UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFA, LETRAS E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA POLÍTICA

NATÁLIA DE PAULA MOREIRA

Três Artigos sobre Eleições e Voto Econômico

(Three Essays on Elections and Economic Voting)

São Paulo

2021

NATÁLIA DE PAULA MOREIRA

Três Artigos sobre Eleições e Voto Econômico

(Three Essays on Elections and Economic Voting)

Versão original

Tese apresentada à Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de Doutor em Ciências pelo Programa de Pós-graduação em Ciência Política.

Área de concentração: Política Comparada, Políticas Públicas

Supervisor: Profa. Dra. Lorena Guadalupe Barberia

Co-supervisor: Prof. Dr. Guy D. Whitten

São Paulo

2021

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

Moreira, Natália de Paula
Mt Três Artigos sobre Eleições e Voto Econômico / Natália de Paula Moreira; orientadora Lorena Guadalupe Barberia - São Paulo, 2021.
155 f.

Tese (Doutorado)- Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Ciência Política. Área de concentração: Ciência Política.

1. Gênero. 2. Aprovação Presidencial. 3. Modelos com Variáveis Dependentes Policotômicas. 4. Efeitos Marginais. 5. Vantagem da Incumbência. I. Barberia, Lorena Guadalupe, orient. II. Título.

MOREIRA, N. P. "**Três Artigos sobre Eleições e Voto Econômico**". 2021. Tese (Doutorado em Ciência Política) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2021.

____

Prof. Dr. ____

Instituição: ____

Presidente

____

Prof. Dr. ____

Instituição: ____

____

Prof. Dr. ____

Instituição: ____

____

Prof. Dr. ____

Instituição: ____

A minha família

Acknowledgements

There are many institutions and people that I would like to thank for supporting me during my dissertation journey.

I would like to thank the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), who has funded my dissertation research including a one-year additional scholarship to train in the Department of Political Science at Texas A &M University. Specifically, I would like to acknowledge these two grants from Fapesp (Grant No. 2016/13199-8 and 2017/11559-0).[1] I am also grateful for the research support (Grant No. 2012/03900-0) that I received from FAPESP that allowed me to complete my training while an undergraduate with a scientific initiation scholarship.

I have been extremely fortunate to have the best advisor and co-advisor to guide me throughout this journey. Lorena Barberia has been an incredible advisor and co-author. When I took her class on methods and research design as an undergraduate student, a new world opened to me, and I fell in love with research and quantitative methods. It is inspiring to see her passion for knowledge. I also would like to express my appreciation to Guy Whitten, who has been a terrific co-advisor and who received me so warmly at College Station during my 3rd year of my PhD for a one-year residence at TAMU. From Lorena and Guy, I learned that the challenges of conducting rigorous scientific research pay off when you do what you love.

Since my Master's Degree, I have presented my research at MPSA, APSA, EPCR, ABCP and LAPOLMETH. The comments I received on my papers and projects during these meetings were extremely valuable and encouraged me to consider a career in academia. I am especially grateful for the comments and critiques that received from Rocío Titiunik, Soledad Prillaman, and Guilherme Russo. I thank the Dynamic Pie Research Group for very helpful comments on the preliminary versions of this dissertation's chapters. I am also grateful to Andrew Philips and Laron Williams for always being available to answer my questions and providing me with invaluable suggestions for my chapters. I am also very thankful for their continuous encouragement along this process. I am thankful to George Avelino and numerous colleagues at CEPESP at the Fundação Getulio Vargas (FGV) who have been very supportive and receptive to my

[1] As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade da autora e não necessariamente refletem a visão da FAPESP.

research. I am also very thankful to the professors in the Department of Political Science at the University of São Paulo for their instruction and dedication. I especially thank Professor Fernando Limongi, Jonathan Philips, Glauco Peres, Marta Arretche, Rogerio Arrantes, Eduardo Marques, and Paolo Ricci. I also want to thank the department for enabling me to participate in the IPSA-USP Summer School. I actively participated in the Summer School each year since my Master's Degree and the international faculty and stellar training helped to broaden my methods training. In serving as a teaching fellow in the IPSA-USP Summer School and in receiving an award for my poster in political methodology, the School provided me with important recognition of my contribution's to academia that provided encouragement at crucial moments in my journey. The School was also fundamental for learning new methodological tools that I employed in my Master's thesis and in my Doctorate's dissertation. Among the Summer School faculty, I want to thank Allyson Benton, Jason Seawright, Clifford Young, and Bruno Cautres.

I am also grateful to Maria Letícia Claro, Isabel Seelander, Luiz Cantarelli, Marcela Zamudio, Ingrid Loureiro, Pedro Schmalz, Gustavo de Paula, and Dara Vilela, who have been supportive friends and amazing colleagues. I especially thank the família Barberia Masiero (Angelo, Ale, Carla, Gilmar, and Lorena) for their tremendous generosity. I would like to thank Gabriela Trindade and Renata Cortez for always being available to discuss ideas and encouraging me. I would like to thank my friends from NECI (USP) for been available to discuss ideas and offer help. I also thank my colleagues and friends from the Department of Political Science: Sergio Simoni, Marina Merlo, Lilian Sendretti, Grazielle Silotto, Joyce Luz, Hellen Guicheney, Barbara Lopes, Thiago Meirelles, Thiago Moreira, Rodrigo Nakahara, Rafael Magalhães, Murilo Junqueira, Gabriel Zanlorenssi, Danilo Praxedes, Gabriel Madeira, and many others that helped me throughout this journey. I also wish to thank Nathalie Mendez, Yoo Sun, Andrea Junqueira, Ali Kagalwala, Thiago Silva, Flavio Souza, Janica Magat, Kostanca Dhima, Samantha Zuhlke, and many other friends that warmly welcomed me in College Station. I wish to thank Clarice Pimentel for helping me to overcome the difficulties encountered during this journey.

Last, I would like to thank Nam Pho, who has supported me throughout these years. I am immensely grateful to my family (Hilda, Luiz, Anna, and Matheus) for their unconditional support and love. I am especially thankful to my parents for making every possible effort to give me the chance to study and to choose my career. I am very grateful to my sister and my mother, who always encourage me to seek new challenges and believe in myself. Thank you!

"Recria tua vida, sempre, sempre. Remove pedras e planta roseiras e faz doces. Recomeça."

(Cora Coralina)

Resumo

MOREIRA, Natália de Paula. **Três Artigos sobre Eleições e Voto Econômico**. 2021. 155 f. Tese (Doutorado em Ciências) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2021.

Esta tese é composta por três capítulos independentes sobre eleições e voto econômico em política comparada. O primeiro capítulo oferece uma resposta à questão sobre se o desalinhamento partidário leva a mudanças na vantagem da incumbência para a Câmara dos Deputados dos EUA ao longo do tempo. O segundo capítulo emprega um projeto de pesquisa comparativa para avaliar a lacuna de gênero nos índices de aprovação presidencial durante a pandemia de COVID-19 no Brasil e nos EUA. O último capítulo foca em como desenhos de pesquisa que se utilizam de variáveis dependentes policotômicas podem ser melhorados com análises pós-estimação mais robustas. Em resumo, os resultados dos capítulos desenvolvidos nesta tese oferecem contribuições substantivas e metodológicas para linhas de investigação distintas.

Palavras-chaves: Gênero; Aprovação Presidencial; Modelos com Variáveis Dependentes Policotômicas; Efeitos Marginais, Vantagem da Incumbência

Abstract

MOREIRA, Natália de Paula. **Three Essays on Elections and Economic Voting**. 2021. 155 p. Dissertation (Doctor of Science) – Faculty of Philosophy, Languages and Human Sciences, University of São Paulo, São Paulo, 2021.

This dissertation comprises three stand-alone chapters on elections and economic voting in comparative politics. The first chapter offers an answer to the question of whether partisan dealignment drive changes in the incumbency advantage in U.S. House elections over time. The second chapter employs a comparative research design to evaluate the gender gap on presidential approval ratings during the COVID-19 pandemic in Brazil and the U.S. A final chapter focuses on how research designs that use polychotomous dependent variables can be improved with more robust post-estimation analyses. In sum, the findings of the research produced in this dissertation provide substantive and methodological contributions to distinct lines of inquiry.

Keywords: Gender; Presidential Approval; Polychotomous Dependent Variable Models; Marginal Effects; Incumbency Advantage.

List of Figures

List of Tables

Contents

1 INTRODUCTION

This dissertation is composed of three stand-alone chapters that address research questions related to elections and economic voting in comparative politics. Two chapters employ a research design that employs dynamic time series analyses to improve hypothesis testing and inferences. A final chapter focuses on how research designs that focus on polychotomous dependent variables can be improved with more robust post-estimation analyses. Each chapter provides substantive and methodological contributions to distinct lines of inquiry. In the following paragraphs, I briefly summarize each chapter. In explaining the contributions of each chapter, I show how the research design I employed makes an original contribution.

In Chapter 2, I offer an answer to the question of whether partisan dealignment drive changes in the incumbency advantage over time. At the beginning of the 1970s, scholars noticed that incumbents enjoyed a substantial advantage when seeking reelection over their challengers in U.S. House of Representatives (Erikson 1971; Mayhew 1974*a*). Later studies then noticed that after the steady growth of incumbency advantage from the 1940s to the mid-1980s, an irregular decline began in the late 1980s (Jacobson 2015). Different explanations were raised to explain the changes in incumbency advantage over time (Krehbiel and Wright 1983; Jacobson and Carson 2019). However, these studies had a significant limitation. Their analysis was limited to examining what might trigger changes in this phenomenon over time without using the incumbency advantage as a dependent variable in a statistical model. Among the factors that were identified as possibly contributing to the incumbency advantage was speculation that partisan dealignment among voters could be central to understanding why the incumbency advantage was experiencing an increase (Erikson 1972; Burnham 1974). According to this view, the boost in the incumbency advantage was a consequence of the disengagement of the electorate with political parties. Voters would be using party labels to an even lesser extent as a reference for their own voting decisions. Instead, voters would be turning to incumbency as a cue to guide their voting choices.

In contrast to earlier research that has employed cross-sectional research designs to verify whether partisan dealignment contributes to the observed dynamics in the incumbency advantage, I propose that a time-series analyses should be employed to fully test the implications of the incumbency advantage across elections. To do so, I introduce a new strategy for studying incumbency advantage over time. In the first stage, I use a series of cross-sectional regression models to generate a continuous measure of incumbency effects for 34 elections. In a second stage, I then employ the estimated incumbency effect across districts in each election as the dependent variable to examine the impact of the partisan dealignment theory on incumbency advantage from 1948 to 2014.

I test four hypotheses that are crucial for identifying if partisan dealignment theory is indeed driving the incumbency advantage observed in U.S. House elections. The U.S. is an especially relevant case to study for at least two reasons. First, the majority of studies on incumbency effects have focused on analyzing U.S. House elections. However, the main focus of these studies has been on how to measure this phenomenon correctly and use this measurement to test hypotheses with incumbency as an explanatory variable (Gelman and King 1990; Cox and Katz 1996; Lee 2008; Gelman and Huang 2008; Caughey and Sekhon 2011; Erikson and Titiunik 2015). Much less emphasis has been directed employing the incumbency effect as the dependent variable and testing which factors have contributed to its behavior across time. Second, the U.S. case provides a relevant example to study the variations in the incumbency advantage across time since there have been marked shifts in the incumbency effects across elections.

Substantively, Chapter 2 addresses a research question that has been explored over the last fifty years. More importantly, it innovates in proposing a novel research design to study the incumbency advantage over time. I build on the research strategy used by Duch and Stevenson (2006) in the study of economic voting in which cross-sectional analyses were estimated separately and the estimated effects were analyzed to discern patterns. Unlike this study, however, I combine the cross-sectional analysis to generate a time series measure which I then analyze with time series econometrics. Another relevant contribution is regarding the measurement of the dependent variable. To test the partisan dealignment theory, I employ two different measures of the incumbency advantage. The first measure is based on observational data (Gelman and King 1990), whereas the second measure is derived from a quasi-experimental research design (Erikson and Titiunik 2015). By employing different measures of the incumbency advantage derived from different research designs, Chapter 2 undertakes a comprehensive analysis of whether the determinants of the incumbency advantage are robust to different measures.

While Chapter 2 analyzes the incumbency advantage in elections, Chapter 3 examines accountability. Particularly, I evaluate the gender gap on presidential approval ratings during the COVID-19 pandemic. Scholars have long been interested in the sources and the impacts of the gender gap on political behavior and voting patterns (Shapiro and Mahajan 1986; Kaufmann and Petrocik 1999; Schlesinger and Heldman 2001; Box-Steffensmeier, De Boef and Lin 2004). In this chapter, I contribute to this literature by assessing whether there are differences between women and men in evaluating presidential job performance in the context of the COVID-19 pandemic. Evidence from previous research shows that women are more likely to assess the COVID-19 pandemic as a severe health crisis than their male counterparts (Galasso et al. 2020), and they are more likely to report wearing masks when leaving home than men (Palmer and Peterson 2020; Cassino and Besen-Cassino 2020). Studies also show that women are more likely to base their attitudes and activities on scientific information (Algara, Fuller and Hare 2020; Algara et al. 2021). Given the substantial gender differences in behavior and attitudes regarding the COVID-19 pandemic, I expect that women will punish the incumbent to a greater degree than their male counterparts as the death toll increases. I evaluate these theoretical expectations using data on presidential approval ratings by gender for Brazil and the United States over time.

Instead of focusing on one case study as most of the previous literature on the gender gap and voting choices (Welch and Hibbing 1992; Chaney, Alvarez and Nagler 1998; Kam 2009) and presidential approval ratings (Clarke et al. 2005; Higgins and Kellstedt 2016) has done, Chapter 3 takes advantage of a comparative research design to evaluate whether statistical findings are robust to different cases during the same pandemic. Since the United States and Brazil are the countries hit hardest by the pandemic so far, they provide relevant cases to study the impact of the pandemic on women's and men's approval ratings. One should expect voters to hold the incumbents accountable given the catastrophic outcome of the pandemic in those places. Another innovation in the research design undertaken by Chapter 3 is in combining evidence from observational studies (i.e., a time series analysis for each country) and a quasi-experimental research design (i.e., a difference-in-difference research design using data for the U.S. case).

While Chapter 2 and 3 focus on research questions related to elections and economic voting, Chapter 4 evaluates the strategies for interpreting the results in nonlinear models, particularly, in polychotomous dependent variable models. Polychotomous dependent variable models have been used to model phenomena in which the dependent variable has more than two categories. These models have been especially relevant in political science for modeling voting behavior in multiparty systems. To mention a few of the most commonly employed models by the literature: multinomial logit (Whitten and Palmer 1996), conditional logit, multinomial probit (Alvarez and Nagler 1998), and mixed logit (Glasgow 2001).[1]

Scholars have dedicated significant efforts to develop more meaningful ways for interpreting the results of nonlinear models (King 1989; Long 1997; King, Tomz and Wittenberg 2000). In these models, interpretation of the regression results is complicated because the explanatory variables are nonlinear functions of the dependent variables. Compared to binary regression models, interpreting the regression results of polychotomous dependent variable models is an even more challenging task. Since in these models the dependent variable has more than two categories, the models produce a high number of coefficients for each dependent variable, and inferences can thus be more complex to discern. Given the nonlinearity of these models, there is no single interpretation approach that completely describes the relationship between a variable and the outcome probability in a multivariate case (Long and Freese 2014). In this chapter, I contribute to this literature by emphasizing the relevance of using a combination of different approaches when evaluating the results from polychotomous dependent variable models. Instead of relying on a single approach for all situations, I argue that the use of different approaches provides a more comprehensive way to interpret the results on several occasions. To show the limitations and advantages of different interpretation approaches, I replicate a part of the analysis undertaken by an article that estimated a multinomial logit model. To evaluate what are the most

[1] Polychotomous dependent variable models have been employed to address research questions about *individual* voting behavior in a static framework. However, there have also been efforts to develop statistical models to examine *aggregate movements* on voting behavior in multiparty elections in a cross-sectional (Katz and King 1999*a*; Tomz, Tucker and Wittenberg 2002; Jackson 2002; Mikhailov, Niemi and Weimer 2002) and in a dynamic framework (Philips, Rutherford and Whitten 2016).

commonly used interpretation strategies in the literature, I surveyed the empirical literature that employed polychotomous dependent variable models published in three major journals in political science from 2006 to 2016. Then, I perform a series of Monte Carlo simulations to analyze whether one of the two most commonly used approaches (marginal effect at means and average marginal effect) produce distinct quantities of interest in the context of polychotomous dependent variable models and which one is more susceptible to bias due to model misspecification.

Chapter 4 proposes important contributions to the discipline. This is the first study that undertakes a comprehensive survey study of the most common interpretation approaches used by the political science literature. Such a review provides a summary of the common patterns across the studies published in the major journals of the area, and it provides a diagnostics of the most relevant issues. The chapter also calls attention to the limits in focusing on only one approach to interpret the results of a polychotomous dependent variable model. By doing so, it contributes by advocating for the use of a combination of approaches to more comprehensively examine the impact of a key explanatory variable on the probability outcome. Methodologically, the chapter provides evidence, in the context of polychotomous dependent variable models, from Monte Carlo simulations about the differences between marginal effect and average marginal effect (AME) and which approach tends to be more susceptible to bias due to model misspecification. Previous literature on this issue had focused on the case of binary probit (Hanmer and Kalkan 2013; Bartus 2005).

Chapter 5 concludes with a discussion of the main findings from each chapter and possible extensions to the research undertaken in this dissertation.

2 UNCOVERING THE DYNAMICS OF INCUMBENCY ADVANTAGE

2.1 Introduction

In the 1970s, the first studies were published showing that incumbents enjoyed a substantial advantage when seeking reelection over challengers in U.S. House elections (Erikson 1971; Mayhew 1974*a*). As Gelman and King (1990) highlighted, such a phenomenon was not recent. Incumbents had enjoyed an electoral advantage since the beginning of the last century, but it grew substantially from the 1950s to the 1980s. After a half-century of research, there is no doubt that incumbency is a relevant cue employed by voters when deciding who to vote for. Despite the recent decline (Jacobson 2015), incumbents – on average – still enjoy a considerable electoral advantage in U.S. House elections.[1] Scholars also agree that the average incumbency advantage has varied significantly over time (Gelman and King 1990; Carson, Sievert and Williamson 2019; LeVeck and Nail 2016). Since 1900, at least three different periods can be identified. In the first period from 1900 to the 1950s, the incumbency advantage was small and constant across elections. In a second period from the 1950s to 1980s, the incumbency advantage increased substantially. And, finally, in a third period from the 1990s to the present, the advantage is decreasing. Although a lot of research has focused on how to measure this phenomenon correctly (Gelman and King 1990; Cox and Katz 1996; Lee 2008; Gelman and Huang 2008; Caughey and Sekhon 2011; Erikson and Titiunik 2015), much less emphasis has been directed to understanding the causes of the incumbency advantage.

One of the main explanations explored by the literature is related to partisan dealignment (Krehbiel and Wright 1983; Jacobson and Carson 2019; Stonecash 2008).[2] Scholars claim that movements in the average incumbency advantage are a result of shifts in the behavior of the electorate (Cover 1977; Ferejohn 1977). According to this view, the increase in partisan dealignment leads to the rise in the incumbency advantage as observed in the 1960s. As Ansolabehere and Snyder Jr (2002) highlight, "psycho-

[1] Although in U.S. House elections incumbents – on average – enjoy an election advantage, scholars have found evidence of negative incumbency effects for some countries (Ariga 2015; Klašnja and Titiunik 2017; Uppal 2009).

[2] An alternative explanation that I will not explore in this chapter argues that changes in the incumbency advantage are a result of changes in the activities of incumbents (e.g., constituency service) (Mayhew 1974*b*; Fiorina 1977; Krehbiel and Wright 1983). As incumbents have access to public resources, they are able to use them to increase their chances of getting reelected.

logically, party and incumbency are thought to be conflicting voting cues, and rising incumbency advantages in the House occur in an era of declining party (316)." Alternatively, as voters become more closely engaged with political parties, they might turn less to incumbency as a cue to inform their vote. In this sense, the recent increasing levels of party loyalty might have affected negatively the incumbency advantage (Jacobson 2015). Although these studies have taken great care in developing their theories about what explains the variation in incumbency advantage, their statistical analyses suffered from two important shortcomings. First, a great number of these studies did not rely on formal tests to evaluate what causes variation in the average incumbency advantage. Second, most of these studies focused on cross-sectional analysis to examine the determinants of variation in incumbency advantage. However, this type of analysis does not provide information about the phenomenon's temporal dynamics since a cross-sectional study offers solely a picture of a single point or period in time. Among those studies that have employed time-series analyses to investigate the causes of incumbency advantage (Ansolabehere and Snyder Jr 2002; Carson, Sievert and Williamson 2019; LeVeck and Nail 2016), none of them has estimated statistical models using incumbency advantage as the dependent variable.

To address these issues, I offer a novel strategy for examining the temporal dynamics of incumbency advantage. In undertaking this effort, I seek to provide robust statistical evidence about the impact of partisan dealignment on the average incumbency advantage. Building on the theoretical contributions of the earlier studies, I test the impact on the average incumbency advantage of four different measures that capture distinct aspects of the partisan dealignment theory. Particularly, I evaluate the effect on incumbency advantage of changes in the following explanatory variables: (i) percentage of independents in the electorate, (ii) percentage of party loyalty, (iii) percentage of challengers' partisan defecting to incumbents, and (iv) polarization. On one hand, I expect that the average incumbency advantage grows as the percentage of independents in the electorate or percentage of challengers' partisan defecting to incumbents increase. On the other hand, I expect the average incumbency advantage to decline when the percentage of party loyalty or polarization increases.

I test my theoretical expectations using data from the U.S. House elections from 1946 to 2014. Using cross-sectional analyses, I estimate the average incumbency advantage for each election. These estimates are then pooled to construct a measure of the incumbency advantage that varies over time. I then employ the estimated incumbency effect as the dependent variable in the second stage of the analysis to evaluate the impact of dealignment on the incumbency advantage. This research strategy is similar to what other studies have done using estimated dependent variables in the second stage of their analysis (Duch and Stevenson 2006; Williams 2013). The novelty of this study is to implement such strategies combining both cross-sectional and time series analysis. Consistent with my theoretical expectations, I find that an increase in polarization or percentage of party loyalty leads to a reduction in the average electoral advantage of the incumbency advantage. I also find evidence that an increase in the percentage of independents or in the percentage of challengers' partisan defecting to incumbents leads to an increase in the incumbency advantage. The findings presented in this chapter contribute to understanding how incumbency advantages are being shaped by party dealignment over the long-term.

The rest of this chapter proceeds as follows. In the next section, I briefly review the literature and discuss its main shortcomings. After reviewing these studies, I present the four hypotheses that I will test in this paper related to how dealignment affects incumbency. I then explain how I generate a time-series measure of incumbency advantage. After showing the procedure to create a time-series measure of incumbency advantage, I then proceed to test my hypotheses. The final section concludes with suggestions of future questions that can be further explored by the literature.

2.2 *Challenges with Measuring the Incumbency Advantage*

The U.S. political system is primarily a two-party system, and third parties and independent parties are minor players. Since the 20th century, the two major parties are the Democrats and the Republicans. The U.S. House of Representatives elections occur every two years (every midterm and presidential election year), and voters directly elect their members to serve a two-year term. More importantly, there are no term limits.

As Erikson (2016) underscores, there are three significant challenges in measuring the incumbency advantage. First, those candidates that win elections are usually of higher quality than those that lose, regardless of the gains obtained once they become incumbents. Or as Lee (2008) explains, "incumbents are, by definition, those politicians who were successful in the previous election. If what makes them successful is somewhat persistent over time, they should be *expected* to be somewhat more successful when running for re-election (683)." Second, incumbents may benefit from the poor quality of their challengers. Third, incumbents' retirement is not a random event. Incumbents may decide to retire when they expect to lose in the next election. In sum, these challenges imply additional difficulties in analyzing the existence of an incumbency advantage.

Since the 1970s, there have been three waves of studies that aim to quantify the incumbency advantage. In the first wave, studies focused on developing measures of the incumbency advantage based on a comparison of the aggregate average of two groups. This is the case of the sophomore surge, retirement slump, and 'slurge.' While the sophomore surge focuses on the electoral gain of incumbents running for reelection for the first time (Cover 1977), the retirement slump focuses on the electoral loss of the incumbent party when the incumbent retires. Both measures are estimated controlling for the partisan trend (Cover and Mayhew 1977). The 'slurge' is an average of these two measures (Alford and Brady 1988). Although the conception of these measures is intuitive to what may be driving incumbency, they suffer some crucial shortcomings. First, these measures are less efficient because they are based on a small fraction of the legislature races. Second, the size and direction of the incumbency effect are biased. Gelman and King (1990) show that, while the sophomore surge underestimated the incumbency advantage, the retirement slump overestimated it.

The second wave of studies focused on estimating the incumbency advantage based on linear regression models. This is the case of the seminal work of Gelman and King (1990).[3] To quantify the incumbency effect, Gelman and King (1990) estimate a linear regression model for each pair of elections. This measure considers all districts with contested elections. The dependent variable is the Democrat party vote share in a district i at election t. The independent variables are the Democrat party vote share in a district i at election $t-1$, the incumbent party (1 if Democrat party and -1 if Republican party), and the candidate incumbency (1 if a Democrat incumbent, 0 if an open seat, and -1 if a Republican incumbent).[4] The incumbency effect is estimated by the coefficient of the candidate incumbency. This approach also suffers some shortcomings. As Erikson (2016) explains, "the lagged vote is intended to control for sources of the $t+1$

3 Cox and Katz (1996) propose a modified version of Gelman and King's (1990) method.
4 In section 5, I further explain the Gelman and King's (1990) method.

vote other than incumbency. However, this assumption leads to an unbiased estimate only if incumbent retirement decisions are unrelated to their expected vote share. This assumption is decidedly untrue, so that the lagged vote is a leaky control for the relevant nonincumbency causes of the time *t+1* vote that it is intended to measure. The consequence is that nonincumbency factors masquerade as part of the incumbency advantage (71)." Indeed, Gelman and King (1990) considered a similar issue. They argue that the candidate incumbency parameter can be understood as the interaction between the incumbent party and the decision of the incumbent to seek reelection (1152). However, as Gelman and King (1990) do not estimate a separate parameter for the decision to seek reelection, the model does not include all the constituent terms of such an interaction. Despite such criticisms, Gelman and King's (1990) method continues to be one of the most influential measures in the literature of incumbency advantage.[5]

Finally, the third wave of studies quantifies the incumbency advantage using regression discontinuity analysis. The work of Lee (2008) was the foundational piece in this ever-increasing literature. Using a regression discontinuity approach, Lee (2008) estimates the partisan incumbency advantage by comparing the electoral performance at time *t* of those districts in which the Democratic party *barely won* at *t-1* to those in which the Democratic party *barely lost* at *t-1*. This method assumes a random chance element (that has a continuous density function) to the final vote share. According to Erikson and Titiunik (2015), the measure proposed by Lee (2008) captures the *partisan* incumbency advantage, but not the *personal* incumbency advantage. Also using regression discontinuity analysis, Erikson and Titiunik (2015) propose a measure of personal incumbency advantage which is half of the size of the partisan incumbency advantage. However, the regression discontinuity design approach has also received some criticism. Caughey and Sekhon (2011) show that covariate imbalance is an issue in close U.S. House elections. That is, the districts in which candidates of a particular party barely win and those in which they barely lost differ substantially in pretreatment covariates. Another criticism about the regression discontinuity analysis is that it relies on pooling all elections and focuses only on measuring a single quantity of interest. By doing so, this research strategy is missing the time trend of the incumbency advantage phenomenon. As Gelman (2005) points out, "modeling vote shares gives you the ef-

5 As of November 2021, there have been 896 citations of Gelman and King (1990) according to Google Scholar.

ficiency to get separate estimates for each election year and thus study time trends. I understand the appeal of simply looking at winning and losing, but there is much to be learned by studying vote shares." Despite these shortcomings, the use of regression discontinuity analysis to estimate the incumbency advantage has becoming widespread – not limited to evaluating the incumbency advantage just in U.S. House elections, but also in other countries (Uppal 2009; Ariga 2015; Eggers et al. 2015; Klašnja and Titiunik 2017; Eggers and Spirling 2017).[6]

In sum, the literature on measuring the incumbency advantage continues to advance and, as this discussion has underscored, there is substantial disagreement over which methods yield unbiased and efficient measures. Despite the challenges, there is no doubt that studies have advanced what we know about this phenomenon and which methodological tools are well-suited for hypothesis testing. However, much less effort has been directed to understand what causes changes in the incumbency advantage over time. In the next section, I will briefly present the literature on the causes of the incumbency advantage and its shortcomings.

2.3 *The Causes of the Incumbency Advantage*

Scholars have claimed that the steady increase in the incumbency advantage observed in the 1960s was a consequence of the disengagement of the electorate with political parties (Erikson 1972; Burnham 1974). According to this view, the boost in the incumbency advantage was rooted in significant changes in the electorate, particularly by the surge in the number of independent voters and the fall in party loyalty (Erikson 1972). Incumbents would target independents to leverage their votes (Erikson 1972; Cover 1977). As Burnham (1974) explains, "party as a referent for voting decisions has disintegrated at a very rapid rate during the past decade; and, 'liberated' from such cue-giving constraints, voters have increasingly turned to other cues (210)." As party and incumbency can be thought of as competing voting cues (Ansolabehere and Snyder Jr 2002), the weakening of partisan ties leads to a growth of incumbency safety. Therefore, this explanation holds that the decline of partisan ties favored the onset of

[6] As of November 2021, there have been 2,005 citations to Lee (2008) according to Google Scholar.

an electoral system that is more candidate-oriented. As a result, incumbents are the winners of this process, seeing a rise in their electoral advantage.

To test the partisan dealignment hypothesis, scholars have presented evidence based on different sources of data. For instance, Ferejohn (1977) examined the propensity of voters to vote for the Democrat party based on their party affiliation and their knowledge about the candidates. To do that, the author estimated linear regression models using data from 1958 to 1970. Ferejohn (1977) found that among those voters who identify with a party, they have been using it to a lesser degree when voting. That is, partisan voters were acting similar to independents. Moreover, the results for 1964, 1966, 1968, and 1970 (three out of four election years analyzed) revealed that voters were using incumbency as a voting cue even when they could not remember the name of the incumbent candidate (171). Cover (1977) arrived at a similar conclusion. The author argues that the decline of party loyalty does not present enough evidence indicating that this phenomenon is helping incumbents. Instead, Cover (1977) breaks the defection rate into two components: pro-incumbent (challengers' partisan defecting to incumbents) and pro-challenger (incumbents' partisan defecting to challengers). Using data from 1956 to 1974, the author shows that, since 1970, about three-fourths of all defections are pro-incumbent. More importantly, "since 1972, about half of those identifying with the challenger's party have deserted their party's congressional candidate in contested elections involving an incumbent (Cover 1977, p. 535)." Then he concludes that such a phenomenon indicates a change in the mass electorate behavior. More recently, Jacobson (2015) also presents evidence that corroborates the hypothesis of shifts in mass electoral behavior. Using data from 1956 to 2012, the author shows that the rise of incumbency advantage was simultaneous with the reduction in partisanship and party loyalty, and a process of denationalization of electoral politics. As in recent decades, elections have become more nationalized, and party loyalty has risen, Jacobson (2015) argues that incumbency has weakened as a voting cue.

Despite these studies, evidence for the partisan dealignment theory has not been unequivocal. In an attempt to test whether the changes in incumbency advantage were explained by changes in partisan composition (i.e., partisan dealignment) or by changes in individuals behavior (i.e., incumbents' constituency service), Krehbiel and Wright (1983) decomposed the total change in the rate of incumbency advantage over time into components reflecting these two explanations. They found evidence that partisan dealignment accounted for little of the increase in the incumbency advantage between 1956 and 1978 (140). Thus, the authors concluded that the second explanation was the most appropriate to explain the rise in incumbency voting. Cox and Katz (1996) also adopted a decomposition analysis to investigate the sources of the incumbency advantage. However, they arrived at a different conclusion than Krehbiel and Wright (1983). Cox and Katz (1996) decomposed the incumbency advantage into direct ("reflecting the value of resources (such as staff) attached to the legislative office"), scare-off ("the ability of incumbents to scare off high-quality challenger"), and quality effect ("reflecting how much electoral advantage a party accrues when it has an experienced rather than an inexperienced candidate"). They found that changes in the quality effect mainly explained the growth of the incumbency advantage for the period between 1948 and 1990. According to Cox and Katz (1996), the decline in partisanship within the electorate is the key factor that explains the increases in the quality effect.

As these studies have shown, the partisan dealignment explanation emphasizes changes in the mass electorate's behavior as central to explaining the growth in the electoral pay-off of incumbency. According to this view, the incumbents' activities play only a secondary role in explaining the incumbency advantage. As Mayhew (1974*b*) underscores, "voters dissatisfied with party cues could be reaching for any other cues that are available in deciding how to vote. The incumbency cue is readily at hand. This hypothesis assumes a current rise in discontent with parties; it assumes nothing about changes in the cues voters have been receiving from congressmen (313)." In this sense, this view has been criticized for understanding voters as 'uninformed' and 'superficial' when deciding who to vote for (Jacobson and Carson 2019). Indeed, other explanations have been raised to explain the rise in the incumbency advantage, placing more emphasis, for example, on institutional changes, and on constituency service (e.g., the role of institutional characteristics of the Congress (Mayhew 1974*b*) or the incumbents' constituency service (Fiorina 1977)). For instance, analyzing the trajectory of two electoral districts, Fiorina (1977) argues that the growth of the bureaucracy has given a unique opportunity for members of Congress to expand constituency services. Consequently, this has allowed incumbents to capture votes that otherwise would go to the opposing party (181).

Despite the relevance of these other explanations, the most consistent body of evidence that explains the variation in the incumbency advantage over time has been presented by the party dealignment hypothesis. First, while the constituency service hypothesis may explain the period of increase in the incumbency advantage, it does not do a good job explaining the period of decrease in the incumbency advantage. This hypothesis attributes the growth in the incumbency advantage to the increasing access of incumbents to resources attached to legislative office. To explain the period of decrease in the incumbency advantage, it would be expected empirical evidence showing that, after the 1990s – which is the period when the incumbency advantage started to decrease –incumbents began to give up their valuable resources. However, there is no evidence showing that. Second, following the hypothesis about the role of institutional characteristics of the Congress, it is expected that the incumbency advantage would be a specific phenomenon occurring in the House of Representatives. However, Ansolabehere and Snyder Jr (2002) show that the incumbency advantage is a nation-wide phenomenon. Hence, the explanation for it has to be related to something broader and not circumscribed to the characteristics of the Congress. Analyzing elections in U.S. House, Senate, and all statewide offices from the 1940s to 1990s, Ansolabehere and Snyder Jr (2002) reveal that incumbency advantage has grown for all offices during the period.

Although these studies have provided valuable contributions to understanding the causes of the incumbency advantage, they have two critical shortcomings. First, a significant number of these studies did not employ a formal test to evaluate their hypothesis about what causes variation in the average incumbency advantage. For instance, some studies have relied on the association between two variables to draw inferences about the causal relationship between them without implementing a formal hypothesis test. For example, discussing a figure that shows the incumbency advantage and the number of independents over time, Krehbiel and Wright (1983) say "according to the compositional change explanation, these trends [increases in incumbency advantage and increases in the proportion of independents in the electorate] are not coincidental, but rather are evidence of partisan dealignment at work, both in the long term and especially during the 1964–1972 'critical period' (142)."

However, drawing inferences based on this type of analysis is prone to spurious conclusions. That is, when two variables seem to be associated, but this is – in fact – due to an artifact of omitting a crucial third variable or failing to acknowledge the serial correlation present in explosive time series. On this point, Granger and Newbold (1974) demonstrated why spurious conclusions occur with multivariate models involving nonstationary variables. As I will show in the following sections, the statistical evidence suggests that the incumbency advantage measure is nonstationary.[7] In the classical regression model, the assumption is that the dependent and independent variables are stationary and that the errors have a zero mean and a finite variance. However, when the variables are nonstationary, the classical linear regression model assumptions are violated, and the tests of statistical inference do not hold (Enders 2008). Hence the regression results may yield findings that do not exist ("spurious"). In sum, analysts must follow a series of rigorous procedures and tests to evaluate if there is a causal relationship between the dependent and the explanatory variables. This is especially important in analyzing social science processes, such as incumbency, that evolve over time.

A second limitation of these studies is related to the type of analysis and the data employed. Most of the evidence about the impact of partisan dealignment on incumbency advantage is based on short time periods. As most of the studies focusing on this issue were published in the late 1970s, the period analyzed is usually between the 1940s to the 1970s. In part, this is explained by a lack of a more extended time series to analyze when these works were published. In this sense, most of the studies captured the period when the incumbency advantage began to increase but did not include the fall in recent decades. As the average incumbency advantage has varied significantly over the last five decades (Jacobson and Carson 2019), it is essential to employ methodological techniques that allow investigating the dynamic movements in this phenomenon over time. Among the studies that employed time series analysis to investigate the causes of incumbency advantage (Ansolabehere and Snyder Jr 2002; Carson, Sievert and Williamson 2019; Peskowitz 2019), none of them estimated statistical models using the incumbency advantage as the dependent variable.

Given the strength of the arguments that partisan dealignment is driving the incumbency advantage and the methodological problems in the empirical evidence to date, in this chapter, I focus on testing the partisan dealignment hypothesis with more stringent analytical tests. In the remaining sections of this chapter, I examine the dynamics of incumbency advantage using time series analysis. To do so, I introduce a

[7] A time series is covariance stationary when it exhibits (i) mean reversion, (ii) constant variance, and (iii) constant covariance over time (Asteriou and Hall 2015).

strategy that permits us to estimate the dynamics of the incumbency advantage. Then, I use this measure to examine whether the dynamics of the incumbency advantage are driven by partisan dealignment. In the next section, I discuss the hypotheses that I will test.

2.4 Hypotheses

Several explanations have been offered about how and why partisan dealignment affects the incumbency advantage. In this chapter, I will test four different causal mechanisms that relate partisan dealignment to incumbency advantage.

A first theoretical expectation is that increases in the proportion of independents in the electorate will lead to increases in the incumbency advantage (Cover 1977; Krehbiel and Wright 1983). As Erikson (1972) suggests, "as voters display greater partisan ambivalence, a factor such as the incumbent's visibility is likely to tip the balance in a greater number of voter decision (1240)." Therefore, the decline in the number of partisan identifiers would favor incumbents since voters will tend to rely more on incumbency than on the party as a cue to vote.

A second theoretical expectation is that incumbency advantage will increase as party loyalty decreases in the elections contested by the two major parties (Ferejohn 1977; Jacobson 2015). Party loyalty is defined as the proportion of voters that consistently vote for candidates in House elections who also share their partisan identity. That is, the expectation is that voters that identify with the Democrat party will vote for a Democratic candidate in the House elections regardless of who is the incumbent. Alternatively, it is expected that voters that identify as Republican to vote for the Republican candidate. Thus, as the proportion of voters who decide their votes independent of their party affiliation increases, the incumbency advantage will increase.

More importantly, I expect that lower levels of party loyalty will especially benefit incumbents when voters betray their parties and vote for the incumbent from the opposite party. For instance, this pattern would be observed in situations in which voters that identify with the Democrat party vote for a Republican incumbent in the House elections. Alternatively, this would also occur if voters that identify with the Republican party vote for a Democrat incumbent in House elections. Table 1 describes these patterns. Thus, a third expectation is that increases in challengers' partisans defecting to the incumbent's party will lead to increases in the incumbency advantage (Jacobson 2015). That is, as the percentage of voters who break away from their party to vote for the incumbent increases, more the incumbents will benefit.

Table 1 – Voting Patterns and Challengers' Partisan Defecting to Incumbents

Voters' Party Affiliation	Incumbent's party	Vote in House elections	Challengers' Partisan Defecting to Incumbents
D	R	D	No
D	R	R	Yes
R	D	R	No
R	D	D	Yes

Note: A Democrat candidate or voter is denoted by 'D' and a Republican candidate or voter by 'R.'

Finally, a fourth theoretical expectation is that ideological polarization will affect the incumbency advantage. Polarization – defined as "movement away from the center toward the extremes (Fiorina and Abrams 2008)" – is a key factor to explore since this phenomenon may amplify the perceived differences between political parties. I expect that the average incumbency advantage will decrease as the partisan ideological divisions increase. That is, voters are more likely to cast their votes following partisan lines than incumbency when they perceive clear differences between the parties running for election.

Table 2 summarizes my theoretical expectations. Although all four independent variables are measures of dealignment, there is an important distinction between them. Except for polarization, all the other three measures are at the behavioral level. By exploring the impact of these three measures (independents in the electorate, party loyalty, challengers' partisans defecting to incumbents), we analyze how changes in mass electorate behavior affect incumbency advantage. In contrast, polarization is a measure at the elite political level. This measure captures how changes in the cues voters receive from congressmen and political parties impact the incumbency advantage. By using these four different measures, I seek to test distinct implications of the partisan dealignment theory.

Table 2 – Summary of Theoretical Expectations of the Partisan Dealignment Theory

Independent Variables	**Expected Effects on Incumbency Advantage**
% Independents in the Electorate	+
% Party Loyalty	−
% Challengers' Partisans Defecting to Incumbents	+
Polarization	−

2.5 *Research Design*

In this section, I explain the two stages of the research design. First, I discuss the modeling strategy – that is how I will generate the time-series cross-sectional observations. Next, I introduce the model specification I will employ in hypothesis tests.

2.5.1 The Data Generating Process Employed to Measure the Incumbency Advantage

The data employed in this study are from the U.S. House elections from 1946 to 2014. The sample comprises 34 Congressional elections which were analyzed by Jacobson (2015). Table 8 shows the summary statistics of the electoral data for the period of analysis, which is the dependent variable in this chapter. The four different measures (% independents in the electorate, % party loyalty, % challengers' partisans defecting to incumbents, and polarization) that explore distinct aspects of partisan dealignment were obtained from the American National Election Studies Cumulative Data File (ANES) and the Lewis et al. (2019).

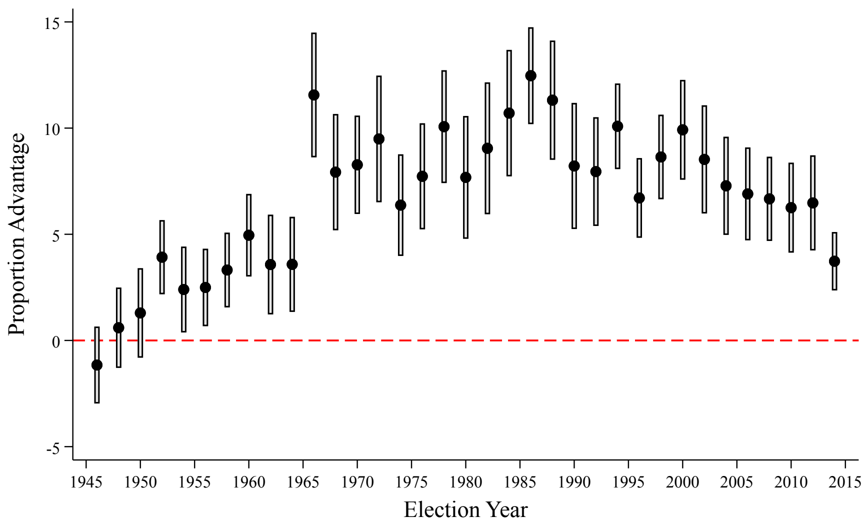
As I have explained, several methods have been proposed to measure the incumbency effect. The dependent variable is based on a first stage analysis where I run cross-sectional models for each election to generate a measure of the incumbency effect. To estimate the incumbency advantage, I employ Gelman and King (1990)'s estimator. The authors define incumbency advantage as "the average gain in the proportion of the district vote that the incumbent party receives if its incumbent candidate runs for reelection (King and Gelman 1991, p. 112)." Gelman and King (1990)'s model can be summarized as follows:

$$\text{Dem. Vote Share}_{it} = \beta_{0t} + \beta_{1t}\ \text{Dem. Vote Share}_{it-1} + \beta_{2t}\ \text{Party}_{it} + \psi_t\ \text{Incumbency}_{it} + \varepsilon_{it}$$

For a pair of elections ($election_t$ and $election_{t-1}$), the incumbency effect measure, ψ_t, is based on a linear regression model in which the dependent variable is Democratic party vote share and the key independent variable is incumbency status. Incumbency is coded as equal 1 if a Democratic incumbent runs for reelection, 0 if it is an open seat (no incumbent runs), and -1 if a Republican incumbent runs for reelection. Democratic party vote share in the previous election and partisan swing (coded as 1 if the Democrat wins the previous election, and -1 if the Republican wins) are added as control variables. Figure 1 depicts the incumbency effects for each election and the respective 95% confidence intervals.

I follow this specification and use it to estimate a regression for each election year. The coefficient estimates for the incumbency advantage represents the average effect across all districts in election t.

Figure 1 – Estimates of House Incumbency Advantage, 1946-2014



There are some important patterns to highlight from Figure 1. First, the effect of incumbency was not statistically distinguishable from zero in the elections that occurred during the 1940s. The 95% confidence interval for the estimated incumbency advantage includes zero in three elections (1946, 1948, and 1950). This means that there was not an electoral advantage or disadvantage of incumbency. Additionally, during the period between 1946 to 2014, there was not any case of electoral disadvantage of incumbency. In the 1946 election, the incumbency advantage was -1.15% (or between -2.97 and 0.66 using a 95% confidence interval), but as I highlighted above the 95% confidence interval overlaps zero. Second, the incumbency advantage achieved its highest levels in the elections between the 1960s and 1980s. The largest effect was observed in the 1986 election with an estimated incumbency advantage of 12.47% (or between 10.19 and 14.75 using a 95% confidence interval). Third, there has been a decrease in the magnitude of the incumbency advantage in recent elections. For instance, in the 2000 House elections, incumbents enjoyed an estimated advantage of 9.91% points (or between 5.59% and 14.24% using a 95% confidence interval) over challengers. However, in the 2014 elections, the magnitude of the incumbency advantage fell to 3.73% (with a 95% confidence interval estimating the effect to be as low as 1.85% and high as 5.61%). In summary, this figure shows that the incumbency effects have varied significantly over time and, for most of the time, incumbents had enjoyed an electoral advantage over challengers.

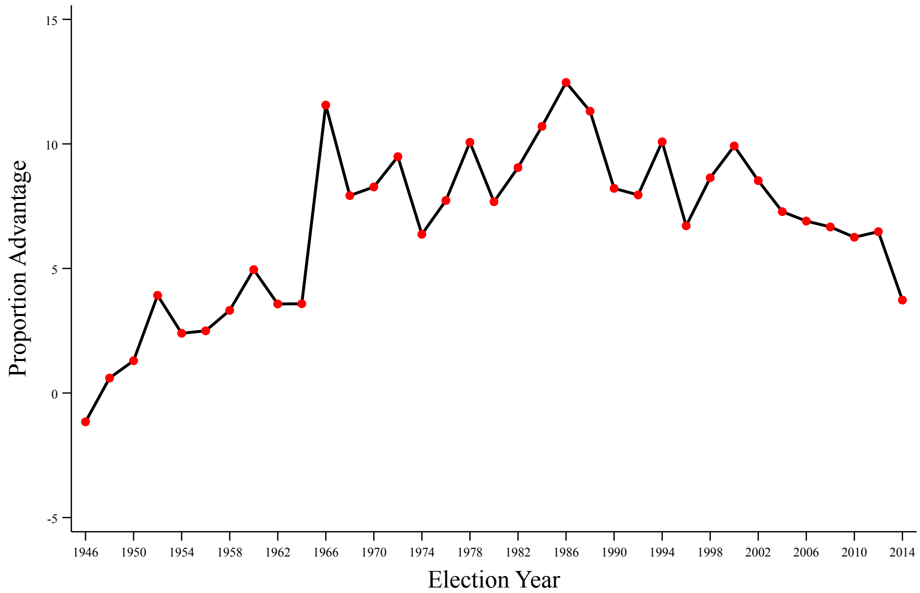
In Figure 1, each point estimate represents the estimated effect across all districts in a specific election. Taken together, these estimates show the variation in the incumbency advantage over time. Therefore, Gelman and King (1990)'s strategy to measure incumbency effects is key to the first stage of this analysis because it allows us to have a measure of the incumbency effect for each election and thus allows to capture its trends. As highlighted by Alesina and Rosenthal (1995), "by allowing for all the parameters to vary from election to election, Gelman and King capture, in addition to structural change in incumbency advantage, many other sources of intertemporal variation (139)."[8]

After estimating incumbency effects for each election, I now pool these estimates to create the dependent variable that I will use in the second stage of the analysis. Figure 2 shows the time series of incumbency effects.

After the 1950s, there is an abrupt change in the trajectory of the series. Some scholars argue that the 1952 election is a break point in the series (LeVeck and Nail 2016).

[8] The Gelman and King (1990)'s strategy assumes that the incumbency effect is constant across the districts in a specific election. Although this may not be the case, exploring the cross-sectional variation in incumbency effects is outside the scope of this chapter.

Figure 2 – Time Series of the Estimates of House Incumbency Advantage, 1946-2014



I tested for a structural break with a known break date to verify if there was a break in the series and in which year. For each year, I performed a test to check if that year was a break point. The null hypothesis of the test is that there was not a structural break. The results show that there is evidence to reject the null hypothesis of no structural break at the 1% level for the year of 1966 and at the 5% level for the year of 1964.[9] Given the evidence of possible structural breaks in the incumbency advantage series, I should not use the usual unit root test, e.g., Augmented Dickey-Fuller, without modification. When we have a series with structural breaks, the mean, variance, and covariance might be significantly different in the period before and after the break. Consequently, the series might be considered as nonstationary. However, it might be the case that its parts are stationary (Levendis 2018, p.172).[10]

[9] I performed a Wald test for a structural break at a known break date for the estimation results of a linear regression of $\widehat{Incumbency}_t$ on $\widehat{Incumbency}_{t-1}$. This routine was implemented in Stata using the command *estat sbknown*.

[10] To test if the series follow a stationary process with a break in the 1966 election, I employ the unit root test developed by Perron (1989). I implement the unit root test using Levendis (2018)'s simplified routine. I proceed by testing a general model that includes: D_L, a dummy variable indicating one for the period after the 1966 election ($D_L = 1$ if $time > 1966$, and 0 otherwise), D_P, a dummy variable indicating the one election right after the 1966 election ($D_P = 1$ if $time = 1968$, and 0 otherwise), a *time* variable, a lag of the dependent variable (y_{t-1}), and three lags of the difference of the dependent variable (Δy_{t-i}). With a much larger sample size, Perron (1989) presents the results for lag parameter k between 1 and 12 (see Perron's explanation in page 1385). For instance, one of the variables tested by the author is GNP. This variable has $T = 62$ observations. As I only have $T = 35$ observations, I choose to report the result of $k = 3$. Besides this model, I also specify a similar test that includes one more term to model a change in the slope (*NewSlope*). The inclusion of this term modifies the test. While in the first test, the hypothesis is that the slope or the intercept has changed, in the second test, the

Since the 95% confidence intervals of the estimated value of α, the coefficient on y_{t-1}, includes 1, we cannot reject the null that the series follows a unit root process. We then have evidence that incumbency advantage is I(1). We also find evidence that the break in 1966 caused a change in both the slope and intercept of the series. Besides Perron's test, I also estimate the Zivot and Andrews' (1992) unit test which allows to test for unit root with an unknown break. As in the Perron's test, we are not able to reject the null of unit root of the test with break in both intercept and trend. The t-statistic is -3.31, at break in 1966, and critical values are -5.57 (for 1%), -5.08 (for 5%), and -4.82 (for 10%).[11] As a robustness check, I have also conducted a series of six unit root tests (Augmented Dickey-Fuller, Augmented Dickey-Fuller (with trend), Phillips-Perron (with trend), Dickey-Fuller GLS, Elliott-Rothenberg-Stock, and Kwiatkowski-Phillips-Schmidt-Shin (KPSS)), see Tables 12 and 13. As I have 1,000 estimated values of incumbency advantage for each election year, in Tables 12 and 13, I report the results of the average of these 1,000 simulations. Given the results of the tests, there is enough evidence to conclude that incumbency advantage is I(1) in levels, and I(0) in the first difference.

2.5.2 Using the Incumbency Advantage DGP to Examine the Effect of Dealignment

After running each of these 34 models, I pool these estimated incumbency effects to construct a dependent variable that captures the mean estimated incumbency effect across districts in each election. I use this dependent variable to test whether partisan dealignment affects the incumbency advantage.

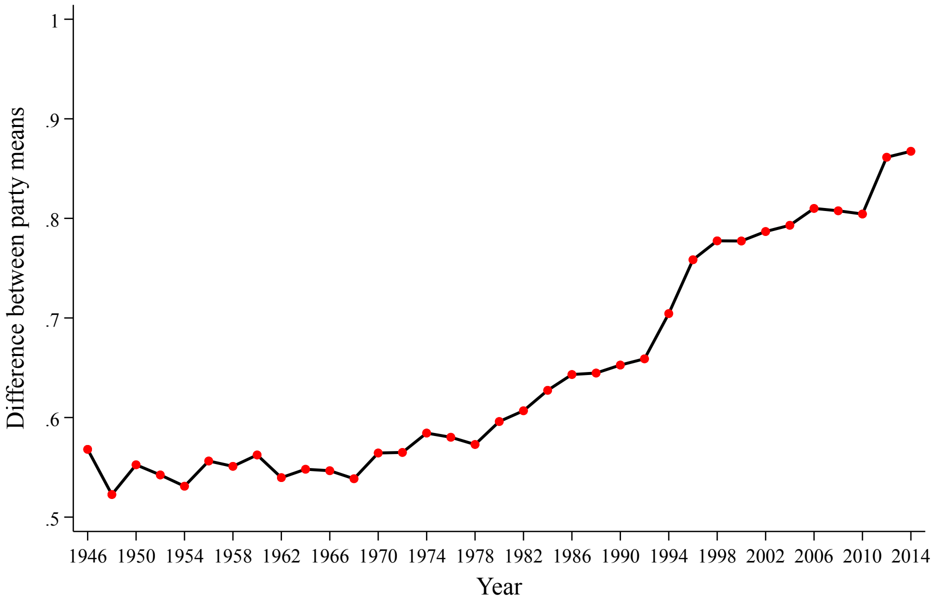
To measure polarization, I use data on polarization in the U.S. House of Representatives from Lewis et al. (2019) from 1946 to 2014. This variable measures the average ideological differences between the policies enacted by Democrats and Republicans over time. Polarization exhibits a clear upward trend as Figure 3 shows. Since the 1970s, there has been a steady increase in polarization in the House of Representatives. The lowest difference between party means was observed in the 1950s with an average

hypothesis is that both has changed. In both tests, the unit root hypothesis ($\alpha < 1$) can be tested using Perron's critical values. However, as highlighted by Levendis (2018), "critical values are larger when the break is in the middle of the series (183)." This is not the case here since the 1966 break is located in one-third of the series (31.4%). Table 11 presents the results of these two specifications.

11 I implemented Zivot and Andrews' (1992) technique using *zandrews* program in Stata.

around 0.55. However, the highest levels were reached in the 2010s with an average of approximately 0.84.

Figure 3 – House Liberal-Conservative Partisan Polarization



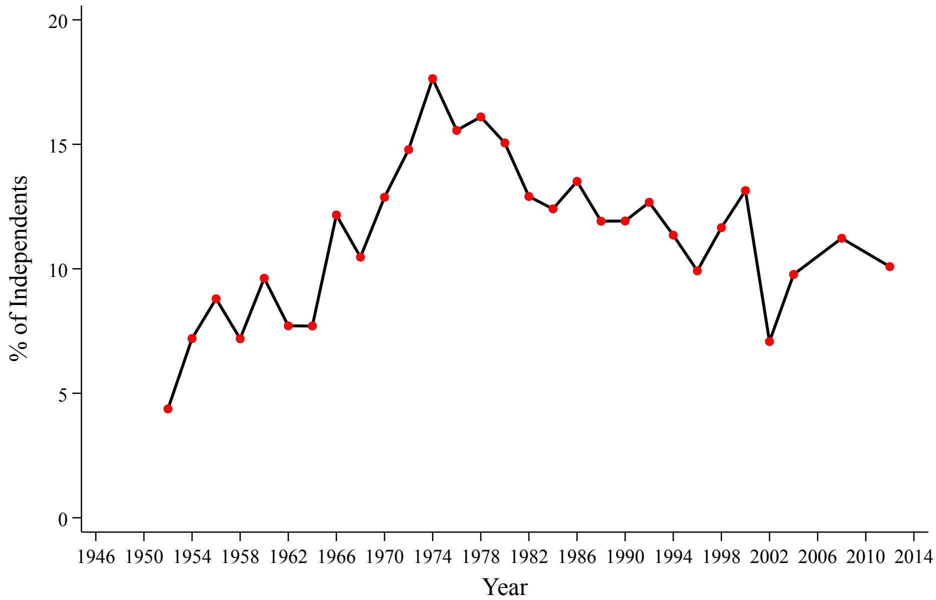
Source: Lewis et al. (2019).

As a measure of the % independents in the electorate, I employ data from the American National Election Studies Cumulative Data File (ANES) from 1952 to 2012. In this measure, I only consider pure independents.[12] Thus, I do not include as independents those leaning towards Democrat or Republican parties. As Jacobson and Carson (2019) underscore, "voters who lean toward a party are nearly as loyal to that party's candidates in House elections as are weak party identifiers." The % independents in the electorate has varied significantly over time as Figure 4 illustrates. There are a growing number of Independents from the 1950s up until the middle of the 1970s. However, since the 1980s, the number of respondents that consider themselves as independents has fallen. During this period, the percentage of independents averaged 11.3. The lowest level of 4.4% was registered in 1952. And, in 1974, the highest level of 17.6% of independents was achieved.

For the measures of % party loyalty and % challengers' partisans defecting to incumbents, I use data from American National Election Studies Cumulative Data File (ANES) and I follow Jacobson's (2015) definitions. A respondent exhibits party loyalty

[12] I coded independents equal to 1 for those respondents that answered as 'Independent - Independent' (item 4) for the question of party identification (VCF0301), otherwise 0.

Figure 4 – Percentage of Independents in the Electorate



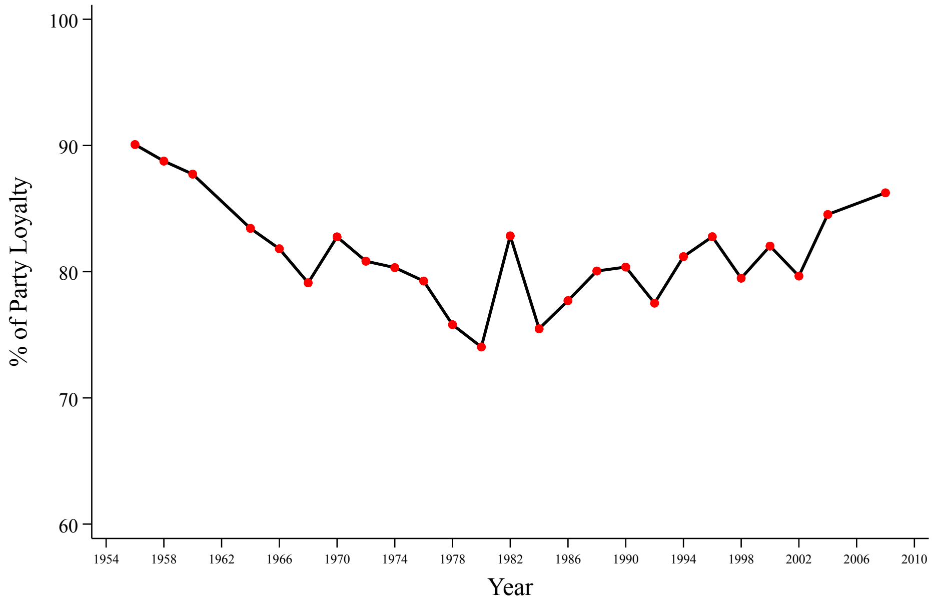
Source: ANES.

when her vote in the Congressional election is consistent with her party identification. When her vote is not consistent, there is no party loyalty. This measure also exhibits a pattern that varies significantly over time as Figure 5 shows. From 1956 to 2008, the lowest level of party loyalty in the House elections was detected in 1980 with 74% of the respondents reporting a Congressional vote consistent with their partisan alignment. The highest level was observed in 1956 with 90% of the respondent being loyal to their parties in the House elections. Over this period, the average % of party loyalty was 81.

Although there is a high level of party loyalty, there are some exceptions. One of those exceptions happens when challengers' partisan voters defect and vote with the incumbent from the opposite party. There has been also a significant variation over time in the % of voters that defect their parties and vote with the incumbent as Figure 6 shows. The lowest level was observed in 1956 with 15% of challengers' partisans defecting to incumbents. And, the highest level was seen in 1990 when 55.5% of respondents abandoned their parties and voted for the incumbent from the opposite party. During 1956 to 2008, the average % of challengers' partisans defecting to incumbents was 35.9%.

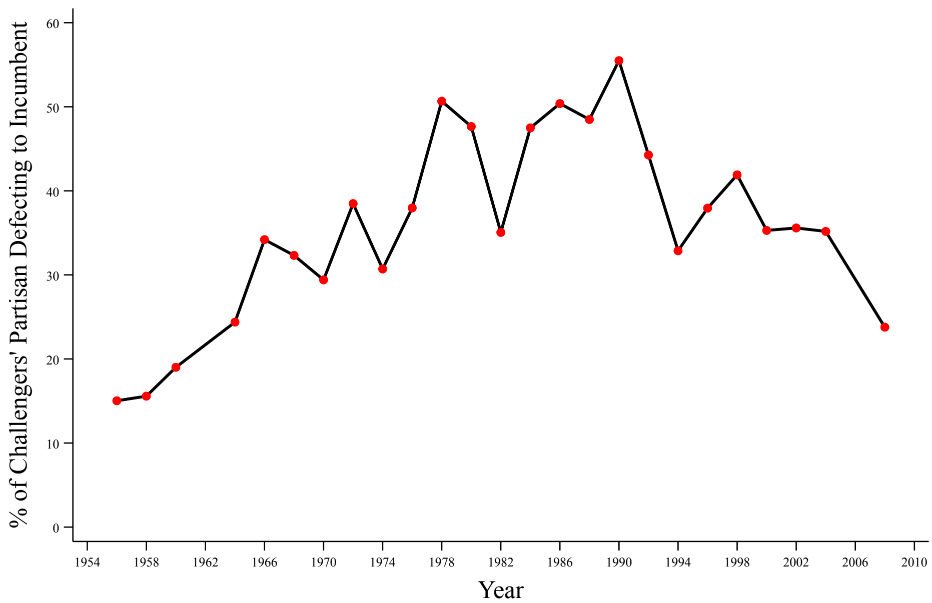
Both measures of party loyalty and challengers' partisans defecting to incumbents are only able to be different from zero in contests in which the two major parties (Democrat and Republican parties) are running and there is an incumbent and a challenger (Jacobson 2015). In this sense, I did not consider the cases in which the election

Figure 5 – House Party Loyalty



Source: ANES and Jacobson (2015).

Figure 6 – Challengers' Partisan Defecting to Incumbents



Source: ANES and Jacobson (2015).

was an open seat, when the election was not contested, and when there was only incumbents or only challengers running as may happen after redistricting. After excluding the cases mentioned above, there are on average 368 districts in each election year.

It should be noted that there are strong correlations among some of the different variables analyzed in this study. This underscores that, although some commonalities, the four independent variables capture distinct aspects of the partisan dealignment theory. The correlation between party loyalty and independents is -0.6 and statistically significant at 1% level. Party loyalty is even more correlated with challengers' partisan defecting to incumbents. There is a correlation between these two variables of -0.85 which is statistically significant at the 1% level. Challengers' partisan defecting to incumbents is correlated with independents to a lesser degree. There is a correlation of 0.52 which is statistically significant at the 1% level. Polarization is the only variable that is not correlated with any of the other three variables. Polarization captures a process happening between the two major parties in the House of Representatives. In turn, the variation in the percent of independents captures how party identification is changing. Finally, party loyalty and challengers' partisan defecting both capture distinct electoral patterns. The first one emphasizes loyalty to the party, while the second one disloyalty to the party that benefits the incumbent.

Second Stage: Dynamic Specification

In the first stage of the analysis, I estimated a linear regression model that examines the average effect of incumbency on vote share across electoral districts in a specific election. Then, I use the incumbency estimate for each of the 34 elections as the dependent variable in the second stage of the analysis. This is not the first study in the field that employs estimated incumbency effects as the dependent variable. To investigate the impact of constituency services on incumbency advantage, King (1991) employs the estimated incumbency advantage as the dependent variable in its analysis.[13] I now turn to examine the dynamic relationship between incumbency advantage and partisan dealignment.

[13] As explanatory variables, King (1991) uses the legislative operating budget, salary, and the lag of the dependent variable (incumbency advantage).

As in the second stage of the analysis, the dependent variable is an estimate from the first stage, it incorporates a measurement error that brings additional uncertainty (Williams 2013; Lewis and Linzer 2005). To account for the additional uncertainty around the estimates of incumbency effects, I employ a procedure similar to Williams (2013).

- First, I generate 1,000 estimates of incumbency advantage for each election using *Clarify* (**?**).
- Then, in the second stage, I estimate the model 1,000 separate times. For each time, I do 100 draws from the multivariate normal distribution using *Clarify* to incorporate uncertainty from the other parameters.[14]
- Finally, I combine the 100,000 estimates and use the percentile method to derive confidence intervals.

As for the dependent variable, I also conducted a series of unit root tests to check the order of integration of the independent variables. For each variable, I performed six unit root tests (Augmented Dickey-Fuller, Augmented Dickey-Fuller (with trend), Phillips-Perron (with trend), Dickey-Fuller GLS, Elliott-Rothenberg-Stock, and Kwiatkowski-Phillips-Schmidt-Shin (KPSS)). For all tests, the null hypothesis is that the series follows a unit root process. The only exception is the KPSS test in which the null hypothesis is that the series follows a trend stationary process. The results of the test with the variables in levels and in the first difference are reported in the appendix. All four variables are nonstationary in level. However, they are stationary in the first difference.

Given problems of multicollinearity and also the limited number of observations, Equation 1 to 4 present the most parsimonious models that allows us to test our hypotheses. As I only have 22 observations for *% of party loyalty* and *% of challengers' partisan defecting to incumbents*, I do not have enough degrees of freedom to estimate a full model including all four explanatory variables at the right hand side. Therefore, I estimate the four following error correction models (ECM). As the dependent and the explanatory variables are I(1), ECM is a particularly interesting model to estimate since it allows us to evaluate the short- and the long-term effects of a shock in each explanatory variable on the dependent variable. In each model, I included the key explanatory variable and

[14] In this second stage, I also have estimated the models 1,000 separate times without running *Clarify*. The results are consistent.

a dummy variable to capture the structural break in the incumbency advantage series that occurred in the 1966 election.

$$\Delta Inc\,Adv_t = \theta_0 + \theta_1 Inc\,Adv_{t-1} + \theta_2 \Delta Polarization_t + \theta_3 Polarization_{t-1} + \theta_4 Election\ 1966_t + \varepsilon_t \tag{1}$$

$$\Delta Inc\,Adv_t = \lambda_0 + \lambda_1 Inc\,Adv_{t-1} + \lambda_2 \Delta Indep_t + \lambda_3 Indep_{t-1} + \lambda_4 Election\ 1966_t + \varepsilon_t \tag{2}$$

$$\Delta Inc\,Adv_t = \alpha_0 + \alpha_1 Inc\,Adv_{t-1} + \alpha_2 \Delta Loyalty_t + \alpha_3 Loyalty_{t-1} + \alpha_4 Election\ 1966_t + \varepsilon_t \tag{3}$$

$$\Delta Inc\,Adv_t = \gamma_0 + \gamma_1 Inc\,Adv_{t-1} + \gamma_2 \Delta Chal\ Defection_t + \gamma_3 Chal\ Defection_{t-1} + \gamma_4 Election\ 1966_t + \varepsilon_t \tag{4}$$

To verify whether the incumbency advantage is cointegrated with each of the four independent variables, I follow the procedures outlined by Philips (2018) and apply the bounds test in the residuals of each error correction model. For the estimated model with *polarization* as the explanatory variable (Equation 1), the 5% critical values for 34 observations and one regressor are 5.29 for the lower bound and 6.17 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the average of the 1,000 estimated F-statistics is 10.1, there is statistical evidence of cointegration between the *polarization* and the incumbency advantage series. The analysis of the t-test also corroborates such conclusion.[15]

For the estimated model with *% of independents* as the explanatory variable (Equation 2), the sample size is smaller (26 observations). The 10% critical values for 26 observations and one regressor are 4.29 for the lower bound and 5.08 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the average of the 1,000 estimated F-statistics is 5.58, there is statistical evidence of cointegration between the *% of independents* and the incumbency advantage series at 10% significance

[15] The 5% critical values for the t-test are -2.86 for the lower bound and -3.22 for the upper bound, as the average t-statistic is -4.45, I can conclude that there is evidence of cointegration between incumbency advantage and *polarization*.

level. The analysis of the t-test also corroborates such conclusion at 1% significance level.[16]

For the estimated model with *% of party loyalty* as the explanatory variable (Equation 3), the sample size is even smaller with only 22 observations. However, in this case I have evidence of cointegration between *% of party loyalty* and incumbency advantage at 1% significance level. The 1% critical values for 22 observations and one regressor are 8.17 for the lower bound and 9.28 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the average of the 1,000 estimated F-statistics is 10.64, there is statistical evidence of cointegration at 1% significance level. The analysis of the t-test also corroborates such conclusion at 5% level.[17]

The sample size for the estimated model with *% of challengers' partisan defecting to incumbents* as the explanatory variable (Equation 4) has also only 22 observations. Here, there is also evidence of cointegration at 1% significance level. As in the previous model, the 1% critical values for 22 observations and one regressor are 8.17 for the lower bound and 9.28 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the average of the 1,000 estimated F-statistics is 9.73, there is statistical evidence of cointegration at 1% significance level. The analysis of the t-test also corroborates such conclusion at 1% level.[18]

2.6 Results

I have argued that as the political scenario becomes more polarized voters will tend to cast their votes following more partisan cues than incumbency. In other words, I expect to find a negative relationship between the incumbency advantage and polarization. To assess this hypothesis, I calculate and test the statistical significance of the short- and the long-run effects of polarization on the incumbency advantage. Table 22 shows the results for this model. There is no evidence to reject the null that the short-term effect is different from zero. However, the long-term effect is statistically significant. In the

[16] As the average t-statistic is -3.24, and the 5% critical values for the t-test are -2.86 for the lower bound and -3.22 for the upper bound, I can conclude that there is evidence of cointegration between incumbency advantage and *% of independents*.

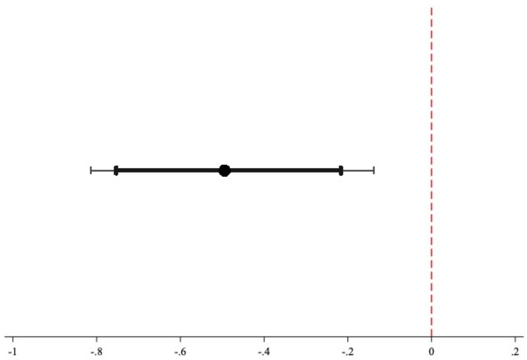
[17] I can conclude that there is evidence of cointegration, since the 1% critical values for the t-test are -3.43 for the lower bound and -3.82 for the upper bound, and the average t-statistic is -4.50.

[18] As in the previous case, the 1% critical values for the t-test are -3.43 for the lower bound and -3.82 for the upper bound. As the average t-statistic is -4.26, there is evidence of cointegration between the incumbency advantage and the *% of challengers' partisan defecting to incumbents* series.

long-run, an increase of 1 unit in polarization (i.e., in the difference between Democrat and Republican in the House of Representatives) can reduce the average incumbency advantage in -0.49 (or between -0.81 and -0.14 using 95% confidence interval), as Figure 7 Panel A confirms. Using dynamic simulations, I then simulate the effect of a shock of 1 standard deviation (11.07) in polarization on the incumbency advantage at time 10 using stochastic simulation (Jordan and Philips 2018; Philips 2018).[19] Figure 7 Panel B shows the expected value of the incumbency advantage before and after the shock. As the figure makes clear, the expected value of the incumbency advantage decreases substantially after the shock.

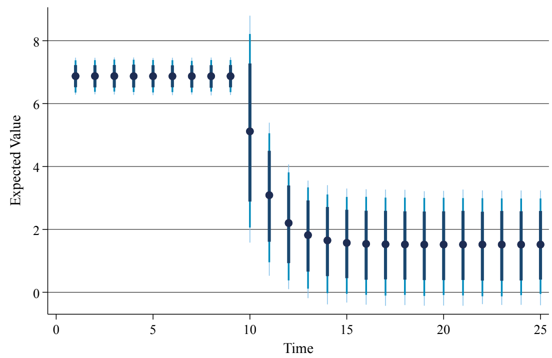
Figure 7 – The Effect of Polarization on Incumbency Advantage

Panel A. Estimated Long-Term Effect



Note: The figure depicts 90% and 95% confidence intervals.

Panel B. Dynamic Simulation



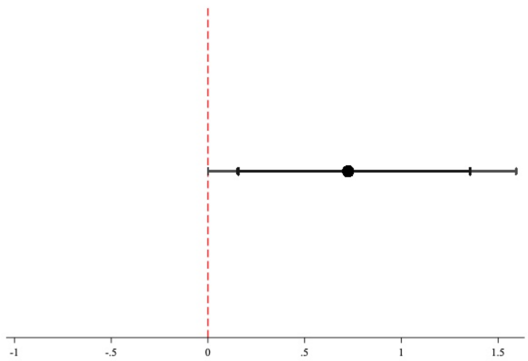
Note: The figure depicts 75%, 90% and 95% confidence intervals.

I also test if the incumbency advantage grows when the % of independents in the electorate expands. Table 23 shows the results for this model. Although we cannot reject the null that the short-term effect is different from zero, the long-term effect is statistically significant. The Figure 8 depicts the long-term effect in Panel A. An increase of 1% of independents in the electorate can lead to an increase of 0.72 in the incumbency advantage (or between 0.01 and 1.59 using 95% confidence interval). In Figure 8 Panel B, I present the expected value of incumbency advantage before and after an increase on % of independents. As the dynamic simulation shows, there is a substantial increase on the expected value of the incumbency advantage after an increase of 1 standard deviation (3.08) on % of independents.

[19] In Stata, I used the command *dynardl* that automatically implement the routine to calculate the dynamic simulations.

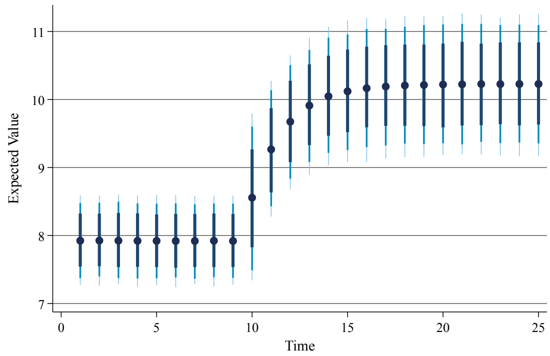
Figure 8 – The Effect of % of Independents in the Electorate on Incumbency Advantage

Panel A. Estimated Long-Term Effect



Note: The figure depicts 90% and 95% confidence intervals.

Panel B. Dynamic Simulation

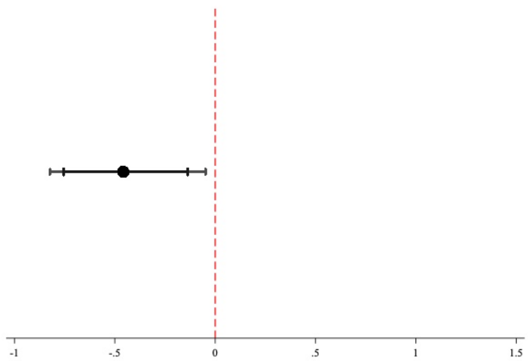


Note: The figure depicts 75%, 90% and 95% confidence intervals.

I have also calculated the effects of % of party loyalty on incumbency advantage. I expect that the average incumbency advantage will grow as increases the percentage of voters who are loyal to their party identification when they place their votes for the House of Representatives. Table 24 presents the results for this model. Despite the fact that the short-term effect is not statistically significant, the long-term effect is negative and statistically significant as expected (Figure 9 Panel A). This indicates that an increase of 1% in party loyalty leads to a decrease of 0.46 (or between -0.05 and -0.82 using a 95% confidence interval). As shows Figure 9 Panel B, the results of the dynamic simulation also corroborates the substantive and statistically significant effect of the impact of party loyalty on the average incumbency advantage.

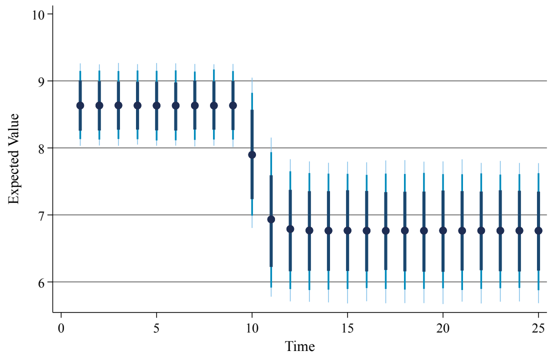
Figure 9 – The Effect of % of Party Loyalty on Incumbency Advantage

Panel A. Estimated Long-Term Effect



Note: The figure depicts 90% and 95% confidence intervals.

Panel B. Dynamic Simulation

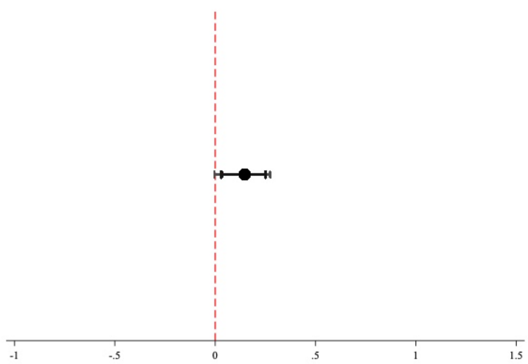


Note: The figure depicts 75%, 90% and 95% confidence intervals.

Finally, I also evaluated the effect on incumbency advantage of changes in % of challengers' partisan defecting to incumbents. I expect to find a positive relationship between these two variables. Table 25 reports the results for this model. While the short-term effect is not statistically significant, the long-run impact is significant only at 10% significance level. The long-term effect on incumbency advantage of an increase in % of challengers' partisan defecting to incumbents is of 0.15 (or between 0.03 and 0.25 using a 90% confidence interval), as shows Panel A of Figure 10. This indicates that an increase on % of challengers' partisan defecting to incumbents leads to a substantial grow in the average incumbency advantage, as shows the dynamic simulations (Panel B of Figure 10).

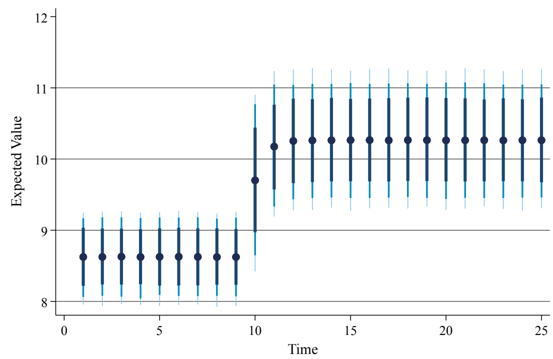
Figure 10 – The Effect of % of Challengers' Partisan Defecting to Incumbents on Incumbency Advantage

Panel A. Estimated Long-Term Effect



Note: The figure depicts 90% and 95% confidence intervals.

Panel B. Dynamic Simulation



Note: The figure depicts 75%, 90% and 95% confidence intervals.

In summary, in this section, I have tested the four hypotheses and the results presented provide robust evidence that corroborate my theoretical expectations. Table 3 summarizes these findings.

Table 3 – Summary of the results

Independent Variables	**Expected Effects on Incumbency Advantage**	**Findings**
% Independents in the Electorate	+	+
% Party Loyalty	–	–
% Challengers' Partisans Defecting to Incumbents	+	+
Polarization	–	–

2.7 *Robustness check: Revisiting Results within a Causal Inference Selected Sub-sample of U.S. House Elections*

Having presented findings that show that party dealignment is contributing to declines in the incumbency effect, this section present a robustness check. Since most recent studies have employed a sub-sample of elections, I test the same hypotheses as advanced earlier focusing on the RDD measure in the first stage instead of Gelman and King's (1990) measure to generate the dependent variable. To test my theoretical expectations, I use the same model specification and data as Erikson and Titiunik (2015) employed in their analyses. This study is considered a seminal study in identifying the causal effect of the personal incumbency advantage. I use the model specification as reported in Table 1, Panel (A), first column (Erikson and Titiunik 2015, p. 114). For this model, Erikson and Titiunik (2015) estimate a parametric linear regression of incumbency in U.S. House elections for all open seat contests at *t* where freshman incumbents run again at *t+1*. Although their dataset includes data from 1968 to 2008, the authors only analyze a sub-sample of this data (399 out of 9,134 cases). Table 9 shows a summary of the dataset and the sub-sample used by the authors.

The running variable used by Erikson and Titiunik (2015) is the Democratic margin of victory at *t* and the outcome variable is the Democratic vote share at *t+1*. Therefore, their analysis is focused on Congressional districts with elections near 50% (vote threshold) at time *t* and which the incumbent run again at time *t+1*. Their model specification can be summarized as follows:

$$\text{Dem. Vote Share}_{it} = \delta_0 + \delta_1 \text{ Dem. Vote Share}_{it-1} + \delta_2 \text{ Dem. Win } t_{it} + \varepsilon_{it}$$

where δ_2 captures the personal incumbency advantage. *Dem. Win* takes on values 1 if at time *t* the winner is a Democrat and -1 if it is a Republican.

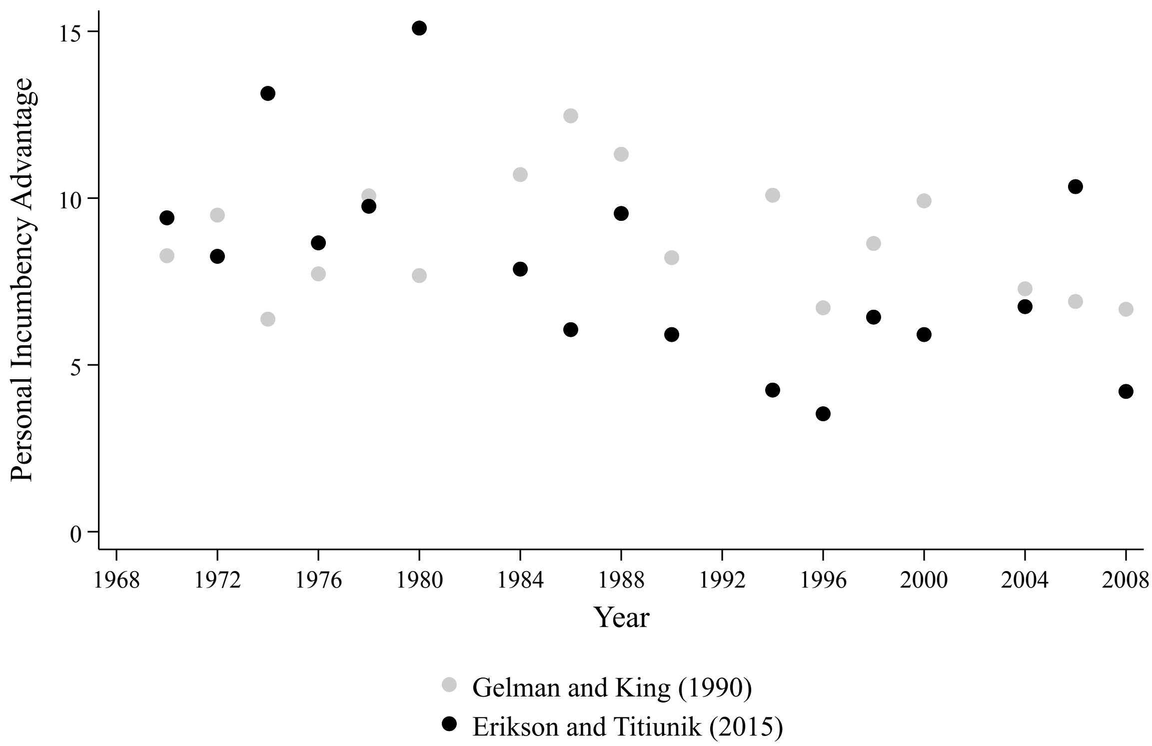
Based on close elections, the authors report a personal incumbency advantage of 6.8. As I seek to analyze what drives changes in the incumbency advantage over time, I run the same linear regression as Erikson and Titiunik (2015), but now I run a regression for each year separately. Then, I save the estimated incumbency advantage for each election year. Figure 12 depicts the incumbency effects for each election year. For four elections, I was not able to estimate a linear regression model because of the limited number of observations available. This is the cases of elections 1968, 1982, 1992, and 2002. The average of the estimated incumbency advantage for the remaining 17 elections is 7.95. This incumbency effect is similar to the 7.66 from a model in which they estimated a linear regression including dummy variables for each election year in Table 1, column 2 (Erikson and Titiunik 2015, p. 144).

Figure 11 – Time Series of the Estimates of House Incumbency Advantage, 1968-2008

Source: Estimates based on the RD incumbency effect model estimated by Erikson and Titiunik (2015).

The estimates from Figure 12 show a distinct pattern that differs from the estimates using the Gelman and King's (1990) measure. In fact, the correlation between the estimates from the Gelman and King (1990) and the Erikson and Titiunik (2015) measures is negative (-0.15), but not statistically significant. The largest differences between the two methods are seen in the years 1980 (diff = -7.41), 1974 (diff = -6.77), and 1986 (diff = 6.41). Despite these differences, I now turn the analysis to test if there is statistical evidence that corroborates my hypotheses with these alternative measure for the dependent variable central to this study.

Figure 12 – Estimates of House Incumbency Advantage: Gelman and King (1990) vs Erikson and Titiunik (2015)



Given the limited number of observations, I am not able to estimate all four models. As I only have 17 data points, I decided to only test the hypothesis about the impact of polarization on incumbency advantage. By adding the lag and the first difference of the dependent and the explanatory variables, the number of observations gets even smaller (13). For the other variables (% party loyalty, % independents, and % challengers' partisan defecting to incumbents), as the number of observations is limited, I was not able to estimate an ECM. I do not find evidence of cointegration between polarization and incumbency advantage. The 10% critical values for 13 observations and one regressor are 4.290 for the lower bound and 5.08 for the upper bound. As the average of the 1,000 estimated F-statistics is 3.863, there is no statistical evidence of cointegration between the series. Thus, I proceed by estimating a first-difference model. However, the average effect of polarization on incumbency advantage is not statistically significant (0.80, or between -0.97 and 2.58 using a 95% confidence interval) in this model.

2.8 Conclusion

In this paper, I have sought to contribute to the debate about the causes of incumbency advantage. By systematically testing four theoretical propositions that highlight different aspects of the partisan dealignment theory, I provide robust statistical evidence that corroborates that shifts in the mass electoral behavior impacts the incumbency advantage. The support for hypothesis about polarization indicates that congressmen and political parties' behavior impacts the incumbency advantage. On the one hand, the incumbency advantage tends to grow when the ideological differences between the Republican and Democratic parties decrease in the House of Representatives. The electoral advantage of the incumbent also tends to grow when the percentage of independents in the electorate increases. On the other hand, the incumbency advantage tends to decline as the percentage of party loyalty and percentage of challengers' partisan defecting to incumbents increase.

To my knowledge, this is the first attempt to evaluate these hypotheses using incumbency advantage as the dependent variable and estimating dynamic models that take into account the temporal variation of the incumbency advantage. While the literature has given a lot of attention to how to measure such phenomenon correctly (Gelman and King 1990; Cox and Katz 1996; Lee 2008; Gelman and Huang 2008; Caughey and Sekhon 2011; Erikson and Titiunik 2015), there have been limited tests of the impact of dealignment on incumbency advantage.

Understanding the causes of incumbency advantage sheds light on important features of how representative government works (Krehbiel and Wright 1983). Although initially scholars were interested in investigating the size and the trend of the incumbency advantage (Gelman and King 1990; Katz and King 1999*b*), recent developments in the field have neglected the temporal variation in this phenomenon (Lee 2008).

In this study, I have focused specifically in analyzing the causes of incumbency advantage in the U.S. House of Representatives. There are at least three avenues that future research can further investigate. First, one possibility for future research is to investigate what drives shifts in the incumbency effects in other countries. This may be particularly interesting to analyze in multiparty systems (e.g., U.K) or in places in which studies have detected negative incumbency effects (e.g., Brazil and India). Second, future research may focus on testing other determinants that have also been argued to be contributing factors. For instance, the impact of redistricting or campaign finance on incumbency advantage. Third, in this paper, I use the Gelman and King's (1990) and Erikson and Titiunik's (2015) measures to evaluate the four hypotheses. Future research may further explore other incumbency advantage measures to verify if the results are robust to additional measures of the dependent variable.

3 GENDER AND ACCOUNTABILITY DURING THE COVID-19 PANDEMIC

3.1 Introduction

Scholars have long been interested in analyzing whether there is a gender gap in presidential approval (Clarke et al. 2005) and vote choice (Chaney, Alvarez and Nagler 1998; Kam 2009). Specifically, they examine whether women and men hold differential economic outlooks and if that translates into how each group evaluates the incumbent. Different reasons have been offered about why such a gender gap might exist. Some scholars argue that men and women hold distinct economic preferences, and this leads to different evaluations of the incumbent's job performance (Welch and Hibbing 1992; Chaney, Alvarez and Nagler 1998; Clarke et al. 2005). In contrast, others contend that such differences in preferences between men and women do not lead to a distinct assessment of the incumbent's job (Kam 2009; Higgins and Kellstedt 2016). While these studies have contributed substantially to our understanding of the gender differences in presidential approval and voting behavior, they have focused their analyses mainly on how economic factors drive executive accountability. However, voters may hold incumbents accountable for policies in other areas as well. For instance, in the wake of the COVID-19 pandemic, voters may pay close attention to their government's actions to control the outbreak and how successful their efforts are in reducing the number of victims. Ultimately, the number of lives lost to the disease is a tangible measure of how (un)successful government efforts to control the outbreak have been (Lipscy 2020).

Since the World Health Organization declared the COVID-19 pandemic, governments worldwide have adopted policies to curb the spread of the virus and mitigate the pressure on their healthcare systems. However, these policies impose significant limitations on people's lives. Researchers have found that women and men differ in assessing the government's policies (Algara et al. 2021; Galasso et al. 2020) and their adherence to these measures (Galasso et al. 2020; Carreras, Vera and Visconti 2021; Palmer and Peterson 2020; Cassino and Besen-Cassino 2020). Given these gender differences in behavior and attitude regarding the COVID-19 pandemic, all else constant, I expect that, as the COVID-19 death toll increases, both women and men will punish the incumbent. However, I expect that women will punish the incumbent to a greater degree than their male counterparts. I empirically test these theoretical expectations using the case of two of the world's largest democracies – Brazil and the United States. These two nations are compelling cases to study, as they have been two of the countries most hard hit by the pandemic and are still striving to control the outbreak even after the vaccination roll-out began in late 2020 in the case of the U.S. and early 2021 in the case of Brazil. On September 2, 2021, the United States had the first-highest number of cumulative cases (39,449,332) and deaths (642,578), and Brazil had the third-highest number of cumulative cases (20,804,215) and the second-highest number of deaths (581,150) (Dong, Du and Gardner 2021).

To test my theoretical expectations, I conducted a separate empirical analysis for each country. For the Brazilian case, I employ newly assembled monthly time-series data on approval by gender for President Jair Bolsonaro across 31 monthly observations from January 2019 to August 2021. For the United States case, I use monthly time-series data on approval by gender for President Trump from the beginning of his mandate up until the end of his term for a total of 46 monthly observations from February 2017 to November 2020. Using time series analysis, the results for both countries suggest that men and women punish the president for increases in deaths due to COVID-19. However, I do not find evidence that the effect of worsening pandemic performance on presidential approval is more sensitive to changes in the death toll for women. This study contributes to the literature by providing empirical evidence on the degree to which there are differences in how women and men hold the incumbent accountable during extraordinary times.

The rest of the chapter proceeds as follows. In the next section, I briefly review the literature on the gender gap and presidential approval. I present evidence showing differences between men and women in the context of the COVID-19 pandemic. After discussing the data sources and research design, I present the results from analyses for each country. Finally, I conclude the chapter by discussing aspects that future research in the area can further develop.

3.2 Previous Literature on the Gender Gap in Incumbent's Evaluation

The impacts of the gender gap on political behavior and voting patterns have long been analyzed by scholars (Shapiro and Mahajan 1986; Kaufmann and Petrocik

1999; Schlesinger and Heldman 2001; Box-Steffensmeier, De Boef and Lin 2004). Indeed, previous research claims that the female vote has been influential to the dynamics of U.S. presidential elections since the 1980s (Carroll 2006). While a solid body of research suggests that female and male voters tend to reward (punish) incumbents when the economy is performing well (poorly) (Nannestad and Paldam 1994; Lewis-Beck and Stegmaier 2000), a particularly relevant question raised by the literature is whether women and men have different perceptions of the health of the economy and if that has an impact on their voting choices (Welch and Hibbing 1992; Chaney, Alvarez and Nagler 1998; Kam 2009) and presidential approval ratings (Clarke et al. 2005; Higgins and Kellstedt 2016).

Scholars generally view a gender gap as an effect that translates into a difference in the overall mean between men and women. As May and Stephenson (1994) explain: "it has long been understood that discernable differences exist between women and men on issues, party identification, and candidate selection (533)." All else equal, women (men) vote for Democrats (Republicans) more (Edlund and Pande 2002). Similarly, women hold a more pessimistic view of the economy than their male counterparts (Chaney, Alvarez and Nagler 1998). However, an even more fundamental question is whether women and men differ in the ways in which they translate performance assessments into voting choices and approval ratings and whether those performance assessments are also gender-dependent. Scholars have distinct views on whether the gender gap goes beyond a difference in *levels*.

On the one hand, scholars argue that different economic preferences of men and women translate into different evaluations of the incumbent government. According to these studies, male and female voters rely on different aspects when evaluating the economy's performance. For instance, Welch and Hibbing (1992) and Chaney, Alvarez and Nagler (1998) find that women tend to cast their votes based on sociotropic economic evaluations, whereas men on egotropic economic evaluations. Similarly, Clarke et al. (2005) present evidence that men and women base their assessment on distinct factors regarding the state of the economy. Analyzing U.S. presidential approval data from 1978 and 1997, Clarke et al. (2005) show that both women and men are future-oriented when they evaluate the president's job performance. That is, they tend to base their evaluation of the president's job performance on their prospective assessment of the economy. However, women tie their approval rating to their perceptions of the country's future economic situation, whereas men tie their evaluation to their perceptions of their future personal economic circumstances.

On the other hand, another group of studies claims that – despite the distinct economic perceptions among male and female voters – there are more similarities than differences in how both genders evaluate the incumbent (Kam 2009; Higgins and Kellstedt 2016). More recently, Higgins and Kellstedt (2016) extended the analysis advanced by Clarke et al. (2005) to evaluate if there is persistence in the relationship once analyses are extended to include President Obama's administration. Their findings suggest that "these differences between the economic experiences and perceptions of men and women, real as they are, do not lead to massively different causal processes in the formation of presidential approval (Higgins and Kellstedt 2016, p. 19-20)." Similarly, Kam (2009) also reports evidence of similarities in the way that men and women cast their votes. Analyzing data from the U.S. presidential elections from 1980 to 2004, Kam (2009) finds that male and female voters both vote sociotropically. Thus, she concludes that gender differences in voting patterns are exceptional.

In sum, the empirical evidence about the impact of gender differences on incumbent's evaluation (vote choice and presidential approval rating) is mixed. It might also be the case that the gender gap, as well as its influence over the incumbent's evaluation, fluctuates over time. In some periods, the gap may be more marked, whereas there might be reduced differences in other times. As Kellstedt, Peterson and Ramirez (2010) highlight, "given that the policy-opinion gap is dynamic, the gender gap in voting should be more prevalent in some elections than in others (19)."

Despite the contribution of these studies to our understanding of the gender differences in presidential approval and voting behavior, they have restricted their analyses to examine whether men and women hold distinct economic outlooks and the impact of such differences on the approval rating or vote choice. Although voters tend to hold incumbents accountable for a country's economic performance during their mandate, they may also weigh these gains (or losses) against performances in other policy areas. For instance, Gilens (1988) finds that the difference among men and women in support for President Reagan was explained mainly by a gender gap in views related to the military and social welfare issues. Examining President Clinton's approval rating, Mattei (2000) reports that attitudes toward the role of the government were the most relevant factor for the gender gap.

Indeed, some policy areas might get special attention from the public and become more salient in specific moments. Given the salience of the COVID-19 pandemic, one might expect voters to pay close attention to their government's actions to control the outbreak and how successful the government's efforts are in reducing the number of victims. Ultimately, the loss of human lives implies a considerable humanitarian cost to society. As survey data from April, May, and July 2020 show, Americans considered the coronavirus or COVID-19 to be the nation's top problem and the government/poor leadership the second most crucial problem for the country (*Gallup* 2020*a*). During the pandemic, the number of lives lost to the disease is a tangible measure of how (un)successful government's efforts to control the outbreak (Lipscy 2020). While most of the literature on gender differences and approval has examined the U.S. case solely, it is essential to evaluate whether the gender gap impacts incumbent evaluations in other democratic countries.

I seek to advance our knowledge in this area by examining whether women and men hold the incumbent accountable for the COVID-19 pandemic deaths toll. In the next section, I will briefly review some of the differences between women and men in the context of the COVID-19 pandemic.

3.3 The Gender Gap in COVID-19 Pandemic Perceptions and Behaviors

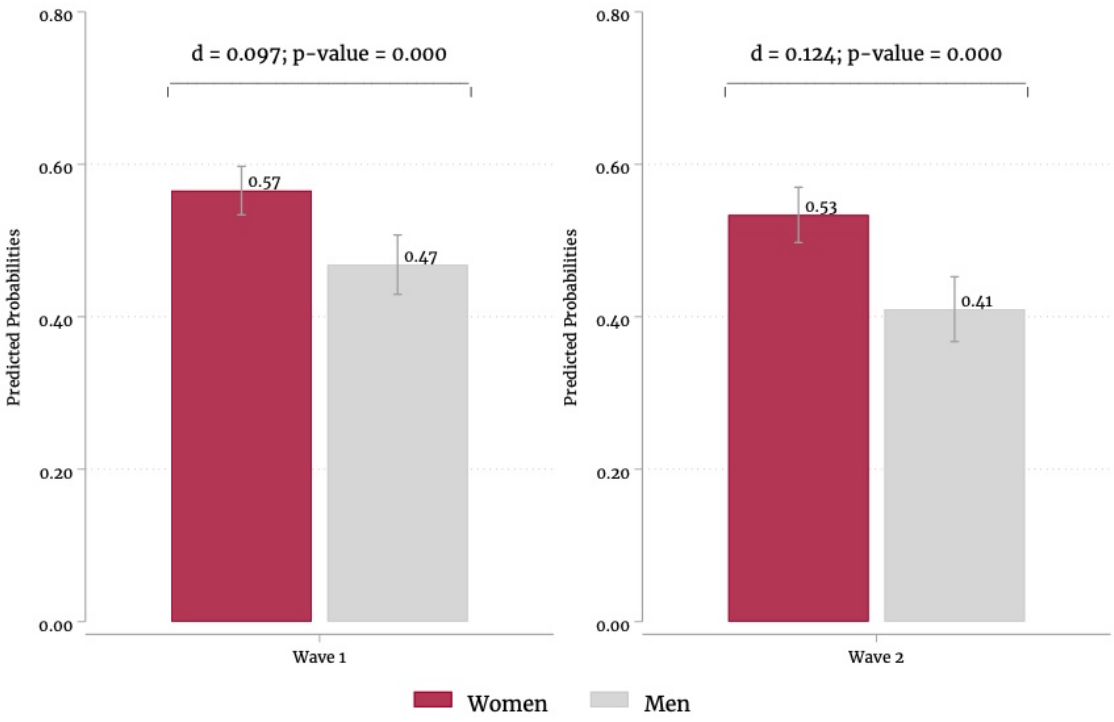
The COVID-19 pandemic has entailed not only a health crisis but also a social and economic crisis that has impacted women and men differently. Evidence shows that since the beginning of the pandemic, women are likely to disproportionately bear the cost of unpaid care and domestic work (Johnston, Mohammed and Van Der Linden 2020; *UN Women* 2020). For instance, a study conducted in five countries (United States, Canada, Denmark, Brazil, and Spain) demonstrates that women are more likely to spend more time on childcare and household chores than their male counterparts during the pandemic (Giurge, Whillans and Yemiscigil 2021). There is also evidence showing that female and male workers have been unequally affected by the pandemic recession, as the female employment rate experienced a more considerable decline in most countries (Alon et al. 2021; Kugler et al. 2021).

Interestingly, women and men hold distinct perceptions regarding the seriousness of the pandemic. Across time, the percentage of women that say they are worried about getting infected is substantively higher than the percentage of males, as shows data from the U.S. (*Gallup* 2020*b*). According to Brooks and Saad (2020), the gender gap in the concern about contracting the disease persists even after controlling for socioeconomic and political factors. A cross-national survey study in eight countries (Australia, Austria, France, Germany, Italy, New Zealand, the United Kingdom, and the United States) demonstrates that women are more likely to assess the COVID-19 pandemic as a severe health crisis (Galasso et al. 2020). Using replication data from Galasso et al. (2020), I show in Figure 13 the predicted probabilities for the U.S. male and female citizens in assessing whether the COVID-19 pandemic has serious health consequences. After controlling for sociodemographic, psychological, and behavioral factors, women in the U.S. are more likely to say they consider COVID-19 as having severe health issues than their male counterparts. In the first wave of the survey conducted by Galasso et al. (2020) (from March 16 to March 30, 2020), the predicted probability of women who consider the COVID-19 as having serious health consequences was 56.55% (p-value = 0.000, 95% CI = 53.37%, 59.73%), whereas it was only 46.84% (p-value = 0.000, 95% CI = 42.94%, 50.73%) for men, holding the control variables at their observed values. On average, the gender difference (*d* = women - men) is 9.7% (p-value = 0.000). Such a pattern persists over time. In the second wave of the study (from April 15 to April 20, 2020), the predicted probability for women to say they consider the COVID-19 as having serious health consequences was 53.36% (p-value = 0.000, 95% CI = 49.74%, 56.99%), whereas it was only 40.99% (p-value = 0.000, 95% CI = 36.74%, 45.25%) for men. Interestingly, from waves 1 to 2, the gap slightly increased from 9.72% (95% CI = 5.24%, 14.18%) to 12.37% (95% CI = 7.18%, 17.56%).[1]

The evidence further suggests that the gender gap in assessing the severity of the COVID-19 is also substantial in Brazil. Using survey data from Petherick et al. (2020), Figure 14 depicts the predicted probabilities for Brazilian male and female citizens in assessing the severity of COVID-19 symptoms as compared to the normal flu for the majority of infected individuals. After controlling for age, schooling, income, and state

[1] Although the second wave had been followed by the first one after only a month, the pandemic situation had deteriorated substantially during this period (between March and April 2020) in the United States. By April 29, 2020, the country had registered more than 1 million confirmed cases, and the death toll was greater than U.S. fatalities in Vietnam War (*CNN* 2020).

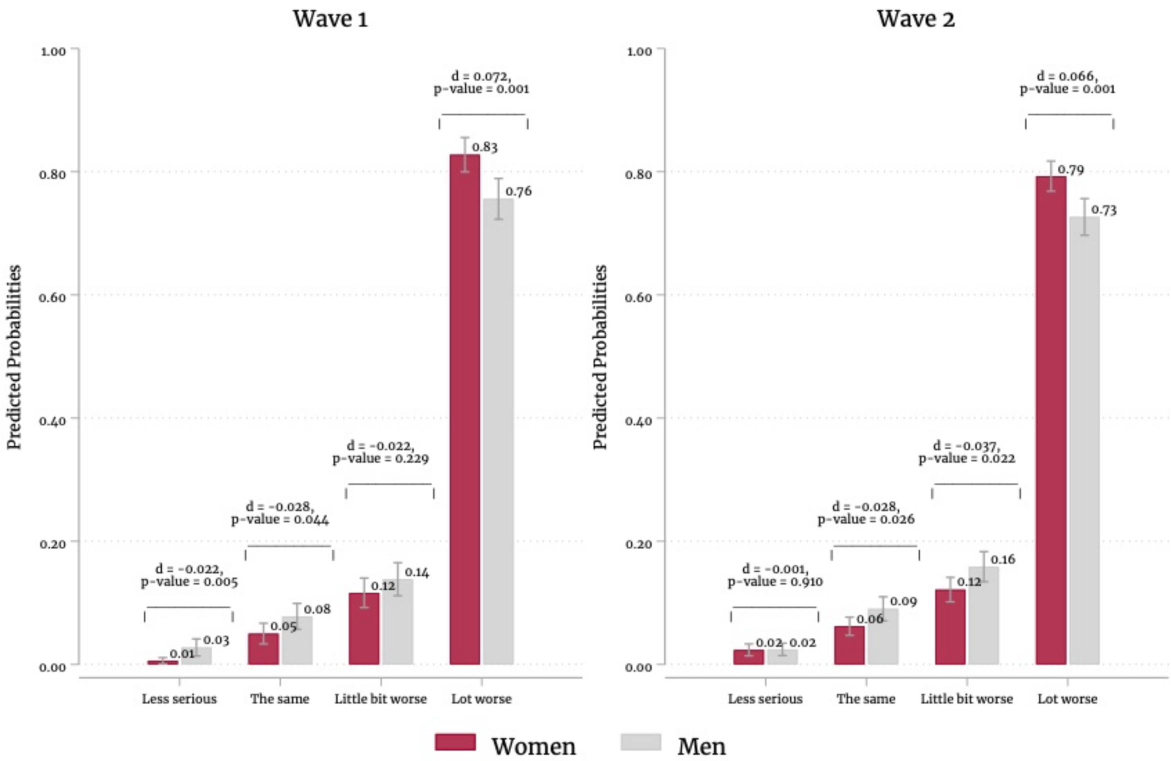
Figure 13 – Predicted Probabilities of Evaluating the Severity of COVID-19 Symptoms by Gender, U.S. citizens, March and April 2020



Note: The predicted probabilities and corresponding 95% CIs were calculated based on the results of an OLS regression model regressing the dependent variable (1 if the respondent said COVID-19 as having serious health consequences, and 0 otherwise) on the key independent variables (female) and control variables (sociodemographic, psychological, and behavioral factors). The predicted probabilities for women and men were calculated holding the independent variables at their observed values. Data source: Galasso et al. (2020).

of residence, Brazilian women are more likely to say that they consider SARS-CoV-2 a lot worse than normal flu than their male counterparts. In the first wave of the survey conducted by Petherick et al. (2020) (May 2020), the predicted probability for women to consider COVID-19 a lot worse than normal flu was 82.77% (p-value = 0.000, 95% CI = 79.99%, 85.52%), whereas for men it was only 75.60% (p-value = 0.000, 95% CI = 72.29%, 78.91%), holding the control variables at their observed values. Such a pattern persists over time. In the second wave (November 2020), 79.28% (p-value = 0.000, 95% CI = 76.84%, 81.72%) of females as compared to 72.66% (p-value = 0.000, 95% CI = 69.68%, 75.64%) of men consider COVID-19 much worse than the flu. In contrast, men are more likely than women to say that the COVID-19 symptoms are less serious, the same, or a little bit worse than normal flu. In both waves, the gender gap in the probability of assessing COVID-19 symptoms as a lot worse than the normal flu has remained around 7%.

Figure 14 – Predicted Probabilities of Evaluating the Severity of COVID-19 Symptoms Compared to the Normal Flu by Gender, Brazilian citizens, May and November 2020



Note: The predicted probabilities and corresponding 95% CIs were calculated based on the results of a multinomial logit regression model. The dependent variable is a categorical variable with four categories which are the responses to the question of "How severe do you think the symptoms of coronavirus are for the majority of people who get it?." Respondents had to choose one of the four following answers: (i) it is less severe than normal flu; (ii) it is about the same as normal flu; (iii) it is a little bit worse than normal flu; and (iv) it is a lot worse than normal flu. The predicted probabilities for women and men were calculated holding the independent variables (age, schooling, income, and state of residence) at their observed values. Data source: Petherick et al. (2020).

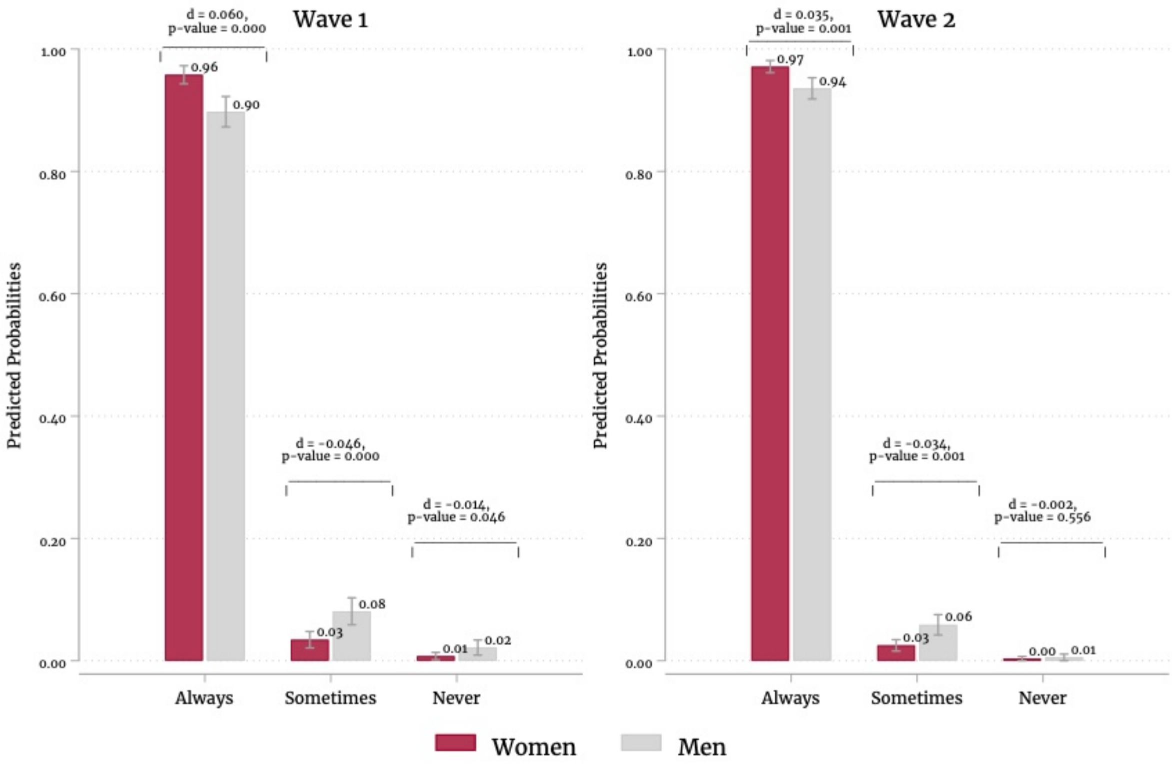
Men and women also differ in their self-reported adherence to social distancing. Governments worldwide adopted social distancing policies to reduce the number of infections and victims caused by the disease (Hsiang et al. 2020). Evidence from survey studies demonstrates that women are more likely than men to endorse (Carreras, Vera and Visconti 2021) and adhere to social distancing policies (Galasso et al. 2020). Among the social distancing measures, the use of masks stands out as one of the most effective policies to curb the virus's spread. Evidence from the U.S. highlights that women are more likely to report using masks compared to their male counterparts (Cassino and Besen-Cassino 2020). Data from Brazil also suggest that there are gender differences in the use of masks. Using survey data from Petherick et al. (2020), Figure 15 depicts the predicted probabilities of mask use for the Brazilian male and female citizens. After controlling for age, income, and state of residence, women are more likely to say they wear a mask when they leave home than their male counterparts. In the first wave of the survey, the predicted probability for women to say they wear a mask when going out was 96% (p-value = 0.000, 95% CI = 94.32%, 97.28%), whereas for men it was 89.76% (p-value = 0.000, 95% CI = 87.27%, 92.24%), holding the independent variables at their observed values. In contrast, men are more likely than women to report wearing a mask sometimes or never when they leave home. The gender difference in reporting always wearing a mask is 6%. Although the difference decreased from the first to the second wave of the study, there is still a substantively and statistically significant difference between the predictions for women and men of 3.5% in the second wave.[2]

In summary, survey evidence from Brazil and the United States suggests that women and men hold distinct views regarding the COVID-19 pandemic. Women are more likely than men to assess the severity of the virus and its effects as much worse than a flu. Most importantly, these differences reflect on their attitude and behavior to prevent the disease and their behavior with respect to social distancing. Women are more likely to obey public health recommendations, especially regarding the use of masks outside the home. Studies also highlight that, in the context of the COVID-19 pandemic, women are more likely to base their attitudes and activities on scientific information (Algara, Fuller and Hare 2020; Algara et al. 2021). This is particularly intriguing given the worldwide evidence suggesting that male patients are more likely to require intensive treatment unit admission and more likely to die from the disease than female patients (Peckham et al. 2020; *Global Health 50/50* 2021).

Given these substantial gender differences, I expect women to be less tolerant of increases in the number of victims and to hold the incumbent responsible for the worsening in the COVID-19 death toll to a greater degree than their male counterparts. Particularly, I expect that women will punish the incumbent president to a greater degree than men for increases in the number of COVID-19 fatalities.

[2] While one can debate whether or not a 3.5% difference for a single incident is substantively significant, this difference is certainly substantively significant when we consider such effects cumulating across weeks and months.

Figure 15 – Predicted Probabilities of Wearing a Mask When Leaving Home by Gender, Brazilian citizens, May and November 2020



Note: The predicted probabilities and corresponding 95% CIs were calculated based on the results of a multinomial logit regression model. The dependent variable is a categorical variable with three categories: the responses to the question of ”Do you wear a mask when you go out?.” Respondents had to choose one of the three following answers: (i) always, (ii) sometimes, and (iii) never. The predicted probabilities for women and men were calculated, holding the independent variables (age, income, and state of residence) at their observed values. Data source: Petherick et al. (2020).

3.4 Data and Methods

To test my theoretical expectations, I employ data from Brazil and the United States. These two countries are relevant cases to study for several reasons. First, both countries have been severely affected by the pandemic and accumulated the most significant number of cases and victims. Given the catastrophic outcome of the pandemic in those places, voters should be expected to hold the incumbents accountable. Second, when the pandemic hit the world, Brazil and the United States were governed by populist right-wing presidents that opposed implementing strong nationwide social distancing policies to curb the spread of the virus, and, in the case of Brazil, even questioned the efficacy of mass vaccinations. As a result, most policies adopted in these countries to fight the pandemic were implemented mainly by sub-national governments

(Barberia et al. 2021; Adolph et al. 2021; Bennouna et al. 2021). Third, President Donald Trump (United States) and President Jair Bolsonaro (Brazil) were both elected in very polarized elections and such polarization extended throughout their mandates. As such, many issues concerning COVID-19 became polarized (e.g., the adoption of social distancing, the use of masks, vaccines, etc.). Beyond these commonalities, both presidents were markedly unsuccessful in obtaining the support of female voters, and such a pattern has persisted across the duration of their mandates (*Pew Research* 2016; *El País Brasil* 2018).[3]

3.4.1 The Brazilian case

To evaluate the impact of COVID-19 deaths on presidential approval by gender, I collected data from January 2019 – the first month of President Bolsonaro's mandate – up to August 2021. This longer time span allows me to assess the trends in presidential approval before and during the pandemic.[4]

The dependent variable employed in the analysis is a monthly measure of presidential approval for female and male adults aged 18 and older. The data were collected from almost 20 polling firms. This is measured as the percentage of respondents in the surveys that evaluated the performance of President Bolsonaro as "Very Good" or "Good" to the question, "Do you evaluate the performance of the current president as Very Good, Good, Regular, Bad or Very Bad?". These are the only data publicly available for the period that disaggregate evaluations of the president's performance by the gender of the respondents. The number of surveys available by month varies substantially during the period of analysis. In general, there is at least one survey per month. However, in a few cases, more than one survey was conducted in the same month. In those cases, I calculated the average by gender of the percentage of respondents in the surveys that evaluated the president's performance as "Very Good" or "Good" across the surveys available for that specific month. Unfortunately, no surveys

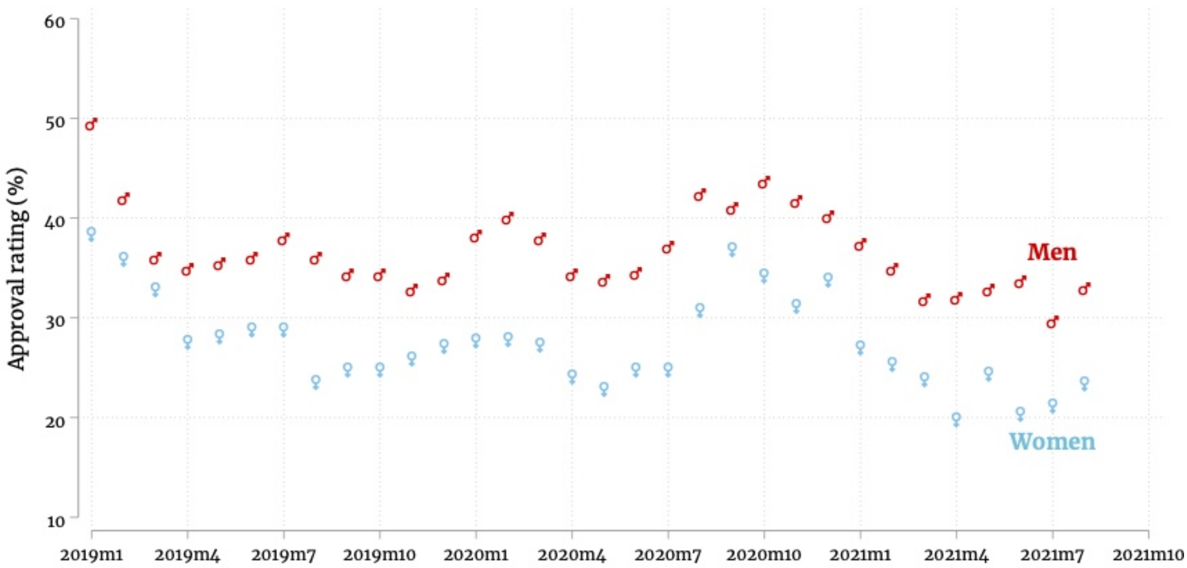
[3] A plausible rival argument that I will not explore in this chapter is that, since the electorates were so polarized in Brazil and in the U.S. and much of this was visible as gender differences, one could argue that we might expect the mounting deaths to have no effect. That is, people who already approved and disapproved of each president were unlikely to be persuaded regardless of the rising body count. A fruitful avenue of future research on this area is in evaluating whether the findings of this chapter hold in less polarized environments.

[4] Brazil reported the first confirmed case by the novel coronavirus in the country on February 26, 2020, and the first confirmed death by the virus on March 17, 2020.

with the needed breakdown were available in three months (May 2019, September 2019, and October 2019). For those cases, I interpolated the data using information from one period before and one period after the missing data.

As Figure 16 confirms, there has been a marked gender gap in the approval rating of President Bolsonaro since the beginning of his mandate. Fewer women consistently evaluate his job as "Very Good" or "Good" as compared to men. President Bolsonaro was elected in the 2018 elections. During the electoral campaign, the women's movement led a campaign against him with the emblem *#elenão* (in English, *#nothim*). The campaign was mainly motivated against Bolsonaro's misogynistic statements. Days before the 2018 elections, demonstrations against Bolsonaro led by women took place in many cities across the country (*BBC News Brasil* 2018). Evidence from national representative surveys shows that Bolsonaro was substantively rejected by females more than by male voters (50% among women against 39% among men) (*Datafolha* 2018). In sum, the difference between women's and men's support for Bolsonaro preceded his election and mandate.

Figure 16 – Presidential Approval of Jair Bolsonaro for Women and Men (%)



To evaluate the impact of the pandemic on the president's job performance by gender, I employ monthly national data on COVID-19 deaths and cases. The data source is the COVID-19 Dashboard from Johns Hopkins University (Dong, Du and Gardner 2021), which compiles COVID-19 indicators based on official government releases. While the COVID-19 pandemic has hit most countries worldwide, its effects have been distributed unevenly across countries and regions. Brazil is one of the hardest hit countries in the world. On September 2, 2021, Brazil had the third-highest number of cumulative cases (20,804,215) and the second-highest number of deaths (581,150) (Dong, Du and Gardner 2021).

As it is well-known that the state of the economy is a relevant predictor of variations in presidential approval, I also include a monthly measure for the inflation rate. I use the *Índice Nacional de Preços ao Consumidor Amplo* (IPCA), which measures the variation in the average cost of living for households with a monthly income from 1 to 40 minimum wages. To capture the effect of changes in unemployment, I employ monthly data on the stock of formal labour jobs for women and men. This measure captures the size of the formal market workforce. These data were collected from the *Cadastro Geral de Empregados e Desempregados* (CAGED) of the Ministry of Economics. Although this measure does not encompass the informal labour market, it still provides an informative picture of the losses and gains on the available jobs in the formal market over the period.

Finally, I employ data on the monthly expenditures (*despesas pagas*) of the Emergency Aid Cash transfer program (*auxílio emergencial*) to capture how disbursements affect executive approval ratings. Data are aggregate transfers to all recipients from the *Tesouro Nacional* and in Brazilian reais. Among the fiscal policies implemented by the federal government to cope with the effect of the pandemic, the emergency aid program was the most important assistance provided to the vulnerable population. By February 9, 2021, the government had spent R$ 293.1 billion (US$ 53.6 billion) on this program. It has benefited more than 67.9 million Brazilians (Caixa 2021). Including the eligible people and their family members, the *auxílio emergencial* has benefited 56.1% of the population (*Ministério da Cidadania* 2020). In fact, scholars have evaluated that this program is the largest social protection program to alleviate the effects of the pandemic in Latin America (Lustig and Trasberg 2021). Initially, it consisted of a monthly emergency cash transfer of R$ 600 (US$ 110.56) per recipient for a total of 3 months. Single mothers received twice this amount, that is, a benefit of R$ 1,200 (US$ 221.12). The benefit was then extended for an additional 2 months of R$ 600/R$1,200, and then an additional 4 months of R$ 300/R$ 600, and more recently, in April 2021, an additional 4 months, but now with a reduced payment sum (R$ 150/R$ 375).[5] Since April 2020, the extensions in the program were followed by reductions in the monthly benefit amount and, in the case of the most recent extension, also new restrictions on access criteria.

The *auxílio emergencial* has been a powerful tool in alleviating hunger and in reducing poverty in the face of the pandemic recession (Duque 2020). As survey data show, the majority of those that received at least one payment of the benefit stated that they used the money for purchasing food (*Datafolha* 2020). Data from the same survey report that, among those that received at least one payment of the benefit, 49% of women stated that this source was the sole income for the household. In contrast, only 36% of the men indicated the same. According to Fares et al. (2021), female-headed households are the ones that most lost income during the pandemic and those that most benefit from the emergency aid program. The emergency aid has been especially relevant for complementing the family income of black female-headed households, which have been deeply affected by the pandemic (Fares et al. 2021). In Brazil, women are the majority of workers in the service sector, and this sector is one of the most severely hit by the pandemic (*Agência Câmara de Notícias* 2020; *Folha* 2021). In addition, as governments worldwide have adopted on-site school closures policies to contain the spread of the virus, the extra burden of taking care of children and their education when schools were closed has rested primarily on women's shoulders. In sum, the *auxílio emergencial* was a powerful mechanism for alleviating hunger, especially for women, in the midst of the economic and health crisis. For these reasons, I expect that women will reward the incumbent president to a greater degree than men for increases in the monthly expenditures with the *auxílio emergencial* program.

Modeling Strategy

For the dependent and explanatory variables, I conducted a series of unit root tests to check the order of integration of each series. For each variable, I performed six unit root tests (Augmented Dickey-Fuller, Augmented Dickey-Fuller (with trend),

[5] The value of the benefit varies according to the household's composition. For instance, in the extension that occurred in April 2021, if the household consists of only one person, the benefit was of R$ 150 per month. However, if the family consists of more than one person, the benefit was of R$ 250 per month. When the family is headed by a woman without a spouse or partner, with at least one person under the age of eighteen, the recipient receives R$ 375 monthly.

Phillips-Perron (with trend), Dickey-Fuller GLS, Elliott-Rothenberg-Stock, and Kwiatkowski-Phillips-Schmidt-Shin (KPSS)). For all tests, the null hypothesis is that the series follows a unit root process. The only exception is the KPSS test, in which the null hypothesis is that the series follows a trend stationary process. The results of the tests with the variables in levels and the first difference are reported in the appendix. All variables are nonstationary in levels. However, they are stationary in the first difference. Given the evidence of nonstationarity in the dependent and independent variables, I then verify whether the women's and men's presidential approval time series are cointegrated with the explanatory variables. To do that, I follow the procedures outlined by Philips (2018) and apply the bounds test in the residuals of the error correction model specified in Equation 5, for men's and women's presidential approval model estimated separately. Because of limitations due to the small sample size features of the data, I present the most parsimonious model possible, which allows me to test my argument. In the appendix, I also include the results for the model specification, including inflation and stock of the formal labor market jobs as control variables.

$$\begin{aligned} \Delta Approval_t = \beta_0 + \alpha_1 Approval_{t-1} + \beta_1 \Delta Deaths_t + \beta_2 Deaths_{t-1} + \\ \beta_3 \Delta Emergency\ Aid_t + \beta_4 Emergency\ Aid_{t-1} + \varepsilon_t \end{aligned} \tag{5}$$

For the model estimated with *presidential approval for men* as the dependent variable and the explanatory variables as described by Eq. 5, the 5% critical values for the F-test with 31 observations and two regressors are 4.183 for the lower bound and 5.333 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the estimated F-statistics is 8.14, there is statistical evidence of cointegration. The analysis of the t-test also corroborates such conclusion.[6] For the estimated model with *presidential approval for women* as the dependent variable, the 5% critical values for the F-test with 31 observations and two regressors are 4.183 for the lower bound and 5.333 for the upper bound. As the estimated F-statistics is 6.140, there is also statistical evidence of cointegration.[7] In the appendix, I also report the results for the model estimated using COVID-19 monthly cases instead of deaths.

6 The 5% critical values for the t-test are -2.86 for the lower bound and -3.53 for the upper bound, as the t-statistic is -4.47, I can conclude that there is evidence of cointegration.

7 The analysis of the t-test corroborates such conclusion at a 10% level. The 10% critical values for the t-test are -2.57 for the lower bound and -3.21 for the upper bound, whereas the t-statistic is -3.46.

To analyze the impact of deaths due to COVID-19 on approval rating for women and men, I follow Clarke et al. (2005) and estimate a separate model for men and women. I estimate the two models simultaneously using seemingly unrelated regression (SUR) procedures for accounting for the cross-equation error correlation (Greene 2008).[8]

3.4.2 The U.S. case

To evaluate the effect of COVID-19 deaths on the U.S. presidential approval by gender, I use monthly data from Gallup for President Trump's mandate from 2017 to 2020. The dependent variable employed in the analysis is a monthly measure of the percentage of respondents in the surveys that "approve" how President Trump has handled his job as president.

Figure 17 – Presidential Approval of Donald Trump for Women and Men (%)

[8] The equations are linked only by the disturbances. Suppose the disturbances across equations are not correlated. In that case, there is no efficiency gain in estimating SUR, and the results obtained estimating the full set of equations will be similar to estimating equation-by-equation using standard ordinary least squares (OLS). However, if the disturbances across equations are correlated, there is efficiency gain by estimating SUR instead of equation-by-equation using OLS. In the models reported in this section on Brazilian presidential approval by gender, each equation does not contain precisely the same set of regressors, as the lag dependent variable depends on which equation (men or female) is being analyzed. In this case, Greene (2008) says that "with unrestricted correlation of the disturbances and different regressors in the equations, the results are complicated and dependent on the data." He then states that (i) the greater the disturbance correlation across equation, the greater the gain in efficiency in estimating SUR, and (ii) the efficiency gain will be larger as there is less correlation between the **X** matrices across the equations (Greene 2008).

As previous studies have shown (Clarke et al. 2020), there is a substantial gender gap during President Trump Administration, with women consistently approving the president's job performance less than men. Figure 17 confirms such a pattern. To evaluate the effect of the death toll due to the COVID-19 pandemic on approval, I use monthly nationally aggregated data on reported deaths and cases. The data source is the COVID-19 Dashboard from Johns Hopkins University (Dong, Du and Gardner 2021). Thus far, the United States reported the highest number of cumulative cases and deaths in the world. To account for the effect of the economy on approval, I use a separate monthly measure for women's and men's unemployment rates (*U.S. Bureau of Labor Statistics* 2021). I also include the U.S. real disposable personal income (seasonally adjusted), consumer price index (*U.S. Bureau of Labor Statistics* 2021), and index of consumer sentiment (*University of Michigan* 2021) as control variables.

Modeling Strategy

For the dependent and independent variables, I conducted a series of unit root tests to check the order of integration of each series. The results of the tests with the variables in levels and the first difference are reported in the appendix. All variables are nonstationary in levels, except for men's presidential approval rating. However, they are all stationary in the first difference. To verify whether the dependent and independent variables are cointegrated, I follow the procedures outlined by Philips (2018) and apply the bounds test in the residuals of the error correction model specified in Equation 6, for men's and women's presidential approval model estimated separately. Because of limitations due to the small sample size features of the data, I present the most parsimonious model possible, which allows me to test my argument. In the appendix, I also include the results for the model specification, including consumer price index, real disposable personal income per capita, and consumer sentiment index as control variables.

$$\begin{aligned} \Delta Approval_t = \beta_0 + \alpha_1 Approval_{t-1} + \beta_1 \Delta Deaths_t + \beta_2 Deaths_{t-1} + \\ \beta_3 \Delta Unemployment_t + \beta_4 Unemployment_{t-1} + \varepsilon_t \end{aligned} \tag{6}$$

For the model estimated with *presidential approval for men* as the dependent variable and the explanatory variables as described by Eq. 6, the 5% critical values for the F-test with 46 observations and two regressors are 4.070 for the lower bound and 5.190 for the upper bound, values for case III as reported by Pesaran, Shin and Smith (2001). As the estimated F-statistics is 11.29, there is statistical evidence of cointegration. The analysis of the t-test also corroborates such conclusion.[9] For the estimated model with *presidential approval for women* as the dependent variable, the 5% critical values for the F-test with 46 observations and two regressors are 4.07 for the lower bound and 5.19 for the upper bound. As the estimated F-statistics is 8.56, there is also statistical evidence of cointegration.[10] I also report the results for the model estimated using COVID-19 monthly cases instead of deaths in the appendix. Similar to the time series analysis for the Brazilian case, I estimate a separate model for men and women simultaneously using seemingly unrelated regression (SUR) procedures for accounting for the cross-equation error correlation.

3.5 Results

I have argued that, as the number of deaths due to COVID-19 increases, voters will be more likely to punish the president by evaluating his job performance as good or very good to a lesser degree. In other words, I expect to find a negative relationship between presidential approval and COVID-19 deaths. More importantly, I especially expect that women will tend to punish the president more than their male counterparts for increases in the death toll. To assess these hypotheses, I calculate and test the statistical significance of the short- and long-run effects of COVID-19 deaths on presidential approval for women and men in the U.S. and Brazil. The short-term effects are calculated using the estimated coefficient on the first-differenced variables. They indicate the immediate impact of a one-unit change in the number of reported COVID-19 fatalities on approval. In contrast, the long-term effects are calculated using the lagged dependent variable (i.e., the rate of adjustment parameter) and the lagged independent variable. The long-term effects show the permanent impact on the approval rating of a one-unit

[9] The 5% critical values for the t-test are -2.86 for the lower bound and -3.53 for the upper bound, as the t-statistic is -5.72, I can conclude that there is evidence of cointegration.

[10] The analysis of the t-test corroborates such conclusion. The 5% critical values for the t-test are -2.86 for the lower bound and -3.53 for the upper bound, whereas the t-statistic is -5.02.

change in the number of COVID-19 deaths. After computing the short- and long-term effects, I will evaluate whether the estimated of an equivalent change in deaths has a differential impact for how women assess the president as compared to men. In the sections that follow, I discuss each case separately.

3.5.1 The Brazilian case

Table 35 shows the results for the SUR models estimated using data for the Brazilian presidential approval.[11] The estimated immediate short-term impact of a one-unit change in the number of deaths due to COVID-19 on support for the president is negative and statistically significant – at the 0.05 significance level – only for women. All else equal, an increase of one standard deviation (105.77) in the reported number of COVID-19 deaths per million reduces the women's approval for President Bolsonaro by 2.13 percentage points (or a reduction between 0.34 and 3.92 using a 95% confidence interval). Although 105.77 deaths per million in a month are many human lives lost, the official record shows that Brazil has registered a number of victims much higher than this amount since January 2021. We cannot reject the null of no immediate short-term effect of deaths on men's presidential approval rating at the 0.05 significance level.

The estimated long-term impact of a one-unit increase in the number of COVID-19 deaths for the president is negative and statistically significant at the 0.05 level for both men and women. All else equal, an increase of one standard deviation (105.77) in the reported number of COVID-19 deaths per million leads to a long-run reduction in women's approval by 2.42 percentage points (or between 0.46 and 4.38 using a 95% confidence interval) and in men's approval for President Bolsonaro by 1.61 percentage point (or between 0.21 and 3.01 using a 95% confidence interval). To understand how these long-term effects should be interpreted, I present the results for a simulation of the impact of a shock of a one standard deviation increase in the death toll on presidential approval for women and for men at time 10 using stochastic simulation (Jordan and Philips 2018; Philips 2018).[12] Figure 18 depicts the expected values of the presidential

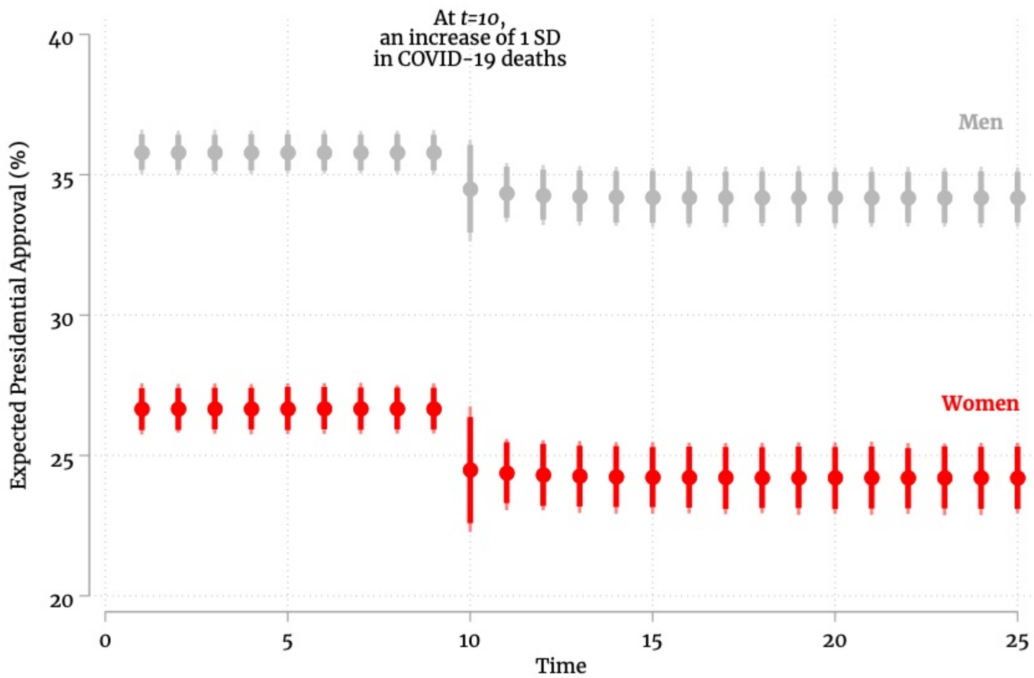
[11] The results reported in this section are robust regardless of whether we assess the impact of the pandemic on the president's support using COVID-19 deaths or cases. The results for the model estimated using COVID-19 cases are shown in Table 37 in the appendix.

[12] To calculate the dynamic simulations, I use the command *dynardl* that automatically implements the routine in Stata. As I was not able to run the dynamic simulations after estimating SUR model, Figure 18 shows the simulations based on the results after estimating OLS models.

approval rating for both men and women before and after the increase in deaths at time period 10. As the figure makes clear, the expected value of the approval rating decreases substantially after the shock. This provides support to our expectation about the negative impact of the COVID-19 deaths on presidential approval.

So far, the evidence suggests that Brazilian female and male voters punish the president in the long run for increases in the death toll. The results also suggest an immediate fall in the women's approval rating after increasing the number of COVID-19 deaths. However, I have argued that rises in the number of deaths would have a more considerable impact on presidential approval for women than for men. To evaluate this hypothesis, I calculate the difference between the estimated short- and long-term effects of deaths on approval by gender. The difference ($d = women - men$) is not statistically different from zero for the short- and long-term effects. Thus, we do not find evidence that corroborates the expectation that women punish the president for increases in the death toll at higher rates than their male counterparts in the case of Brazil.

Figure 18 – The Effect of COVID-19 deaths (per million) on Brazilian Presidential Approval for Women and Men



Note: The results depict the dynamic effects of a one standard deviation increase in deaths at time 10. Predicted values and 90 and 95 percent confidence intervals.

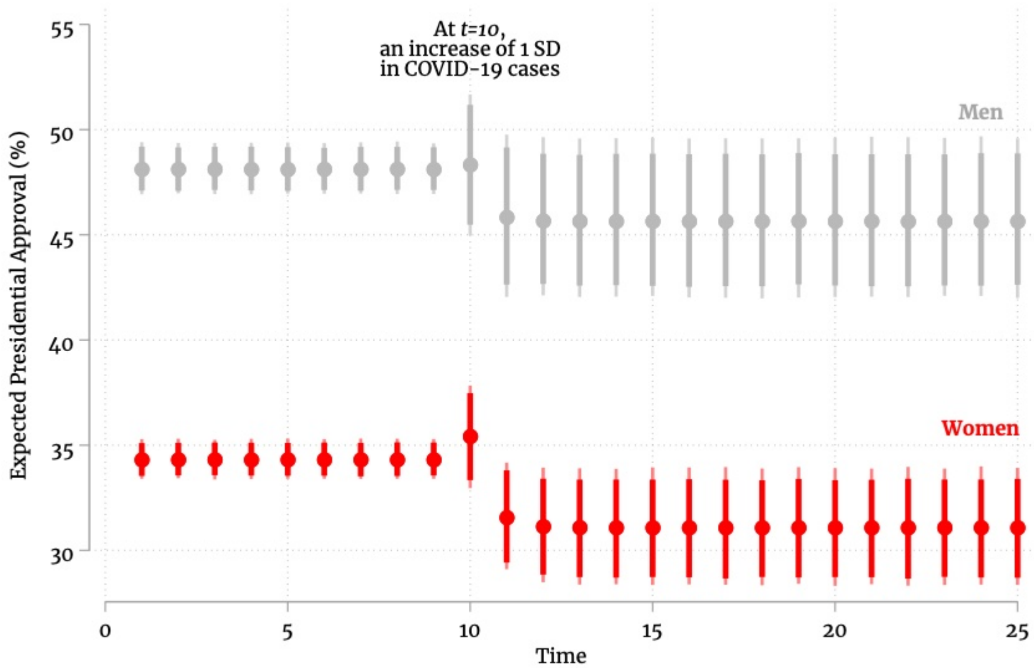
3.5.2 The U.S. case

The results for the SUR models estimated using data for the U.S presidential approval rating are shown in Table 48. The estimated immediate short-term impact of a one-unit increase in the number of deaths due to COVID-19 on support for the president is negative and statistically significant – at the 0.05 significance level – for both men's and women's approval. All else equal, an increase of one standard deviation (52.54) in the reported number of COVID-19 deaths per million (within a month) reduces women's approval for President Trump by 2.67 percentage points (or a reduction between 0.13 and 5.22 using a 95% confidence interval). The same shock in deaths per million would reduce men's approval for President Trump by 3.98 percentage points (or a decrease between 1.10 and 6.87 using a 95% confidence interval). Except for the initial months of the pandemic (January, February, and March 2020), the U.S. has registered more than 52.54 deaths per million during most of the year of 2020. If we consider the standard deviation on COVID-19 deaths for the period between January and December 2020, an increase of one standard deviation (72.35) in the reported number of COVID-19 deaths per million would reduce women's approval by 3.68 percentage points (or a reduction between 0.17 and 7.19 using a 95% confidence interval) and men's approval for President Trump by 5.49 percentage points (or a reduction between 1.51 and 9.46 using a 95% confidence interval).

We cannot reject the null of no long-term effect of deaths on men's and women's presidential approval rating at the 0.05 significance level. Thus, the evidence suggests that U.S. female and male citizens punish the president only in the short run for increases in the death toll. This finding persists even after including consumer price index, index of consumer sentiment, and disposable income per capita as control variables (see Table 50 in appendix). However, I have argued that rises in the number of COVID-19 deaths would have a larger effect on presidential approval for women than for men. I do not find evidence that supports this expectation since the difference between the short-term effects for women's and men's approval is not statistically different from zero in the U.S. case.

Table 49 displays the results for the SUR models estimated using COVID-19 cases per million instead of deaths. The estimated permanent long-term impact of a one-unit change in the number of COVID-19 cases on women's support for the president is negative and statistically significant at the 0.05 significance level and on men's support at the 0.10 significance level. All else equal, an increase of one standard deviation (3617.12) in the reported number of COVID-19 cases per million leads to a permanent reduction in the women's approval by 3.09 percentage points (or between 1 and 5.18 using a 95% confidence interval) and in the men's approval for President Trump by 2.38 (or between 0.23 and 4.53 using a 90% confidence interval). Using stochastic simulation, Figure 19 depicts the expected values of the presidential approval rating for both men and women before and after a shock at the period 10 of a one standard deviation increase in the number of cases on presidential approval (Jordan and Philips 2018; Philips 2018).[13] The predicted values of presidential approval for women decrease substantially after the change in cases at time period 10. However, we do not find evidence that corroborates our expectation that women's approval would be more sensitive to increases in the number of COVID-19 cases than men's presidential approval.

Figure 19 – The Effect of COVID-19 cases (per million) on the U.S. Presidential Approval for Women and Men

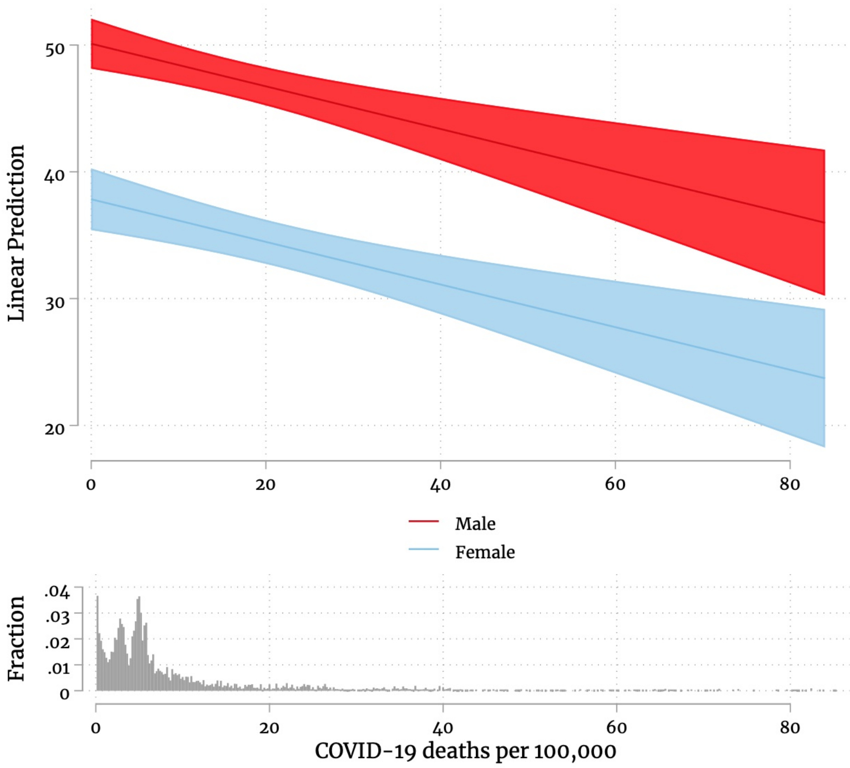


Note: The results depict the dynamic effects of a one standard deviation increase in deaths at time 10. Predicted values and 90 and 95 percent confidence intervals.

[13] Figure 19 shows the simulations based on the results after estimating OLS models.

Using replication data from Warshaw, Vavreck and Baxter-King (2020), I now assess the impact of COVID-19 deaths on approval by gender using a difference-in-differences regression research design approach. Table 51 shows the results, controlling for race, education, gender, vote in 2016, survey wave, and state. Figure 20 depicts the predicted probabilities of approving President Trump's job performance by gender and across different amounts of COVID-19 deaths over the past 30 days, holding the other independent variables at their observed values. Both women and men are less likely to approve of the president's job performance when there are a large number of COVID-19 fatalities over the past 30 days than when there are a relatively smaller number of victims. Similar to the time series analysis, women are less likely to approve of the president than men, and the gender difference persists across different values of COVID-19 deaths.

Figure 20 – Predicted Probabilities of Approving President Trump's Job Performance by Gender, U.S. citizens, July 2019 to July 2020



Note: The predicted probabilities and corresponding 95% CIs were calculated based on the results of a regression model regressing the dependent variable (1 if the respondent approves President Trump's job performance, and 0 otherwise) on the key independent variables (female and state-level COVID-19 deaths per 100,000 over the last 30 days) and control variables (race, education, vote in 2016 elections, wave, and state). The predicted probabilities for women and men were calculated holding the independent variables at their observed values. Data source: Warshaw, Vavreck and Baxter-King (2020).

In sum, evidence from Brazil and the US suggests that both men and women punish the president for mismanaging the crisis. The results further indicate that there is no evidence that female citizens' presidential approval rating is more sensitive to changes in the death toll. Remarkably, the evidence from the time series analysis shows similarity in the magnitude of the effect of COVID-19 fatalities on the president's support in Brazil and the US.

3.6 Conclusion

This chapter has presented evidence to show that there are differences between women and men regarding the subjective perception of the severity of the COVID-19 pandemic. Most importantly, these gender differences are correlated with how individuals act to prevent infection from the virus. Women are consistently more likely to report feeling afraid of contracting the disease and following and supporting social distancing policies. Yet, these gender differences do not translate into differential rates of accountability during the pandemic. Although females' approval ratings for President Trump and Bolsonaro are substantively lower than among their male counterparts, the direction and magnitude of the patterns that underlie the dynamics of support for these presidents across men and women are similar. Since both of these tests pertain to right-wing populist leaders, future research should verify if these findings hold for other democracies in which there is variation in leadership gender, leadership stance towards COVID-19, and other relevant factors.

There are some limitations to the questions that I have sought to explore in this chapter. First, the data availability imposed restrictions on the analysis. For instance, data on approval by different social groups (e.g., race, income, unemployed, formal or informal marked employed, etc.) would be relevant to investigate the roots of the gender gap in approval. However, this type of data is not publicly available, especially for Brazil. Second, the small number of observations restricts the extent to which more fully specified multivariate models could be used to explore the findings in further depth.

Earlier studies have documented that political leaders received a boost in their approval ratings during the first months of the COVID-19 pandemic (Yam et al. 2020), in line with what was documented by the literature as the "rally around the flag" effect (Mueller 1970). Nonetheless, more recent studies have reported a negative impact of COVID-19 cases and deaths on support for the executive. Analyzing the U.S. case, Warshaw, Vavreck and Baxter-King (2020) found that COVID-19 fatalities decreased support for President Trump and Republican candidates for House and Senate. This chapter corroborates this finding and shows that, in democracies hit hard by the pandemic, female and male voters are equally likely to hold incumbents accountable for the worsening of the crisis.

The United States and Brazil poorly managed the pandemic at the federal level, and state and local governments played a larger role in pandemic containment. The efforts to control the outbreak in the United States and Brazil were markedly decentralized and poorly coordinated. In those places, voters may find it difficult to assess who to blame or reward for the efforts to mitigate the crisis. A fruitful avenue for future research is to test whether these findings are different in places where the coordination to combat the outbreak was centralized and straightforward. For example, future research should explore whether female and male voters reward governments that handled the crisis well in other federal democracies as compared to unitary regimes.

4 TESTING THEORIES WITH POLYCHOTOMOUS DEPENDENT VARIABLES

4.1 Introduction

The development of more meaningful ways to interpret the results of nonlinear models has been a central preoccupation in political science and the social sciences more generally (King 1989; Long 1997; King, Tomz and Wittenberg 2000). As these models estimate the probability of a specific outcome given the values of a set of regressors, the estimated coefficient does not directly show the magnitude or sign of the marginal effect on the dependent variable of a change in a key explanatory variable. The magnitude of the marginal effect will depend on the level of the other explanatory variables included in the model (Long and Freese 2014). Compared to binary regression models (e.g., binomial logit and binomial probit), interpreting the regression results of polychotomous dependent variable models (e.g., multinomial logit, conditional logit, multinomial probit, mixed logit, or nested logit) is an even harder task. As these models produce many coefficients, the numerical output may be overwhelming. For instance, a model with a dependent variable with three categories and six explanatory variables generates fourteen different parameter estimates, including the intercept terms. Given this complexity in the context of nonlinearity, no interpretation approach completely describes the relationships between an independent variable and the outcome probability in a multivariate case, which greatly challenges hypothesis testing and inferences from these models. In this chapter, I contribute to this literature by emphasizing the benefits from using a combination of different approaches when evaluating the results produced by polychotomous dependent variable models. Instead of relying on a single approach for all situations, I argue that using a combination of interpretation approaches provides more meaningful ways to interpret the results.

Marginal effects, predicted probabilities, and first differences are some of the most powerful tools employed to evaluate the impact of a key explanatory variable on outcome probabilities. This chapter explores these and other approaches used to make inferences from polychotomous dependent variable models. To show the insights gained from different approaches, I replicate part of the analysis undertaken by an article in which results were presented from a multinomial logit model. The results from the replication dispute some of the original authors' conclusions and provide an interesting case to show the insights gained when researchers employ a combination of interpretation approaches to test their hypotheses.

To analyze what are the most commonly employed approaches in the literature, I reviewed and classified all sixty-five papers published in the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics* from 2006 to 2016 that employed a polychotomous dependent variable model in some part of their analysis. In the majority of the papers, researchers based their inferences in part on the analysis of coefficients tables. Only a small fraction of the surveyed papers employed a combination of coefficient tables and different marginal effects to evaluate whether statistical and substantive evidence corroborates their theoretical predictions. More importantly, in 25% of the papers, the researchers relied solely on coefficient tables to report and discuss their results. This means that, in these articles, the authors conducted their hypothesis testing based on the sign and statistical significance of the estimated coefficients. Such a strategy of interpreting the results based solely on the sign and statistical significance of the estimated coefficients may lead to flawed inferences (Paolino 2020). In contrast, marginal effects and predicted probabilities provide valuable tools to evaluate the substantive and statistical significance of an analysis' results (Long and Freese 2014; Cameron and Trivedi 2010; Williams 2012; Hanmer and Kalkan 2013; Paolino 2020). These findings highlight that there are considerable improvements that researchers can make to enhance their analyses by adopting more meaningful ways to interpret their results.

The literature survey also shows that the marginal effect at means (MEM) and the average marginal effect (AME) are two of the most commonly used marginal effect approaches to interpret the results of polychotomous dependent variable models in political science. While the MEM describes the average case, the AME portrays the average impact on the sample. Most of the discussion about the differences between the MEM and the AME approaches has been focused on the case of binary dependent variable models. I review the main aspects raised by the literature about the differences between these two approaches. Then, I perform a series of Monte Carlo simulations to analyze whether these two approaches produce distinct quantities of interest in the context of polychotomous dependent variable models and which one is more susceptible to bias due to model misspecification. In line with the findings from the previous literature that had focused on the binary probit case (Hanmer and Kalkan 2013;

Bartus 2005), I show that the AME and the MEM produce distinct estimates. The results further suggest that the MEM is more sensitive to bias due to model misspecification than the AME.

The rest of the chapter proceeds as follows. In the next section, I briefly review some of the most important methodological contributions to interpreting results from nonlinear models in the discipline. Then, I summarize the most commonly used interpretation approaches. To illustrate the insights gained from each approach, I replicate an article in which results were presented from a multinomial logit model. After discussing the results from a survey of the literature that I conducted, I present the results from the Monte Carlo simulations. Finally, I conclude the chapter by providing recommendations for improving the interpretation of the results from polychotomous dependent variable models. I briefly discuss some aspects that future research in the area can further develop.

4.2 *Brief Review of Interpreting the Results of Nonlinear Models in Political Science*

Social scientists have long been interested in more theoretically substantive ways for interpreting the results of nonlinear models (Hanushek and Jackson 1977; King 1989; Long 1997; Petersen 1985). As Hanushek and Jackson (1977) emphasize, "we are interested in analyzing the underlying probability of a given event or choice; more specifically, how a series of exogenous variables influences the underlying probabilities (215)." In order to constrain the predicted outcomes so that they are between 0 and 100 and add up to 100%, we need to change the functional form to be nonlinear. However, the interpretation of the regression results is especially more complicated because the dependent variable is a nonlinear function of the explanatory variables. Thus, the estimated coefficient does not directly show the magnitude or sign of the marginal effect on the dependent variable of a change in a key explanatory variable. In political science, the work of King (1989) was seminal in proposing ways to evaluate the results from non-linear discrete regression models substantively. In summary, King (1989) suggests four general procedures to meaningfully interpret the coefficient estimates of discrete regression models (i.e., graphical methods, fitted values, first differences, and derivatives). When describing these interpretation approaches, King (1989) acknowledges the advantages and limitations of each one and suggests the use of substantive research knowledge to evaluate the effect of an explanatory variable on the outcome.

The work of Long (1997) has also been fundamental on this line of inquiry. Long (1997) highlights that, given the nonlinearity of most discrete regression models, no approach completely describes the relationships between an independent variable and the outcome probability in a multivariate case. Similar to King (1989), Long (1997) proposes five different interpretation approaches (i.e., predicted probabilities, partial changes, discrete changes, interpretation using odds ratio, and plotting the coefficients) for showing the substantive implications for the estimated models. Both King (1989) and Long (1997) recognize that interpreting the results of nonlinear models for inference is a complex task. The goal is to use strategies that help a researcher convey the substantive significance of the reported results. To do so, researchers can take advantage of different approaches in evaluating hypotheses.

However, in this first wave of studies, a thorough discussion of the relevance of reporting uncertainty alongside the estimates was absent. Coefficient estimates, marginal effects, and predicted probabilities suffer from different sources of uncertainty. The work of Herron (1999) and King, Tomz and Wittenberg (2000) was central in calling attention to this issue. Herron (1999) emphasized the importance that estimated quantities should be reported accompanied by confidence intervals or standard errors. Back then, the common practice was reporting estimated marginal effects or predicted probabilities without any measure of uncertainty. Notably, such practice tends to overstate the estimates' precision. Standard errors can be obtained using the delta method, bootstrapping, or via simulations (Mize, Doan and Long 2019).

King, Tomz and Wittenberg (2000) highlight that there are two forms of uncertainty that were not addressed by the earlier studies when reporting predicted probabilities, derivatives, or first differences. The first one is the estimation uncertainty which arises from not knowing the true values of the estimated parameters. This type of uncertainty is acknowledged when researchers report standard errors or confidence intervals. The second type is the fundamental uncertainty that arises from other events that may influence the dependent variable but are not taken into account by the explanatory variables (King, Tomz and Wittenberg 2000). To account for uncertainty, researchers have developed parametric and non-parametric methods. King, Tomz and Wittenberg (2000) developed a statistical package in Stata, *Clarify*, which facilitated the computing of both forms of uncertainty when calculating a quantity of interest based on estimates from a statistical model (Tomz, Wittenberg and King 2003). *Clarify* has been influential among practitioners and, more recently, it was upgraded by a new expanded version of the same protocol now called for R *Zelig* (Imai, King and Lau 2009).[1]

In the last decades, scholars have directed attention to other issues involving the interpretation of results from nonlinear regression models. For instance, how to correctly interpret interactions results in nonlinear models (Ai and Norton 2003; Berry, DeMeritt and Esarey 2010; Mize 2019), how to compare the effects across models (Mize, Doan and Long 2019), and which interpretation approach provides a better summary of the impact of an explanatory variable on the outcome probability (Hanmer and Kalkan 2013). Undoubtedly, there has been substantial progress in the discipline towards using more meaningful statistical ways to communicate and interpret the results of nonlinear models. However, such advances have not been widely incorporated by practitioners in the field. Despite the limitations, it is still a common practice to interpret the results of nonlinear models based solely on the coefficient estimates (Paolino 2020). Moreover, there is still a belief among scholars that the AME is the best approach for conducting theoretically driven hypothesis testing when evaluating results from limited dependent variable models (Hanmer and Kalkan 2013). In the next section, I will present and discuss the limitations and the advantages of different interpretation approaches. As I will show, even the AME can be misleading, especially when the distribution of effects is skewed. Thus, I argue that researchers should focus on using a combination of different approaches rather than a single one that fits all situations.

4.3 Brief Review of Some Approaches for Interpreting the Results of Multinomial Logit Models

Interpretation of regression results is more complicated in nonlinear models. As Long and Freese (2014) explain, "the challenge of interpreting results, then, is to find a summary of how changes in the independent variables are associated with changes in the outcome that best reflects critical substantive processes without overwhelming yourself or your readers with distracting detail (227)." While in linear regression models, interpreting the results is straightforward (i.e., a unit change in an independent variable

[1] As of November 2021, *Clarify* (Tomz, Wittenberg and King 2003) already has 1,835 citations and *Zelig* (Imai, King and Lau 2009) 358 citations according to Google Scholar.

leads to an increase or decrease in the dependent variable, holding all the other variables constant). In contrast, as nonlinear regression models estimate the probability of a specific outcome given a set of regressors, the estimated coefficient does not directly show the marginal effect on the dependent variable of a change in a key explanatory variable. Researchers then need to use the estimated coefficients to calculate the marginal effects or discrete changes of a key independent variable on the probability of a specific outcome. Most importantly, the value of the marginal effect will depend on the level of the other explanatory variables included in the model.

Compared to binary regression models, interpreting the regression results of polychotomous dependent variable models is an even harder task since these models produce many coefficients. In this section, I will explore some approaches for interpreting the results of polychotomous dependent variable models.

To show the shortcomings and advantages of different interpretation approaches, I will replicate part of the analysis undertaken by an article in which results were presented from a multinomial logit models to test its hypotheses. I chose to replicate the article, "Economic Discontent as a Mobilizer: Unemployment and Voter Turnout," from Burden and Wichowsky published in 2014 at the *Journal of Politics*.[2] In that article, the authors argue that "a worse economy actually mobilizes voters, thus making turnout a key mechanism of economic accountability that connects the economy to electoral outcomes (887)." Among the hypotheses that the authors test, they evaluate whether "unemployment affects both the decision to vote and for whom to vote (894)." Conversely, they also propose a rival hypothesis, "higher turnout is merely a response to a more competitive electoral environment rather than unemployment per se (893)."

To test these hypotheses, Burden and Wichowsky (2014) employ the pooled American National Election Studies (ANES) data from 1978 to 1998. They estimate a multinomial logit model in which the dependent variable is vote choice in the gubernatorial elections. The dependent variable has three categories: vote for the Democratic candidate, Republican candidate, and abstain. As explanatory variables, they include state unemployment rates, competitive elections (measured by "the amount of money the national party committees transferred to state and local parties in each year"), socioeconomic indicators (age, female, education, income, etc.), state-level indicators and fixed

2 This article has been influential in the economic voting and electoral studies literature. As of November 2021, it has 110 citations, according to Google Scholar.

effects for states and years. The results from the multinomial logit model are shown in their Table 5 (Burden and Wichowsky 2014, p. 895). The authors made the replication files available, and I could reproduce the same results from their multinomial logit model as those reported in the article.

4.3.1 Coefficients Table and Plot

Researchers commonly use tables or plots to display their models' estimated coefficients (in log odds). Some studies also opt to display the odds ratio or the relative-risk ratio. In their analysis, Burden and Wichowsky (2014) report the results from the multinomial logit model using a coefficient table showing the estimated coefficients in log odds. Table 4 reproduce the results from their analysis estimating multinomial logit regression model.

Although researchers may be tempted to infer whether there is evidence supporting or not their hypotheses by evaluating the size, sign, and statistical significance of the estimated coefficients, this strategy may lead to misleading inferences. The estimated coefficients from a multinomial logit model provide limited information about a key explanatory variable's effect size. This occurs because the size and the statistical significance of a coefficient are contingent on the baseline category (Paolino 2020). Consider the following multinomial logit model:

$$ln\Omega_{m|b}(\mathbf{x}) = ln\frac{Pr(y=m|\mathbf{x})}{Pr(y=b|\mathbf{x})} = \mathbf{x}\beta_{m|b} \quad \text{for } m = 1 \text{ to } J, \tag{7}$$

where m refers to the dependent variable categories that range from 1 to J, and b refers to the reference or baseline category.

As Equation 7 shows, the results from a multinomial logit model describe how a one-unit increase in an explanatory variable, x_k, leads to an increase or decrease in the relative log ratio of category m versus category b (the baseline). This means that the size and statistical significance of the coefficient estimates depend not only on the changes in the probability of category m, but also of changes on the baseline category, b. When the probability of the baseline category changes, the size, sign of the coefficient estimates, and the standard errors may change substantively. For instance, a baseline category with few observations leads to estimates with higher standard errors and consequently

Table 4 – Reproduction of Burden and Wichowsky's (2014) Multinomial Logit Analysis

	Vote for Democrat	Vote for Republican
State unemployment rate	6.639**	2.505
	(3.254)	(4.845)
Campaign expenditures	0.229**	0.335***
	(0.090)	(0.120)
Democrat	1.378***	0.351**
	(0.075)	(0.138)
Republican	0.226*	1.688***
	(0.133)	(0.121)
African American	0.374***	-0.626***
	(0.123)	(0.221)
Latino	0.116	-0.316***
	(0.077)	(0.113)
Other race/ethnicity	-0.664***	-0.584***
	(0.151)	(0.152)
Female	-0.011	-0.027
	(0.049)	(0.054)
Married	0.193***	0.353***
	(0.065)	(0.071)
Age	0.041***	0.042***
	(0.002)	(0.002)
Education	0.635***	0.639***
	(0.042)	(0.046)
Income	0.175***	0.228***
	(0.034)	(0.030)
Income not reported	-0.302***	-0.124
	(0.112)	(0.144)
Unemployed	-0.338***	-0.570***
	(0.127)	(0.135)
Constant	-6.051***	-6.811***
	(0.414)	(0.446)
Observations	7620	
State-fixed effects	Yes	

Note: Results from a multinomial logit regression model. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

wider confidence intervals. More importantly, the signs of the estimated coefficients may not even show the direction of the marginal effects (Cameron and Trivedi 2010, 488).

While the estimated coefficients' size, sign, and statistical significance change as we change the baseline category, the same does not occur when we calculate the predicted probabilities. That is, the probabilities will be the same *regardless* of the baseline category employed in the estimation of the model. Equation 8 shows the probability equation for the multinomial logit model with category b as the baseline.

$$Pr(y = m|\mathbf{x}) = \frac{exp(\mathbf{x}\beta_{m|b})}{\sum_{j=1}^{J} exp(\mathbf{x}\beta_{j|b})} \tag{8}$$

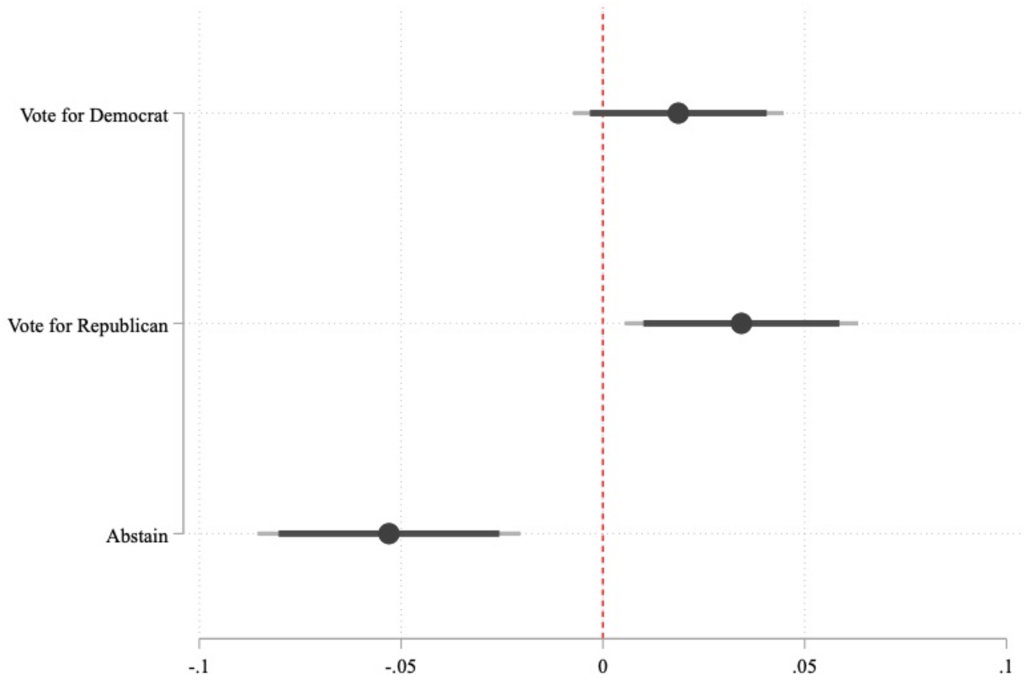
In the log odds equation (Equation 7), the denominator only considers the probability for the baseline category. Thus, when the baseline category changes, the estimated parameters also change. In contrast, in the probability equation (Equation 8), the denominator considers the information from all categories of a model since it sums up the exponent from the estimated log odds of all categories. That is, regardless of the baseline category, the estimated parameters produce the same predictions.

Analyzing the impact of competitive elections on vote, Burden and Wichowsky (2014) conclude that "competitive races are positively correlated with casting a ballot for either candidate relative to the baseline category of not voting (Burden and Wichowsky 2014, p. 896)." In fact, by evaluating the results in terms of log odds, as Table 4 shows, a one-unit increase in competitive elections is associated with an increase in the relative log odds of voting for the Democrat candidate and also for voting for the Republican candidate relative to the baseline category of not voting. Both log odds are statistically significant at the 0.05 level. However, an analysis of the marginal effects of competitive elections on vote leads to a different conclusion. As Figure 21 shows, once we evaluate the effect of a marginal change in campaign expenditures on the probability of voting for the Democrat candidate, Republican, and abstain, the effect on voting for the Democrat candidate is not statistically significant at the 0.05 or 0.10 levels. That is, a marginal change in competitive races leads to an increase in the probability of voting for the Republican candidate and a decrease in the probability of not voting. However, such a marginal change does not impact the probability of voting for the Democrat candidate.

As the discussion about Equation 7 showed, the size, sign, and statistical significance of the estimated coefficients in log odds depend not only on the changes in the probability of category *m*, but also on changes in the baseline category, *b*. In contrast, the predicted probabilities are the same regardless of the baseline category of the estimated model. To calculate the marginal effects (ME) as reported in Figure 21, I evaluated the difference between (i) the predicted probability of outcome *m after* an increase in competitive elections compared to (ii) the predicted probability of outcome *m before* an increase in competitive elections, that is:

$$ME = Pr(y = m|\mathbf{x}, x_k = x_k^{after}) - Pr(y = m|\mathbf{x}, x_k = x_k^{before})$$

Figure 21 – The Effect of a Marginal Change in Campaign Expenditures on Probability of Voting for a Democrat, a Republican, and Abstain



Note: While the gray lines indicate a 95% confidence interval, the black lines represent a 90% confidence interval. Marginal effects calculated holding the explanatory variables at their observed values for each unit in the sample.

As the marginal effects are calculated based on the predicted probabilities, they will be the same *regardless* of the baseline category. In sum, by replicating Burden and Wichowsky (2014)'s analysis of the impact of competitive races on voting, it is clear that statistically significant estimated coefficients do not translate into statistically significant marginal effects. Thus, analysts need to pay extra attention, especially when evaluating their results based solely on the estimated coefficients, as they may incur misleading inferences. This is a relevant issue given that political scientists usually set their hypothesis in terms of marginal effects or predicted probabilities rather than log odds, odds ratio, or relative-risk ratio. In the following subsection, I will discuss marginal effects in more detail.

4.3.2 Marginal Effects

Marginal effects (or partial effects) measure the change in the probability of an outcome as a response to a change in x_k, holding constant the other explanatory variables at specific values (Long and Freese 2014, p. 239). Marginal effects can be computed for marginal changes (infinitely small change) or discrete changes (discrete or finite change) in x_k. More importantly, the magnitude of the marginal effects depends on how the analyst sets the values of the other covariates in the model. While the value at which the covariates are held constant is irrelevant in the case of linear regression models, this is a relevant issue in the case of nonlinear models. As King (1989) explains: "as in all nonlinear equations, a single unit change in X_j will have a different effect on the expected value of Y depending on the points at which the curve is evaluated (108)."

In sum, there are three main approaches to calculate the marginal effects given how the values of the covariates are set: (a) marginal effect at the mean, (b) average marginal effect, and (c) marginal effect at representative values. I will briefly show how these three distinct types of marginal effects produce different results. To illustrate the differences between them, I will explore the analysis undertaken by Burden and Wichowsky (2014) about the effect of unemployment rates on the probability of voting.

Marginal Effect at the Mean

The marginal effect at the mean (MEM) approach shows the marginal effect on the dependent variable outcomes of a change in a key explanatory variable *holding* all the covariates at their means, medians, or modes. While the MEM for a marginal change in x_k on the probability of outcome m holding all covariates at their means is:

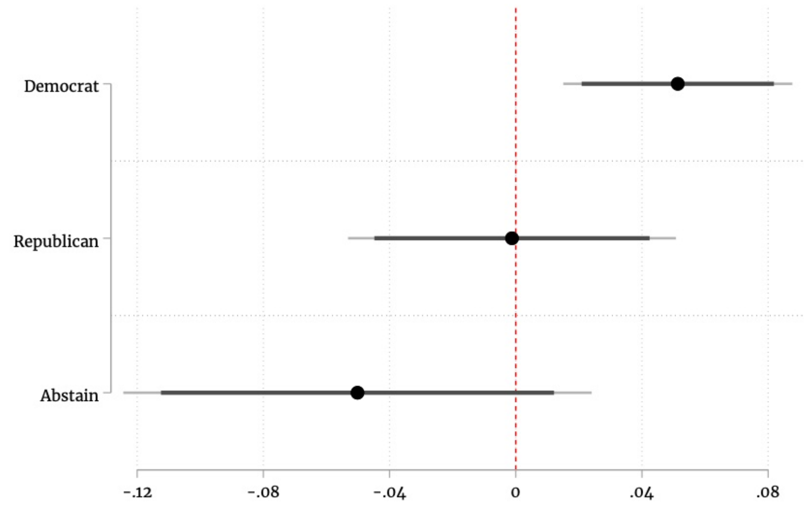
$$\frac{\partial Pr(y=m|\bar{\mathbf{x}}, x_k=\bar{x}_k)}{\partial x_k}$$

The MEM for a discrete change is:

$$\frac{\Delta Pr(y=m|\bar{\mathbf{x}}, x_k=\bar{x}_k)}{\Delta x_k}$$

To test Burden and Wichowsky's (2014) hypothesis about the impact of unemployment on the probability of voting, Figure 22 shows the marginal effect at means for a discrete change in state unemployment rates. The authors conclude that "all else equal, a one standard deviation swing around the mean state unemployment rates is associated with a four percentage point increase in the probability of voting for the Democratic candidate and a four percentage point decrease in the probability of abstaining (Burden and Wichowsky 2014, p. 896)." However, the analysis of the marginal effects at means does not corroborate this conclusion. A one standard deviation swing around the mean state unemployment rate (i.e., Δ = (mean + 1 SD) - (mean - 1 SD)) is associated with a 5.1 percentage point (95% CI 1.5%, 8.8%) increase in the probability of voting for the Democratic candidate. Such a change does not lead to a marginal effect in voting for the Republican candidate that is statistically different from zero (95% CI -5.3%, 5.1%) or abstaining (95% CI -12.4, 2.4).

Figure 22 – Marginal Effect at Means for a Discrete Change in State Unemployment Rate on the Probability of Voting for a Democrat, a Republican, and Abstain



Note: While the gray lines indicate a 95% confidence interval, the black lines represent a 90% confidence interval. Marginal effects calculated for a one standard deviation swing around the mean state unemployment rate with all other explanatory variables at their mean values.

Average Marginal Effect

Another approach commonly employed by researchers interested in interpreting the results of nonlinear models is the average marginal effect (AME). The AME approach calculates the marginal effect holding the covariates at their existing values in the data set and then computes the sample average of these changes. As Williams (2012) explains, "the logic is similar to that of a matching study, where subjects have identical values on every independent variable except one (p. 326)." The average marginal effect is calculated as follows:

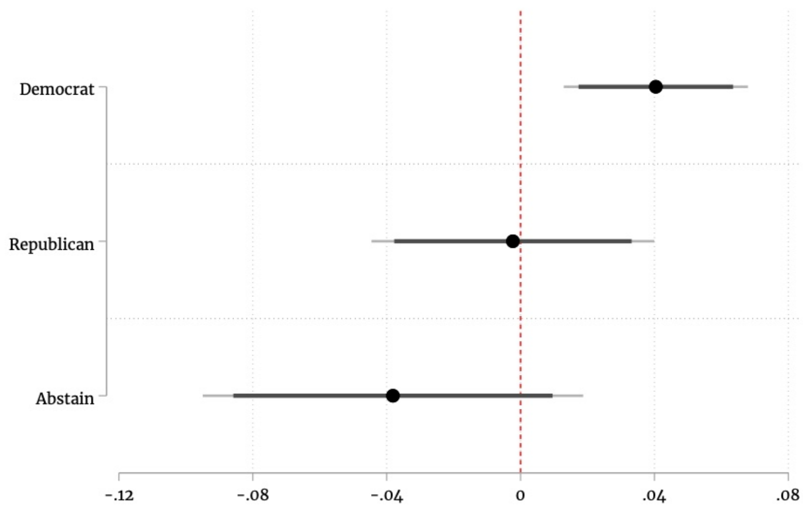
$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial Pr(y_i = m|\mathbf{x} = \mathbf{x}_i)}{\partial x_k}$$

And, for a discrete change, the AME is:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\Delta Pr(y_i = m|\mathbf{x} = \mathbf{x}_i)}{\Delta x_k}$$

Figure 23 shows the average marginal effect for a discrete change in state unemployment rates on the probability of voting. Burden and Wichowsky (2014) had argued that such a change would lead to an increase of 4 percentage points in the probability of voting for the Democratic candidate and a decrease of 4 percentage points in the probability of abstaining. Instead, the AME calculations confirm that a one standard deviation swing around the mean state unemployment rate (i.e., Δ = (mean + 1 SD) - (mean - 1 SD)) is associated with a 4 percentage points (95% CI 1.3%, 6.8%) increase in the probability of voting for the Democratic candidate. However, such a change does not lead to an average marginal effect statistically different from zero in voting for the Republican candidate (95% CI -4.4%, 4.0%) or abstaining (95% CI -9.5, 1.9). Here again, the statistical inferences in the study could have befitted from additional and further exploration.

Figure 23 – Average Marginal Effect for a Discrete Change in State Unemployment Rate on the Probability of Voting for a Democrat, a Republican, and Abstain



Note: While the gray lines indicate a 95% confidence interval, the black lines represent a 90% confidence interval. Marginal effects calculated for a one standard deviation swing around the mean state unemployment rate with all other explanatory variables at their observed values for each unit in the sample.

From the article and the replication files, it is not clear whether Burden and Wichowsky (2014) employed the MEM or the AME approach. This is because both approaches provide similar results, but there is more uncertainty around the estimated effects calculated via MEM and therefore larger confidence intervals. Since these were not reported in the study, it is therefore difficult to know with certainty which approach was followed by the authors.

Marginal Effect at Representative Values

The marginal effect at representative values (MER) approach presents the marginal effect on the dependent variable outcomes of a change in a key explanatory variable *holding* the other variable(s) at *representative values* of theoretical interest. For instance, the marginal effect is calculated for every decile or range of values of a covariate. A relevant advantage of MER over MEM and AME is that it does not rely on a single estimate to evaluate the impact of the key explanatory variable on the outcome probability. As Williams (2012) explains, "the biggest problem with both of the last two approaches [MEM and AME], however, may be that they only produce a single estimate of the ME [marginal effect]. No matter how 'average' is defined, averages can obscure differences in effects across cases (326)."

For the MER approach, the marginal and discrete changes are calculated in the same way as in the MEM. Instead of holding all covariates to their means ($\bar{\mathbf{x}}$), in the case of MER, the covariates are set to other specific values, not necessarily their means. Formally, the marginal change is

$$\frac{\partial Pr(y = m|\mathbf{x} = \mathbf{x}^*)}{\partial x_k}$$

While for a discrete change, it is:

$$\frac{\Delta Pr(y = m|\mathbf{x} = \mathbf{x}^*)}{\Delta x_k}$$

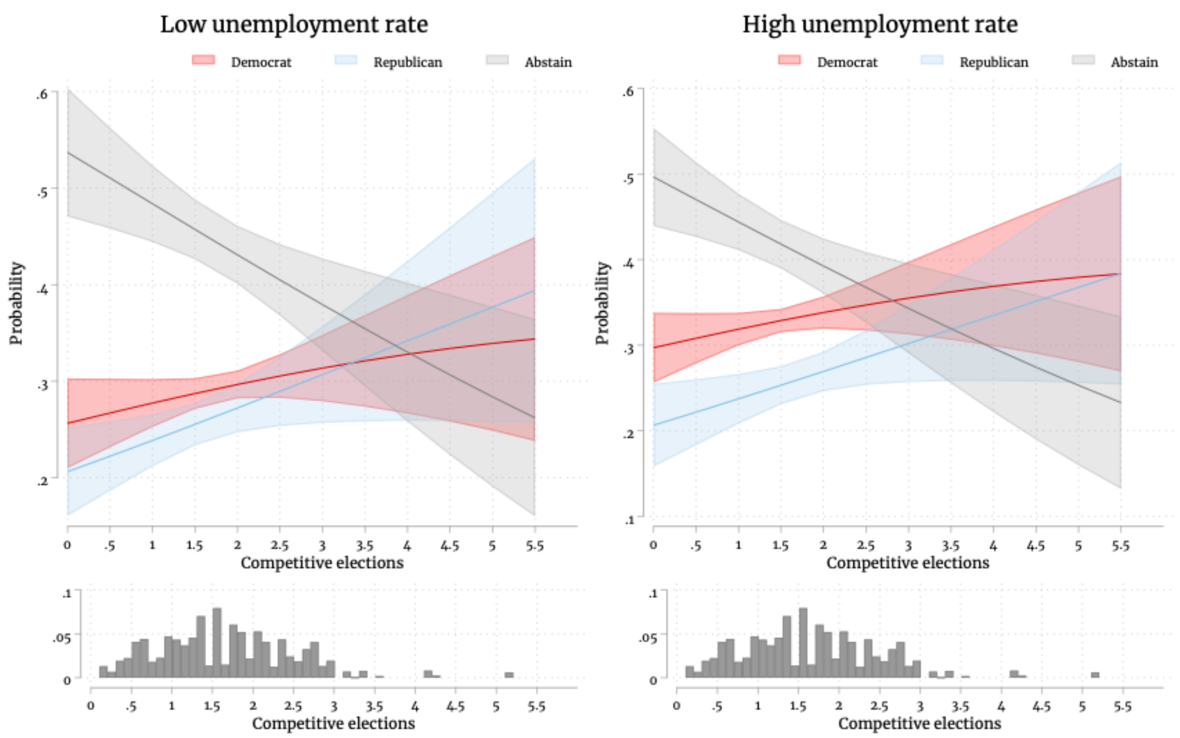
To illustrate the insights gained using MER compared to the previously discussed marginal effects, I will evaluate whether there is evidence corroborating one of the hypotheses advanced by Burden and Wichowsky (2014). As discussed earlier in this section, Burden and Wichowsky (2014) have hypothesized that "higher turnout is merely a response to a more competitive electoral environment rather than unemployment per se (893)." To test this hypothesis, Figure 24 shows the effect of competitive elections on the probability of abstaining under low and high state unemployment rates. As Figure 21 showed, a marginal change in competitive elections (holding the control variables at their observed values) leads to an increase in the probability of voting for the Republican candidate of 3.3 percentage points (95% CI 0.4, 6.2) and a decrease for abstaining of 5.2 percentage points (95% CI -8.4, -1.9). The effect on voting for the Democrat candidate is not statistically significant at the 0.05 or 0.10 levels (95% CI -0.7, 4.5). Figure 24 corroborates such findings. Regardless of experiencing low or high state unemployment rates, the probability of abstaining decreases as the elections get more competitive.

Taken together, the evidence from Figure 22, 23, and 24 corroborates the hypothesis that abstention is a response to a competitive electoral environment rather than to unemployment. In summary, such finding disputes the conclusion made by the authors that "we find no evidence that the relationship [between the unemployment rates and voter turnout] is mediated by campaign competition (...). This supports our contention that the effect is due to the behavior of the public rather than campaign elites (896)." Burden and Wichowsky (2014) based such conclusions on the evidence provided by several different analyses undertaken throughout the article. However, the evidence from the analysis of the marginal effects based on the multinomial logit model demonstrates that the effect on voter turnout is more likely due to campaign elites rather than the behavior of the public as a reaction to higher unemployment rates.

4.3.3 Additional Approaches

Scholars may opt to evaluate the impact of a key explanatory variable on the outcome probability by examining the distribution of marginal effects for each observation in the estimation sample. This strategy allows one to evaluate if the distribution is skewed and where specific marginal effects (e.g., AME and MEM) are placed in the distribution (Long and Freese 2014). Such analysis can provide additional information that is not readily available from the evaluation of the AME or MEM. As Long and Freese (2014) explain, "just as the means of the independent variables used to compute

Figure 24 – The Effect of Competitive Elections on the Probability of Voting for a Democrat, a Republican, and Abstain under Low and High State Unemployment Rate



Note: The effects and 95 percent confidence intervals were calculated, holding the control variables at their observed values. A low unemployment rate is defined as a standard deviation below the mean, and a high unemployment rate as a standard deviation above the average.

the MEM might not correspond even approximately to anyone in the sample, the AME might not correspond to the magnitude of the marginal effect for anyone in the sample." The distribution of marginal effects is particularly useful to evaluate the variation in the size of the effects within the sample, and it can provide valuable insights (Long and Freese 2014).

To illustrate the insights gained with this approach, Figure 25 shows the distribution of discrete changes in the probability of voting for the Democrat candidate (Panel a), Republican candidate (Panel b), and abstaining (Panel c) for a discrete change (from one standard deviation below to one above the mean) in the state unemployment rate. The distributions of effects are markedly bimodal for the probability of voting for the Democrat candidate and abstaining. In both cases, the AME corresponds to the magnitude of the effect only for a small share of the sample. In contrast, the distribution of effects for a discrete change on unemployment rates in the probability of voting for the Republican candidate is around zero, with both AME and MEM assuming similar values.

To further illustrate how the AME can be misleading in some situations, Figure 26 shows the predicted probabilities of voting for the Democrat, Republican, and abstaining calculated via AME and MEM. The distributions of effects for the probability of voting for the Democrat and Republican candidate are markedly right-skewed. Given the skew, the MEM is a better indicator of what is expected for most respondents than is the AME, particularly in the case of the distribution for the Republican candidate. In the case of the probability for the Democrat candidate, neither the AME nor the MEM provides a representative indicator of what is expected for the majority of the respondents.

Analysts may also opt to display the first and the second differences of marginal effects or predicted probabilities when their interest is in evaluating the impact of changing one or more explanatory variables at the same time over the outcome probability. Table 5 shows the first difference of discrete changes in the unemployment rates and competitive elections (one standard deviation below compared to one above the mean) and the second difference of these effects. The results show that more competitive elections lead to a significantly higher probability of citizens voting. Corroborating the findings from Figure 23, the effect of increasing the unemployment rate does not translate into a higher turnout. In addition, the impact of rising unemployment rates and elections competition at the same time does not lead to a statistically significant effect on the probability of abstaining (second difference 0.086 0.039 = -0.047; p = 0.204).

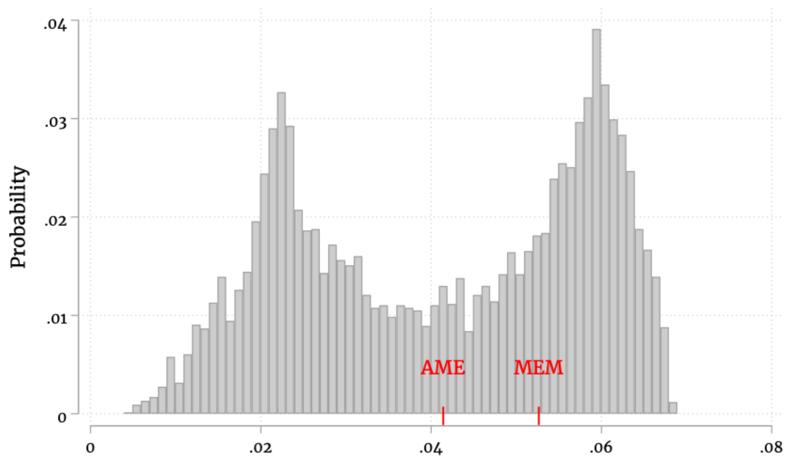
Table 5 – Probability of Abstaining by a Low and a High Level of Unemployment and Election Competition

	Pr(Abstain)	First differences	Second differences
Low unemployment rates	0.450***		
	(0.015)	-0.039	
High unemployment rates	0.411***	(0.030)	
	(0.014)		-0.047
Low elections competition	0.476***		(0.037)
	(0.015)	-0.086***	
High elections competition	0.390***	(0.027)	
	(0.013)		

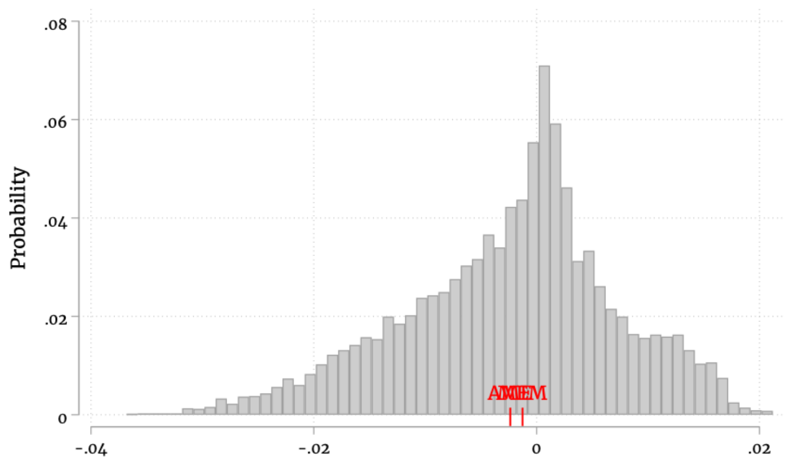
Note: A low (high) level of unemployment or election competition is defined as a standard deviation below (above) the mean. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 25 – The Effect of a Discrete Change (from one standard deviation below to one above the mean) in Unemployment Rate on the Probability of Voting

(a) Democrat Candidate



(b) Republican Candidate



(c) Abstain

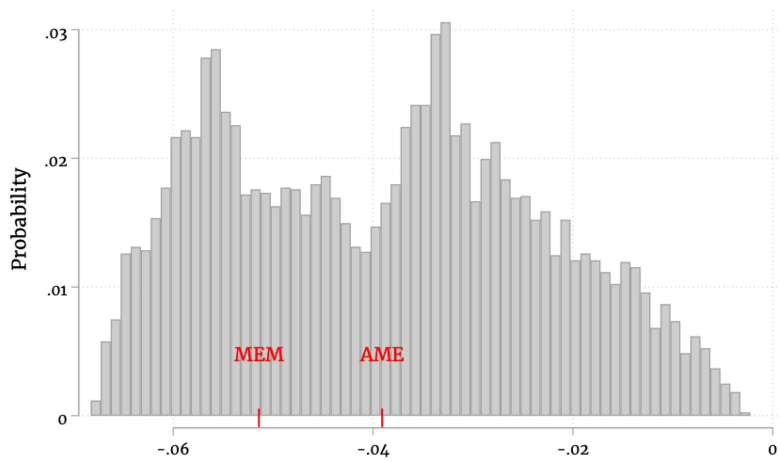
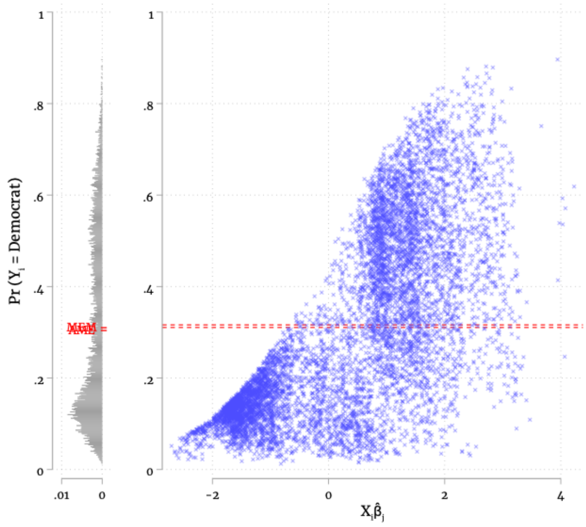
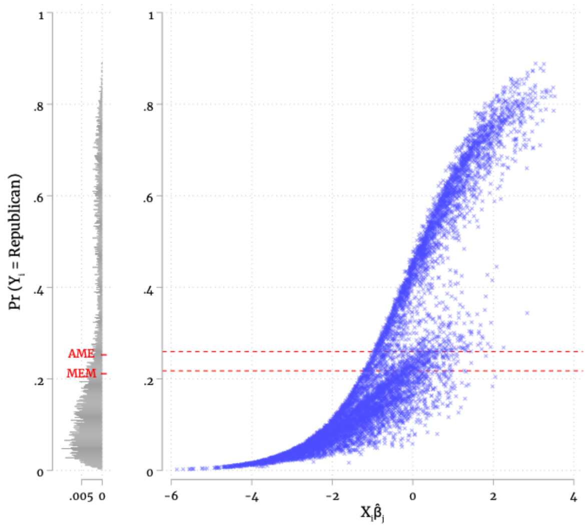


Figure 26 – Predicted Probabilities of Voting for the Democrat Candidate, Republican or Abstain

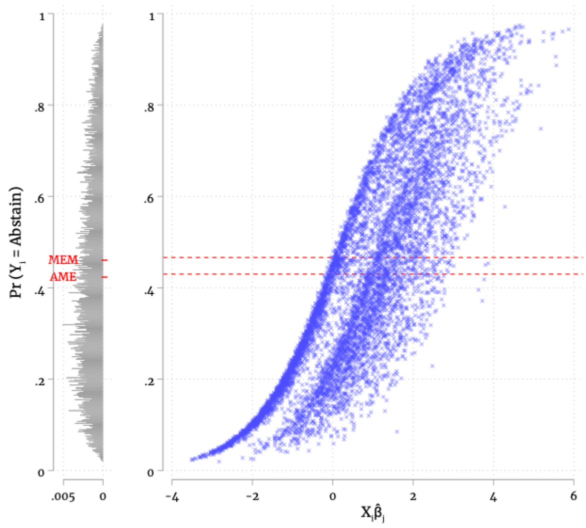
(a) Democrat Candidate



(b) Republican Candidate



(c) Abstain



In summary, the replication I have presented of Burden and Wichowsky's (2014) analyses provides evidence that there are risks that misleading or even incorrect inferences when a more robust and complete approach is not adopted in the context of multinomial models. As the findings I have presented here underscore, there are considerable insights gained when researchers use a combination of interpretation approaches to test their hypotheses. In the next section, I will briefly discuss the results of a survey of the literature about the most commonly employed interpretation approaches.

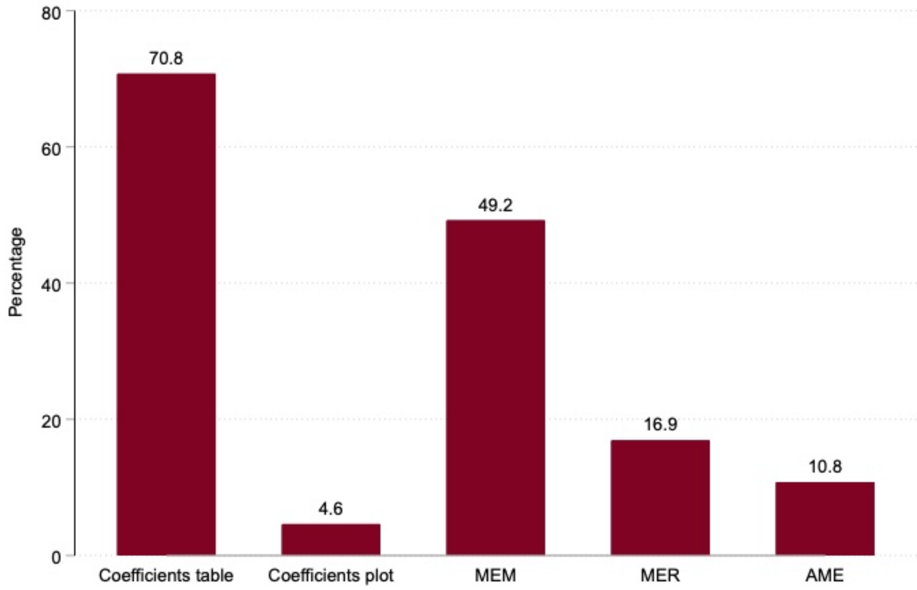
4.4 Patterns of Hypothesis Testing in Articles

To evaluate how researchers employ and interpret the results of polychotomous dependent variable models, I conducted a survey of the literature of empirical studies published in the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics*. A total of 157 articles were published between 2006 and 2016 that mentioned at least one of the following expressions: 'conditional logit', 'multinomial logit', 'multinomial probit', 'nested logit', and 'mixed logit' were identified. As I am interested in analyzing how researchers have employed different interpretation approaches to examine the results of statistical models with polychotomous dependent variables, I excluded from this pool methodological articles as this study is interested in examining how hypotheses are tested in empirical applications in the discipline. Thus, the final sample has 65 articles that employed at least one of these methods to test their hypotheses. From these 65 selected articles, 25 (38.5%) were published in AJPS, 11 (16.9%) in APSR, and 29 (44.6%) in JOP.

Figure 27 shows the distribution of the most common approaches to hypothesis testing in the 65 articles in the sample. Most of the studies rely on coefficients tables to present and interpret their results (70.8%). A small group of studies reports coefficient estimates in a plot instead of a table (4.6%). Half of the articles employ MEM to analyze hypotheses (49.2%). Only 16.9% and 10.8% of the surveyed articles used, respectively, MER and AME. In summary, most of the 65 articles relied on the coefficients table and on MEM to examine their findings.

Figure 28 summarizes the average number of interpretation approaches reported by the articles. Most of the papers employed either one or two strategies to analyze

Figure 27 – Distribution of Approaches used in Hypothesis Testing (%)

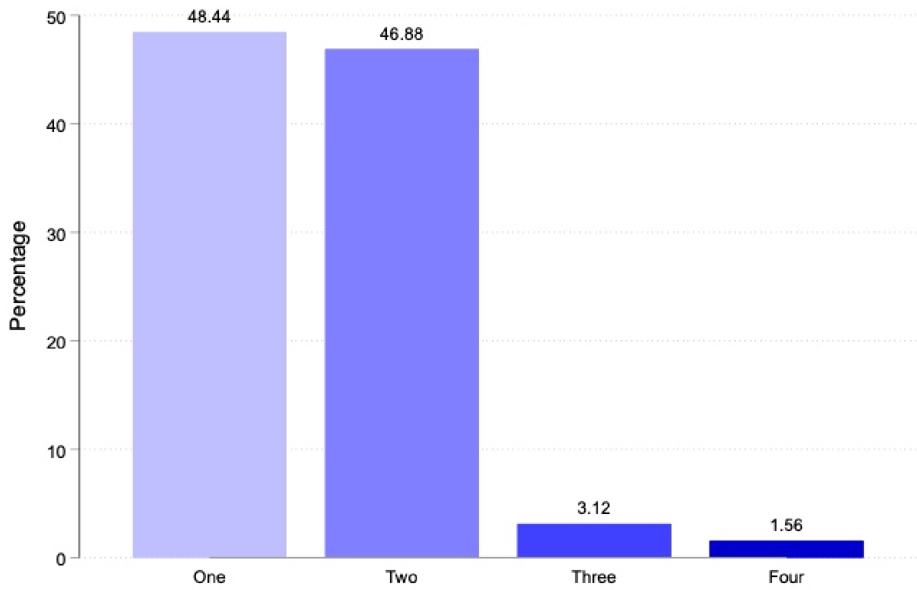


Source: author's compilation.

hypotheses. A small group of studies employed more than two approaches (4.7%). To get a sense of how this pattern is different across journals, Figure 29 shows the mean number of approaches reported in each of the three journals that were included in the sample. The figure shows a considerable variation in the average number of interpretation approaches employed by the articles published in these journals. Most of the articles (60%) published by AJPS employed one interpretation approach when conducting hypothesis testing and a smaller share of studies (40%) based their conclusions on the reporting of a combination of two approaches. Among the reviewed articles published in the APSR, slightly more than half (54.5%) used a combination of two approaches, 36.4% only employed one approach, and 9.1% used a combination of four different interpretation approaches. Finally, among the articles published at JOP, 42.9% used only one approach, 50% reported two approaches, and 7.1% used a combination of three different approaches.

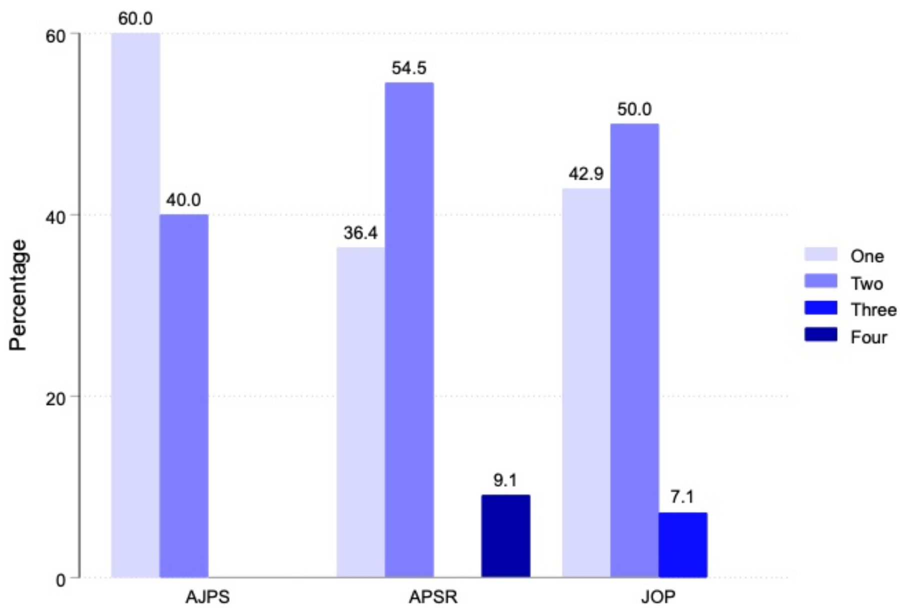
There is also significant variation in the way that researchers combine different interpretation approaches to test their theories. Most of the articles (31%) reported a combination of the coefficients table and the marginal effects at means. A small share of the articles employed a combination of a coefficients table and marginal effects at representative values (6.15%) or a combination of a coefficients table and average

Figure 28 – Number of Hypothesis Testing Approaches Reported by Article (%)



Source: author's compilation.

Figure 29 – Number of Interpretation Approaches Reported by Journal



Source: author's compilation.

marginal effects (4.62%). A small number of articles also reported other combinations, but they were much less common. More importantly, about one quarter (24.62%) of the articles reported only a coefficients table. This means that, in these articles, the authors conducted their hypothesis testing based on the sign and statistical significance of the estimated coefficients. However, researchers may incur in misleading inferences when they based their conclusions solely on the sign and statistical significance of the estimated coefficients (Paolino 2020). In contrast, marginal effects and predicted probabilities provide valuable tools to evaluate the substantive and statistical significance of an analysis' results (Long and Freese 2014; Cameron and Trivedi 2010; Williams 2012; Hanmer and Kalkan 2013; Paolino 2020).

4.5 MEM vs. AME in the context of Limited Dependent Variable Models

As the survey of the literature discussed in the last section shows, the MEM is the most commonly used marginal effect approach to interpret the results of polychotomous dependent variable models in political science. While the MEM describes the average case, the AME portrays the average effect in the sample. According to Greene and Hensher (2010), the difference between these two marginal effects is likely to be small, especially in large samples (36). However, this is not a point of consensus in the literature.

There has been some debate about whether the AME and the MEM produce different quantities of interest (Bartus 2005; Verlinda 2006; Long and Freese 2014; Hanmer and Kalkan 2013). To show when these two approaches produce distinct estimates, scholars have focused on the case of binary logit and binary probit models (Bartus 2005; Verlinda 2006; Hanmer and Kalkan 2013). Essentially, the absolute difference between the AME and the MEM in the binary probit case can be calculated as (Bartus 2005, p. 312):[3]

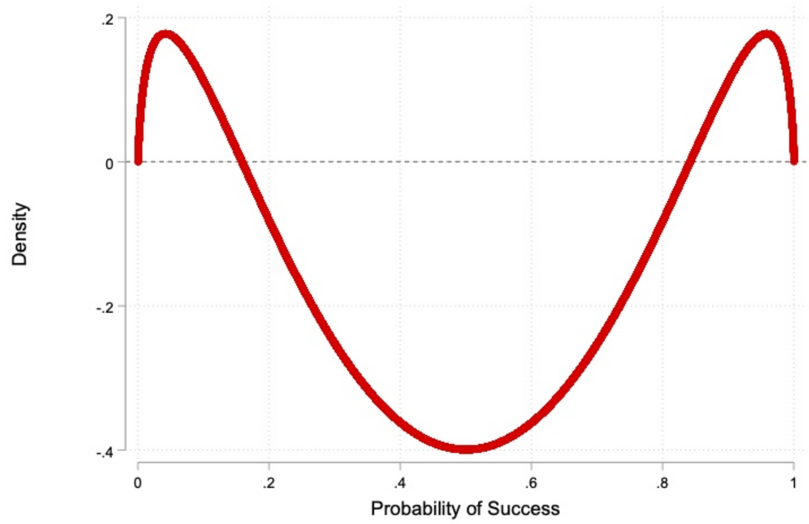
$$AME_i - MEM_i \approx \frac{1}{2}\beta_i f''(\beta\bar{x})Var(\beta\mathbf{x}),$$

where $f''(\beta\bar{x})Var(\beta\mathbf{x})$ refers to the second derivative of the density function evaluated at the sample mean and the sample variance.

[3] See Bartus (2005) and Hanmer and Kalkan (2013) for the full derivation of the difference between the two approaches.

The magnitude of the difference between the predictions calculated via AME and MEM will depend on the size of the coefficient estimates, the size of the variance of the linear predictions, and on the value of the second derivative of the probability density function (PDF) evaluated at the sample mean and the sample variance of the linear prediction (Verlinda 2006). Figure 30 shows the second derivative of the PDF for the binary probit. When the second derivative of the PDF is negative (i.e., the area of the curve below zero) or from 0.16 to 0.84, the MEM will be greater in absolute value than the AME. Alternatively, when the second derivative of the PDF is positive (i.e., the area of the curve above zero) or from 0 to 0.15 and from 0.85 to 1, the AME will be greater in absolute value than the MEM. As Figure 30 demonstrates, the largest difference between the two approaches is found when $Pr(y|\bar{x}) = 0.5$.[4]

Figure 30 – Second Derivative of the Binary Probit PDF by the Probability of Success for the MEM



There are four aspects that have been highlighted by the literature that make the AME approach superior to the MEM. First, the combination of values used in the estimation of the MEM (total sample average) may not exist in the population, thus the predictions might be representing an unrealistic case in the population (Long 1997; Hanmer and Kalkan 2013). This can be even more problematic when it is averaging dummy or categorical variables. Second, the AME tends to be more aligned with the theoretical arguments advanced in the articles. As Hanmer and Kalkan (2013) point out, "estimates from the observed-value approach [AME] connect directly to the original

[4] For the binary logit case, the second derivative of the logit PDF is negative from 0.21 to 0.79 (Hanmer and Kalkan 2013).

hypotheses and more easily allow researchers to evaluate the substantive implications of their theory (p.269)." Third, the AME is more robust to model misspecification than the MEM. Drawing on King and Zeng's (2006) contributions, Hanmer and Kalkan (2013) argue that "the farther one moves from the support of the data, the more sensitive to model misspecification the results will be. By definition, with the observed-value approach [AME], the values of all of the other independent variables will be on the support of the data; however, the mean [MEM] of all the other independent variables might not represent a case that is present or common in the population or sample (271)."

On the issue of model misspecification, Hanmer and Kalkan (2013) show that, in general, the MEM estimates tend to be more sensitive to bias due to model misspecification than the AME estimates. Their Monte Carlo simulation results show three interesting patterns.[5] First, when the correlation between x_2 and x_3 is equal to 0, the omission of a variable from the estimated probit model does not affect the predictions and marginal effects for the AME, but it *does* for the MEM. Even when the correlation is equal to 0, the MEM is biased. Second, in some cases, the bias from the AME and the MEM estimates are quite similar. Particularly the bias for the marginal effects when the correlation is equal to 0.5. Third, the bias for the MEM approach does not follow a clear pattern as correlation increases. For instance, while in some cases, the bias of the MEM estimates gets larger as the correlation increases, in other cases, it gets smaller as the correlation increases (especially the bias for the predicted effects). In sum, while Hanmer and Kalkan's (2013) findings show that the MEM estimates are more sensitive to bias due to models misspecification than the AME ones, it is not clear why the size of the bias does not vary with the level of correlation between the independent variables.

Another issue that has been highlighted by the literature that makes the AME approach superior to the MEM is related to model dependence. The MEM is more sensitive to model dependence and particularly to the problem of extreme counterfactuals (i.e., the risk of extrapolation) than the AME approach (King and Zeng 2006; Williams 2018). Extrapolation means that the counterfactual predicted probabilities are far from the observed outcomes. A consequence of the extrapolation issue is unreliable inferences. The AME mitigates such a problem since it is calculated based on the existing values in

[5] The authors simulated the following DGP $y* = 2 - x_1 + x_2 + 0.5x_3 + \varepsilon$. While x_1 is a categorical variable that takes the values 1, 2, and 3, x_2 and x_3 are continuous variables. To evaluate the bias due to model misspecification, the authors explore three scenarios: (i) when the correlation between x_2 and x_3 is equal to 0, (ii) the correlation is equal to 0.5, and (iii) the correlation is equal to 0.8.

the data set (Williams 2018). In contrast, the MEM is based on the average case that may not exist in the population. Although this issue is not easy to be identified, scholars have provided some recommendations to evaluate the quality of a study's counterfactual (King and Zeng 2006; Gelman and Pardoe 2007; Hanmer and Kalkan 2013; Williams 2018). A commonly recommended strategy is to visually compare the graph of the distribution of the marginal effects (or predicted probabilities) across the values of an independent variable and the histogram of such a variable (King and Zeng 2006; Williams 2018). Evidence of extrapolation will show up as areas that do not overlap in the two graphs. This task can be especially more laborious when the estimated model includes many explanatory variables.

In this section, I have briefly reviewed some of the most salient differences between the AME and the MEM approaches, mainly in the context of binary logit and binary probit models. However, it is not clear yet if these two approaches produce substantively different quantities of interest in the context of polychotomous dependent variable models. This is a relevant issue to investigate since the conclusions for the binary models may not fully apply to a more complex scenario with polychotomous dependent variable models. In the next section, I employ Monte Carlo simulations to show whether the AME and the MEM produce distinct estimates in the context of polychotomous dependent variable models. As I will show, the two approaches not only produce distinct quantities of interest, but also the MEM tends to be more sensitive to bias due to model misspecification.

4.6 Monte Carlo Simulations

I perform a series of Monte Carlo simulations to demonstrate the differences between the results obtained using the average marginal effect (AME) and the marginal effect at means (MEM) and the amount of bias due to model misspecification in the context of polychotomous dependent variable models.

The Monte Carlo simulations are based on 1,000 repetitions using 1,000 observations. The dependent variable, y, has three distinct outcomes. I start by generating the systematic component. First, I generate the log-odds ($x\beta$).

$$
\begin{aligned}
x\beta_1 &= ln(.5) + ln(1.5) * x_1 + ln(2.5) * x_2 + ln(3) * x_3 + ln(.7) * x_4 \\
x\beta_2 &= ln(2) + ln(.33) * x_1 + ln(.5) * x_2 + ln(5) * x_3 + ln(2.8) * x_4,
\end{aligned} \tag{9}
$$

where x_2 and x_3 are continuous variables generated from a normal distribution. x_1 was drawn from a uniform distribution but recoded into a dummy variable that assigns 1 when the original values were smaller than -1, and 0 otherwise. x_4 was drawn from a beta distribution (i.e., *beta(1,10)*, where 1 and 10 are the beta distribution shape parameters) but recoded into a dummy variable that assigns 1 when the original values were smaller than 0.1, and 0 otherwise.

Next, I generate the probabilities for the three outcomes (P_1, P_2, and P_3) based on the log-odds. Finally, I combine the systematic and the stochastic components:

$$
\begin{aligned}
y_i &= cond(u_i < P_1, 1, \\
&cond(u_i < P_1 + P_2, 2, 3)),
\end{aligned} \tag{10}
$$

where u_i is drawn from a uniform distribution.

According to Equation 10, y_i assumes value of 1 when $u_i < P_1$, value of 2 when $u_i < P_1 + P_2$, otherwise 3. Finally, I also allow the correlation between x_2 and x_3 to vary across the range from 0 to 0.9.

Table 6 summarizes the differences regarding the size of the marginal effects calculated via the AME and the MEM approaches for each of the dependent variable's categories. In line with the findings from the literature in binary probit (Hanmer and Kalkan 2013), Table 6 shows that overall the MEM produces larger estimates (in absolute value) than the AME approach.

Table 6 – Marginal Effects calculated via AME and MEM Approaches for the True Model

	Category 1			*Category 2*			*Category 3*		
	AME	MEM	Δ	AME	MEM	Δ	AME	MEM	Δ
x_1	0.109	0.159	0.050	0.123	0.132	0.009	-0.232	-0.291	0.059
x_2	0.050	0.086	0.036	0.128	0.118	0.011	-0.179	-0.204	0.025
x_3	-0.214	-0.261	0.047	0.001	-0.011	0.012	0.213	0.272	0.059
x_4	-0.107	-0.147	0.040	-0.105	-0.104	0.001	0.212	0.252	0.040

To evaluate the bias due to model misspecification, I estimated four different models, and in each model, I exclude one of the independent variables. For instance, model 1 omits x_1 from the estimation, model 2 omits x_2, and so on. Table 7 shows the average amount of bias (in absolute value) of the AME and MEM estimates across the four models. There are two aspects to highlight from this table. First, the MEM approach consistently produces a larger bias than the AME approach. Second, as expected, models 2 and 3 present the largest bias across the four models. This is an expected finding since, in the DGP, I have varied the level of correlation only between x_2 and x_3. As models 2 and 3 specifically estimate models omitting these variables, I expect the bias to be larger for those cases. In model 2, the omission of x_2 is expected to increase the amount of bias for the estimate of x_3 as the level of correlation between x_2 and x_3 increases.

Table 7 – Average amount of bias (in absolute value) of AME and MEM estimates across Models

Estimated model	Omitted variable from the model	AME	MEM
Model 1	x_1	0.000	0.006
Model 2	x_2	0.019	0.039
Model 3	x_3	0.022	0.048
Model 4	x_4	0.000	0.008

The only cases in which the amount of bias is approximately equal to zero are the marginal effects calculated via the AME approach for model 1 and model 4. In the case of the marginal effects estimated through the AME approach, the size of the bias increases as the correlation between x_2 and x_3 gets larger. In contrast, the marginal effects estimated via MEM are biased even when there is no correlation between x_2 and x_3. This means that the MEM estimates are biased regardless of the estimated model (model 1, 2, 3, or 4). Interestingly, the size of the bias of the MEM estimates does not vary as a function of the correlation between x_2 and x_3.

In sum, the simulation results suggest that, in general, the MEM estimates tend to be larger (in absolute value) than the ones calculated via AME. Additionally, the MEM estimates are more sensitive to model misspecification. While the bias of the AME estimates tends to vary as a function of the correlation between the independent variables, the pattern of bias in the marginal effects calculated via MEM is not a clear function of the correlation. Overall, these findings corroborate the argument that the AME estimates are less model-dependent than the MEM ones.

4.7 *Summary of Recommendations*

There are some relevant recommendations that practitioners should consider when conducting hypothesis testing with polychotomous dependent variable models. These suggestions may also apply to other limited dependent variable models in general.

- Be careful when interpreting the results from coefficient estimates. Given that these models are nonlinear in the parameters, the estimated coefficients' size, sign, and statistical significance might not translate into statistically and substantively significant marginal effects. If interpretation is based on regression coefficient estimates, there is a high chance of incur in incorrect inferences and scholars should report this possibility in discussing their findings.
- When evaluating whether the statistical results from hypotheses tests provide or do not support for the theoretical expectations, analysts should preferably employ more meaningful ways to interpret the estimates (e.g., marginal effects, predicted probabilities, first difference, etc.) rather than coefficient estimates.
- Given the nonlinearity of polychotomous dependent variable models, no single interpretation approach can fully describe the relationships between a key explanatory variable and the outcome probabilities. Therefore, scholars should undertake more than one approach and discuss why each approach provides complimentary insights that together contribute to more robust insights. In sum, as there is no perfect approach, researchers should rely on a combination of different interpretation approaches to evaluate their results more comprehensively.
- While the combination of values used in the estimation of the MEM (total sample average) may not exist in the population, the estimate calculated via AME might not correspond to the magnitude of the effect for anyone in the sample (Long and Freese 2014). Despite of that, some scholars have focused on the idea that the AME is the best approach for conducting theoretically driven hypothesis testing when evaluating results from limited dependent variable models (Hanmer and Kalkan 2013). In contrast, this study has shown that even the AME can be misleading, especially when the distribution of effects is skewed.
- To avoid extrapolation issues, a commonly recommended strategy is to visually compare the graph of the distribution of the marginal effects across the values of

an independent variable and the histogram of such a variable (King and Zeng 2006; Williams 2018).

4.8 Conclusion

This chapter contributes to the discussion about hypothesis testing in limited dependent variable models more generally. This study advocates that researchers should focus on using hypothesis testing approaches best suited to their theories instead of relying on a "one size fits all" approach. Despite the challenges for interpreting the results of nonlinear models, there are benefits from employing a combination of different strategies to analyze if the results corroborate with one's theoretical expectations.

While most of the literature has focused on binomial logit and probit models when evaluating the differences between AME and MEM (Hanmer and Kalkan 2013; Bartus 2005; Verlinda 2006), in this chapter, I present novel evidence about this issue in the context of polychotomous dependent variable models. The results from the Monte Carlo simulations show that the AME and the MEM approaches produce distinct quantities of interest. In addition, the MEM tends to be more sensitive to bias due to model misspecification even when there is no correlation between the covariates. This is a relevant concern that scholars should take into account when deciding which interpretation approach to choosing. Interestingly, in the presence of model misspecification, it is not clear why the bias in the AME varies according to the correlation between the covariates, while the same does not hold for the MEM. Further studies should investigate why the MEM is more sensitive to bias even when there is no correlation between the covariates.

5 CONCLUSIONS

The three preceding chapters addressed research questions in distinct lines of inquiry and provided substantive and methodological contributions to those areas. In the following paragraphs, I will summarize the main findings from each chapter.

In Chapter 2, I tested whether partisan dealignment contributes to the observed dynamics on the incumbency advantage. To do so, I employed a research design that proposes a new strategy for studying incumbency advantage over time. In the first stage, I used a series of cross-sectional analyses to generate a continuous measure of incumbency effects that varies across elections. In the second stage, I then employed this estimated incumbency effect across districts in each election as the dependent variable to examine the impact of the partisan dealignment theory on incumbency advantage. Using data from the U.S. House elections from 1948 to 2014, I tested four hypotheses that involve distinct aspects of the partisan dealignment theory. By systematically testing these four theoretical propositions, I provide robust statistical evidence that corroborates that shifts in dealignment impact the incumbency advantage in all four explanatory variables used to measure partisan dealignment.

In Chapter 3, I evaluated the gender gap on presidential approval during the COVID-19 pandemic. Using survey data for Brazil and the United States, I showed that women and men hold distinct views regarding the seriousness of the COVID-19 pandemic. Women are more likely than men to assess the severity of the virus and its effects as much worse than a flu. The evidence further suggests that these differences reflect on their attitude and behavior to prevent the disease and their behavior concerning social distancing. Women are more likely to obey public health recommendations, especially regarding the use of masks outside the home. Based on these findings, I argued that women would punish the incumbent president to a greater degree than men as the COVID-19 death toll increases. Using presidential approval rating data for Brazil and the United States, the results suggest that men and women punish the president for mismanaging the crisis. Contrary to my expectation, I did not find evidence suggesting that women's presidential approval rating is more sensitive to changes in the death toll.

In Chapter 4, I showed the relevance of using a combination of different approaches when evaluating the results from polychotomous dependent variable models.

Given the nonlinearity of these models, there is no single interpretation approach that completely describes the relationship between a variable and the outcome probability in a multivariate case. Instead of relying on a single approach for all situations, I argued that the use of different approaches provides a more comprehensive way to interpret the results on several occasions. To exemplify this point, I conducted a replication of part of the analysis undertaken by an article that estimated a multinomial logit model. The results from the replication challenged some of the conclusions made by the original authors. They provided an interesting case to show the insights gained when researchers use a combination of interpretation approaches to test their hypotheses. In Chapter 4, I also conducted a survey of the literature to evaluate what are the most commonly used interpretation strategies in the empirical literature that employs polychotomous dependent variable models. The survey results highlight that there are considerable improvements that researchers can make to improve their analyses by adopting more meaningful ways to interpret their results. In line with the literature in binary probit, the results of the Monte Carlo simulations showed that the marginal effect at means and average marginal effect approaches produce distinct quantities of interest in the context of polychotomous dependent variable models, and the former is more sensitive to bias due to model misspecification.

In summary, the first two chapters speak to the opportunities to answer challenging research questions involving time series. In contrast, the third chapter emphasizes how substantive and statistical communication of the results' interpretation can improve when researchers combine different approaches to evaluate the results from polychotomous dependent variable models.

Bibliography

Adolph, Christopher, Kenya Amano, Bree Bang-Jensen, Nancy Fullman and John Wilkerson. 2021. "Pandemic politics: Timing state-level social distancing responses to COVID-19." *Journal of Health Politics, Policy and Law* 46(2):211–233.

Agência Câmara de Notícias. 2020. Mulheres são mais afetadas por crise econômica provocada pela Covid-19. https://www.camara.leg.br/noticias/681393-mulheres-sao-mais-afetadas-por-crise-economica-provocada-pela-covid-19/. Accessed: 2021-07-03.

Ai, Chunrong and Edward C Norton. 2003. "Interaction terms in logit and probit models." *Economics Letters* 80(1):123–129.

Alesina, Alberto and Howard Rosenthal. 1995. *Partisan Politics, Divided Government, and the Economy*. Cambridge University Press.

Alford, John and David W Brady. 1988. "Partisan and incumbent advantage in US House elections, 1846-1986." *Center for the Study of Institution and Values, Rice University* .

Algara, Carlos, Sam Fuller and Christopher Hare. 2020. "The Conditional Effect of Scientific Knowledge and Gender on Support for COVID-19 Government Containment Policies in a Partisan America." *Politics & Gender* 16(4):1075–1083.

Algara, Carlos, Sam Fuller, Christopher Hare and Sara Kazemian. 2021. "The interactive effects of scientific knowledge and gender on COVID-19 social distancing compliance." *Social Science Quarterly* 102(1):7–16.

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll and Michèle Tertilt. 2021. From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions. Technical report National Bureau of Economic Research.

Alvarez, R Michael and Jonathan Nagler. 1998. "When politics and models collide: Estimating models of multiparty elections." *American Journal of Political Science* 42(1):56–96.

Ansolabehere, Stephen and James M Snyder Jr. 2002. "The Incumbency Advantage in US Elections: An Analysis of State and Federal Offices, 1942–2000." *Election Law Journal* 1(3):315–338.

Ariga, Kenichi. 2015. "Incumbency Disadvantage under Electoral Rules with Intraparty Competition: Evidence from Japan." *The Journal of Politics* 77(3):874–887.

Asteriou, Dimitrios and Stephen G Hall. 2015. *Applied Econometrics*. Macmillan International Higher Education.

Barberia, Lorena G, Luiz GR Cantarelli, Maria Leticia Claro de Faria Oliveira, Natália de Paula Moreira and Isabel Seelaender Costa Rosa. 2021. "The effect of state-level social distancing policy stringency on mobility in the states of Brazil." *Revista de Administração Pública* 55(1):27–49.

Bartus, Tamás. 2005. "Estimation of marginal effects using margeff." *The Stata Journal* 5(3):309–329.

BBC News Brasil. 2018. EleNão: A manifestação histórica liderada por mulheres no Brasil vista por quatro ângulos. https://www.bbc.com/portuguese/brasil-45700013. Accessed: 2021-06-28.

Bennouna, Cyril, Agustina Giraudy, Eduardo Moncada, Eva Rios, Richard Snyder and Paul Testa. 2021. "Pandemic Policymaking in Presidential Federations: Explaining Subnational Responses to Covid-19 in Brazil, Mexico, and the United States." *Publius: The Journal of Federalism* 51(4):570–600.

Berry, William D, Jacqueline HR DeMeritt and Justin Esarey. 2010. "Testing for interaction in binary logit and probit models: is a product term essential?" *American Journal of Political Science* 54(1):248–266.

Box-Steffensmeier, Janet M, Suzanna De Boef and Tse-Min Lin. 2004. "The dynamics of the partisan gender gap." *American Political Science Review* pp. 515–528.

Brooks, Deborah Jordan and Lydia Saad. 2020. "Double Whammy: Why the Underrepresentation of Women among Workplace and Political Decision Makers Matters in Pandemic Times." *Politics & Gender* 16(4):1110–1122.

Burden, Barry C and Amber Wichowsky. 2014. "Economic discontent as a mobilizer: unemployment and voter turnout." *The Journal of Politics* 76(4):887–898.

Burnham, Walter Dean. 1974. "Communication." *American Political Science Review* 68:207–213.

Caixa. 2021. "Caixa Management Report 4Q2O.".

Cameron, Adrian Colin and Pravin K Trivedi. 2010. *Microeconometrics using Stata*. Vol. 2 Stata Press College Station, TX.

Carreras, Miguel, Sofia Vera and Giancarlo Visconti. 2021. "Who does the caring? Gender disparities in COVID-19 attitudes and behaviors." *Politics & Gender* .

Carroll, Susan J. 2006. "Voting choices: Meet you at the gender gap." *Gender and elections: Shaping the future of American politics* pp. 74–96.

Carson, Jamie L, Joel Sievert and Ryan D Williamson. 2019. "Nationalization and the Incumbency Advantage." *Political Research Quarterly* .

Cassino, Dan and Yasemin Besen-Cassino. 2020. "Of Masks and Men? Gender, Sex, and Protective Measures during COVID-19." *Politics & Gender* 16(4):1052–1062.

Caughey, Devin and Jasjeet S Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008." *Political Analysis* 19(4):385–408.

Chaney, Carole Kennedy, R Michael Alvarez and Jonathan Nagler. 1998. "Explaining the gender gap in US presidential elections, 1980-1992." *Political Research Quarterly* 51(2):311–339.

Clarke, Harold D, Marianne C Stewart, Mike Ault and Euel Elliott. 2005. "Men, women and the dynamics of presidential approval." *British Journal of Political Science* pp. 31–51.

Clarke, Harold, Marianne Stewart, Paul Whiteley and Guy D Whitten. 2020. "The large gender gap in Trump's support threatens his reelection. Here's what's driving it." *USApp–American Politics and Policy Blog* .

CNN. 2020. US coronavirus cases surpass 1 million and the death toll is greater than US losses in Vietnam War. https://edition.cnn.com/2020/04/28/health/us-coronavirus-tuesday/index.html. Accessed: 2021-11-10.

Cover, Albert D. 1977. "One Good Term Deserves Another: The Advantage of Incumbency in Congressional Elections." *American Journal of Political Science* pp. 523–541.

Cover, Albert D and David R Mayhew. 1977. Congressional dynamics and the decline of competitive congressional elections. In *Congress reconsidered*. Praeger New York pp. 62–82.

Cox, Gary W and Jonathan N Katz. 1996. "Why Did the Incumbency Advantage in US House Elections Grow?" *American Journal of Political Science* 40(2):478–497.

Datafolha. 2018. Bolsonaro chega à véspera da eleição com 55% dos votos válidos. October 28th, 2018. https://datafolha.folha.uol.com.br/eleicoes/2018/10/1983786-bolsonaro-chega-a-vespera-da-eleicao-com-55-dos-votos-validos.shtml. Accessed: 2021-06-28.

Datafolha. 2020. Compra de alimentos é principal gasto de auxílio emergencial. https://datafolha.folha.uol.com.br/opiniaopublica/2020/08/1988833-compra-de-alimentos-e-principal-gasto-de-auxilio-emergencial.shtml. Accessed: 2021-07-03.

Dong, E, H Du and L Gardner. 2021. "COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)." *Johns Hopkins University. Available online: https://coronavirus.jhu.edu/map.html (accessed on September 2, 2021)* .

Duch, Raymond M and Randy Stevenson. 2006. "Assessing the Magnitude of the Economic Vote over time and across nations." *Electoral Studies* 25(3):528–547.

Duque, Daniel. 2020. Auxílio emergencial para de crescer em agosto, e pobreza cai de novo; com redução do benefício a partir de setembro, porém, pobreza e desigualdade devem aumentar. https://blogdoibre.fgv.br/posts/auxilio-emergencial-para-de-crescer-em-agosto-e-pobreza-cai-de-novo-com-reducao-do-beneficio. Accessed: 2021-07-02.

Edlund, Lena and Rohini Pande. 2002. "Why have women become left-wing? The political gender gap and the decline in marriage." *The Quarterly Journal of Economics* 117(3):917–961.

Eggers, Andrew C, Anthony Fowler, Jens Hainmueller, Andrew B Hall and James M Snyder Jr. 2015. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races." *American Journal of Political Science* 59(1):259–274.

Eggers, Andrew C and Arthur Spirling. 2017. "Incumbency effects and the strength of party preferences: Evidence from multiparty elections in the United Kingdom." *The Journal of Politics* 79(3):903–920.

El País Brasil. 2018. Um milhão de mulheres contra Bolsonaro: a rejeição toma forma nas redes. https://brasil.elpais.com/brasil/2018/09/12/actualidad/1536768048$_3$21164*.html.Accessed* : 2021 – 10 – 27.

Enders, Walter. 2008. *Applied Econometric Time Series*. John Wiley & Sons.

Erikson, Robert S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3(3):395–405.

Erikson, Robert S. 1972. "Malapportionment, Gerrymandering, and Party Fortunes in Congressional Elections." *American Political Science Review* 66(4):1234–1245.

Erikson, Robert S. 2016. "The Congressional Incumbency Advantage over Sixty Years." *Governing in a Polarized Age: Elections, Parties, and Political Representation in America* p. 65.

Erikson, Robert S and Rocío Titiunik. 2015. "Using Regression Discontinuity to Uncover the Personal Incumbency Advantage." *Quarterly Journal of Political Science* 10(1):101–119.

Fares, Lygia Sabbag, Ana Luíza Matos de Oliveira, Luísa Cardoso and Luiza Nassif-Pires. 2021. "As políticas econômicas implementadas no Brasil durante a pandemia sob a perspectiva de gênero." *Nota de Política Econômica - número 006* .

Ferejohn, John A. 1977. "On the Decline of Competition in Congressional Elections." *American Political Science Review* 71(1):166–176.

Fiorina, Morris P. 1977. "The Case of the Vanishing Marginals: The Bureaucracy Did It." *The American Political Science Review* 71(1):177–181.

Fiorina, Morris P and Samuel J Abrams. 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11:563–588.

Folha. 2021. Pandemia deixa mais da metade das mulheres fora do mercado de trabalho. https://www1.folha.uol.com.br/mercado/2021/02/pandemia-deixa-mais-da-metade-das-mulheres-fora-do-mercado-de-trabalho.shtml. Accessed: 2021-07-03.

Galasso, Vincenzo, Vincent Pons, Paola Profeta, Michael Becher, Sylvain Brouard and Martial Foucault. 2020. "Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries." *Proceedings of the National Academy of Sciences* 117(44):27285–27291.

Gallup. 2020*a*. More Americans Cite COVID-19 as Most Important U.S. Problem. https://news.gallup.com/poll/316751/americans-cite-covid-important-problem.aspx. Accessed: 2021-05-21.

Gallup. 2020*b*. Gallup Blog: The COVID-19 Responses of Men vs. Women. https://news.gallup.com/opinion/gallup/321698/covid-responses-men-women.aspx. Accessed: 2021-10-15.

Gelman, Andrew. 2005. "Estimating the incumbent-party advantage and the incumbency advantage in House elections." https://statmodeling.stat.columbia.edu/2005/10/14/estimating_the_1/. Accessed: 2021-05-03.

Gelman, Andrew and Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* pp. 1142–1164.

Gelman, Andrew and Iain Pardoe. 2007. "Average predictive comparisons for models with nonlinearity, interactions, and variance components." *Sociological Methodology* 37(1):23–51.

Gelman, Andrew and Zaiying Huang. 2008. "Estimating Incumbency Advantage and its Variation, as an Example of a Before–After Study." *Journal of the American Statistical Association* 103(482):437–446.

Gilens, Martin. 1988. "Gender and support for Reagan: A comprehensive model of presidential approval." *American Journal of Political Science* pp. 19–49.

Giurge, Laura M, Ashley V Whillans and Ayse Yemiscigil. 2021. "A multicountry perspective on gender differences in time use during COVID-19." *Proceedings of the National Academy of Sciences* 118(12).

Glasgow, Garrett. 2001. "Mixed logit models for multiparty elections." *Political Analysis* 9(2):116–136.

Global Health 50/50. 2021. COVID-19 Sex-Disaggregated Data Tracker. https://globalhealth5050.org/the-sex-gender-and-covid-19-project/the-data-tracker/. Accessed: 2021-09-16.

Granger, Clive and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2(2):111–120.

Greene, William H. 2008. *Econometric analysis*. Vol. 6th edition Pearson Education, Inc.

Greene, William H and David A Hensher. 2010. *Modeling ordered choices: A primer*. Cambridge University Press.

Hanmer, Michael J and Kerem Ozan Kalkan. 2013. "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *American Journal of Political Science* 57(1):263–277.

Hanushek, Eric A and John E Jackson. 1977. *Statistical methods for social scientists*. Academic Press.

Herron, Michael C. 1999. "Postestimation uncertainty in limited dependent variable models." *Political Analysis* 8(1):83–98.

Higgins, Alison and Paul M Kellstedt. 2016. *Bankers or Bankers? Exploring Gender Differences in Approval Dynamics During the Obama Presidency*. Working paper.

Hsiang, Solomon, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich et al. 2020. "The effect of large-scale anti-contagion policies on the COVID-19 pandemic." *Nature* 584(7820):262–267.

Imai, Kosuke, Gary King and Olivia Lau. 2009. "Zelig: Everyone's statistical software." *R package version* 3(5).

Jackson, John E. 2002. "A seemingly unrelated regression model for analyzing multiparty elections." *Political Analysis* 10(1):49–65.

Jacobson, Gary C. 2015. "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections." *The Journal of Politics* 77(3):861–873.

Jacobson, Gary C and Jamie L Carson. 2019. *The Politics of Congressional Elections*. Rowman & Littlefield Publishers.

Johnston, Regan M, Anwar Mohammed and Clifton Van Der Linden. 2020. "Evidence of Exacerbated Gender Inequality in Child Care Obligations in Canada and Australia During the COVID-19 Pandemic." *Politics & Gender* pp. 1–16.

Jordan, Soren and Andrew Q Philips. 2018. "Cointegration Testing and Dynamic Simulations of Autoregressive Distributed Lag Models." *The Stata Journal* 18(4):902–923.

Kam, Cindy D. 2009. "Gender and economic voting, revisited." *Electoral Studies* 28(4):615–624.

Katz, Jonathan N and Gary King. 1999*a*. "A statistical model for multiparty electoral data." *American Political Science Review* pp. 15–32.

Katz, Jonathan N and Gary King. 1999*b*. "A statistical model for multiparty electoral data." *American Political Science Review* 93(1):15–32.

Kaufmann, Karen M and John R Petrocik. 1999. "The changing politics of American men: Understanding the sources of the gender gap." *American Journal of Political Science* pp. 864–887.

Kellstedt, Paul M, David AM Peterson and Mark D Ramirez. 2010. "The macro politics of a gender gap." *Public opinion quarterly* 74(3):477–498.

King, Gary. 1989. *Unifying political methodology: The likelihood theory of statistical inference*. Cambridge University Press.

King, Gary. 1991. "Constituency Service and Incumbency Advantage." *British Journal of Political Science* 21(1):119–128.

King, Gary and Andrew Gelman. 1991. "Systemic Consequences of Incumbency Advantage in US House Elections." *American Journal of Political Science* pp. 110–138.

King, Gary and Langche Zeng. 2006. "The dangers of extreme counterfactuals." *Political Analysis* 14(2):131–159.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the most of statistical analyses: Improving interpretation and presentation." *American Journal of Political Science* pp. 347–361.

Klašnja, Marko and Rocio Titiunik. 2017. "The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability." *American Political Science Review* 111(1):129–148.

Krehbiel, Keith and John R Wright. 1983. "The Incumbency Effect in Congressional Elections: A Test of Two Explanations." *American Journal of Political Science* pp. 140–157.

Kugler, Maurice David, Mariana Viollaz, Daniel Vasconcellos Archer Duque, Isis Gaddis, David Locke Newhouse, Amparo Palacios-Lopez, Michael Weber et al. 2021. How Did the COVID-19 Crisis Affect Different Types of Workers in the Developing World? Technical report The World Bank.

Lee, David S. 2008. "Randomized Experiments from Non-random Selection in US House Elections." *Journal of Econometrics* 142(2):675–697.

LeVeck, Brad L and Stephanie A Nail. 2016. "Evidence for a Scale Invariant Relationship between the Incumbency Advantage and the Nationalization of US House Elections 1866–2014." *Research & Politics* 3(4).

Levendis, John D. 2018. Structural Breaks. In *Time Series Econometrics*. Springer pp. 171–196.

Lewis-Beck, Michael S and Mary Stegmaier. 2000. "Economic determinants of electoral outcomes." *Annual Review of Political Science* 3(1):183–219.

Lewis, Jeffrey B and Drew A Linzer. 2005. "Estimating Regression Models in which the Dependent Variable is Based on Estimates." *Political Analysis* 13(4):345–364.

Lewis, Jeffrey B, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2019. "Voteview: Congressional Roll-Call Votes Database." *See https://voteview.com/. Accessed 25 March 2021* .

Lipscy, Phillip Y. 2020. "COVID-19 and the Politics of Crisis." *International Organization* 74(S1):E98–E127.

Long, J Scott. 1997. *Regression models for categorical and limited dependent variables*. Vol. 7 Sage.

Long, J Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables using Stata*. Third Edition. Stata Press.

Lustig, Nora and Mart Trasberg. 2021. "How Brazil and Mexico Diverged on Social Protection in the Pandemic." *Current History* 120(823):57–63.

Mattei, Franco. 2000. "The gender gap in presidential evaluations: Assessments of Clinton's performance in 1996." *Polity* 33(2):199–228.

May, Ann Mari and Kurt Stephenson. 1994. "Women and the great retrenchment: the political economy of gender in the 1980s." *Journal of Economic Issues* 28(2):533–542.

Mayhew, David R. 1974*a*. *Congress: The Electoral Connection*. Vol. 26 Yale University Press.

Mayhew, David R. 1974*b*. "Congressional Elections: The Case of the Vanishing Marginals." *Polity* 6(3):295–317.

Mikhailov, Nikolai, Richard G Niemi and David L Weimer. 2002. "Application of Theil group logit methods to district-level vote shares: tests of prospective and retrospective voting in the 1991, 1993, and 1997 Polish elections." *Electoral Studies* 21(4):631–648.

Ministério da Cidadania. 2020. Auxílio Emergencial 2020. https://aplicacoes.mds.gov.br/sagi/vis/data3/index.php?g=2. Accessed: 2021-07-02.

Mize, Trenton D. 2019. "Best practices for estimating, interpreting, and presenting nonlinear interaction effects." *Sociological Science* 6:81–117.

Mize, Trenton D, Long Doan and J Scott Long. 2019. "A general framework for comparing predictions and marginal effects across models." *Sociological Methodology* 49(1):152–189.

Mueller, John E. 1970. "Presidential popularity from Truman to Johnson." *The American Political Science Review* 64(1):18–34.

Nannestad, Peter and Martin Paldam. 1994. "The VP-function: A survey of the literature on vote and popularity functions after 25 years." *Public Choice* 79(3):213–245.

Palmer, Carl L and Rolfe D Peterson. 2020. "Toxic Mask-ulinity: The link between masculine toughness and affective reactions to mask wearing in the COVID-19 era." *Politics & Gender* 16(4):1044–1051.

Paolino, Philip. 2020. "Predicted Probabilities and Inference with Multinomial Logit." *Political Analysis* pp. 1–6.

Peckham, Hannah, Nina M de Gruijter, Charles Raine, Anna Radziszewska, Coziana Ciurtin, Lucy R Wedderburn, Elizabeth C Rosser, Kate Webb and Claire T Deakin. 2020. "Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission." *Nature Communications* 11(1):1–10.

Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica: Journal of the Econometric Society* pp. 1361–1401.

Pesaran, M Hashem, Yongcheol Shin and Richard J Smith. 2001. "Bounds Testing Approaches to the Analysis of Level Relationships." *Journal of Applied Econometrics* 16(3):289–326.

Peskowitz, Zachary. 2019. "Ideological Signaling and Incumbency Advantage." *British Journal of Political Science* 49(2):467–490.

Petersen, Trond. 1985. "A comment on presenting results from logit and probit models." *American Sociological Review* 50(1):130–131.

Petherick, Anna, Rafael Goldzmidt, Beatriz Kira and Lorena Barberia. 2020. "As medidas governamentais adotadas em resposta ao COVID-19 no Brasil atendem aos critérios da OMS para flexibilização de restrições." *Universidade de Oxford* .

Pew Research. 2016. Behind Trump's victory: Divisions by race, gender, education. https://www.pewresearch.org/fact-tank/2016/11/09/behind-trumps-victory-divisions-by-race-gender-education. Accessed: 2021-10-27.

Philips, Andrew Q. 2018. "Have Your Cake and Eat It Too? Cointegration and Dynamic Inference from Autoregressive Distributed Lag Models." *American Journal of Political Science* 62(1):230–244.

Philips, Andrew Q, Amanda Rutherford and Guy D Whitten. 2016. "Dynamic pie: A strategy for modeling trade-offs in compositional variables over time." *American Journal of Political Science* 60(1):268–283.

Schlesinger, Mark and Caroline Heldman. 2001. "Gender gap or gender gaps? New perspectives on support for government action and policies." *Journal of Politics* 63(1):59–92.

Shapiro, Robert Y and Harpreet Mahajan. 1986. "Gender Differences in Policy Preferences: A Summary of Trends from the 1960s to the 1980s." *Public Opinion Quarterly* 50(1):42–61.

Stonecash, Jeffrey M. 2008. *Reassessing the Incumbency Effect*. Cambridge University Press New York.

Tomz, Michael, Jason Wittenberg and Gary King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* .

Tomz, Michael, Joshua A Tucker and Jason Wittenberg. 2002. "An easy and accurate regression model for multiparty electoral data." *Political Analysis* pp. 66–83.

University of Michigan. 2021. Index of Consumer Sentiment. Surveys of Consumers. http://www.sca.isr.umich.edu/tables.html. Accessed: 2021-08-04.

UN Women. 2020. UN Women Rapid Assessment Survey on the socio-economic consequences of COVID-19 on women's and men's economic empowerment (23 April 2020). https://data.unwomen.org/resources/surveys-show-covid-19-has-gendered-effects-asia-and-pacific. Accessed: 2021-05-21.

Uppal, Yogesh. 2009. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Public Choice* 138(1-2):9–27.

U.S. Bureau of Labor Statistics. 2021. Unemployment Rate - Men [LNS14000001], retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/LNS14000001. Accessed: 2021-11-04.

Verlinda, Jeremy A. 2006. "A comparison of two common approaches for estimating marginal effects in binary choice models." *Applied Economics Letters* 13(2):77–80.

Warshaw, Christopher, Lynn Vavreck and Ryan Baxter-King. 2020. "The effect of local COVID-19 fatalities on Americans' political preferences." *Science Advances* .

Welch, Susan and John Hibbing. 1992. "Financial conditions, gender, and voting in American national elections." *The Journal of Politics* 54(1):197–213.

Whitten, Guy D and Harvey D Palmer. 1996. "Heightening comparativists' concern for model choice: Voting behavior in Great Britain and the Netherlands." *American Journal of Political Science* 40(1):231.

Williams, Laron K. 2013. "Flexible Election Timing and International Conflict." *International Studies Quarterly* 57(3):449–461.

Williams, Laron K. 2018. "Temporal Dependence and the Sensitivity of Quantities of Interest: A Solution for a Common Problem." *International Studies Quarterly* 62(4):892–902.

Williams, Richard. 2012. "Using the margins command to estimate and interpret adjusted predictions and marginal effects." *Stata Journal* 12(2):308.

Yam, Kai Chi, Joshua Conrad Jackson, Christopher M Barnes, Jenson Lau, Xin Qin and Hin Yeung Lee. 2020. "The rise of COVID-19 cases is associated with support for world leaders." *Proceedings of the National Academy of Sciences* 117(41):25429–25433.

Zivot, Eric and Donald W K Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business & Economic Statistics* 20(1):25–44.

APPENDIX A –

Table 8 – Summary statistics

Election Year	N Districts	Average % of Dem. Votes	Seats Dem. Party Won	Seats Rep. Party Won	Seats Dem. Inc. Won	Seats Rep. Inc. Won	Open Seats	Seats Inc. Won	% Seats Inc. Won
1946	347	46.9	167	180	135	162	50	297	85.6
1948	346	54.4	116	229	96	208	42	304	87.9
1950	331	50.1	169	162	156	142	33	298	90.0
1952	337	47.9	142	185	119	160	55	279	82.8
1954	344	51.2	128	215	117	200	26	317	92.2
1956	358	49.9	159	198	147	181	29	328	91.6
1958	335	54.9	135	198	125	167	43	292	87.2
1960	356	53.5	206	150	193	131	32	324	91.0
1962	372	52.2	183	161	173	143	49	316	84.9
1964	390	57.1	207	176	189	159	42	348	89.2
1966	376	50.3	234	136	221	125	29	346	92.0
1968	385	50.8	200	176	185	163	33	348	90.4
1970	372	53.2	186	184	169	164	38	333	89.5
1972	380	51.8	196	163	175	144	58	319	83.9
1974	375	57.2	188	186	161	162	52	323	86.1
1976	383	56.2	245	138	210	122	51	332	86.7
1978	365	54.5	237	128	203	109	53	312	85.5
1980	380	51.8	236	144	210	128	42	338	88.9
1982	378	54.6	183	166	163	152	57	315	83.3
1984	367	50.5	213	154	201	140	26	341	92.9
1986	360	53.8	196	164	177	142	41	319	88.6
1988	354	52.7	196	158	185	143	26	328	92.7
1990	349	55.0	210	139	199	121	29	320	91.7
1992	401	53.8	229	145	192	120	85	312	77.8
1994	382	49.3	239	143	208	122	52	330	86.4
1996	408	51.2	186	221	157	200	51	357	87.5
1998	340	52.2	167	173	150	156	34	306	90.0
2000	370	51.9	179	191	170	166	34	336	90.8
2002	352	49.9	162	168	151	151	46	302	85.8
2004	366	51.3	174	189	161	172	31	333	91.0
2006	376	53.3	157	219	145	198	33	343	91.2
2008	379	55.3	195	184	188	156	35	344	90.8
2010	408	49.5	252	156	233	134	41	367	90.0
2012	387	51.7	159	205	138	187	60	325	84.0
2014	358	47.3	160	198	143	174	41	317	88.5

Table 9 – Summary statistics of the entire Erikson and Titiunik (2015)'s dataset and the sub-sample used in their RDD analysis

Election Year	Erikson and Titiunik (2015)'s Entire dataset **Number of Districts**	Sub-sample used **Number of Districts**	Erikson and Titiunik (2015)'s Entire dataset **Average % of Dem. Votes** (1)	Sub-sample used **Average % of Dem. Votes** (2)	**Difference** (1 - 2)
1968	383	0	50.8		
1970	372	10	53.3	57.6	-4.3
1972	381	5	51.9	43.4	8.5
1974	376	28	57.2	49.3	7.9
1976	383	37	56.2	55.0	1.2
1978	370	32	54.5	58.9	-4.4
1980	385	32	51.7	46.7	5.1
1982	378	1	54.8	72.1	-17.3
1984	369	19	50.5	48.1	2.4
1986	362	13	53.9	49.5	4.4
1988	358	22	52.7	49.1	3.7
1990	354	16	54.9	50.0	4.9
1992	399	1	53.6	65.0	-11.4
1994	389	46	49.2	47.8	1.3
1996	404	29	51.5	51.8	-0.3
1998	340	24	52.3	49.4	2.9
2000	371	27	51.8	52.4	-0.6
2002	354	1	49.8	33.6	16.3
2004	370	23	50.9	46.5	4.5
2006	380	13	53.2	56.1	-3.0
2008	379	20	55.2	57.0	-1.8

Table 10 – Structural Break Test with a know break point

Year	Wald test χ^2	LR test χ^2
1954	0.47	0.53
1956	1.88	2.07
1958	3.21	3.45
1960	4.06	4.32
1962	3.55	3.81
1964	6.69	6.84
1966	10.97	10.59
1968	0.87	0.98
1970	1.26	1.40
1972	0.81	0.91
1974	0.32	0.36
1976	0.95	1.06
1978	0.55	0.62
1980	0.05	0.06
1982	0.33	0.37
1984	0.13	0.15
1986	0.05	0.06
1988	0.72	0.80
1990	1.21	1.35
1992	0.28	0.31
1994	0.29	0.33
1996	0.95	1.06
1998	0.28	0.31
2000	0.94	1.05
2002	1.47	1.62
2004	1.31	1.46
2006	1.14	1.27
2008	1.08	1.20

Note: For each one of these 28 election years (from 1954 to 2008), I performed a Wald test for a structural break at a known break date for the estimation results of a linear regression of $\widehat{Incumbency}_t$ on $\widehat{Incumbency}_{t-1}$. This routine was implemented in Stata using the command *estat sbknown*.

Table 11 – Perron's tests for a unit root, 1946 to 2014

	(1)	(2)
	y_t	y_t
D_L	0.183	11.18**
	(0.07)	(2.35)
time	-0.0777	0.939**
	(-1.06)	(2.42)
D_P	-0.174	-2.325
	(-0.06)	(-0.80)
y_{t-1}	0.878**	0.516
	(2.39)	(1.46)
Δy_{t-1}	-0.473	-0.0940
	(-1.30)	(-0.27)
Δy_{t-2}	-0.449*	-0.233
	(-1.75)	(-0.97)
Δy_{t-3}	-0.000373	0.107
	(-0.00)	(0.55)
New Slope		-1.046**
		(-2.66)
Intercept	2.482	-4.711
	(1.52)	(-1.54)
N	31	31
R^2	0.547	0.657

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table 12 – Unit-root tests for ***Incumbency Advantage$_t$***

Unit Root Test	*Average* Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.55	-2.98	I(1)
Augmented Dickey-Fuller (with trend)	-1.93	-3.57	I(1)
Phillips-Perron (with trend)	-2.71	-3.56	I(1)
Dickey-Fuller GLS	-1.82	-3.29	I(1)
Elliott-Rothenberg-Stock	-1.82	-3.19	I(1)
Kwiatkowski-Phillips-Schmidt-Shin	0.36	0.15	I(1)

Table 13 – Unit-root tests for Δ***Incumbency Advantage$_t$***

Unit Root Test	*Average* Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-6.37	-2.98	I(0)
Augmented Dickey-Fuller (with trend)	-7.13	-3.57	I(0)
Phillips-Perron (with trend)	-8.95	-3.57	I(0)
Dickey-Fuller GLS	-6.88	-3.30	I(0)
Elliott-Rothenberg-Stock	-6.88	-3.19	I(0)
Kwiatkowski-Phillips-Schmidt-Shin	0.03	0.15	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 1 lag. The test statistics reported above are the average of the test statistics of each test conducted for the 1,000 simulations of *Incumbency Advantage$_t$*.

Table 14 – Unit-root tests for $\boldsymbol{Polarization_t}$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	1.37	-2.98	I(1)
Augmented Dickey-Fuller (with trend)	-1.23	-3.57	I(1)
Phillips-Perron (with trend)	-2.24	-3.56	I(1)
Dickey-Fuller GLS	-0.81	-3.29	I(1)
Elliott-Rothenberg-Stock	-0.81	-3.19	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.42	0.15	I(1)

Table 15 – Unit-root tests for $\Delta\boldsymbol{Polarization_t}$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.51	-2.98	I(0)
Augmented Dickey-Fuller (with trend)	-4.32	-3.57	I(0)
Phillips-Perron (with trend)	-7.28	-3.57	I(0)
Dickey-Fuller GLS	-3.66	-3.30	I(0)
Elliott-Rothenberg-Stock	-3.66	-3.19	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.06	0.15	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 1 lag.

Table 16 – Unit-root tests for $\boldsymbol{Independents_t}$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.90	-3.00	I(1)
Augmented Dickey-Fuller (with trend)	-1.23	-3.60	I(1)
Phillips-Perron (with trend)	-2.14	-3.60	I(1)

Table 17 – Unit-root tests for $\Delta \boldsymbol{Independents_t}$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-4.55	-3.00	I(0)
Augmented Dickey-Fuller (with trend)	-5.22	-3.60	I(0)
Phillips-Perron (with trend)	-7.42	-3.60	I(0)
Dickey-Fuller GLS	-5.04	-3.40	I(0)
Elliott-Rothenberg-Stock	-5.04	-3.19	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.05	0.15	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 1 lag.

Table 18 – Unit-root tests for ***Party Loyalty***$_t$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.98	-3.00	I(1)
Augmented Dickey-Fuller (with trend)	-1.58	-3.60	I(1)
Phillips-Perron (with trend)	-2.15	-3.60	I(1)

Table 19 – Unit-root tests for Δ***Party Loyalty***$_t$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.66	-3.00	I(0)
Augmented Dickey-Fuller (with trend)	-5.93	-3.60	I(0)
Phillips-Perron (with trend)	-8.31	-3.60	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. Tests were conducted with 1 lag.

Table 20 – Unit-root tests for ***Challengers' Partisan Defecting to Incumbent***$_t$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.97	-3.00	I(1)
Augmented Dickey-Fuller (with trend)	-1.65	-3.60	I(1)
Phillips-Perron (with trend)	-1.92	-3.60	I(1)

Table 21 – Unit-root tests for Δ***Challengers' Partisan Defecting to Incumbent***$_t$

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.26	-3.00	I(0)
Augmented Dickey-Fuller (with trend)	-5.89	-3.60	I(0)
Phillips-Perron (with trend)	-5.51	-3.60	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. Tests were conducted with 1 lag.

Table 22 – The Effect of Polarization on Incumbency Advantage, 1946 – 2014

	ΔIncumbency Advantage$_t$
Incumbency Advantage$_{t-1}$	-0.74***
	(0.18)
ΔPolarization$_t$	-0.20
	(0.21)
Polarization$_{t-1}$	-0.37**
	(0.14)
Time trend	0.48**
	(0.17)
Election 1966	5.61**
	(2.19)
Constant	20.01**
	(6.86)
Observations	34
Adjusted R^2	0.47
Breusch-Godfrey χ^2 of:	
AR(1)	0.65
AR(2)	1.70
AR(3)	2.34
Durbin's Alternative χ^2 of:	
AR(1)	0.55
AR(2)	1.44
AR(3)	1.94
Cumby-Huizinga χ^2 of:	
AR(1)-AR(3)	2.15
Shapiro-Wilk z	-0.03

Note: Dependent variable is ΔIncumbency Advantage$_t$. Regression model with error correction specification. Standard errors in parentheses. All the estimates and the statistics were averaged across the 1,000 simulations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 23 – The Effect of % of Independents on Incumbency Advantage, 1952 – 2012

	ΔIncumbency Advantage$_t$
Incumbency Advantage$_{t-1}$	-0.57***
	(0.17)
Δ% of Independents$_t$	0.21
	(0.23)
% of Independents$_{t-1}$	0.39*
	(0.18)
Election 1966	6.33**
	(2.49)
Constant	-0.34
	(1.88)
Observations	26
Adjusted R^2	0.45
Breusch-Godfrey χ^2 of:	
AR(1)	0.90
AR(2)	2.18
AR(3)	3.40
Durbin's Alternative χ^2 of:	
AR(1)	0.77
AR(2)	1.88
AR(3)	2.94
Cumby-Huizinga χ^2 of:	
AR(1)-AR(3)	2.69
Shapiro-Wilk z	-0.24

Note: Dependent variable is ΔIncumbency Advantage$_t$. Regression model with error correction specification. Standard errors in parentheses. All the estimates and the statistics were averaged across the 1,000 simulations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 24 – The Effect of % of Party Loyalty on Incumbency Advantage, 1956 – 2008

	ΔIncumbency Advantage$_t$
Incumbency Advantage$_{t-1}$	-0.92***
	(0.21)
Δ% of Party Loyalty$_t$	-0.18
	(0.15)
% of Party Loyalty$_{t-1}$	-0.43**
	(0.17)
Election 1966	4.42
	(2.18)
Constant	42.26**
	(15.0)
Observations	22
Adjusted R^2	0.61
Breusch-Godfrey χ^2 of:	
AR(1)	0.77
AR(2)	2.07
AR(3)	2.73
Durbin's Alternative χ^2 of:	
AR(1)	0.62
AR(2)	1.68
AR(3)	2.14
Cumby-Huizinga χ^2 of:	
AR(1)-AR(3)	2.34
Shapiro-Wilk z	-0.14

Note: Dependent variable is ΔIncumbency Advantage$_t$. Regression model with error correction specification. Standard errors in parentheses. All the estimates and the statistics were averaged across the 1,000 simulations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 25 – The Effect of % of Challengers' Partisans Defecting to Incumbent on Incumbency Advantage, 1956 – 2008

	ΔIncumbency Advantage$_t$
Incumbency Advantage$_{t-1}$	-0.92***
	(0.22)
Δ% of Challengers' Partisans Defecting to Incumbent$_t$	0.10
	(0.07)
% of Challengers' Partisans Defecting to Incumbent$_{t-1}$	0.14*
	(0.06)
Election 1966	4.40
	(2.29)
Constant	2.39
	(1.87)
Observations	22
Adjusted R^2	0.59
Breusch-Godfrey χ^2 of:	
AR(1)	0.91
AR(2)	3.03
AR(3)	3.58
Durbin's Alternative χ^2 of:	
AR(1)	0.75
AR(2)	2.63
AR(3)	2.96
Cumby-Huizinga χ^2 of:	
AR(1)-AR(3)	2.97
Shapiro-Wilk z	-0.16

Note: Dependent variable is ΔIncumbency Advantage$_t$. Regression model with error correction specification. Standard errors in parentheses. All the estimates and the statistics were averaged across the 1,000 simulations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX B –

Supplementary Appendix: the Brazilian case

Table 26 – Summary Statistics

	Mean	SD	Min	Max	N
Presidential Approval Rating among Women	27.6	4.7	20.0	38.5	32
Presidential Approval Rating among Men	36.7	4.2	29.7	49.5	32
COVID-19 Monthly Deaths per Million	84.8	105.8	0	384.4	32
COVID-19 Monthly Cases per Million	3034.1	3483.6	0	10269	32
Inflation	0.44	0.41	-0.38	1.35	32
Stock of Formal Labor Jobs among Women	190,009.7	348,799.9	-398,423	1,062,200	32
Stock of Formal Labor Jobs among Men	465,416.9	528,797.3	-362,332	1,777,713	32
Spending with the Emergency Aid Program	10.5	15.6	0	45.9	32

Table 27 – Unit-root tests for ***Presidential Approval$_t$: Women***

Levels

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.639	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-2.575	-3.580	I(1)
Phillips-Perron (with trend)	-2.615	-3.576	I(1)
Dickey-Fuller GLS	-2.214	-3.325	I(1)
Elliott-Rothenberg-Stock	-2.214	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.140	0.146	I(0)

First-difference

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.662	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-3.606	-3.584	I(0)
Phillips-Perron (with trend)	-5.301	-3.580	I(0)
Dickey-Fuller GLS	-3.622	-3.336	I(0)
Elliott-Rothenberg-Stock	-3.622	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.093	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 28 – Unit-root tests for ***Presidential Approval$_t$: Men***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.037	-2.986	I(0)
Augmented Dickey-Fuller (with trend)	-3.030	-3.580	I(1)
Phillips-Perron (with trend)	-3.346	-3.576	I(1)
Dickey-Fuller GLS	-2.536	-3.325	I(1)
Elliott-Rothenberg-Stock	-2.536	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.136	0.146	I(0)
First-difference			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-4.176	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-4.052	-3.584	I(0)
Phillips-Perron (with trend)	-4.189	-3.580	I(0)
Dickey-Fuller GLS	-2.882	-3.336	I(1)
Elliott-Rothenberg-Stock	-2.882	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.117	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 29 – Unit-root tests for ***COVID-19 Monthly Deaths$_t$***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.922	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-3.105	-3.580	I(1)
Phillips-Perron (with trend)	-2.256	-3.576	I(1)
Dickey-Fuller GLS	-3.172	-3.325	I(1)
Elliott-Rothenberg-Stock	-3.172	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.113	0.146	I(0)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.205	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-3.102	-3.584	I(1)
Phillips-Perron (with trend)	-3.612	-3.580	I(0)
Dickey-Fuller GLS	-3.274	-3.336	I(1)
Elliott-Rothenberg-Stock	-3.274	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.081	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 30 – Unit-root tests for ***COVID-19 Monthly Cases***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.391	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-2.190	-3.580	I(1)
Phillips-Perron (with trend)	-1.949	-3.576	I(1)
Dickey-Fuller GLS	-2.296	-3.325	I(1)
Elliott-Rothenberg-Stock	-2.296	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.164	0.146	I(1)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.900	-2.989	I(1)
Augmented Dickey-Fuller (with trend)	-2.741	-3.584	I(1)
Phillips-Perron (with trend)	-4.213	-3.580	I(0)
Dickey-Fuller GLS	-2.988	-3.336	I(1)
Elliott-Rothenberg-Stock	-2.988	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.118	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 31 – Unit-root tests for ***Inflation***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.199	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-2.766	-3.580	I(1)
Phillips-Perron (with trend)	-3.583	-3.576	I(0)
Dickey-Fuller GLS	-2.834	-3.325	I(1)
Elliott-Rothenberg-Stock	-2.834	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.128	0.146	I(0)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.309	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-5.307	-3.584	I(0)
Phillips-Perron (with trend)	-7.696	-3.580	I(0)
Dickey-Fuller GLS	-5.364	-3.336	I(0)
Elliott-Rothenberg-Stock	-5.364	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.036	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 32 – Unit-root tests for ***Stock of Formal Labor Jobs among Men***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-0.282	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-1.268	-3.580	I(1)
Phillips-Perron (with trend)	-0.571	-3.576	I(0)
Dickey-Fuller GLS	-1.637	-3.325	I(1)
Elliott-Rothenberg-Stock	-1.637	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.301	0.146	I(0)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.150	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-3.438	-3.584	I(1)
Phillips-Perron (with trend)	-3.307	-3.580	I(1)
Dickey-Fuller GLS	-3.536	-3.336	I(0)
Elliott-Rothenberg-Stock	-3.536	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.115	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 33 – Unit-root tests for ***Stock of Formal Labor Jobs among Women***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-0.554	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-0.917	-3.580	I(1)
Phillips-Perron (with trend)	-0.296	-3.576	I(1)
Dickey-Fuller GLS	-1.417	-3.325	I(1)
Elliott-Rothenberg-Stock	-1.417	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.288	0.146	I(1)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.741	-2.989	I(1)
Augmented Dickey-Fuller (with trend)	-3.085	-3.584	I(1)
Phillips-Perron (with trend)	-3.289	-3.580	I(1)
Dickey-Fuller GLS	-3.106	-3.336	I(1)
Elliott-Rothenberg-Stock	-3.106	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.159	0.146	I(1)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 34 – Unit-root tests for ***Federal Government Spending with the Emergency Aid Program$_t$***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.844	-2.986	I(1)
Augmented Dickey-Fuller (with trend)	-1.767	-3.580	I(1)
Phillips-Perron (with trend)	-1.614	-3.576	I(1)
Dickey-Fuller GLS	-1.864	-3.325	I(1)
Elliott-Rothenberg-Stock	-1.864	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.218	0.146	I(1)

First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.062	-2.989	I(0)
Augmented Dickey-Fuller (with trend)	-3.040	-3.584	I(1)
Phillips-Perron (with trend)	-4.520	-3.580	I(0)
Dickey-Fuller GLS	-3.149	-3.336	I(1)
Elliott-Rothenberg-Stock	-3.149	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.079	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 35 – The Effect of COVID-19 Monthly Deaths on Presidential Approval for Women and Men, Brazil, January 2019 - August 2021

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-0.515***	
	(0.094)	
Approval Women$_{t-1}$		-0.419***
		(0.099)
ΔCOVID-19 Deaths per Million$_t$	-0.012	-0.020**
	(0.008)	(0.009)
COVID-19 Deaths per Million$_{t-1}$	-0.008**	-0.010**
	(0.004)	(0.004)
ΔEmergency Aid Spending$_t$	-0.035	-0.076
	(0.045)	(0.052)
Emergency Aid Spending$_{t-1}$	0.078***	0.075***
	(0.024)	(0.027)
Constant	18.296***	11.253***
	(3.575)	(2.982)
Observations	31	31
R^2	0.533	0.499
Breusch-Godfrey χ^2 of:		
AR(1)	0.332	1.938
AR(2)	0.341	5.126
AR(3)	7.166	6.259
Durbin's Alternative χ^2 of:		
AR(1)	0.260	1.600
AR(2)	0.256	4.556
AR(3)	6.614	5.566
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	6.793	5.623
Shapiro-Wilk z	0.185	-0.565

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Table 36 – The Effect of COVID-19 Monthly Cases on Presidential Approval for Women and Men, Brazil, January 2019 - August 2021

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-0.501***	
	(0.091)	
Approval Women$_{t-1}$		-0.419***
		(0.097)
ΔCOVID-19 Cases per Million$_t$	-0.000	-0.000
	(0.000)	(0.000)
COVID-19 Cases per Million$_{t-1}$	-0.000*	-0.000**
	(0.000)	(0.000)
ΔEmergency Aid Spending$_t$	-0.050	-0.101*
	(0.046)	(0.055)
Emergency Aid Spending$_{t-1}$	0.079***	0.079***
	(0.025)	(0.029)
Constant	17.767***	11.267***
	(3.484)	(2.929)
Observations	31	31
R^2	0.509	0.455
Breusch-Godfrey χ^2 of:		
AR(1)	0.239	3.756
AR(2)	0.247	6.465
AR(3)	5.551	6.858
Durbin's Alternative χ^2 of:		
AR(1)	0.187	3.309
AR(2)	0.185	6.061
AR(3)	4.799	6.250
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	6.228	5.244
Shapiro-Wilk z	-0.103	-0.193

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Table 37 – The Effect of COVID-19 Monthly Deaths on Presidential Approval for Women and Men, Brazil, January 2019 – August 2021

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-0.443***	
	(0.084)	
Approval Women$_{t-1}$		-0.348***
		(0.080)
ΔCOVID-19 Deaths per Million$_t$	-0.010	-0.016*
	(0.008)	(0.008)
ΔEmergency Aid Spending$_t$	-0.004	-0.144**
	(0.062)	(0.065)
Emergency Aid Spending$_{t-1}$	0.069***	0.040
	(0.023)	(0.025)
ΔInflation$_t$	-1.433*	2.261**
	(0.853)	(0.912)
ΔStock of formal labor Men$_t$	0.000	
	(0.000)	
ΔStock of formal labor Women$_t$		-0.000*
		(0.000)
Constant	15.025***	9.065***
	(3.176)	(2.326)
Observations	31	31
R^2	0.529	0.557
Breusch-Godfrey χ^2 of:		
AR(1)	1.863	0.221
AR(2)	4.620	3.791
AR(3)	7.313	5.900
Durbin's Alternative χ^2 of:		
AR(1)	1.470	0.165
AR(2)	3.853	3.065
AR(3)	6.484	4.936
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	7.733	7.328
Shapiro-Wilk z	1.555	-1.372

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Supplementary Appendix: the U.S. case

Table 38 – Summary Statistics

	Mean	SD	Min	Max	N
Presidential Approval Rating among Women	34.94	3.52	29.00	44.00	47
Presidential Approval Rating among Men	48.32	3.70	40.00	56.00	47
COVID-19 Monthly Deaths per Million	22.50	52.54	0	243.85	47
COVID-19 Monthly Cases per Million	1288.00	3617.12	0	19440.96	47
Unemployment rate among Women	5.04	2.84	3.30	16.10	47
Unemployment rate among Men	4.98	2.33	3.50	13.60	47
Consumer Price Index	252.88	5.34	243.77	261.56	47
Index of Consumer Sentiment	93.05	8.69	71.80	101.40	47
Real Disposable Personal Income (per capita)	44914.70	1950.65	42504.00	52070.00	47

Table 39 – Unit-root tests for ***Presidential Approval$_t$: Women***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.440	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-3.025	-3.520	I(1)
Phillips-Perron (with trend)	-5.263	-3.516	I(0)
Dickey-Fuller GLS	-2.239	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.239	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.15	0.146	I(1)
First-difference			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-6.289	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-6.172	-3.524	I(0)
Phillips-Perron (with trend)	-11.878	-3.520	I(0)
Dickey-Fuller GLS	-4.477	-3.202	I(0)
Elliott-Rothenberg-Stock	-4.477	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0806	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 40 – Unit-root tests for ***Presidential Approval$_t$: Men***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-3.494	-2.944	I(0)
Augmented Dickey-Fuller (with trend)	-4.030	-3.520	I(0)
Phillips-Perron (with trend)	-5.797	-3.516	I(0)
Dickey-Fuller GLS	-4.058	-3.195	I(0)
Elliott-Rothenberg-Stock	-4.058	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0862	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 41 – Unit-root tests for ***COVID-19 Monthly Deaths$_t$***

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-0.248	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-1.568	-3.520	I(1)
Phillips-Perron (with trend)	-1.559	-3.516	I(1)
Dickey-Fuller GLS	-2.078	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.078	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.344	0.146	I(1)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.468	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-5.793	-3.524	I(1)
Phillips-Perron (with trend)	-5.696	-3.520	I(0)
Dickey-Fuller GLS	-4.994	-3.202	I(1)
Elliott-Rothenberg-Stock	-4.994	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0628	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 42 – Unit-root tests for ***COVID-19 Monthly Cases***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	2.954	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	2.153	-3.520	I(1)
Phillips-Perron (with trend)	4.335	-3.516	I(1)
Dickey-Fuller GLS	-0.878	-3.195	I(1)
Elliott-Rothenberg-Stock	-0.878	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.323	0.146	I(1)

First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.652	-2.947	I(1)
Augmented Dickey-Fuller (with trend)	-3.477	-3.524	I(1)
Phillips-Perron (with trend)	-3.054	-3.520	I(1)
Dickey-Fuller GLS	-3.230	-3.202	I(0)
Elliott-Rothenberg-Stock	-3.230	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.187	0.146	I(1)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 43 – Unit-root tests for ***Unemployment Rate among Men***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.195	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-2.732	-3.520	I(1)
Phillips-Perron (with trend)	-2.661	-3.516	I(1)
Dickey-Fuller GLS	-2.686	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.686	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.26	0.146	I(1)

First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.102	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-5.048	-3.524	I(0)
Phillips-Perron (with trend)	-6.411	-3.520	I(0)
Dickey-Fuller GLS	-5.170	-3.202	I(0)
Elliott-Rothenberg-Stock	-5.170	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0402	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 44 – Unit-root tests for ***Unemployment Rate among Women***$_t$

Levels

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-2.427	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-2.930	-3.520	I(1)
Phillips-Perron (with trend)	-2.799	-3.516	I(1)
Dickey-Fuller GLS	-2.898	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.898	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.226	0.146	I(1)

First-differences

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.266	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-5.207	-3.524	I(0)
Phillips-Perron (with trend)	-6.295	-3.520	I(0)
Dickey-Fuller GLS	-5.327	-3.202	I(0)
Elliott-Rothenberg-Stock	-5.327	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0385	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 45 – Unit-root tests for ***Consumer Price Index***$_t$

Levels

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-0.884	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-3.363	-3.520	I(1)
Phillips-Perron (with trend)	-2.226	-3.516	I(1)
Dickey-Fuller GLS	-3.413	-3.195	I(0)
Elliott-Rothenberg-Stock	-3.413	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.296	0.146	I(1)

First-differences

Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-4.976	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-4.916	-3.524	I(0)
Phillips-Perron (with trend)	-4.260	-3.520	I(0)
Dickey-Fuller GLS	-4.838	-3.202	I(0)
Elliott-Rothenberg-Stock	-4.838	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0467	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 46 – Unit-root tests for ***Index of Consumer Sentiment***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.715	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-2.679	-3.520	I(1)
Phillips-Perron (with trend)	-2.515	-3.516	I(1)
Dickey-Fuller GLS	-2.612	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.612	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.329	0.146	I(1)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-5.642	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-5.619	-3.524	I(0)
Phillips-Perron (with trend)	-6.033	-3.520	I(0)
Dickey-Fuller GLS	-5.690	-3.202	I(0)
Elliott-Rothenberg-Stock	-5.690	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0333	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 47 – Unit-root tests for ***Real Disposable Personal Income Per Capita***$_t$

Levels			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-1.611	-2.944	I(1)
Augmented Dickey-Fuller (with trend)	-2.928	-3.520	I(1)
Phillips-Perron (with trend)	-3.942	-3.516	I(0)
Dickey-Fuller GLS	-2.995	-3.195	I(1)
Elliott-Rothenberg-Stock	-2.995	-3.190	I(1)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.109	0.146	I(0)
First-differences			
Unit Root Test	Test Statistic	5% Critical Value	Conclusion
Augmented Dickey-Fuller	-6.287	-2.947	I(0)
Augmented Dickey-Fuller (with trend)	-6.221	-3.524	I(0)
Phillips-Perron (with trend)	-9.548	-3.520	I(0)
Dickey-Fuller GLS	-6.375	-3.202	I(0)
Elliott-Rothenberg-Stock	-6.375	-3.190	I(0)
Kwiatkowski-Phillips-Schmidt-Shin (Ho: Trend Stationary)	0.0381	0.146	I(0)

Note: For all tests, the null hypothesis is that the series contains a unit-root. The only exception is the expect KPSS in which the null hypothesis is that the series follows a trend stationary process. Tests were conducted with 3 lags.

Table 48 – The Effect of COVID-19 Monthly Deaths on Presidential Approval for Women and Men, U.S., February 2017 – December 2020

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-1.054***	
	(0.130)	
Approval Women$_{t-1}$		-0.919***
		(0.132)
ΔCOVID-19 Deaths per Million$_t$	-0.076***	-0.051**
	(0.028)	(0.025)
COVID-19 Deaths per Million$_{t-1}$	0.076	-0.009
	(0.058)	(0.042)
ΔUnemployment Rate Men$_t$	1.282**	
	(0.568)	
Unemployment Rate Men$_{t-1}$	-1.379	
	(1.034)	
ΔUnemployment Rate Women$_t$		1.043***
		(0.378)
Unemployment Rate Women$_{t-1}$		0.535
		(0.642)
Linear time trend	0.142***	0.112***
	(0.051)	(0.038)
March 2020	-1.099	6.179**
	(3.138)	(2.601)
June 2020	-5.767	-10.196***
	(3.874)	(3.045)
July 2020	0.450	0.143
	(5.209)	(4.295)
Constant	53.488***	27.037***
	(8.341)	(4.741)
Observations	46	46
R^2	0.550	0.634
Breusch-Godfrey χ^2 of:		
AR(1)	0.390	1.251
AR(2)	0.659	1.289
AR(3)	1.671	1.291
Durbin's Alternative χ^2 of:		
AR(1)	0.299	0.979
AR(2)	0.494	0.980
AR(3)	1.244	0.953
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	1.351	1.834
Shapiro-Wilk z	-0.441	-0.037

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Table 49 – The Effect of COVID-19 Monthly Cases on Presidential Approval for Women and Men, U.S., February 2017 – December 2020

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-0.989***	
	(0.131)	
Approval Women$_{t-1}$		-0.925***
		(0.131)
ΔCOVID-19 Cases per Million$_t$	0.000	0.000
	(0.000)	(0.000)
COVID-19 Cases per Million$_{t-1}$	-0.001*	-0.001***
	(0.000)	(0.000)
ΔUnemployment Rate Men$_t$	-0.045	
	(0.348)	
Unemployment Rate Men$_{t-1}$	0.290	
	(0.330)	
ΔUnemployment Rate Women$_t$		0.279
		(0.217)
Unemployment Rate Women$_{t-1}$		0.756***
		(0.214)
Linear time trend	0.177***	0.122***
	(0.049)	(0.034)
March 2020	-1.350	5.948**
	(3.157)	(2.553)
June 2020	-7.112*	-11.059***
	(3.805)	(2.920)
July 2020	-7.640**	-3.203
	(3.839)	(2.994)
Constant	43.014***	26.173***
	(6.140)	(4.124)
Observations	46	46
R^2	0.548	0.650
Breusch-Godfrey χ^2 of:		
AR(1)	1.977	0.642
AR(2)	2.048	0.675
AR(3)	3.032	0.693
Durbin's Alternative χ^2 of:		
AR(1)	1.572	0.496
AR(2)	1.585	0.506
AR(3)	2.329	0.505
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	2.667	0.781
Shapiro-Wilk z	-0.518	0.554

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Table 50 – The Effect of COVID-19 Monthly Cases on Presidential Approval for Women and Men, U.S., February 2017 – December 2020

	ΔApproval Men$_t$	ΔApproval Women$_t$
Approval Men$_{t-1}$	-1.090***	
	(0.132)	
Approval Women$_{t-1}$		-0.935***
		(0.145)
ΔCOVID-19 Deaths per Million$_t$	-0.065**	-0.046*
	(0.029)	(0.026)
COVID-19 Deaths per Million$_{t-1}$	0.059	-0.016
	(0.058)	(0.044)
ΔUnemployment Rate Men$_t$	2.288	
	(1.412)	
Unemployment Rate Men$_{t-1}$	-1.324	
	(1.039)	
ΔUnemployment Rate Women$_t$		0.788
		(0.879)
Unemployment Rate Women$_{t-1}$		0.624
		(0.663)
ΔDisposable Personal Income (per capita)$_t$	-0.001	0.000
	(0.002)	(0.001)
ΔConsumer Price Index$_t$	-1.355	-0.655
	(1.264)	(1.036)
ΔIndex of Consumer Sentiment$_t$	0.207	0.021
	(0.161)	(0.124)
Linear time trend	0.153***	0.115***
	(0.051)	(0.038)
March 2020	-2.556	6.086*
	(4.393)	(3.498)
June 2020	-3.081	-9.999**
	(4.786)	(4.075)
July 2020	3.866	0.030
	(5.812)	(4.788)
Constant	55.510***	27.423***
	(8.523)	(5.295)
Observations	46	46
R^2	0.568	0.637
Breusch-Godfrey χ^2 of:		
AR(1)	0.000	0.571
AR(2)	0.085	0.614
AR(3)	1.404	0.648
Durbin's Alternative χ^2 of:		
AR(1)	0.000	0.402
AR(2)	0.058	0.420
AR(3)	0.945	0.428
Cumby-Huizinga χ^2 of:		
AR(1)-AR(3)	1.162	0.671
Shapiro-Wilk z	-0.856	0.081

Note: Dependent variable is ΔApproval$_t$ for men (column 1) and women (column 2). Results from a seemingly unrelated regression model with error correction specification. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All the diagnostic tests were conducted after estimating a separate OLS model for each gender.

Table 51 – The Effect of COVID-19 Monthly Deaths on the Probability of Approving President Trump's Job Performance

	Approval
State-level COVID-19 deaths per 100,000 over the last 30 days	-0.168***
	(0.042)
Race	9.939***
	(0.454)
Education	1.638***
	(0.196)
Vote in 2016 Elections	-3.242***
	(0.065)
Female	-10.652***
	(0.551)
Wave	0.075***
	(0.016)
State	0.026
	(0.078)
Constant	20.407***
	(2.874)
Observations	329679
R^2	0.097

Note: The dependent variable is a binary variable indicating whether the respondent approves or disapproves of President Trump's job performance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.