

Ruan da Silva Vianna

**Modelos Dinâmicos Hierárquicos
Generalizados: Uma aplicação com dados
eleitorais brasileiros**

Rio de Janeiro - RJ, Brasil

19 de agosto de 2021



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Ruan da Silva Vianna

Modelos Dinâmicos Hierárquicos Generalizados: Uma aplicação com dados eleitorais brasileiros

Dissertação de Mestrado

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Orientadora: Prof^a. Dr^a Flávia Maria P. Ferreira Landim - UFRJ

Co-Orientadores: Prof^a. Dr^a. Marina Silva Paez - UFRJ
Prof. Dr. Vinícius Pinheiro Israel - UNIRIO

Rio de Janeiro - RJ, Brasil

19 de agosto de 2021



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Ruan da Silva Vianna

**Modelos Dinâmicos Hierárquicos
Generalizados: Uma aplicação com dados
eleitorais brasileiros**

Dissertação de Mestrado sob o título “*Modelos Dinâmicos Hierárquicos Generalizados: Uma aplicação com dados eleitorais brasileiros*”, defendida por Ruan da Silva Vianna e aprovada em 19 de agosto de 2021, na cidade do Rio de Janeiro, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof^a. Dr^a. Marina Silva Paez

Departamento de Métodos Estatísticos - UFRJ

Prof^o. Dr^o. Carlos Antonio Abanto-Valle

Departamento de Métodos Estatísticos - UFRJ

Prof^o. Dr^o. Steven Dutt Ross

Departamento de Métodos Quantitativos - UNIRIO

Rio de Janeiro, 19 de agosto de 2021

Resumo

No contexto do mundo moderno, o qual a globalização já é uma realidade, a política possui um papel fundamental nas relações e diálogos entre pessoas e também entre países. O Brasil possui um sistema multipartidário, que hoje é o quarto maior em termos de eleitorado, apenas atrás da Índia, Estados Unidos e Indonésia, respectivamente. Os partidos brasileiros são divididos em partidos trabalhistas, socialistas, comunistas, ecológicos, democráticos, liberais, entre outros.

O interesse dessa dissertação é utilizar uma das classes de modelos dinâmicos, mais especificamente, modelos lineares dinâmicos hierárquicos generalizados. O modelo utilizado foi proposto por um grupo de pesquisadores da Alemanha em 2017, a fim de prever o resultado da eleição presidencial do mesmo ano. O modelo é especificado para prever resultados de eleições em cenários multipartidários como o caso brasileiro. Ele consiste na combinação de outros dois modelos, sendo assim, feito em duas etapas. A primeira etapa é responsável pela previsão de votos compartilhados de cada partido com base em resultados de eleições passadas e de variáveis explicativas. A segunda etapa é responsável por incluir informações para a previsão da intenção de votos de cada partido, candidato ou grupo (que depende da modelagem especificada), a partir de pesquisas eleitorais que são divulgadas durante o período eleitoral.

O modelo proposto foi considerado para ser aplicado a um conjunto de dados sobre as eleições presidenciais brasileiras de 2018, especificamente para dados referente ao primeiro turno. Uma reorganização desse modelo foi proposta nesta dissertação para esse conjunto de dados. Após ajustes, os resultados são apresentados e é possível verificar que o modelo ajustado foi eficiente em prever o resultado final da eleição de 2018 de acordo com os pressupostos apresentados.

Palavras-chave: **Modelos dinâmicos hierárquicos generalizados, Inferência Bayesiana, MCMC, Monte Carlo Hamiltoniano (HMC), previsão em eleições multipartidárias.**

Abstract

In the context of the modern world, in which globalization is already a reality, politics has a fundamental role in relations and dialogues between people and also between countries. Brazil has a multiparty system, which today is the fourth largest in terms of electorate, just behind India, the United States and Indonesia. Brazilian parties are divided into labor, socialist, communist, ecological, democratic, liberal, among others.

The interest of this dissertation is to use one of the classes of dynamic models, more specifically, generalized hierarchical dynamic linear models. The model that we used was proposed by a group of researchers from Germany in 2017, in order to predict the result of the presidential election of the same year. The model is specified to predict election results in multiparty scenarios such as the Brazilian case. It consists of a combination of two other models, thus being done in two steps. The first stage is responsible for predicting the shared votes of each party based on the results of past elections and explanatory variables. The second stage is responsible for including information to predict the intention of votes for each party, candidate or group (which depends on the specified modeling), based on electoral surveys that are released during the electoral period.

The proposed model was applied to the dataset of the 2018 Brazilian presidential elections, specifically for data referring to the first round. A reorganization of this model was proposed in this dissertation for this dataset. After adjustments, the results are presented and it is possible to verify that the applied model was efficient in predicting the final result of the 2018 election according to the presented assumptions.

Keywords: Generalized hierarchical dynamic models, Bayesian Inference, MCMC, Monte Carlo Hamiltoniano (HMC), multiparty advance prediction .

Dedicatória

Dedico esta, como todas as outras minhas conquistas a minha família.

Agradecimentos

Primeiramente quero agradecer a Deus por estar ao meu lado em todo o tempo, por ter me escutado e compartilhado todas as minhas angústias e aflições aos logos desses anos. Agradeço a minha família que é a minha base, em particular aos meus pais que desde muito pequeno me incentivavam e me mostraram o valor da educação e a minha esposa por todo o suporte durante esses anos, sempre me apoiando e incentivando com carinho e amor.

Agradeço aos amigos que fiz, tanto na graduação quanto no mestrado que fizeram com que esses períodos fossem mais alegres e as jornadas mais leves. Vocês foram essenciais na minha trajetória até aqui.

Agradeço aos meus orientadores, Marina, Flávia e Vinícius por toda ajuda e suporte ao escrever esta dissertação, por toda paciência que tiveram comigo durante esse período, por todas as conversas e incentivos. Deixo a vocês todo o meu carinho, respeito e admiração.

Por fim, agradeço a toda a população brasileira e a CAPES que me auxiliaram durante esse período com uma bolsa de pesquisa. Aqui deixo meu agradecimento por todo o suporte financeiro, e não menos importante a UFRJ e todo o departamento da Pós-Graduação em Estatística por ter me aceitado e me acolhido.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
1.1	Modelagem proposta	p. 13
1.2	Estrutura da dissertação	p. 15
2	Conceitos preliminares	p. 16
2.1	Inferência Bayesiana	p. 16
2.1.1	Teorema de Bayes	p. 17
2.1.2	Predição	p. 19
2.2	Métodos de Monte Carlo via Cadeias de Markov (MCMC)	p. 20
2.2.1	Monte Carlo via Cadeias de Markov (MCMC)	p. 20
2.2.2	Algoritmo Monte Carlo Hamiltoniano (HMC)	p. 21
2.3	Modelos dinâmicos: Uma introdução	p. 25
2.3.1	Modelos Lineares Dinâmicos	p. 25
2.3.2	Modelos Lineares Dinâmicos Hierárquicos	p. 27
2.3.3	Modelos Lineares Dinâmicos Hierárquicos Generalizados	p. 28
3	O modelo	p. 30
4	Implementação do modelo completo e resultados	p. 36
4.1	Nova República: Um resumo	p. 36

4.2	Aplicação para as Eleições Executivas de 2018	p. 39
4.2.1	Aplicação do modelo completo	p. 46
5	Conclusões	p. 58
	Referências	p. 60
	Apêndice 1 – Distribuições	p. 63
1.1	Distribuição Multinomial	p. 63
1.2	Distribuição de Dirichlet	p. 63
	Apêndice 2 – Divisão Ideológica: Eleições 1989 - 2018	p. 65
	Apêndice 3 – Códigos	p. 69
3.1	Modelo A	p. 69
3.2	Modelo completo	p. 71

Lista de Figuras

1	Porcentagem de votos da eleições presidenciais passadas	p. 39
2	Taxas médias em proporção das variáveis explicativas mencionadas, PIB, Inflação e Taxa de Desemprego em anos eleitorais de 2002 à 2018. . . .	p. 41
3	Evolução dos candidatos à presidência.	p. 43
4	Evolução dos votos em outros candidatos.	p. 44
5	Evolução dos votos em brancos/nulos.	p. 44
6	Evolução de todas as séries durante o 1º turno.	p. 45
7	Previsão das eleições de 2018, dois dias antes da eleição.	p. 47
8	Desenvolvimento das previsões do compartilhamento de votos do modelo completo ao longo do tempo para a Eleição Federal de 2018.	p. 49
9	Relação entre a variável Votos Compartilhados (%) vs variáveis preditoras.	p. 50
10	Relação entre a variável Votos Compartilhados (%) vs variáveis preditoras.	p. 52
11	Evolução do coeficiente β_e relativo a cada variável explicativa durante as últimas 7 eleições.	p. 53
12	Predição dos Votos Compartilhados para a direita durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.	p. 54
13	Predição dos Votos Compartilhados para a esquerda durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.	p. 56
14	Predição dos Votos Compartilhados para o centro durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.	p. 57

Lista de Tabelas

1	Classificação partidária segundo a ideologia	p. 14
2	Total de pessoas entrevistadas por instituto	p. 42
3	Porcentagem real para cada grupo, estimativa pontual para cada um dos grupos e desvio padrão.	p. 48
4	Llimites do intervalo de credibilidade de 83% e de 95% respectivamente.	p. 48
5	Classificação das principais coligações/partidos da eleição de 1989 . . .	p. 65
6	Classificação das principais coligações/partidos da eleição de 1994 . . .	p. 66
7	Classificação das principais coligações/partidos da eleição de 1998 . . .	p. 66
8	Classificação das principais coligações/partidos da eleição de 2002 . . .	p. 66
9	Classificação das principais coligações/partidos da eleição de 2006 . . .	p. 67
10	Classificação das principais coligações/partidos da eleição de 2010 . . .	p. 67
11	Classificação das principais coligações/partidos da eleição de 2014 . . .	p. 68

1 Introdução

O Brasil é um dos maiores países do mundo, e não se destaca somente pelo seu tamanho territorial, ou pela quantidade de habitantes, hoje cerca de 210 milhões, segundo o IBGE - Instituto Brasileiro de Geografia e Estatística, mas também pelo seu multiculturalismo. Essa característica da população brasileira estende-se desde a raça, religião até a política, que é o foco de interesse dessa dissertação.

Segundo Bizzaro e Coppedge (2017), um dos grandes desafios da ciência política é medir e analisar corretamente os fenômenos de natureza complexa e multifacetada que compõem os objetos da disciplina. Os principais dentre esses objetos são a natureza, a causa e as consequências que os regimes políticos sofrem devido aos desafios de mensuração e análise, e um bom modelo deve ser capaz de fazer as previsões mesmo introduzindo essas incertezas no mesmo. Os regimes políticos são multifacetados, compostos por várias práticas, instituições e atores. Eles são também o resultado do desenvolvimento histórico e das disputas políticas pelas quais se constroem o Estado Nacional e as instituições políticas de um país.

O sistema eleitoral brasileiro é composto de duas formas de eleição: a majoritária e a proporcional. Na majoritária, são eleitos os prefeitos, os governadores, o presidente da República e os senadores, enquanto que na outra forma, são eleitos os deputados e vereadores. O Brasil possui um sistema multipartidário que segundo Reis (2015) é a exagerada proliferação de agremiações partidárias ou, ainda, o sistema pluripartidarista extremado, a utilização desenfreada do princípio do pluripartidarismo insculpido no art. 17 da Constituição Federal. Os partidos brasileiros são divididos em partidos trabalhistas, socialistas ou comunistas, ecológicos, democráticos, liberais, entre outros. Em termos de população, o Brasil é a quarta maior democracia do mundo, ficando atrás da Índia, Estados Unidos da América e Indonésia.

As pesquisas de opinião são úteis em diversos cenários e na política são importantes para que os candidatos possam criar um plano de governo que atenda a demanda e a

necessidade do seu eleitorado. As pesquisas geralmente são feitas por empresas especializadas, chamadas de institutos de pesquisas e no âmbito político, tornam-se pesquisas eleitorais.

Quando as pesquisas são divulgadas nos canais de comunicação, a população é informada sobre uma possível tendência do eleitorado, naquele determinado momento de tempo, permitindo que cada eleitor possa usar essa informação na escolha do seu candidato. Além disso, as pesquisas também afetam os candidatos e as suas campanhas, seja pela confiança e exposição na mídia adquirida pelos que estão na liderança, ou pela desmotivação daqueles que não têm chances reais de ganhar a eleição. Maiores detalhes podem ser vistos em El-Dash (2010).

Como mencionado anteriormente, o Brasil é um país com grande diversidade partidária e o interesse dessa dissertação é implementar um modelo que será discutido mais a frente para modelar os votos compartilhados de cada partido ou candidato através de distribuições multivariadas como a Dirichlet ou multinomial. Pode-se definir como votos compartilhados a proporção de votos válidos de cada um dos grupos ou candidatos que participam do processo eleitoral.

A constante mudança de partidos e a criação de novos partidos no cenário nacional fazem que mesmo um modelo que tenha como principal objetivo prever resultados para uma eleição multipartidária apresente problemas, uma vez que um dos requisitos para a criação destes modelos é obter informações sobre esses partidos/candidatos nas eleições anteriores. A fim de contornar esse problema, nesta dissertação os partidos serão divididos de acordo com dois critérios distintos: cunho ideológico (direita, esquerda e centro) e a situação, em que tenta-se reeleger o governo atual (incumbente), o não incumbente e a terceira via.

É necessário para a implementação do modelo proposto, segundo as abordagens mencionadas, que seja feita uma classificação (seguindo algum critério estabelecido) dos partidos políticos brasileiros. Há diversos artigos que sugerem uma forma como é feita essa classificação como Tarouco e Madeira (2015) que sugerem como classificação uma escala que varia entre 1 e 20, em que os extremos indicam que o partido é de extrema esquerda (com valor 1) e de extrema direita (com valor 20). Essa escala é conhecida como escala de Wiesehomeier e Benoit elaborada em 2007. No artigo de Maciel, Alarcon e Gimenes (2017) é apresentada a escala mencionada anteriormente e também a escala da *Experts Surveys* que varia entre 1 e 7, em que os extremos representam extrema esquerda e extrema direita respectivamente. Outros artigos mencionam outras formas de classificação

partidária como Israel e Bachini (2018).

Como proposto por Bolognesi, Ribeiro e Codato (2020), há alguns caminhos para a classificação partidária. Em seu artigo, os autores aplicaram um questionário para 519 cientistas políticos, residentes no país e no exterior, a fim de realizar essa classificação. A pesquisa foi realizada em julho de 2018, antes da campanha eleitoral daquele ano. Uma das questões levantadas pelo grupo é que geralmente essas pesquisas não são feitas por brasileiros, o que pode acarretar e desconsiderar as especificidades do Brasil. Outro método de classificação é perguntar aos parlamentares como eles classificam seus próprios partidos e os partidos dos adversários na escala esquerda-centro-direita. Outra opção é analisar a composição social do partido, quem são seus quadros e o perfil dos seus eleitores.

Com base nos artigos mencionados sobre classificação partidária, foi proposta uma classificação para esta dissertação que é apresentada na Tabela 1. Ela ilustra uma possível forma de classificação dos partidos segundo sua ideologia.

São considerados apenas três grupos neste trabalho: direita, esquerda e centro. A esquerda é composta dos grupos de extrema esquerda e esquerda, o centro é composto da centro esquerda, centro e centro direita e por fim, a direita que é composta pelos partidos de direita e extrema direita.

1.1 Modelagem proposta

O interesse desta dissertação é implementar o modelo proposto por Stoetzer, Marcel, Neunhoeffers e Sternberg (2018), que será chamado de modelo completo daqui em diante e que tem como objetivo prever o resultado de uma certa eleição envolvendo múltiplos partidos/candidatos, ou seja, modelar os votos compartilhados da eleição de interesse.

O modelo completo, que pode ser utilizado em diferentes cenários, se baseia na ideia de que três fatores principais preveem os resultados de uma dada eleição, são eles: apego partidário a longo prazo, dinâmica da campanha em curto prazo, e características institucionais conforme descrito por Stoetzer, Marcel, Neunhoeffers e Sternberg (2018), o que faz com que esse modelo possa ser utilizado em diferentes cenários.

O modelo de Stoetzer, Marcel, Neunhoeffers e Sternberg (2018) provém de uma combinação de outros dois modelos: o primeiro deles é o modelo baseado em fundamentos, que será chamado de modelo A, e o segundo é um modelo baseado em pesquisas eleitorais, que chamaremos de modelo B. O modelo A, proposto por Norpoth e Gschwend (2010) e

Tabela 1: Classificação partidária segundo a ideologia

Partido	Classificação
PSTU	Esquerda
PCO	Esquerda
PCB	Esquerda
PSOL	Esquerda
PCdoB	Esquerda
PT	Esquerda
PDT	Centro
PSB	Centro
REDE	Centro
CIDANANIA	Centro
PV	Centro
AVANTE	Centro
SDD	Centro
PMN	Centro
PMB	Centro
MDB	Centro
DEM	Centro
PTB	Direita
PSDB	Direita
PSD	Direita
PODEMOS	Direita
PSL	Direita
PRTB	Direita
PROS	Direita
PL	Direita
PRB	Direita
PTC	Direita
NOVO	Direita
PROGRESSISTAS	Direita
PATRIOTA	Direita

Magalhães, Aguiar-Conraria e Lewis-Beck (2012), fornece uma previsão para a proporção de votos válidos de cada partido, desde que o mesmo seja relevante bem antes do início da campanha eleitoral. Essa proposta considera que para prever o resultado de eleições multipartidárias é necessário uma combinação entre os resultados das eleições passadas \mathbf{V} , uma matriz de preditores \mathbf{X} e um vetor de parâmetros $\boldsymbol{\theta}$ que faz o *link* entre \mathbf{V} e \mathbf{X} . Através da distribuição preditiva a posteriori ¹ são fornecidos os resultados futuros \mathbf{v}_E dados os preditores \mathbf{x}_E .

¹A distribuição preditiva a posteriori é a distribuição associada a uma nova observação y , marginalizada (integrada) sobre a distribuição a posteriori.

O modelo B depende da distribuição a posteriori do modelo A e da distribuição a posteriori preditiva do modelo A. É necessário construir um modelo que estime o nível atual de apoio partidário com base em pesquisas pré-eleitorais publicadas durante a campanha eleitoral, e que combine com as previsões do modelo fundamental teórico (modelo A). Para isso são utilizados os resultados das pesquisas publicadas por diferentes institutos de pesquisa.

1.2 Estrutura da dissertação

No primeiro capítulo foi apresentada a introdução e uma breve descrição da modelagem que será empregada neste trabalho; no capítulo 2 são introduzidos os conteúdos essenciais para uma melhor leitura e entendimento do modelo. No terceiro capítulo é apresentado detalhadamente o modelo proposto e no capítulo seguinte sua aplicação, bem como os resultados. Por último, é apresentada a conclusão do trabalho e propostas de desenvolvimento futuro.

2 Conceitos preliminares

A estatística tem como um de seus fundamentos elaborar técnicas a fim de extrair informações a respeito de um determinado assunto. Como descrito por Costa (2011), com o avanço da modernização global, a modelagem de fenômenos cada vez mais complexos requer maior flexibilidade, ou seja, existe a necessidade de modelar um fenômeno através de funções que não sejam de distribuições necessariamente conhecidas. Há uma série de técnicas utilizadas para encontrar valores como probabilidades, valores esperados, entre outros, quando não há forma fechada (analítica) para estes problemas. Alguns exemplos são: método de Laplace, *bootstrap*, Monte Carlo, Monte Carlo via Cadeia de Markov (MCMC), entre outros.

Neste capítulo será feita uma revisão de conceitos e métodos necessários para a implementação e compreensão dos modelos dinâmicos hierárquicos que serão aplicados aos dados eleitorais.

2.1 Inferência Bayesiana

Em meados do século XVIII, desenvolvido por Thomas Bayes, surgiu o Teorema de Bayes e posteriormente Laplace estendeu esse conhecimento, desenvolvendo a inferência Bayesiana, que apresenta uma teoria subjetiva de probabilidade, fundamentada em um conhecimento a priori em associação a incertezas envolvidas no estudo, ou seja, utiliza-se o conhecimento prévio do pesquisador (informação a priori) adicionalmente aos dados na análise estatística, diferentemente do que ocorre com a abordagem clássica que restringe-se apenas às aquelas informações obtidas com a observação dos dados amostrais.

Seja θ um escalar, um vetor, ou até mesmo uma matriz sobre o qual deseja-se fazer inferência, sendo este parâmetro (ou coleção de parâmetros) desconhecido. A ideia da inferência Bayesiana é tentar minimizar a falta de informações sobre θ , e ainda quantificar o grau de incerteza acerca da quantidade de interesse através de modelos probabilísticos

para θ . De maneira mais simplista é agregar informações já conhecidas sobre um determinado problema, a distribuição a priori, e através de uma regra (teorema de Bayes) com informações provenientes dos dados. Os conceitos de distribuição a priori e distribuição a posteriori bem como o teorema de Bayes serão apresentados nas próximas seções.

2.1.1 Teorema de Bayes

Como descrito por Migon, Gamerman e Louzada (2014), seja H as informações iniciais disponíveis sobre algum parâmetro de interesse. Suponha ainda que esta informação inicial seja expressa em termos probabilísticos, ou seja, pode ser resumido através de $p(\theta | H)$ e, se o conteúdo de H é suficiente para o propósito inferencial, isso é tudo o que é preciso. Sendo assim, a incerteza sobre θ está completa.

Na maioria das vezes a informação sobre θ através de H não é suficiente e, nesse caso, obter mais informações é necessário, sendo este o problema a ser solucionado. Seja \mathbf{X} um vetor de quantidades aleatórias relacionado com θ que pode ser observado, fornecendo mais informações sobre θ . Se \mathbf{X} não for aleatório, uma relação funcional relacionada a θ deve existir. Pode-se então avaliar o valor de θ e o problema é resolvido. Antes de observar \mathbf{X} , deve-se conhecer a distribuição amostral de \mathbf{X} dada por $p(\mathbf{X} | \theta, H)$, em que a dependência de θ é explicitada. Depois de observar o valor de \mathbf{X} , a quantidade de informações sobre θ é refinada de H para $H^* = H \cap \{\mathbf{X} = \mathbf{x}\}$. De fato, H^* é um subconjunto de H .

Agora a informação que se tem sobre θ é resumida por $p(\theta | \mathbf{x}, H)$. Utilizando a probabilidade condicional, é fácil obter tal expressão:

$$p(\theta | \mathbf{x}, H) = \frac{p(\theta, \mathbf{x} | H)}{p(\mathbf{x} | H)} = \frac{p(\mathbf{x} | \theta, H)p(\theta | H)}{p(\mathbf{x} | H)} \quad (2.1)$$

em que,

$$p(\mathbf{x} | H) = \int_{\Theta} p(\mathbf{x}, \theta | H) d\theta. \quad (2.2)$$

O conjunto Θ , chamado de espaço paramétrico, é o conjunto formado por todos os valores que θ pode assumir. O resultado mostrado em (2.1) é conhecido como Teorema

de Bayes. Como o principal interesse é sobre $\boldsymbol{\theta}$, o resultado é aproximado por

$$p(\boldsymbol{\theta} \mid \mathbf{x}) \propto p(\mathbf{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.3)$$

A dependência em H é omitida para simplificar a notação, uma vez que é fator comum em todos os termos, no entanto, não deve ser esquecida. A fórmula acima é válida tanto para casos discretos e contínuos, como para escalares, vetores e matrizes. O teorema fornece uma regra para atualizar probabilidades de $\boldsymbol{\theta}$, a partir da observação dos dados, partindo de $p(\boldsymbol{\theta})$ e levando a $p(\boldsymbol{\theta} \mid \mathbf{x})$, por isso essas quantidades são chamadas de distribuição a priori e distribuição a posteriori respectivamente.

Para recuperar a constante em (2.3), basta integrar a expressão apresentada em (2.4) e igualar a 1.

$$p(\boldsymbol{\theta} \mid \mathbf{x}) = kp(\mathbf{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (2.4)$$

em que,

$$1 = \int_{\Theta} p(\boldsymbol{\theta} \mid \mathbf{x})d\boldsymbol{\theta} = k \int_{\Theta} p(\mathbf{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (2.5)$$

Sendo assim,

$$\begin{aligned} k^{-1} &= p(\mathbf{x} \mid H) \\ &= \int_{\Theta} p(\mathbf{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= E_{\boldsymbol{\theta}}[p(\mathbf{x} \mid \boldsymbol{\theta})] \end{aligned} \quad (2.6)$$

Esta é a chamada distribuição preditiva. Após remover a dependência de H em termos de notação, a distribuição pode ser denotada por $p(\mathbf{x})$. Neste caso ela é o valor esperado da distribuição de \mathbf{X} e se comporta como uma previsão para um determinado H . Portanto, antes de observar \mathbf{X} , é útil verificar a adequação prévia através das previsões que são fornecidas para \mathbf{X} .

A distribuição a priori para um parâmetro θ deve ser elicitada levando em consideração o grau de informação que o pesquisador possui sobre tal parâmetro, podendo esta ser informativa ou não-informativa. A informação a priori requer uma forma funcional, ou seja, a distribuição de probabilidade e o valor dos seus parâmetros, que são chamados de hiperparâmetros. Nos casos em que há informação para um intervalo de um ou mais

valores do parâmetro no espaço paramétrico, define-se a distribuição a priori como sendo uma distribuição informativa. Por outro lado, quando não há priorização de nenhum valor do parâmetro no espaço paramétrico pode-se propor o uso de uma distribuição a priori não-informativa, que pode ser obtida por métodos tais como: método de Bayes-Laplace, método de Jeffreys, método de Box-Tiao e método de Berger-Bernardo, entre outros. Além dessas, existem outras formas de especificação da distribuição a priori como prioris subjetivas e prioris conjugadas. Mais detalhes podem ser encontrados em Casella e Berger (2004) e Larson (1982).

A função a posteriori é a distribuição do parâmetro de interesse θ , após a coleta de informação que os dados trazem junto das informações prévias fornecidas através da priori. Na inferência Bayesiana, a distribuição de densidade a posteriori contém toda a informação disponível em termos de descrição probabilística sobre a quantidade de interesse θ , e é a base da estimação em inferência Bayesiana.

Nos casos em que há mais de um parâmetro de interesse, ou seja, θ é um vetor paramétrico $\boldsymbol{\theta}$, para obter as funções marginais de um determinado θ_i e assim prosseguir com a inferência sobre o mesmo, deve-se integrar a distribuição conjunta a posteriori em relação aos demais parâmetros que não sejam o de interesse. Através de cada posteriori marginal, obtém-se um “leque” de possibilidades para cada um dos parâmetros, possibilitando uma variedade de técnicas de análise sobre o mesmo. Pode-se utilizar a média da distribuição a posteriori como estimativa, a moda, ou até mesmo uma região de credibilidade construída a partir da distribuição a posteriori. Na maioria das vezes, essas distribuições não possuem forma analítica como afirma Gamerman (1996), principalmente nos casos em que há vários parâmetros abordados, sendo necessário o uso de métodos numéricos e de aproximação, como o método de Laplace e métodos de Monte Carlo via Cadeia de Markov (MCMC), esta última classe utilizada nesta dissertação.

2.1.2 Predição

No mundo moderno, o uso de técnicas de predição é de muita valia, e permite identificar certos padrões a fim de uma melhor estratégia para a realização de uma determinada tarefa. A inferência Bayesiana através do teorema de Bayes é uma técnica eficaz que produz automaticamente maneiras de fazer previsões para futuras observações. Se o desejo é prever Y , cuja descrição probabilística é $p(y | \boldsymbol{\theta})$, tem-se que

$$\begin{aligned}
p(y \mid \mathbf{x}) &= \int_{\Theta} p(y, \boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} \\
p(y \mid \mathbf{x}) &= \int_{\Theta} p(y \mid \boldsymbol{\theta}, \mathbf{x}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} \\
p(y \mid \mathbf{x}) &= \int_{\Theta} p(y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}.
\end{aligned} \tag{2.7}$$

A última igualdade ocorre devido à independência entre \mathbf{X} e Y , uma vez que o $\boldsymbol{\theta}$ é dado. Essa suposição de independência condicional é observada em vários problemas estatísticos. Além disso, segue-se da última equação que,

$$p(y \mid \mathbf{x}) = E_{[\boldsymbol{\theta} \mid \mathbf{x}]}[p(y \mid \boldsymbol{\theta})]. \tag{2.8}$$

2.2 Métodos de Monte Carlo via Cadeias de Markov (MCMC)

Como dito anteriormente, na prática a distribuição a posteriori do vetor paramétrico dificilmente possui forma analítica conhecida, e através da utilização de métodos de simulação estocástica é possível obter amostras, de forma aproximada, da distribuição a posteriori. Nesse capítulo é feita uma revisão dos métodos de MCMC, primeiramente fazendo uma introdução ao método e, posteriormente, apresentando três algoritmos: os dois métodos de MCMC mais usados na literatura (Metropolis-Hastings e Amostrador de Gibbs) e o método utilizado nesta dissertação (Monte Carlo Hamiltoniano).

2.2.1 Monte Carlo via Cadeias de Markov (MCMC)

O método de Monte Carlo via cadeia de Markov (MCMC) é um método iterativo amplamente utilizado quando deseja-se obter a distribuição posteriori em problemas mais complexos, em que a distribuição a posteriori não possui forma explícita, ou de fácil cálculo. O método baseia-se em construir uma cadeia de Markov que seja fácil de ser simulada e tenha distribuição de equilíbrio dada pela distribuição de interesse. Na inferência Bayesiana, o objetivo torna-se obter uma amostra da distribuição a posteriori e calcular estimativas amostrais de características desta distribuição.

Definição 2.1 *Uma cadeia de Markov é um processo estocástico X_0, X_1, \dots , tal que a distribuição de X_t , dados todos os valores anteriores X_0, \dots, X_{t-1} , depende apenas de*

X_{t-1} . Probabilisticamente, $P(X_t \in A \mid X_0, \dots, X_{t-1}) = P(X_t \in A \mid X_{t-1})$ para qualquer subconjunto A .

Os métodos MCMC requerem ainda que a cadeia seja homogênea, irredutível e aperiódica. Uma cadeia é dita ser,

1. **Homogênea** se as probabilidades de transição de um estado para outro são invariantes;
2. **Irredutível** se cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
3. **Aperiódica** se não há estados absorventes, ou seja, em k passos a cadeia pode estar em qualquer estado.

O método é baseado na seguinte forma: seja $P(j, i) = P(i \mid j)$ a função de transição da cadeia, ou seja, a probabilidade de em um passo o processo ir para o estado i , dado que estava no estado j no passo anterior. Seja $\pi(j), j \in \mathbb{R}^d$ uma distribuição, conhecida a menos de uma constante multiplicativa, porém complexa o suficiente para não ser possível obter uma amostra diretamente dela. Para gerar amostras de $\pi(j)$, a função $P(j, i)$ que converge para $\pi(j)$ na k -ésima iteração (k passos) é calculada e utilizada. O processo é iniciado em um estado qualquer j e após um número suficientemente grande de simulações, as observações são aproximadamente geradas da distribuição de interesse $\pi(\cdot)$. A convergência da cadeia de Markov acontece depois de um período chamado de aquecimento. Conforme o número de iterações vai aumentando, os valores iniciais são esquecidos pela cadeia até que aconteça a convergência para a distribuição de equilíbrio $\pi(j)$. Maiores informações e detalhes podem ser encontrados em Robert e Casella (2004).

Os três algoritmos que são apresentados na dissertação satisfazem a essas condições, o algoritmo de Metropolis-Hastings e o amostrador de Gibbs, os métodos mais usuais de MCMC e, por fim, o algoritmo de Monte Carlo Hamiltoniano (HMC) que é utilizado pelo Stan, linguagem usada nessa dissertação e que será apresentado em maiores detalhes a seguir.

2.2.2 Algoritmo Monte Carlo Hamiltoniano (HMC)

Como descrito por Xavier (2019), a ideia original do algoritmo Monte Carlo Hamiltoniano (HMC) é baseada na dinâmica Hamiltoniana, uma técnica geral da física teórica

em que a energia total de um sistema de partículas é descrito por meio de uma função chamada função Hamiltoniana.

Inicialmente o método foi proposto por Dueane, Kennedy, Penedleton, e Roweth (1987) e chamado de Monte Carlo Híbrido, sendo utilizado para a simulação dinâmica de sistemas moleculares. Na década de 90, o método foi utilizado pela primeira vez em aplicações estatísticas por Neal (1995), mais especificamente no contexto de redes neurais, logo após Neal (2011) e Gelman *et al.* (2014) apresentaram propriedades importantes desse algoritmo.

O Monte Carlo Hamiltoniano pode ser entendido a partir de duas perspectivas: do ponto de vista de dinâmicas moleculares o HMC surge como um método que corrige erros introduzidos por uma discretização das equações de Hamilton. Já do ponto de vista estatístico, tem-se uma proposta para o uso da dinâmica Hamiltoniana como um mecanismo eficaz para realização de métodos MCMC, de tal forma que o novo estado proposto estará longe do estado atual da cadeia e será aceito com uma alta probabilidade. Mais detalhes sobre a dinâmica de Hamilton podem ser encontrados em Neal (2011).

Para exemplificar a dinâmica Hamiltoniana considere um disco sem atrito deslizando sobre uma superfície de altura variável. O estado deste sistema consiste no posicionamento do disco, dada por um vetor $\boldsymbol{\theta}$, e o movimento do disco (sua massa vezes sua velocidade), dada por um vetor \mathbf{q} . A energia potencial, $U(\boldsymbol{\theta})$, do disco é proporcional à altura da superfície na posição atual. Já sua energia cinética, $K(\mathbf{q})$, é igual a $\mathbf{q}'\mathbf{q}/2m$, em que m é a massa do disco.

Em alguma parte do nível da superfície, o disco move-se a uma velocidade constante. Entretanto, se este encontrar uma inclinação ascendente, o movimento do disco permite que este continue com a sua energia cinética diminuindo e sua energia potencial crescendo até que sua energia cinética (e, portanto, \mathbf{q}) chegue a zero. Em seguida, o disco iniciará um movimento descendente com energia cinética aumentando e energia potencial diminuindo. Mais formalmente a Função Hamiltoniana pode ser definida por:

Definição 2.2 (*Função Hamiltoniana*) A energia total de um sistema é representada pela função Hamiltoniana que é dada por:

$$\begin{aligned} H(\boldsymbol{\theta}, \mathbf{q}) &= U(\boldsymbol{\theta}) + K(\mathbf{q}) \\ &= U(\boldsymbol{\theta}) + \mathbf{q}'M^{-1}\mathbf{q}, \end{aligned} \tag{2.9}$$

em que $U(\boldsymbol{\theta})$ é a energia potencial e é definida como menos o log da densidade de probabi-

lidade da distribuição de $\boldsymbol{\theta}$ a qual se quer amostrar e $K(\mathbf{q})$ é a energia cinética, sendo \mathbf{M} uma matriz de massa simétrica positiva definida, que é tipicamente diagonal, e é muitas vezes um múltiplo escalar da matriz identidade.

Definida a função Hamiltoniana deve-se determinar como $\boldsymbol{\theta}$ e \mathbf{q} se movem durante um tempo t . Abaixo segue a definição de equações de movimento.

Definição 2.3 (*Equações de movimento*) Para determinar o movimento de $\boldsymbol{\theta}$ e \mathbf{q} , considere as seguintes derivadas parciais da função Hamiltoniana

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \frac{\partial H}{\partial \mathbf{q}} = \nabla_{\mathbf{q}} K(\mathbf{q}) \quad (2.10)$$

$$\frac{\partial \mathbf{q}}{\partial t} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}), \quad (2.11)$$

em que o símbolo ∇_x indica o vetor gradiente em relação a uma variável x . As equações (2.10) e (2.11) determinam a posição e a velocidade de uma partícula para qualquer intervalo de tempo de duração s , ou seja, estas equações definem um mapeamento T_s , do estado, em qualquer instante t para o estado no tempo $t + s$.

Considere um vetor de parâmetros de interesse $\boldsymbol{\theta} \in \mathbb{R}^d$ distribuído de acordo com uma densidade a posteriori $p(\boldsymbol{\theta})$. Seja $\mathbf{q} \in \mathbb{R}^d$ um vetor auxiliar de parâmetros com $\mathbf{q} \sim N_d(\mathbf{0}, \mathbf{M})$, distribuído como uma distribuição Normal com média $\mathbf{0}$ e com uma matriz de covariâncias \mathbf{M} , independente de $\boldsymbol{\theta}$. Logo, o negativo do logaritmo da distribuição conjunta de $(\boldsymbol{\theta}, \mathbf{q})$ é dado pela equação Hamiltoniana

$$H(\boldsymbol{\theta}, \mathbf{q}) = -\ln(p(\boldsymbol{\theta})) + \frac{1}{2} \ln((2\pi)^d |\mathbf{M}|) + \frac{1}{2} \mathbf{q}' \mathbf{M}^{-1} \mathbf{q}$$

Em aplicações práticas as equações (2.10) e (2.11), em geral, não podem ser resolvidas de forma analítica, sendo necessário a utilização de métodos numéricos. Um dos métodos mais populares na literatura é o método de Störmer-Verle, que aproxima as equações por:

$$\mathbf{q}(t + \epsilon/2) = \mathbf{q}(t) + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}(t)) \quad (2.12)$$

$$\boldsymbol{\theta}(t + \epsilon) = \boldsymbol{\theta}(t) + \epsilon \mathbf{M}^{-1} \mathbf{q}(t + \epsilon/2) \quad (2.13)$$

$$\mathbf{q}(t + \epsilon) = \mathbf{q}(t + \epsilon/2) + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}(t + \epsilon)) \quad (2.14)$$

para algum valor de $\epsilon > 0$ pequeno. Ao gerar \mathbf{q}^* de uma distribuição normal proporcione-se uma inicialização estocástica. Este passo é necessário a fim de tornar a cadeia de Markov resultante irredutível e aperiódica. Utilizando as equações (2.12), (2.13) e (2.14) pode-se definir o algoritmo HMC abaixo.

1. Forneça uma posição inicial $\boldsymbol{\theta}^{(0)}$ e o tamanho do passo ϵ ;
2. Inicie um contador em $i = 1$;
3. Gere $\mathbf{q}^* \sim N^d(\mathbf{0}, \mathbf{M})$ e $u \sim U(0, 1)$;
4. Faça $(\boldsymbol{\theta}^I, \mathbf{q}^I) = (\boldsymbol{\theta}^{(i-1)}, \mathbf{q}^*)$, $H_0 = H(\boldsymbol{\theta}^I, \mathbf{q}^I)$;
5. De $j = 1, \dots, k$, em que k é o número de vezes para resolver as equações de Störmer-Verlet, faça:

$$\begin{aligned} \mathbf{q}^* &= \mathbf{q}^* + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}^{(j-1)}) \\ \boldsymbol{\theta}^{(j-1)} &= \boldsymbol{\theta}^{(j-1)} + \epsilon \mathbf{M}^{-1} \mathbf{q}^* \\ \mathbf{q}^* &= \mathbf{q}^* + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}^{(j-1)}) \end{aligned} \quad (2.15)$$

6. Faça $(\boldsymbol{\theta}^S, \mathbf{q}^S) = (\boldsymbol{\theta}^{(j-1)}, \mathbf{q}^*)$ e $H_1 = H(\boldsymbol{\theta}^S, \mathbf{q}^S)$;
7. Calcule $\alpha[(\boldsymbol{\theta}^S, \mathbf{q}^S), (\boldsymbol{\theta}^I, \mathbf{q}^I)] = \min[\exp(H_0 - H_1), 1]$;
8. Se $u < \alpha[(\boldsymbol{\theta}^S, \mathbf{q}^S), (\boldsymbol{\theta}^I, \mathbf{q}^I)]$, então

$$\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^S; \quad (2.16)$$

caso contrário,

$$\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^I; \quad (2.17)$$

9. Faça $i = i + 1$ e retorne para o passo 3.

A expressão $\alpha[(\boldsymbol{\theta}^S, \mathbf{q}^S), (\boldsymbol{\theta}^I, \mathbf{q}^I)]$, que é a probabilidade de aceitação do HMC, é dada pela diferença das equações Hamiltonianas H_0 e H_1 . Para manter esta probabilidade próxima de um, a diferença entre as equações Hamiltonianas deve tender a zero. Outra característica é que esta probabilidade só será mantida próxima a um se a evolução da trajetória for contínua. No entanto, devido a sua discretização em etapas individuais,

a probabilidade de aceitação sempre se desvia de um (devido a erros de discretização). Desta forma os parâmetros $\epsilon > 0$ e k devem ser ajustados para atingir a taxa de aceitação desejada.

2.3 Modelos dinâmicos: Uma introdução

Como já mencionado, o interesse desta dissertação é propor um modelo que consiga prever de forma eficaz o resultado de eleições multipartidárias. Uma das vantagens de utilizar modelos dinâmicos se dá pelo fato de que em casos em que há fenômenos com estrutura temporal, como no caso desta aplicação, pode-se incorporar a incerteza associada à passagem do tempo através de uma evolução suave. Serão vistas de forma resumida três classes do extenso tema de modelagem dinâmica, são elas: modelos lineares dinâmicos, modelos lineares hierárquicos e modelos lineares dinâmicos hierárquicos generalizados.

2.3.1 Modelos Lineares Dinâmicos

O tema modelos dinâmicos é extenso, sendo dividido em várias sub-categorias, como a dos modelos lineares dinâmicos normais, conhecidos simplesmente por modelos lineares dinâmicos (MLD). Esses modelos consideram que a variável resposta, bem como os parâmetros de evolução, são distribuídos de forma normal. Além disso, as equações são compostas de equações lineares.

Em um modelo de regressão usual, temos uma variável Y , chamada de variável resposta ou variável dependente, que é explicada por um conjunto de variáveis regressoras \mathbf{X} e um conjunto de parâmetros denotados por $\boldsymbol{\theta}$:

$$\begin{aligned} Y_1 &= \theta_0 + \theta_1 x_{1,0} + \dots + \theta_k x_{1,k} + e_1 \\ Y_2 &= \theta_0 + \theta_1 x_{2,0} + \dots + \theta_k x_{2,k} + e_2 \\ &\vdots \\ Y_n &= \theta_0 + \theta_1 x_{n,0} + \dots + \theta_k x_{n,k} + e_n \end{aligned}$$

em que geralmente $e_i \sim N(0, \sigma^2)$, para $i = 1, \dots, n$. Ainda pode-se escrever a equação acima na forma matricial.

$$\mathbf{Y} = \mathbf{X}'\boldsymbol{\theta} + \mathbf{e},$$

em que,

$$\mathbf{X}' = \begin{pmatrix} x_{1,0} & x_{1,1} & \dots & x_{1,k} \\ x_{2,0} & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,k} & x_{n,1} & \dots & x_{n,k} \end{pmatrix} \text{ e } \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}.$$

As variáveis explicativas podem ser de diferentes naturezas, bastando apenas serem quantificáveis. Os coeficientes $\boldsymbol{\theta}$ informam a influência das regressoras sobre \mathbf{Y} . Os valores de $\boldsymbol{\theta}$ geralmente são valores desconhecidos e estimados a partir de uma amostra.

Em modelos dinâmicos, os valores dos coeficientes mudam com o passar do tempo, sendo assim,

$$Y_t = \mathbf{X}'_t \boldsymbol{\theta}_t + e_t, t = 1, \dots, n,$$

em que n é o número total de períodos de tempo em que os dados foram observados.

A formulação acima gera uma quantidade maior de parâmetros a serem estimados, além de ser um modelo mais informativo. A informação vem do fato que os parâmetros sucessivos são intimamente relacionados. Em geral, um parâmetro é igual ao seu antecessor mais uma perturbação, que é causada pelas mudanças as quais o sistema está sujeito.

Se o sistema é estático, $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}$, aí volta-se para o modelo de regressão. Caso contrário, a fórmula mais utilizada é,

$$\boldsymbol{\theta}_t = \mathbf{G}_{t-1} \boldsymbol{\theta}_{t-1} + \mathbf{w}_t,$$

em que

- \mathbf{G}_t é a matriz de evolução, que contém todos os valores conhecidos e controla a parte determinística do processo.
- \mathbf{w}_t é a perturbação responsável pela introdução de incertezas ao longo do tempo e consequentemente a perda de informação.
- Se $\mathbf{G}_t = \mathbf{I}$ e $\mathbf{w}_t = 0$, volta-se para o caso de regressão linear múltipla.

O modelo, sendo assim, pode ser especificado através de duas equações:

- A equação de observações

$$Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, v_t \sim N(0, V)$$

- A equação de sistema

$$\boldsymbol{\theta}_t = \mathbf{G}_{t-1}\boldsymbol{\theta}_{t-1} + w_t, w_t \sim N(0, W),$$

em que uma possível apresentação para \mathbf{F}'_t é $\mathbf{F}'_t = \{1, X'\}$

A equação do sistema informa a partir da posteriori de ontem como pode-se chegar na priori de hoje. Através do método Bayesiano pode-se combinar a priori de hoje com a informação do sistema, assim obtendo a posteriori do atual momento. Para amanhã e os dias futuros, o ciclo se repete. Para maiores informações e detalhes ver Gamerman e Migon (1993).

2.3.2 Modelos Lineares Dinâmicos Hierárquicos

A classe dos modelos lineares dinâmicos hierárquicos (MLDH), como o próprio nome sugere, incorpora uma estrutura hierárquica na modelagem ao longo do tempo. O modelo especificado por Gamerman e Migon (1993) é composto de três partes: a equação de observação, a equação de estrutura e o sistema de equação. A primeira equação descreve a distribuição das observações, a equação de estrutura descreve a estrutura hierárquica dos parâmetros e a equação do sistema descreve a evolução dos parâmetros através do tempo. Essa construção generaliza os modelos hierárquicos de Lindley e Smith (1972), considerando uma variação temporal para os parâmetros e os modelos lineares dinâmicos de Harrison e Stevens (1976), considerando uma estrutura paramétrica refinada.

O modelo linear dinâmico hierárquico com k estágios pode ser escrito como:

- a) A equação de observação:

$$\mathbf{Y}_t = \mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t} + \mathbf{v}_{1,t}, \text{ em que } \mathbf{v}_{1,t} \sim N(0, \mathbf{V}_{1,t}), \quad (2.18)$$

- b) As equações de estrutura:

$$\begin{aligned} \boldsymbol{\theta}_{1,t} &= \mathbf{F}_{2,t}\boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \text{ em que } \mathbf{v}_{2,t} \sim N(0, \mathbf{V}_{2,t}), \\ \boldsymbol{\theta}_{2,t} &= \mathbf{F}_{3,t}\boldsymbol{\theta}_{3,t} + \mathbf{v}_{3,t}, \text{ em que } \mathbf{v}_{3,t} \sim N(0, \mathbf{V}_{3,t}), \\ \boldsymbol{\theta}_{3,t} &= \mathbf{F}_{4,t}\boldsymbol{\theta}_{4,t} + \mathbf{v}_{4,t}, \text{ em que } \mathbf{v}_{4,t} \sim N(0, \mathbf{V}_{4,t}), \\ &\vdots \\ \boldsymbol{\theta}_{k-1,t} &= \mathbf{F}_{k,t}\boldsymbol{\theta}_{k,t} + \mathbf{v}_{k,t}, \text{ em que } \mathbf{v}_{k,t} \sim N(0, \mathbf{V}_{k,t}), \end{aligned} \quad (2.19)$$

c) E por fim, a equação do sistema:

$$\boldsymbol{\theta}_{k,t} = \mathbf{G}_t \boldsymbol{\theta}_{k,t-1} + \mathbf{w}_t, \text{ em que } \mathbf{w}_t \sim N(0, \mathbf{W}_t). \quad (2.20)$$

Os termos $\mathbf{v}_{1,t}, \mathbf{v}_{2,t}, \dots, \mathbf{v}_{k,t}$ e \mathbf{w}_t são independentes com matrizes de covariância, em geral, desconhecidas e $\mathbf{F}_{1,t}, \mathbf{F}_{2,t}, \dots, \mathbf{F}_{k,t}$ são matrizes conhecidas. O modelo especificado acima considera k estágios na hierarquia dos parâmetros. De acordo com Lindley e Smith (1972), modelos com mais do que três estágios raramente são utilizados. Frequentemente, modelos hierárquicos têm dois estágios que podem ser obtidos a partir de um modelo com três estágios, fazendo $\mathbf{F}_{3,t} = \mathbf{I}$, a matriz identidade, e $\mathbf{V}_{3,t} = 0$, uma matriz de entradas nulas.

O aspecto essencial das equações de estrutura acima é a redução progressiva na dimensão dos parâmetros conforme o nível aumenta. Desta forma, as dimensões de $\boldsymbol{\theta}_{i,t}$ são r_i , para $i = 1, 2, \dots, k$, satisfazendo $r_1 > r_2 > \dots > r_k$. Adicionalmente, as matrizes $\mathbf{F}_{i,t}$ são de posto completo. Caso haja alguma matriz $\mathbf{F}_{i,t}$ que não seja de posto completo, então o modelo é sobre-parametrizado no estágio i e $\theta_{i,t}$ pode ser reduzido através da remoção de parâmetros redundantes.

2.3.3 Modelos Lineares Dinâmicos Hierárquicos Generalizados

Harrison e Migon (1985) introduziram o conceito de modelos lineares dinâmicos generalizados (MLDG), tendo como diferença que esta classe não necessita do pressuposto de normalidade presente nos modelos dinâmicos lineares.

Como no caso dos MLDG, há a classe de modelos lineares dinâmicos hierárquicos generalizados (MLDHG) que surge com o relaxamento das hipóteses de linearidade e normalidade, muitas vezes não suportadas pelos dados.

O modelo linear dinâmico hierárquico generalizado é composto de três equações: a equação de observação, a equação de estrutura e a equação de sistema. Essa construção, por sua vez, generaliza os modelos hierárquicos de Lindley e Smith (1972) e os modelos lineares dinâmicos generalizados de West, Harrison e Migon (1985). O modelo com k estágios pode ser escrito como:

a) A equação de observação:

$$\begin{aligned} Y_t &\propto \exp \left\{ \phi_t [Y_t \eta_t - b(\eta_t)] \right\} \\ g(\eta_t) &= F_{1,t} \boldsymbol{\theta}_{1,t} + v_{1,t} \text{ com } v_{1,t} \sim N(0, V_{1,t}), \end{aligned} \quad (2.21)$$

ou seja, Y_t deve ter distribuição que pertença a família exponencial ¹.

b) As equações de estrutura:

$$\begin{aligned}
 \boldsymbol{\theta}_{1,t} &= \mathbf{F}_{2,t}\boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \text{ em que } \mathbf{v}_{2,t} \sim N(0, \mathbf{V}_{2,t}), \\
 \boldsymbol{\theta}_{2,t} &= \mathbf{F}_{3,t}\boldsymbol{\theta}_{3,t} + \mathbf{v}_{3,t}, \text{ em que } \mathbf{v}_{3,t} \sim N(0, \mathbf{V}_{3,t}), \\
 \boldsymbol{\theta}_{3,t} &= \mathbf{F}_{4,t}\boldsymbol{\theta}_{4,t} + \mathbf{v}_{4,t}, \text{ em que } \mathbf{v}_{4,t} \sim N(0, \mathbf{V}_{4,t}), \\
 &\vdots \\
 \boldsymbol{\theta}_{k-1,t} &= \mathbf{F}_{k,t}\boldsymbol{\theta}_{k,t} + \mathbf{v}_{k,t}, \text{ em que } \mathbf{v}_{k,t} \sim N(0, \mathbf{V}_{k,t}),
 \end{aligned} \tag{2.22}$$

c) E por fim, a equação do sistema:

$$\boldsymbol{\theta}_{k,t} = \mathbf{G}_t\boldsymbol{\theta}_{k,t-1} + \mathbf{w}_t, \text{ em que } \mathbf{w}_t \sim N(0, \mathbf{W}_t). \tag{2.23}$$

A distribuição amostral é agora possivelmente não Normal e a regressão linear afeta a distribuição das observações através de uma função possivelmente não linear da função de resposta média $\mu_t = E(Y_t | \eta_t, \varphi_t) = b(\eta_t)$. Os termos de erro $\mathbf{v}_{1,t}, \mathbf{v}_{2,t}, \dots, \mathbf{v}_{k,t}$ e w_t são independentes com matrizes de covariância, em geral, desconhecidas e $\mathbf{F}_{1,t}, \mathbf{F}_{2,t}, \dots, \mathbf{F}_{k,t}$ são matrizes de regressão conhecidas. Por fim, $g(\cdot)$ é uma função de ligação conhecida que relaciona o parâmetro natural da distribuição η_t ao parâmetro de estado $\boldsymbol{\theta}_{1,t}$.

No próximo capítulo é apresentado o modelo proposto para este trabalho, que necessita das informações até aqui apresentadas para uma melhor compreensão. O modelo é uma aplicação direta da classe de MLDG que foi recém explicada.

¹Considere Y uma variável aleatória que segue uma distribuição na família exponencial, definida da seguinte forma $f(y | \theta, \phi) = \exp\{\phi - 1[y\theta - \psi(\theta)] + c(y, \phi)\}$, para θ e ϕ parâmetros desconhecidos e $\psi(\cdot)$ e $c(\cdot, \cdot)$ funções conhecidas. θ é conhecido como parâmetro natural e ϕ é conhecido como parâmetro de escala. A família exponencial também estende-se a vetores aleatórios.

3 O modelo

Como dito na motivação, o modelo proposto, também chamado de modelo completo para esta dissertação é o modelo apresentado por Stoetzer, Marcel, Neunhoeffers e Sternberg (2018), que é uma combinação do modelo A e um modelo que é baseado em pesquisas eleitorais que ocorrem durante o processo eleitoral, o modelo B. Desta forma, é apresentado o modelo A e posteriormente será apresentado a adição/aperfeiçoamento deste modelo ao ser combinado com o modelo baseado nas pesquisas, modelo B, resultando no modelo completo.

Seja \mathbf{v}_{pe} o número de votos do partido p na eleição e , com $p = 1, \dots, P$, em que P é o número total de partidos e $e = 1, \dots, E$, em que E representa a próxima eleição.

O modelo A propõe utilizar dados/informação anteriores de cada partido/candidato a fim de prever os resultados da próxima eleição. Sendo assim, como a soma total de votos em cada eleição deve somar um (100%), a distribuição de Dirichlet pode ser apropriada para modelar as porcentagens de votação em cada partido/candidato. Dessa forma, o modelo assume que o vetor de votos compartilhados $\mathbf{v}_e = (\mathbf{v}_{1e}, \mathbf{v}_{2e}, \dots, \mathbf{v}_{Pe})$ na eleição e é distribuído por:

$$\mathbf{v}_e \sim D(\mathbf{a}_e), \quad (3.1)$$

em que $\mathbf{a}_e = (a_{1e}, \dots, a_{Pe})$ são os parâmetros da Dirichlet. Na eleição E , eleição que pretende-se estimar o número de votos compartilhados, o componente sistemático $\mathbf{a}_e = \boldsymbol{\alpha}_e$ modela o log dos parâmetros de escala como uma função aditiva de covariáveis $x_{pe}^{(k)}$, ou seja,

$$\alpha_{pe} = \exp \left(\beta_e^0 + \sum_{k=1}^K \beta_e^k x_{pe}^{(k)} \right), \quad p = 1, \dots, P \text{ e } e = E. \quad (3.2)$$

O modelo contém K covariáveis que variam de acordo com cada partido/candidato e eleição, sendo assim, $x_{pe}^{(k)}$ é o valor da covariável k na eleição e referente ao partido/candidato p . Grande parte dos modelos teóricos de previsão utilizam dados de muito tempo atrás. Este modelo, diferentemente do que ocorre nos outros, tem como característica variar os efeitos de diferentes preditores ao longo do tempo. Sendo assim os parâmetros β_e^0 e β_e^k , $k = 1, \dots, K$ variam de acordo com a eleição. Esse aspecto da modelagem pode ser explicada pelo fato de que a sociedade muda com o passar dos anos. Por exemplo, um país que concentrava seus votos em apenas dois partidos pode, a partir de certo momento, distribuir os votos entre diversos partidos. O modelo A é flexível o suficiente para modelar de forma adequada situações como essa, com a incorporação de uma estrutura hierárquica nos coeficientes da equação 3.2, tal que:

$$\beta_e^k \sim N(\tilde{\beta}_e^k, \tau_k^2) \quad (3.3)$$

$$\tilde{\beta}_e^k = \beta_{e-1}^k + \gamma_d^k. \quad (3.4)$$

Ou seja, os coeficientes β_e^k seguem uma distribuição normal com média $\tilde{\beta}_e^k$, que é uma combinação linear do parâmetro β correspondente da eleição anterior e de um parâmetro de desvio.

De posse dessas informações, agora pode-se calcular a distribuição a posteriori e distribuição a posteriori preditiva para a eleição futura.

Para calcular a distribuição a posteriori para o modelo A é necessário especificar a distribuição a priori para θ , que é definido como o conjunto de todos os parâmetros do modelo, ou seja: $\theta = (\tau_k^2, \beta_e^k, \gamma_d^k, \text{ para } k = 1, \dots, K)$. Para τ_k^2 e γ_d^k são sugeridas prioris independentes da forma $\tau_k^2 \sim N^+(0, 1)$, $\gamma_d^k \sim N(0, 10)$ e para $\beta_1^k \sim N(0, \sigma_{k1}^2)$, em que $\sigma_{k1}^2 \approx 0$. Lembrando que no modelo β_e^k são os parâmetros de efeito, γ_d^k são os parâmetros de variação da média e τ_k^2 os parâmetros da variância. Assumindo prioris independentes para esses argumentos, tem-se que

$$P(\theta) = \prod_{k=0}^K P(\tau_k^2) P(\beta_1^k) P(\gamma_d^k). \quad (3.5)$$

Os dados da função de verossimilhança são os resultados das eleições passadas, ou seja, $(\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{E-1}])$ e uma matriz de preditores também relativos às eleições passadas, $(\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{E-1}])$, em que $\mathbf{x}_e = [x_{1e}, \dots, x_{Pe}]$.

Pelo teorema de Bayes, a posteriori é dada por

$$P(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{X}) \propto P(\mathbf{V} \mid \boldsymbol{\theta}, \mathbf{X})P(\boldsymbol{\theta})$$

Como a função preditiva prediz as ações futuras, neste caso, prediz o percentual de votos da eleição futura, \mathbf{v}_E dado o respectivo conjunto de co-variáveis \mathbf{x}_E .

$$P(\mathbf{v}_E \mid \mathbf{x}_E) = \int_{\boldsymbol{\theta}} P(\mathbf{v}_E \mid \boldsymbol{\theta}, \mathbf{x}_E)P(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{X})d\boldsymbol{\theta} \quad (3.6)$$

Como citado anteriormente, o modelo completo se baseia na ideia de que três fatores preveem os resultados das eleições, que são: apego partidário a longo prazo, dinâmica de campanha em curto prazo e características institucionais, o que faz com que esse modelo possa ser usado em diferentes e amplos cenários. Porém esses fatores podem não ser os únicos na modelagem para previsão de resultados de uma eleição. De fato ao nos depararmos com as eleições presidências brasileiras recentes, podemos observar que fatores externos e aleatórios estão ligados diretamente a estimação dos votos compartilhados, por exemplo, a morte do candidato Eduardo Campos em 2014 ou o atentado que o então candidato Jair Bolsonaro sofreu no decorrer da campanha eleitoral, que mudaram o rumo das eleições.

É necessário construir um modelo que estime o nível atual de apoio partidário com base em pesquisas pré-eleitorais publicadas durante a campanha eleitoral, e que combine com as previsões do modelo A. Sendo assim, são utilizados os resultados das pesquisas publicadas por diferentes institutos de pesquisa.

Seja $\mathbf{y}_{ct} = (y_{ct1}, \dots, y_{ctP})$ uma variável aleatória p-dimensional que indica a porcentagem de intenção de votos no tempo t , $t = 1, \dots, T$, dado pelo instituto de pesquisa c , $c = 1, \dots, C$, em que p representa o partido/candidato, e $p = 1, \dots, P$. Sendo assim, é considerado que \mathbf{y}_{ct} é gerado por um processo multinomial, em que $\boldsymbol{\pi}_{ct}^* = (\pi_{ct1}^*, \dots, \pi_{ctP}^*)$ é o vetor de suporte esperado no dia t do instituto de pesquisa c .

$$\mathbf{y}_{ct} \sim \text{Multinomial}(\boldsymbol{\pi}_{ct}^*, N_{ct}), \quad (3.7)$$

em que N_{ct} representa o tamanho amostral da pesquisa eleitoral no tempo t publicada pelo instituto de pesquisa c .

A soma da proporção de todos os votos em uma pesquisa é de 100%. Para explicar isso e converter as proporções em um vetor de quantidades ilimitadas e com valor real, deve-se empregar uma transformação de log-razão. Cada entrada do vetor de suporte de votos compartilhados π_{ct}^* no tempo t pelo instituto de pesquisa c é dividido pelo último elemento do vetor π_{ctP}^* , depois é aplicado o logaritmo:

$$\pi_{ct} = alr(\pi_{ct}^*) = \left(\log\left(\frac{\pi_{ct1}^*}{\pi_{ctP}^*}\right), \dots, \log\left(\frac{\pi_{ct(P-1)}^*}{\pi_{ctP}^*}\right) \right) = (\pi_{ct1}, \dots, \pi_{ct(P-1)}) \quad (3.8)$$

Pela teoria da confiabilidade (mais informações sobre esse tema podem ser vistas em Drenick (1963)) pode-se escrever π_{ct} como uma combinação linear da seguinte forma, $\pi_{ct} = \alpha_t + \delta_c$, em que α_t indica o nível de apoio ao partido/candidato ao longo do tempo, “a verdadeira taxa” de apoio. Já δ_c representa o “nível de apoio” de cada instituto de pesquisa para cada um dos partidos/candidatos, ou seja, é como de fosse uma tendência do instituto com o partido/candidato (de modo geral, não intencional, que pode ser gerada pelo plano amostral adota, por exemplo). O vetor $\alpha_t = (\alpha_{t1}, \dots, \alpha_{tP})$ tem sua evolução dada por:

$$\alpha_t = \alpha_{t+1} + \mathbf{w}_t, \text{ em que } \mathbf{w}_t \sim N(0, \mathbf{W}) \quad (3.9)$$

Note que o α_t é igual a α_{t+1} mais um erro aleatório, ou seja temos um passeio aleatório *backward*, que é um passeio aleatório só que de modo reverso. Isso permite estimar os níveis de apoio do partido para cada dia, mesmo sem o lançamento de uma nova pesquisa eleitoral.

A variância deste erro aleatório, também é chamada de variância evolucional \mathbf{W} , descreve a taxa de mudança entre quaisquer dois dias consecutivos. A matriz \mathbf{W} é construída de modo que seja constante ao longo do tempo, mas permite explicitamente que os estados latentes covariem:

$$\mathbf{W} = \begin{bmatrix} w_1^2 & \dots & w_{1,P-1} \\ \vdots & \ddots & \vdots \\ w_{1,P-1} & \dots & w_{P-1}^2 \end{bmatrix}. \quad (3.10)$$

O processo do passeio aleatório para trás na distribuição conjunta a posteriori isola

a priori no dia da eleição. Isso, por sua vez, permite que haja uma integração entre a previsão do modelo A ao modelo B, definindo o estado latente do apoio partidário no dia da eleição α_T igual ao posteriori transformada pela razão logarítmica da distribuição preditiva a partir do modelo A:

$$P(\alpha_T | \mathbf{x}_E) = alr \left(\int_{\theta} P(\mathbf{v}_E | \theta, \mathbf{x}_E) P(\theta | \mathbf{V}, \mathbf{X}) d\theta \right). \quad (3.11)$$

Relembrando, seja $\mathbf{X} = (X_1, \dots, X_n)$, a função alr é definida pelo logaritmo da divisão dos (X_1, \dots, X_{n-1}) por (X_n) . Ou seja,

$$alr(\mathbf{X}) = \left(\log\left(\frac{X_1}{X_n}\right), \dots, \log\left(\frac{X_{n-1}}{X_n}\right) \right). \quad (3.12)$$

Com essa integração, pode-se obter a distribuição a posteriori conjunta do modelo A e do modelo B que é proporcional ao produto de suas verossimilhanças, da distribuição preditiva a posteriori, do passeio aleatório para trás e as prioris dos parâmetros dos dois modelos. Ao integrar e estimar todas as partes do modelo em uma especificação conjunta, obtêm-se várias vantagens, pois assim através de métodos MCMC pode-se encontrar aproximadamente a distribuição a posteriori para qualquer um dos parâmetros de interesse. A mais importante delas é que assim considera-se automaticamente a completa incerteza do modelo A, que é relevante quando busca-se ponderar com precisão as pesquisas e os fundamentos para uma previsão sintética.

Temos assim a seguinte distribuição a posteriori para os parâmetros desconhecidos:

$$\begin{aligned} P(\alpha, \delta, \mathbf{W}, \theta | \mathbf{Y}, \mathbf{V}, \mathbf{X}) &\propto P(\mathbf{Y} | \alpha, \delta, \mathbf{W}) P(\mathbf{V} | \theta, \mathbf{X}) P(\alpha_T | \mathbf{x}_E) \\ &\times \prod_{t=1}^{T-1} P(\alpha_t | \alpha_{t+1}, \mathbf{W}) P(\theta) P(\mathbf{W}) P(\delta), \end{aligned} \quad (3.13)$$

em que \mathbf{Y} é uma matriz empilhada com todas as pesquisas observadas, \mathbf{y}_{ct} no tempo t referente ao instituto c . Já $\alpha = [\alpha_1, \dots, \alpha_{T-1}]'$ e $\delta = [\delta_1, \dots, \delta_C]'$.

$P(\delta)$ é especificada como sendo a priori de “*house effects*”, a priori que informa possíveis tendências não intencionais dos institutos de pesquisas. Supondo que em média

espera-se que não haja “*house effects*”, pode-se especificar $\delta_{cp} \overset{ind}{\sim} N(0, 1)$, para cada elemento de $\boldsymbol{\delta}$ e assumindo que um instituto de pesquisa não interfere no outro. $P(\mathbf{W})$ ¹ é especificada a priori fracamente informativa LKJ com $\nu = 50$. Mais detalhes sobre esta priori podem ser vistos em Lewandowski, Kurowicka e Joe (2009). A priori para $\boldsymbol{\theta}$ foi especificada na equação (3.5).

De fato, após a combinação dos dois modelos, nosso interesse concentra-se na estimação do $\boldsymbol{\alpha}$ e $\boldsymbol{\delta}$, que são os parâmetros utilizados para a estimação dos votos compartilhados.

No próximo capítulo é feito o ajuste do modelo completo para um conjunto de dados referente as eleições do poder executivo no ano de 2018, mais precisamente a eleição presidencial. Como o foco deste modelo é para cenários multipartidários, a análise é feita apenas para o primeiro turno.

¹A priori para uma matriz de correlação é chamada de LKJ, cuja densidade é proporcional ao determinante da matriz de correlação elevada à potência de uma regularização positiva de um parâmetro menos um.

4 Implementação do modelo completo e resultados

Neste capítulo é proposta a aplicação do modelo completo estudado a um conjunto de dados de previsões feitas ao longo do ano de 2018, em que o principal interesse era prever o resultado da eleição presidencial, dados estes que foram cedidos por Neale El-Dash (2010).

4.1 Nova República: Um resumo

O período atual da democracia brasileira é conhecido como Nova República, que após a formação da constituição de 1988, implementou novamente as eleições diretas no Brasil.

A eleição presidencial de 1989 elegeu como presidente Fernando Collor de Melo, após 29 anos sem eleições diretas. O plano Collor foi caracterizado por: substituir a moeda circulante pelo cruzeiro; abertura do comércio brasileiro à importações e capital estrangeiro, o que ocasionou a fusão e/ou quebra de várias empresas brasileiras que não estavam aptas a competir com o mercado internacional; empresas estatais foram privatizadas e órgãos públicos extintos, ocasionando milhares de demissões. Estas foram algumas medidas implementadas durante o governo Collor.

Em 1991 começou o processo de *impeachment*, em que milhares de pessoas foram a rua protestar contra o atual governo. Em dezembro de 1992 Fernando Collor renunciou ao cargo de presidente, porém ainda assim foi julgado e teve seus direitos políticos cassados por oito anos.

Entre final de 1992 e 1994, após a renúncia de Collor, o então vice-presidente, Itamar Franco assumiu o cargo de presidente. Em abril de 1993, Itamar Franco convocou o plebiscito, previsto na Constituição de 1988, para que a população decidisse sobre a forma de governo (monarquia ou república) e o sistema de governo (presidencialismo ou parlamentarismo), tendo a maioria votado pela continuação da república presidencialista.

Em 94, Itamar lançou o plano real, plano que tornava viável a estabilização econômica criado pela equipe do então, ministro da Fazenda, Fernando Henrique Cardoso (FHC).

Após a criação do plano real, o resultado foi imediato. Houve a queda da inflação, que passou de cerca de 50% em junho de 1994 caiu 3% no final do ano. Com o sucesso do plano real, Fernando Henrique Cardoso (FHC) tornou-se muito popular, fazendo com que o mesmo concorresse ao cargo mais alto do poder executivo, ganhando as eleições de 1994 no primeiro turno, com cerca de 54% dos votos. O seu primeiro mandato caracterizou-se pelo controle da inflação brasileira.

Ainda durante o primeiro governo FHC algumas questões sociais tornaram-se bastante evidentes, como a batalhas entre forças policiais e integrantes do Movimento dos Trabalhadores Rurais Sem Terra (MST) que evidenciaram o problema da questão da terra no Brasil. Também foram criadas medidas de incentivo à cultura, como a ampliação do orçamento do Ministério da Cultura e mudanças na lei Rouanet, que beneficiaram o cinema nacional.

O segundo mandato de Fernando Henrique Cardoso iniciou-se em primeiro de janeiro de 1998 finalizando em primeiro de janeiro de 2002. O presidente eleito derrotou Luiz Inácio Lula da Silva no segundo turno com 53% dos votos válidos, tornando-se o primeiro presidente a ter dois mandatos consecutivos. Por outro lado houve a estagnação da economia, perpetuação da desigualdade social, em que grande parte da população tinha pouco poder aquisitivo e a saúde brasileira também foi duramente criticada pelos órgãos internacionais.

Após três tentativas, Luiz Inácio Lula da Silva, mais conhecido como Lula, assumiu o cargo de presidente do Brasil em primeiro de janeiro de 2003. O primeiro governo Lula conseguiu manter a estabilidade da moeda real criada pelo antigo governo, além de fortalecer e elaborar programas de cunho social, como bolsa família, a expansão do crédito e a geração de empregos formais. Em 2005, foi descoberto um escândalo político, que posteriormente, foi chamado de Mensalão, esquema político que beneficiava parlamentares que apoiavam o governo no congresso, segundo denúncias.

Durante seu primeiro mandato, o Brasil ingressou no G-20; ampliou as relações diplomáticas do país com a África e o Oriente Médio; teve início a campanha pela inclusão do Brasil no Conselho de Segurança da Organização das Nações Unidas (ONU) como membro permanente.

Lula conseguiu se reeleger presidente do Brasil dado o seu alto índice de aprovação

perante a população brasileira. Seu segundo mandato começou em primeiro de janeiro de 2007 e finalizou em primeiro de janeiro de 2011. Este mandato de Lula caracterizou-se pelo crescimento econômico nacional, redução da pobreza e o início da ampliação do acesso de pessoas com menos recursos ao ensino superior/tecnológico. Também foi marcado por crises de corrupção.

Dilma Rousseff foi a primeira mulher a assumir o cargo de presidente do país e com o apoio do até então presidente Lula, se elegeu pelo PT, tendo seu primeiro mandato iniciado em primeiro de janeiro de 2011. Durante o primeiro mandato deu continuidade as ações do antigo presidente, investiu na educação, principalmente no ensino superior, criando mais institutos de pesquisa e universidades. Ampliou o Enem (Exame Nacional do Ensino Médio) que passou a ser a principal porta de acesso ao ensino superior. Além disso criou o PRONATEC (Programa Nacional de Acesso ao Ensino Técnico e Emprego) e deu maior visibilidade e verba aos programas de incentivo, como FIES (Financiamento Estudantil) e ProUni (Programa Universidade para Todos). Porém como nos outros governos PTistas, houve grande crítica a corrupção política.

Durante meados de 2013 aconteceram várias manifestações em todo o país iniciadas contra o aumento da passagem no transportes coletivos, levando milhares de brasileiros às ruas. Contudo, as reivindicações tornaram-se mais amplas, como o fim da corrupção, melhorias na saúde e na educação públicas. Dilma foi reeleita para o segundo mandato em uma das eleições presidências mais acirradas da história brasileira. O segundo mandato de Dilma foi marcado pelo *impeachment* que a mesma sofreu, tendo seu mandato interrompido em 31 de agosto de 2016. As acusações versaram sobre desrespeito à lei orçamentária e à lei de improbidade administrativa por parte da presidente, além de lançarem suspeitas de envolvimento da mesma em atos de corrupção na Petrobras.

Michel Temer assumiu como presidente do Brasil logo após o *impeachment* de Dilma Rousseff, tendo seu mandato finalizado em primeiro de janeiro de 2018. Seu governo foi marcado por diversas críticas, apesar disso economicamente falando, o país começou a reagir da forte recessão econômica que enfrentava.

As eleições do poder executivo de 2018, mais especificamente ao cargo de presidente foi marcada por um processo bastante tumultuado. No início das campanhas eleitorais, o até então candidato do Partido dos Trabalhadores (PT), Luíz Inácio Lula da Silva, encontrava-se em primeiro lugar nas pesquisas de intenção de votos, porém o mesmo tornou-se inegável, após ser preso, o que ocasionou grande confusão no cenário político. Os dados disponibilizados para esta dissertação levam em conta apenas as pesquisas com o

candidato Fernando Haddad, substituto de Lula na campanha eleitoral. . Todavia, ainda em 2018, tivemos o evento da “Facada do Bolsonaro” que modificou as regras e limitou os debates dos candidatos na TV. Além disso, tivemos a influência da internet, das *fake news*, que interferiram nas eleições e influenciaram (ou não) na capacidade preditiva dos modelos estatísticos.

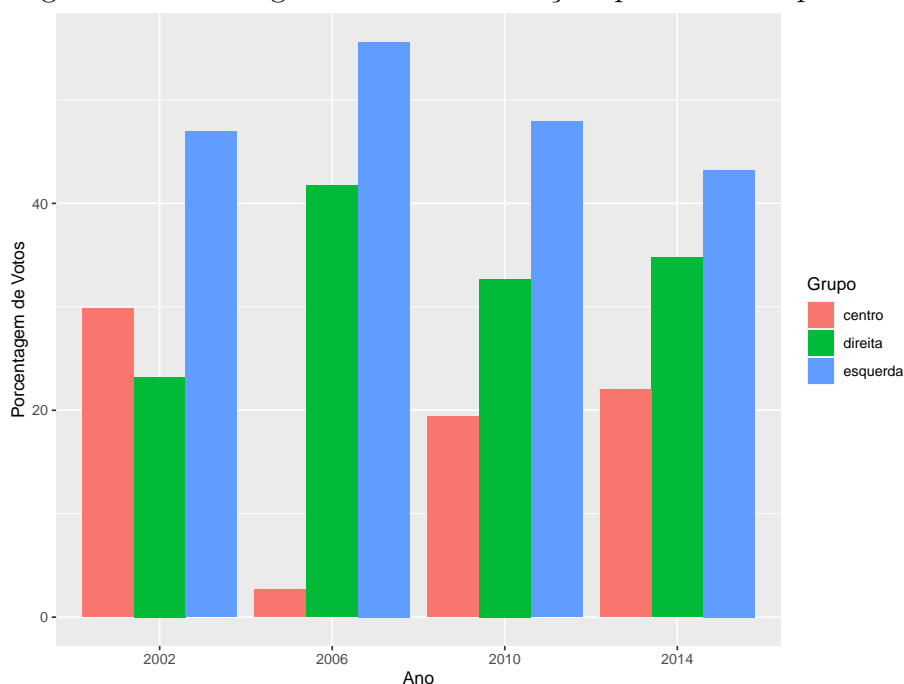
Como já ressaltado anteriormente nesse texto, o Brasil é um país multi-partidário, ou seja, há inúmeros partidos. Nesta dissertação esses partidos são divididos em três grupos baseados em ideologia: direita, esquerda e centro. Para cada pesquisa eleitoral e resultado de eleições passadas foi feito um agrupamento dos partidos em um destes três grupos com base na Tabela 1.

Mais detalhes sobre este período podem ser encontrados em Mota e Braick (2016).

4.2 Aplicação para as Eleições Executivas de 2018

A análise exploratória dos dados é de grande importância para resumir e entender as principais características das variáveis de um conjunto de dados. A Figura 1 apresenta as proporções de votos em cada grupo (direita, centro e esquerda) no primeiro turno das eleições presidenciais desde 2002. Toda análise foi feita utilizando o RStudio (2020) e os pacotes RStan (2020) e o Ggplot2 (2016).

Figura 1: Porcentagem de votos da eleições presidenciais passadas



Geralmente na previsão de votos de um certo candidato ou partido, existem algumas variáveis que podem ser utilizadas a fim de explicar a porcentagem de votos - medida de crescimento do produto interno bruto (PIB), taxa de inflação (IPCA) e taxa de desemprego (TD). Essas quantidades são definidas a seguir:

- PIB - O produto interno bruto é um indicador econômico amplamente utilizado que apresenta a soma de todos os bens e serviços produzidos por um país (região), durante um período de tempo. Nesta dissertação foi utilizado o PIB médio anual.
- IPCA - É um índice econômico que tem por objetivo medir a inflação, um dos índices mais tradicionais e importantes do Brasil. Criado em 1979, ele mede a variação dos preços de um conjunto de produtos e serviços vendidos no varejo e consumidos pelas famílias brasileiras.
- TD - A taxa de desemprego é um indicador socioeconômico divulgado pela PNAD contínua. Os valores são determinados a partir de estudos feitos em determinado período de tempo com a população economicamente ativa (PEA) ¹. A taxa de desemprego é calculada da seguinte forma

$$\frac{PD}{PEA} \times 100,$$

em que PD representa a taxa de desocupados.

A Figura 2 apresenta os dados médios referentes as variáveis mencionadas acima em cada ano eleitoral.

¹A Pnad considera o PEA como a soma das pessoas ocupadas ou não, pessoas com 15 anos de idade ou mais.

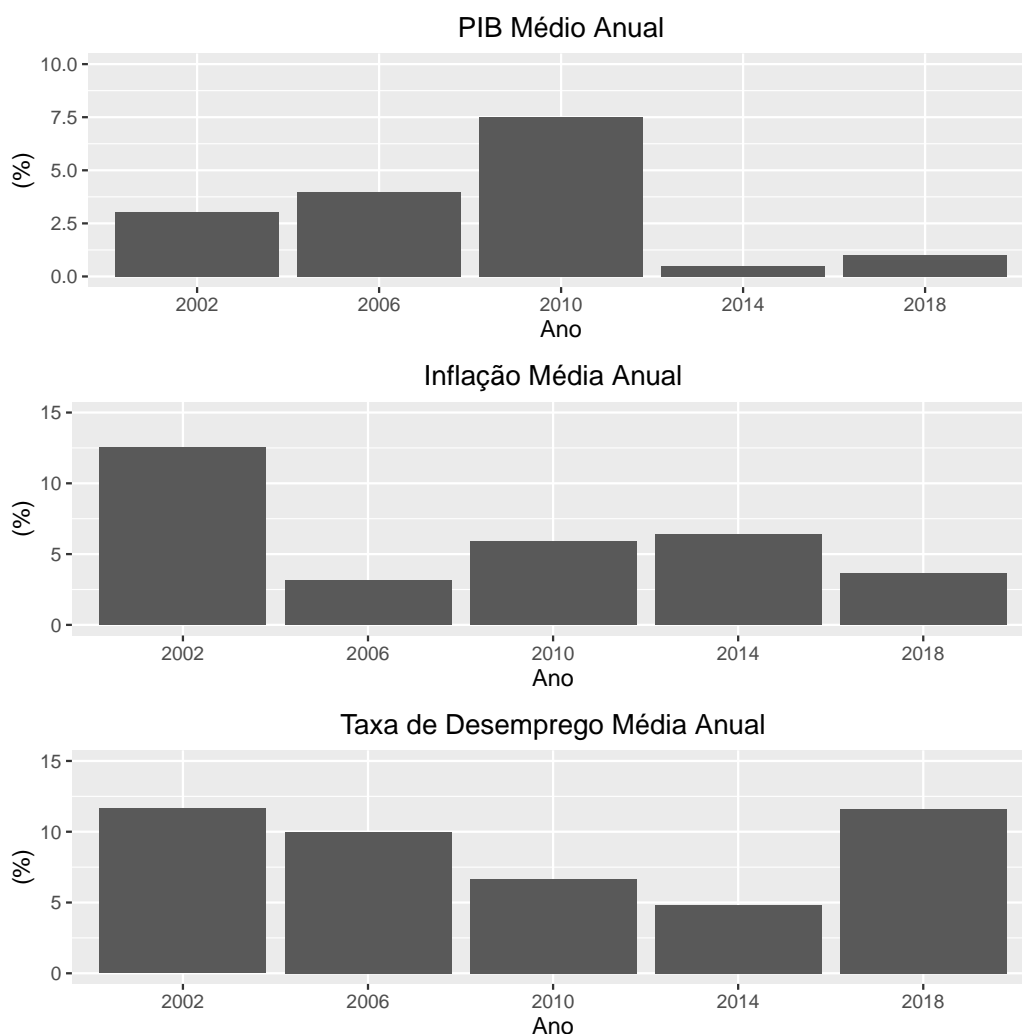


Figura 2: Taxas médias em proporção das variáveis explicativas mencionadas, PIB, Inflação e Taxa de Desemprego em anos eleitorais de 2002 à 2018.

É importante ressaltar que ao longo dos anos a maneira como é calculada a taxa de desemprego foi alterada, o que pode interferir no resultado final ou não.

O banco de dados contempla 12 institutos de pesquisas. São eles: Pesquisa 365, MDA, Data Folha, Data Poder 360, Parana Pesquisas, Ipespe, Ibope, FSB, Vox Populi, Real Time Big Data, Virtu Analise e Sensus.

Ao todo 13 candidatos disputaram as eleições presidências de 2018: Álvaro Dias (PODE - Podemos), coligação: PODE, PRP, PSC e PTC; Cabo Daciolo (PATRI - Patriota), coligação: partido isolado; José Maria Eymael (DC - Democracia Cristã), coligação: partido isolado; Ciro Gomes (PDT - Partido Democrático Trabalhista), coligação: PDT e AVANTE; Geraldo Alckmin (PSDB - Partido Social da Democracia Brasileira), coligação: PSDB, PTB, PP, PR, DEM, SOLIDARIEDADE, PPS, PRB e PSD; Guilherme Boulos

(PSOL - Partido Socialismo e Liberdade), coligação: PSOL e PCB; Henrique Meirelles (MDB - Movimento Democrático Brasileiro), coligação: MDB e PHS; Jair Bolsonaro (PSL - Partido Social Liberal), coligação: PSL e PRTB; João Amoêdo (NOVO - Partido Novo), coligação: partido isolado; João Goulart Filho (PPL - Partido Pátria Livre), coligação: partido isolado; Fernando Haddad (PT - Partido dos Trabalhadores), coligação: PT, PCdoB e PROS; Marina Silva (REDE - Rede Sustentabilidade), coligação: REDE e PV; e por ultimo, Vera Lúcia (PSTU - Partido Socialista dos Trabalhadores Unificado), coligação: partido isolado.

No banco de dados e consequentemente na análise exploratória, apenas os candidatos com maior intenção de votos foram considerado de forma singular. São eles: Ciro Gomes (PDT), Geraldo Alckmin (PSDB), Jair Bolsonaro (PSL) e Fernando Haddad (PT), enquanto os demais foram enquadrados na categorias “outros”.

A Tabela 2 apresenta quantos indivíduos foram entrevistados por cada instituto de pesquisa durante a corrida presidencial.

Tabela 2: Total de pessoas entrevistadas por instituto

Instituto de pesquisa	Pessoas entrevistadas
Pesquisa 365	4000
MDA	26026
Data Folha	159270
Data Poder 360	75000
Parana Pesquisas	16624
Ipespe	42000
Ibope	52620
FSB	22000
Vox Populi	8000
Real Time Big Data	41000
Virtu Analise	4000
Senses	4000

Para efeito de comparação entre as pesquisas dos institutos mencionados foi feito o agrupamento de alguns institutos, sendo assim considerados os institutos Data Folha, Data Poder 360, Ibope, Real Time Big Data, Ipespe e outros. O Grupo “outros” formado pelos institutos Pesquisa 365, MDA, Parana Pesquisas, FSB, Vox Populi, Virtu Analise e Senses. Nesta primeira parte do modelo essa informação não será utilizada.

A próxima figura apresenta a evolução da proporção de votos de cada candidato semanas antes das eleições até o dia da votação, no dia 07 de outubro de 2018, ajustando um modelo através do método *loess*².

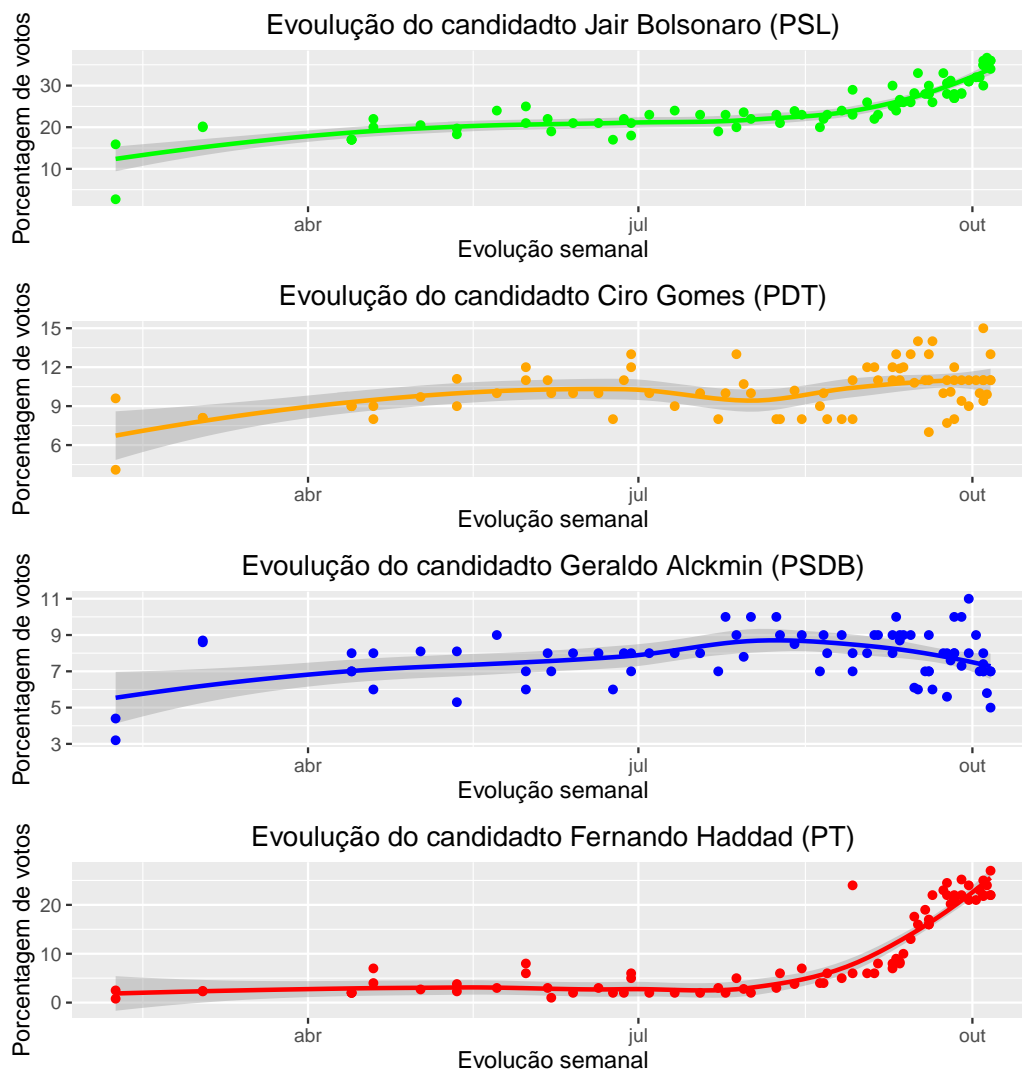


Figura 3: Evolução dos candidatos à presidência.

Note que apenas os candidatos do PSL e PT, ao longo da corrida presidencial, obtiveram um crescimento em relação ao início das pesquisas, enquanto que os outros continuaram de forma estacionária, sem grandes mudanças.

Vale destacar que em uma eleição, os votos brancos, indecisos e nulos são extremamente importantes. Eles podem ser um componente decisivo em uma eleição. As Figuras 4 e 5 apresentam o comportamento da porcentagem de votos em outros candidatos (Figura 4) e de votos brancos e nulos (Figura 5) ao longo dos dias antecedentes ao dia da

²Regressão Local (Loess) é um método não paramétrico que estima curvas e superfícies através de suavização (*smoothing*). Mais informações podem ser vistas em Cleveland (1981).

eleição.

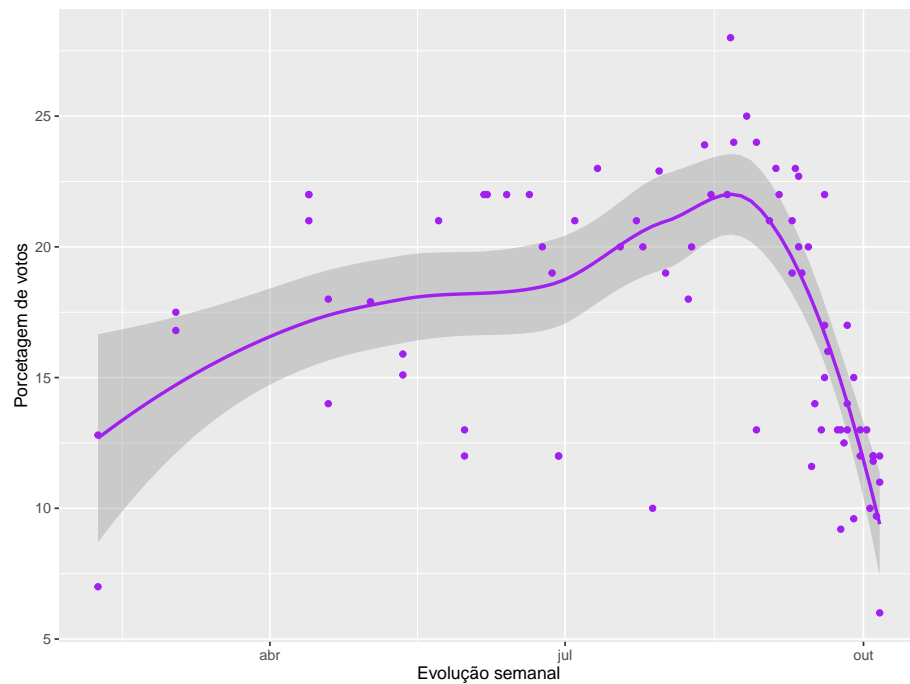


Figura 4: Evolução dos votos em outros candidatos.

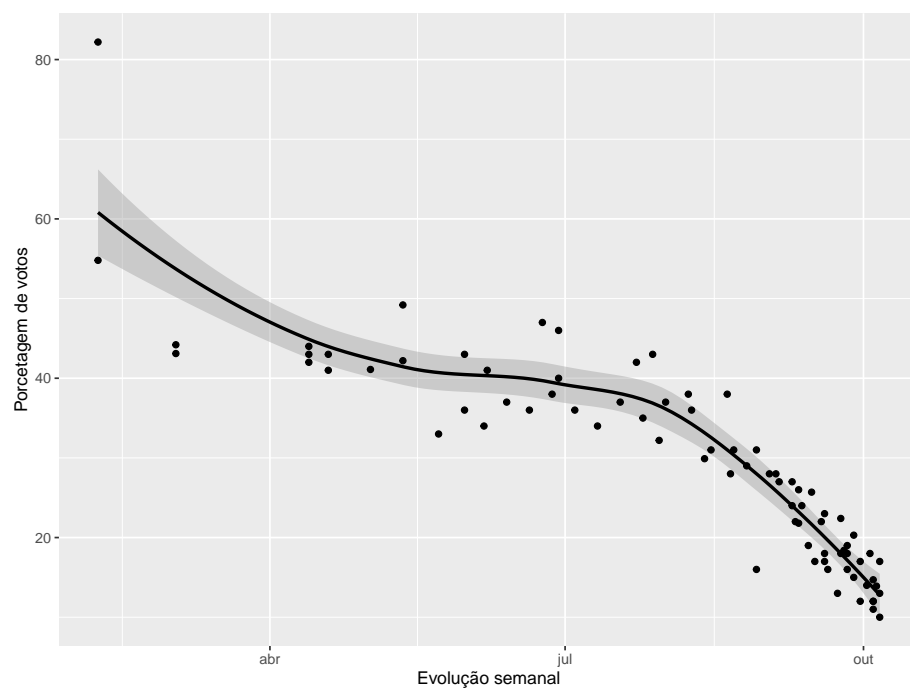


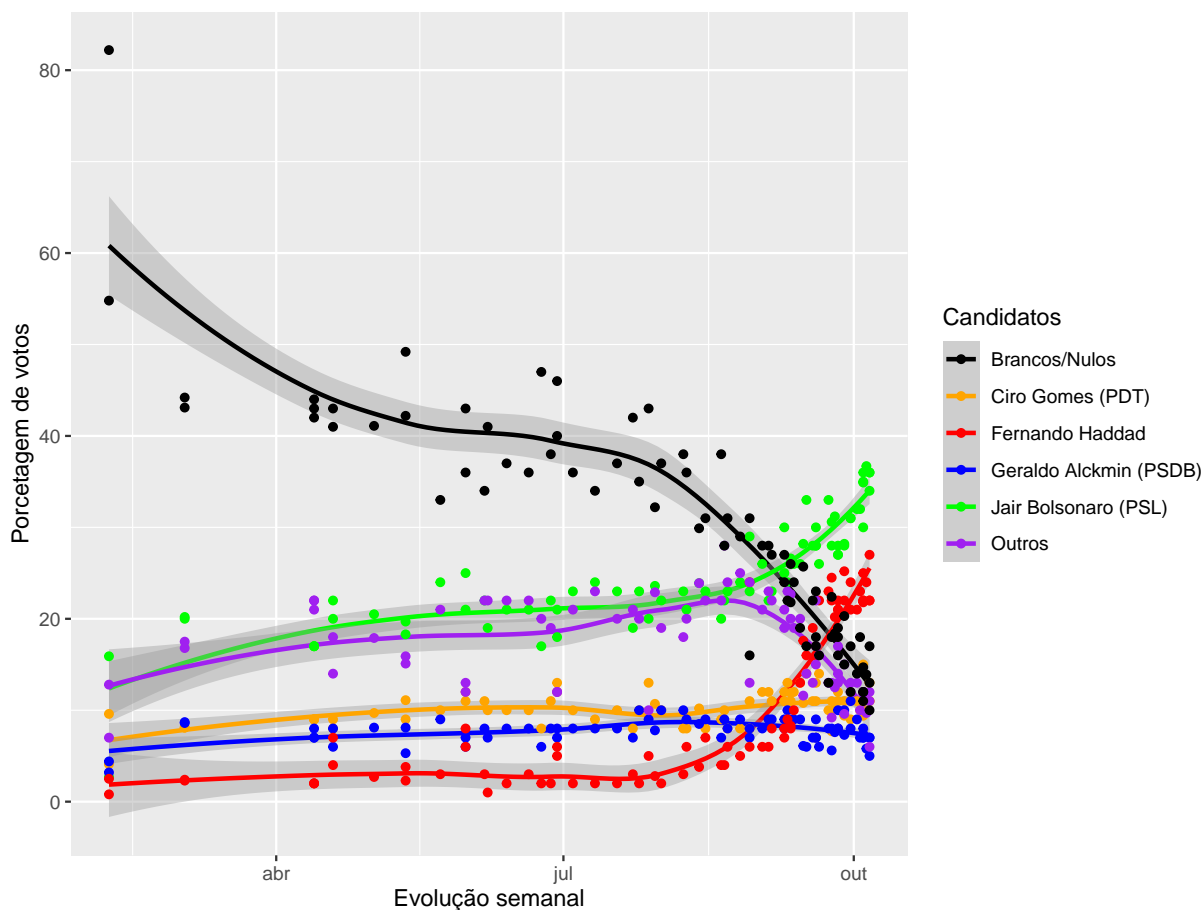
Figura 5: Evolução dos votos em brancos/nulos.

A Figura 4 apresenta a evolução dos votos em outros candidatos, ou seja, aqueles que não foram analisados individualmente. A Figura 5 apresenta a evolução dos votos brancos

e nulos. Observe que os votos nos outros candidatos parece, apresentar um crescimento até meados de agosto. Porém, a partir desse momento, os votos nesse grupo de candidatos decresce bruscamente com a aproximação do dia da votação. Os votos brancos e nulos, por sua vez, começam de forma bem elevada, ou seja, a população estava indecisa sobre em quem votar, e com o passar do tempo esses indivíduos escolham seus respectivos candidatos, fazendo com que essa índice diminua de forma significativa.

A Figura 6 apresenta a evolução de todos os candidatos durante a disputa eleitoral somado à evolução dos votos em outros candidatos e branco/nulos. Também é apresentado um modelo ajustado através do método *lowess* para cada uma dessas series e seus respectivos intervalos de confiança.

Figura 6: Evolução de todas as séries durante o 1º turno.



É interessante frisar que a eleições de 2018 marcaram uma nova era no país, na qual o candidato que aparecia com maior índice não era do PT e nem do PSDB, partidos que disputavam a preferência brasileira há quase 20 anos.

4.2.1 Aplicação do modelo completo

Nesta seção o modelo completo apresentado no capítulo anterior será aplicado aos dados das eleições presidenciais de 2018. As variáveis explicativas usadas são: votos compartilhados da eleição passada para cada um dos grupos abordados, intenção de votos entre 200 e 230 dias antes da data da eleição, variável indicadora (o valor 1 representa candidato/partido no poder, e 0 caso contrário), PIB médio, taxa média de inflação e taxa média de desemprego.

São apresentados a seguir os resultados obtidos para previsão de votos compartilhados com base no modelo completo e dados de eleições passadas (a partir de 1989). A previsão será feita para a porcentagem de votos em cada um dos três grupos propostos: direita, esquerda e centro.

A Figura 7 apresenta as previsões finais e seus respectivos intervalos de credibilidade (83% e 95%) do modelo proposto, dois dias antes da eleição.

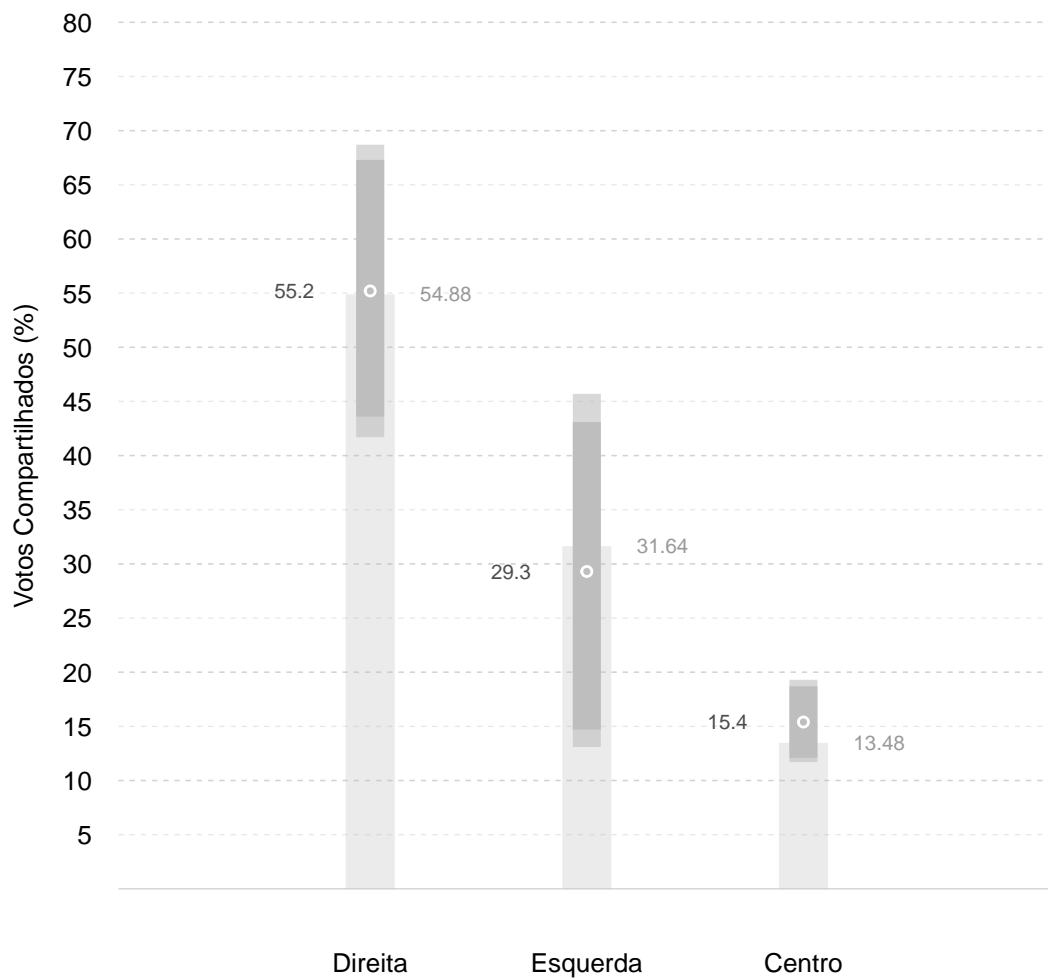


Figura 7: Previsão das eleições de 2018, dois dias antes da eleição.

A estimativa pontual é representada pelo círculo branco e é acompanhada dos intervalos de credibilidade (83% - cinza escuro) e (95% - cinza claro). A barra em cinza claro representa o resultado da eleição para cada um dos grupos, direita com 54.88%, esquerda com 31.64% e centro com 13.48%. Como apresentado na Figura 7, essas previsões finais estão razoavelmente perto dos resultados finais.

As Tabelas 3 e 4 apresentam os votos compartilhados para cada grupo após o resultado da eleição, a estimativa pontual (média a posteriori) obtida através do modelo, o desvio padrão e os limites dos dois intervalos mencionados acima. Nas Tabelas, V. C. é a abreviação de Voto Compartilhado, L. I. a abreviação de Limite e Inferior e L. S. a abreviação de Limite Superior.

Tabela 3: Porcentagem real para cada grupo, estimativa pontual para cada um dos grupos e desvio padrão.

Grupo	V. C. (%) Real	V. C. (%) Estimado	Desvio Padrão
Direita	54.88	55.2	9.25
Esquerda	31.64	29.3	11.58
Centro	13.48	15.4	2.60

Tabela 4: Llimites do intervalo de credibilidade de 83% e de 95% respectivamente.

Grupo	L. I. (83%)	L. S. (83%)	L. I. (95%)	L. S. (95%)
Direita	43.62	67.31	41.65	68.71
Esquerda	14.74	43.12	13.14	45.68
Centro	12.08	18.69	11.71	19.32

Note que a previsão do modelo completo se aproxima bastante do real valor e que em ambos cenários apresentados, os dois intervalos contém o real valor do parâmetro. Contudo, pode-se observar que a amplitude intervalar é expressiva em ambos os casos, indicando que há variabilidade alta no modelo.

Pode ser também de interesse investigar como os resultados do modelo se comportam ao longo da campanha. A Figura 8 destaca a característica central do modelo. Ao longo do tempo, mais perto do dia da eleição, as pesquisas se tornam mais informativas e fazem com que o peso do modelo A seja menos forte. Os filtros de componentes dinâmicos, as novas informações das pesquisas, refletem a dinâmica de curto prazo não explicada pelo componente baseado em fundamentos, e a sua incorporação contribui para melhorar a capacidade preditiva do modelo.

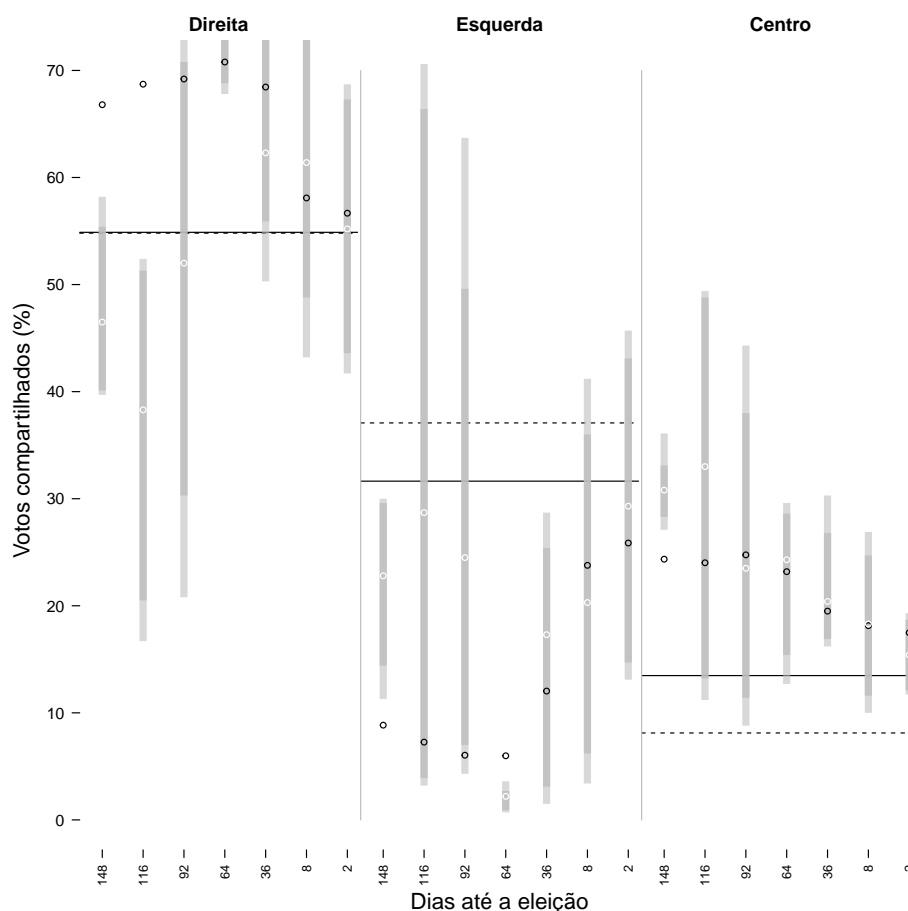


Figura 8: Desenvolvimento das previsões do compartilhamento de votos do modelo completo ao longo do tempo para a Eleição Federal de 2018.

A Figura 8 apresenta o ajuste do modelo completo iniciando em 148 dias até a previsão final, dois dias antes da eleição. Os pontos claros apresentam a previsão média; as barras em cinza escuro representam os intervalos de 83% de credibilidade e as barras em cinza claro representam os intervalos de 95% de credibilidade. A porcentagem de votos observada por cada grupo é indicada pela linha horizontal sólida. A previsão do modelo A é marcada pela linha horizontal tracejada. Os pontos escuros indicam a média mensal das pesquisas incluídas na estimativa do modelo B.

Note que para o grupo direita tanto o modelo completo quanto o modelo A apresentam estimativas próximas, enquanto que para os outros dois grupos, esquerda e centro, as estimativas para o modelo completo, ou seja, aquele que leva em consideração as pesquisas eleitorais, as estimativas pontuais (média a posteriori) mostram-se mais próximas dos valores reais.

A Figura 9 apresenta a relação dos preditores (votos compartilhados da eleição pas-

sada, intenção de votos - pesquisas eleitorais 200-230 dias antes da eleição e a variável dicotômica que indica se o partido/candidato esta no poder ou não) com a variável resposta que é a porcentagem de votos compartilhados, levando em conta os dados desde 1989.

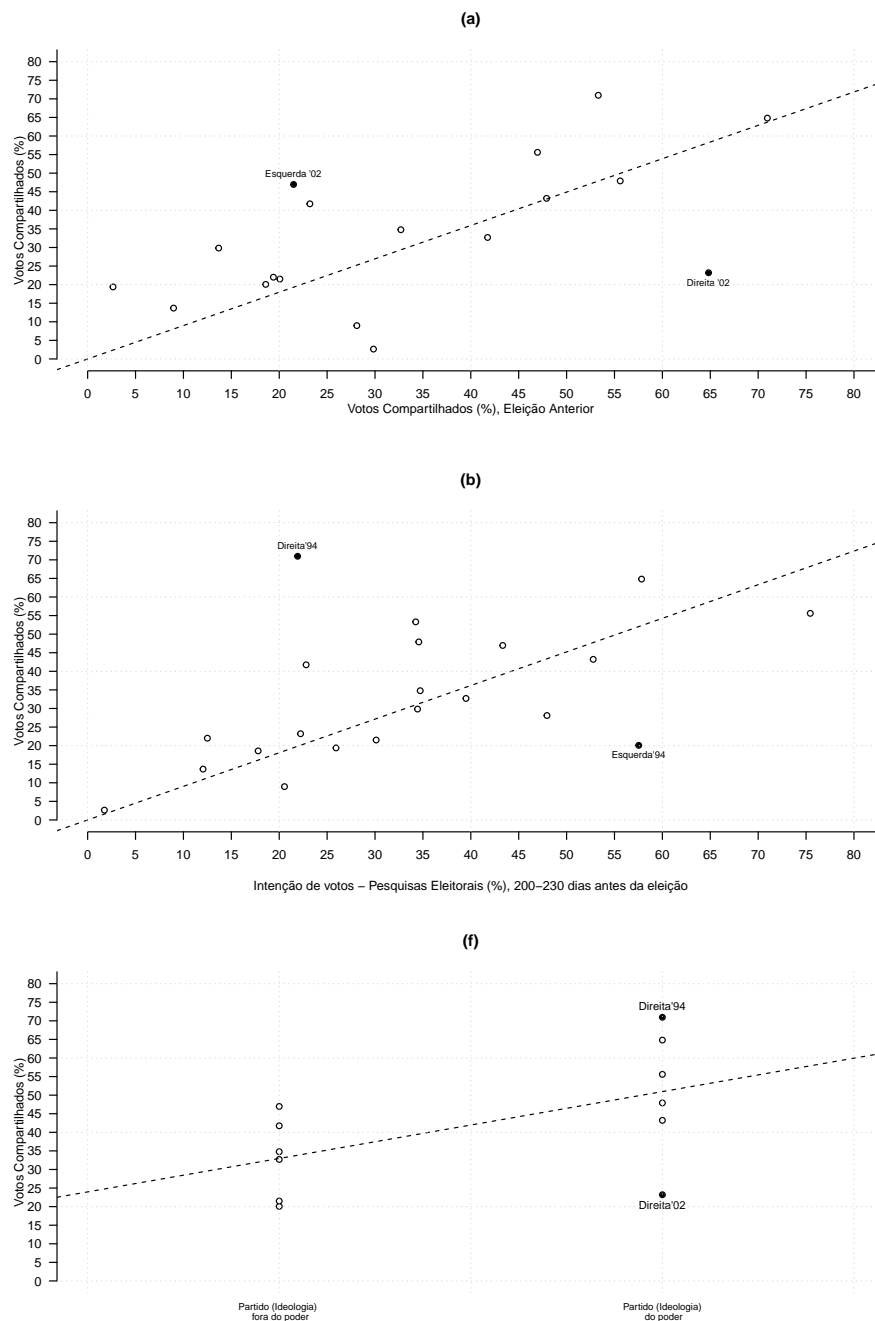


Figura 9: Relação entre a variável Votos Compartilhados (%) vs variáveis predictoras.

Em (a) e (b) nota-se existência de uma relação entre as duas variáveis de forma linear e forte. O painel (c) revela que o partido em exercício tem, em média, uma parcela de

votos maior do que o outro partido principal que não detém o poder. Os pontos em preto são os *outliers* obtidos na amostra.

A Figura da 10 apresenta a mesma relação entre a variável dependente com o restante das outras variáveis explicativas usadas no ajuste do modelo como explicado na Figura 9. Nesta figura, os pontos azuis representam as informações para o grupo direita, os pontos vermelhos para o grupo esquerda e os pontos verdes para o grupo centro. Note que as variáveis explicativas apresentadas na Figura 10 não apresentam relação aparente com a variável de interesse, ou seja, os votos compartilhados. É provável, entretanto, que exista uma relação entre essas variáveis e a porcentagem de votos no candidato/partido incumbente, que poderia ser explorada em um trabalho futuro.

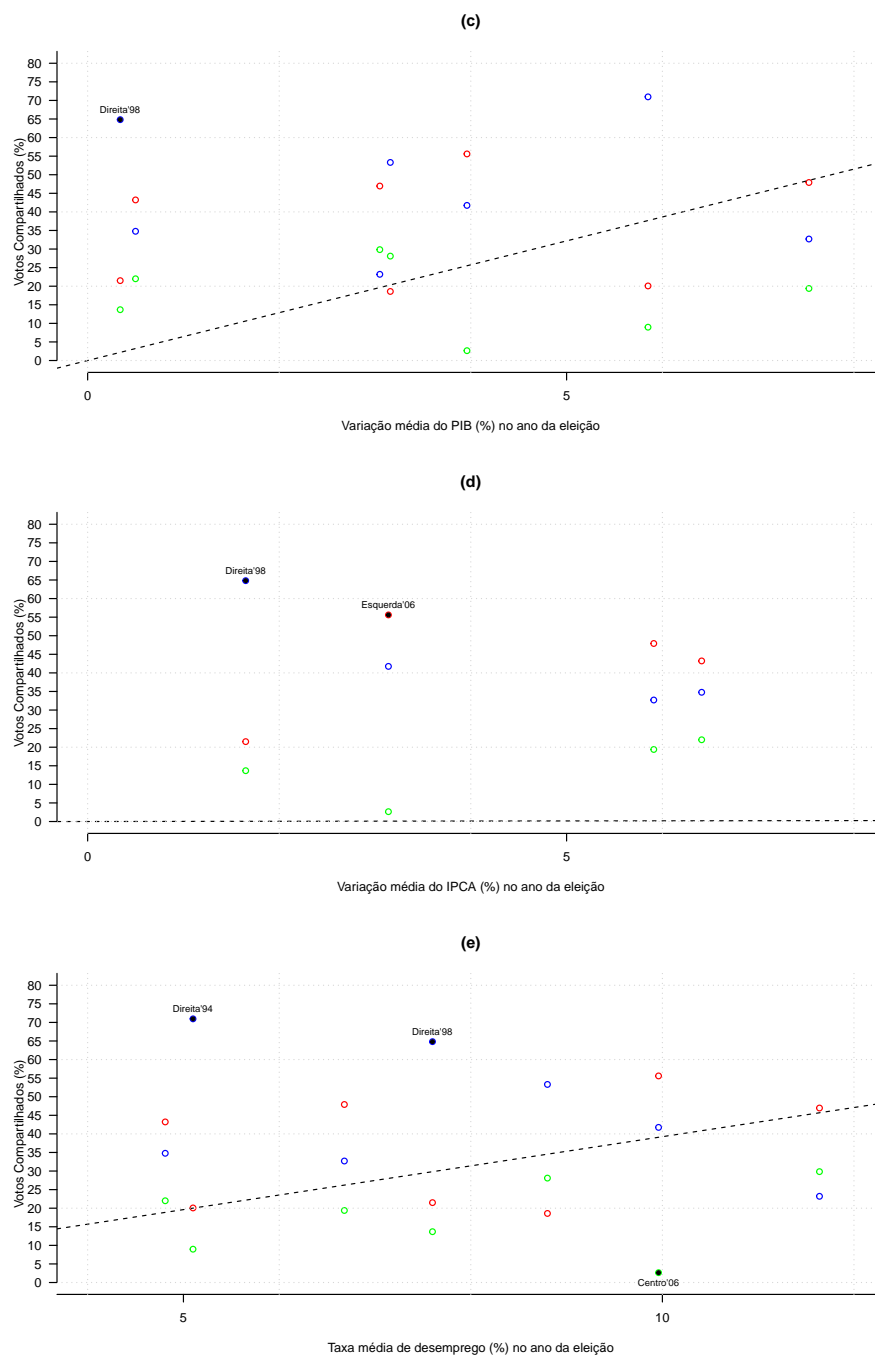


Figura 10: Relação entre a variável Votos Compartilhados (%) vs variáveis preditoras.

A Figura 11 apresenta as estimativas do coeficiente β_e relativo a cada variável explicativa que foi significativa no ajuste do modelo, variando por eleição (considerando as últimas sete). Através dos gráficos apresentados podemos analisar a flutuação da influência das variáveis regressoras para explicar as porcentagens de voto.

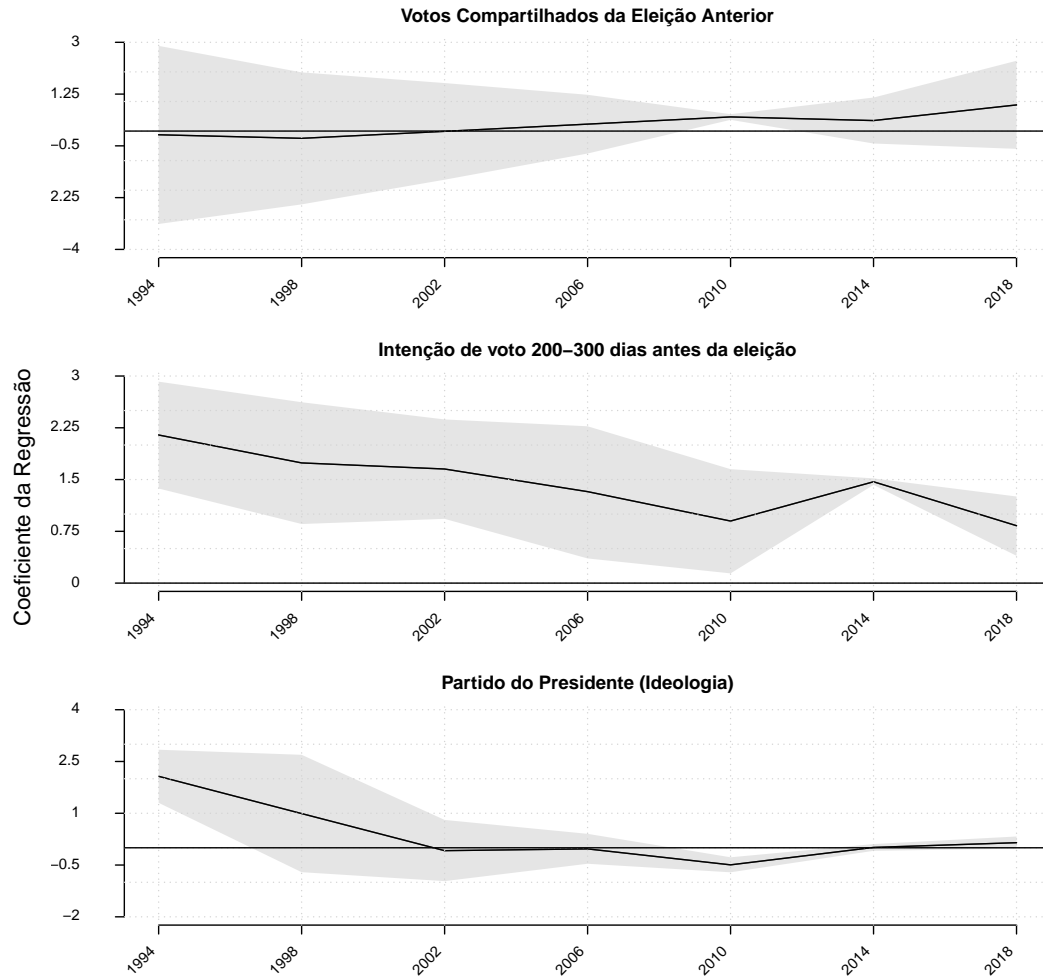


Figura 11: Evolução do coeficiente β_e relativo a cada variável explicativa durante as últimas 7 eleições.

Os resultados das eleições passadas mostrou-se um fator que oscila bastante ao longo do tempo na tentativa de explicar o resultado da eleição de 2018, na maior parte do tempo analisado o intervalo de confiança contempla o valor 0. Conforme visto na figura 11, com o passar do tempo, o resultado passa a interferir positivamente na predição da eleição atual.

Uma das variáveis explicativas que o modelo A utiliza é intenção de votos em pesquisas realizadas entre 200 e 230 dias antes da eleição. O coeficiente para essa variável indica que com o passar das eleições a importância dele é menos significativa, o que pode ser explicado através da adição de pesquisas eleitorais que o modelo B introduz na previsão dos votos compartilhados.

A variável de Bernoulli que indica se um partido está ou não no poder é uma das duas variáveis que o coeficiente apresentou resultados positivos e negativos ao decorrer do tempo. É interessante notar que esse sinal se altera de acordo com o período político no qual o país encontra-se.

A introdução do modelo B faz com que a previsão seja atualizada sequencialmente. Para ilustrar como funciona o processo de atualização no passeio aleatório reverso, apresentamos os resultados obtidos para o grupo direita durante a campanha para as eleições federais de 2018. A Figura 12 apresenta a previsão dos votos compartilhados para o grupo direita ao longo do tempo, iniciando em 148 e finalizando em 2 dias antes da eleição.

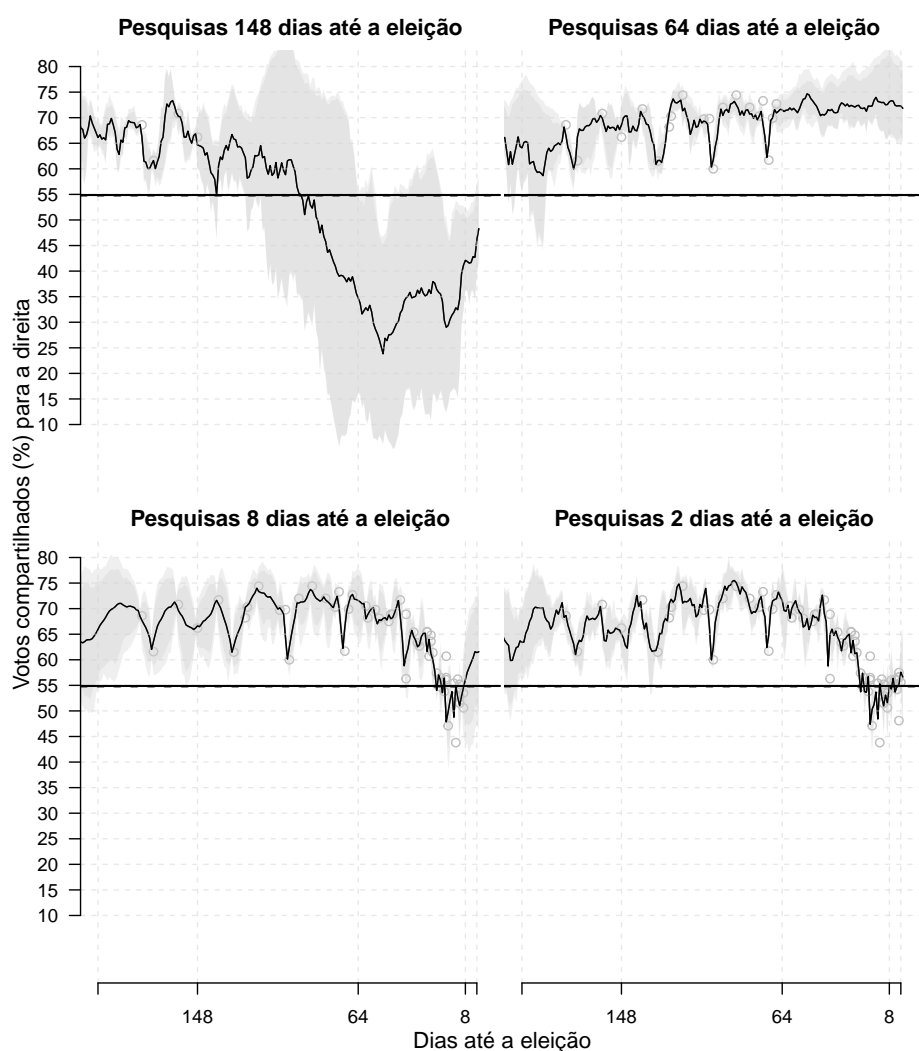


Figura 12: Predição dos Votos Compartilhados para a direita durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.

A linha sólida representa a média latente do apoio a direita da distribuição a posteriori;

a área sombreada representa os intervalos de credibilidade de 83% e 95%. Os votos compartilhados estimados em 2018 são indicados pela linha horizontal sólida (54,88%), e a previsão do modelo A é marcada pela linha horizontal tracejada (54,80%).

A partir da Figura 12 existem duas características do modelo que podem ser notadas. A primeira é que a incerteza ao longo do tempo vai sendo diluída. Observa-se que no início faltando 148 dias para eleição a incerteza sobre a previsão é maior, pois os intervalos de credibilidade tem uma amplitude muito maior se comparado ao modelo faltando 2 dias para a eleição. O segundo é a influência que as pesquisas eleitorais introduzem no modelo proposto, fazendo com que as previsões baseadas do modelo fundamental teórico agregue menos ao final da campanha eleitoral.

As Figuras 13 e 14 apresentam o mesmo gráfico apresentado na Figura 12, porém desta vez, para os grupos esquerda e centro respectivamente.

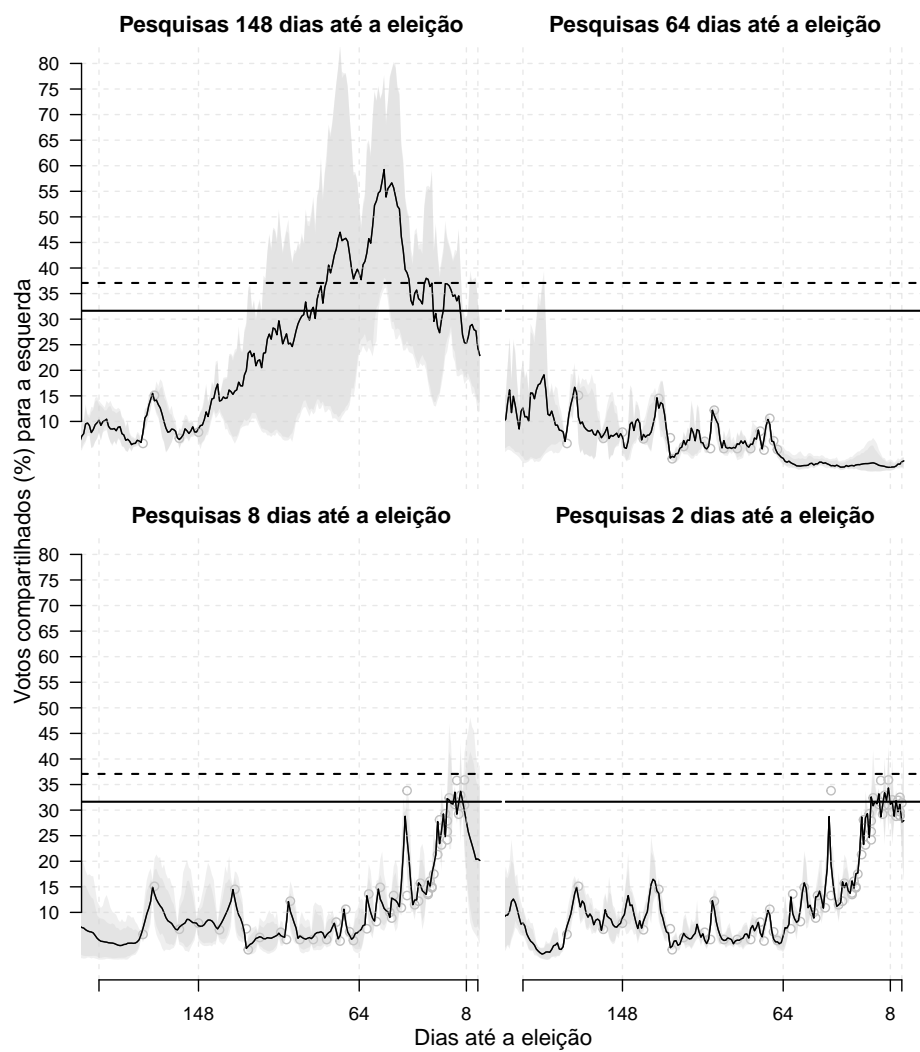


Figura 13: Predição dos Votos Compartilhados para a esquerda durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.

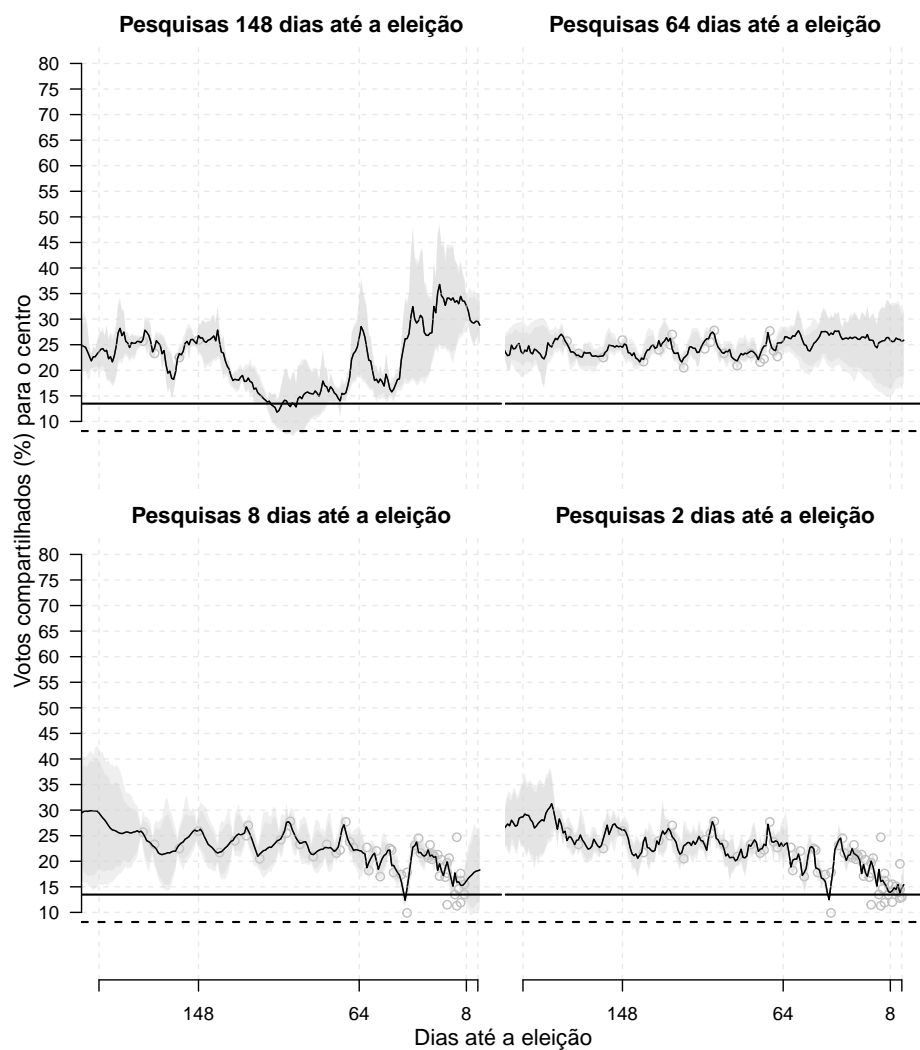


Figura 14: Predição dos Votos Compartilhados para o centro durante as eleições presidenciais de 2018 baseado no modelo Bayesiano dinâmico de previsão.

Como ocorreu para o grupo direita, as Figuras 13 e 14 evidenciam que com a aproximação do dia da votação e com a inclusão das pesquisas eleitorais, o modelo torna-se mais preciso e menos variável.

5 Conclusões

Esta dissertação teve como objetivo implementar e ajustar o modelo proposto por Stoetzer, Marcel, Neunhoeffers e Sternberg (2018), usando dados das eleições presidenciais brasileiras de 2018. Foi necessário revisar conceitos preliminares para maior entendimento do modelo, como inferência Bayesiana, métodos de MCMC e modelagem dinâmica.

As variáveis explicativas utilizadas na modelagem foram: votos compartilhados da eleição passada para cada um dos grupos abordados, intenção de votos entre 200 e 230 dias antes da data da eleição, variável indicadora (o valor 1 representa candidato/partido no poder, e 0 caso contrário), PIB médio, taxa média de inflação e taxa média de desemprego. Dentre estas, mostraram-se significativas para o modelo as três primeiras variáveis acima citadas.

A variável sobre a intenção de votos em pesquisas realizadas entre 200 e 230 dias antes da eleição apresentou-se como um fator relevante na estimação do modelo. Porém, com o passar do tempo, mais especificamente das eleições, ela torna-se menos significativa, como ilustrado na Figura 11 (b). Essa característica é esperada, uma vez que o modelo proposto vai sendo atualizado com informações de pesquisas eleitorais constantemente.

As variáveis votos compartilhados de cada um dos grupos abordados na eleição anterior e a variável de Bernoulli que indica se um partido está ou não no poder mudam significativamente o sinal de sua contribuição na previsão. A variável votos compartilhados da eleição anterior tem se tornado com o passar das eleições mais significativa, enquanto a variável indicadora tem apresentado menos impacto com o passar das eleições.

As variáveis que representam a taxa média de desemprego, a taxa de inflação e PIB, não mostraram-se significativas para a estimação do compartilhamento de votos. É provável, entretanto, que exista uma relação entre essas variáveis e a porcentagem de votos no candidato/partido incumbente, que poderia ser explorada em um trabalho futuro.

Como resultado, obtivemos que para todos os grupos (direita, centro e esquerda)

o valor previsto para a porcentagem de votos se diferencia em menos de 3% do valor real. Além disso, os intervalos de credibilidade de 83% compreenderam o real valor do parâmetro em seu domínio.

Se compararmos o modelo completo, ou seja, aquele em que há introdução das pesquisas eleitorais de forma dinâmica Bayesiana no decorrer do processo eleitoral (modelo B) com a primeira parte deste modelo, ou seja, o modelo fundamental teórico (modelo A), observamos que para esse conjunto de dados o modelo completo apresentou previsões melhores, destacando-se que com o decorrer do tempo o modelo em duas etapas ganha em precisão preditiva.

Futuramente, pretende-se comparar o modelo ajustado nesta dissertação com outros modelos já propostos na literatura, como Lewis-Beck (2005), Norpoth e Gschwend (2010), Magalhães, Aguiar-Conraria e Lewis-Beck (2012).

Referências

- BIZZARRO, F.; COPPEDGE, M. Variedades da democracia no Brasil. *Opinião Pública, Campinas*, n. 23, 2017.
- BOLOGNESI, B.; RIBEIRO, E. A.; CODATO, A. Esquerda, centro ou direita? como classificar os partidos no Brasil. *Observatório das eleições*, 2020.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. [S.l.]: Duxbury, 2004.
- CLEVELAND, W. S. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, American Statistical Association, v. 35, p. 54, 1981. ISSN 0003-1305,1537-2731. Disponível em: <http://doi.org/10.2307/2683591>.
- COSTA, A. C. C. da. *Modelos Dinâmicos Hierárquicos Espaço-Temporais para Dados na Família Exponencial*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Instituto de Matemática, Pós-Graduação em Estatística, 2011.
- DRENICK, R. F. Statistical theory of reliability. *Technometrics*, American Statistical Association, v. 5, p. 415–415, 1963. ISSN 0040-1706,1537-2723. Disponível em: <http://doi.org/10.1080/00401706.1963.10490117>.
- DUANE S.; KENNEDY, A. D. P. B. J.; ROWETH, D. Hybrid monte carlo. *Physics Letter B*, n. 195(2), p. 216–222, 1987.
- EL-DASH, N. A. *Avaliação metodológica das pesquisas eleitorais brasileiras*. Tese (Doutorado) — Universidade de São Paulo - USP, 2010.
- GAMERMAN, D.; MIGON, H. S. Dynamic hierarchical models. *Journal of the Royal Society*, v. 3, n. 3, p. 629–642, 1993.
- GELMAN A.; CARLIN, J. B. S. H. S. R. D. Bayesian data analysis. *Taylor and Francis*, 2014.
- HARRISON, P. J.; STEVENS, C. F. Bayesian forecasting. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 38, n. 3, p. 205–228, 1976. Disponível em: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1976.tb01586.x>.
- ISRAEL, V. P.; PEREIRA, N. B. Estudo sobre a distribuição das taxas de encarceramento nos estados brasileiros e principais variáveis associadas: Influências socioeconômicas e ideológicas. *DILEMAS*, v. 3, n. 11, p. 385–411, 2018.
- LARSON, H. J. *Introduction to Probability Theory and Statistical Inference*. 3. ed. [S.l.]: John Wiley and Sons, 1982.

- LEWANDOWSKI, D.; KUROWICKA, D.; JOE, H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, v. 100, n. 9, p. 1989–2001, 2009. ISSN 0047-259X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0047259X09000876>.
- LEWIS-BECK, M. Election forecasting: Principles and practice. *BJPIR*, v. 7, p. 145–164, 2005.
- LINDLEY, D. V.; SMITH, A. F. M. Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 34, n. 1, p. 1–18, 1972. Disponível em: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972.tb00885.x>.
- MACIEL, A. d. O. A. A. P. B.; GIMENES Éder R. Partidos políticos e espectro ideológico: Parlamentares, especialistas, esquerda e direita no brasil. *Revista Eletrônica de Ciência Política*, n. 3, 2017.
- MAGALHÃES, P.; AGUIAR-CONRARIA, L.; LEWIS-BECK, M. Forecasting spanish elections. *International Journal of Forecasting*, v. 28, 10 2012.
- MIGON, D. G. H. S.; LOUZADA, F. *Statistical Inference (An Integrated Approach)*. 2. ed. [S.l.]: CRC Press, 2014.
- MOTA, M. B.; BRAICK, P. R. *História da caverna ao terceiro milênio*. [S.l.]: Moderna, 2016. v. 3.
- NEAL, R. *Bayesian Learning of Neural Networks*. Dissertação (Mestrado) — University of Toronto, 1995.
- NEAL, R. *MCMC using hamiltonian dynamics*. [S.l.]: Handbook of Markov Chain Monte Carlo, 2011.
- NORPOTH, H.; GSCHWEND, T. The chancellor model: Forecasting german elections. *International Journal of Forecasting*, v. 26, p. 42–53, 03 2010.
- P., J. H. M. W.; MIGON, H. S. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 1985.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <https://www.R-project.org/>.
- REIS, M. A. dos. Multipartidarismo brasileiro: o excesso de partidos políticos e suas implicações. *Jus*, 2015.
- ROBERT, G. C. C. *Monte Carlo Statistical Methods*. 2nd. ed. [S.l.]: Springer, 2004. (Springer Texts in Statistics). ISBN 0387212396,9780387212395.
- Stan Development Team. *RStan: the R interface to Stan*. 2020. R package version 2.21.2. Disponível em: <http://mc-stan.org/>.
- STOETZER MARCEL NEUNHOEFFER, T. G. S. M. L. F.; STERNBERG, S. Forecasting election in multiparty systems: A bayesian approach combining polls and fundamentals. *Cambridge University Press on behalf of the Society for Political Methodology*, 2018.

TAROUCO, G. d. S.; MADEIRA, R. M. Os partidos brasileiros segundo seus estudiosos: análise de um expert survey. *Civitas - Revista de Ciências Sociais*, v. 15, n. 1, p. e24–e39, abr. 2015. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/civitas/article/view/18077>.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <https://ggplot2.tidyverse.org>.

XAVIER, C. M. *Métodos de Monte Carlo Hamiltoniano aplicados em modelos GARCH*. Dissertação (Mestrado) — Universidade de São Paulo, 2019.

APÊNDICE 1 – Distribuições

Como descrito em Migon e Louzada (2014) e Berger e Casella (2004), nesta seção é apresentado um breve resumo da distribuição multinomial e distribuição de Dirichlet que são utilizadas como as distribuições das variáveis resposta multivariadas do modelo proposto nesta dissertação.

1.1 Distribuição Multinomial

Considere n repetições independentes de um experimento aleatório, sendo que em cada repetição há k possíveis resultados com probabilidades p_1, p_2, \dots, p_k de ocorrer. Seja X_i o número de ocorrências do i -ésimo resultado, então o vetor $\mathbf{X} = (X_1, X_2, \dots, X_k)$ tem distribuição multinomial com parâmetros $n, k, p_1, p_2, \dots, p_k$ de ocorrer, cuja distribuição de probabilidades é dada por

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k \mid n, p_1, \dots, p_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} \quad (1.1)$$

em que, x_1, \dots, x_k são números inteiros não negativos satisfazendo $\sum_{i=1}^k x_i = n$ e $\sum_{i=1}^k p_i = 1$.

O valor esperado e a variância de $X_i, 1 = 1, \dots, k$ são dados respectivamente por,

- $\mu_i = E(X_i) = np_i$;
- $\sigma_i^2 = Var(X_i) = np_i(1 - p_i)$.

1.2 Distribuição de Dirichlet

Uma variável aleatória $\mathbf{X} = (X_1, X_2, \dots, X_k)$, em que $X_j \in [0, 1], j = 1, \dots, k$, e $\sum_{j=1}^k X_j = 1$ tem distribuição de Dirichlet com parâmetros $\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_j > 0, j =$

$1, \dots, k$, se sua função de densidade é dada por,

$$f(X_1, X_2, \dots, X_k \mid \alpha_1, \dots, \alpha_k) = \frac{\Gamma(A)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_k)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} \dots x_k^{\alpha_k-1}. \quad (1.2)$$

O valor esperado e variância são dados respectivamente por,

- $\mu_i = E(X_i) = \frac{\alpha_i}{A}$;
- $\sigma_i^2 = Var(X_i) = \frac{\alpha_i(A-\alpha_i)}{A^2(A+1)} = \frac{E(X_i)(1-E(X_i))}{(A+1)}$.

Em que, $A = \sum_{i=1}^k \alpha_i$.

APÊNDICE 2 – Divisão Ideológica: Eleições 1989 - 2018

Como visto ao longo da dissertação o modelo proposto é composto de duas partes, ou de forma mais informal, dois "sub modelos". O primeiro sub modelo utiliza informações de eleições passadas e como foi sugerido apenas utilizar a questão ideológica e não partidária: direita, esquerda e centro, foi necessário fazer uma adaptação dos resultados das eleições passadas para que enquadrassem se no cenário proposto.

É mostrado neste apêndice como foi feita essa divisão. Lembrando que para esta parte foram utilizados desde 1989 e também foram apenas considerados os votos válidos, excluindo brancos, nulos, justificativas e ausentes.

Tabela 5: Classificação das principais coligações/partidos da eleição de 1989

Coligação (Partidos)	Votos (%)	Classificação
Movimento Brasil Novo (PRN, PSC, PST, PTR)	30,7	Direita
Frente Brasil Popular (PT, PSB, PCdoB)	17,18	Esquerda
Partido Democrático Trabalhista	16,51	Centro
Partido da Social Democracia Brasileira	11,51	Direita
Partido Democrático Social	8,85	Direita
Aliança Liberal Cristã (PL, PDC)	4,83	Centro
Partido do Movimento Democrático Brasileiro	4,73	Centro
Partido Comunista Brasileiro	1,13	Esquerda
Partido da Frente Liberal	0,88	Centro
União Cidade Campo (PSD, PDN)	0,72	Centro
Partido Trabalhista Brasileiro	0,56	Direita
Partido de Reedificação da Ordem Nacional	0,53	Direita
Partido Social Progressista	0,35	Direita

Tabela 6: Classificação das principais coligações/partidos da eleição de 1994

Coligação (Partidos)	Votos (%)	Classificação
União, Trabalho e Progresso (PSDB, PFL, PTB)	54,24	Direita
Frente Brasil Popular pela Cidadania (PT, PSB, PPS, PV, PCdoB, PCB, PSTU)	27,07	Esquerda
Partido de Reedificação da Ordem Nacional	7,38	Direita
Desenvolvimento do Brasil (PMDB, PSD)	4,38	Centro
Força do Povo (PDT, PMN)	3,19	Centro
Partido Progressista Reformador	2,75	Direito
Partido da Reconstrução Nacional	0,61	Centro
Partido Social Cristão	0,38	Direita

Tabela 7: Classificação das principais coligações/partidos da eleição de 1998

Coligação (Partidos)	Votos (%)	Classificação
União, Trabalho e Progresso (PSDB, PFL, PPB, PTB, PSD)	53,06	Direita
União do Povo Muda Brasil (PT, PDT, PSB, PCdoB, PCB)	31,71	Esquerda
Brasil Real e Justo (PPS, PL, PAN)	10,97	Centro
Partido de Reedificação da Ordem Nacional	2,14	Direita
Partido da Mobilização Nacional	0,37	Centro
Partido Verde	0,31	Centro
Partido Socialista dos Trabalhadores Unificado	0,30	Esquerda

Tabela 8: Classificação das principais coligações/partidos da eleição de 2002

Coligação (Partidos)	Votos (%)	Classificação
Lula Presidente (PT, PL, PMN, PCdoB, PCB)	46,44	Esquerda
Grande Aliança (PSDB, PMDB))	23,19	Direita
Brasil Esperança (PSB, PTC, PGT)	17,86	Centro
Frente Trabalhista (PPS, PDT, PTB)	11,97	Centro
Partido Socialista dos Trabalhadores Unificado	0,47	Esquerda
Partido da Causa Operária	0,04	Esquerda

Tabela 9: Classificação das principais coligações/partidos da eleição de 2006

Coligação (Partidos)	Votos (%)	Classificação
A Força do Povo (PT, PRB, PCdoB)	48,61	Esquerda
Por um Brasil Decente (PSDB, PFL)	41,64	Direita
Frente de Esquerda (PSOL, PCB, PSTU)	6,85	Esquerda
Partido Democrático Trabalhista	2,64	Centro

Tabela 10: Classificação das principais coligações/partidos da eleição de 2010

Coligação (Partidos)	Votos (%)	Classificação
Para o Brasil seguir mudando (PT, PMDB, PDT, PCdoB, PSB, PR, PRB, PSC, PTC e PTN)	46,91	Esquerda
O Brasil pode mais (PSDB, DEM, PPS, PMN, PTdoB e PTB)	32,61	Direita
Partido Verde	19,33	Centro
Partido Socialismo e Liberdade	0,87	Esquerda

Tabela 11: Classificação das principais coligações/partidos da eleição de 2014

Coligação (Partidos)	Votos (%)	Classificação
Com a Força do Povo (PT, PMDB, PSD, PP, PR, PDT, PRB, PROS, PCdoB)	41,54	Esquerda
Muda Brasil (PSDB, SD, PMN, PEN, PTN, PTC, DEM, PTdoB, PTB)	33,50	Direita
Unidos pelo Brasil (PSB, PPS, PSL, PHS, PPL, PRP)	21,32	Centro
Partido Socialismo e Liberdade	1,55	Esquerda
Partido Social Cristão	0,75	Direita
Partido Verde	0,61	Centro
Partido Renovador Trabalhista Brasileiro	0,43	Direita

As porcentagens de cada coligação ou partido foram calculadas a partir dos votos válidos e refeita a porcentagem de acordo com a soma dos principais partidos ou coligação.

APÊNDICE 3 – Códigos

Neste anexo estão os *scripts* que foram utilizados para a simulação, utilizando o RStudio (2020) e os pacotes RStan (2020) e o Ggplot2 (2016).

3.1 Modelo A

O Modelo A é o primeiro modelo a ser rodado, sendo que a sua previsão funciona como a priori para o modelo completo.

```
data {
  int<lower=0> LA;
  int<lower=0> L;
  int<lower=0> Nobs;
  int<lower=0> Nmis;
  int<lower=0> N;
  int<lower=0> election[N];
  int<lower=0> K;
  int<lower=0> nParties;
  int<lower = 1, upper = Nobs + Nmis> ii_obs[Nobs];
  int<lower = 1, upper = Nobs + Nmis> ii_mis[Nmis];
  vector[Nobs] y_obs;
  matrix[N,K] x;
  int s[LA];
}
parameters {
  simplex[Nmis] y_mis;
  vector[L] b0;
  matrix[L,K] b;
```

```

    real<lower=0> sigma_b0;
    real<lower=0> sigma_b[K];
    real drift_b0;
    real drift[K];
}
transformed parameters {
    vector[N] y;
    vector[N] a;

    real mu_b0[L];
    matrix[L,K] mu_b;

    y[ii_obs] = y_obs;
    y[ii_mis] = y_mis;

    for (i in 1:N)
        a[i] = exp(b0[election[i]] + b[election[i],] * x[i,]');

    mu_b0[1] = 0;
    for (k in 1:K)
        mu_b[1,k] = 0;

    for (j in 2:L)
        mu_b0[j] = b0[j-1] + drift_b0;

    for (k in 1:K)
        for(j in 2:L)
            mu_b[j,k] = b[j-1,k] + drift[k];
}
model {
    int pos;
    pos = 1;

    b0[1] ~ normal(0, 10); // initialization
    drift_b0 ~ normal(0, 10);

```

```

sigma_b0 ~ normal(0, 1);

for (j in 2:L)
  b0[j] ~ normal(mu_b0[j], sigma_b0);

for (k in 1:K) {
  b[1,k] ~ normal(0, 10);
  drift[k] ~ normal(0, 10);
  sigma_b[k] ~ normal(0, 1);
}

for (k in 1:K)
  for(j in 2:L)
    b[j,k] ~ normal(mu_b[j,k], sigma_b[k]);

for (l in 1:LA) {
  segment(y, pos, s[l]) ~ dirichlet(segment(a, pos, s[l]));
  pos = pos + s[l];
}
}

```

3.2 Modelo completo

Uma vez que já rodamos o modelo A, rodamos a seguir o modelo completo, cujo código segue abaixo:

```

data {
  // Data for dynamic component of the model
  int<lower=0> nParties; // Number of Parties
  int<lower=0> nInst; // Number of Polling Institutes
  int<lower=0> nPolls; // Number of published Polls
  int<lower=0> nPeriods; // Number of days the model should account for
  int<lower=0, upper = nInst> iid[nPolls]; // Numeric IDs of the Polling Institutes
  int<lower=0, upper = nPeriods> date[nPolls]; // Date (as days from 1 to nPeriods)
  int<lower=0> y[nPolls,nParties]; // The poll results

```



```

// Data for sturctural part of the model
int<lower=0> LA;
int<lower=0> L;
int<lower=0> Nobs; // Number of observed party-election results
int<lower=0> Nmis; // Number of "missing" party-election results (No. of parties
int<lower=0> N; // Total Number of Observations
int<lower=0> election[N]; // Election ID
int<lower=0> K; // Number of covariates in structural model
vector[Nobs] v_obs; // observed vote shares
matrix[N,K] x; // Matrix of Covariates
matrix[Nmis,K] xE; // Matrix of Covariates for upcoming election

// Auxilliary data
int<lower = 1, upper = Nobs + Nmis> ii_obs[Nobs]; // index for observed party-ele
int<lower = 1, upper = Nobs + Nmis> ii_mis[Nmis]; // index for missing party-elec
int s[LA]; // Number of parties per election
matrix[L,K] b_prior; // beta values from pre-trained structural model
vector[L] b0_prior; // beta values for b0 from pre-trained structural model
}
parameters {
  simplex[Nmis] v_mis;
  simplex[nParties] vE;
  vector[L] b0;
  matrix[L,K] b;
  //real<lower=0> tau_b0;
  //real<lower=0> tau_b[K];
  real drift_b0;
  real drift[K];
  cholesky_factor_corr[nParties-1] S_cor;
  cholesky_factor_corr[nParties-1] S_shock_cor;
  vector<lower=0>[nParties-1] sigma_evo;
  vector<lower=0>[nParties-1] sigma_shock;
  vector[nParties-1] alphastarforecast;
  matrix[nPeriods-1,nParties-1] alphastar;
  matrix[nInst-1,nParties-1] house_effect_raw;

```

```

    real<lower=0> sigma_b0;
    real<lower=0> sigma_b[K];
}
transformed parameters {
    vector[N] v;
    vector[N] a;
    vector[nParties] a_pred;
    real mu_b0[L];
    matrix[L,K] mu_b;
    matrix[nPeriods,nParties] alphastar_prior;
    vector[nParties] alphastarforecast_prior;
    matrix[nInst,nParties] house_effect;
    matrix[nPeriods,nParties] alpha;
    matrix[nPeriods,nParties] ea;

    vector[nParties] ef;
    vector[nParties] forecast;

    alphastar_prior[1:(nPeriods-1),1:(nParties-1)] = alphastar;
    alphastarforecast_prior[1:nParties-1] = alphastarforecast;

    house_effect[2:nInst, 2:nParties] = house_effect_raw;

    for (j in 2:nParties)
        house_effect[1, j] = 0 - sum(house_effect[2:nInst, j]);

    for(c in 1:nInst)
        house_effect[c, 1] = 0 - sum(house_effect[c, 2:nParties] );

    v[ii_obs] = v_obs;
    v[ii_mis] = v_mis;

```

```

for (i in 1:N)
  a[i] = exp(b0[election[i]] + b[election[i],] * x[i,]');

for (j in 1:nParties)
  a_pred[j] = exp(b0[L] + b[L,] * xE[j,]');

mu_b0[1] = 0;
for (k in 1:K)
  mu_b[1,k] = 0;

for (j in 2:L)
  mu_b0[j] = b0_prior[j-1] + drift_b0;

for (k in 1:K)
  for(j in 2:L)
    mu_b[j,k] = b_prior[j-1,k] + drift[k];

// Transform structural forecast to log-ratio
for (j in 1:(nParties))
  alphastar_prior[nPeriods,j] = log(vE[j]/vE[nParties]);

for (i in 1:(nPeriods-1))
  alphastar_prior[i,nParties] = 0;

for(k in 1:nPolls)
  alphastar_prior[date[k],] = (alphastar_prior[date[k],] + house_effect[iid[k],])

alphastarforecast_prior[nParties] = 0;

// Transform values from log-ratio space back to vote shares
for (i in 1:(nPeriods))
  for (j in 1:(nParties))
    ea[i,j] = exp(alphastar_prior[i,j]);

```

```

for (i in 1:(nPeriods))
  for (j in 1:(nParties))
    alpha[i,j] = ea[i,j]/sum(ea[i,]);

for (j in 1:(nParties))
  ef[j] = exp(alphastarforecast_prior[j]);

for (j in 1:(nParties))
  forecast[j] = ef[j]/sum(ef[1:nParties]);

}

model {

  int pos; // Auxilliary counter
  pos = 1; // Pre-set counter to 1

  // Structural component of the model

  b0[1] ~ normal(b0_prior[1], 10); // Initialization of b0 in the structural compon
  drift_b0 ~ normal(0, 10);
  sigma_b0 ~ normal(0, 1);

  for (j in 2:L)
    b0[j] ~ normal(mu_b0[j], sigma_b0);

  for (k in 1:K) {
    b[1,k] ~ normal(b_prior[1,k], 10);
    drift[k] ~ normal(0, 10);
    sigma_b[k] ~ normal(0, 1);
  }

  for (k in 1:K)

```

```

    for(j in 2:L)
        b[j,k] ~ normal(mu_b[j,k], sigma_b[k]);

for (l in 1:LA) {
    segment(v, pos, s[l]) ~ dirichlet(segment(a, pos, s[l]));
    pos = pos + s[l];
}

vE ~ dirichlet(a_pred);

// Dynamic component of the model

// Evolution variance prior
sigma_evo ~ normal(0, 0.1);

// Forecast variance shock prior
sigma_shock ~ normal(0, 0.18);

// Evolution covariance prior
S_cor ~ lkj_corr_cholesky(50);

// Forecast covariance prior
S_shock_cor ~ lkj_corr_cholesky(100);

// Backward Random Walk of latent support (in log-ratio space)
for (i in 1:(nPeriods-1)){
    alphastar[i,1:(nParties-1)] ~ multi_normal_cholesky(alphastar_prior[(i+1),1:(nParties-1)])
}

// Add forecast variance shock
alphastarforecast[1:(nParties-1)] ~ multi_normal_cholesky(alphastar_prior[nPeriods,1:(nParties-1)])

// Model Polls based on latent state (in vote share space)
for (k in 1:nPolls)
    y[k,1:nParties] ~ multinomial(alpha[date[k],,]);

```

```
// Estimate house effects
for (j in 1:(nParties-1))
  for (c in 1:(nInst-1))
    house_effect_raw[c, j] ~ normal(0, 0.001); // Prior 1 percent point sd
}
```