



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

CAMILA CRISTINA LOPES

UMA AVALIAÇÃO DO IMPACTO ELEITORAL DO PROGRAMA BOLSA FAMÍLIA

Recife

2017

CAMILA CRISTINA LOPES

UMA AVALIAÇÃO DO IMPACTO ELEITORAL DO PROGRAMA BOLSA FAMÍLIA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco como requisito parcial para a obtenção do título de Mestre em Estatística.

Orientador: Francisco Cribari-Neto

Coorientadora: Tatiene Correia de Souza

Recife

2017

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

L864a Lopes, Camila Cristina
 Uma avaliação do impacto eleitoral do Programa Bolsa Família / Camila
 Cristina Lopes. – 2017.
 81 f.: il., fig., tab.

 Orientador: Francisco Cribari Neto.
 Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN,
 Estatística, Recife, 2017.
 Inclui referências.

 1. Modelo de regressão. 2. Regressão beta. I. Cribari Neto, Francisco
 (orientador). II. Título.

 519.536 CDD (23. ed.) UFPE- MEI 2017-69

CAMILA CRISTINA LOPES

UMA AVALIAÇÃO DO IMPACTO ELEITORAL DO PROGRAMA BOLSA FAMÍLIA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 16 de fevereiro de 2017.

BANCA EXAMINADORA

Prof. Francisco Cribari Neto
UFPE

Prof.^a Patrícia Leone Espinheira Ospina
UFPE

Prof.^a Verônica Maria Cadena Lima
UFBA

*Aos meus pais,
Bernadete e Gilmar,
com amor.*

Agradecimentos

Aos meus pais, Bernadete e Gilmar, por serem meus maiores incentivadores e os primeiros a acreditarem na minha capacidade. É por vocês e graças a vocês que cheguei até aqui.

Ao meu orientador, professor Francisco Cribari-Neto, pela orientação, seriedade e confiança, por todos os ensinamentos e contribuições prestadas. És um exemplo de professor, pesquisador e profissional.

À minha coorientadora, Tatiane Correia de Souza, pela dedicação, paciência, compreensão e por todas as contribuições dadas a esta pesquisa. Fica aqui registrada toda a minha gratidão e admiração.

À minha amada avó, Maria, por toda a fé, amor e torcida pelo meu sucesso.

Ao meu irmão, Lucas, pelo carinho, parceria e pelas alegrias de cada reencontro.

À minha grande, querida e eterna amiga, Elisandra, pelos incríveis momentos vividos, pelo companheirismo, parceria, lealdade e cumplicidade, pelas lágrimas e angústias compartilhadas, por toda a força e apoio que me mantiveram firme em meu propósito e me possibilitaram alcançar essa vitória. Amiga, essa conquista também é tua!

Aos professores do Departamento de Estatística da UFPE, pelos conhecimentos, ensinamentos e contribuições à minha formação acadêmica, pessoal e profissional.

Aos colegas do mestrado, pelos momentos compartilhados.

À CAPES, pelo apoio financeiro.

Aos membros da banca examinadora, pelas sugestões.

*“Que os vossos esforços desafiem as
impossibilidades, lembrai-vos de que as
grandes coisas do homem foram
conquistadas do que parecia impossível.”*

(Charles Chaplin)

Resumo

O modelo de regressão beta foi proposto por [Ferrari e Cribari-Neto \(2004\)](#) e é largamente utilizado na modelagem de taxas e proporções, isto é, tais que a variável de interesse é contínua e restrita ao intervalo $(0, 1)$. Os parâmetros do modelo de regressão beta são interpretáveis em termos da média da resposta, sendo esta relacionada a um preditor linear por meio de uma função de ligação. Mediante a utilização de um modelo de regressão beta, o interesse desta dissertação é avaliar e comparar a magnitude do impacto que os gastos com programas assistenciais (especialmente o Programa Bolsa Família) tiveram sobre a proporção de votos válidos no segundo turno das eleições presidenciais dos anos 2006, 2010 e 2014. Também, objetiva-se obter uma previsão da quantidade de votos perdidos caso o gasto *per capita* com o Bolsa Família tivesse permanecido no nível em que se encontrava na eleição anterior, ou seja, se o gasto *per capita* não houvesse aumentado nos últimos 4 anos. A comparação entre os impactos estimados dos gastos com programas assistenciais sobre a proporção de votos válidos recebidos pelos presidentes eleitos nas três eleições consideradas indica uma diminuição do impacto com o passar do tempo, isto é, a quantidade de votos recebida “graças” aos benefícios do assistencialismo é menor a cada nova eleição. No que tange à previsão de votos perdidos se o gasto *per capita* tivesse permanecido no nível em que estava há 4 anos, as únicas duas situações possíveis de ocorrer foram constatadas: houve o caso em que, apesar da perda estimada ser razoável, o resultado da eleição não sofreria alteração, como também verificou-se o caso em que o resultado final da eleição teria mudado devido à quantidade de votos perdidos em função dos gastos com assistencialismo.

Palavras-chave: Modelo de regressão beta. Eleições presidenciais. Programa Bolsa Família.

Abstract

The beta regression model was proposed by [Ferrari e Cribari-Neto \(2004\)](#) and it is used for modelling rates and proportions or, more generally, when the variable of interest is continuous and restricted to the interval $(0, 1)$. The parameters of the beta regression model are interpretable in terms of the mean response, which is related to a linear predictor through a link function. Using a beta regression model, the interest of this work is to evaluate and compare the magnitudes of the impact of assistance programs spending (especially the *Bolsa Família* Program) on the proportion of valid votes in the second round of the 2006, 2010 and 2014 presidential elections. Additionally, we obtain an estimate of the number of lost votes if the per capita spending on *Bolsa Família* had remained at the level at the previous election, that is, if per capita spending had not increased in the last 4 years. The comparison between the estimated impacts of the spending on assistance programs on the proportion of valid votes for the elected presidents in the three elections considered indicates a reduced impact over time, i.e., the number of votes received due to the benefits of assistance programs is lowered at each new election. About the forecast of lost votes if per capita spending had remained at the level it was 4 years ago, the only two possible situations were found: there was the case where the estimated loss was reasonable, but the result of the election did not change and the case where the final result of the election would have changed because of the number of lost votes if spending with assistance programs had not increased.

Keywords: Beta regression model. Presidential elections. *Bolsa Família* Program.

Lista de ilustrações

Figura 1 – Histograma e <i>boxplot</i> da proporção de votos válidos do ex-presidente Lula no segundo turno da eleição presidencial de 2006.	34
Figura 2 – Histograma e <i>boxplot</i> da proporção de votos válidos da ex-presidente Dilma no segundo turno da eleição presidencial de 2010.	35
Figura 3 – Histograma e <i>boxplot</i> da proporção de votos válidos da ex-presidente Dilma no segundo turno da eleição presidencial de 2014.	37
Figura 4 – Gráficos das medidas de alavancagem <i>versus</i> os valores preditos do modelo para o ano 2006 e das distâncias de Cook <i>versus</i> os valores preditos.	43
Figura 5 – Gráficos dos resíduos ponderados padronizados 2 <i>versus</i> os índices das observações e dos valores observados <i>versus</i> os valores preditos do modelo para o ano 2006.	44
Figura 6 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2006.	45
Figura 7 – Gráficos das medidas de alavancagem <i>versus</i> os valores preditos do modelo para o ano 2010 e das distâncias de Cook <i>versus</i> os valores preditos.	46
Figura 8 – Gráficos dos resíduos ponderados padronizados 2 <i>versus</i> os índices das observações e dos valores observados <i>versus</i> os valores preditos do modelo para o ano 2010.	47
Figura 9 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2010.	47
Figura 10 – Gráficos das medidas de alavancagem <i>versus</i> os valores preditos do modelo para o ano 2014 e das distâncias de Cook <i>versus</i> os valores preditos.	49
Figura 11 – Gráficos dos resíduos ponderados padronizados 2 <i>versus</i> os índices das observações e dos valores observados <i>versus</i> os valores preditos do modelo para o ano 2014.	49
Figura 12 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2014.	50
Figura 13 – Impactos estimados do gasto <i>per capita</i> nos programas assistenciais em 2006 sobre a resposta média.	55
Figura 14 – Impactos estimados do gasto <i>per capita</i> nos programas assistenciais em 2006 sobre a resposta média fixando <i>analf.</i>	56
Figura 15 – Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2010 sobre a resposta média quando $dnord = 0$	64

Figura 16 – Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2010 sobre a resposta média quando $dnord = 1$	65
Figura 17 – Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2014 sobre a resposta média.	69
Figura 18 – Impactos estimados do gasto <i>per capita</i> nos programas assistenciais em 2014 sobre a resposta média fixando $panalf2010$	70
Figura 19 – Impactos estimados do gasto <i>per capita</i> nos programas assistenciais sobre a resposta média quando $dnord=0$ e $DU=0$	74
Figura 20 – Impactos estimados do gasto <i>per capita</i> nos programas assistenciais sobre a resposta média quando $dnord=1$ e $DU=1$	75

Lista de tabelas

Tabela 1	– Descrição das variáveis consideradas no ajuste dos modelos.	31
Tabela 2	– Estatísticas descritivas das variáveis consideradas no ajuste dos modelos.	32
Tabela 3	– Estimativas pontuais dos parâmetros do modelo para o ano 2006 e p -valores correspondentes.	34
Tabela 4	– Estimativas pontuais dos parâmetros do modelo para o ano 2010 e p -valores correspondentes.	36
Tabela 5	– Estimativas pontuais dos parâmetros do modelo para o ano 2014 e p -valores correspondentes.	38
Tabela 6	– Variações percentuais sobre as estimativas dos parâmetros excluindo as observações influentes.	44
Tabela 7	– Impactos estimados do gasto <i>per capita</i> com assistencialismo em 2006 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.	53
Tabela 8	– Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto <i>per capita</i> nos programas assistenciais em 2006 sobre a resposta média.	53
Tabela 9	– Impactos estimados do gasto <i>per capita</i> com assistencialismo em 2006 sobre a resposta média fixando <i>analf</i> e seus respectivos erros-padrão obtidos via bootstrap.	54
Tabela 10	– Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto <i>per capita</i> nos programas assistenciais em 2006 sobre a resposta média fixando <i>analf</i>	57
Tabela 11	– Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2010 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.	60
Tabela 12	– Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto <i>per capita</i> no programa Bolsa Família em 2010 sobre a resposta média.	61
Tabela 13	– Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2014 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.	67
Tabela 14	– Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto <i>per capita</i> no programa Bolsa Família em 2014 sobre a resposta média.	67

Tabela 15 – Impactos estimados do gasto <i>per capita</i> no programa Bolsa Família em 2014 sobre a resposta média fixando <i>panalf2010</i> e seus respectivos erros-padrão obtidos via bootstrap.	71
Tabela 16 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto <i>per capita</i> no programa Bolsa Família em 2014 sobre a resposta média fixando <i>panalf2010</i>	71
Tabela 17 – Resultados das eleições, reais e estimados, para os anos 2006, 2010 e 2014.	76

Sumário

1	INTRODUÇÃO	15
1.1	Programa Bolsa Família	15
2	MODELO DE REGRESSÃO BETA	23
3	ESPECIFICAÇÃO DOS MODELOS	30
3.1	Introdução	30
3.2	Descrição dos dados	30
3.3	Especificação do modelo de regressão beta para a eleição presiden- cial de 2006	33
3.4	Especificação do modelo de regressão beta para a eleição presiden- cial de 2010	35
3.5	Especificação do modelo de regressão beta para a eleição presiden- cial de 2014	36
4	ANÁLISE DE DIAGNÓSTICO E ADEQUABILIDADE	39
4.1	Introdução	39
4.2	Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2006	42
4.3	Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2010	45
4.4	Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2014	48
5	IMPACTOS ESTIMADOS	51
5.1	Introdução	51
5.2	Impactos estimados do modelo de regressão beta para a eleição presidencial de 2006	52
5.3	Impactos estimados do modelo de regressão beta para a eleição presidencial de 2010	58
5.4	Impactos estimados do modelo de regressão beta para a eleição presidencial de 2014	66
5.5	Comparação entre as eleições presidenciais de 2006, 2010 e 2014 .	72
6	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	77

REFERÊNCIAS 79

1 Introdução

1.1 Programa Bolsa Família

As primeiras iniciativas de implantação de Programas de Transferência de Renda no Brasil datam do início dos anos 1990 com o Projeto de Lei do Senado n° 80, de 1991, de autoria do senador Eduardo Suplicy. No Projeto, o senador propôs o Programa de Garantia de Renda Mínima (PGRM), o qual objetivava amparar integralmente os brasileiros maiores de 25 anos residentes no país e detentores de renda inferior a 2.25 salários mínimos, cerca de 45 mil cruzeiros na época. Contudo, somente a partir de 1996 é que se registraram experiências em nível nacional, iniciadas pelo Programa de Erradicação do Trabalho Infantil (PETI), com a pretensão de incentivar a frequência escolar ao abolir o trabalho infantil por meio de uma transferência em dinheiro às famílias carentes e pelo Benefício de Prestação Continuada (BPC), oferecido às pessoas idosas com mais de 65 anos e aos portadores de deficiência, incapacitados para o trabalho e para a vida independente, com renda familiar mensal *per capita* inferior a $\frac{1}{4}$ do salário mínimo.

Sequencialmente, foi idealizado o Programa Bolsa Escola, vinculado ao Ministério da Educação e instituído pela Lei n° 10.219, de 2001, no governo do então presidente Fernando Henrique Cardoso, filiado ao Partido da Social Democracia Brasileira (PSDB), marcando o início da consolidação dos programas de transferência de renda no país. O Bolsa Escola surgiu com o propósito de estimular a permanência de crianças carentes na escola, via pagamento de R\$ 15 mensais por criança às famílias com renda *per capita* inferior a $\frac{1}{2}$ salário mínimo (R\$ 90), auxílio limitado a três crianças entre 6 e 15 anos por unidade familiar. Para garantir o recebimento do recurso, elas deveriam estar matriculadas no ensino fundamental regular de alguma instituição de ensino e apresentar frequência escolar igual ou superior a 85%.

A partir de então, outros programas passaram a ser implementados, como por exemplo o Bolsa Alimentação, estabelecido pela Medida Provisória n° 2.206, de 2001, vinculado ao Ministério da Saúde, cujo objetivo consistia em promover as condições de saúde e nutrição de gestantes, nutrizes (mães em fase de amamentação com filhos de até 6 meses de idade) e crianças de 6 meses a 6 anos e 11 meses de idade, em risco nutricional, nas mesmas condições de renda e limite de beneficiários por família do Programa Bolsa Escola. No ano seguinte, em 2002, encerrando a gama de programas sociais propostos no governo Fernando Henrique Cardoso, foi regulamentado o Programa Auxílio-Gás por meio do Decreto n° 4.102, com vistas a subsidiar o preço do botijão de gás junto às famílias de baixa renda integrantes dos programas Bolsa Escola ou Bolsa Alimentação do Governo Federal. O Auxílio-Gás foi vinculado ao Ministério de Minas e Energia. O repasse

individual mensal era de R\$ 7.50, pago bimestralmente.

Em 2003, delineou-se uma nova fase na história dos programas sociais federais com a posse de Luiz Inácio Lula da Silva, filiado ao Partido dos Trabalhadores (PT), na presidência do Brasil. O combate à fome e a redução da pobreza e da desigualdade se configuraram como eixos de várias ações durante os anos de governo do presidente (2003/2010), que instituiu uma série de programas voltados a esse fim. Durante o primeiro mandato de Lula, o grande destaque recaiu sobre o Programa Fome Zero, cujo propósito era garantir a segurança alimentar e nutricional¹ de toda a população brasileira por meio de uma política permanente. De acordo com informações extraídas de [Silva, Grossi e França \(2010\)](#), as principais operações a serem realizadas para atingir esse objetivo compreendiam o barateamento da alimentação por meio de convênios com supermercados, cooperativas de consumo, canais alternativos de comercialização, restaurantes populares, entre outros; o aumento da oferta de alimentos básicos mediante o apoio à agricultura familiar, incentivo e produção para autoconsumo e políticas agrícolas; a melhoria na renda via políticas de emprego e renda, microcrédito, previdência social universal, reforma agrária e programas Bolsa Escola e Renda Mínima; e, por fim, um conjunto de ações específicas, tais como cupom de alimentos, cesta básica emergencial, merenda escolar, estoques de segurança e combate à desnutrição materno-infantil.

Uma das primeiras medidas do Fome Zero foi a implantação do Cartão Alimentação através do Programa Nacional de Acesso à Alimentação (PNAA), firmado pela Lei nº 10.689, de 2003. O Cartão Alimentação foi vinculado ao Ministério Extraordinário de Segurança Alimentar e Combate à Fome (MESA) e foi projetado como cartão unificado para o recebimento dos benefícios financeiros do PNAA, concedidos por transferência monetária no valor de R\$ 50 por família com rendimento mensal *per capita* inferior a $\frac{1}{2}$ salário mínimo, bem como de outros programas de transferência de renda.

Conforme descrito, são notáveis os avanços conquistados em relação aos programas sociais de combate à fome e à pobreza no país, contudo, por estarem sob responsabilidade de diferentes ministérios e atuarem de forma independente um do outro, alguns problemas começaram a emergir e foram detectados pelo Relatório de Governo de Transição sobre os Programas Sociais, de 2002. [Silva \(2004\)](#) pontua tais problemas, nos quais se enquadram a existência de programas concorrentes com consequente sobreposição de objetivos e público alvo, o desperdício de recursos causado pela ausência de uma coordenação geral, a ausência de planejamento gerencial dos programas, a insuficiência de orçamentos alocados e o não alcance do público alvo conforme os critérios de elegibilidade dos programas.

Nesse contexto, sancionado pela Lei nº 10.836, de 2004, nasceu o Programa Bolsa Família (PBF), em vigor até os dias atuais, com a proposta de unificar os Programas de

¹ Entende-se por segurança alimentar e nutricional o acesso a uma alimentação diária, em quantidade e qualidade satisfatórias.

Transferência de Renda brasileiros, originalmente abrangendo os programas Bolsa Escola, Bolsa Alimentação, Auxílio-Gás e Cartão Alimentação. Gerenciado pelo Ministério do Desenvolvimento Social e Combate à Fome (MDS), que foi criado em 2004 em substituição ao MESA, o Bolsa Família atua em prol da superação da pobreza mediante três eixos, conforme informações de [Brasil \(2015\)](#): transferência direta de renda às famílias; ampliação do acesso a serviços públicos que representam direitos básicos nas áreas de Saúde, Educação e Assistência; articulação com outras ações e com outros programas dos governos, nas suas três esferas, e da sociedade.

Segundo a determinação legal, para garantir o benefício, as famílias selecionadas devem cumprir determinadas condicionalidades, fixadas nas áreas de Saúde e Educação, que englobam consultas de exame pré-natal, acompanhamento nutricional e de saúde, matrícula e frequência escolar mínima de 85% em estabelecimento regular de ensino para crianças e adolescentes com idade entre 6 e 15 anos e 75% para jovens de 16 e 17 anos.

Para se beneficiarem do recurso, as famílias devem estar inscritas no Cadastro Único para Programas Sociais do Governo Federal, instituído pelo Decreto nº 3.877, de 2001. Popularmente conhecido como CadÚnico, este cadastro está sob responsabilidade do MDS e foi concebido como uma ferramenta a ser utilizada por todos os órgãos públicos federais para a concessão de programas focalizados do governo, dado que ele armazena informações sobre as famílias brasileiras de baixa renda (receptoras de até $\frac{1}{2}$ salário mínimo *per capita*), esmiuçando dados sobre seus membros, renda, condições de moradia, entre outros. A responsabilidade pelo cadastramento das famílias e a atualização dos dados está a cargo dos municípios. Todavia, o processo de seleção para a concessão do auxílio é automatizado e a preferência recai sobre as famílias que apresentam menor renda.

O Bolsa Família oferta quatro tipos de benefício: Benefício Básico, Benefício Variável, Benefício Variável Vinculado ao Adolescente (BVJ) e Benefício para a Superação da Extrema Pobreza (BSP). Em conformidade com o Ministério do Desenvolvimento Social e Combate à Fome, o Benefício Básico, no valor de R\$ 85, é destinado às famílias consideradas em situação de extrema pobreza, isto é, com renda mensal *per capita* de até R\$ 85; o Benefício Variável, no valor de R\$ 39, é concedido às famílias em situação de pobreza (renda mensal *per capita* de até R\$ 170) e extrema pobreza constituídas por gestantes, nutrizes e/ou crianças e adolescentes com até 15 anos de idade, havendo o limitante de 5 benefícios por unidade familiar; o BVJ, no valor de R\$ 46, é remetido às famílias com jovens de 16 e 17 anos frequentadores da escola, limitado a dois auxílios por unidade familiar; por fim, o BSP é reservado às famílias que não alcançam o limiar de R\$ 85 *per capita* após o recebimento dos recursos do PBF que lhes são de direito, de modo que ao valor do auxílio é acrescentada a quantia referente ao alcance desse patamar.

Aos beneficiários é permitido o acúmulo de mais de um tipo de benefício, conforme a sua composição familiar. Assim sendo, as diferentes famílias integrantes do Programa

podem receber valores de auxílio distintos, variando de R\$ 85 a R\$ 372, aí excluída a contagem do BSP visto que é calculado caso a caso. As quantias acima referidas foram ajustadas pelo Decreto nº 8.794, de 2016, corrigindo os antigos valores vigentes e são as que vigoram atualmente.

Desde a sua implantação até o ano 2015, o Bolsa Família já atendeu cerca de 142 milhões de beneficiários e destinou um total de aproximadamente 182 bilhões de reais a famílias em situação de pobreza e extrema pobreza. Segundo dados da Matriz de Informação Social² (MI Social) do MDS, no primeiro ano de vigência do programa foram repassados R\$ 3,791,785,038.00 para atender 6,571,839 famílias inscritas. Daí para adiante verifica-se uma tendência de crescimento no total de recursos anuais destinados ao PBF, bem como no número de beneficiados, todavia, em relação a esse último, registrou-se uma pequena redução no ano 2008 (10,557,996 famílias) quando comparado aos dois anos anteriores (11,043,076 e 10,965,810 famílias, 2007 e 2006, respectivamente) e também no ano 2015 (13,936,791 famílias) em comparação a 2014 e 2013 (14,003,441 e 14,086,199 famílias).

Após cumprir dois mandatos, Lula ficou impossibilitado de concorrer a uma nova eleição presidencial. Em 2010, o PT lançou a candidatura de Dilma Vana Rousseff ao cargo máximo do Poder Executivo Federal. Dilma, que até então estava à frente da Casa Civil, concorreu com 8 candidatos e venceu José Serra, do PSDB, em segundo turno, tendo obtido 56.05% dos votos válidos, cerca de 55.7 milhões de votos, tornando-se a primeira mulher Presidente da República Federativa do Brasil. Em 2014, candidatou-se à reeleição e disputou o segundo turno com Aécio Neves da Cunha, filiado ao PSDB, derrotando-o com 51.64% dos votos válidos, aproximadamente 54.5 milhões de votos.

Ao fazer uma comparação entre as eleições de 2010 e 2014, é possível verificar que o resultado de ambas, no que tange ao candidato vencedor por estado, é praticamente igual. Dos 26 estados brasileiros, Dilma foi vitoriosa em 15, os mesmos nas duas eleições, enquanto que os candidatos oponentes, Serra e Aécio, em 2010 e 2014, respectivamente, venceram nos 11 estados restantes. A única discordância ocorreu no Distrito Federal: em 2010 Dilma foi a vencedora e em 2014 a vitória foi de seu concorrente. As urnas indicaram que a presidente obteve aprovação absoluta na região Nordeste, na maior parte da região Norte e em metade da região Sudeste, enquanto o candidato oponente dominou as regiões Sul e Centro-Oeste.

Em seu governo, Dilma deu continuidade às ações sociais de combate à pobreza e desigualdade introduzidas por seu antecessor, além de implantar outros programas. De acordo com a pesquisa divulgada pela Confederação Nacional da Indústria (CNI) em 17

² A Matriz de Informação Social é uma ferramenta de gestão da informação que permite consulta a dados e indicadores gerenciais dos programas sociais do MDS. Os resultados da pesquisa estão disponíveis na forma de planilhas eletrônicas, relatórios, tabelas, gráficos e mapas, listados por municípios ou estados WWP (2015).

de dezembro de 2014, CNI/Ibope, o combate à fome e à pobreza e os investimentos em programas sociais destacaram-se como os pontos fortes do primeiro mandato. De uma situação de elevada popularidade, ante uma diversidade de fatores políticos e econômicos apurados ao longo de sua gestão, a presidente passou a enfrentar uma crise de governo de proporções crescentes, fazendo sua aceitação deteriorar.

Na época das eleições de Lula, levantou-se uma questão polêmica sobre a influência que o investimento em programas sociais exerceu sobre o resultado da votação, destaque para o Bolsa Família. [Marques et al. \(2009\)](#) apontaram que o PBF foi determinante para o resultado positivo do ex-presidente na reeleição ao passo que [Shikida et al. \(2009\)](#) entenderam que o programa não foi decisivo.

[Souza e Cribari-Neto \(2013\)](#) avaliaram os impactos dos programas assistenciais e do crescimento da economia sobre o resultado da eleição presidencial de 2006, chegando à conclusão que se os gastos com programas assistenciais fossem mantidos nos níveis de 2002 haveria uma redução de cerca de 7.116 milhões de votos no presidente e não ocorrendo crescimento econômico a redução seria de 2 milhões de votos, números que não reverteriam o resultado da eleição.

Trabalhos anteriores, e.g. [Zucco \(2008\)](#), indicaram que houve uma mudança na base eleitoral do presidente Lula da eleição de 2002 para a de 2006. Na primeira eleição, o candidato obteve maior votação nas regiões brasileiras mais desenvolvidas, ao passo que na reeleição seu melhor desempenho se deu nas áreas mais pobres do país. No mesmo estudo, o autor assinalou que essa ocorrência pode estar ligada à implantação do Bolsa Família e, ademais, sugere também que existe uma forte tendência das regiões menos desenvolvidas a votarem no candidato a presidente ligado ao governo que já detém o poder. [Zucco \(2010\)](#) reiterou a mutação do caráter eleitoral em Lula, como também a criação do PBF como fator de importância e acrescenta a informação que o padrão de votação observado em 2006 manteve-se em 2010 na eleição da presidente Dilma Rousseff (que deu continuidade ao programa de transferência de renda), fato esse atestado pelos resultados eleitorais e dados provenientes da pesquisa em questão.

O autor ainda salientou que a alteração no eleitorado do ex-presidente não é algo inusitado, visto que parece haver um padrão no qual, independente de quem sejam, os candidatos à presidência do partido no poder são melhor avaliados em locais menos favorecidos e os candidatos da oposição conquistam mais votos onde o nível de desenvolvimento é maior, culminando na percepção que o eleitorado de Lula teria mudado de uma eleição para outra com ou sem a instauração do PBF.

[Pereira et al. \(2015\)](#) levantaram duas hipóteses para a similaridade no cenário de votação entre a reeleição de Lula e a eleição de Dilma. Em primeiro lugar, os fatores econômicos e sociais que garantiram a vitória dele em regiões desfavorecidas (a disseminação dos programas de transferência de renda e o crescimento da economia) configuraram-

se como condições relevantes na eleição dela. Em segundo lugar, o apoio de Lula à candidata Dilma a amparou na votação em geral, acarretando o pressuposto que o novo governo estenderia as ações de seu precursor, o que poderia ter preservado a lealdade do eleitorado petista à margem de qualquer situação econômica ou obtenção de proventos. Em concordância com os resultados da pesquisa, há indícios de que a votação expressiva na presidente na maioria dos municípios do Brasil se deve, em grande parte, à dependência da população aos programas sociais mais do que a constatação de uma efetiva melhora na economia municipal, de forma que, ante ao que foi apurado, o PT parece ter seguido à frente do governo em 2010 graças à ampla magnitude que o PBF atingiu desde sua implantação.

[Almeida e Souza \(2015\)](#) investigaram o impacto dos gastos com o Programa Bolsa Família nas eleições presidenciais de 2010 na região Nordeste e as conclusões retratam que os maiores percentuais de votos em prol da presidente ocorreram nos municípios com os mais elevados gastos com o PBF, com maior proporção de pobres, em que mais pessoas residem na zona rural e com maior Produto Interno Bruto (PIB) *per capita*. Além disso, certificou-se a boa notoriedade de Lula como influência favorável à eleição de Dilma e atestou-se que a candidata petista sairia vitoriosa na região Nordeste ainda que inexistissem os gastos com o PBF, dado que os impactos do gasto *per capita* calculados em dois diferentes cenários revelam reduções de 2.125 e 1.233 milhões de votos, respectivamente, números insuficientes para alterar o resultado da eleição de 2010 na região Nordeste.

Outros trabalhos investigaram e avaliaram os efeitos eleitorais que o assistencialismo exerce sobre as eleições presidenciais no Brasil, tais como os de [Zucco \(2013\)](#), no qual são estimados os efeitos dos programas de transferência de renda nas eleições de 2002, 2006 e 2010 a partir da análise de resultados eleitorais a nível municipal; [Zucco \(2015\)](#), que estendeu o trabalho anterior examinando os impactos eleitorais desses programas na eleição de 2014 e comparando-os aos resultados das três eleições anteriores; [Canêdo-Pinheiro \(2015\)](#), que investigou o papel do Programa Bolsa Família e do desempenho econômico como determinantes dos resultados da eleição presidencial de 2006, principalmente na questão da migração da base eleitoral de Lula para as regiões menos desenvolvidas; e [Corrêa \(2015b\)](#), o qual reavaliou os impactos do Bolsa Família sobre a eleição de 2006, demonstrando que à medida que a cobertura do Programa cresce, o desempenho eleitoral de Lula entre 2002 e 2006 melhora e à medida que o tamanho da classe alta cresce, pior é o desempenho.

Ampliando o contexto de investigação, [Corrêa \(2015a\)](#) analisou dados relativos às eleições presidenciais na América Latina que ocorreram entre 1990 e 2010 e revela que os presidentes latino-americanos que investiram em programas de transferência de renda condicionada não obtiveram melhor desempenho eleitoral do que os presidentes que não investiram nesses programas. Seguindo essa mesma linha de pesquisa, [Corrêa \(2016\)](#)

estendeu o estudo, considerando os dados das eleições presidenciais latino-americanas de 1990 a 2014 e confirma os resultados do trabalho anterior, reforçando que a probabilidade de reeleição dos presidentes não aumenta com investimentos em programas de transferência de renda condicionada. Também nesse cenário, [Pavão \(2016\)](#) avaliou os efeitos dos programas de transferência de renda condicionada na América Latina sobre a responsabilidade pela economia e corrupção dos países e os resultados apontam que, embora tais programas gerem retornos eleitorais para os governos responsáveis por eles, não chegam a encobrir as consequências eleitorais do mau desempenho do governo em termos de economia e corrupção.

Ainda, uma comparação entre o Programa Bolsa Família e a *Asignación Universal por Hijo*, programa de transferência de renda na Argentina, foi feita por [Marques \(2013\)](#) e diferenças expressivas foram detectadas, entre elas a de que, enquanto o Bolsa Família se configura como um programa, a *Asignación Universal por Hijo* se constitui em um direito e paga valores maiores aos seus beneficiados. Contrariamente à grande parte dos trabalhos que estudam os impactos dos programas de transferência de renda, em que os resultados são limitados a curto prazo, [Jones \(2016\)](#) revisou as evidências existentes sobre os impactos a longo prazo para a redução da pobreza.

Diante do exposto, torna-se cabível investigar, a nível nacional, a influência que os investimentos na continuidade dos programas sociais tiveram sobre a votação favorável de Luiz Inácio Lula da Silva e Dilma Rousseff no segundo turno das últimas três eleições presidenciais – 2006, 2010 e 2014 – e avaliar se os resultados das eleições sofreriam alteração caso essas ações não viessem a ser concretizadas. Essa é a linha norteadora da presente dissertação.

Para cumprir tal objetivo, a modelagem foi feita por meio da utilização do modelo de regressão beta, indicado para dados de taxas e proporções, de forma que para cada ano eleitoral considerado um modelo para explicar a proporção de votos válidos no segundo turno da eleição foi ajustado. Para avaliar e comparar a magnitude dos impactos que os gastos com programas assistenciais tiveram sobre a proporção de votos válidos, foram considerados os valores de gasto *per capita* anual com assistencialismo por município brasileiro. Uma análise sobre a votação final de cada eleição foi feita com base na previsão do modelo para a perda de votos caso o gasto *per capita* com o Programa Bolsa Família tivesse permanecido no nível em que se encontrava na eleição anterior, isto é, no nível em que estava há 4 anos.

Todos os resultados aqui apresentados foram obtidos através da linguagem de programação R. O software estatístico é distribuído gratuitamente e pode ser encontrado no endereço <http://www.r-project.org>.

Mais 5 capítulos compõem o trabalho, além desta introdução. No Capítulo 2 é definido o modelo de regressão beta, bem como a forma de estimação dos parâmetros que

indexam este tipo de modelo. Os dados são descritos no Capítulo 3, o qual contempla também a definição das variáveis utilizadas na modelagem e a especificação de cada um dos modelos de regressão ajustados. Para verificar a qualidade do ajuste e a sua boa adequação aos dados, no Capítulo 4 encontra-se a análise de diagnóstico feita através da realização do teste RESET de especificação correta, cálculo de medidas de alavancagem e influência, coeficiente de determinação e utilização de ferramentas gráficas. O Capítulo 5 apresenta os resultados obtidos com o cálculo dos impactos juntamente com uma análise sobre o que eles representam no contexto em que estão inseridos. Finalmente, no Capítulo 6 constam algumas considerações gerais a respeito do trabalho.

2 Modelo de Regressão Beta

O presente capítulo visa definir o modelo de regressão beta, bem como o logaritmo da função de verossimilhança, as funções (vetores) escore úteis no processo de obtenção dos estimadores de máxima verossimilhança do modelo de regressão beta e a matriz de informação de Fisher associada. Assim, dado que as variáveis de interesse na modelagem aqui proposta caracterizam-se como a proporção de votos válidos obtidos no segundo turno pelos presidentes eleitos nas eleições de 2006, 2010 e 2014, deve-se fazer uso de uma distribuição que represente adequadamente esses dados. Logo, a escolha pela distribuição beta parece ser apropriada, dado ser uma distribuição de grande importância na modelagem de variáveis que pertencem ao intervalo unitário $(0, 1)$. Ademais, os modelos de regressão em que a variável resposta segue a distribuição beta acomodam naturalmente dados assimétricos, caso típico quando se trata de taxas e proporções.

Na classe de modelos de regressão beta assume-se que a média da resposta é relacionada a um preditor linear, este envolvendo covariáveis e parâmetros de regressão desconhecidos. Tal relação é feita por meio de uma função de ligação, que pode assumir formas variadas:

- logit: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$;
- probit: $g(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ é a distribuição acumulada de uma variável aleatória normal padrão;
- log-log (loglog): $g(\mu) = -\log[-\log(\mu)]$;
- complemento log-log (cloglog): $g(\mu) = \log[-\log(1 - \mu)]$;
- Cauchy: $g(\mu) = \tan[\pi(\mu - \frac{1}{2})]$.

A escolha da função de ligação a ser utilizada no ajuste é feita, na prática, por aquela que promove o melhor entre os ajustes. Maiores informações e detalhes sobre tais funções de ligação podem ser encontradas em [McCullagh e Nelder \(1989\)](#).

O modelo de regressão beta surgiu a partir de uma reparametrização da densidade beta, sendo que a função densidade de uma variável aleatória y que segue distribuição beta é dada por

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1},$$

em que $0 < y < 1$, $p > 0$, $q > 0$ e $\Gamma(\cdot)$ é a função gama. A depender dos valores escolhidos para p e q , diferentes distribuições no intervalo $(0, 1)$ podem ser obtidas, como é o caso da distribuição uniforme, que ocorre quando $p = q = 1$.

A variável aleatória com distribuição beta tem média e a variância dadas, respectivamente, por:

$$\mathbb{E}(y) = \frac{p}{p+q}$$

e

$$\text{Var}(y) = \frac{pq}{(p+q)^2(p+q+1)}.$$

A reparametrização acima citada foi proposta por [Ferrari e Cribari-Neto \(2004\)](#), de forma que fosse possível modelar a média (μ) da resposta juntamente com um parâmetro de precisão (ϕ). Com essa nova parametrização, tem-se $\mu = \frac{p}{p+q}$ e $\phi = p+q$, o que implica $p = \mu\phi$ e $q = (1-\mu)\phi$. Logo, a média e a variância de y passam a ser escritas como

$$\mathbb{E}(y) = \mu$$

e

$$\text{Var}(y) = \frac{V(\mu)}{1+\phi},$$

onde $V(\mu) = \mu(1-\mu)$ é a função de variância. Nota-se que, para μ fixo, quanto maior o valor de ϕ , menor será a variância de y .

Assim, a função densidade é reescrita como

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1},$$

em que $0 < y < 1$, $0 < \mu < 1$ e $\phi > 0$.

Tal como feito por [Souza \(2011\)](#), neste trabalho admitiu-se um parâmetro de dispersão (σ) na modelagem ao invés do parâmetro de precisão. Desse modo, sendo $\sigma^2 = \frac{1}{1+\phi}$, ou $\phi = \frac{1-\sigma^2}{\sigma^2}$, a densidade passa a ser

$$f(y; \mu, \sigma) = \frac{\Gamma\left(\frac{1-\sigma^2}{\sigma^2}\right)}{\Gamma\left(\mu\frac{1-\sigma^2}{\sigma^2}\right)\Gamma\left((1-\mu)\frac{1-\sigma^2}{\sigma^2}\right)} y^{\mu\left(\frac{1-\sigma^2}{\sigma^2}\right)-1} (1-y)^{(1-\mu)\left(\frac{1-\sigma^2}{\sigma^2}\right)-1}, \quad (2.1)$$

em que $0 < y < 1$, $0 < \mu < 1$ e $0 < \sigma < 1$.

Ademais, admitiu-se também que o parâmetro de dispersão é variável. Assim, torna-se necessário modelar σ por meio de uma estrutura de regressão envolvendo covariáveis, parâmetros desconhecidos e uma função de ligação, similarmente ao modo que é feito para

a média da resposta. Esta extensão do modelo original, denominada modelo de regressão beta com dispersão variável, foi apresentada por [Simas, Barreto-Souza e Rocha \(2010\)](#).

Sejam $y_1, \dots, y_n$ variáveis aleatórias independentes, tal que cada y_t , $t = 1, \dots, n$, segue a densidade em (2.1) com média μ_t e dispersão σ_t , desconhecidas. As estruturas de regressão para a média e para a dispersão de y_t são definidas, respectivamente, por

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = \eta_t$$

e

$$h(\sigma_t) = \sum_{j=1}^q z_{tj}\gamma_j = v_t,$$

em que $\beta = (\beta_1, \dots, \beta_k)^T \in \mathbb{R}^k$ e $\gamma = (\gamma_1, \dots, \gamma_q)^T \in \mathbb{R}^q$ são vetores de parâmetros desconhecidos, $x_{t1}, \dots, x_{tk}$ e $z_{t1}, \dots, z_{tq}$ são observações de k e q covariáveis ($k + q < n$), respectivamente, assumidas fixas e conhecidas, $\eta_t = \sum_{i=1}^k x_{ti}\beta_i$ e $v_t = \sum_{j=1}^q z_{tj}\gamma_j$ são os preditores lineares e $g(\cdot)$ e $h(\cdot)$ são funções de ligação estritamente monótonas e duas vezes diferenciáveis, em que $g : (0, 1) \rightarrow \mathbb{R}$ e $h : (0, 1) \rightarrow \mathbb{R}$. A escolha dessas funções normalmente ocorre entre as cinco citadas anteriormente, podendo ser a mesma para os dois submodelos ou, então, uma função diferente para cada estrutura de regressão. Por fim, tem-se que $\mu_t = g^{-1}(\eta_t)$, $\sigma_t = h^{-1}(v_t)$ e $\text{Var}(y) = \mu_t(1 - \mu_t)\sigma_t^2$.

A estimação dos parâmetros que indexam a distribuição é feita mediante a maximização do logaritmo da função de verossimilhança. Para uma amostra de n observações independentes, a log-verossimilhança decorrente da densidade (2.1) é dada por

$$\ell(\beta, \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \sigma_t),$$

em que

$$\begin{aligned} \ell_t(\mu_t, \sigma_t) = & \log \Gamma\left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right) - \log \Gamma\left(\mu_t \frac{1 - \sigma_t^2}{\sigma_t^2}\right) - \log \Gamma\left((1 - \mu_t) \frac{1 - \sigma_t^2}{\sigma_t^2}\right) \\ & + \left(\mu_t \frac{1 - \sigma_t^2}{\sigma_t^2} - 1\right) \log y_t + \left((1 - \mu_t) \frac{1 - \sigma_t^2}{\sigma_t^2} - 1\right) \log(1 - y_t). \end{aligned}$$

A derivação da função de log-verossimilhança em relação a cada um dos parâmetros resulta na função escore de β e γ . Assim sendo, a função escore para β_i , $i = 1, \dots, k$, é dada por

$$\frac{\partial \ell(\beta, \gamma)}{\partial \beta_i} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i}, \quad (2.2)$$

em que $\frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \mu_t} = \frac{1 - \sigma_t^2}{\sigma_t^2} (y_t^* - \mu_t^*)$, $\frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)}$ e $\frac{\partial \eta_t}{\partial \beta_i} = x_{ti}$. Aqui, y_t^* e μ_t^* são definidos como $y_t^* = \log\left(\frac{y_t}{1 - y_t}\right)$ e $\mu_t^* = \psi\left(\mu_t \frac{1 - \sigma_t^2}{\sigma_t^2}\right) - \psi\left((1 - \mu_t) \frac{1 - \sigma_t^2}{\sigma_t^2}\right)$, em que $\psi(\cdot)$ é a função digama

definida por $\psi(u) = \frac{d \log \Gamma(u)}{du}$ e $\Gamma(u) = \int_0^\infty t^{u-1} e^{-t} dt$ é a função gama, para $u > 0$. A função escore pode assim ser reescrita como

$$\frac{\partial \ell(\beta, \gamma)}{\partial \beta_i} = \sum_{t=1}^n \frac{1 - \sigma_t^2}{\sigma_t^2} (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} x_{ti}.$$

Matricialmente, o vetor escore $U_\beta(\beta, \gamma)$ assume a forma

$$U_\beta(\beta, \gamma) = X^\top \Sigma T (y^* - \mu^*),$$

em que X é uma matriz $n \times k$ cuja t -ésima linha é $x_t^\top = (x_{t1}, \dots, x_{tk})$, $\Sigma = \text{diag} \left\{ \frac{1 - \sigma_1^2}{\sigma_1^2}, \dots, \frac{1 - \sigma_n^2}{\sigma_n^2} \right\}$, $T = \text{diag} \left\{ \frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right\}$, $y^* = (y_1^*, \dots, y_n^*)^\top$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$.

Similarmente, a função escore para γ_j , $j = 1, \dots, q$, é dada por

$$\frac{\partial \ell(\beta, \gamma)}{\partial \gamma_j} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t} \frac{d\sigma_t}{dv_t} \frac{\partial v_t}{\partial \gamma_j}, \quad (2.3)$$

em que $\frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t} = a_t = -\frac{2}{\sigma_t^3} \left[\mu_t (y_t^* - \mu_t^*) + \log(1 - y_t) + \psi \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) - \psi \left((1 - \mu_t) \frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right]$, $\frac{d\sigma_t}{dv_t} = \frac{1}{h'(\sigma_t)}$ e $\frac{\partial v_t}{\partial \gamma_j} = z_{tj}$. Logo, a função escore é reescrita como

$$\frac{\partial \ell(\beta, \gamma)}{\partial \gamma_j} = \sum_{t=1}^n a_t \frac{1}{h'(\sigma_t)} z_{tj}.$$

O vetor escore $U_\gamma(\beta, \gamma)$ em forma matricial é:

$$U_\gamma(\beta, \gamma) = Z^\top H a,$$

em que Z é uma matriz $n \times q$ cuja t -ésima linha é $z_t^\top = (z_{t1}, \dots, z_{tq})$, $H = \text{diag} \left\{ \frac{1}{h'(\sigma_1)}, \dots, \frac{1}{h'(\sigma_n)} \right\}$ e $a = (a_1, \dots, a_n)^\top$.

Os estimadores de máxima verossimilhança do modelo de regressão beta são obtidos com a resolução do sistema

$$\begin{cases} U_\beta(\beta, \gamma) = 0, \\ U_\gamma(\beta, \gamma) = 0. \end{cases}$$

Todavia, os estimadores de β e γ não possuem forma fechada e, portanto, devem ser obtidos numericamente via algum algoritmo de otimização não-linear, sendo o método quasi-Newton BFGS¹ o mais usado entre eles.

A matriz de informação de Fisher é obtida a partir do cálculo da esperança das segundas derivadas do logaritmo da função de verossimilhança. Desse modo, partindo de (2.2), a segunda derivada da log-verossimilhança em relação a β_i e β_p , $i, p = 1, \dots, k$, é dada por

¹ Maiores informações a respeito deste método podem ser encontradas em [Press et al. \(1992\)](#).

$$\begin{aligned}\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \beta_p} &= \sum_{t=1}^n \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_p} x_{ti} \\ &= \sum_{t=1}^n \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} + \frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{1}{g'(\mu_t)} x_{ti} x_{tp}.\end{aligned}$$

Sabendo que $\mathbb{E} \left(\frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \mu_t} \right) = 0$, tem-se que

$$\begin{aligned}\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \beta_p} \right) &= \sum_{t=1}^n \mathbb{E} \left[\left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t^2} \frac{1}{g'(\mu_t)} \right) \frac{1}{g'(\mu_t)} x_{ti} x_{tp} \right] \\ &= \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t^2} \left(\frac{1}{g'(\mu_t)} \right)^2 x_{ti} x_{tp} \right),\end{aligned}$$

em que $\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t^2} = - \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right)^2 \left[\psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) + \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right]$ e $\psi'(\cdot)$ é a função trigama definida por $\psi'(u) = \frac{d^2 \log \Gamma(u)}{du^2}$, para $u > 0$.

Portanto,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \beta_p} \right) = - \sum_{t=1}^n \frac{1-\sigma_t^2}{\sigma_t^2} w_t x_{ti} x_{tp},$$

em que $w_t = \frac{1-\sigma_t^2}{\sigma_t^2} \left[\psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) + \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right] \left(\frac{1}{g'(\mu_t)} \right)^2$. Ou, em forma matricial,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \beta_p} \right) = -X^\top \Sigma W X$$

em que $W = \text{diag}\{w_1, \dots, w_n\}$.

Ademais, a segunda derivada da função de log-verossimilhança em relação a β_i e γ_j , $i = 1, \dots, k$ e $j = 1, \dots, q$, é dada por

$$\begin{aligned}\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \gamma_j} &= \sum_{t=1}^n \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t \partial \sigma_t} \frac{d\sigma_t}{dv_t} \frac{\partial v_t}{\partial \gamma_j} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i} \\ &= \sum_{t=1}^n \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t \partial \sigma_t} \frac{1}{h'(\sigma_t)} z_{tj} \right) \frac{1}{g'(\mu_t)} x_{ti},\end{aligned}$$

em que $\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \mu_t \partial \sigma_t} = -\frac{2}{\sigma_t^3} (y_t^* - \mu_t^*) + \frac{2}{\sigma_t^3} \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) \left[\mu_t \psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) - (1-\mu_t) \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right]$.

Portanto,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \gamma_j} \right) = - \sum_{t=1}^n c_t \frac{1}{g'(\mu_t)} \frac{1}{h'(\sigma_t)} x_{ti} z_{tj},$$

em que $c_t = \frac{2}{\sigma_t^3} \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) \left[\mu_t \psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) - (1-\mu_t) \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right]$. Em forma matricial,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \beta_i \partial \gamma_j} \right) = -X^\top CTHZ,$$

em que $C = \text{diag}\{c_1, \dots, c_n\}$.

Em relação a γ_j e γ_l , $j, l = 1, \dots, q$, a segunda derivada da função de log-verossimilhança é dada por:

$$\begin{aligned} \frac{\partial^2 \ell(\beta, \gamma)}{\partial \gamma_j \partial \gamma_l} &= \sum_{t=1}^n \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t^2} \frac{d\sigma_t}{dv_t} + \frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t} \frac{\partial}{\partial \sigma_t} \frac{d\sigma_t}{dv_t} \right) \frac{d\sigma_t}{dv_t} \frac{\partial v_t}{\partial \gamma_l} \frac{\partial v_t}{\partial \gamma_j} \\ &= \sum_{t=1}^n \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t^2} \frac{1}{h'(\sigma_t)} + \frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t} \frac{\partial}{\partial \sigma_t} \frac{d\sigma_t}{dv_t} \right) \frac{1}{h'(\sigma_t)} z_{tl} z_{tj}. \end{aligned}$$

Sabendo que $\mathbb{E} \left(\frac{\partial \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t} \right) = 0$, tem-se que

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \gamma_j \partial \gamma_l} \right) = \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t^2} \right) \left(\frac{1}{h'(\sigma_t)} \right)^2 z_{tl} z_{tj},$$

$$\begin{aligned} \text{em que } \frac{\partial^2 \ell_t(\mu_t, \sigma_t)}{\partial \sigma_t^2} &= -\frac{4}{\sigma_t^6} \left[-\psi' \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) + \mu_t^2 \psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) + (1-\mu_t)^2 \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right] \\ &\quad + \frac{6}{\sigma_t^4} \left[\mu_t (y_t^* - \mu_t^*) + \log(1-y_t) + \psi \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) - \psi \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right]. \end{aligned}$$

Portanto,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \gamma_j \partial \gamma_l} \right) = -\sum_{t=1}^n d_t z_{tl} z_{tj},$$

em que $d_t = \frac{4}{\sigma_t^6} \left[-\psi' \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) + \mu_t^2 \psi' \left(\mu_t \frac{1-\sigma_t^2}{\sigma_t^2} \right) + (1-\mu_t)^2 \psi' \left((1-\mu_t) \frac{1-\sigma_t^2}{\sigma_t^2} \right) \right] \left(\frac{1}{h'(\sigma_t)} \right)^2$. Ou, em forma matricial,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\beta, \gamma)}{\partial \gamma_j \partial \gamma_l} \right) = -Z^\top DZ,$$

em que $D = \text{diag}\{d_1, \dots, d_n\}$.

Logo, dado o vetor paramétrico $\theta = (\beta^\top, \gamma^\top)^\top$, a matriz de informação de Fisher conjunta assume a seguinte forma:

$$K(\theta) = K(\beta, \gamma) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\gamma} \\ K_{\gamma\beta} & K_{\gamma\gamma} \end{pmatrix},$$

em que $K_{\beta\beta} = X^\top \Sigma W X$, $K_{\beta\gamma} = (K_{\gamma\beta})^\top = X^\top CTHZ$ e $K_{\gamma\gamma} = Z^\top DZ$. Por conseguinte, a inversa da matriz de informação de Fisher é dada por

$$K(\theta)^{-1} = K(\beta, \gamma)^{-1} = \begin{pmatrix} K_*^{\beta\beta} & K_*^{\beta\gamma} \\ K_*^{\gamma\beta} & K_*^{\gamma\gamma} \end{pmatrix},$$

em que $K_*^{\beta\beta} = (X^\top \Sigma W X - X^\top CTHZ(Z^\top DZ)^{-1}Z^\top HTCX)^{-1}$,

$$K_*^{\beta\gamma} = (K_*^{\gamma\beta})^\top = -K_*^{\beta\beta} X^\top CTHZ(Z^\top DZ)^{-1} \text{ e}$$

$$K_*^{\gamma\gamma} = (Z^\top DZ)^{-1}[I_q + (Z^\top HTCX)K_*^{\beta\beta}X^\top CTHZ(Z^\top DZ)^{-1}].$$

Para grandes tamanhos de amostra e sob certas condições de regularidade, a distribuição conjunta de $\hat{\beta}$ e $\hat{\gamma}$ é aproximadamente normal $(k + q)$ -multivariada:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \beta \\ \gamma \end{pmatrix}, K(\theta)^{-1} \right),$$

aproximadamente, em que $\hat{\beta}$ e $\hat{\gamma}$ são os estimadores de máxima verossimilhança de β e γ , respectivamente.

3 Especificação dos Modelos

3.1 Introdução

Este capítulo objetiva definir cada um dos três modelos empíricos referentes às eleições presidenciais tomadas em estudo. Por se tratar de dados de proporção, o modelo de regressão beta foi utilizado para o ajuste dos dados e posterior obtenção dos impactos estimados que os gastos *per capita* com assistencialismo exerceram sobre a proporção de votos válidos no segundo turno das eleições dos anos 2006, 2010 e 2014.

Todas as regressões foram estimadas utilizando o software estatístico R¹, por meio do pacote `gamlss`. A classe de modelos de regressão denominada Modelos Aditivos Generalizados para Posição, Escala e Forma (*Generalized Additive Models for Location, Scale and Shape* – GAMLSS), proposta por [Rigby e Stasinopoulos \(2005\)](#), corrige algumas limitações dos Modelos Lineares Generalizados (*Generalized Linear Models* – GLM) e dos Modelos Aditivos Generalizados (*Generalized Additive Models* – GAM). Essa classe caracteriza-se por ser semiparamétrica e permitir o ajuste de modelos em que a variável resposta não segue distribuição pertencente à família exponencial, bem como possibilita a modelagem de todos os parâmetros da distribuição condicional da variável dependente e não apenas da média². Sendo assim, além da média da resposta, modelou-se também a dispersão.

O processo de seleção das covariáveis ocorreu mediante o uso da função *stepGAIC* disponível no pacote `gamlss` e, também, minuciosa investigação manual. A ferramenta *stepGAIC* emprega o método *stepwise* e utiliza o Critério de Informação de Akaike Generalizado (*Generalized Akaike Information Criterion* – GAIC) para a seleção de variáveis. Este critério foi apresentado por [Takeuchi \(1976\)](#) que generalizou o Critério de Informação de Akaike (*Akaike's Information Criterion* – AIC), proposto por [Akaike \(1974\)](#). A função *stepGAIC* também permite escolher entre os métodos *forward* ou *backward* e oferece, ainda, a opção do Critério de Informação Bayesiano (*Bayesian Information Criterion* – BIC) para o processo de seleção de variáveis.

3.2 Descrição dos dados

As variáveis utilizadas no presente estudo, bem como suas respectivas descrições encontram-se na Tabela 1. Os dados foram obtidos por meio dos bancos de dados dis-

¹ Para maiores informações sobre o software R, consultar [Venables e Smith \(2016\)](#).

² Maiores detalhes relacionados à classe de modelos de regressão GAMLSS podem ser obtidos em [Stasinopoulos, Rigby e Akantziliotou \(2008\)](#).

ponibilizados pelo Tribunal Superior Eleitoral (TSE), Ministério do Desenvolvimento Social e Combate à Fome (MDS), Atlas do Desenvolvimento Humano no Brasil e Instituto Brasileiro de Geografia e Estatística (IBGE). Vale salientar que os dados referentes ao modelo ajustado para a eleição do ano 2006 foram obtidos de Souza (2011), trabalho em que a autora avaliou o impacto dos programas assistenciais e do crescimento da economia sobre o resultado da eleição presidencial de 2006.

Tabela 1 – Descrição das variáveis consideradas no ajuste dos modelos.

Variável	Descrição
<i>eleicoes2014</i>	Proporção de votos válidos de Dilma no segundo turno da eleição de 2014
<i>gastopc2014</i>	Gasto <i>per capita</i> com o PBF em 2014
<i>eleicoes2010</i>	Proporção de votos válidos de Dilma no segundo turno da eleição de 2010
<i>gastopc2010</i>	Gasto <i>per capita</i> com o PBF em 2010
<i>rur2010</i>	População rural do município em 2010
<i>panalf2010</i>	Taxa de analfabetismo entre pessoas com 15 anos ou mais em 2010
<i>ppent2010</i>	Percentual da população evangélica de origem pentecostal em 2010
<i>rendapc2010</i>	Renda <i>per capita</i> do município em 2010
<i>govpt2010</i>	Variável dummy: 1 se o governador era do PT em 2010, 0 caso contrário
<i>dnord</i>	Variável dummy: 1 se o município é do nordeste, 0 caso contrário
<i>dpb</i>	Variável dummy: 1 se o município é da Paraíba, 0 caso contrário
<i>lula2006</i>	Proporção de votos válidos de Lula no segundo turno da eleição de 2006
<i>gastopc2006</i>	Gasto <i>per capita</i> com os programas assistenciais em 2006
<i>analf</i>	Proporção de analfabetos entre pessoas acima de 15 anos em 2000
<i>renda</i>	Renda <i>per capita</i> do município em 2000
<i>DU</i>	Variável dummy: 1 se população 2006 > 50,000 habitantes, 0 caso contrário

A fim de caracterizar as variáveis acima mencionadas, na Tabela 2 podem ser visualizadas as correspondentes estatísticas descritivas calculadas tomando-se 5,560 municípios brasileiros para os anos 2010 e 2014 e 4,864 municípios para o ano 2006. É possível verificar que o valor médio do gasto *per capita* aumentou com o tempo. Por exemplo, em 2006 o valor registrado foi R\$ 64.08, passando a R\$ 111.40 em 2010 e R\$ 201.50 em 2014, o que representou um aumento médio de R\$ 137.42 *per capita* em um período de 8 anos. Da mesma forma, a renda *per capita* também aumentou à medida que os anos avançaram. Para notar tal efeito, basta avaliar as variáveis *renda* e *rendapc2010*, sendo que em 75% dos municípios brasileiros a renda era menor ou igual a R\$ 235.00 em 2000 e passou a R\$ 650.40 em 2010.

Em contrapartida, constatou-se que o percentual de analfabetismo entre pessoas com 15 anos ou mais diminuiu com o tempo, de modo que a taxa média registrada na variável *analf* foi de 21.38% enquanto que em *panalf2010* o percentual baixou para 16.17%. Ademais, tem-se que em 50% dos municípios estudados a votação dos presidentes eleitos no segundo turno das eleições foi cerca de 61% em 2006, 58% em 2010 e 57% em 2014.

Entre os municípios brasileiros avaliados, o maior percentual de votos válidos

Tabela 2 – Estatísticas descritivas das variáveis consideradas no ajuste dos modelos.

Variável	Mínimo	Q_1	Mediana	Média	Q_3	Máximo
<i>eleicoes2014</i>	0.1186	0.4376	0.5692	0.5775	0.7316	0.9393
<i>gastopc2014</i>	1.58	79.24	162.20	201.50	311.90	832.90
<i>eleicoes2010</i>	0.1967	0.4757	0.5818	0.5949	0.7143	0.9650
<i>gastopc2010</i>	0.25	49.42	99.64	111.40	166.90	3850.00
<i> rur2010</i>	0	1600	3234	5348	6768	125300
<i>panalf2010</i>	0.00950	0.08088	0.13120	0.16170	0.24320	0.44400
<i>ppent2010</i>	0.0005	0.0595	0.0979	0.1099	0.1480	0.5211
<i>rendapc2010</i>	96.25	281.00	467.40	493.10	650.40	2044.00
<i>lula2006</i>	0.1960	0.4861	0.6143	0.6178	0.7574	0.9626
<i>gastopc2006</i>	2.40	31.24	54.22	64.08	95.06	384.20
<i>analf</i>	0.00907	0.11490	0.17570	0.21380	0.31500	0.60660
<i>renda</i>	28.38	88.78	163.90	173.70	235.00	954.60

recebidos por Lula no segundo turno da eleição de 2006 ocorreu no município de Belágua (MA), 96.26%, enquanto a menor votação deu-se em São Marcos (RS), 19.60%. Em 2010, verificou-se que o maior e o menor percentuais de votos que Dilma recebeu no segundo turno da eleição foram em Calumbi (PE), 96.50%, e Barra Longa (MG), 19.67%, respectivamente. Na última eleição presidencial, em 2014, a maior votação em Dilma ocorreu novamente no município de Belágua (MA), 93.93%, ao passo que o menor registro foi de Nova Pádua (RS), 11.86%.

Em relação aos gastos *per capita*, verificou-se que os maiores valores foram registrados nos municípios de Campos Verdes (GO) em 2006, R\$ 384.20, Santa Rita (PB) em 2010, R\$ 3850.00, e Junco do Maranhão (MA) em 2014, R\$ 832.90. Os menores gastos pertencem aos municípios de Águas de São Pedro (SP), R\$ 2.40, Monte Belo do Sul (RS), R\$ 0.25, e Alto Feliz (RS), R\$ 1.58, em 2006, 2010 e 2014, respectivamente.

Considerando os dois casos da variável taxa de analfabetismo, *analf* e *panalf2010*, averiguou-se que os maiores percentuais foram dos municípios de Jordão (AC) e Alagoinha do Piauí (PI), 60.66% e 44.40%, respectivamente, enquanto as menores taxas foram atingidas em São João do Oeste (SC) e Feliz (RS), 0.91% e 0.95%. Se a variável renda *per capita* for tomada para avaliação, tem-se que os maiores valores foram registrados em Águas de São Pedro (SP), R\$ 954.60, e São Caetano do Sul (SP), R\$ 2044.00, para *renda* e *rendapc2010* correspondentes aos anos 2000 e 2010. As menores rendas ocorreram em Centro do Guilherme (MA), R\$ 28.38, e Marajá do Sena (MA), R\$ 96.25, respectivamente.

Diante das informações acima citadas, cabe salientar que os maiores percentuais de votos válidos recebidos pelos ex-presidentes, Lula e Dilma, no segundo turno das três eleições estudadas ocorreram de forma absoluta em municípios pertencentes à região Nordeste, ao passo que as menores votações concentraram-se nas regiões Sul e Sudeste.

Comportamento similar foi observado com os valores do gasto *per capita* com o Bolsa Família, em que, com exceção do ano 2006, os maiores gastos ocorreram na região Nordeste e os menores deram-se nas regiões Sul e Sudeste.

As regiões Norte e Nordeste apareceram com as maiores taxas de analfabetismo, tendo sido registrada no Nordeste a menor renda, ao passo que a região Sul apareceu com o menor percentual de analfabetos e a região Sudeste com a maior renda *per capita* do país. Este cenário sugere que as regiões contempladas com os maiores recursos provenientes dos programas assistenciais, que apresentaram menor renda *per capita* e maior taxa de analfabetismo tiveram influência direta sobre o resultado final das eleições presidenciais em estudo.

3.3 Especificação do modelo de regressão beta para a eleição presidencial de 2006

Definindo a variável de interesse (y) como a proporção de votos válidos recebidos pelo ex-presidente Lula no segundo turno da eleição presidencial de 2006, a análise descritiva indica que o valor médio na amostra é 0.6178, ao passo que a mediana se encontra em 0.6143 e os valores mínimo e máximo assumidos por y são 0.1960 e 0.9626, respectivamente. A fim de detectar algumas características iniciais dos dados e da possível distribuição da variável resposta, construiu-se o histograma e o *boxplot* dos dados, os quais podem ser visualizados na Figura 1. Os gráficos indicam uma leve assimetria da distribuição, que deve ser levada em consideração no ajuste do modelo.

As covariáveis selecionadas para explicar a proporção de votos válidos do ex-presidente Lula no segundo turno da eleição presidencial de 2006 foram: *analf*, *renda*, *DU* e *gastopc2006*, de forma que o modelo escolhido assume a forma:

$$\text{logit}(\mu_t) = \beta_0 + \beta_1 \text{analf}_t + \beta_2 \text{renda}_t + \beta_3 \text{DU}_t + \beta_4 \text{gastopc2006}_t + \beta_5 \text{gastopc2006}_t \times \text{analf}_t \\ + \beta_6 \text{gastopc2006}_t \times \text{renda}_t,$$

$$\text{probit}(\sigma_t) = \gamma_0 + \gamma_1 \text{analf}_t + \gamma_2 \text{renda}_t, \quad (3.1)$$

para $t = 1, 2, \dots, 4864$.

As funções de ligação *logit* e *probit*, para a média e dispersão, respectivamente, foram utilizadas por terem proporcionado o ajuste mais satisfatório. Os coeficientes estimados das covariáveis do modelo e os correspondentes p -valores podem ser consultados na Tabela 3, na qual é possível averiguar que o efeito positivo sobre a proporção de votos válidos do

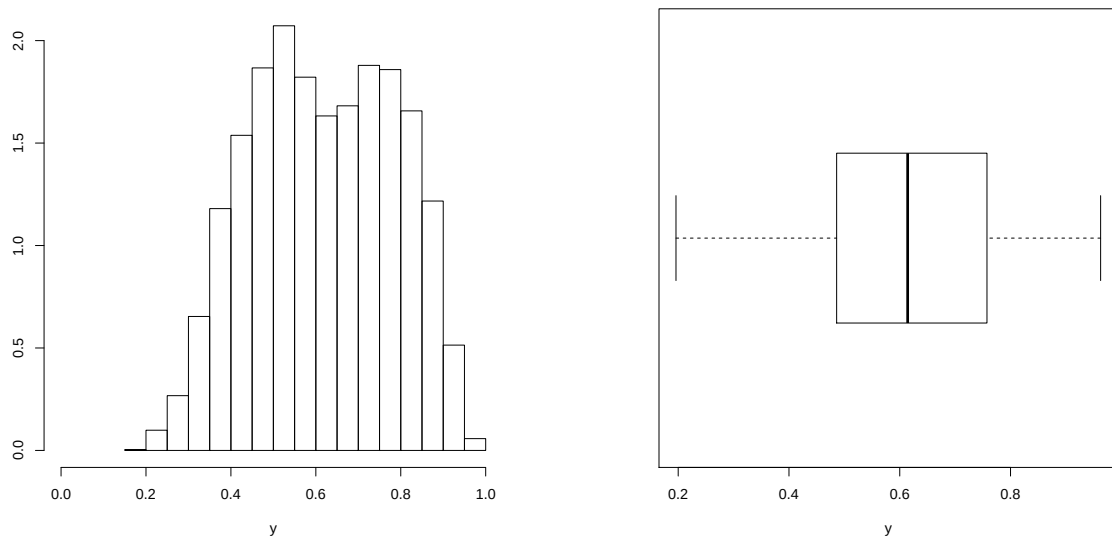


Figura 1 – Histograma e *boxplot* da proporção de votos válidos do ex-presidente Lula no segundo turno da eleição presidencial de 2006.

presidente eleito no segundo turno da eleição de 2006 é proveniente da variável independente *DU*, ou seja, municípios com população maior que 50,000 habitantes registraram maior proporção de votos válidos em Lula na eleição de 2006. Todavia, diante da presença de interações no modelo, o mesmo não pode ser afirmado em relação às covariáveis *analf* e *gastopc2006*, bem como *renda*, visto que o efeito de interação mascara o efeito principal de cada um desses regressores. Para fins de exemplificação, tomando a interação *gastopc2006* $\times$ *renda*, tem-se que o efeito de *gastopc2006* sobre a proporção média de votos válidos em Lula depende do nível de *renda* e, analogamente, o efeito de *renda* sobre a resposta média depende do nível de *gastopc2006*. A mesma interpretação vale para a interação *gastopc2006* $\times$ *analf*.

Tabela 3 – Estimativas pontuais dos parâmetros do modelo para o ano 2006 e *p*-valores correspondentes.

	$\hat{\beta}$	<i>p</i> -valor	$\hat{\gamma}$	<i>p</i> -valor
<i>intercepto</i>	−0.202500	0.003530	−0.548100	0.000000
<i>analf</i>	2.067000	0.000000	−0.334000	0.000105
<i>renda</i>	−0.000997	0.000000	−0.000656	0.000000
<i>DU</i>	0.470400	0.000000	—	—
<i>gastopc2006</i>	0.015530	0.000000	—	—
<i>gastopc2006</i> $\times$ <i>analf</i>	−0.017150	0.000000	—	—
<i>gastopc2006</i> $\times$ <i>renda</i>	−0.000035	0.000000	—	—

A modelagem do parâmetro de dispersão indica que maiores valores de renda *per*

capita e taxas de analfabetismo mais elevadas implicam menor dispersão na proporção de votos válidos no presidente eleito no ano considerado. Ademais, todas as variáveis mostram-se significativas ao nível de significância de 1%.

3.4 Especificação do modelo de regressão beta para a eleição presidencial de 2010

A variável de interesse (y), neste caso, é definida como a proporção de votos válidos recebidos pela ex-presidente Dilma no segundo turno da eleição presidencial de 2010. O valor médio na amostra é 0.5949, enquanto a mediana de y vale 0.5818 e os valores mínimo e máximo são 0.1967 e 0.9650, respectivamente. O histograma e o *boxplot* (Figura 2) mostram que a distribuição da resposta é assimétrica, o que corrobora com a escolha da regressão beta para a modelagem.

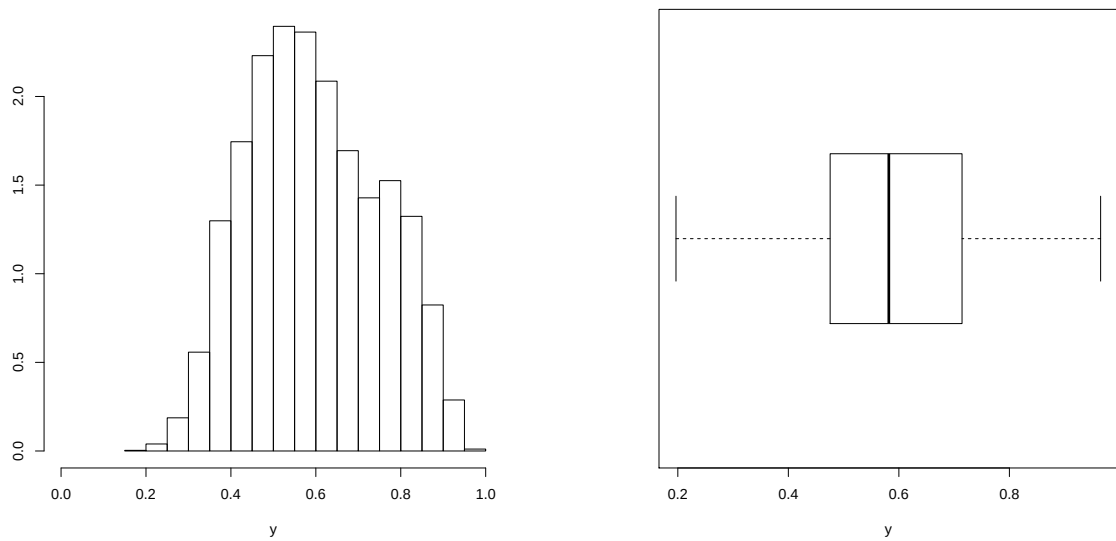


Figura 2 – Histograma e *boxplot* da proporção de votos válidos da ex-presidente Dilma no segundo turno da eleição presidencial de 2010.

As covariáveis selecionadas para explicar a proporção de votos válidos da ex-presidente Dilma no segundo turno da eleição presidencial de 2010 foram: *rendapc2010*, *govpt2010*, *ppent2010*, *gastopc2010*, *dnord* e *dpb*. O modelo escolhido assume a forma:

$$\begin{aligned} \text{cloglog}(\mu_t) &= \beta_0 + \beta_1 \text{rendapc2010}_t + \beta_2 \text{govpt2010}_t + \beta_3 \text{ppent2010}_t + \beta_4 \text{gastopc2010}_t + \beta_5 \text{dnord}_t \\ &\quad + \beta_6 \text{dpb}_t + \beta_7 \text{gastopc2010}_t \times \text{rendapc2010}_t + \beta_8 \text{gastopc2010}_t \times \text{dnord}_t, \\ \text{probit}(\sigma_t) &= \gamma_0 + \gamma_1 \text{rendapc2010}_t + \gamma_2 \text{ppent2010}_t + \gamma_3 \text{dnord}_t, \end{aligned} \quad (3.2)$$

para $t = 1, 2, \dots, 5560$.

As funções de ligação *cloglog* e *probit*, para a média e dispersão, respectivamente, asseguraram o melhor ajuste e, por isso, foram escolhidas. Os coeficientes estimados das covariáveis do modelo e os correspondentes p -valores podem ser consultados na Tabela 4. Verificou-se que as covariáveis *govpt2010*, *ppent2010* e *dpb* operam negativamente sobre a proporção de votos válidos da presidente eleita no segundo turno da eleição de 2010, enquanto que o efeito de interação entre *rendapc2010* e *dnord* com *gastopc2010* não permite qualquer afirmativa nesse sentido, dado que os efeitos de *rendapc2010* e *dnord* sobre a proporção média de votos válidos obtidos por Dilma dependem do nível de *gastopc2010* e vice-versa.

Tomando as estimativas resultantes do ajuste feito para o parâmetro de dispersão, notou-se a influência positiva das covariáveis *ppent2010* e *dnord* sobre a dispersão de y . Isto significa que municípios pertencentes à região Nordeste e com maior percentual de evangélicos pentecostais favorecem resposta mais dispersa, diferentemente do que ocorre com a *rendapc2010* que minimiza a dispersão à medida que seu valor aumenta. A análise dos p -valores revela que todas as variáveis são significativas ao nível nominal de 0.1%.

Tabela 4 – Estimativas pontuais dos parâmetros do modelo para o ano 2010 e p -valores correspondentes.

	$\hat{\beta}$	p -valor	$\hat{\gamma}$	p -valor
<i>intercepto</i>	-0.347500	0.000000	-0.765800	0.000000
<i>rendapc2010</i>	-0.000128	0.000142	-0.000230	0.000000
<i>govpt2010</i>	-0.183900	0.000000	—	—
<i>ppent2010</i>	-0.634600	0.000000	0.597400	0.000000
<i>gastopc2010</i>	0.004887	0.000000	—	—
<i>dnord</i>	0.756900	0.000000	0.072370	0.000068
<i>dpb</i>	-0.257000	0.000000	—	—
<i>gastopc2010</i> $\times$ <i>rendapc2010</i>	-0.000004	0.000000	—	—
<i>gastopc2010</i> $\times$ <i>dnord</i>	-0.003695	0.000000	—	—

3.5 Especificação do modelo de regressão beta para a eleição presidencial de 2014

Tomando a variável de interesse como a proporção de votos válidos recebidos pela ex-presidente Dilma no segundo turno da eleição presidencial de 2014, as estatísticas descritivas mostram que o valor médio de tal proporção é 0.5775, a mediana se estabelece em 0.5692 e os valores mínimo e máximo de y ficam em 0.1186 e 0.9393, respectivamente. A observação do histograma e do *boxplot* apresentados na Figura 3 sugere uma distribuição

assimétrica para a variável resposta, sendo a assimetria uma característica da distribuição beta utilizada no ajuste do modelo.

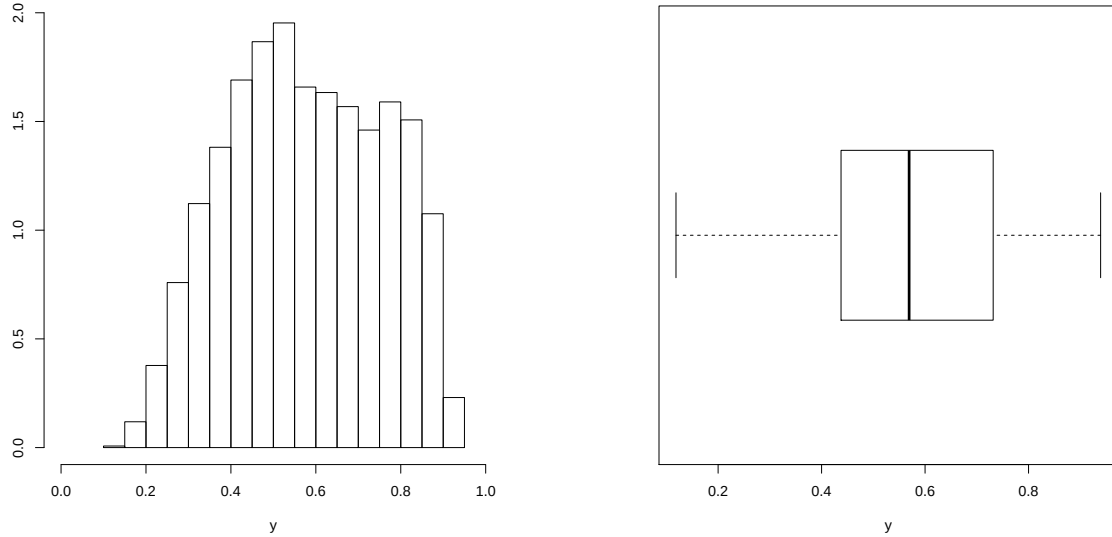


Figura 3 – Histograma e *boxplot* da proporção de votos válidos da ex-presidente Dilma no segundo turno da eleição presidencial de 2014.

As covariáveis selecionadas para explicar a proporção de votos válidos recebidos pela ex-presidente Dilma no segundo turno da eleição presidencial de 2014 foram: *rur2010*, *panalf2010*, *ppent2010*, *dnord*, *gastopc2014* e *rendapc2010*, o modelo selecionado tendo sido:

$$\begin{aligned} \text{logit}(\mu_t) = & \beta_0 + \beta_1 \text{rur2010}_t + \beta_2 \text{panalf2010}_t + \beta_3 \text{ppent2010}_t + \beta_4 \text{dnord}_t + \beta_5 \text{gastopc2014}_t \\ & + \beta_6 \text{rendapc2010}_t + \beta_7 \text{panalf2010}_t \times \text{ppent2010}_t + \beta_8 \text{panalf2010}_t \times \text{gastopc2014}_t, \end{aligned}$$

$$\text{probit}(\sigma_t) = \gamma_0 + \gamma_1 \text{ppent2010}_t + \gamma_2 \text{gastopc2014}_t, \quad (3.3)$$

para $t = 1, 2, \dots, 5560$.

As funções de ligação *logit* e *probit*, para a média e dispersão, respectivamente, foram selecionadas por terem conduzido ao melhor ajuste. Os coeficientes estimados das covariáveis do modelo e os correspondentes *p*-valores podem ser consultados na Tabela 5. De acordo com os resultados, as covariáveis *rur2010* e *dnord* influenciam positivamente a proporção de votos válidos da presidente eleita no segundo turno da eleição de 2014, o que indica que nos municípios com maior população rural e pertencentes à região Nordeste a proporção de votos válidos em Dilma tende a ser maior, contrariamente ao que ocorre com *rendapc2010* que exerce efeito negativo sobre *y*. As covariáveis *panalf2010*, *ppent2010* e

gastopc2014 estão sob o efeito de interação, o que significa que os efeitos de *ppent2010* e *gastopc2014* sobre a resposta média dependem do nível de *panalf2010*, bem como o efeito de *panalf2010* depende do nível de *ppent2010* e *gastopc2014*.

Em relação ao parâmetro de dispersão, um maior percentual da população evangélica de origem pentecostal acarreta resposta mais dispersa, ao passo que maiores valores de gasto *per capita* com programas assistenciais reduzem a dispersão. Conforme indicam os *p*-valores, as variáveis são significativas ao nível de significância de 5%.

Tabela 5 – Estimativas pontuais dos parâmetros do modelo para o ano 2014 e *p*-valores correspondentes.

	$\hat{\beta}$	<i>p</i> -valor	$\hat{\gamma}$	<i>p</i> -valor
<i>intercepto</i>	−0.238500	0.000142	−0.863100	0.000000
<i>rur2010</i>	0.000005	0.000002	—	—
<i>panalf2010</i>	2.161000	0.000000	—	—
<i>ppent2010</i>	−0.423200	0.023300	0.455300	0.000010
<i>dnord</i>	0.374700	0.000000	—	—
<i>gastopc2014</i>	0.003915	0.000000	−0.000123	0.007260
<i>rendapc2010</i>	−0.000530	0.000000	—	—
<i>panalf2010</i> × <i>ppent2010</i>	−5.123000	0.000007	—	—
<i>panalf2010</i> × <i>gastopc2014</i>	−0.006739	0.000000	—	—

4 Análise de Diagnóstico e Adequabilidade

4.1 Introdução

A fim de verificar a qualidade dos ajustes e garantir a boa representação dos dados, este capítulo descreve a análise de diagnóstico e adequabilidade, juntamente com as representações gráficas, realizada para cada um dos modelos de regressão beta estimados. Paralelamente ao ajuste de um modelo de regressão, é de extrema importância investigar sua adequação aos dados, ou em outras palavras, buscar eventuais afastamentos das suposições firmadas para o modelo. A análise de resíduos constitui uma técnica amplamente utilizada e foi pioneira em se tratando de análise de diagnóstico, sendo que tanto resíduos ordinários como versões padronizadas podem ser empregadas. Os resíduos medem a discrepância entre os valores verdadeiros dos dados e os valores estimados, gerados pelo modelo de regressão proposto. Sendo assim, o principal objetivo do uso dessa metodologia é identificar a presença de pontos discrepantes e verificar se há violação dos pressupostos inerentes ao modelo.

No contexto da classe de modelos de regressão beta, o resíduo ordinário padronizado foi introduzido por [Ferrari e Cribari-Neto \(2004\)](#), sendo dado por

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{\text{Var}}(y_t)}},$$

onde $\hat{\mu}_t = g^{-1}(x_t^\top \hat{\beta})$ e $\widehat{\text{Var}}(y_t) = \frac{\hat{\mu}_t(1-\hat{\mu}_t)}{1+\hat{\phi}}$, em que $\hat{\phi} = \frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}$ e $\hat{\sigma}_t^2 = h^{-1}(z_t^\top \hat{\gamma})$. Aqui, $\hat{\beta}$ e $\hat{\gamma}$ são os estimadores de máxima verossimilhança de β e γ , respectivamente.

Não obstante, [Espinheira, Ferrari e Cribari-Neto \(2008b\)](#) chegaram a uma nova definição de resíduo que se mostrou superior ao acima citado, principalmente na detecção de observações influentes. Os autores o nomearam resíduo ponderado padronizado 2 e assume a forma:

$$r_t^{ww} = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\hat{v}_t(1 - h_{tt}^*)}},$$

onde $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ e, de acordo com a parametrização adotada para o parâmetro de precisão, $\hat{\mu}_t^* = \psi\left(\hat{\mu}_t \frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}\right) - \psi\left((1-\hat{\mu}_t) \frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}\right)$, tal que $\psi(\cdot)$ é a função digama. Ademais, $\hat{v}_t = \psi'\left(\hat{\mu}_t \frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}\right) + \psi'\left((1-\hat{\mu}_t) \frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}\right)$, onde $\psi'(\cdot)$ é a função trigama, e h_{tt}^* é o t -ésimo elemento diagonal de $(\widehat{W}\widehat{\Sigma})^{\frac{1}{2}}X(X^\top\widehat{\Sigma}\widehat{W}X)^{-1}X^\top(\widehat{\Sigma}\widehat{W})^{\frac{1}{2}}$, em que $\widehat{\Sigma} = \text{diag}\left\{\frac{1-\hat{\sigma}_1^2}{\hat{\sigma}_1^2}, \dots, \frac{1-\hat{\sigma}_n^2}{\hat{\sigma}_n^2}\right\}$ e $\widehat{W} = \text{diag}\{\hat{w}_1, \dots, \hat{w}_n\}$, sendo que os elementos $\hat{w}_t$ são definidos por $\hat{w}_t = \left(\frac{1-\hat{\sigma}_t^2}{\hat{\sigma}_t^2}\right) \hat{v}_t \left[\frac{1}{g'(\hat{\mu}_t)}\right]^2$. As

funções digama e trigama mencionadas são dadas por: $\psi(z) = \frac{d \log \Gamma(z)}{dz}$ e $\psi'(z) = \frac{d^2 \log \Gamma(z)}{dz^2}$, com $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$, para $z > 0$, respectivamente.

Esse foi o resíduo utilizado no desenvolvimento da presente pesquisa. O gráfico dos resíduos ponderados padronizados 2 *versus* o índice das observações é bastante útil para verificar a adequação do ajuste do modelo. Para ser considerado um bom ajuste, a distribuição dos pontos no gráfico não deve apresentar padrão algum, isto é, os pontos devem estar dispostos aleatoriamente em torno de zero.

A medida h_{tt}^* é conhecida como alavancagem generalizada, tendo sido proposta inicialmente por [Wei, Hu e Fung \(1998\)](#) podendo ser entendida como uma medida que avalia a importância individual de cada observação. O cálculo da alavancagem generalizada para o modelo de regressão beta foi apresentado por [Ferrari e Cribari-Neto \(2004\)](#) e, posteriormente, [Ferrari, Espinheira e Cribari-Neto \(2011\)](#) estenderam a ideia para modelos de regressão beta com dispersão variável. A análise gráfica da alavancagem permite a identificação de pontos sujeitos à distorção das estimativas dos parâmetros, os quais podem comprometer a qualidade do ajuste.

O grau de influência das observações é determinado por meio da distância de Cook, uma ferramenta de diagnóstico que consiste na retirada individual de cada observação do conjunto de dados para mensurar o impacto exercido por ela sobre as estimativas. Essa medida foi proposta por [Cook \(1977\)](#) e sua versão para os modelos de regressão beta foi dada por [Espinheira, Ferrari e Cribari-Neto \(2008a\)](#):

$$LD_t = \frac{h_{tt}^*}{1 - h_{tt}^*} (r_t^{ww})^2.$$

Observações que apresentam valor elevado de LD_t são consideradas influentes e passíveis de investigação. O gráfico das distâncias de Cook *versus* os valores preditos pelo modelo pode ser usado para identificar pontos potencialmente influentes.

O gráfico dos valores observados *versus* os valores preditos e o gráfico de probabilidade meio-normal com envelopes simulados aparecem para complementar a análise e assegurar a validade do modelo proposto. Para que isso se confirme, os pontos devem ficar distribuídos em torno da reta bissetriz (45°) no contexto do primeiro gráfico e os resíduos devem permanecer dentro das bandas de confiança no segundo. Esse último é uma ferramenta gráfica de bastante utilidade na análise de diagnóstico, visto que a distribuição dos resíduos não é conhecida e por meio dele pode-se decidir se as respostas observadas são consistentes com o modelo estimado.

Gráficos de probabilidade meio-normal são assim chamados por considerarem os resíduos em valor absoluto no seu processo de construção. A ideia de adicionar envelopes simulados surgiu de [Atkinson \(1985\)](#) como uma forma de possibilitar a melhor interpretação do gráfico meio-normal habitual. Para traçar o gráfico com os envelopes simulados, alguns

passos devem ser seguidos, a saber:

- (1) Ajustar o modelo com os dados originais e gerar uma amostra simulada de n observações independentes usando o modelo estimado como se fosse o modelo verdadeiro;
- (2) Ajustar o modelo com a amostra gerada e calcular os valores absolutos ordenados dos resíduos;
- (3) Repetir os passos (1) e (2) r vezes;
- (4) Tomar os r conjuntos de resíduos ordenados e, então, calcular a média, valor mínimo e valor máximo de cada conjunto;
- (5) Construir o gráfico desses valores e dos resíduos ordenados da amostra de dados original contra os escores meio-normais $\Phi^{-1}\left(\frac{t+n-1/8}{2n+1/2}\right)$, em que $\Phi(\cdot)$ é a função de distribuição acumulada normal padrão.

O envelope é produzido pelos valores mínimos e máximos dos r conjuntos de resíduos ordenados. As observações correspondentes aos resíduos absolutos que ficarem fora das bandas de confiança estabelecidas devem ser investigadas com maior atenção. A não adequação do modelo ajustado é certificada quando uma quantidade considerável de pontos estiver fora do envelope simulado.

A correta especificação do modelo pode ser verificada por meio do Teste de Erro de Especificação da Regressão (*Regression Specification Error Test – RESET*), originalmente proposto por Ramsey (1969) para modelos de regressão linear. Esse teste identifica a presença de dois tipos de erros de especificação, variável omitida e forma funcional incorreta, e admite a hipótese nula de que o modelo está corretamente especificado contra a hipótese alternativa de que há um erro de especificação no modelo. O teste RESET para regressão beta foi apresentado por Lima (2007) e consiste em adicionar à regressão uma nova variável de teste ao submodelo da média, a sugestão dada é que seja o preditor linear estimado elevado à segunda potência, $\hat{\eta}^2$. A hipótese nula $H_0 : \beta_{k+1} = 0$, em que β é o parâmetro referente à variável de teste $\hat{\eta}^2$, é testada mediante a exclusão do termo adicional via algum teste assintótico, neste trabalho a razão de verossimilhanças foi empregada e os três modelos propostos revelaram-se corretamente especificados ao nível de significância de 5%.

A qualidade do ajuste é avaliada pelo cálculo do coeficiente de determinação, conhecido como R^2 e representa a proporção da variabilidade dos dados que é explicada pelo modelo estimado. Entretanto, esta medida é válida apenas no contexto dos modelos de regressão linear e, diante de tal limitação, Ferrari e Cribari-Neto (2004) propuseram o coeficiente de determinação ajustado para modelos de regressão beta. O pseudo- R^2 (R_p^2), como foi denominado, é uma medida análoga ao R^2 e é definido como o quadrado

do coeficiente de correlação entre $\hat{\eta}$ e $g(y)$. Cabe ressaltar que o valor deste coeficiente restringe-se ao intervalo $0 \leq R_p^2 \leq 1$ e quanto mais próximo de 1 ele estiver, melhor será a capacidade de explicação da variabilidade total de y pelo modelo. Ademais, calculou-se também o pseudo- R_{RV}^2 , que é baseado na razão de verossimilhanças e foi apresentado por Nagelkerke (1991), sua expressão sendo dada por

$$R_{RV}^2 = 1 - \left(\frac{L_{null}}{L_{fit}} \right)^{\frac{2}{n}},$$

em que L_{null} e L_{fit} denotam as funções de verossimilhança maximizadas do modelo sem regressores e do modelo ajustado, respectivamente.

4.2 Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2006

O pseudo- R^2 obtido para o modelo ajustado para o ano 2006 foi de 0.57, o que significa que as variáveis independentes explicam cerca de 57% da variabilidade total da variável resposta. Similarmente, o pseudo- R_{RV}^2 foi de 0.58.

Os gráficos de diagnóstico construídos para confirmar a validade e a adequação do modelo aos dados podem ser observados nas Figuras 4, 5 e 6. Na Figura 4 é apresentado o gráfico de alavancagem generalizada e das distâncias de Cook, por meio dos quais se destacam algumas observações mais afastadas das demais. Em relação à alavancagem, o caso mais extremo compreende a observação 58, identificada como o município de Águas de São Pedro (SP). Tal como citado no capítulo anterior, o município paulista registrou a maior renda *per capita* em 2000 e, conforme salientado por Souza (2011), também o menor valor de gasto em programas assistenciais em 2006. As demais observações evidenciadas, 4032, 2900, 3910 e 4010, correspondem aos municípios de São Caetano do Sul (SP), Niterói (RJ), Santana de Parnaíba (SP) e Santos (SP), respectivamente.

A fim de averiguar o efeito que as observações de maiores alavancagens exercem sobre as estimativas dos parâmetros, o modelo foi ajustado sem o ponto 58 e, ainda, sem os pontos 58 e 4032, os dois mais discrepantes. As estimativas obtidas com tais ajustes não apresentaram variações percentuais muito elevadas, sendo que a maior diferença nos dois casos ocorreu no intercepto do submodelo da média, que ficou cerca de 5.14% menor sem o ponto 58 e 5.68% menor ao serem excluídos os dois pontos de alavanca.

Os três pontos sinalizados no gráfico das distâncias de Cook – 2900, 3910 e 4834 – representam os municípios de Niterói (RJ), Santana de Parnaíba (SP) e Vitória (ES), os quais foram excluídos individual e conjuntamente para investigação da influência exercida por eles sobre as estimativas dos parâmetros.

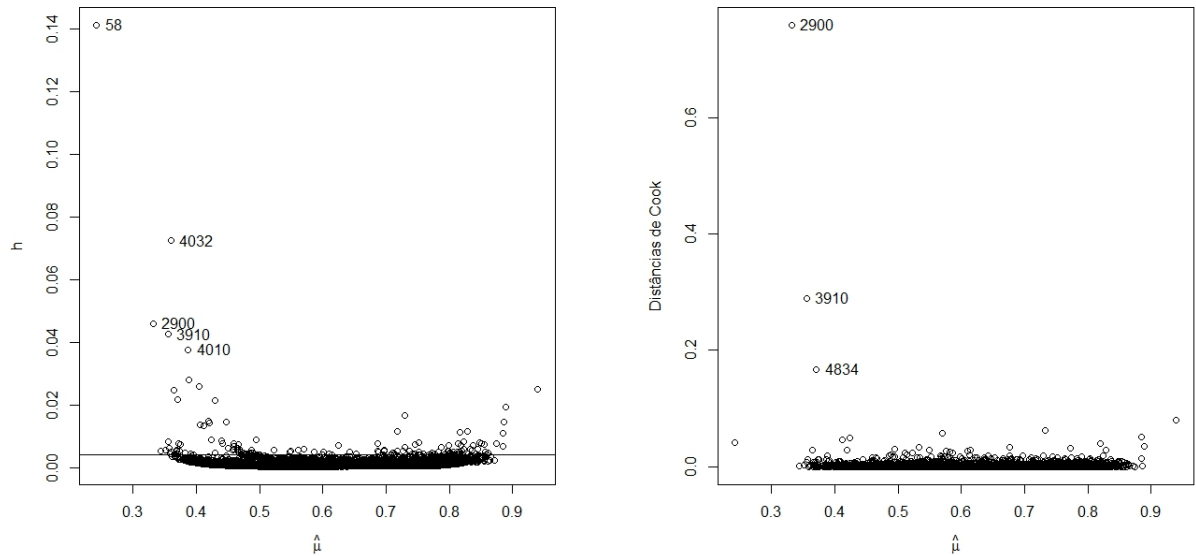


Figura 4 – Gráficos das medidas de alavancagem *versus* os valores preditos do modelo para o ano 2006 e das distâncias de Cook *versus* os valores preditos.

A Tabela 6 apresenta as variações percentuais obtidas com os ajustes realizados após a remoção dos pontos em destaque dos dados. Atesta-se que o maior impacto da exclusão da observação 2900 no submodelo da média ocorre em $\hat{\beta}_0$, correspondente ao intercepto, que sofre aumento de cerca de 14.62%. Em relação à dispersão, o ajuste sem tal observação provoca redução de aproximadamente 21.11% em $\hat{\gamma}_1$ (taxa de analfabetismo) e 20.27% em $\hat{\gamma}_2$ (renda *per capita*), implicando diminuição na dispersão das respostas.

Ao se excluir o município de Santana de Parnaíba (SP), ponto 3910, e Vitória (ES), ponto 4834, as conclusões são semelhantes, contudo, as variações no caso do município paulista equivalem a 12% no intercepto e -4.97% e -4.53% na taxa de analfabetismo e na renda do submodelo da dispersão e no município capixaba tem-se 6.52%, -3.80% e -4.04% , respectivamente. A observação paralela dos três casos reafirma o fato de que a influência de um ponto é menor à medida que o valor da distância de Cook é mais baixo.

Quando as três observações são excluídas do conjunto de dados, o impacto é maior em todos os parâmetros se comparado às exclusões individuais, tanto positiva quanto negativamente. Da mesma forma, as maiores influências ocorrem em $\hat{\beta}_0$, $\hat{\gamma}_1$ e $\hat{\gamma}_2$, onde os percentuais atingem as marcas de 37.23%, -43.17% e -38.52% , respectivamente.

Além disso, os gráficos dos resíduos ponderados padronizados 2 e dos valores observados *versus* os valores preditos também foram tomados para análise, sendo mostrados na Figura 5. Os resíduos parecem dispor-se de forma aleatória em torno de zero, sugerindo bom ajuste do modelo. Apesar de alguns pontos estarem além dos limites estabelecidos

Tabela 6 – Variações percentuais sobre as estimativas dos parâmetros excluindo as observações influentes.

		2900	3910	4834	2900,3910,4834
<i>intercepto</i>	$\hat{\beta}_0$	14.62	12.00	6.52	37.23
<i>analf</i>	$\hat{\beta}_1$	-3.48	-2.76	-1.21	-8.61
<i>renda</i>	$\hat{\beta}_2$	-9.06	-6.15	-3.24	-20.20
<i>DU</i>	$\hat{\beta}_3$	-0.32	-0.13	-0.26	-1.00
<i>gastopc2006</i>	$\hat{\beta}_4$	-2.64	-1.22	-0.39	-4.57
<i>analf</i> × <i>gastopc2006</i>	$\hat{\beta}_5$	5.25	2.97	0.93	10.20
<i>renda</i> × <i>gastopc2006</i>	$\hat{\beta}_6$	4.08	0.60	-0.57	4.08
<i>intercepto</i>	$\hat{\gamma}_0$	6.77	1.51	1.26	13.68
<i>analf</i>	$\hat{\gamma}_1$	-21.11	-4.97	-3.80	-43.17
<i>renda</i>	$\hat{\gamma}_2$	-20.27	-4.53	-4.04	-38.52

em $(-3, 3)$, 23 especificamente, eles representam uma quantidade menor que 0.5% do total de municípios brasileiros. A disposição dos pontos em torno da reta bissetriz no gráfico dos valores observados *versus* os valores preditos sinaliza bom ajuste do modelo, fato que é reforçado pelo gráfico de probabilidade meio-normal com envelopes simulados mostrado na Figura 6, onde percebe-se os resíduos dentro das bandas de confiança, não havendo indícios de afastamento das suposições feitas para o modelo e, portanto, a regressão beta estimada mostra-se apropriada aos dados.

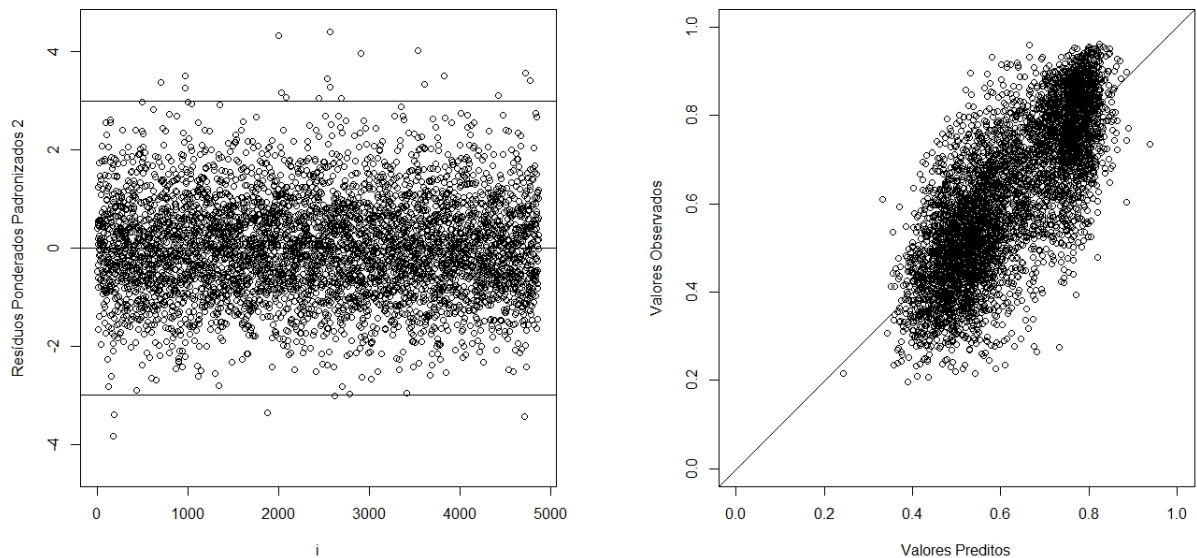


Figura 5 – Gráficos dos resíduos ponderados padronizados 2 *versus* os índices das observações e dos valores observados *versus* os valores preditos do modelo para o ano 2006.

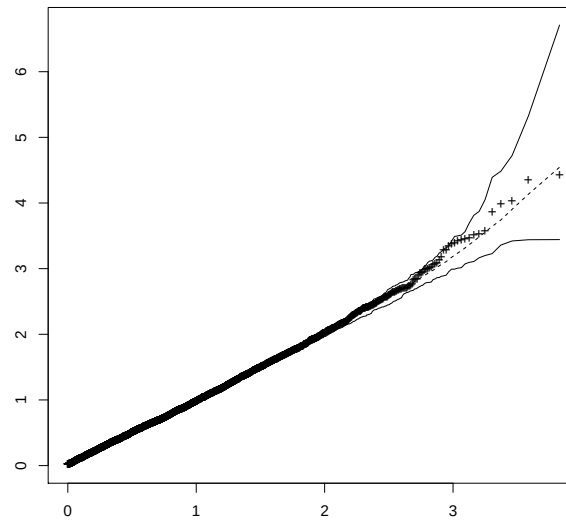


Figura 6 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2006.

4.3 Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2010

O modelo ajustado para o ano 2010 resultou em um pseudo- R^2 de 0.54, ou seja, cerca de 54% da variabilidade total da proporção de votos válidos da presidente Dilma no segundo turno da eleição presidencial é explicada pelas variáveis independentes, ao passo que o pseudo- R^2_{RV} foi de 0.56.

Os gráficos de diagnóstico que conferem a validade e a adequação do modelo aos dados são apresentados nas Figuras 7, 8 e 9. Duas observações se destacam no gráfico de alavancagem generalizada, Figura 7, as quais representam os municípios de Queimadas (PB), observação 2713, e Santa Rita (PB), observação 2729. Este último registrou o maior valor de gasto *per capita* com o Programa Bolsa Família no ano de 2010 entre todos os municípios brasileiros em estudo.

Os ajustes realizados sem as observações citadas, tanto considerando as exclusões individuais quanto a exclusão coletiva, não implicaram variações percentuais relevantes. Entre todos, o maior impacto ocorrido representa um aumento de 6.09% na estimativa no parâmetro $\hat{\beta}_1$, associado à covariável *rendapc2010* do submodelo da média, para o caso em que as duas observações foram retiradas da amostra.

Na mesma Figura encontra-se o gráfico com as distâncias de Cook, em que o ponto marcado – 2713 – é uma das observações que se mostrou discrepante também no gráfico de

alavancagem e representa o município de Queimadas (PB), como citado anteriormente. A retirada dessa observação não provocou variações percentuais expressivas nas estimativas dos parâmetros, a saber, o caso mais influente foi de um aumento de 5.46% na estimativa $\hat{\beta}_1$, submodelo da média. Para todas as demais estimativas dos dois submodelos, média e dispersão, o impacto manteve-se inferior a 3%, seja no aumento ou redução das estimativas.

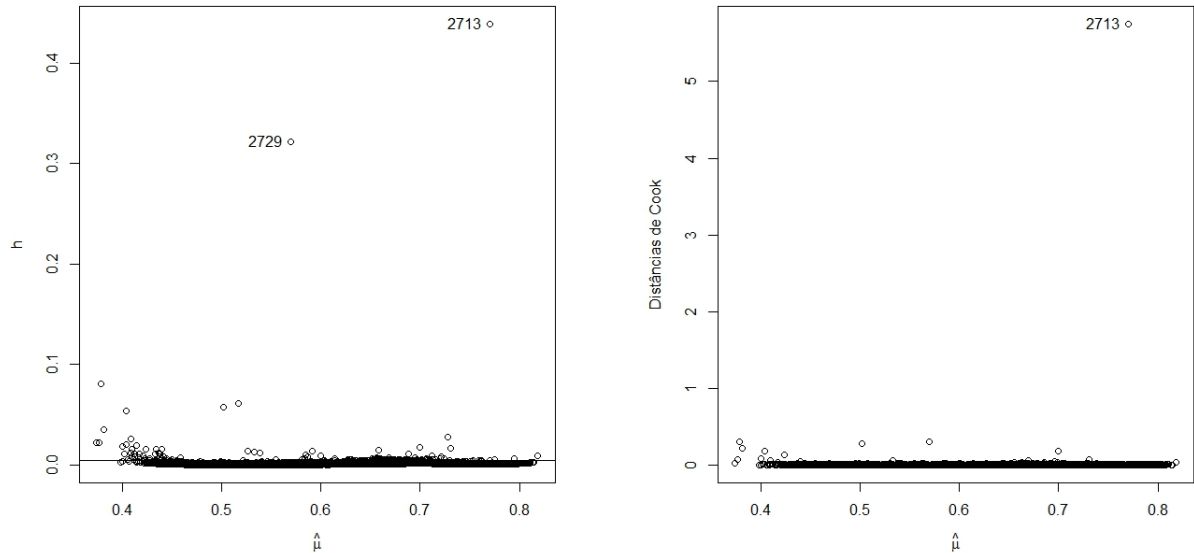


Figura 7 – Gráficos das medidas de alavancagem *versus* os valores preditos do modelo para o ano 2010 e das distâncias de Cook *versus* os valores preditos.

No gráfico dos resíduos ponderados padronizados 2 observam-se 22 pontos fora do intervalo $(-3, 3)$, contudo, eles representam uma quantidade inferior a 0.4% do total de municípios brasileiros e a disposição aleatória dos pontos em torno de zero indica o bom ajuste do modelo. Os pontos distribuídos em volta da reta bissetriz no gráfico dos valores observados *versus* os valores preditos sugerem um ajuste satisfatório do modelo de regressão beta. Os dois gráficos mencionados podem ser visualizados na Figura 8.

Pelo gráfico de probabilidade meio-normal com envelopes simulados, apresentado na Figura 9, é possível certificar que não há indícios de afastamento da suposição de que o modelo estimado é adequado para representar os dados, visto que, em geral, os resíduos encontram-se dentro das bandas de confiança dos envelopes.

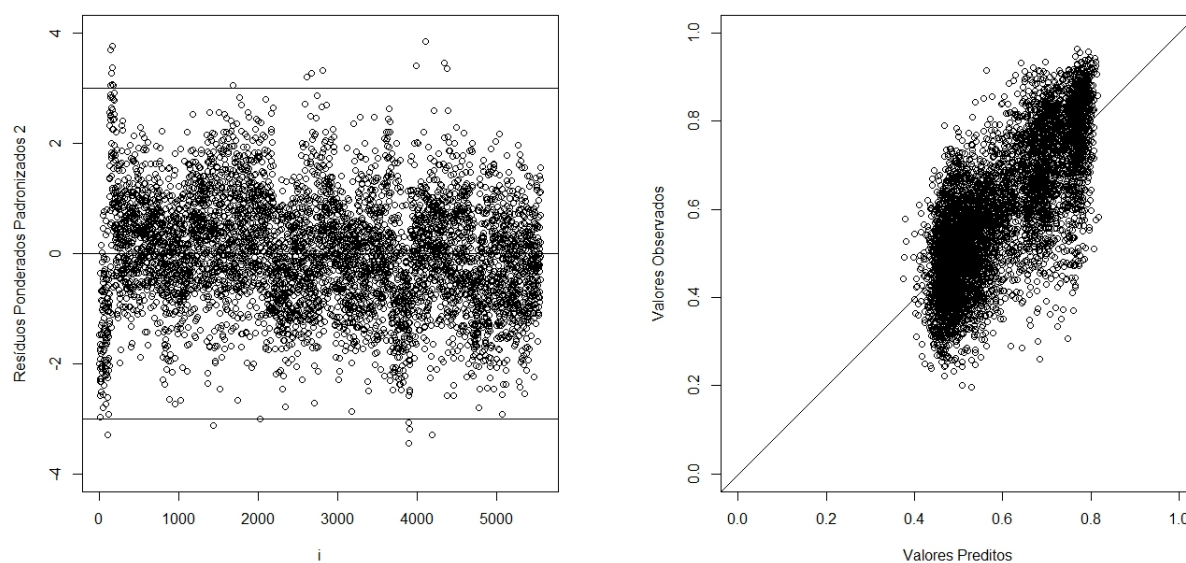


Figura 8 – Gráficos dos resíduos ponderados padronizados 2 *versus* os índices das observações e dos valores observados *versus* os valores preditos do modelo para o ano 2010.

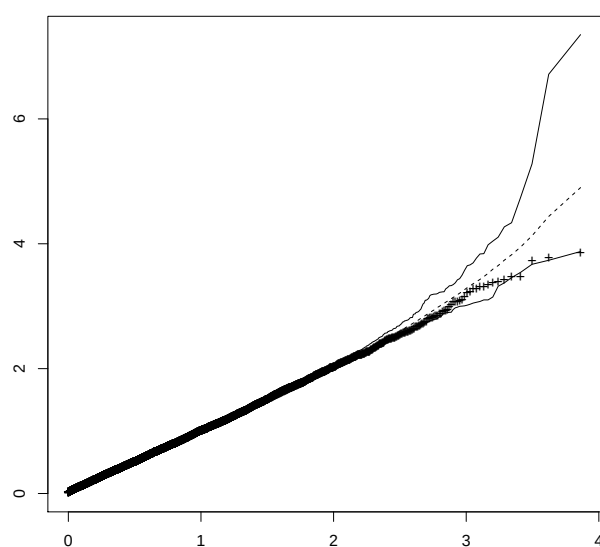


Figura 9 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2010.

4.4 Análise de diagnóstico do modelo de regressão beta para a eleição presidencial de 2014

Para o modelo ajustado para o ano 2014, o pseudo- R^2 foi de 0.71, isto é, a variabilidade de y é explicada em cerca de 71% pelas variáveis independentes. O pseudo- R^2_{RV} manteve-se em 0.72.

As Figuras 10, 11 e 12 apresentam os gráficos de diagnóstico que certificam a validade e adequação do modelo aos dados. O gráfico de alavancagem generalizada, apresentado na Figura 10, indica a existência de duas observações discrepantes, 1310 e 5341, correspondentes aos municípios de São José de Ribamar (MA) e São Paulo (SP). A retirada dessas observações, de modo individual e coletivo, não provocou grandes impactos sobre as estimativas dos parâmetros. A maior variação representou um aumento de 1.60% em $\hat{\beta}_1$, associado à covariável *rur2010* do submodelo da média, no caso em que apenas o ponto 5341 foi desconsiderado. As variações percentuais de todos os outros parâmetros, seja do submodelo da média ou da dispersão, para todas as exclusões consideradas, mantiveram-se inferiores a 1%.

No gráfico das distâncias de Cook, que pode ser observado na mesma Figura, nota-se um único ponto possivelmente influente – 2528 – que corresponde ao município de Santarém (PA). O ajuste sem essa observação não ocasionou mudanças consideráveis nas estimativas, sendo que a maior variação percentual ocorreu em $\hat{\beta}_1$, associado à variável *rur2010* do submodelo da média, em que a estimativa sofreu aumento de 8.21%. Como no caso da alavancagem generalizada, o impacto em todos os demais parâmetros foi inferior a 1% nos dois submodelos, média e dispersão.

Na Figura 11 estão os gráficos dos resíduos ponderados padronizados 2 e dos valores observados *versus* os valores preditos. Os pontos posicionados aleatoriamente em volta de zero no gráfico dos resíduos ponderados padronizados 2 sugerem ajuste adequado do modelo. Os pontos além do limite dado por $(-3, 3)$ representam uma quantidade menor que 0.25% do total de municípios brasileiros, 13 no total. Por meio do gráfico dos valores observados *versus* os valores preditos, o bom ajuste do modelo estimado é percebido pela disposição dos pontos em torno da reta bissetriz.

Os resíduos posicionados dentro das bandas de confiança no gráfico de probabilidade meio-normal com envelopes simulados, Figura 12, indicam que o modelo de regressão beta assegura uma boa representação para os dados.

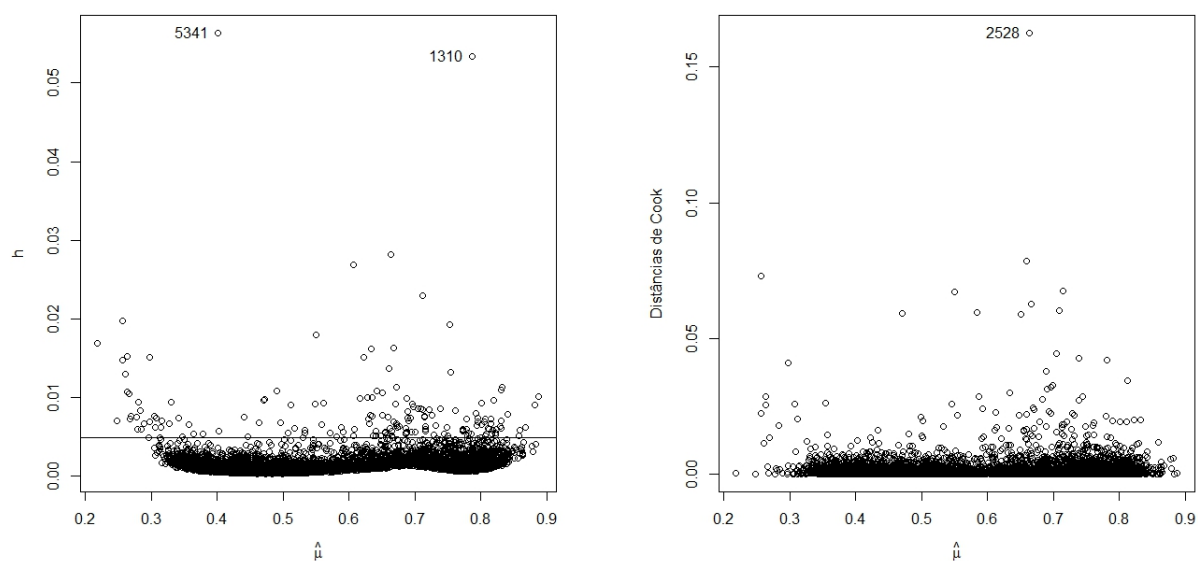


Figura 10 – Gráficos das medidas de alavancagem *versus* os valores preditos do modelo para o ano 2014 e das distâncias de Cook *versus* os valores preditos.

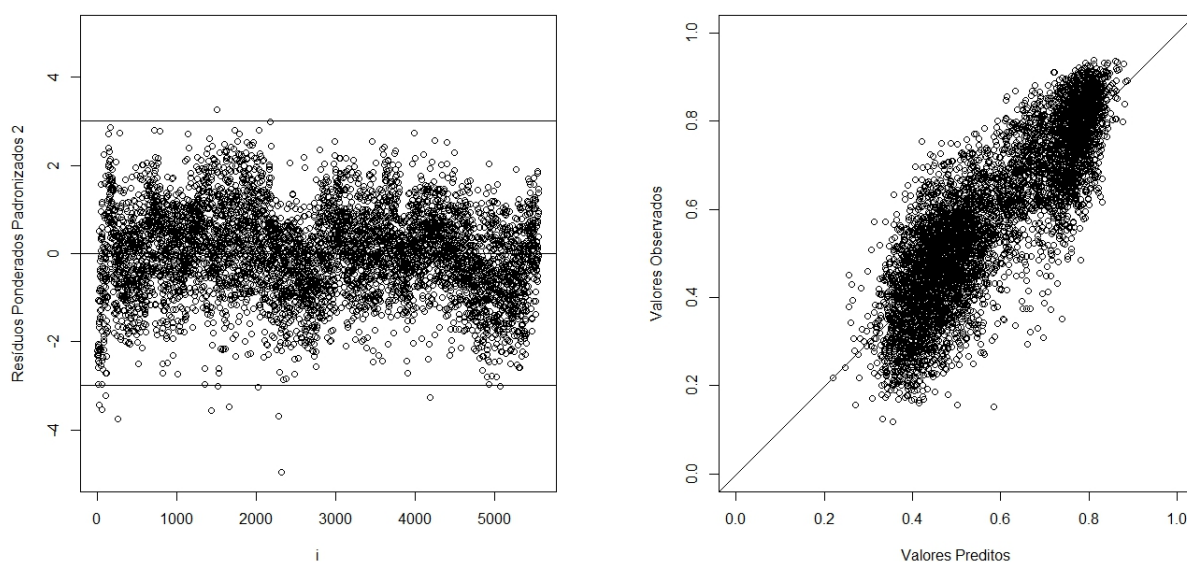


Figura 11 – Gráficos dos resíduos ponderados padronizados 2 *versus* os índices das observações e dos valores observados *versus* os valores preditos do modelo para o ano 2014.

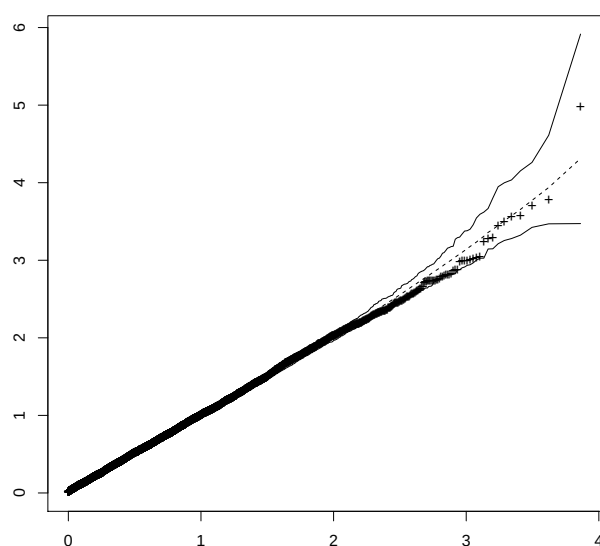


Figura 12 – Gráfico de probabilidade meio-normal com envelopes simulados do modelo para o ano 2014.

5 Impactos Estimados

5.1 Introdução

O objetivo principal deste estudo é avaliar os impactos exercidos pelos gastos *per capita* com programas assistenciais (especialmente o Programa Bolsa Família) nos anos 2006, 2010 e 2014 sobre a proporção média de votos válidos recebidos pelos presidentes eleitos no segundo turno das três eleições presidenciais ocorridas nos anos citados, fazendo uma comparação entre os resultados dos diferentes anos para verificar a evolução da influência que o assistencialismo teve sobre o resultado final das eleições. Dado que $\mu = \mathbb{E}(y)$, em que y representa a proporção de votos válidos no segundo turno das eleições presidenciais, o impacto (I) é obtido por meio da derivada da média da resposta em relação à covariável de interesse, neste caso, os gastos *per capita* ($gastopc$). Desse modo, I é calculado da seguinte forma:

$$I_t = \frac{\partial \mathbb{E}(y_t)}{\partial gastopc_t} = \frac{\partial \mu_t}{\partial gastopc_t}. \quad (5.1)$$

Para cada impacto estimado, foi obtido o erro-padrão correspondente. Obteve-se ainda os intervalos de confiança associados com nível de 95% de confiança. Esses dois processos foram realizados por meio de bootstrap paramétrico com a utilização de 1000 réplicas, sendo que os quantis 0.025 e 0.975 dos 1000 impactos calculados foram usados como limites inferior e superior dos intervalos construídos. Os passos para a aplicação do método bootstrap empregado são descritos abaixo:

- (1) Para cada $t = 1, \dots, n$, gerar y_t^* da densidade beta com parâmetros $\hat{\mu}_t$ e $\hat{\sigma}_t$;
- (2) Estimar a regressão beta utilizando como resposta $y = (y_1^*, \dots, y_n^*)^\top$ conjuntamente com os regressores originais;
- (3) Com as estimativas obtidas em (2), calcular o impacto da covariável $gastopc$:

$$I_{t,b} = \frac{\partial \mathbb{E}(y_{t,b})}{\partial gastopc_t} = \frac{\partial \mu_{t,b}}{\partial gastopc_t} \quad ;$$

- (4) Executar os passos (1), (2) e (3) 1000 vezes;
- (5) Calcular o desvio-padrão/selecionar os quantis 0.025 e 0.975 dos 1000 impactos bootstrap estimados.

Adicionalmente, deseja-se também obter uma previsão da quantidade de votos perdidos caso o gasto *per capita* com o Bolsa Família tivesse permanecido no nível em que

se encontrava na eleição anterior, ou seja, se o gasto *per capita* não houvesse aumentado nos últimos 4 anos. Para isso, consideraram-se dois cenários para a variável *gastopc* utilizada nos ajustes:

- I) Gasto *per capita* de 4 anos antes a preços da data da eleição, inflacionado pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA);
- II) Gasto *per capita* de 4 anos antes sem correção pelo IPCA.

5.2 Impactos estimados do modelo de regressão beta para a eleição presidencial de 2006

De acordo com o modelo ajustado para o ano 2006, definido em (3.1), μ_t é dado por

$$\mu_t = g^{-1}(\beta_0 + \beta_1 \text{analf}_t + \beta_2 \text{renda}_t + \beta_3 \text{DU}_t + \beta_4 \text{gastopc2006}_t + \beta_5 \text{gastopc2006}_t \times \text{analf}_t + \beta_6 \text{gastopc2006}_t \times \text{renda}_t).$$

A função de ligação utilizada no submodelo da média foi *logit*, assim, de (5.1), o impacto exercido pela variável *gastopc2006* sobre a proporção média de votos válidos do segundo turno da eleição presidencial do ano 2006 é dado por

$$I_t = \frac{\partial \mathbb{E}(y_t)}{\partial \text{gastopc2006}_t} = (\beta_4 + \beta_5 \text{analf}_t + \beta_6 \text{renda}_t) \frac{\exp(\beta_0 + \beta_1 \text{analf}_t + \beta_2 \text{renda}_t + \beta_3 \text{DU}_t)}{[1 + \exp(\beta_0 + \beta_1 \text{analf}_t + \beta_2 \text{renda}_t + \beta_3 \text{DU}_t)]^2} \frac{+ \beta_4 \text{gastopc2006}_t + \beta_5 \text{gastopc2006}_t \times \text{analf}_t + \beta_6 \text{gastopc2006}_t \times \text{renda}_t}{+ \beta_4 \text{gastopc2006}_t + \beta_5 \text{gastopc2006}_t \times \text{analf}_t + \beta_6 \text{gastopc2006}_t \times \text{renda}_t}]^2$$

A Tabela 7 apresenta os impactos estimados e os respectivos erros-padrão obtidos via bootstrap. Três casos são avaliados, a saber:

- CASO 1: *renda* é fixada no primeiro quartil; *DU*, covariável *dummy*, é fixada em 0 e 1, alternadamente; *gastopc2006* é fixado no primeiro, segundo e terceiro quartis; por fim, *analf* é fixada na mediana;
- CASO 2: *renda* é fixada no segundo quartil e as demais covariáveis seguem a configuração descrita no Caso 1;
- CASO 3: *renda* é fixada no terceiro quartil e as demais covariáveis seguem a configuração descrita no Caso 1.

Nota-se que, nos três casos estudados, o impacto de *gastopc2006* é maior quando $DU=0$, isto é, quando se considera que a população de todos os municípios é menor do que 50,000 habitantes. Ademais, para todos os casos de *renda* os maiores impactos são registrados quando *gastopc2006* é fixado no primeiro quartil.

Os resultados apontam que com *renda* fixada no primeiro quartil (R\$ 88.78) o gasto com Programas Assistenciais representa um aumento médio de 0.2101% na votação do presidente eleito em 2006, Luiz Inácio Lula da Silva, quando se tomam todos os municípios com população menor do que 50,000 habitantes; com *renda* fixada no segundo e terceiro quartis (R\$ 163.90 e R\$ 235.00) e mantendo $DU=0$, o aumento médio é de 0.1606% e 0.1044%, respectivamente. Para todos os casos, os erros-padrão bootstrap para os impactos estimados do gasto sobre a proporção média de votos válidos são valores pequenos.

Tabela 7 – Impactos estimados do gasto *per capita* com assistencialismo em 2006 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.

<i>renda</i>	<i>DU</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	0.002269 (0.000121)	0.002159 (0.000111)	0.001875 (0.000081)
	1	0.001980 (0.000113)	0.001801 (0.000097)	0.001459 (0.000064)
Q_2	0	0.001665 (0.000114)	0.001629 (0.000110)	0.001523 (0.000094)
	1	0.001504 (0.000108)	0.001422 (0.000097)	0.001257 (0.000075)
Q_3	0	0.001055 (0.000198)	0.001049 (0.000153)	0.001028 (0.000144)
	1	0.000985 (0.000145)	0.000959 (0.000138)	0.000904 (0.000120)

Os intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos são apresentados na Tabela 8. É possível constatar que nenhum dos intervalos estabelecidos contém o valor zero, o que indica que todos os impactos estimados são estatisticamente diferentes de zero ao nível nominal de 5%, isto é, tais impactos não são estatisticamente nulos.

Tabela 8 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto *per capita* nos programas assistenciais em 2006 sobre a resposta média.

<i>renda</i>	<i>DU</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	(0.002028, 0.002501)	(0.001937, 0.002370)	(0.001710, 0.002028)
	1	(0.001752, 0.002204)	(0.001607, 0.001987)	(0.001326, 0.001579)
Q_2	0	(0.001446, 0.001885)	(0.001417, 0.001840)	(0.001340, 0.001699)
	1	(0.001296, 0.001711)	(0.001234, 0.001607)	(0.001109, 0.001395)
Q_3	0	(0.000929, 0.001693)	(0.000764, 0.001351)	(0.000754, 0.001308)
	1	(0.000713, 0.001265)	(0.000699, 0.001219)	(0.000672, 0.001126)

O comportamento dos impactos estimados para cada caso descrito anteriormente pode ser visualizado na Figura 13. As curvas mostram que, nos três casos considerados

para *renda*, o impacto quando $DU=0$ é sempre maior para qualquer valor de *gastopc2006* do que quando $DU=1$.

Constata-se ainda que, quando *renda* é fixada no primeiro e segundo quartis, para $DU=0$ e $DU=1$, os valores estimados do impacto são predominantemente decrescentes à medida que *gastopc2006* aumenta, comportamento também observado quando *renda* é fixada no terceiro quartil e $DU=1$. Todavia, quando *renda* é fixada no terceiro quartil e $DU=0$, verifica-se que o impacto é crescente para valores de *gastopc2006* até R\$ 17.45 enquanto que para valores de *gastopc2006* acima do referenciado, o impacto assume forma decrescente. Evidencia-se que, para valores altos do gasto, o impacto estimado é pequeno. Para fins de ilustração, tomando o caso em que *renda* é fixada no primeiro quartil e *gastopc2006* é fixado no terceiro quartil (R\$ 95.06), a Tabela 7 indica que os impactos estimados quando $DU=0$ e $DU=1$ são, respectivamente, 0.001875 e 0.001459. No entanto, ao tomar *gastopc2006* fixado em valores maiores, R\$ 300 por exemplo, os impactos passam a ser 0.000469 e 0.000305, o que representa algo em torno de 4 e 4.78 vezes menor. Nos casos em que *renda* é fixada no segundo e terceiro quartis, a redução no impacto para valores altos do gasto toma proporções mais amenas, cerca de 1.8 e 1.6 vezes menores, respectivamente.

A observação da Tabela 3, que apresenta as estimativas pontuais dos parâmetros do modelo 2006, revela que o coeficiente estimado da covariável *analf* é bastante elevado quando comparado às demais covariáveis. Diante disso, buscou-se verificar o comportamento dos impactos ao fixar esta covariável no primeiro, segundo e terceiro quartis, semelhante ao modo como foi feito para *renda*. Os resultados são apresentados nas Tabelas 9 e 10 e na Figura 14, em que é possível verificar que os impactos estimados são valores ligeiramente menores do que os obtidos quando a *renda* foi fixada, contudo, o comportamento é muito similar conforme mostra a Figura 14. Os intervalos de confiança bootstrap ao nível de 95% de confiança apontam que todos os impactos estimados são estatisticamente diferentes de zero ao nível nominal de 5%.

Tabela 9 – Impactos estimados do gasto *per capita* com assistencialismo em 2006 sobre a resposta média fixando *analf* e seus respectivos erros-padrão obtidos via bootstrap.

<i>analf</i>	<i>DU</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	0.001938 (0.000130)	0.001903 (0.000126)	0.001773 (0.000105)
	1	0.001787 (0.000125)	0.001686 (0.000112)	0.001471 (0.000083)
Q_2	0	0.001665 (0.000114)	0.001629 (0.000110)	0.001523 (0.000094)
	1	0.001504 (0.000108)	0.001422 (0.000097)	0.001257 (0.000075)
Q_3	0	0.001039 (0.000129)	0.001015 (0.000124)	0.000964 (0.000111)
	1	0.000896 (0.000115)	0.000857 (0.000107)	0.000786 (0.000090)

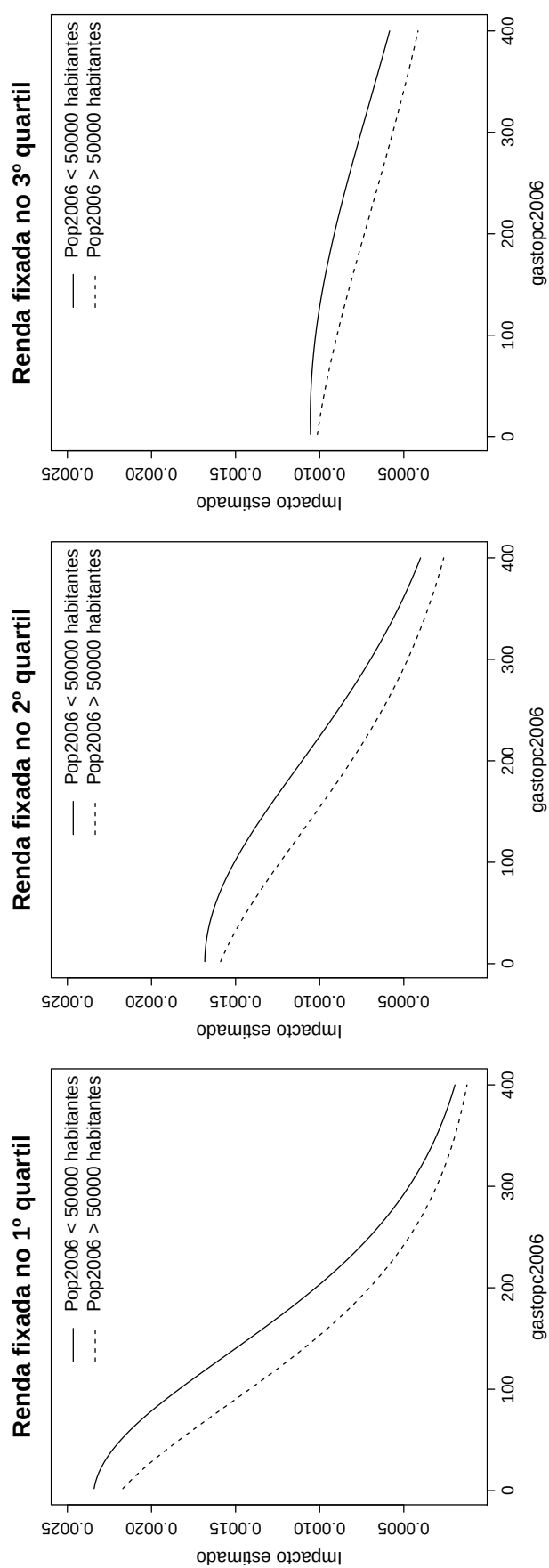


Figura 13 – Impactos estimados do gasto *per capita* nos programas assistenciais em 2006 sobre a resposta média.

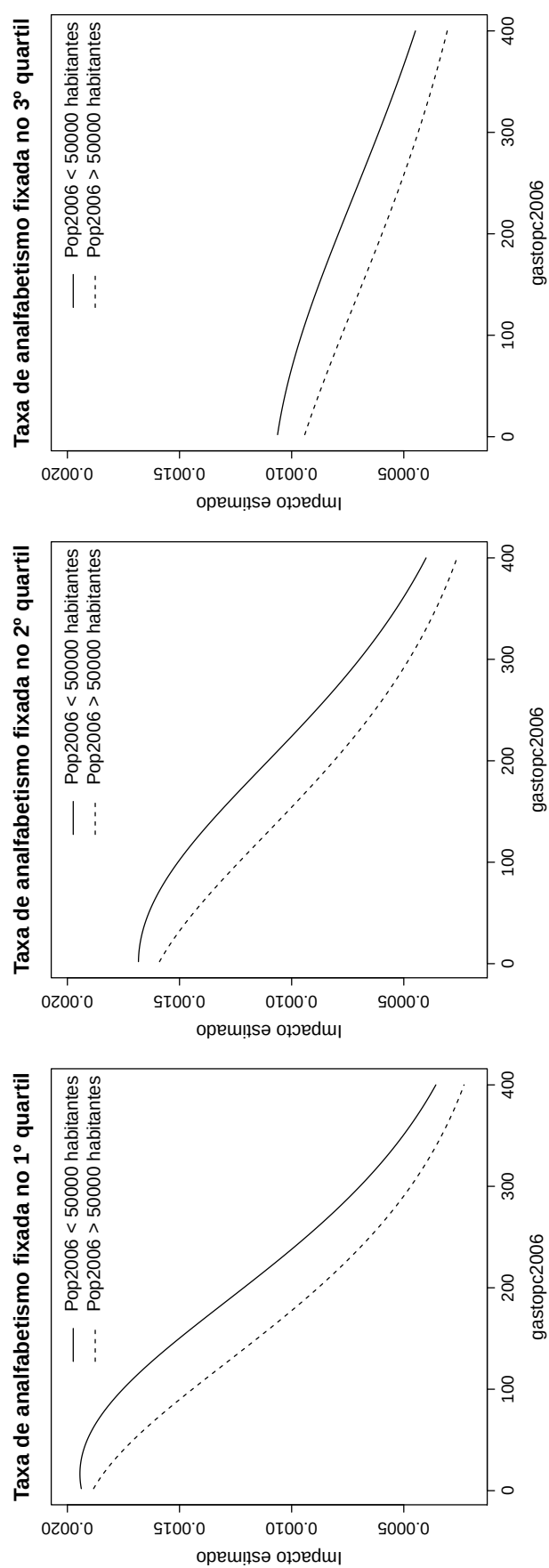


Figura 14 – Impactos estimados do gasto *per capita* nos programas assistenciais em 2006 sobre a resposta média fixando *analf*.

Tabela 10 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto *per capita* nos programas assistenciais em 2006 sobre a resposta média fixando *analf*.

<i>analf</i>	<i>DU</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	(0.001681, 0.002184)	(0.001653, 0.002144)	(0.001561, 0.001973)
	1	(0.001536, 0.002024)	(0.001458, 0.001901)	(0.001297, 0.001627)
Q_2	0	(0.001446, 0.001885)	(0.001417, 0.001840)	(0.001340, 0.001699)
	1	(0.001296, 0.001711)	(0.001234, 0.001607)	(0.001109, 0.001395)
Q_3	0	(0.000780, 0.001291)	(0.000766, 0.001258)	(0.000736, 0.001178)
	1	(0.000666, 0.001118)	(0.000643, 0.001061)	(0.000602, 0.000954)

Por fim, de acordo com o modelo estimado para 2006, os resultados apontam que se o gasto *per capita* com os Programas Assistenciais neste ano tivesse permanecido no nível em que se encontrava em 2002, considerando o cenário I (*gastopc2002* corrigido pelo IPCA), Lula teria obtido 52,712,657 votos, isto é, 5,582,385 votos a menos do que a quantidade conquistada por ele na eleição presidencial, cujo total foi de 58,295,042 votos. Geraldo Alckmin, que disputou a presidência com Lula, obteve 37,543,178 votos, o que implica uma diferença de 20,751,864 votos entre os dois. Admitindo que na votação de Lula haveria tal redução e se todos os 5,582,385 votos perdidos pelo ex-presidente se *gastopc2006*=*gastopc2002* tivessem migrado para o candidato Geraldo Alckmin, este teria obtido 43,125,563 votos válidos e, portanto, segundo o modelo estimado, Lula ainda assim teria vencido a eleição presidencial ocorrida no ano 2006 mesmo se o gasto com assistencialismo tivesse permanecido no nível em que se encontrava há 4 anos. O intervalo com 95% de confiança associado obtido via bootstrap é $(-12,057,902; 16,066,978)$. Como o número estimado de votos perdidos pertence ao intervalo, evidencia-se que tal quantidade é significativa ao nível de 95% de confiança.

Tomando o cenário II (*gastopc2002* não corrigido pelo IPCA) para análise, os resultados apontam que Lula teria conquistado 5,945,638 votos a menos, o que culminaria um total de 52,349,404 votos recebidos na eleição em vez dos 58,295,042 votos que obteve. Da mesma forma que no cenário I, essa quantidade não seria suficiente para alterar o resultado final da eleição. Salienta-se que o número estimado de votos perdidos no cenário em que não foi considerada a correção na variável *gastopc2002* foi maior, com 363,253 votos a mais. O intervalo de confiança associado, ao nível de confiança de 95% obtido via bootstrap, é $(-10,887,263; 16,238,203)$, sendo que o número estimado de votos perdidos pertence ao intervalo.

5.3 Impactos estimados do modelo de regressão beta para a eleição presidencial de 2010

De acordo com o modelo ajustado para o ano 2010, definido em (3.2), μ_t é dado por

$$\mu_t = g^{-1}(\beta_0 + \beta_1 \text{rendapc2010}_t + \beta_2 \text{govpt2010}_t + \beta_3 \text{ppent2010}_t + \beta_4 \text{gastopc2010}_t + \beta_5 \text{dnord}_t + \beta_6 \text{dpb}_t + \beta_7 \text{gastopc2010}_t \times \text{rendapc2010}_t + \beta_8 \text{gastopc2010}_t \times \text{dnord}_t).$$

A função de ligação utilizada no submodelo da média foi *cloglog*, assim, de (5.1), o impacto exercido pela variável *gastopc2010* sobre a proporção média de votos válidos do segundo turno da eleição presidencial do ano 2010 é dado por

$$I_t = \frac{\partial \mathbb{E}(y_t)}{\partial \text{gastopc2010}_t} = \exp[-\exp(\beta_0 + \beta_1 \text{rendapc2010}_t + \beta_2 \text{govpt2010}_t + \beta_3 \text{ppent2010}_t + \beta_4 \text{gastopc2010}_t + \beta_5 \text{dnord}_t + \beta_6 \text{dpb}_t + \beta_7 \text{gastopc2010}_t \times \text{rendapc2010}_t + \beta_8 \text{gastopc2010}_t \times \text{dnord}_t)] \times \exp(\beta_0 + \beta_1 \text{rendapc2010}_t + \beta_2 \text{govpt2010}_t + \beta_3 \text{ppent2010}_t + \beta_4 \text{gastopc2010}_t + \beta_5 \text{dnord}_t + \beta_6 \text{dpb}_t + \beta_7 \text{gastopc2010}_t \times \text{rendapc2010}_t + \beta_8 \text{gastopc2010}_t \times \text{dnord}_t) \times (\beta_4 + \beta_7 \text{rendapc2010}_t + \beta_8 \text{dnord}_t).$$

A Tabela 11 apresenta os impactos estimados e os respectivos erros-padrão obtidos via bootstrap. Três casos são avaliados, a saber:

- CASO 1: *rendapc2010* é fixada no primeiro quartil; *govpt2010*, *dnord* e *dpb*, covariáveis *dummies*, são fixadas em 0 e 1, alternadamente; *gastopc2010* é fixado no primeiro, segundo e terceiro quartis; por fim, *ppent2010* é fixado na mediana;
- CASO 2: *rendapc2010* é fixada no segundo quartil e as demais covariáveis seguem a configuração descrita no Caso 1;
- CASO 3: *rendapc2010* é fixada no terceiro quartil e as demais covariáveis seguem a configuração descrita no Caso 1.

Cabe ressaltar que, do total de oito cenários resultantes da combinação das três *dummies*, dois cenários foram desconsiderados por não apresentarem real sentido de ocorrência. Tais cenários assumiriam *govpt2010* alternando entre 0 e 1, *dnord*=0 e *dpb*=1, estes dois últimos representando a situação fictícia de que todos os municípios não pertenceriam à região Nordeste, mas pertenceriam ao estado da Paraíba, o que não seria possível.

Nota-se que, nos três casos estudados, é evidente a diferença entre os resultados quando se considera *dnord*=0 e *dnord*=1, sendo que o impacto de *gastopc2010* é maior

quando $dnord=0$, isto é, quando se considera que todos os municípios não pertencem à região Nordeste. Destaca-se o fato de que quando $dnord=1$ os impactos têm efeito negativo para $rendapc2010$ fixada no segundo e terceiro quartis. Ademais, quando $dnord=0$ os maiores impactos são registrados quando $gastopc2010$ é fixado no terceiro quartil para os três casos considerados da renda. Uma única exceção ocorre para o cenário em que $govpt2010=0$, $dnord=0$, $dpb=0$ e $rendapc2010$ está fixada no primeiro quartil, pois neste caso o impacto maior ocorre com $gastopc2010$ fixado no segundo quartil. Quanto à variável $rendapc2010$, os maiores impactos são registrados quando está fixada no primeiro quartil.

Por meio da Tabela 11 é possível averiguar que se todos os estados não fossem governados pelo PT e se todos os municípios não pertencessem à região Nordeste, portanto, nem ao estado da Paraíba, com a renda e o gasto *per capita* fixados no primeiro quartil, o gasto com o Programa Bolsa Família representaria um aumento de 0.136531% na votação de Dilma Rousseff. Tomando o mesmo cenário, mas assumindo que os estados fossem governados pelo PT, o aumento teria sido cerca de 0.129364%. Entre todos os cenários considerados, o maior impacto é registrado para o caso em que todos os estados seriam governados pelo PT e todos os municípios não pertenceriam à região Nordeste nem ao estado da Paraíba, com $rendapc2010$ fixada no primeiro quartil (R\$ 281) e $gastopc2010$ fixado no terceiro quartil (R\$ 166.90), situação em que o aumento seria de 0.140720%. Por outro lado, o menor impacto ocorre quando se considera todos os estados governados pelo PT e todos os municípios pertencentes à região Nordeste, porém não ao estado da Paraíba, com $rendapc2010$ fixada no terceiro quartil (R\$ 650.40) e $gastopc2010$ fixado no primeiro quartil (R\$ 49.42). Neste caso, haveria uma redução de 0.046514% na votação.

Os resultados apontam que com $rendapc2010$ fixada no primeiro quartil (R\$ 281) o gasto com o Programa Bolsa Família representa um aumento médio de 0.138308% na votação da presidente eleita em 2010, Dilma Rousseff, quando $govpt2010$, $dnord$ e dpb assumem o valor 0, caso que registra os maiores impactos; com $rendapc2010$ fixada no segundo e terceiro quartis (R\$ 467.40 e R\$ 650.40) este aumento médio é de 0.112643% e 0.086657%, respectivamente. De modo geral e principalmente quando $dnord=0$, nota-se que o impacto estimado aumenta à medida que o valor do gasto *per capita* aumenta, visto que os impactos quando $gastopc2010$ está fixado no terceiro quartil são maiores do que os registrados quando se fixa o segundo quartil, que por sua vez são maiores do que os valores no primeiro quartil, comportamento este observado para os três casos avaliados da renda. Para todos os casos, os erros-padrão bootstrap para os impactos estimados do gasto sobre a proporção média de votos válidos são valores pequenos.

A Tabela 12 mostra os intervalos de confiança bootstrap ao nível de 95% de confiança obtidos para os impactos. De acordo com o que é apresentado, verifica-se que todos os impactos estimados são estatisticamente diferentes de zero ao nível nominal de 5%, visto que o valor zero não está contido em nenhum dos intervalos supracitados.

Tabela 11 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2010 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.

<i>rendapc2010</i>	<i>govpt2010</i>	<i>dnord</i>	<i>dpb</i>	GASTO PER CAPITA		
				Q_1	Q_2	Q_3
Q_1	0	0	0	0.00136531 (0.00005261)	0.00140440 (0.00005690)	0.00137954 (0.00005343)
	1	0	0	0.00129364 (0.00004825)	0.00136782 (0.00003560)	0.00140720 (0.00005703)
	0	1	0	0.00004544 (0.00001879)	0.00004533 (0.00001872)	0.00004518 (0.00001858)
	1	1	0	0.00004763 (0.00001961)	0.00004758 (0.00001957)	0.00004752 (0.00001952)
	0	1	1	0.00004798 (0.00001973)	0.00004796 (0.00001972)	0.00004793 (0.00001970)
Q_2	1	1	1	0.00004773 (0.00001955)	0.00004776 (0.00001959)	0.00004781 (0.00001962)
	0	0	0	0.00109792 (0.00005104)	0.00113467 (0.00005481)	0.00114671 (0.00005434)
	1	0	0	0.00103262 (0.00004716)	0.00108961 (0.00004126)	0.00113870 (0.00005626)
	0	1	0	-0.00020331 (0.00003002)	-0.00020494 (0.00003077)	-0.00020681 (0.00003100)
	1	1	0	-0.00021030 (0.00003116)	-0.00021068 (0.00003127)	-0.00021090 (0.00003129)
Q_3	0	1	1	-0.00021091 (0.00003127)	-0.00021083 (0.00003128)	-0.00021047 (0.00003116)
	1	1	1	-0.00020766 (0.00003099)	-0.00020659 (0.00003070)	-0.00020496 (0.00003022)
	0	0	0	0.00084057 (0.00006093)	0.00086845 (0.00006477)	0.00089068 (0.00006634)
	1	0	0	0.00078517 (0.00005664)	0.00082349 (0.00005700)	0.00086440 (0.00006686)
	0	1	0	-0.00045520 (0.00005347)	-0.00046059 (0.00005522)	-0.00046455 (0.00005493)
Q_3	1	1	0	-0.00046514 (0.00005458)	-0.00046473 (0.00005435)	-0.00046141 (0.00005310)
	0	1	1	-0.00046451 (0.00005458)	-0.00046206 (0.00005401)	-0.00045624 (0.00005238)
	1	1	1	-0.00045306 (0.00005348)	-0.00044628 (0.00005192)	-0.00043533 (0.00004919)

Tabela 12 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto *per capita* no programa Bolsa Família em 2010 sobre a resposta média.

	<i>rendapc2010</i>	<i>govpt2010</i>	<i>dnord</i>	<i>dpb</i>	GASTO PER CAPITA		
					Q_1	Q_2	Q_3
Q_1	0	0	0	0	(0.00126533, 0.00146830)	(0.00129597, 0.00151671)	(0.00127727, 0.00148465)
	1	0	0	0	(0.00120284, 0.00138442)	(0.00129441, 0.00143547)	(0.00129811, 0.00152005)
	0	1	0	0	(0.00000760, 0.00008211)	(0.00000752, 0.00008205)	(0.00000759, 0.00008125)
	1	1	0	0	(0.00000796, 0.00008581)	(0.00000796, 0.00008567)	(0.00000796, 0.00008545)
	0	1	1	1	(0.00000803, 0.00008654)	(0.00000803, 0.00008652)	(0.00000803, 0.00008647)
Q_2	1	1	1	1	(0.00000800, 0.00008550)	(0.00000800, 0.00008565)	(0.00000800, 0.00008582)
	0	0	0	0	(0.00100023, 0.00119861)	(0.00103107, 0.00124246)	(0.00104221, 0.00125341)
	1	0	0	0	(0.00094111, 0.00112648)	(0.00101032, 0.00117501)	(0.00103165, 0.00124895)
	0	1	0	0	(-0.00025967, -0.00014305)	(-0.00026263, -0.00014366)	(-0.00026523, -0.00014487)
	1	1	0	0	(-0.00026834, -0.00014814)	(-0.00026883, -0.00014830)	(-0.00026898, -0.00014841)
Q_3	0	1	1	1	(-0.00026885, -0.00014835)	(-0.00026898, -0.00014822)	(-0.00026867, -0.00014796)
	1	1	1	1	(-0.00026555, -0.00014563)	(-0.00026374, -0.00014511)	(-0.00026090, -0.00014428)
	0	0	0	0	(0.00072618, 0.00096437)	(0.00074804, 0.00100022)	(0.00076774, 0.00102560)
	1	0	0	0	(0.00068103, 0.00089830)	(0.00071769, 0.00093438)	(0.00074232, 0.00099861)
	0	1	0	0	(-0.00055981, -0.00034473)	(-0.00056867, -0.00035002)	(-0.00057111, -0.00035011)
Q_3	1	1	0	0	(-0.00057065, -0.00035188)	(-0.00057172, -0.00035194)	(-0.00056562, -0.00035079)
	0	1	1	1	(-0.00057177, -0.00035179)	(-0.00056777, -0.00035027)	(-0.00055863, -0.00034713)
	1	1	1	1	(-0.00055809, -0.00034390)	(-0.00054772, -0.00034034)	(-0.00053097, -0.00033373)

O comportamento dos impactos estimados para cada caso descrito anteriormente pode ser visualizado nas Figuras 15 e 16. As curvas mostram que, nas três situações consideradas para a renda, os impactos assumem formas diferentes quando se considera $dnord=0$ e $dnord=1$. Tomando o caso em que $dnord=0$, com $rendapc2010$ fixada no primeiro quartil e $dpb=0$, o impacto quando $govpt2010=0$ é maior para valores de $gastopc2010$ menores que R\$140.70, enquanto que para gastos mais altos que esse valor, o maior impacto é observado para $govpt2010=1$. Para renda fixada no segundo e terceiro quartis, esse comportamento se repete, contudo, os valores de referência do gasto *per capita* em que os impactos passam a ser maiores com $govpt2010=1$ são R\$ 179.71 e R\$ 240.31, respectivamente. Neste mesmo cenário constata-se que, tanto para $govpt2010=0$ quanto para $govpt2010=1$, o impacto é crescente para valores de $gastopc2010$ até R\$ 116.32 e R\$ 164.39, respectivamente, com $rendapc2010$ fixada no primeiro quartil, R\$ 150.46 e R\$ 208.97 com $rendapc2010$ fixada no segundo quartil e R\$ 202.70 e R\$ 278.62 com $rendapc2010$ fixada no terceiro quartil. Para valores de $gastopc2010$ acima dos referenciados, como a Figura 15 sinaliza, o impacto assume forma decrescente até tornar-se nulo para gastos aproximadamente maiores que R\$ 700.00, R\$ 900.00 e R\$ 1200.00 para a renda fixada no primeiro, segundo e terceiro quartis.

Ao se considerar o caso em que $dnord=1$, a Figura 16 aponta comportamentos distintos mesmo entre os diferentes casos considerados para $rendapc2010$. Quando está fixada no primeiro quartil, verifica-se que os impactos assumem forma decrescente, além de que, para valores de $gastopc2010$ menores que R\$ 277.23, o cenário em que $govpt2010=0$, $dnord=1$ e $dpb=1$ registra o maior impacto. A partir deste valor, o maior impacto ocorre quando $govpt2010$, $dnord$ e dpb assumem o valor 1, sendo este o único cenário em que se observa um comportamento inicialmente crescente para valores de gasto até R\$ 993.98, quando, então, começa a diminuir suavemente. Nos outros três cenários analisados, os valores estimados do impacto são predominantemente decrescentes à medida que $gastopc2010$ aumenta. O caso em que $govpt2010=0$, $dnord=1$ e $dpb=0$ destaca-se por estar relativamente mais afastado dos demais, indicando que o menor impacto é obtido quando se considera que todos os estados não são governados pelo PT e todos os municípios pertencem à região Nordeste, mas não ao estado da Paraíba.

Com $rendapc2010$ fixada no segundo e terceiro quartis, os impactos caracterizam-se por apresentarem forma crescente, além de que, para valores de $gastopc2010$ menores que R\$ 131.65 e R\$ 41.10, respectivamente, o cenário em que $govpt2010=0$, $dnord=1$ e $dpb=0$ registra o maior impacto. A partir deste valor, o maior impacto ocorre quando $govpt2010$, $dnord$ e dpb assumem o valor 1, sendo este o único cenário em que os valores estimados do impacto são predominantemente crescentes à medida que $gastopc2010$ aumenta. Nos outros três cenários analisados, se observa um comportamento inicialmente decrescente para valores de gasto até R\$ 53.63 e R\$ 5.57 quando $govpt2010=0$, $dnord=1$ e $dpb=1$, R\$ 181.10 e R\$ 63.39 quando $govpt2010=1$, $dnord=1$ e $dpb=0$ e R\$502.21 e R\$208.97 quando

$govpt2010=0$, $dnord=1$ e $dpb=0$ para $rendapc2010$ fixada no segundo e terceiro quartis, respectivamente. Para valores de $gastopc2010$ acima dos referenciados, o impacto começa a aumentar conforme aumentam os valores do gasto *per capita*. Nota-se que os impactos estimados nos quatro cenários avaliados quando $rendapc2010$ é fixada no terceiro quartil parecem convergir para o mesmo valor quando o gasto alcança os registros mais altos, característica não observada nos demais casos considerados para a renda. Contudo, em todos os casos constata-se que o efeito maior do impacto ocorre quando se considera que todos os estados seriam governados pelo PT e todos os municípios pertenceriam à região Nordeste e também ao estado da Paraíba. O segundo maior impacto é observado para a situação em que todos os estados não seriam governados pelo PT e todos os municípios pertenceriam à região Nordeste e ao estado da Paraíba, seguido pelo caso em que todos os estados seriam governados pelo PT e todos os municípios pertenceriam à região Nordeste, mas não ao estado da Paraíba. O cenário que apresenta os menores efeitos do impacto de $gastopc2010$ sobre a proporção de votos válidos é quando se assume que todos os estados não seriam governados pelo PT e todos os municípios pertenceriam à região Nordeste, mas não ao estado da Paraíba.

Por fim, para fazer a previsão do número de votos perdidos caso o gasto *per capita* com o Programa Bolsa Família no ano 2010 tivesse permanecido no nível em que se encontrava em 2006, igualou-se o número de municípios brasileiros disponíveis para esses dois anos, de forma que 698 municípios foram desconsiderados do conjunto de dados de 2010, totalizando, assim, 4864 municípios. Então, a partir do número estimado de votos perdidos calculados com base nessa amostra reduzida, projetou-se a proporção dos votos em relação ao ano 2006 para os 5560 municípios disponíveis em 2010, assumindo que tal proporção se manteve constante de 2006 para 2010. Logo, os resultados apontam que, considerando o cenário I ($gastopc2006$ corrigido pelo IPCA), Dilma Rousseff teria obtido 52,075,946 votos, isto é, 3,676,583 votos a menos do que a quantidade conquistada por ela na eleição presidencial, cujo total foi de 55,752,529 votos. José Serra, que disputou a presidência com Dilma, obteve 43,711,388 votos, o que implica uma diferença de 12,041,141 votos entre os dois. Admitindo que na votação de Dilma haveria tal redução e se todos os 3,676,583 votos perdidos pela ex-presidente se $gastopc2010=gastopc2006$ tivessem migrado para o candidato José Serra, este teria obtido 47,387,971 votos válidos e, portanto, segundo o modelo estimado, Dilma ainda assim teria vencido a eleição presidencial ocorrida no ano 2010 mesmo se o gasto com o Bolsa Família tivesse permanecido no nível em que se encontrava há 4 anos. O intervalo de confiança associado obtido via bootstrap é $(-19,923,956; 11,010,940)$. Como o número estimado de votos perdidos pertence ao intervalo, evidencia-se que tal quantidade é significativa ao nível de 95% de confiança.

Tomando o cenário II ($gastopc2006$ não corrigido pelo IPCA) para análise, os resultados apontam que Dilma teria conquistado 3,964,046 votos a menos, o que culminaria um total de 51,788,483 votos recebidos na eleição em vez dos 55,752,529 votos que obteve.

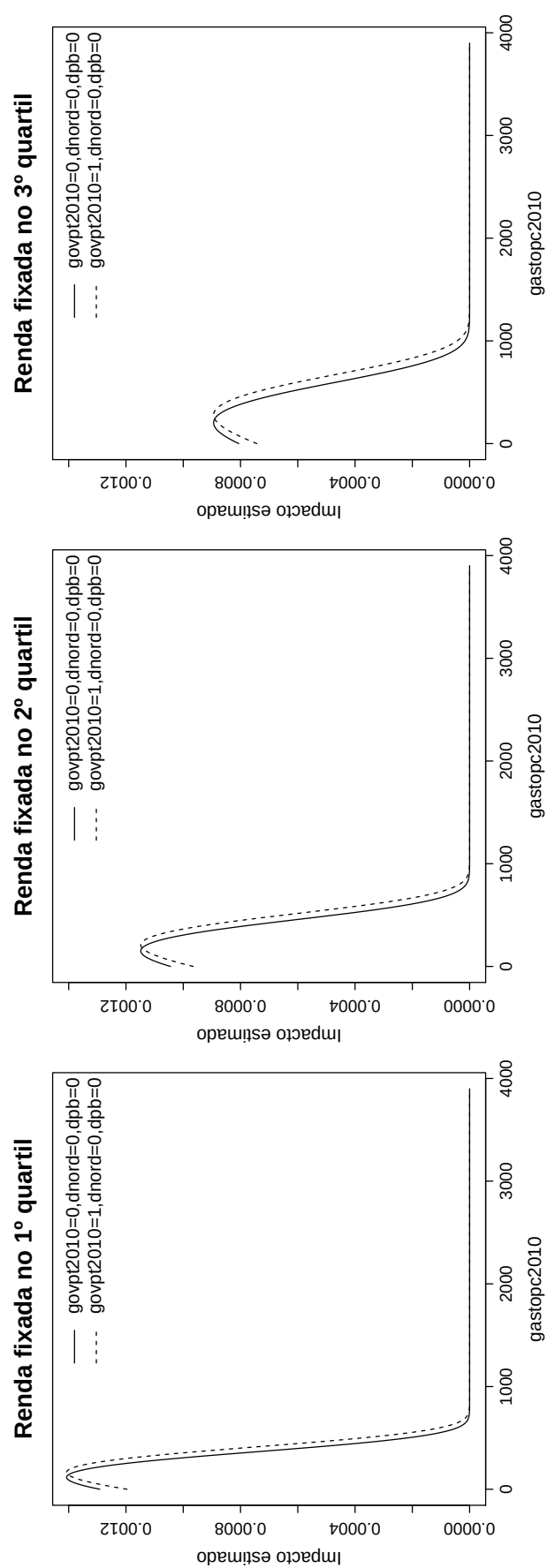


Figura 15 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2010 sobre a resposta média quando $dnord = 0$.

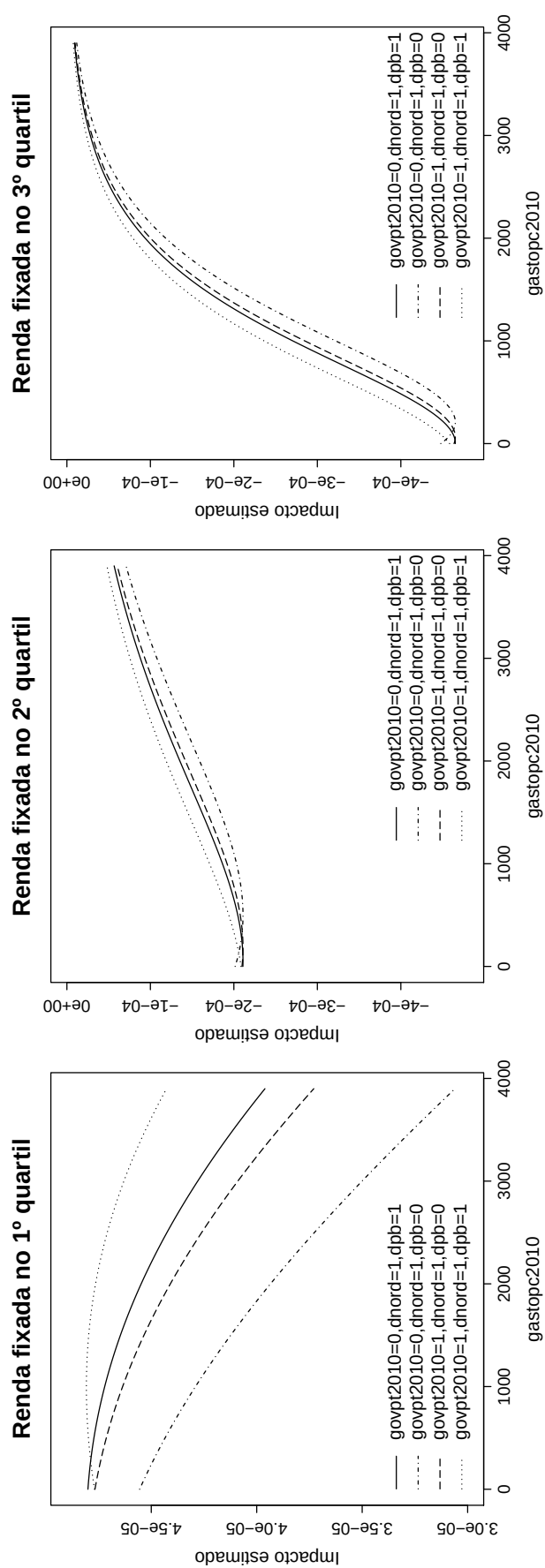


Figura 16 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2010 sobre a resposta média quando $dnord = 1$.

Da mesma forma que no cenário I, essa quantidade não seria suficiente para alterar o resultado final da eleição. Salienta-se que o número estimado de votos perdidos no cenário em que não foi considerada a correção na variável *gastopc2006* foi maior, com 287,463 votos a mais. O intervalo de confiança associado, ao nível de confiança de 95% obtido via bootstrap, é $(-19,603, 768; 11,164, 128)$, sendo que o número estimado de votos perdidos pertence ao intervalo.

5.4 Impactos estimados do modelo de regressão beta para a eleição presidencial de 2014

De acordo com o modelo ajustado para o ano 2014, definido em (3.3), μ_t é dado por

$$\mu_t = g^{-1}(\beta_0 + \beta_1 rur2010_t + \beta_2 panalf2010_t + \beta_3 ppent2010_t + \beta_4 dnord_t + \beta_5 gastopc2014_t + \beta_6 rendapc2010_t + \beta_7 panalf2010_t \times ppent2010_t + \beta_8 panalf2010_t \times gastopc2014_t).$$

A função de ligação utilizada no submodelo da média foi *logit*, assim, de (5.1), o impacto exercido pela variável *gastopc2014* sobre a proporção média de votos válidos do segundo turno da eleição presidencial do ano 2014 é dado por

$$I_t = \frac{\partial \mathbb{E}(y_t)}{\partial gastopc2014_t} = (\beta_5 + \beta_8 panalf2010_t) \frac{\exp(\beta_0 + \beta_1 rur2010_t + \beta_2 panalf2010_t + \beta_3 ppent2010_t + \beta_4 dnord_t + \beta_5 gastopc2014_t + \beta_6 rendapc2010_t + \beta_7 panalf2010_t \times ppent2010_t + \beta_8 panalf2010_t \times gastopc2014_t)}{[1 + \exp(\beta_0 + \beta_1 rur2010_t + \beta_2 panalf2010_t + \beta_3 ppent2010_t + \beta_4 dnord_t + \beta_5 gastopc2014_t + \beta_6 rendapc2010_t + \beta_7 panalf2010_t \times ppent2010_t + \beta_8 panalf2010_t \times gastopc2014_t)]^2}.$$

A Tabela 13 apresenta os resultados para os impactos estimados e os respectivos erros-padrão obtidos via bootstrap. Três casos são avaliados, a saber:

- CASO 1: *rendapc2010* é fixada no primeiro quartil; *dnord*, covariável *dummy*, é fixada em 0 e 1, alternadamente; *gastopc2014* é fixado no primeiro, segundo e terceiro quartis; por fim, as demais covariáveis são fixadas na mediana;
- CASO 2: *rendapc2010* é fixada no segundo quartil e as demais covariáveis seguem a configuração descrita no Caso 1;
- CASO 3: *rendapc2010* é fixada no terceiro quartil e as demais covariáveis seguem a configuração descrita no Caso 1.

Nota-se que, nos três casos estudados, o impacto de *gastopc2014* é maior quando *dnord*=0, isto é, quando se considera que todos os municípios não pertencem à região Nordeste. Ademais, quando *rendapc2010* é fixada no primeiro e segundo quartis, os maiores impactos são observados para *gastopc2014* fixado no primeiro quartil, enquanto que para *rendapc2010* fixada no terceiro quartil, o maior impacto é registrado para *gastopc2014* fixado no segundo quartil.

Os resultados apontam que com *rendapc2010* fixada no primeiro quartil (R\$ 281) o gasto com o Programa Bolsa Família representa um aumento médio de 0.0720% na votação da presidente eleita em 2014, Dilma Rousseff, quando se toma todos os municípios como não pertencentes à região Nordeste; com *rendapc2010* fixada no segundo e terceiro quartis (R\$ 467.40 e R\$ 650.40) e mantendo *dnord*=0, o aumento médio é de 0.0730% e 0.0737%, respectivamente. Para todos os casos, os erros-padrão bootstrap para os impactos estimados do gasto sobre a proporção média de votos válidos são valores pequenos.

Tabela 13 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2014 sobre a resposta média e seus respectivos erros-padrão obtidos via bootstrap.

<i>rendapc2010</i>	<i>dnord</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	0.000757 (0.000026)	0.000741 (0.000026)	0.000661 (0.000020)
	1	0.000725 (0.000027)	0.000679 (0.000025)	0.000561 (0.000018)
Q_2	0	0.000757 (0.000026)	0.000750 (0.000025)	0.000683 (0.000020)
	1	0.000739 (0.000026)	0.000699 (0.000024)	0.000589 (0.000017)
Q_3	0	0.000753 (0.000026)	0.000756 (0.000025)	0.000703 (0.000020)
	1	0.000748 (0.000026)	0.000717 (0.000024)	0.000615 (0.000017)

Na Tabela 14 estão os intervalos de confiança bootstrap ao nível de 95% de confiança obtidos para os impactos. Nenhum dos intervalos apresentados inclui o valor zero, portanto, todos os impactos estimados são estatisticamente diferentes de zero ao nível nominal de 5%, isto é, não são estatisticamente nulos.

Tabela 14 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto *per capita* no programa Bolsa Família em 2014 sobre a resposta média.

<i>rendapc2010</i>	<i>dnord</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	(0.000707, 0.000806)	(0.000691, 0.000790)	(0.000621, 0.000699)
	1	(0.000673, 0.000777)	(0.000630, 0.000726)	(0.000527, 0.000595)
Q_2	0	(0.000707, 0.000805)	(0.000701, 0.000798)	(0.000644, 0.000721)
	1	(0.000687, 0.000788)	(0.000652, 0.000745)	(0.000556, 0.000621)
Q_3	0	(0.000704, 0.000801)	(0.000706, 0.000803)	(0.000663, 0.000740)
	1	(0.000698, 0.000797)	(0.000670, 0.000762)	(0.000583, 0.000647)

O comportamento dos impactos estimados para cada caso descrito anteriormente pode ser visualizado na Figura 17. As curvas mostram que, com *rendapc2010* fixada no primeiro quartil, o impacto quando *dnord*=1 é maior para valores de *gastopc2014* menores que R\$ 3.09, enquanto que para gastos mais altos que esse valor, o maior impacto é observado para *dnord*=0. Para renda fixada no segundo e terceiro quartis, este comportamento se repete, contudo, os valores de referência do gasto *per capita* em que os impactos passam a ser maiores com *dnord*=0 são R\$ 35.68 e R\$ 67.63, respectivamente.

Constata-se ainda que, para *rendapc2010* fixada no primeiro e segundo quartis e tomando *dnord*=1, os valores estimados do impacto são predominantemente decrescentes à medida que *gastopc2014* aumenta. Todavia, quando *dnord*=0 para esses casos e quando *rendapc2010* é fixada no terceiro quartil, observa-se que o impacto é crescente para valores de *gastopc2014* até R\$ 64.90 no primeiro caso, R\$ 97.50 no segundo caso, R\$ 5.82 e R\$ 129.45 no terceiro caso para *dnord*=1 e *dnord*=0, respectivamente. Para valores de *gastopc2014* acima dos referenciados, o impacto assume forma decrescente tal que para valores altos do gasto, o impacto estimado é bem menor. Para fins de ilustração, tomando o caso em que *rendapc2010* é fixada no primeiro quartil e *gastopc2014* é fixado no terceiro quartil (R\$311.90), a Tabela 13 indica que os impactos estimados quando *dnord*=0 e *dnord*=1 são, respectivamente, 0.000661 e 0.000561. No entanto, ao tomar *gastopc2014* fixado em valores maiores, R\$ 700 por exemplo, os impactos passam a ser 0.000337 e 0.000251, o que representa algo em torno de 1.96 e 2.24 vezes menor. Nos casos em que *rendapc2010* é fixada no segundo e terceiro quartis, a redução no impacto para valores altos do gasto se mantém aproximadamente nas mesmas proporções.

Tanto no modelo ajustado para o ano 2006 quanto no modelo para o ano 2014, a taxa de analfabetismo se mostrou uma variável significativa e, igualmente ao que ocorreu com o modelo 2006, o coeficiente estimado desta covariável é elevado (vide Tabela 5). Logo, o comportamento dos impactos ao fixar *panalf2010* no primeiro, segundo e terceiro quartis foi investigado e os resultados são encontrados nas Tabelas 15 e 16 e na Figura 18. Nota-se que os impactos estimados quando a taxa de analfabetismo é fixada no primeiro quartil são um pouco maiores do que quando a renda é fixada, ao passo que fixando *panalf2010* no terceiro quartil os impactos estimados são ligeiramente menores. Os intervalos de confiança bootstrap ao nível de 95% de confiança indicam que todos os impactos estimados são estatisticamente diferentes de zero ao nível nominal de 5% e o comportamento dos impactos pode ser visualizado na Figura 18, por meio da qual percebe-se que não ocorre grandes mudanças em relação ao que é observado quando a covariável *rendapc2010* é fixada.

Por fim, de acordo com o modelo estimado para 2014, os resultados apontam que se o gasto *per capita* com o Programa Bolsa Família neste ano tivesse permanecido no nível em que se encontrava em 2010, considerando o cenário I (*gastopc2010* corrigido pelo IPCA), Dilma Rousseff teria obtido 49,288,870 votos, isto é, 5,212,248 votos a menos do

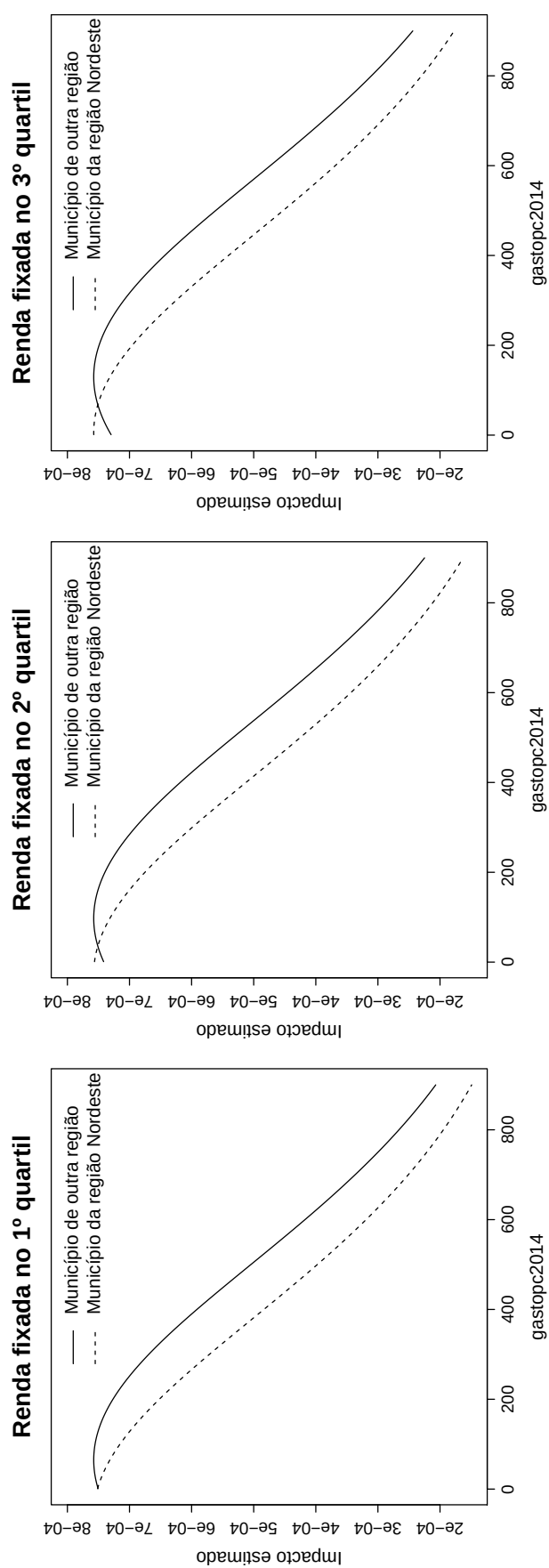


Figura 17 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2014 sobre a resposta média.

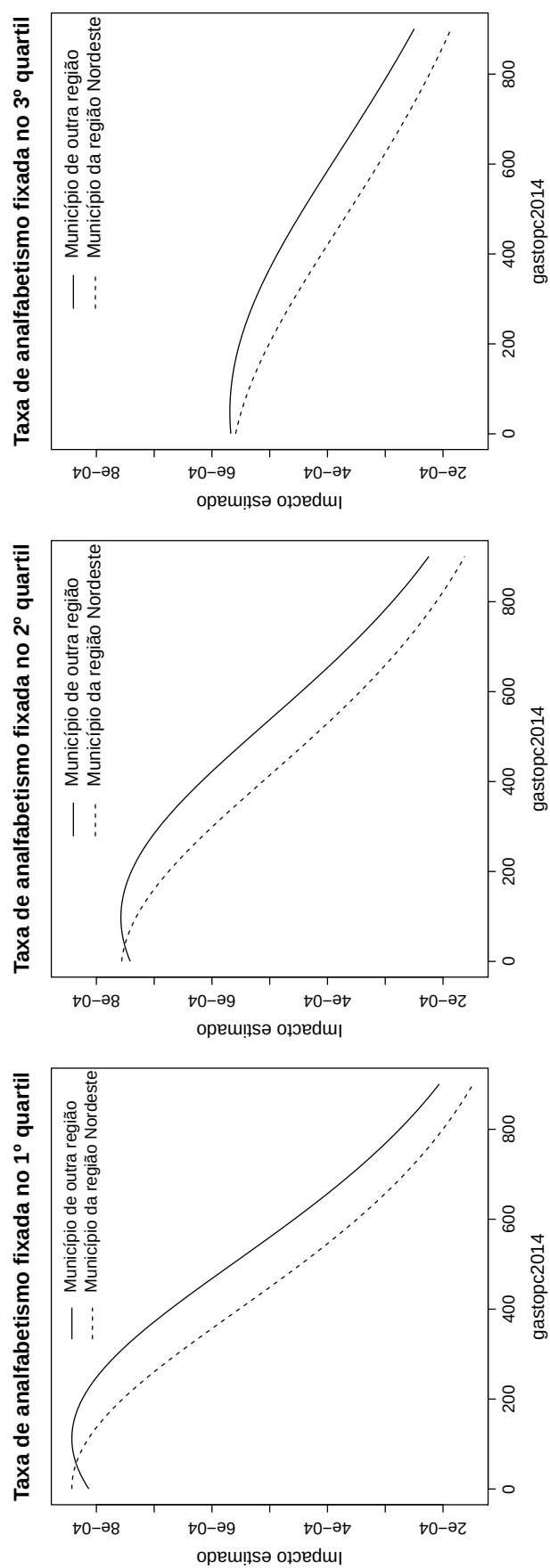


Figura 18 – Impactos estimados do gasto *per capita* nos programas assistenciais em 2014 sobre a resposta média fixando *panalf2010*.

Tabela 15 – Impactos estimados do gasto *per capita* no programa Bolsa Família em 2014 sobre a resposta média fixando *panalf2010* e seus respectivos erros-padrão obtidos via bootstrap.

<i>panalf2010</i>	<i>dnord</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	0.000840 (0.000030)	0.000837 (0.000030)	0.000754 (0.000023)
	1	0.000828 (0.000032)	0.000783 (0.000029)	0.000648 (0.000020)
Q_2	0	0.000757 (0.000026)	0.000750 (0.000025)	0.000683 (0.000020)
	1	0.000739 (0.000026)	0.000699 (0.000024)	0.000589 (0.000017)
Q_3	0	0.000568 (0.000022)	0.000559 (0.000022)	0.000521 (0.000019)
	1	0.000542 (0.000023)	0.000515 (0.000022)	0.000453 (0.000017)

Tabela 16 – Intervalos de confiança bootstrap ao nível de 95% de confiança para os impactos do gasto *per capita* no programa Bolsa Família em 2014 sobre a resposta média fixando *panalf2010*.

<i>panalf2010</i>	<i>dnord</i>	GASTO PER CAPITA		
		Q_1	Q_2	Q_3
Q_1	0	(0.000784, 0.000896)	(0.000780, 0.000893)	(0.000710, 0.000797)
	1	(0.000771, 0.000887)	(0.000730, 0.000839)	(0.000611, 0.000686)
Q_2	0	(0.000707, 0.000805)	(0.000701, 0.000798)	(0.000644, 0.000721)
	1	(0.000687, 0.000788)	(0.000652, 0.000745)	(0.000556, 0.000621)
Q_3	0	(0.000526, 0.000608)	(0.000517, 0.000600)	(0.000484, 0.000556)
	1	(0.000498, 0.000585)	(0.000473, 0.000557)	(0.000419, 0.000486)

que a quantidade conquistada por ela na eleição presidencial, cujo total foi de 54,501,118 votos. Aécio Neves, que disputou a presidência com Dilma, obteve 51,041,155 votos, o que implica uma diferença de 3,459,963 votos entre ambos. Admitindo que na votação de Dilma haveria tal redução e se todos os 5,212,248 votos perdidos pela ex-presidente se $gastopc2014 = gastopc2010$ tivessem migrado para seu oponente, este teria obtido 56,253,403 votos válidos e, portanto, segundo o modelo estimado, o resultado final da eleição sofreria alteração e Aécio Neves teria vencido a eleição presidencial ocorrida no ano 2014. Salienta-se, também, que ainda que o total de votos perdidos por Dilma tivesse se transformado em votos brancos e nulos, Aécio Neves sairia vencedor da eleição. O intervalo com 95% de confiança associado obtido via bootstrap é (4, 547, 243; 5, 889, 436). Como o número estimado de votos perdidos pertence ao intervalo, evidencia-se que tal quantidade é significativa ao nível de 95% de confiança.

Tomando o cenário II (*gastopc2010* não corrigido pelo IPCA) para análise, os resultados apontam que Dilma Rousseff teria conquistado 6,559,936 votos a menos, o que culminaria um total de 47,941,182 votos recebidos na eleição em vez dos 54,501,118 votos que obteve. Da mesma forma que no cenário I, essa quantidade alteraria o resultado final da eleição. Salienta-se que o número estimado de votos perdidos no cenário em que não foi considerada a correção na variável *gastopc2010* foi maior, com 1,347,688 votos a mais.

O intervalo de confiança associado, ao nível de confiança de 95% obtido via bootstrap, é (5, 898, 374; 7, 238, 545), sendo que o número estimado de votos perdidos pertence ao intervalo.

Ademais, buscou-se determinar o menor percentual do total de votos perdidos por Dilma Rousseff de modo que se tais votos tivessem migrado para Aécio Neves, o resultado da eleição teria sido revertido. No contexto do cenário I, verificou-se que se 33.6186% dos votos válidos perdidos por Dilma tivesse migrado para seu oponente, essa quantidade bastaria para que Aécio Neves vencesse a eleição do ano em questão. Quando considerado o cenário II, a migração de 47.2561% dos votos perdidos por Dilma seria suficiente para reverter o resultado da eleição.

5.5 Comparação entre as eleições presidenciais de 2006, 2010 e 2014

O comportamento dos impactos estimados para cada ano estudado pode ser visualizado nas Figuras 19 e 20. A Figura 19 considera o caso em que as covariáveis *dummies*, *dnord* e *DU*, assumem o valor 0 e por meio dela verifica-se que, nos três casos considerados para a renda, o impacto dos gastos com programas assistenciais sobre a proporção de votos recebidos pelos presidentes eleitos em 2006, 2010 e 2014 diminui com o passar do tempo, isto é, a quantidade de votos recebida “graças” aos benefícios do assistencialismo é menor a cada nova eleição.

Em todos os casos avaliados, valores menores de gasto implicam impactos de maiores proporções. Tomando o gasto *per capita* e a renda de cada ano fixados no primeiro quartil, a proporção estimada de votos ganhos devido ao assistencialismo é 0.2269% em 2006, uma média (entre *govpt2010*=0 e *govpt2010*=1) de 0.13295% em 2010 e 0.0757% em 2014, o que representa uma redução de 0.09395% de 2006 para 2010 e 0.05725% de 2010 para 2014, impactos de aproximadamente 1.7 vez menor de um ano para o outro. Mantendo a renda fixada no primeiro quartil e tomando o gasto fixado no terceiro quartil, a proporção de votos ganhos passa a ser de 0.1875% em 2006, uma média de 0.13935% em 2010 e 0.0661% em 2014, indicando reduções de 0.04815% de 2006 para 2010 e 0.07325% de 2010 para 2014, impactos de cerca de 1.3 e 2.1 vezes menor, respectivamente.

Contudo, conforme o valor do gasto aumenta, a proporção de votos ganhos devido ao assistencialismo diminui consideravelmente nos três anos. Isto pode ser explicado pelo fato de que o número de municípios com valores altos de gasto *per capita* é pequeno, o que implica menores proporções de votos. A saber, o terceiro quartil da variável *gastopc* de cada ano considerado é R\$ 95.06 em 2006, R\$ 166.90 em 2010 e R\$ 311.90 em 2014, o que indica que 75% dos municípios têm o gasto *per capita* registrado abaixo desses valores em seus respectivos anos.

Na Figura 20 observa-se o caso em que as covariáveis *dummies*, *dnord* e *DU*, assumem o valor 1. As curvas de impacto mostram que, nos três casos considerados para a renda, o impacto dos gastos com programas assistenciais sobre a proporção de votos recebidos pela presidente eleita em 2010 difere bastante do comportamento apresentado pelos anos 2006 e 2014 em dois aspectos: os impactos estimados são relativamente menores, registrando valores negativos quando a renda é fixada no segundo e terceiro quartis e, para os mesmos quartis, o impacto é crescente à medida que o gasto *per capita* aumenta, ao passo que em todos os quartis dos outros dois anos avaliados tem-se um comportamento decrescente. Os impactos estimados para 2006 e 2014 comportam-se de forma similar, ambos com maior efeito para valores pequenos de *gastopc* e efeito reduzido para os gastos mais elevados, o que também é percebido no caso em que as *dummies* assumem valor 0.

Por fim, averigua-se que a diferença entre os impactos registrados nos anos 2006 e 2014 para os menores valores do gasto *per capita* diminuem conforme a renda aumenta. A saber, tomando a renda fixada no primeiro quartil e a média entre os valores de gasto do primeiro, segundo e terceiro quartis, o impacto em 2006 representa 0.2101% de votos ganhos devido ao assistencialismo, enquanto que em 2014 esse impacto representa 0.0720%, isto é, há uma redução de 0.1381%, o que caracteriza um impacto aproximadamente 2.9 vezes menor. Todavia, se a renda é tomada no terceiro quartil, em 2006 o impacto é de 0.1044% e em 2014 cerca de 0.0737%. Neste caso, a redução é de apenas 0.0307%, ou 1.4 vez menor.

A Tabela 17 apresenta os resultados gerais das eleições para os três anos tomados em estudo, incluindo os impactos estimados no que tange à quantidade de votos perdidos por cada presidente se os gastos *per capita* com os programas assistenciais tivessem permanecido no nível em que se encontravam nas eleições anteriores. É possível verificar com facilidade que, enquanto nas eleições de 2006 e 2010 o resultado final não sofreria alteração, isto é, Lula e Dilma continuariam eleitos mesmo se perdessem a quantidade de votos prevista pelos modelos estimados, em 2014 a situação seria outra. O número estimado de votos perdidos, apesar de não ser uma quantidade muito expressiva, bastaria para fazer com que o rumo da eleição tivesse sido outro e Aécio Neves teria saído vencedor da disputa ao cargo presidencial.

Em relação às variáveis selecionadas para cada um dos modelos definidos pelas Equações (3.1), (3.2) e (3.3), nota-se que o gasto *per capita* nos programas assistenciais e a renda *per capita* estão presentes nos três modelos, o que mostra a sua relevância e forte influência sobre a proporção de votos válidos no segundo turno das três eleições.

Canêdo-Pinheiro (2015) retrata a mudança na base eleitoral do presidente Lula que ocorreu entre as votações de 2002 e 2006, as duas eleições nas quais se elegeu. Na primeira, seu eleitorado concentrava-se nas regiões mais desenvolvidas do Brasil, enquanto em 2006 as regiões de menor desenvolvimento foram as responsáveis pelo melhor desempenho de Lula.

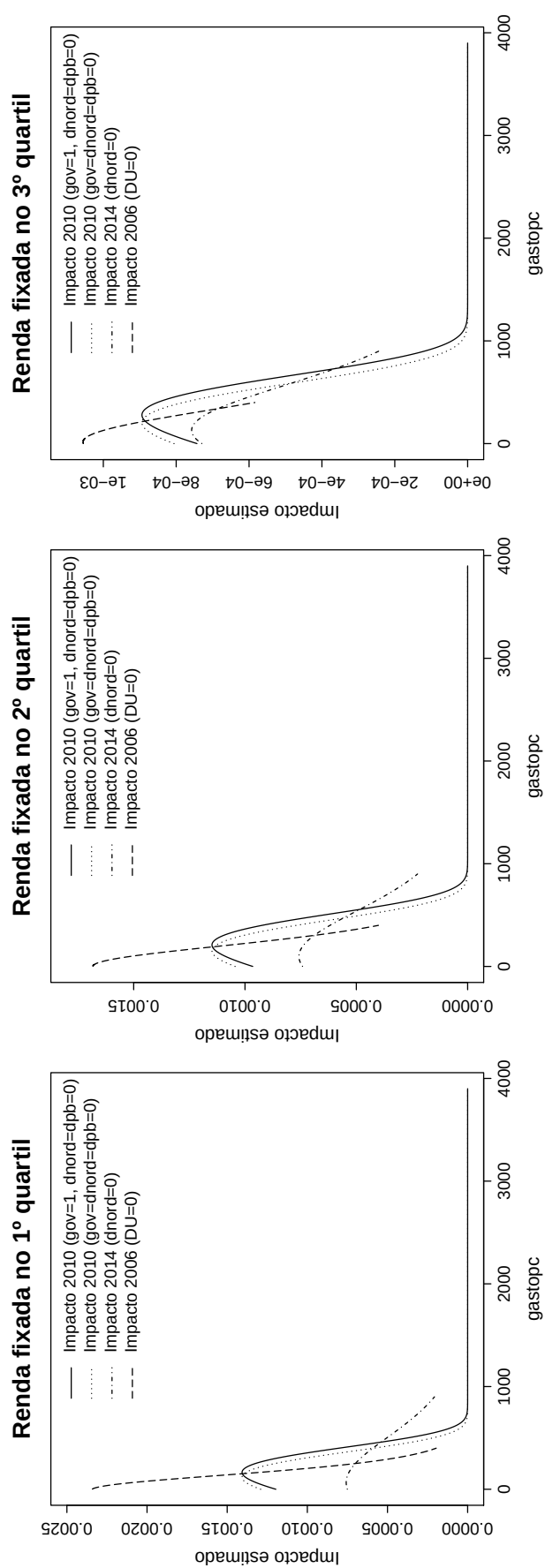


Figura 19 – Impactos estimados do gasto *per capita* nos programas assistenciais sobre a resposta média quando $dnord=0$ e $DU=0$.

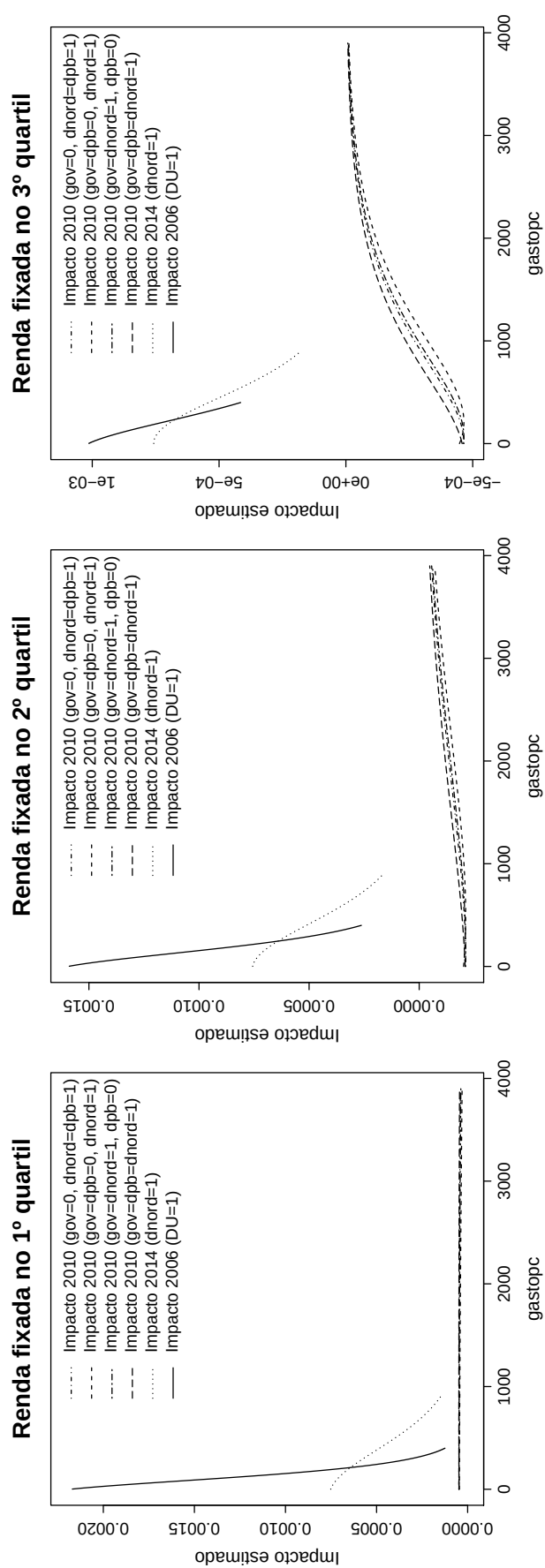


Figura 20 – Impactos estimados do gasto *per capita* nos programas assistenciais sobre a resposta média quando $dnord=1$ e $DU=1$.

Tabela 17 – Resultados das eleições, reais e estimados, para os anos 2006, 2010 e 2014.

		2006	2010	2014
Nº de votos do presidente eleito	–	58, 295, 042	55, 752, 529	54, 501, 118
Nº de votos do adversário	–	37, 543, 178	43, 711, 388	51, 041, 155
Diferença no nº de votos	–	20, 751, 864	12, 041, 141	3, 459, 963
Nº estimado de votos perdidos	<i>Cenário I</i> ^a	5, 582, 385	3, 676, 583	5, 212, 248
	<i>Cenário II</i> ^b	5, 945, 638	3, 964, 046	6, 559, 936
Nº de votos do presidente considerando a perda estimada dos votos	<i>Cenário I</i> ^a	52, 712, 657	52, 075, 946	49, 288, 870
	<i>Cenário II</i> ^b	52, 349, 404	51, 788, 483	47, 941, 182
Nº de votos do adversário caso ele tivesse recebido os votos perdidos	<i>Cenário I</i> ^a	43, 125, 563	47, 387, 971	56, 253, 403
	<i>Cenário II</i> ^b	43, 488, 816	47, 675, 434	57, 601, 091

^a Gasto *per capita* de 4 anos antes inflacionado pelo IPCA;

^b Gasto *per capita* de 4 anos antes sem correção pelo IPCA.

Segundo o autor, essa melhora ocorreu, principalmente, na grande maioria dos municípios das regiões Norte e Nordeste. Logo, percebe-se que a variável *dummy dnord*, que considera se o município pertence ou não à região Nordeste, foi significativa nos ajustes realizados para as duas últimas eleições, 2010 e 2014, situação que reflete o fato de o Nordeste ser uma das regiões mais contempladas com os benefícios dos programas assistenciais, visto que é uma das regiões brasileiras menos favorecidas social e economicamente, concentrando os maiores índices de pobreza do país como também elevadas taxas de analfabetismo, estas presentes nos modelos 2006 e 2014.

A covariável *ppent*, que se mostrou significativa para explicar a proporção de votos válidos recebidos por Dilma Rousseff nas eleições dos anos 2010 e 2014, foi considerada com base em Zucco (2008), ao salientar que, na época, o crescimento das igrejas pentecostais e a sua perceptível capacidade de organização política tornou-se alvo de debate no Brasil por vários anos e que o fato do presidente ter desfrutado de boas relações com as igrejas pentecostais sugere que elas possam ter investido seu capital eleitoral no ex-presidente.

Além disso, as covariáveis *DU* e *rur2010*, que consideram a população rural, presentes nos modelos para os anos 2006 e 2014, respectivamente, contribuem positivamente com a proporção de votos válidos e as *dummies govpt2010* e *dpb*, presentes no modelo definido para o ano 2010 e não consideradas nos demais ajustes, exercem efeito contrário sobre a proporção de votos válidos do segundo turno da eleição presidencial de 2010.

6 Considerações Finais e Trabalhos Futuros

A avaliação da magnitude dos impactos exercidos pelo aumento no gasto *per capita* com programas assistenciais, especialmente o Programa Bolsa Família, sobre o resultado das três últimas eleições presidenciais ocorridas nos anos 2006, 2010 e 2014, proposta deste trabalho, foi feita por meio da utilização do modelo de regressão beta com dispersão variável, uma vez que a variável de interesse é a proporção de votos válidos recebidos pelos ex-presidentes Lula (2006) e Dilma Rousseff (2010 e 2014) no segundo turno das três eleições consideradas. Desse modo, obteve-se um submodelo para a média e outro para a dispersão da variável resposta de cada ano.

A partir do modelo ajustado e corretamente especificado, conforme atestou a análise de diagnóstico e adequabilidade, para cada eleição obteve-se uma previsão do número de votos perdidos caso o gasto *per capita* com assistencialismo não tivesse aumentado nos últimos 4 anos. Tal previsão foi estimada com base em dois cenários: gasto *per capita* no nível da eleição anterior corrigido pelo IPCA e gasto *per capita* no nível da eleição anterior sem correção pelo IPCA. Quando considerado o ano 2006, previu-se um total de 5,582,385 votos perdidos pelo ex-presidente Lula quando o gasto *per capita* de 2002 foi corrigido e 5,945,638 votos a menos quando o gasto não foi corrigido. Assim sendo, mesmo ocorrendo a redução prevista, independente da variável *gastopc* ser corrigida ou não pelo IPCA, o resultado da eleição não sofreria alteração e Lula seria reeleito.

Já em 2010, a previsão para o número de votos perdidos pela ex-presidente Dilma Rousseff foi de 3,676,583 votos no caso em que o gasto *per capita* de 2006 foi corrigido pelo IPCA e, quando não corrigido, um total de 3,964,046 votos seriam perdidos. Em qualquer dos cenários, a redução na votação final da ex-presidente não seria significativa a ponto de alterar o resultado da eleição.

Finalmente, quando o ano 2014 foi tomado para avaliação, a quantidade de votos perdidos pela ex-presidente Dilma Rousseff prevista pelo modelo foi de 5,212,248 votos no cenário em que o gasto *per capita* de 2010 recebeu a correção pelo IPCA e 6,559,936 votos perdidos no segundo cenário. Ocorrendo uma redução nesses níveis na votação final da eleição, seja em qualquer dos cenários, Dilma não conseguiria um número suficiente de votos para se reeleger e, portanto, Aécio Neves teria se tornado o presidente do Brasil em 2014. Diante dos resultados acima descritos, ressalta-se que em todos os anos a previsão do número de votos perdidos foi maior quando o gasto *per capita* não foi corrigido pelo IPCA.

A comparação entre os impactos estimados dos gastos com programas assistenciais sobre a proporção de votos válidos recebidos pelos presidentes eleitos no segundo turno

das três eleições consideradas foi realizada sob dois aspectos: quando as covariáveis *dnord* e *DU*, variáveis *dummies*, assumem valor 0 e quando as mesmas *dummies* assumem valor 1. Verificou-se que o comportamento dos impactos é semelhante entre os diferentes anos nas duas situações acima especificadas, salvo o caso particular dos impactos relativos ao ano 2010 com as *dummies* valendo 1.

De modo geral, o que se constatou é que o impacto dos gastos com programas assistenciais sobre a proporção de votos recebidos pelos presidentes eleitos em 2006, 2010 e 2014 diminui com o passar do tempo, isto é, a quantidade de votos recebida “graças” aos benefícios do assistencialismo é menor a cada nova eleição e, ainda, valores menores de gasto implicam impactos de maiores proporções. Ademais, notou-se que à medida que o valor do gasto aumenta, a proporção de votos ganhos devido ao assistencialismo diminui nos três anos, o que pode ser explicado pelo fato de que o número de municípios com valores elevados de gasto *per capita* é pequeno, implicando menores proporções de votos.

Em relação à situação particular do ano 2010, que diferiu das demais, além de registrar valores de impacto relativamente menores, atingindo marcas negativas em alguns casos, também o impacto revelou-se crescente à medida que o gasto *per capita* aumenta, enquanto um comportamento decrescente foi registrado para todas as outras situações avaliadas.

Nesse contexto, evidencia-se que os gastos com assistencialismo, principalmente com o Programa Bolsa Família, foco desta pesquisa, exercem influência sobre a votação dos candidatos à presidência. Em períodos mais remotos a influência era maior, isto é, acarretava maior quantidade de votos favoráveis aos presidentes, entretanto, com o passar dos anos, este cenário foi enfraquecendo, tanto que o resultado da eleição mais recente poderia ter sido alterado com base no número de votos perdidos caso tais gastos não tivessem aumentado nos últimos 4 anos.

Para pesquisas futuras, objetiva-se desenvolver uma análise longitudinal dos dados aqui utilizados.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974.
- ALMEIDA, P. M.; SOUZA, T. C. de. Estimativas de votos da presidente Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do Bolsa Família. *Ciência & Natura*, v. 37, n. 1, p. 12–22, 2015.
- ATKINSON, A. C. *Plots, transformations and regression: an introduction to graphical methods of diagnostic regression analysis*. New York: Oxford University Press, 1985.
- BRASIL. Ministério do Desenvolvimento Social e Combate à Fome - MDS. *Bolsa Família: Transferência de Renda e Apoio à Família no Acesso à Saúde, à Educação e à Assistência Social*. Brasília, 2015.
- CANÊDO-PINHEIRO, M. Bolsa Família ou desempenho da economia? Determinantes da reeleição de Lula em 2006. *Economia Aplicada*, v. 19, n. 1, p. 31–61, 2015.
- COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, v. 19, n. 1, p. 15–18, 1977.
- CORRÊA, D. S. Conditional cash transfer programs, the economy, and presidential elections in Latin America. *Latin American Research Review*, v. 50, n. 2, p. 63–85, 2015.
- CORRÊA, D. S. Os custos eleitorais do Bolsa família: Reavaliando seu impacto sobre a eleição presidencial de 2006. *Opinião Pública*, v. 21, n. 3, p. 514–534, 2015.
- CORRÊA, D. S. Transfêrencia de renda e eleições: A trajetória latino-americana. *Revista Brasileira de Ciências Sociais*, v. 31, n. 90, 2016.
- ESPINHEIRA, P. L.; FERRARI, S. L.; CRIBARI-NETO, F. Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, v. 52, n. 9, p. 4417–4431, 2008.
- ESPINHEIRA, P. L.; FERRARI, S. L.; CRIBARI-NETO, F. On beta regression residuals. *Journal of Applied Statistics*, v. 35, n. 4, p. 407–419, 2008.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004.
- FERRARI, S. L.; ESPINHEIRA, P. L.; CRIBARI-NETO, F. Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, v. 65, n. 3, p. 337–351, 2011.
- JONES, H. More education, better jobs? A critical review of CCTs and Brazil's Bolsa Família Programme for long-term poverty reduction. *Social Policy and Society*, v. 15, n. 03, p. 465–478, 2016.
- LIMA, L. B. de. *Um teste de especificação correta em modelos de regressão beta*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2007.
- MARQUES, R. M. Políticas de transferência de renda no Brasil e na Argentina. *Brazilian Journal of Political Economy/Revista de Economia Política*, v. 33, n. 2, 2013.

- MARQUES, R. M. et al. Discutindo o papel do Programa Bolsa Família na decisão das eleições presidenciais brasileiras de 2006. *Revista de Economia Política*, v. 29, n. 1, p. 114–132, 2009.
- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. London: Chapman and Hall, 1989.
- NAGELKERKE, N. J. A note on a general definition of the coefficient of determination. *Biometrika*, v. 78, n. 3, p. 691–692, 1991.
- PAVÃO, N. Conditional Cash Transfer Programs and electoral accountability: Evidence from Latin America. *Latin American Politics and Society*, v. 58, n. 2, p. 74–99, 2016.
- PEREIRA, A. E. G. et al. A eleição de Dilma em 2010 e seus determinantes: evidências empíricas do Programa Bolsa Família. *Análise Econômica*, v. 33, n. 64, p. 111–142, 2015.
- PRESS, W. H. et al. *Numerical recipes in C: the art of scientific computing*. [S.l.]: Cambridge University Press, 1992.
- RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B*, v. 31, n. 2, p. 350–371, 1969.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 54, n. 3, p. 507–554, 2005.
- SHIKIDA, C. D. et al. “It is the economy, companheiro!”: an empirical analysis of Lula’s re-election based on municipal data. *Economics Bulletin*, v. 29, n. 2, p. 976–991, 2009.
- SILVA, J. G. d.; GROSSI, M. E. D.; FRANÇA, C. G. d. (Orgs.). *Fome Zero: a experiência brasileira*. Brasília: Ministério do Desenvolvimento Agrário - MDA, 2010.
- SILVA, M. O. da. Os programas de transferência de renda na política social brasileira: seu desenvolvimento, possibilidades e limites. *Revista de Políticas Públicas*, v. 8, n. 2, 2004.
- SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, v. 54, n. 2, p. 348–366, 2010.
- SOUZA, T. C. de. *Ensaio sobre modelos de regressão com dispersão variável*. Tese (Doutorado) — Universidade Federal de Pernambuco, 2011.
- SOUZA, T. C. de; CRIBARI-NETO, F. Uma estimativa do impacto eleitoral do Programa Bolsa-Família. *Revista Brasileira de Biometria*, v. 31, n. 1, p. 79–103, 2013.
- STASINOPOULOS, D. M.; RIGBY, R. A.; AKANTZILIOTOU, C. *Instructions on how to use the gamlss package in R*: Second edition. [S.l.], 2008.
- TAKEUCHI, K. Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematical Sciences)*, v. 153, n. 1, p. 12–18, 1976.
- VENABLES, W. N.; SMITH, D. M. *An introduction to R: Notes on R: A programming environment for data analysis and graphics*. [S.l.], 2016.

WEI, B.-C.; HU, Y.-Q.; FUNG, W.-K. Generalized leverage and its applications. *Scandinavian Journal of Statistics*, v. 25, n. 1, p. 25–37, 1998.

WWP. Brazil Learning Initiative for a World Without Poverty. *A estruturação de monitoramento e avaliação de políticas sociais no Brasil: A experiência da Secretaria de Avaliação e Gestão da Informação (SAGI)*. Brasília, 2015.

ZUCCO, C. The president's 'new' constituency: Lula and the pragmatic vote in Brazil's 2006 presidential elections. *Journal of Latin American Studies*, v. 40, n. 1, p. 29–49, 2008.

ZUCCO, C. Poor voters vs. poor places: persisting patterns in presidential elections in Brazil. In: *Workshop on Political Consequences of Declining Inequality in Brazil*. Oxford: [s.n.], 2010.

ZUCCO, C. When payouts pay off: Conditional cash transfers and voting behavior in Brazil 2002–10. *American Journal of Political Science*, v. 57, n. 4, p. 810–822, 2013.

ZUCCO, C. The impacts of conditional cash transfers in four presidential elections (2002–2014). *Brazilian Political Science Review*, v. 9, n. 1, p. 135–149, 2015.