanas⁶. No mesmo ano, a *American Political Science Association* (APSA) abordou o tema em um encontro profissional⁷. Enquanto Michael Lewis-Beck é um dos principais estudiosos do assunto (LEWIS-BECK, 1985; 1995; 2005; LEWIS-BECK e RICE, 1984; 1992), Nate Silver⁸, Drew Linzer⁹, entre outros, contribuíram para a disseminação dessa área de pesquisa entre o público não especializado.

Curiosamente, ao se *googlar* “Previsões Eleitorais Brasil”, observa-se registros inusitados. Por exemplo, o primeiro *link* de retorno anunciava que: “Previsões místicas para a disputa eleitoral entre Dilma e Aécio: Veja o que tarô, astrologia, búzios e numerologia dizem sobre a eleição”¹⁰. Em particular, Pai Carlinhos¹¹, Pai Uzêda, entre outros, apresentam previsões sobre temas variados: eleições, futebol, celebridades. Enquanto Pai Carlinhos apostava em Aécio Neves para as eleições presidenciais de 2014, a previsão de Pai Uzêda não foi muito objetiva. A reportagem relatou que “o PT terá sucesso apenas se conseguir vencer a eleição em primeiro turno porque a união entre PSB, Rede Sustentabilidade e PSDB pode ganhar força caso a decisão chegue ao segundo turno”¹².

Essas previsões sofrem de vários problemas. Em primeiro lugar, não se sabe exatamente de onde elas vêm. Segundo, inexiste teoria e não há critério de replicabilidade (KING, 1995; PARANHOS et al, 2014; JANZ, 2015). Terceiro, não há nenhuma estimativa de incerteza associada à previsão. Quarto, afirmar que o partido apenas ganhará se vencer no primeiro turno e caso contrário enfrentará uma competição mais acirrada é uma afirmativa óbvia. Logo, não se enquadra como previsão. Para além de todos esses problemas, a questão importante é: por que confiar em prognóstico dessa natureza se já existem modelos teóricos capazes de oferecer previsões mais confiáveis?

Este artigo apresenta um modelo de predição para as eleições presidenciais no Brasil. O foco repositivo na correlação entre a proporção de beneficiários do programa Bolsa Família e o percentual de votos recebido pelo candidato *incumbent* nas eleições de 2006, 2010 e 2014. A hipótese sustenta que quanto maior a dependência do Bolsa Família, maior a quantidade de votos auferidos pelo candidato mandatário. O desenho de pesquisa combina estatística descritiva, multivariada e espacial para analisar um banco de dados original elaborado a partir de fontes secundárias.

O restante do artigo está dividido da seguinte forma. A próxima parte traz uma revisão da literatura sobre modelos de previsão eleitoral na Ciência Política, enfatizando os critérios de avaliação e as principais dificuldades na elaboração desses modelos. A seção posterior apresenta as principais dificuldades na elaboração. Depois disso, apresenta-se todos os procedimentos metodológicos. A seção seguinte, apresentamos a metodologia e os principais resultados respectivamente. Por fim, a última seção sumariza as conclusões.

⁶ Ver: <<http://journals.cambridge.org/action/displayIssue?decade=2010&jid=PSC&volumeId=47&issueId=02&iid=9235620>>.

⁷ Para detalhes sobre o evento, ver: <<https://www.youtube.com/watch?v=VU69zcAW5uk&feature=youtu.be>>. Além disso, em 2007, a APSA criou o *Political Forecasting Group*.

⁸ Ver: <<http://fivethirtyeight.blogs.nytimes.com/fivethirtyeights-2012-forecast/>>.

⁹ Ver: <<http://votamatic.org/about-me/>>.

¹⁰ Ver: <<http://oglobo.globo.com/brasil/previsoes-misticas-para-disputa-eleitoral-entre-dilma-aecio-14343321#ixzz3vAXIOkRF>>. Acessado em 23 de dez. 2015

¹¹ Ver: <<http://www.youtube.com/watch?v=97JJgJJaQkk>>.

¹² Ver: <<http://www.jornaldebrasil.com.br/noticias/politica/518162/uzeda-o-pai-de-santo-da-politica-faz-previsoes-para-2014/>>.

OS MODELOS DE PREVISÃO ELEITORAL NA CIÊNCIA POLÍTICA¹³

A literatura sobre previsões eleitorais começou a ser produzida sistematicamente no início da década de 1980 nos Estados Unidos e depois foi replicada em outros países como França, Alemanha e Canadá (LADEAU, LEWIS-BECK e BELANGER, 2012). Em geral, existem três principais abordagens para estimar modelos de previsão eleitoral. O Quadro 1 sumariza essas informações.

Quadro 1 - Modelos de previsão eleitoral

Abordagem	Objetivo	Modelo	Exemplos
Eleitoral	Analisar o impacto da receita de campanha sobre o percentual de votos obtido pelo candidato	Regressão linear de mínimos quadrados ordinários	Jacobson (1978); Green e Krasno (1988); Gerber (1998); Abromowitz (1988; 2004)
Econômica	Observar o efeito de variáveis econômicas, como variação do PIB per capita, sobre o percentual de votos obtido pelo candidato	Regressão linear de mínimos quadrados ordinários	Fair (1978); Tufte (1978); Lewis-Beck e Rice (1982, 1992); Hibbs (1982); Norpoth (2004); Cuzán e Bundrick (1999);
<i>Survey</i>	Examinar as chances de sucesso de um candidato a partir de pesquisas de intenção de voto	Modelo de probabilidade condicional	Lamb (2007), Silver (2012) e Drew Linzer (2012)

Fonte: elaborado pelos autores

A abordagem eleitoral assume que a votação do candidato é função de variáveis eleitorais, em particular, a receita de campanha. A hipótese típica de pesquisa sustenta que quanto maior a quantidade de recursos investidos na campanha, maior a quantidade de votos. Exemplos de trabalhos nesse sentido podem ser encontrados em Jacobson (1978), Abromowitz (1988), Green e Krasno (1988) e Gerber (1998; 2004).

A abordagem econômica assume que a votação dos candidatos é uma função de variáveis econômicas, em particular, a variação do Produto Interno Bruto (PIB) per capita no período pré-eleitoral. A hipótese clássica é de que quanto melhor a economia, maior a quantidade de votos conquistada pelo candidato mandatário. Exemplos de trabalho nesse sentido podem ser encontrados em Fair (1978), Hibbs (1982), Cuzán e Bundrick (1990), Lewis-Beck e Rice (1982, 1992) e Norpoth (2004).

Por fim, a abordagem baseada em *surveys* utiliza modelos de probabilidade condicional baseado em pesquisas de intenção de voto para prever as chances de vitória de um determinado candidato em um processo eleitoral. Tecnicamente, essa abordagem utiliza séries históricas de pesquisas de intenção de votos. São realizados cálculos condicionais com intuito de mensurar as chances de um determinado candidato vencer ou perder, levando em consideração os falsos positivos e negativos e o peso dos institutos de pesquisa a partir das

¹³ Ver Lewis-Beck (1985), Lewis-Beck e Rice (1982; 1984; 1992). Campbell (2008) faz uma ampla análise das críticas feitas aos modelos de predição presidencial, assim como oferece uma leitura especial para a distinção entre eleições com *incubents* e aquelas com assentos livres (*open-seat elections*). Somos gratos ao parecerista anônimo por chamar nossa atenção para este último trabalho. Para trabalhos didáticos, ver Jones (2002), Lewis-Beck (2005), Campbell e Lewis-Beck (2008) e o PollyVote <<http://pollyvote.ifkw.uni-muenchen.de/en/>>.

médias de erros em trabalhos passados. Essa abordagem ganhou grande notoriedade após a divulgação das previsões de Silver (2012) em relação às eleições presidenciais estadunidenses.

No Brasil, alguns trabalhos de predição eleitoral se destacam¹⁴. As eleições de 2006 – primeira após a implementação do programa Bolsa Família e que culminou na reeleição de Lula – é uma fonte especialmente prolifera de trabalhos. Nicolau e Peixoto (2007), ao analisar dados a nível municipal, perceberam algumas pistas – como os próprios autores colocam – sobre a tendência de voto no candidato à reeleição. A primeira foi a alta correlação entre gastos do Bolsa Família e as áreas mais pobres do país. A segunda foi a concentração do programa nas regiões Norte e Nordeste. Os autores concluem que três questões foram fundamentais para entender o voto no país: desenvolvimento social, valores do Bolsa Família e a região em que o município está localizado.

Com ferramentas da análise espacial, Soares e Terron (2008) também analisam a participação do Bolsa Família na mudança geográfica dos votos obtidos por Lula em 2006, se comparado com 2002. Os autores percebem que Lula ainda recebia apoio da base anterior – embora com menor –, mas com uma maior coesão regional em municípios menos populosos e mais pobres. Soares e Terron (2008) concluem, que o programa foi o fator com maior peso na explicação dos votos municipais.

Zucco (2008) também mostra que a forte relação entre Bolsa Família e a votação obtida por Lula não está restrita ao nível estadual. Quando levado em conta os municípios, essa relação é ainda mais forte. O esforço feito por Zucco (2008) é similar ao de Nicolau e Peixoto (2007), mas com a inclusão de outras variáveis independentes. A conclusão é de que os votos concentrados nas regiões Norte e Nordeste não se deu por polarização ideológica, nem por rivalidade regional. Segundo Zucco (2008) é “*merely the response of voters to material changes in their living conditions, much of which can be traced to the Bolsa Família cash transfer programme*” (ZUCCO, 2008, p. 47).

Já com uma análise dados ao nível individual, Licio, Rennó e Castro (2009) usam uma amostra probabilística nacional não só para analisar os votos em Lula, mas também sua avaliação durante o segundo governo. Os achados ratificam a ideia de que não só os beneficiários tiveram uma tendência de votar na reeleição, mas também avaliavam melhor o governo.

COMO AVALIAR OS MODELOS DE PREVISÃO ELEITORAL?¹⁵

De acordo com Lewis-Beck (2005), existem quatro elementos para julgar a qualidade de um modelo de previsão¹⁶: 1) acurácia; 2) antecipação; 3) parcimônia e 4) reproduzibilidade. A precisão da previsão é essencial. Quanto mais precisa a estimativa, melhor. Contrariamente, quanto mais distante do valor observado, pior é a previsão. A antecipação diz respeito a quantidade de tempo que antecede a previsão. Para o autor, quanto antes a previsão, melhor. O princípio da parcimônia não é específico aos modelos de previsão, mas diz respeito à atividade científica de forma geral. Mantendo todos outros elementos constantes, quanto mais simples o modelo, melhor. Tecnicamente, como a maior parte das amostras é pequena ($n < 30$), o critério da parcimônia é essencial já que a cada variável adicional perde-se muito grau de liberdade. No limite, um modelo saturado pode ser impossível de ser estimado, já que a quantidade de parâmetros supera o número de

¹⁴ Gostaríamos de agradecer a um parecerista anônimo por chamar nossa atenção para esta literatura.

¹⁵ Lewis-Beck (2005) diferencia previsão, que se refere ao futuro, de predição, que remete a algo que já ocorreu. Formalmente, todavia, o tratamento técnico dos dados é o mesmo. Nosso modelo foi originalmente elaborado antes do resultado das eleições. No entanto, o processo editorial apenas terminou após a eleição de Dilma, o que nós obrigada a trocar a terminologia conceitual, mas não altera o propósito central do artigo que é difundir a utilização de modelos de previsão no Brasil.

¹⁶ Esquemáticamente, de acordo com Lewis-Beck (2005), a qualidade de um modelo pode ser mensurada a partir da seguinte equação: $Q = [(3a + P + R)A] / M$. Em que Q é a qualidade do modelo, a acurácia tem peso 3, os demais componentes (antecipação, parcimônia e reproduzibilidade) tem peso 1. Todas as variáveis variam entre zero e dois. Dessa forma, M representa o valor máximo, no caso, 20, para que o indicador fique entre zero e um.

observações. Por fim, tem-se a reproduzibilidade. Um bom modelo de previsão é aquele que qualquer pessoa pode reproduzir e atingir os mesmos resultados. Não adianta propor um modelo preciso, com antecipação adequada e parcimonioso se não é possível reproduzir publicamente os resultados.

AS PRINCIPAIS DIFICULDADES ENFRENTADAS NA PREVISÃO ELEITORAL

A despeito da variedade e principalmente sofisticação metodológica dos modelos estimados, ainda são raros os estudos sistemáticos fora do contexto norte-americano. A literatura comparada geralmente é formada por estudos de caso que replicam os modelos elaborados para os Estados Unidos. Por exemplo, em 2012, o *International Journal of Forecasting* publicou um número especialmente voltado para modelos preditivos. Lewis-Beck e Tien (2012) analisam o caso do Japão. Magalhães, Aguiar-Conraria e Lewis-Beck (2012) examinam o caso espanhol enquanto Dassonneville e Hooghe (2012) examinam o contexto belga. Salvo melhor catalogação, Turgeon e Rennó (2012) constituem o trabalho pioneiro no Brasil a replicar modelos consolidados.

Tecnicamente, o principal impeditivo para o exame mais aprofundado desse tema em outros países é o reduzido número de observações disponíveis para a variável dependente (TURGEON e RENNÓ, 2012). Por exemplo, no Brasil, apenas existem informações para seis casos: 1994, 1998, 2002, 2006, 2010 e 2014, o que reduz a confiabilidade de qualquer estimativa. Isso porque quanto menor o número de casos, menos informação estará disponível para estimar os parâmetros de interesse. Este artigo segue as orientações de Turgeon e Rennó (2012) sobre como lidar com os problemas de “n” reduzido, ou seja, amostras pequenas. A próxima seção apresenta a metodologia do artigo.

METODOLOGIA

Esta seção descreve todos os procedimentos metodológicos com o objetivo de aumentar a transparência e garantir a replicabilidade dos resultados (KING, 1995; JANZ, 2015). O Quadro 2 sintetiza as principais características do desenho de pesquisa.

Quadro 2 - Desenho de pesquisa

População	Candidatos mandatários nas eleições presidenciais de 2006, 2010 e 2014
Variáveis	<i>Dependente:</i> percentual de votos recebido pelo candidato mandatário <i>Independente:</i> índice bolsa família (IBF)
Unidade de análise	Estado/Ano
Hipótese	Quanto maior a dependência do programa Bolsa Família, maior é a quantidade de votos auferidos pelo candidato mandatário
Técnicas	Estatística descritiva, multivariada e análise espacial
Fontes	Tribunal Superior Eleitoral (TSE), Ministério de Desenvolvimento Social e Combate à Fome (MDS) e Instituto Brasileiro de Geografia e Estatística (IBGE)

Fonte: elaboração dos autores

A população de interesse são os candidatos *incumbents* nas eleições presidenciais de 2006, 2010 e 2014. A variável dependente é o percentual de votos recebido pelo candidato em cada Estado. A variável explicativa é o nível de dependência de de cada ente estatal em relação aos recursos do Programa Bolsa Família

(IBF). Esse indicador foi calculado a partir da razão entre o número de famílias beneficiadas e o tamanho da população de cada unidade da federação. A hipótese sustenta que quanto maior a dependência do programa Bolsa Família, maior é a quantidade de votos auferidos pelo candidato mandatário. Tecnicamente, o desenho de pesquisa combina estatística descritiva, multivariada e análise espacial para examinar uma base de dados original.

Os dados foram coletados a partir do TSE, MDS e IBGE. A tabulação foi realizada a partir do *Microsoft Excel* e os modelos foram estimados a partir do *Statistical Package for Social Science*, versão 20. A análise espacial foi realizada com o auxílio do Geoda e do QGIS. Como procedimento de armazenamento e tratamento das informações, seguimos o protocolo TIER 2.0 que está disponível publicamente no endereço eletrônico da *Haverford College*¹⁷.

RESULTADOS

A média do IBF, em 2006, foi de 6,98. Já em 2010, foi de 8,03 e, em 2014, 8,36, conforme mostra a Tabela 1. Podemos verificar um aumento médio na dependência do programa bolsa família entre os anos eleitorais.

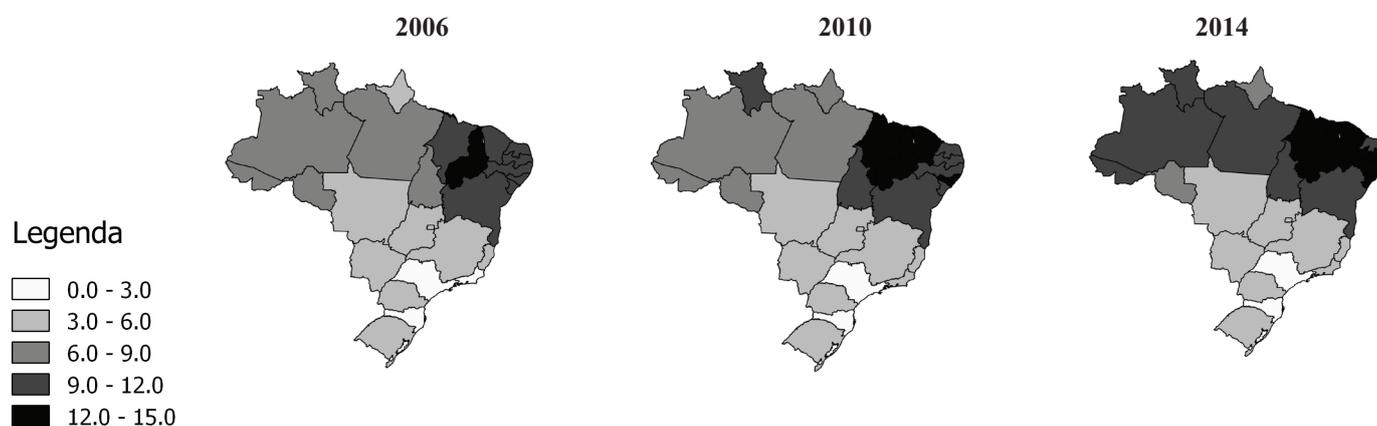
Tabela 1 - Estatística descritiva do IBF

Ano	Média	Mediana	Mínimo	Máximo	Desvio Padrão
2006	6,98	6,16	2,33	12,08	3,05
2010	8,03	8,00	2,30	13,48	3,50
2014	8,36	9,27	2,03	14,38	3,93

Fonte: elaboração dos autores

Analisando espacialmente IBF através do Mapa 1, podemos constatar que, em 2006, o Piauí (12,08) possuiu a maior média, enquanto que Santa Catarina (2,33), São Paulo (2,71) e Rio de Janeiro (2,84) as menores. Em 2010, Piauí (13,48), Alagoas (13,27), Maranhão (13,25) e Ceará (12,09) possuíram as mais elevadas dependências, enquanto que Santa Catarina (2,30) e São Paulo (2,85) as menores. Por fim, em 2014, Maranhão (14,38), Piauí (14,31), Paraíba (13,35), Alagoas (13,29), Sergipe (12,69), Pernambuco (12,42) e Ceará (12,36) possuíram os maiores valores do IBF, do outro lado, novamente, Santa Catarina (2,03) e São Paulo (2,88) os menores.

Mapa 1 – IBF no espaço

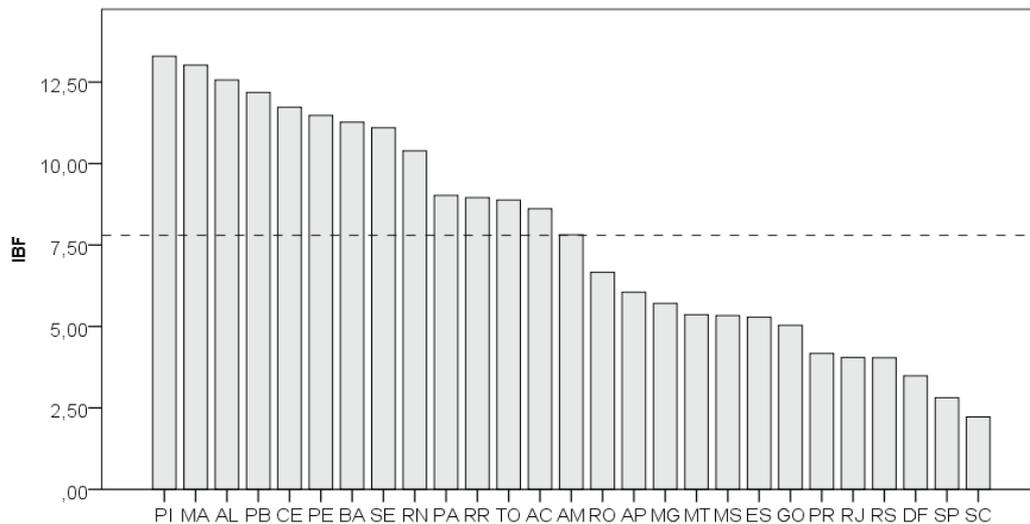


Fonte: elaborado pelos autores

¹⁷ Ver: < <https://www.haverford.edu/project-tier/protocol-v2> >

Além disso, constata-se uma alta correlação espacial do IBF em 2006 (Moran's I: 0,735; p-valor<0,001; 999 permutações), 2010 (Moran's I: 0,780; p-valor<0,001; 999 permutações) e 2014 (Moran's I: 0,762; p-valor<0,001; 999 permutações). Podemos identificar um padrão espacial na distribuição do IBF. Isso quer dizer que casos com maior dependência do Bolsa Família estão próximos de outros na mesma situação, notadamente os estados das regiões Norte e Nordeste. O Gráfico 1 ilustra a média do IBF por estado no período analisado.

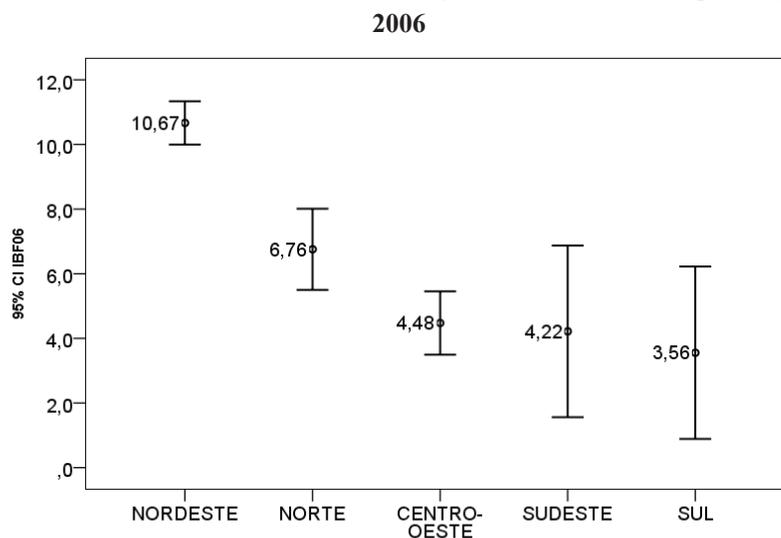
Gráfico 1 – IBF médio (2006, 2010 e 2014)



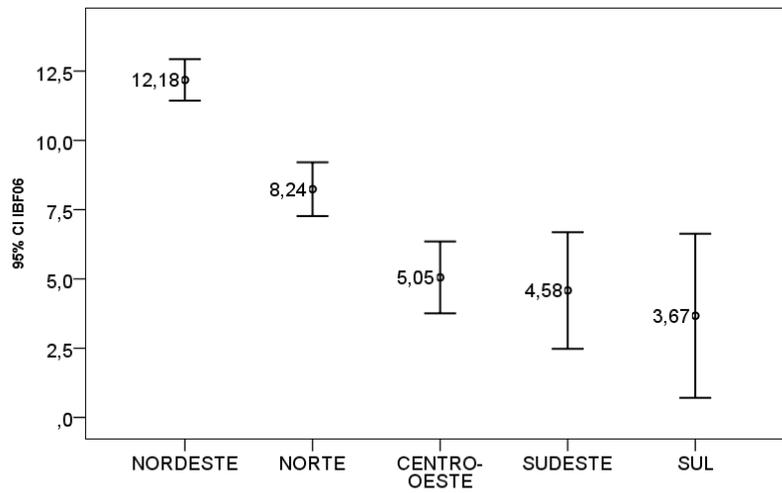
Fonte: elaboração dos autores.

Comparativamente, por região geográfica, observamos, a partir do Gráfico 2, um certo padrão do IBF. Nordeste e Norte possuem as maiores médias, enquanto Sul e Sudeste as menores. O Centro-Oeste ocupa uma posição intermediária. A diferença média entre Norte e Nordeste e as demais regiões é estatisticamente significativa para todos os anos, já que não existe sobreposição dos intervalos de confiança

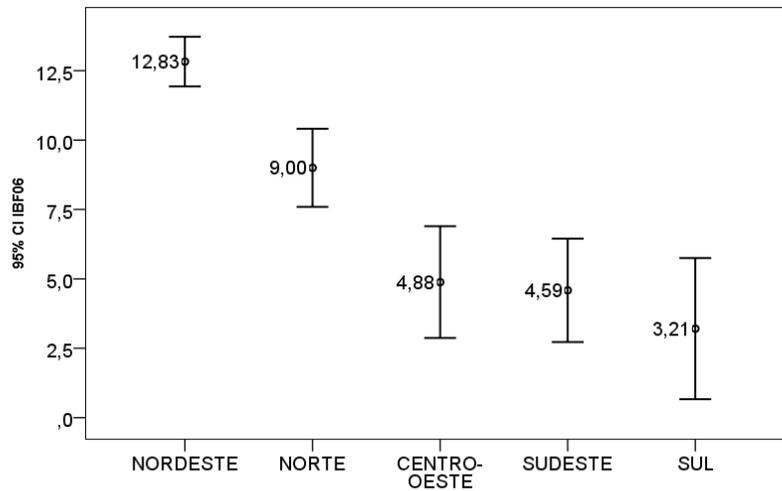
Gráfico 2 – Diferença média entre as regiões (IC 95%)



2010



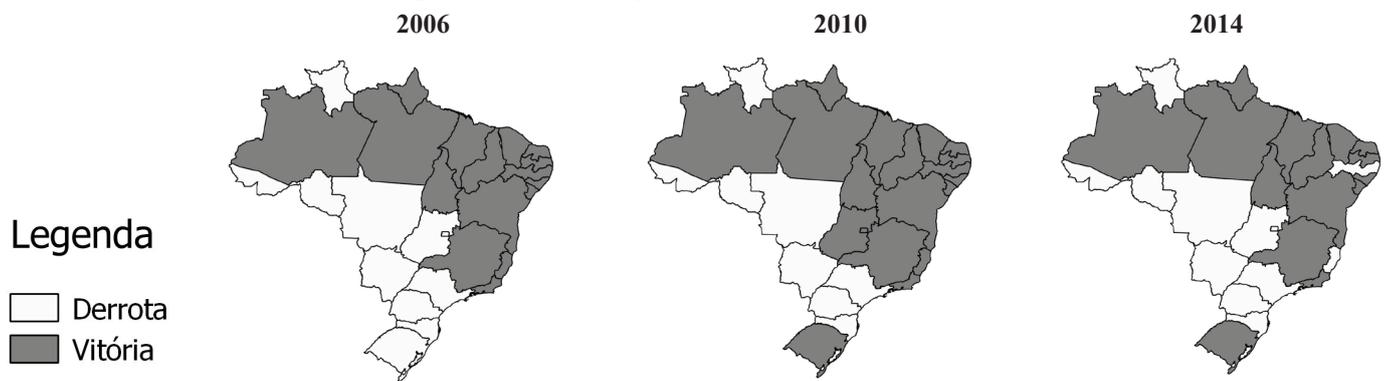
2014



Fonte: elaborado pelos autores

Analisando a votação do candidato *incumbent* no período observado, em 2006, ele venceu em 16 estados e obteve os maiores percentuais de votos no Amazonas (78,06), Maranhão (75,50) e Ceará (71,22). Já em 2010, a candidata do governo ganhou em 19 estados, tendo conquistado as maiores votações no Maranhão (70,65), Piauí (67,09) e Ceará (66,30). Por fim, a *incumbent*, em 2014, ganhou em 15 estados e alcançou as maiores votações, novamente, no Piauí (70,61), Maranhão (69,56) e Ceará (68,30). O Mapa 2 ilustra as vitórias dos candidatos da situação no país.

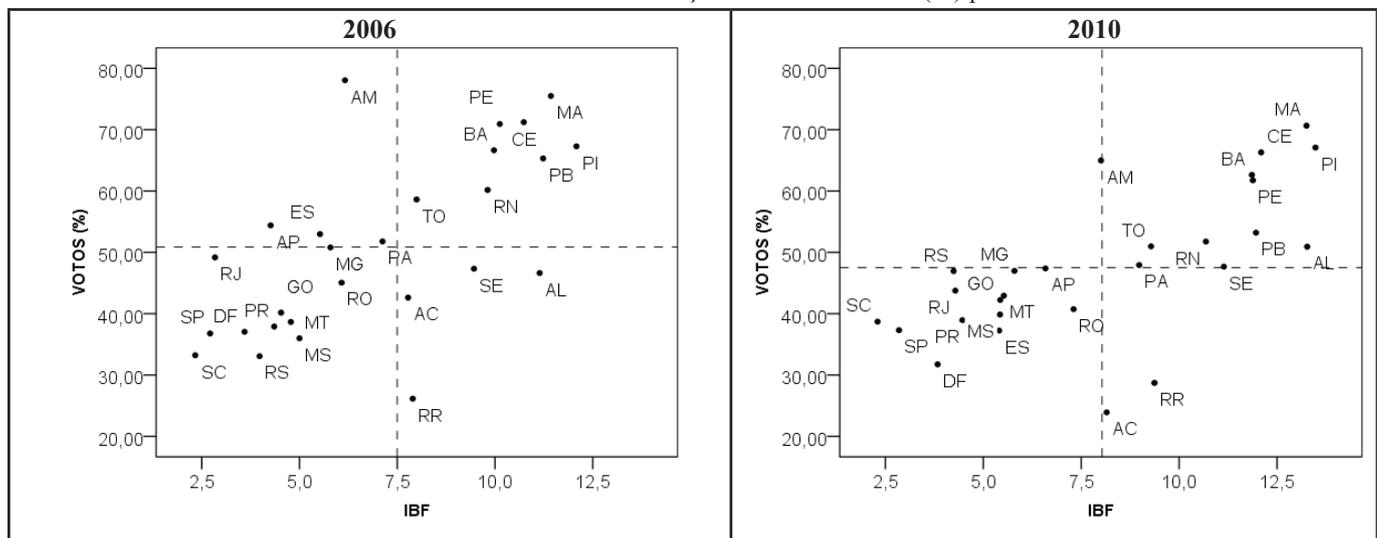
Mapa 2 – Estados em que o incumbente saiu vitorioso

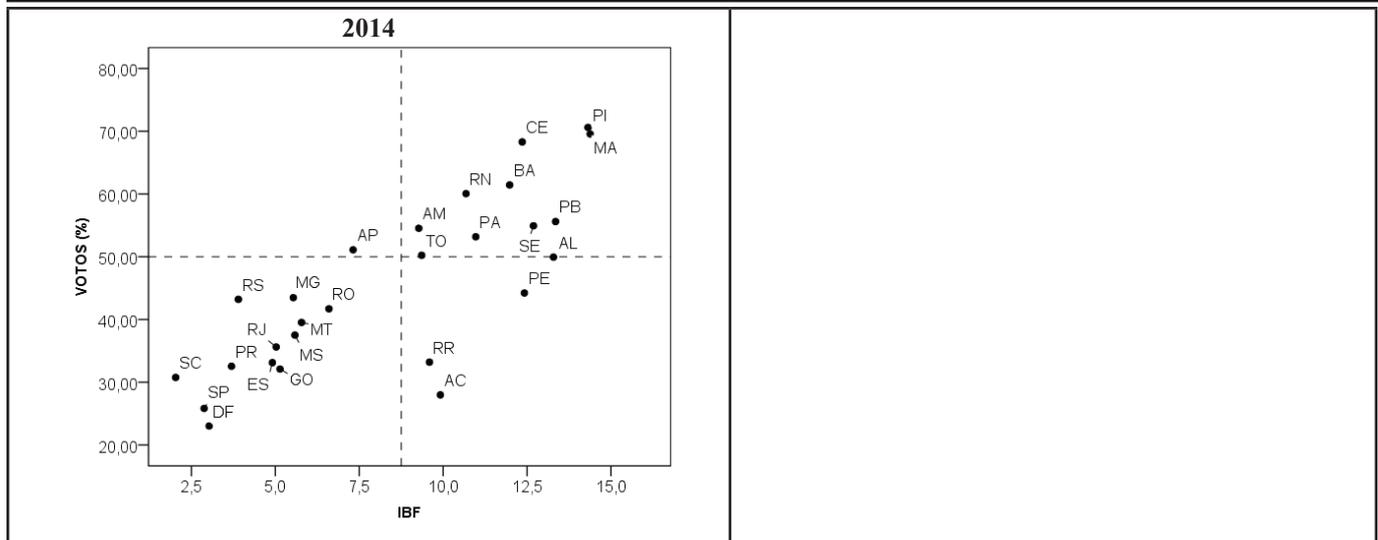


Fonte: elaborado pelos autores

O Gráfico 3 ilustra a correlação entre o nível de dependência do Bolsa Família (IBF) e a votação observada do candidato *incumbent* em 2006, 2010 e 2014 por estado. Em 2006, a correlação entre IBF e o percentual de votos válidos foi de 0,657 (p -valor $<0,001$; $n = 27$). O coeficiente de determinação (r^2) indica que o IBF explica cerca de 41% da variação dos votos nesse período. Em 2010, o padrão se manteve consistente com um coeficiente de 0,674 (p -valor $<0,001$; $n=27$) e um percentual de explicação de 45,4%. Por fim, em 2014, a correlação entre a porcentagem de votos e o nível de dependência do Bolsa Família foi de 0,805 (p -valor $<0,001$; $n = 27$), 19% maior do que na eleição anterior. O percentual de explicação do modelo passou de 45,4% para, aproximadamente, 65%.

Gráfico 3 - Correlação entre IBF e votos (%) por estado





Fonte: elaboração dos autores

O padrão é muito consistente: quando o IBF é alto, o percentual de votos recebido pelo candidato mandatário aumenta. À medida que a dependência do Bolsa Família diminui, reduz-se também a proporção de votos conquistada pelo candidato mandatário.

O modelo estimado assume que a votação do candidato mandatário é uma função do Índice Bolsa Família (IBF). O modelo foi definido assim: $Y = \alpha + \beta \cdot X + E$ onde Y representa o percentual de votos recebido pelo candidato *incumbent*, α representa a constante do modelo, ou seja, a melhor estimativa de Y na ausência de variáveis independentes, β representa a variação observada em Y ao se elevar X_1 em uma unidade. Por fim, o E representa o erro médio do modelo. A variável independente do modelo é o Índice Bolsa Família (X_1). A Tabela 2 sumariza as principais estatísticas de interesse do modelo linear de mínimos quadrados ordinários (MQO)¹⁸.

Tabela 2 - Estimativas do modelo MQO (dados observados)

Ano	α	β	IBF	Predito	Observado	Erro
2006	29,05	3,12	6,98	50,82	48,61	2,21
2010	28,95	2,31	8,03	47,49	46,91	0,58
2014	21,85	2,80	8,36	45,25	41,59	3,66

Fonte: elaboração dos autores

Em 2006, o modelo previu 50,82% *versus* um observado de 48,61% (erro de 2,21%). Em 2010, o modelo previu 47,49% *versus* um observado de 46,91% (erro de 0,58%). Por fim, em 2014, o modelo previu 45,25% contra um observado de 41,59% (erro de 3,66%). Ou seja, em todas as eleições, o modelo superestimou o verdadeiro valor do parâmetro.

Curiosamente, as pesquisas de intenção de voto realizadas pelo Datafolha e Ibope, ambas divulgadas em 04 de outubro de 2014, mostravam a intenção de voto no candidato mandatário em 44% e 46%, respectivamente¹⁹. Uma diferença de apenas 1,25% em relação ao DataFolha e 0,75% ao Ibope no que se refere ao modelo apresentado.

Por fim, ao se comparar o modelo desenvolvido neste artigo com outros propostos pela literatura, tem-se o seguinte cenário apresentado na Tabela 3.

¹⁸ Para uma introdução intuitiva para o modelo de mínimos quadrados ordinários (MQO), ver: Figueiredo Filho et al (2011).

¹⁹ Para informações sobre essas pesquisas, ver: <<http://eleicoes.uol.com.br/2014/pesquisas-eleitorais/brasil/1-turno/>>.

Tabela 3 - Comparação da previsão dos modelos

Ano	Abromowitz (2008) A (erro)	Abromowitz (2008) B (erro)	Campbell (2008) (erro)	Lewis-Beck <i>et al</i> (2008) (erro)	Lins et al (2016) (erro)
2006	52,09 (3,48)	51,68 (3,07)	47,23 (1,38)	53,67 (5,06)	50,82 (2,21)
2010	21,47 (25,50)	10,36 (36,60)	44,48 (2,49)	77,21 (30,24)	47,49 (0,58)
2014	-	-	-	-	45,25 (3,66)
ERRO MÉDIO	14,49	19,83	1,93	17,65	2,15

Fonte: elaboração dos autores a partir de Turgeon e Rennó (2012)

Como pode ser observado, o modelo A de Abromowitz (2008) apresentou uma previsão de 52,09% *versus* um observado de 48,61%, ou seja, um erro de 3,48%. Para 2010, o erro foi muito maior já que o modelo subestimou a real votação em 25,50 pontos percentuais. O nosso modelo apresentou um erro de predição médio de apenas 2,15%, ficando muito próximo do modelo de Campbell (2008) com um erro de 1,93%.

CONCLUSÃO

Este artigo apresentou um modelo de predição para as eleições presidenciais no Brasil. Os resultados indicam que quanto maior a dependência do programa Bolsa Família, maior é a quantidade de votos recebidos pelo candidato mandatário. Do ponto de vista geográfico, as regiões Nordeste e Norte são as mais vinculadas ao programa, enquanto que Sudeste e Sul são as menos vinculadas. Além disso, identificamos um padrão espacial em relação ao programa: estados mais dependentes estão mais próximos no espaço.

Por um lado, as estimativas do modelo se aproximam fortemente das pesquisas de intenção de voto divulgadas pelo Datafolha e Ibope no dia 04 de outubro de 2014. No entanto, nosso modelo tem três principais vantagens: (1) logisticamente mais barato; (2) mais facilmente replicável e (3) mais rapidamente implementado. A pesquisa do Datafolha demorou dois dias para ser realizada (03 e 04 de outubro) e entrevistou 18.116 eleitores. A pesquisa do Ibope foi realizada em três dias (02 a 04 de outubro), mas entrevistou uma amostra muito menor: 3.030 eleitores. Comparativamente, o nosso modelo apresenta um erro médio de 2,15%, sendo mais preciso do que outros modelos presentes na literatura.

É importante diferenciar modelos de previsão eleitoral de projeções de cenários eleitorais. Não existe mérito científico algum em antecipar um eventual segundo turno entre candidato X e Y ou X e Z. Isso não é previsão. Similarmente, não existe mérito algum em afirmar que durante um eventual segundo turno a competição será mais acirrada e/ou as eleições serão mais dependentes do apoio dos governadores. Isso são conjecturas possíveis, ou seja, fazem parte do espaço amostral. Um modelo de previsão eleitoral deve apresentar exatamente o valor da variável de interesse (previsão). Ou seja, deve reportar uma estimativa específica da variável dependente, além de medidas de variabilidade e incerteza da previsão.

Caso contrário, continuaremos dependendo de Pais de Santo que não só fazem previsões, como também orientam as estratégias eleitorais dos candidatos. Entre Pai Carlinhos e Pai Uzêda, é melhor ficar com Pai

REFERÊNCIAS

- ABRAMOVITZ, A. I. (1988). **Explaining Senate Election Outcomes.** *American Political Science Review*, Washington-DC, v. 82, n. 2, p. 385-403, May.
- ABRAMOVITZ, A. I. (1991). **Incumbency, campaign spending, and the decline of competition in U. S. house elections.** *Journal of Politics*, West Nyack, v. 53, p. 34-56, February.
- ABRAMOVITZ, A. I. (2008). **Forecasting the 2008 presidential election with the time-for-change model.** *PS: Political Science and Politics*, 41, 691-695.
- CAMPBELL, J. E. (2008). **The trial-heat forecast of the 2008 presidential vote: performance and value considerations in an open-seat election.** *PS: Political Science and Politics*, 41, 697-701.
- CAMPBELL, J. E.; LEWIS-BECK, M. (2008). **Special Issue: US Presidential Election Forecasting.** *International Journal of Forecasting*, 24, (April/June).
- CUZÁN, A. G. and BUNDRICK, C. M. (1999). **Fiscal Policy as a Forecasting Factor in Presidential Elections.** *American Politics Quarterly*, 29 (3): 338-353.
- DASSONNEVILLE, R.; HOOGHE, M. (2012). **Election Forecasting under Opaque Conditions.** A Model for Francophone Belgium, 1981-2010. *International Journal of Forecasting*, 28 (4), 777-788.
- FAIR, R. C. (1978). **The Effect of Economic Events on Votes for President.** *Review of Economics and Statistics* 60 (May): 322-325.
- FIGUEIREDO FILHO, D. et al (2011). **O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (MQO).** *Revista Política Hoje*, v. 20, n. 1.
- GERBER, A. (1998). **Estimating the effect of campaign spending on Senate election outcomes using instrumental variables.** *American Political Science Review*, 92, 401-411.
- GREEN, D. P.; KRASNO, J. S. (1988). **Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections.** *American Journal of Political Science*, Hoboken-NJ, v. 32, p. 884-907.
- HIBBS, D. A., Jr. (1982). **President Reagan's mandate from the 1980 elections: a shift to the right?** *American Politics Quarterly*, 10, 387-420.
- JACOBSON, G. C. (1978). **The Effect of Campaign Spending in Congressional Elections.** *American Political Science Review*, Washington-DC, v. 72, p. 469-491, May.
- JONES, R. J., Jr. (2002). **Who Will Be in the White House? Predicting Presidential Elections.** New York: Longman.
- KING, G. (1995). **Replication, Replication.** *PS: Political Science and Politics* 28: 443-499. Disponível: <<http://gking.harvard.edu/gking/files/replication.pdf>> Acesso em: 27 set. 2014.
- LEWIS-BECK, M. (1985). **Pocketbook voting in US national election studies: fact or artifact?** *American Journal of Political Science*. 29:348-56.
- LEWIS-BECK, M. (1995). **Data analysis.** Thousand Oaks, CA: Sage, Sage University papers series on quantitative applications in the social sciences; n. 07-103.
- LEWIS-BECK, M. (2005). **Election Forecasting: Principles and Practice.** *British Journal of Politics and International Relations*, v. 7, n° 2, p.145-164.
- LEWIS-BECK, M. S.; BÉLANGER, É., & FAUVELLE-AYMAR, C. (2008). **Forecasting the 2007 French presidential election: Ségolène Royel and the Iowa model.** *French Politics*, 6, 106-115.
- LEWIS-BECK, M.; RICE, T. W. (1982). **Presidential popularity and presidential vote.** *The Public Opinion Quarterly*, 46, 534-537.

- LEWIS-BECK, M.; RICE, T. (1984). **Forecasting Presidential Elections: A Comparison of Naïve Models.** *Political Behavior*, 6:9–21.
- LEWIS-BECK, M.; RICE, T. (1992). **Forecasting Elections.** Washington, D.C.: CQ Press.
- LEWIS-BECK, M.; TIEN, C. (2012). **Election forecasting for turbulent times.** *PS: Political Science & Politics*, 45,625-629.
- LICIO, Elaine Cristina; RENNÓ, Lucio R.; CASTRO, Henrique Carlos de O. de (2009). **Bolsa família na eleição presidencial de 2006: em busca do elo perdido.** *Opinião Pública*, v. 15, n° 15, p. 31-54.
- MAGALHÃES, P. C.; AGUIAR-CONRARIA, L. & LEWIS-BACK, M. (2012). **Forecasting Spanish elections.** *International Journal of Forecasting*, Elsevier, vol. 28(4), pages 769-776.
- NADEAU, R.; LEWIS-BECK, M, and BÉLANGER, E. (2012). **Economics and Elections Revisited.** *Comparative Political Studies*. First published online on October 30, 2012.
- NICOLAU, Jairo; PEIXOTO, Vitor (2007). **Uma disputa em três tempos: Uma análise das bases municipais das eleições presidenciais de 2006.** *XXXI Encontro Anual da ANPOCS, ST “Partidos e Sistemas Partidários”*.
- NORPOTH, H. (2004). **From Primary to General Election: A Forecast of the Presidential Vote.** *PS: Political Science and Politics*. 37 (October): 737-740.
- PARANHOS, R. (2014). **A Importância da Replicabilidade na Ciência Política: O Caso do SIGOBR.** *Revista Política Hoje*, 22(2), 213-229.
- SILVER, N. (2012). **The signal and the noise: Why so many predictions fail-but some don't.** Penguin.
- SOARES, Gláucio Ary Dillon; TERRON, Sonia Luiza (2009). **Dois Lulas: A geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise).** *Opinião Pública*, v. 15, n. 1, p. 31-54.
- TURGEON, M e RENNÓ, L. (2012). **Forecasting Brazilian Presidential Elections: Solving the N Problem.** *International Journal of Forecasting* 28: 804-812.
- ZUCCO, Cesar (2008). **The President's 'New' Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections.** *Journal of Latin American Studies*, vol. 40, p. 29-49.

(Footnotes)

¹ Ver: <http://aplicacoes.mds.gov.br/sagi-data/misocial/tabelas/mi_social.php>.