THE UNIVERSITY OF CHICAGO

POLITICAL COGNITION IN A LIBERAL DEMOCRACY: THE EFFECTS OF ELECTION OUTCOMES ON PERCEIVED CORRUPTION, PERCEIVED LEGITIMACY, AND VOLUNTARY COMPLIANCE WITH THE LAW

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
AND
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
DEPARTMENT OF PSYCHOLOGY
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
ANIRUDH TIWATHIA

CHICAGO, ILLINOIS
DECEMBER 2020

To my parents,

The most fascinating, noble, generous people:

Without you, none of this would have been possible

SUPREME COURT OF THE UNITED STATES

Opinions of the Court

On Perceived Corruption and Democratic Legitimacy

Buckley* v. *Valeo

424 U.S. 1 (1976)

Per Curiam Opinion

Of almost equal concern as the danger of actual quid pro quo arrangements is the impact of the appearance of corruption stemming from public awareness of the opportunities for abuse …safeguarding against [it] requires that *the opportunity for abuse inherent in the process of raising large monetary contributions* be eliminated.

Citizens United* v. *FEC

558 U.S. 310 (2010)

5-4 Majority Opinion

This Court now concludes that [unrestricted election spending], including …by corporations, does not give rise to corruption or the appearance of corruption …That [donors] may have influence over or access to elected officials does not mean that those officials are corrupt: *"Favoritism and influence are not …avoidable in representative politics."*

(emphasis added)

CONTENTS

APPENDIX

GLOSSARY

REFERENCES

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Writing a dissertation is the best way to disabuse oneself of any romanticized ideas about doctoral studies. The image of a young scholar toiling alone in their ivory tower in pursuit of an esoteric yet noble cause gives way to a truth that is far more interesting: We exist as a set of inter-dependencies —inseparably enmeshed within networks that support our efforts and give our life meaning. We are beneficiaries of love, attention, and care. And, only as such can our efforts come to fruition. This work —like all works worth undertaking —would never have seen the light of day without the innumerable efforts of countless people. Here, I hope to acknowledge a few of them for the integral role they played every step of the way.

First and foremost, I cannot thank my family enough —especially my parents: no words can truly capture my gratitude. I know that I would never have made it to the finish line if you had not worked tirelessly to ensure that I always had all the support I needed. To my mother, who repeatedly leapt to my aid in a myriad ways: I still cannot believe you learned how to code in R just so that you could keep me company on those long nights of data analysis. I will always cherish the days we spent together in the trenches. Your energy, optimism and perseverance humble us all. To my father, who remained unflappable in his faith that I would finish: thank you for always being the voice that kept me tethered to the world —for repeatedly helping me overcome the myopia that accompanies every dissertation and for always reminding me of the daily pleasures that give life meaning. I could not have navigated the many pitfalls without your patient guidance. But mostly, I am grateful that the dissertation gave us the chance to sit together on Scrivener and spend those early morning hours writing and working together. To my brother, Aditya: thank you for always being a source of good cheer and optimism —and, for making sure that I never lost faith. And, thank you for putting up with all my absences while I was lost in research. To my grandmothers: thank you for the countless prayers you made in my name

—ensuring that I had God in my corner. To Evelina, my dearest partner: thank you for all the countless ways in which you made life easier, for giving me the luxury of pursuing this esoteric goal, and for all your patience as I slowly moved towards the finish. To Molly, my dearest friend and lifelong ally: thank you for all the invaluable support you provided as I started down this journey many years ago. And, thank you for being my coach for the final months of writing —you were always the voice of confidence advocating on my behalf. To Joshua: you have improved each one of my research projects over the years —enriching my experience as a scientist in innumerable ways. From the early days of launching studies on Amazon's Mechanical Turk (when it was still a 'wild' frontier) to the design of immersive online study environments to our recent efforts constructing an 'end-to-end' reproducible research pipeline —your insights, guidance, enthusiasm and generosity have been invaluable in helping me navigate the hurdles involved in a fully digitized research program. I only wish we could have had more time to pursue research together.

At the University of Chicago, I must take a moment to thank all the faculty, students and staff that so generously shared their time and their insights. I must especially thank my dissertation committee: Devin Pope, Reid Hastie, George Wu at Booth School of Business and William Goldstein at the Psychology Department. Devin: thank you for serving as the chair of my committee, for being so generous with your time, for being patient as I finalized the manuscript, and for all your encouragement along the way. Reid and George: thank you for all your input and feedback along the way. Bill: I feel especially indebted to you for all the time you took over the many years that I have known you. I learned most of what I know about the field over the course of many Friday afternoons spent in Socratic dialogues —discovering the contours of the field and developing the insights needed to pursue new lines of inquiry. In many ways, those discussions were the epitome of how I imagined a doctoral education. To Hiroki, my course-mate and friend: I am grateful for all the discussions we had on a wide-range of matters, especially the hours we spent together thinking about how to construct measures of perceived corruption. Our early discussions and work together helped me immensely as I tried to transform my long-standing interest in corruption into a tangible research project. Finally, thank you for finding research funding that helped launch the initial studies on perceived corruption. I would also like to thank all the donors that created research funds to support student research, including Richard Thaler for the *Thaler-Tversky Independent Research Grant* and Susan Einhorn for the *Hillel J. Einhorn Memorial Scholarship.* Their generosity was critical in funding my dissertation research. Finally, I would also like to thank all the members of the Ph.D Office who work tirelessly to ensure that students have the resources they need to complete their program. Malaina, Amity, and Cynthia: thank you for all that you have done over the years.

Finally, this work could never have been completed without the countless named and unnamed (and, collectively under-recognized) mathematicians, software developers, scientists, open-knowledge and open-source crusaders that facilitated the development of languages like *R, markdown, pandoc, LaTeX, Git, html, javascript*; software like *Zotero* and *R-Studio*; R packages like *data.table, tidyverse, ggplot2, knitr, bookdown, citr,* and countless others. Without their efforts, science would remain forever trapped in the hands of elite institutions and remain the sole province of grant-recipients. This work could never have been completed without the efforts of all those individuals that fought for open-access to government-collected datasets. This work (and, the majority of recent psychology studies) would also not be possible without the consent and participation of (often, underpaid and carelessly-treated) online denizens of Amazons Mechanical Turk (Williamson, 2016) and other crowdsourcing platforms. It is my sincere hope that we —as scientists —move aggressively to protect our online participants' interests, including their right to privacy (Xia et al., 2017), their right for collective bargaining (Irani & Silberman, 2013, 2016; Salehi et al., 2015), and their right for fair and equitable compensation. Finally, I must also thank the great staff at Amtrak that indulged me as I spent a month riding the trains —working to finish the final chapters. Tony: your generosity exceeded any reasonable expectations.

It takes a village to complete a dissertation. I hope this document can be a worthy testament to the unwavering faith, endless encouragements, and generous care that the world provided me. This dissertation belongs to you all!

ABSTRACT

The current work presents two complementary programs of research that aim to: (i) clarify the psychological mechanisms governing perceptions of legitimate authority; (ii) examine how US Presidential election outcomes may significantly impact these perceptions; (iii) provide empirical evidence consistent with the notion that such perceptions of legitimate authority can have significant causal effects on ethical and political behavior. Part I examines the effect of election outcomes on *perceptions of corruption*: (i) examining the implications of these findings for perception-based indices of good governance; and, (ii) establishing a robust predictive relationship between an individual's perception of corruption and their willingness to cheat. Part II builds upon these findings and looks at tax-compliance as a case study to test whether election outcomes causally influence people's willingness to meet their tax obligations. These election-based effects provide empirical evidence in support of theories hypothesizing a link between perceived legitimacy and voluntarily compliance with the law.

Keywords: cheating, corruption, elections, legitimacy, political cognition, tax compliance

Part I

(Biased) Perceptions of Corruption and the Consequences for Cheating and Unethical Behavior

CHAPTER 1

INTRODUCTION

In 2014, Putin decided to invade Crimea. Surprisingly, in the months following the invasion, the levels of perceived corruption reported by the Russian populace dropped significantly. Why? It seems unlikely *prima facie* that – because of the Russian military's covert actions on the Crimean Peninsula —the average Russian citizens had fewer encounters with crooked cops or graft-demanding bureaucrats. By 2015, despite the worsening effects of international economic sanctions and a downturn in the economy (Pew Research Center, 2017c), public confidence in Putin soared, with 9 out of 10 Russians reporting high or very high levels of confidence in Putin —a leap of 20% from the start of the conflict (Pew Research Center, 2015, 2017b).

Far from being a mere outlier, I propose that the above-cited case of Putin and the Crimean Invasion is just one instance of a broad class of apparent incongruencies that appear frequently in the measurement of political opinion. Closer to home, for example, numerous recent surveys show that Americans report dismal levels of trust in democratic institutions —with current levels at historic lows (Pew Research Center, 2017a). Longitudinal data from the World Values Survey has shown a steady decline in the percentage of Americans that support democratic governance: in 1930, around 75% of Americans reported that it was "essential to live under democratic governance;" by 1980, only 30% agreed. By 2011, almost 1-in-4 younger Americans (16–34) responded that a "democratic political system" is a "bad" or "very bad" way to run the country – a dramatic increase from 1995, when only 16% had said the same (Foa & Mounk, 2017b).[1]

Given these findings, there appears to be a growing crisis in the perceived legitimacy

[1]Although a rich online exchange has emerged around the conclusions drawn by the Foa & Mounk (2017b) paper, the general pattern of data remains largely uncontested. Instead, the debate is centered on the degree to which the phenomenon is widespread across democracies in general and what can be appropriately concluded from such survey data.

of democratic institutions and the forms of government supported by these institutions. However, while perceptions have deteriorated at an alarming rate, there does not appear to be evidence supporting a corresponding decline in the *actual quality* of the democratic institutions being evaluated. Studies by prominent, independent monitors such as Freedom House, CIRI, V-Dem, and Polity IV show no evidence of dramatic declines in institutional quality. Foa & Mounk (2017b) have argued such survey result are early signs of "democratic deconsolidation" – whereas others suggest they may be less worrisome than they initially appear (for contrasting viewpoints, see rebuttal articles published as part of editor-sanctioned online web discussion at the Journal of Democracy website: Alexander & Welzel, 2017; Foa & Mounk, 2017a; Howe, 2017; Inglehart, 2016; Norris, 2017; Voeten, 2017).

In the current work, rather than focusing on the appropriate conclusions one ought to draw from any of these findings, I propose that these "anomalies" raise some general theoretical concerns about the measurement of political opinion; concerns that serve as the basis for the work developed in this chapter, specifically:

(a) *compliance*: are respondents complying with the survey or are people knowingly and intentionally using their responses as a vehicle for political expression (i.e. do such measures of political opinion capture the "true beliefs" of the respondents)? For the current purposes, *respondent compliance* serves as the necessary precondition for any meaningful subsequent analyses.

(b) *correspondence with reality*: do such measures serve as reasonable proxies for the state of reality (for example, do the perception-based surveys of institutional quality tell us something about the actual quality of democratic institutions in America)? Note, this is different from the usual concern in survey measures i.e. it is not a question of whether the instrument (e.g. survey about changes in price of household goods) reasonably measures the perception of interest (e.g. perceived changes in affordability of daily life). Instead, it asks whether the perceptions () are responsive enough to conditions on the ground to serve as reasonable proxy measures of real conditions (e.g. do measures of *perceived changes in affordability* correspond with reality enough to serve as a meaningful proxy measure of *actual inflation*). In terms of this concern, the process model proposed in Part I of this dissertation argues that there may be aspects of political cognition that should threaten this link for many common political perceptions. For example, one of the premises for the current model argues that most respondents usually do not have well-formed prior beliefs about most political domains – especially, issues like corruption across the various domains of government and society. And, that this lack of prior beliefs leaves such political opinions susceptible to biases common to the formation of spontaneous judgments, especially attribution substitution.

(c) *consequentiality or predictive validity:* do these political perceptions causally drive and shape subsequent political behavior? Or, for a weaker test, we could ask —to what degree do they reliably predict the behavior of respondents (i.e. predictive validity)?

(d) *intervention opportunities:* are there simple, easy-to-implement changes that could improve the reliability of political perceptions, with reliability largely focused on *correspondence* and *consequentiality* (points *b* and *c* from above).

1.1 Perceptions of Corruption: A Case Study

Within this broader context, the current Chapter focuses specifically on the determinants and the consequences of perceived corruption. Perceived corruption provides one of the more practical and theoretically profitable case-studies for exploring the theoretical questions outlined above. It is a case where the issues of "correspondence to reality" and the downstream behavioral consequences ("consequentiality") are central to policy decisions —often with far-reaching consequences. Moreover, given the large number of global institutions that produce and consume measures of perceived corruption annually, there are numerous opportunities for applying interventions that could result in meaningful improvements.

1.1.1 Measuring Corruption: The Perception-Based Approach

Corruption is frequently identified as "the single greatest obstacle to economic and social development" by the World Bank and the UN (Johnson & Sharma, 2004) —with the estimated cost of corruption potentially exceeding more than 5 per cent of global GDP and an estimated $1 trillion in bribes paid each year (Heywood & Rose, 2014). Corruption is also widely identified as the one of the foremost concerns among voters, both in America (Wu & Muñana, 2018) and across the world (Pew Research Center, 2017a, 2014). Between 2010 and 2014, three in four of Americans perceived corruption as "widespread throughout the government in this country" —with the reported rate largely stable (73-79%) in each annual survey (Gallup, 2015). Even on central issues like perceived integrity of our elections, a startlingly large number of Americans expressed doubt —for example, in 2016, around 70% said that they did not feel confident about "the honesty of elections" (Reinhart, 2020).

Corruption also appears to loom large as emotional concerns for most Americans. For example, the *Chapman University Survey of American Fears*, an annual survey measuring the top fears of Americans found that —since 2015 —fear about "corrupt government officials" became the most widely reported fear by the American public. It continue to top the list of American fears for each of the past 5 annual waves of the survey and the prevalence keep increasing each year: growing from 58% of Americans reporting *"corruption"* as a major fear in 2015 to 77% of Americans doing so in 2019 (Sheth, 2019).

Despite its importance, the study of corruption was long hampered by the fact that the incidence of actual corruption can be difficult to measure directly (Hawken & Munck, 2008). Transparency International (TI) was one of the first prominent institutions to tackle this measurement issue by launching the Corruption Perceptions Index (CPI) in 1995. Their work advocating for the use of *perceived corruption* as a reasonable proxy measure for *actual corruption* —along with their work administering annual, worldwide surveys —helped trigger an explosion of academic and policy work on corruption (Treisman, 2007), almost all of which relied upon this perceptions-as-a-proxy-measure approach. The clearest indicator of the success and wide-spread acceptance enjoyed by the perception-based approach to measuring corruption is the sheer prevalence of such measures as decision inputs across a wide-range of applications. For example, even now, almost all major indicators of good governance still rely wholly or partially upon measures of perceived corruption (Olken & Pande, 2012) (see Appendix A.1 for details).

1.1.2 Correspondence with Reality: Are Perceptions of Corruption Good Proxies for Actual Corruption?

Why Correspondence Matters: The Wide-Spread Use of Perception-Based Indicators as Decision Inputs for Policy and Business

Unlike most measures of political opinion, measures of perceived corruption have become a significant *decision input* for businesses, NGOs, government officials and intergovernmental organizations like the World Bank (WB) and the International Monetary Fund (IMF) (Arndt, 2008; Malito, 2014). Previous work has shown that the ratings on indices of perceived corruption can dramatically influence the economic fortunes of a developing country by significantly altering local economic conditions on the ground (Mauro, 1995; Melgar et al., 2010). For example, corruption indicators can influence a country's ability to access private foreign investment (Wei, 1999), since investors and investment banks incorporate these indicators in their investment frameworks; especially when investing in *"less developed countries"* (LDCs) (Oman & Arndt, 2006). In addition, these indicators can also determine a country's ability to access development aid from donors (Andersson & Heywood, 2009).

For example, the United States Agency for Interna1tional Development (USAID), which is one of the largest aid agencies in the world, explicitly states that a country's ratings on Transparency International's Corruption Perceptions Index (TI-CPI) is a factor used to determine eligibility for funding decisions (Treisman, 2007); the Organization for Economic Co-operation and Development (OECD) uses these indicators to target foreign investment (Arndt & Oman, 2006); and, the European Union uses these measures to guide law enforcement actions and to coordinate adherence to EU anti-corruption policies —as specified in the Stockholm treaty (Council of the European Union, 2011).

Is there Empirical Evidence for Correspondence?

The justification for using perception-based measures of corruption (and, perception-based measures of good governance more broadly) relies on the assumption that these indices – often constructed by aggregating national and international surveys of perceived corruption – can serve as a reasonable proxy for actual incidence of corruption in a country (Kaufmann et al., 2004; Ko & Samajdar, 2010; Lambsdorff, 2006; Treisman, 2000). Not surprisingly, concerns about the correspondence between perceived and actual corruption has been a topic of ongoing importance in the field of development studies – with mixed results (Heywood & Rose, 2014).

Evidence against the correspondence between perception and actual corruption The body of evidence that has caused the greatest concern about the utility of perception-based measures primarily comes from developing countries. Multiple field studies in countries like Russia (Belousova et al., 2016), Mexico (Morris & Klesner, 2010), Indonesia (Olken, 2009), as well as a sample of 8 countries from Sub-Saharan Africa (Razafindrakoto & Roubaud, 2010) found low-to-negligible correlations between objective and perception-based measures of corruption. Specifically, they found that perception-based indices tend to correlate poorly with people's direct experiences of corruption, i.e. they did not capture the likelihood, frequency, or necessity to engage in corruption to receive basic civic services (Treisman, 2007). Such findings have led researchers to critique the construct validity (Thomas, 2010) and the methodological frameworks used to produce perception-based scores of good governance (United Nations Development Programme & Global Anti-corruption Initiative, 2015).

Another major factor that may undermine faith in the correspondence between perception and reality is the recent set of findings showing that demographic characteristics of the respondent, like education, can sometimes be the strongest predictors of perceived corruption, even more predictive than the actual incidence of corruption (Donchev & Ujhelyi, 2014; Olken, 2009; Olken & Pande, 2012). Similarly, there is evidence suggesting that the perceptions of corruption can be dramatically influenced by variation in cultural attitudes (Banuri & Eckel, 2012; Barr & Serra, 2010; Cameron et al., 2009), socio-economic characteristics (Pázmándy, 2011) or structural features of the economy (Ulman, 2014). In each of those cases, as perceptions become susceptible to external or contextual factors, there is an increasing threat to the validity of perceived corruption as a proxy measure for actual corruption.

That said, meta-analyses suggest that perception-based indices have been increasing in reliability over the years (Ko & Samajdar, 2010). Moreover, the CPI-based indices correlate remarkably well with measures of judicial effectiveness (Treisman, 2000), measures of fairness in the rule of law (Rothstein & Uslaner, 2005), the quality of political institutions (Lederman et al., 2005), economic liberalization (Goldsmith, 1999), economic growth and openness to trade (Ades & Di Tella, 1999; Mauro, 1995; Svensson, 2005), and inequality (Jong-sung & Khagram, 2005a, 2005b).

Evidence supporting the validity of perceived corruption as a proxy for actual corruption In addition to these robust correlations across indices, other recent work also provides good reasons to reject a "strongly pessimistic" approach to the perception-based measures of corruption. A review these findings makes a persuasive case that concerns regarding perception-based indices may be overblown. Systematic, cross-national reviews suggest that evidence from specific developing countries may not always generalize, and that perception-based indices – despite their problems – perform reasonably well. For example, a large-scale, cross-national study drawing upon a sample of 85,000 respondents across 24 European countries provided evidence that (a) perceptions of corruption among lay citizens and subject matter experts are remarkably consistent; (b) these perceptions correspond well with actual, reported incidences of corruption; and (c) such perceptions are swayed little by 'outside noise' (Charron, 2016).

Additionally, perception-based indicators of corruption correlate reasonable well with indicators of corruption constructed from real-world government actions —for example, correlating quite well with a measure of corruption computed from the analysis of large-scale government procurement datasets (Fazekas & Kocsis, 2017a, 2017b; Fazekas & Tóth, 2017). Similarly, for most cases, perception-based measures of corruption did not differ substantially from estimates of corruption that were calculated based upon legal actions taken by the United States under the *Foreign Corrupt Practices Act of 1977 (FCPA)*; this was true for both individual country estimates as well as cross-national corruption rankings (Escresa & Picci, 2015, 2016; Golden & Picci, 2005).

1.1.3 Consequentiality: Does Perceived Corruption Influence Behavior?

Economic Behavior

In addition to their use as proxy-measures, the perception of corruption is considered important in and of itself, with some theorists arguing that the *perception of corruption* may have as serious consequences for economic development as *corruption itself* (see, Treisman, 2000 for a review). There is at least some evidence that perceived corruption inhibits economic growth due to decreased willingness to invest (Mauro, 1995), an effect that persists above and beyond any consequences that may arise from the use of perception-based measures by financial institutions to determine their willingness to lend or the effect of perception-based measures on a donor's willingness to provide development aid *i.e. the perception of corruption is detrimental enough to inhibit economic activity on its own.*

Trust in Public Institutions

There has been a long-standing opinion in law and political science that believes that perception of corruption has a corrosive effect on the public trust in institutions. This sentiment is best captured in the U.S. Supreme Court's famous ruling in (*Buckley v. Valeo*, 1976), where the court unanimous held that:

> [T]he avoidance of the appearance of improper influence 'is also critical ...if confidence in the system of representative Government is not to be eroded to a disastrous extent.' ...*The case thus raises issues not less than basic to a democratic society.*
>
> – *Buckley v. Valeo*, 424 U.S. 1 (1976)

Based upon this reasoning, US Supreme Court deemed the perception of corruption to be enough of a threat that it ruled unanimously in favor of restrictions on political spending i.e. according to the court, avoiding the appearance of corruption is so crucial to representative democracy that it even justified restrictions on the most sacrosanct of rights, the right to participate freely in political speech as guaranteed by the First Amendment to the Constitution. The Supreme Court reaffirmed this stance again in (*Nixon v. Shrink Missouri Government PAC*, 2000), where it again upheld the central importance of minimizing and reducing *perceptions of corruption* wherever possible:

> Leave the perception of impropriety unanswered, and the cynical assumption that large donors call the tune could jeopardize the willingness of voters to take part in democratic governance.
>
> Democracy works "only if the people have faith in those who govern, and that faith is bound to be shattered when high officials and their appointees engage in activities which arouse suspicions of malfeasance and corruption."
>
> – *Nixon v. Shrink Missouri Gov't PAC*, 528 U.S. 377 (2000)

In a parallel tradition in political philosophy, academics across a wide ideological spectrum —from Democracy theorists (Warren, 2004) to Political Realists (Philp & Dávid-Barrett, 2015) —argue forcefully to place *perceptions of corruption* at the center of theories of governance and are unified in highlighting the importance of mitigating the *"appearance of corruption"* in and of itself.

Adding to these long-standing legal and philosophical bodies of work, there is now initial empirical evidence suggesting that perceived corruption can influence local norms, hinder people's trust in public institutions and undermine democratic norms (Warren, 2017, 2006). For example, high-levels of perceived corruption can discourage victims from reporting crimes to the police (Soares, 2004) or, may encourage individuals seeking public services to engage in corruption by offering bribe themselves (Čábelková & Hanousek, 2004), although other studies appear to find contrasting results (Rose & Mishler, 2010). On the political level, exposure to corruption scandals can decrease people's general trust in government (Citrin & Stoker, 2018) and influence their interpretation of future scandals (Dancey, 2012).

Voluntary Compliance with the Law

Finally, to the degree that perceived corruption undermines the perceived legitimacy of a government, it may also influence people's willingness to comply with the law more broadly

(Jackson et al., 2012; Sunshine & Tyler, 2003; Tankebe et al., 2016; Tankebe & Gutierrez-Gomez, 2018; Tyler, 2004). A set of recently published studies have found innovative ways to measure whether people who originate from countries with high levels of corruption are more willing to engage in dishonest behavior themselves, even when they are no longer in the country of origin. For example, a study examining the number of parking violations by U.N. diplomats[2] found that diplomats from more corrupt countries accrued substantially more parking tickets and left more of those tickets unpaid (Fisman & Miguel, 2007, 2014; Matheus & Paulo, 2016). A study of tax compliance by American corporations found higher rates of tax violations for companies with owners who were originally from countries with higher rates of corruption (DeBacker et al., 2015). Finally, a cross-societal study involving 23 countries showed a strong link between the prevalence of rule violations in a country (measured as an aggregate of political practices, tax evasion, and perceived corruption) and the willingness of individuals from that country to cheat on a laboratory task (Gächter & Schulz, 2016), although not all cross-national studies have found such such a link (Pascual-Ezama et al., 2015). Overall, the findings suggest that for individuals from a given country, perception-based corruption indices could – on average – be predictive of their willingness to violate the law.

The current studies add to this body of work by examining the impact of perceived corruption on willingness to cheat at the level of the individual respondent. This approach allows the current work to extend the above-cited findings, which primarily show a correlation between (i) levels of corruption in the country of origin and (ii) the levels of rule-violation shown by residents of those countries. Instead, as described in the methods section, the current work allows me to link perceptions of corruption to willingness-to-cheat at the individual level, independent of the effects of national culture and respondent background. In addition, by taking advantage of variation in perceived corruption, the current experimental design

[2]Until 2002, U.N. diplomats had diplomatic immunity from New York City law, so their adherence to parking regulations was based solely on their willingness to comply with the law voluntarily.

also allows me to test whether perceived corruption causally determines willingness-to-cheat.

1.2 Biases in Perceptions of Corruption and their Consequences

Notwithstanding the large body of work on corruption, there is a clear need to specify the microlevel determinants of perceived corruption (Citrin & Stoker, 2018; Dong & Torgler, 2009). Recent work has focused on identifying demographic characteristics of the respondent that appear to influence perceptions of corruption, including age, gender, religion, income, occupational status, level of education, and political interest (Andersson & Heywood, 2009; Bai et al., 2016; Dong & Torgler, 2009; Gutmann et al., 2015; Pázmándy, 2011; Rivas, 2013; Swamy et al., 2001). And, as mentioned above, there are reasonable concerns that these respondent characteristics could determine the degree of correspondence between perceived and actual levels of corruption (Olken, 2009). In fact, Abramo (2008), Gutmann et al. (2020), and Corrado et al. (2017) provide convincing initial evidence suggesting that various demographic, geographic, cultural, political, institutional and socio-economic factors may determine both (a) in what context and (b) to what degree the perception of corruption accurately reflect the incidence and experience of corrupt practices (objective corruption). This is still a growing body of research, but —if true —one should expect that a wide-range of incidental / situational factors may significantly influence the validity of perception-based corruption indices, often in unpredictable ways (Donchev & Ujhelyi, 2014; Gutmann et al., 2020; Olken & Pande, 2012).

1.2.1 Attribute Substitution as Another Potential Source of Bias

Complementing research on demographics-driven biases in perceived corruption, the current paper proposes that an additional, significant, and, as of yet, overlooked source of bias in the measures of perceived corruption may simply arise from the complexity of the judgment itself. Judgments of corruption require the integration of numerous – and often conflicting, weak, and overlapping – pieces of information. As such, it is an inherently ambiguous and complex judgment to assess the level of corruption present in the government. Given the complexity of this task, most individuals are likely to take a heuristic approach to judgments in this domain (Kahneman & Frederick, 2002).

While many factors may determine the character of this heuristic evaluation, the current study argues that, as with other complex judgments, "attribute substitution" is likely to be one of the central influences (and, potential biases) on judgments of corruption – especially if most participants did not have a pre-existing attitude about corruption before they were asked (Bertrand & Mullainathan, 2001). Thus, I suggest that, under a non-deliberative response mode, when people are asked to make the complex, ambiguous judgment *"how corrupt is my government?"* they replace it with an easier, more accessible one: *"how much do I support the current government?"* This model would also suggest that respondents with more knowledge and interest in politics would be less susceptible to such effects.

Some initial evidence that supports the attribute substitution hypothesis comes from a study examining global responses to Transparency International's (TI) 2004 Global Corruption Barometer (GCB) —the same one as is used in the current work (see Section 2.3.3 [Measuring Perceived Corruption][Stage 3: Measuring Perceived Corruption —The Global Corruption Barometer] and Section 3.1 Adapting the GCB).[3] The authors of that study analyzed the 53K responses from 64 countries —using the cross-national data to assess the relationship between (i) *perceptions* of corruption and (ii) self-reported *experiences* of corruption[4] (e.g. being asked to pay a bring). Across multiple sets of analyses, they repeatedly found either a non-existent or an unexpectedly weak relationship between the perception measures and the experience measure. Moreover, the "disparities" between perceptions and

[3] Although, as explained in Section 3.1.1 [Excluded Questions][ExcludedQs], the 2004 Questions focused on "Experiences of Corruption" were excluded in the current work. For full details, also see Appendix A.4 for exact list of questions that were excluded.

[4] For example, participants were asked: *In the past 12 months, have you or anyone living in your household paid a bribe in any form?* [Yes/No/DK/DA]

experience did not show a systematic pattern across countries —instead, the distance between the two measures varied haphazardly from country to country —leading them to conclude:

> "Personal or household experience of bribery is not a good predictor of perceptions held about corruption... On the other hand, perceptions are mostly good predictors (sometimes excellent predictors) of other perceptions, not only related to corruption but also to other, apparently unrelated, matters.[5] It seems that opinions operate in a coherent world. The problem is that such imaginary world of opinions and guesses seems not to hold a close relationship with the world of reality, at least in ... regards corruption" (p.5, Abramo, 2008).

Although this is only an analysis of a single corruption survey (GCB) from a single year (2004), other work has also found similar patterns, for example: Mocan (2004) compared four widely-used perception-based corruption indices to a corruption index created by a weighted proportion of individuals who were asked for a bribe. For all four indices, after controlling for the quality of the institutions in the country (measured by the risk of expropriation), there was no statistical relationship between actual corruption and the perceived corruption. This begs the question: if the *perceptions of corruption* are not derived from one's experiences with corruption, what do these perceptions come from and why are they so coherent with opinions on a wide array of matters?

Here, I propose that (at least some portion of) my perceptions of corruption result from attribute substitution and more often capture *"how much do I support the current government?"* and not *"how corrupt is the current government?"* To test this hypothesis, I ran two studies that measured perceived corruption one-day before and after two elections in

[5] Matters like their opinion about (i) cost-of-living/inflation; (ii) poverty; (iii) environmental problems; (iv) human rights; (v) insecurity/crime/violence/terrorism; (vi) jobs.

2016: (i) the Democratic Primary in California,[6] which decided the democratic nomination between Hillary Clinton and Bernie Sanders; and (ii) the 2016 U.S. Presidential election.

In the current framework, each election was treated as a naturalistic shock that altered people's feelings towards their government by either placing authority in the hands of a candidate they support or one they oppose. By comparing pre-election and post-election groups, I arrived at an initial test of whether support of the government was a significant factor driving people's perceptions of corruption and whether this effect is moderated by the level of political interest of the respondent. In addition, as mentioned above in the section on *consequentiality*, I can take advantage of this potential variation in levels of perceived corruption to examine whether perceptions of corruption causally influence people's willingness-to-cheat.

Finally, it should be noted that —if perceptions of corruption are significantly impacted by electoral loss – it would undermine a central tenet in the political science literature – namely that people meaningfully distinguish between *regime support* (trust in the form of government) and *government support* (trust in the specific administration) (Citrin & Stoker, 2018; Foa & Mounk, 2016). In fact, evidence from four Latin American countries that used self-reported *first-person encounters* with corruption showed a significant link between regime support and first-person encounters with corruption (Schneider, 2017) and evidence from China suggests that such a link also exists for perceived corruption (Wang & Dickson, 2018).

[6]Widely discussed in the news at the time as the make-or-break moment for the candidacy of Senator Bernie Sanders for nomination by the Democratic Party.

CHAPTER 2

METHODS

2.1 Rationale and Central Hypothesis

The current work begins with the hypothesis that assessing the level of corruption present in the government is an inherently ambiguous, complex, and distant judgment. Based upon past work, one should expect such judgments to be highly susceptible to attribute substitution: under a non-deliberative response mode, people are likely to replace the complex judgment with an easier one (Kahneman & Frederick, 2002). Specifically, instead of forming the judgment *"how corrupt is my government,"* they may be substituting it with the easier judgment *"how much do I support my government?"*[1] I argue that these attribute substitution effects may be an unaccounted yet prominent source of variability in current measures of perceived corruption – and, that this source of variability has significant implications for the use of perceived corruption as a reliable measure of good governance.

In addition, the current studies add to the literature on perceived corruption and intrinsic honesty, while holding the effect of culture constant. The current work also addresses the notion that perceptions of corruption may influence people's willingness to engage in corruption themselves. Specifically, I test whether perceptions of corruption were accompanied by an increased willingness to cheat and, if so, whether this relationship is causal.

2.1.1 Elections as Naturalistic Shocks

In the current design, elections were treated as naturalistic shocks that I hypothesized would significantly alter people's feelings towards their government. I ran two studies that measured

[1]Although the current work cannot speak to this, I believe that the actual substitution is with the judgment "how do I feel about my government" rather than "how much do I support my government" – and, in cases where these two diverge, it is the feeling rather than the professed support that will dominate judgments of corruption.

perceived corruption one day before and one day after two prominent elections in 2016: (i) the Democratic Primary in California, which decided the contest between Hillary Clinton and Bernie Sanders for the Democratic Presidential nomination; and (ii) the 2016 U.S. Presidential election.

The California Primary was widely discussed in the news at the time as the make-or-break moment for the candidacy of Senator Bernie Sanders; i.e. it would determine which candidate would be nominated to be the presidential candidate for the Democratic Party. Due to the broad-interest and clear implications for the outcome, the California Primary served as the first "naturalistic shock" to test our hypothesis. The models developed from the data in Study 1 were then confirmed via an exact replication in Study 2 using the 2016 U.S. Presidential Election as the second naturalistic shock. Using the Presidential Election as a replication allowed me to test the statistical and theoretical generalizabili0ty of the model developed here, since Hillary Clinton won in the first election and lost in the second.

The current study design addresses both key components of the current project: (i) identifying a significant determinant of perceived corruption; and, (ii) examining the consequences of perceived corruption on individual behavior. By comparing perceptions of corruption pre-and-post election, the current design tested whether support of the winning candidate significantly influenced people's reported perceptions of corruption across the political, institutional and economic spheres. In doing so, it identified a source of (potentially biased or spurious) variability that could undermine the level of correspondence between *perceived* and *actual* corruption.

In addition, I used this quasi-experimental variation in perceived corruption to test the link between perceived corruption and willingness to cheat. I measured willingness to cheat by seeing whether participants were willing to lie about the number of dependents they have in order to receive an inflated "bonus" payment for having participated in the study (the bonus is framed as additional payment to participants on the basis of the number of people that they are financially supporting —and, thus, it is structured as a needs-based compensation scheme, depends upon the self-reported number of dependents, and the bonus increases by $0.25 for each additional dependent reported). Using this measure of cheating, the current experimental design allowed me to test whether *perception of corruption* causally influences people's willingness-to-cheat (as a proxy for their willingness to engage in petty forms of corruption themselves).

2.2 Experimental Procedure

2.2.1 Timing

Two studies were conducted during the 2016 Presidential election season. The first study was run one day before and after the California Primary for the Democratic Party (6th June / 8th June, 2016, with the primary taking place on the 7th of June). The second study was run one day before and after the U.S. 2016 Presidential election (7th / 9th November, 2016 with the election taking place on the 8th of November). Except for some minor differences, participants in both studies were exposed to the same experimental protocol —allowing Study 2 to serve as a replication of the models and findings established in Study 1. The procedure for both studies is shown in Figure 2.1 below and described in the following subsections.

2.2.2 Sequence of Stimuli

In both studies, participants experienced the same sequence of stimuli and survey instruments, which came in four stages. First, they were asked to complete a moderately lengthy demographics form, which also included a question about the number of dependents (this information was used to create an individually-tailored opportunity to cheat). Second, after the demographic form, participants completed a measure of political attitudes, political orientation, and engagement with the election. Third, after completing the survey about their political characteristics, participants completed a measure of perceived corruption that was adapted from Transparency International's *Global Corruption Barometer* (GCB). Fourth, after completing the corruption survey, participants were given an opportunity to cheat such that if they lied they could increase the amount of payment they received for participating in the study. Each of these four stages are described in greater detail in the respective subsections below.

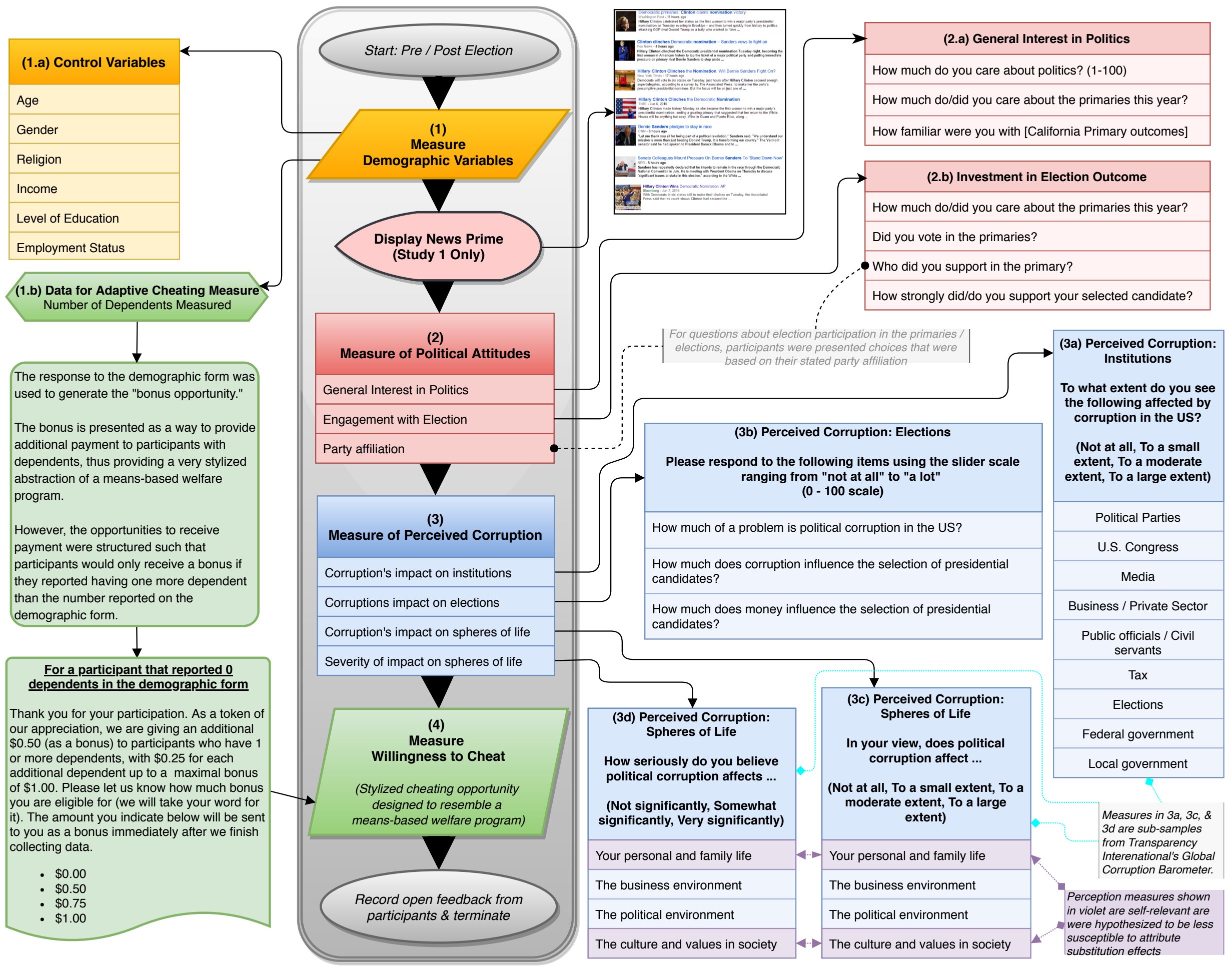
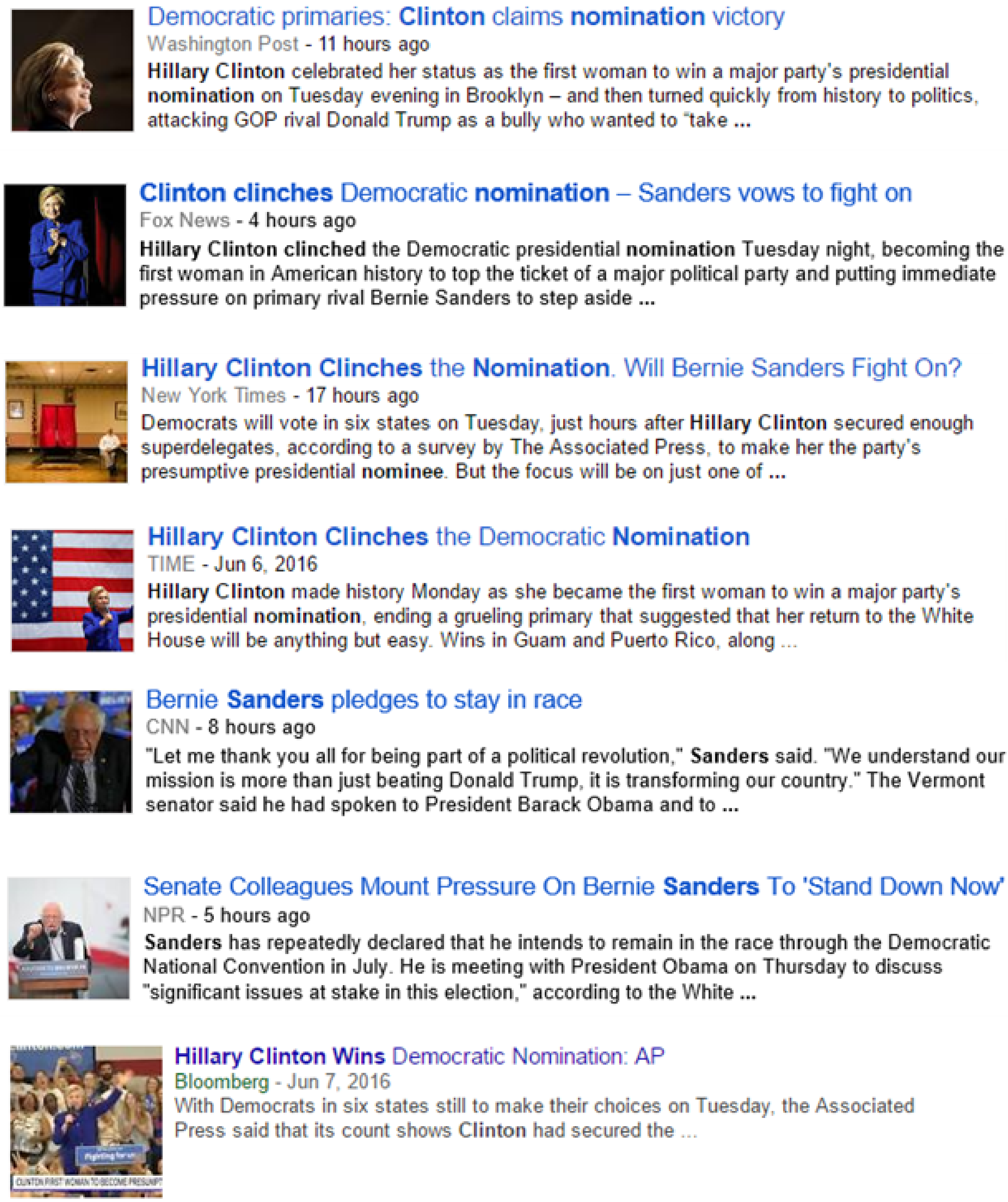


Figure 2.1: For both studies, the same procedure was applied to all participants. First, participants completed a demographics form. Then, they completed a questionnaire designed to measure their political orientation and levels of engagement. After that, participants completed a measure of perceived corruption that was adapted from Transparency International's Global Corruption Barometer. Finally, they were provided with an opportunity to increase their reward by lying. This served as a stylized measure for their willingness to engage in 'petty corruption.'

For the post-election sample in Study 1 (California Primary), I wanted to ensure that all participants were aware of the election outcome and were aware that, for all intents and purposes, the election outcome meant the end of Senator Bernie Sanders' bid for the party nomination. As such, in between the demographics form and the measure of political attitudes, Study 1 participants in the post-election sample were exposed to a sample of newspaper headlines that discussed the outcome of the election and its consequences for

Senator Bernie Sanders' campaign (see Figure 2.2 below). In Study 2, it was assumed that everyone was aware of the outcome of the Presidential Election, so no news media prime was included.


Democratic primaries: **Clinton** claims **nomination** victory
Washington Post - 11 hours ago
Hillary Clinton celebrated her status as the first woman to win a major party's presidential **nomination** on Tuesday evening in Brooklyn – and then turned quickly from history to politics, attacking GOP rival Donald Trump as a bully who wanted to "take ...

Clinton clinches Democratic **nomination** – Sanders vows to fight on
Fox News - 4 hours ago
Hillary Clinton clinched the Democratic presidential **nomination** Tuesday night, becoming the first woman in American history to top the ticket of a major political party and putting immediate pressure on primary rival Bernie Sanders to step aside ...

Hillary Clinton Clinches the **Nomination**. Will Bernie Sanders Fight On?
New York Times - 17 hours ago
Democrats will vote in six states on Tuesday, just hours after **Hillary Clinton** secured enough superdelegates, according to a survey by The Associated Press, to make her the party's presumptive presidential **nominee**. But the focus will be on just one of ...

Hillary Clinton Clinches the Democratic **Nomination**
TIME - Jun 6, 2016
Hillary Clinton made history Monday as she became the first woman to win a major party's presidential **nomination**, ending a grueling primary that suggested that her return to the White House will be anything but easy. Wins in Guam and Puerto Rico, along ...

Bernie **Sanders** pledges to stay in race
CNN - 8 hours ago
"Let me thank you all for being part of a political revolution," **Sanders** said. "We understand our mission is more than just beating Donald Trump, it is transforming our country." The Vermont senator said he had spoken to President Barack Obama and to ...

Senate Colleagues Mount Pressure On Bernie **Sanders** To 'Stand Down Now'
NPR - 5 hours ago
Sanders has repeatedly declared that he intends to remain in the race through the Democratic National Convention in July. He is meeting with President Obama on Thursday to discuss "significant issues at stake in this election," according to the White ...

Hillary Clinton Wins Democratic Nomination: AP
Bloomberg - Jun 7, 2016
With Democrats in six states still to make their choices on Tuesday, the Associated Press said that its count shows **Clinton** had secured the ...

Figure 2.2: For the post-election sample, participants were shown a sample of headlines about the outcome of the California Primary and its conclusive consequences for Senator Bernie Sanders. This was done to ensure that all participants were aware of the election outcome and its consequences.

2.3 Survey Instruments

2.3.1 Stage 1: Measure of Demographic Variables

The purpose of the demographics form was twofold: (i) measure demographic variables to serve as controls; and (ii) measure the number of dependents for each participant so that it could be used to construct an individually-tailored measure of the participant's *willingness-to-cheat*.

Control Variables

Recent work has emphasized some respondent characteristics that appear to reliably influence perceptions of corruption – including: age, gender, religion, income, occupational status, employment type, and level of education (Andersson & Heywood, 2009; Bai et al., 2016; Dong & Torgler, 2009; Gutmann et al., 2015; Swamy et al., 2001). Gender has also been previously identified as a relevant factor in some studies of cheating behavior (Fosgaard et al., 2013). Considering these earlier findings, I measured and included these variables in our analyses as controls; however, demographic factors neither drove any ancillary hypotheses nor were they central to the tests of our primary hypothesis.

Number of Dependents as a Cheating Measure

In addition to the above-mentioned variables, the demographic form also included other questions as fillers. These questions only served to increase the length of the demographics survey and thus to provide an opportunity to inconspicuously embed a question about number of dependents. This information about dependents was used to adaptively modify the cheating question presented to the participant. It was hoped that most participants would not recall that they had already answered a question about number of dependents in the demographic form when they encountered the cheating opportunity at the end of the study.

However, since I am primarily interested in voluntary compliance, participant naivete was not essential to the cheating measure.

2.3.2 Stage 2: Political Affiliation and Engagement with the Election Process

After completing the demographics survey, participants completed a brief measure of political characteristics included items covering the following areas.

Political Affiliation

I asked participants to report their party affiliation and the candidate they supported in the election [*Study 1*: primaries for selection of the presidential candidate; *Study 2*: the 2016 U.S. Presidential election].

Election Investment

I also measured how much they care about the election (measured on a 0-100 scale). In addition, they also indicated whether they intended to vote / had voted in the election. And, independent of whether they voted, how strongly they supported their preferred candidate (0-100 scale).

In Study 1, for the post-election sample, participants were first shown a sample of newspaper headlines with the results of the Democratic California Primary (See Figure 2.2). Along with this sample of headlines, I also asked participants to indicate how familiar they were with the results prior to seeing the sample of headlines (awareness also measured on a 0-100 scale).

Self-Reported Political Interest

Participants also indicated how much they care about politics in general (0-100 scale).

Expectations and Emotional Investment

In Study 2, I also asked about (a) participant's expectations about the election outcome as well as (b) their feelings about the victory / loss of their preferred candidate for President (for, pre-election, anticipated feelings; for post-election, current feelings). For additional details, please see Appendix A.2.

2.3.3 Stage 3: Measuring Perceived Corruption - The Global Corruption Barometer

To measure perception of corruption, I relied upon a subset of the *Global Corruption Barometer* (hereinafter, GCB[2]) administered by Transparency International (TI) annually across the globe. The exact questions are included in Appendix A.3. A detailed treatment of the topic —including the choice of survey items and the procedure for scale construction —is presented in Chapter 3 Perception of Corruption: Measurement and Scale Construction.

Components of the Corruption Measure

Questions on political and institutional corruption The primary set of questions taken from the GCB ask participants to indicate the extent to which they "see the following affected by corruption in the US?": (1) Political parties; (2) U.S. Congress; (3) Media; (4) Business / Private Sector; (5) Public officials / Civil servants; (6) Taxes; (7) Elections; (8) Federal Government; (9) Local Government.

[2] The full survey was last administered in the US by TI in 2013 using an online sample of 1000 participants

Questions 1-4 have been administered by TI in every annual GCB from 2004 to 2013. Questions 5-6 were only present for some of the GCB samples (Q5: 2009-2013; Q6: 2004-2007). Questions 7-9 were created by me to replace three measures from the GCB (Military, Police, & Judiciary), which were left off because of a lack of a clear predicted hypothesis about these institutions (Citrin & Stoker, 2018; Zmerli & Meer, 2017).

Questions on corruption in the election process Finally, in addition to the GCB questions, I included three questions specifically directed towards the US presidential elections. To measure perceptions of the election process, specifically the perceived fairness and representativeness of the voting process, I included two additional questions (0-100 scale) about the U.S. Presidential elections that asked respondents about (i) the effect of corruption on the selection of [*Study 1:* presidential candidates / *Study 2:* the U.S. President]; (ii) the influence of money in the selection of [*Study 1:* presidential candidates / *Study 2:* the U.S. President]. In addition, I included a question about political corruption on the same 0-100 scale to complement similar questions from the GCB.

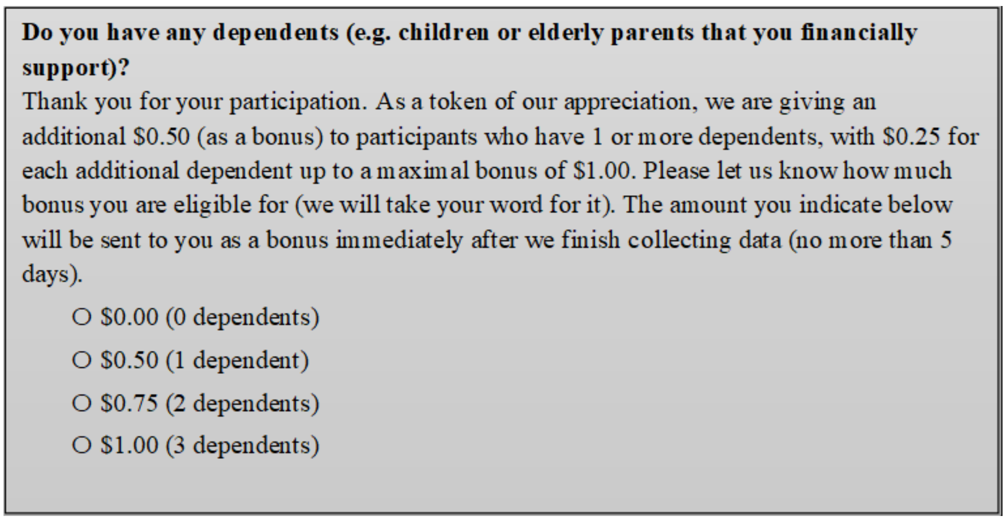
Questions on corruption in different spheres of life: Varying the ease of attribute evaluability Critically, in both studies, I included a set of questions from earlier versions of the GCB (2003-04) because they allowed for a test of our proposed hypothesis.[3] Specifically, these questions split the effect of corruption into separate "spheres of life" and asked respondents to indicate whether and to what degree corruption affects: "the political environment," "the business environment", "the culture and values in society," and "your personal and family life." Since judgments about the effects of corruption on my personal and family are more immediate and easier to answer, it was hypothesized these questions would be less susceptible to attribute substitution effects (exact questions highlighted in

[3] These questions were only asked in full for GCB 2003. They were partially administered for GCB 2004-2006, as such do not allow for longitudinal comparisons.

Appendix A.3).

2.3.4 Stage 4: Measures of Willingness-to- Cheat

To approximate whether levels of perceived corruption may impact a person's willingness to engage in petty corruption in their daily life, I also provided participants the opportunity to engage in minor cheating for a nominal bonus. To speak to the public-sector context of petty corruption, I designed a novel paradigm for measuring cheating that was structured as a stylized abstraction of a means-based benefits program.


Do you have any dependents (e.g. children or elderly parents that you financially support)?
Thank you for your participation. As a token of our appreciation, we are giving an additional $0.50 (as a bonus) to participants who have 1 or more dependents, with $0.25 for each additional dependent up to a maximal bonus of $1.00. Please let us know how much bonus you are eligible for (we will take your word for it). The amount you indicate below will be sent to you as a bonus immediately after we finish collecting data (no more than 5 days).

- O $0.00 (0 dependents)
- O $0.50 (1 dependent)
- O $0.75 (2 dependents)
- O $1.00 (3 dependents)

Figure 2.3: At the end of the study, the participant was told that they would be rewarded a bonus based upon the number of dependents.The survey was responsive, however, and participants were only eligible to receive a bonus if they reported more dependents than they had in the demographic form at the start of the study. The $0 option was matched with the number of dependents previously reported. If participants had indicated 1 dependent, then they would receive the bonus of $0.50 for reporting 2 dependents, and so forth. About 75% of the sample reported 0 dependents.

The measure of cheating was as follows: prior to beginning the survey, the participants completed a moderately lengthy demographics form, which included a question about their number of dependents. This information was used to adaptively modify the survey. After participants had completed the corruption survey, they were told that I would provide them an additional bonus based upon the number of dependents that they were supporting —

described as "children or elderly parents that you financially support" (see Figure 2.3 for exact text).

However, since the survey was adaptive, the $0 option always matched the number of dependents previously reported by the participant. Thus, participants were only eligible to receive a bonus if they reported an inflated number of dependent than they had previously done in the demographic form. They received an additional $0.50 for reporting one additional fictitious dependent and they received $0.25 for each additional fictitious dependent that they were willing to report, up to a maximum of a $1.00.

This measure of cheating not only attempted to capture some surface features of a means-based benefit, it also had the additional benefits, like, (i) for most participants, it was successful at obscuring the fact that it was a cheating measure, (ii) it was a cheating measure that mapped easily onto widely-known fairness norms,[4] and (iii) the cheating measure utilized personally-relevant information, where lying is felt to be particularly aversive for most people (Cappelen et al., 2013).

In addition, the current approach also avoids the pitfalls of most distribution-based measures of cheating[5] that make it difficult or impossible to identify the amount of cheating for any specific individual (thus making it harder to link cheating behavior with individual-level, micro-determinants of behavior). The current task avoided this problem —allowing me to measure not just willingness to cheat (as a binary variable) but also the amount / severity of cheating for each participant individually —which was simply captured by the number of fake dependents reported (which translates into incremental dollar amounts, in terms of

[4]I received lots of feedback regarding this bonus strategy, either praising me for taking "need" into account when distributing bonuses or criticizing me for using "number of dependents" as a valid measure of "need." Despite the specific opinion, it was clear that the vast majority of participants took the bonus payment structure at face-value and not as an explicit measure of cheating —although, a small minority of participants did appear to catch on.

[5]For example, studies that use cheating tasks involving chance devices (e.g. rolling a six-sided die or flipping a coin in private and then reporting the resulting outcome), the only way to detect cheating is to compare the sample distribution to the expected distribution from a fair device.

cheating amount). This flexibility provided a significant advantage, since it allowed me to link perceptions of corruption to willingness to cheat at the participant-level rather than group-level.

All these features were attempts to increase the number of surface and structural features that may be analogous between this lab-based cheating measure and the real-world decision to engage in minor, routine, "petty" corruption. This is not to say that the current measure is equivalent to a measure of petty corruption or even that it could be considered a reasonable / reliable proxy measure. There is simply not enough evidence to connect lab-measures with decisions to engage in corruption in the real-world. In the absence of such validity measures, the design of the current cheating measure was simply an attempt to increase the odds that this novel cheating measure is a bit better able to capture similarities between lab and real-world behavior.

Based upon feedback provided by participants at the end of the study, it seems that most participants accepted the premise of the bonus system at face-value without suspecting anything and provided comments like *"I think that's pretty cool that you are giving out more money to people with dependents, even though I didn't qualify for the bonus"*, *"That is pretty unfair! Getting more money because you have kids/dependents!!"* or *"That was very nice of you to offer a bonus for workers with dependents. I do not have any dependents, so I do not qualify. But that is definitely a nice gesture!"* Overall, the participant feedback gives the impression that the cheating measure provided a reliable cover-story that decreased suspicion that it was a measure of cheating and allowed for a more naturalistic measure. In Section 5.3.2 [Differences in Transparency: Opportunity to Cheat vs Test of Honesty][DiffInTransparency], this issue is examined in greater detail.

CHAPTER 3

PERCEPTION OF CORRUPTION: MEASUREMENT AND SCALE CONSTRUCTION

The organization of the following sections is as follows. In Section 3.1, I describe the survey items in the corruption measure and present the rationale behind the choices guiding survey design and construction. Then, in Section 3.2.1, I lay out the rationale behind the choice of aggregation methods that allow me to construct a composite scale for perceived corruption. After that, in Section 3.2.2, I address the issue of construct dimensionality using the response patterns from Study 1. In Section 3.2.2, I begin by first assuming unidimensionality and, under this assumption, I proceed with a classic Item Response Analysis[1] as a preliminary analysis of the consistency, reliability and discriminability of the current instrument used to measure perceived corruption. By examining the primary indicators constructed from this analysis, I present initial evidence that (a) a multi-dimensional model is both *a priori* reasonable; and, (b) the existence of a multi-dimensional construct is supported the patterns seen in the data. In Section 3.2.3, I present findings from Exploratory Factor Analysis (EFA) that assess whether item responses are sufficiently granular to justify a multi-dimensional construct i.e. I examine whether participants are capable and willing to respond to the different items in the corruption survey in a manner that is *sensitive / responsive to the particularities of the different items;* and, thus, responses to different "clusters of items" are meaningfully and reliably differentiable from each other. Having addressed these concerns regarding the reliability, granularity, and dimensionality of the measured construct, in Section 3.3, I detail the SEM model and present results assessing the quality of the fitted model.

[1]Item response theory (IRT) assumes unidimensionality across all items - i.e. a single latent variable underlies the probability of responses to all items in the scale (Meijer & Tendeiro, 2018). Although, some have advocated for "multidimensional item response theory" (MIRT) (Liu et al., 2018) - due to the significant increase in complexity that MIRT demands, the majority of IRT research that is broadly used relies upon the unidimensional model. Here, unidimensionality is simply assumed to be the starting point - in order to construct the primary IRT indicators - and, based upon those indicators, I present preliminary evidence supporting the construction of a multi-dimensional model.

3.1 Survey Instruments: Adapting the Global Corruption Barometer

To measure perception of corruption, I relied upon a subset of the *Global Corruption Barometer*[2] (hereinafter, GCB) administered by Transparency International (TI) annually across the globe. The exact questions are included in Appendix A.3.

3.1.1 Using the Global Corruption Barometer (GCB) as the Basis for the Corruption Survey

Relevance of the GCB

The GCB, which compliments TI's Corruption Perceptions Index (CPI), was selected as our primary source for three crucial reasons:

Questions in the GCB remain more stable across each annual administration of the survey, thereby allowing me to compare our findings with earlier results; whereas the CPI is a composite index formed by aggregating different sources of data; The GCB specifically focuses on measuring "the extent and impact of corruption, as judged by the general public" as opposed to the CPI, which often includes measures limited to business leaders and government experts; Analyses of early results from the GCB showed that it correlated very highly with the composite measures, like the CPI, as well as the Worldwide Governance Indicators (WGI) produced by the World Bank (Treisman, 2007).

Selecting the Appropriate Subset from the GCB

Excluded Questions In constructing the subset, I excluded the portions of the questionnaire that were not applicable to the current project, specifically (see Appendix A.4 for more

[2]The full Global Corruption Barometer survey was last administered in the US by Transparency International in 2013 using an online sample of 1000 participants

details about the specific questions that were excluded):

(i). *questions focused on perceived changes in corruption*: I was interested in examining the determinants of current levels of perceived corruption, not evaluations of change across time;

(ii). *questions focused on perceived efficacy of / willingness to engage with anti-corruption efforts*;

(iii). *questions that measured conspiratorial mindsets* (Uscinski et al., 2016; Uscinski, 2017);

(iv). *questions that aimed to measure participants' past encounters with bribery:* I excluded questions about actual payments of bribes both because (a) it requires disclosure of illegal activity and (b) the current work is specifically focused on the biases underlying the perceptions of corruption with the goal of estimating whether those perceptions would impact people's willingness to engage in *actual petty corruption*, rather than past-reported encounters with bribery, a lot of which —due to its relative infrequency (around 7% report having bribed) —may be determined by situational variables (encounter with corrupt officials, nature of one's profession, etc.).

Questions included from previous versions of the GCB that allow for a test of attribute substitution Critically, in both studies, I included questions from earlier versions of the GCB surveys (Q3 is taken from GCB-2004 and Q4 from GCB-2003; all the rest from GCB-2013). These questions from the earlier version of the GCB were included because they allowed for a test of our attribute substation hypothesis as described below.

3.2 Prior to Scale Construction: Establishing Methodological and Empirical Justification for SEM Modeling Approach

Before I can proceed to fitting the structural model on the corruption survey data, it is necessary to establish a robust and well-justified methodology for constructing a "perception of corruption" scale. This requires both methodological choices about statistical approaches and it requires an examination of the patterns found in the data. For example, if the data show no evidence of a multi-dimensional structure, fitting the SEM model may produce highly biased estimates. Similarly, if participant responses to different survey items do not show sensitivity to specific characteristics of the item, it may be useful to apply a different approach to item aggregation. By replicating the model using the fresh data collected in Study 2, I have useful guards against concerns of over-fit and biased model estimates. However, the replication in Study 2 can only truly be confirmatory if I address important concerns and establish some evidence to suggest that the Study 1 data meet the necessary preconditions prior to fitting the model.

Having just reviewed the procedure for selecting the survey items, I now turn my attention to three primary concerns: (i) approach to item aggregation; (ii) questions of construct dimensionality; and, (iii) the need to test whether the responses show sufficient granularity to distinguish different items / should show evidence of sensitivity to item-characteristics (i.e. participant responses to different items in the survey should be reasonably nuanced, responsive to specific characteristics of the item, and meaningfully differentiable from each other when the items differ in meaningful ways).

After addressing these concerns, I then present the SEM model used for scale construction. I fit the SEM model to the Study 1 data and present initial evidence to establish the SEM model as a valid approach to scale construction in the current context. I then outline how the SEM model fits estimated from Study 1 are used to construct the Study 2 measures and how this approach can address any possible concerns regarding over-fit.

3.2.1 Concerns Regarding the Aggregation of Measures

Concerns Surrounding the Choice of Statistical Methods for Aggregation and the Construction of Composite Indicators

First, there are significant issues and problems underlying the choice of method for aggregating diverse measures to create a composite indicator (see the *OECD Handbook on constructing composite indicators* for an excellent review of the set of choices and their respective advantages and disadvantages: OECD (2008); Oman & Arndt (2006)). Within the corruption literature, especially among institutions like the World Bank and Transparency International, there has been an active debate about the optimal approach to aggregate data —with special concern aimed at minimizing the inflation of error due to sampling idiosyncrasies, instrument biases, or the overweighting of small unrepresentative data (Atkeson et al., 2015; Escresa & Picci, 2015; Fazekas & Kocsis, 2017a, 2017b; Fazekas & Tóth, 2017; Neumann & Graeff, 2010).

Although these discussion influenced the choice of models deployed in the paper here —which was done in accordance with the guidelines promulgated in the OECD *Handbook on the Construction of Composite Indicators* (OECD, 2008; Oman & Arndt, 2006; Saltelli et al., 2005) —the current work does not face the same burdens of generalizability across cultural contexts nor does it need to sustain longitudinal comparisons over decades. As such, rather than spending resources comparing the relative merits of different approaches, the current work side-steps this debate and only presents the results from the aggregate model constructed using structural equation modeling (SEM) (McArdle & Kadlec, 2013).

Any concerns with the use of the SEM-based approach regarding the potential for overfit or confirmation bias can be robustly addressed by the fact that I replicate the model presented in Study 1 by using the estimates extracted from the current sample and using them to construct the model in Study 2. This is especially reassuring given the significant difference in the context surrounding the two elections as well as the fact that very different populations were on the winning side of the election (Hillary won in the primaries but lost in the Presidential election). This will be addressed in greater detail in Study 2.

3.2.2 Concerns Regarding Construct Dimensionality

Construct Dimensionality: The Potential Influence of Psychological and Decision Processes

There have been numerous, rigorous debates and disagreements surrounding the choice of methods to create a "unidimensional" composite indicator for perceived corruption (Andersson & Heywood, 2009; Brand & Bradley, 2012; Donchev & Ujhelyi, 2014; Kaufmann et al., 2010, 2011; Lambsdorff, 2006; Olken, 2009). At the same time, there have also been significant calls to refine the construct and move away from a unidimensional composite (Andersson, 2017; Hawken & Munck, 2011; Heath et al., 2016).

In agreement with such calls towards a multidimensional construct, the current work presents theoretical and evidentiary grounds that suggest a unidimensional model of perceived corruption would occlude crucial patterns in the data. Although, prior work has raised similar concerns regarding the inadequacy of a unidimensional construct on theoretical and empirical grounds (Alemann, 2004; Andersson, 2017; Heath et al., 2016; Langbein & Knack, 2010), the current work proposes that there may be cognitive factors accompanying abstract political judgments that may impose multi-dimensionality on perception-based measures in and of themselves.

Thus, in contrast to work from political science and development studies, which tends to emphasize the deconstruction of a unidimensional measure of corruption into subcomponents that reflect political constructs (Alemann, 2004; Bauer & Freitag, 2018; Mendes Fialho, 2017; Neumann & Graeff, 2010; Warren, 2017), the current work proposes that —given the reliance on perception-based measures —the search for latent dimensions may be well informed by accounting for the psychological processes driving complex and ambiguous judgments like perceived corruption.

As such, I propose a two-factor model that primarily distinguishes between perceived corruption that is personally or socially-relevant (hereinafter, *social corruption*) and perceived corruption that impacts the abstract structure of government, civil society and industry (hereinafter, *systemic corruption*). I then demonstrate the separability of these constructs by showing that they are differentially responsive to extraneous factors —namely, election effects. As additional confirmation of the separability of these two dimensions, I also present evidence showing that both these factors (a) have a significant and robust relationship to an individual's likelihood of cheating; and, (b) most crucially, these two factors influences the likelihood of cheating in entirely opposite directions.

Taken together, the evidence from the current studies suggests that these cognitively-driven constraints are not readily dismissed as participant inattention and do not appear to reflect the use of survey responses by participants to air grievances about unfavorable election outcomes. The measured perceptions are remarkably well-structured across different samples, appear to capture genuine beliefs and appear to have reliable downstream consequences.

Assuming Unidimensionality: Item Response Analysis and Measures of Reliability and Differentiation

As a starting point, I begin by first assuming unidimensionality. Under this assumption, I conducted an item response analysis of the 20 individual items in the corruption survey —which serves as a preliminary analysis of the consistency, reliability and discriminability of the current measures of perceived corruption. The item analysis is based upon four primary indicators: (i) Cronbach's α (Cronbach, 1951), (ii) mean inter-item correlation (IIC), (iii) item difficulty (also known as item location), and (iv) item discrimination (see Table 3.1 for the analysis details).

Table 3.1: Item Response Analysis showed a high Cronbach's $\alpha = 0.92$ but a low mean inter-item-correlation = 0.384. Assuming a unidimensional scale, the item discrimination index showed that items about elections and political corruption were the most able to discriminate between participants on behavior being measures (i.e. unidimensional scale of perceived corruption), whereas item difficulty showed that response to questions about the personal impact of corruption provide the most ability to discriminate between participants in general.

Items	Missings	Mean	SD	Skew	Item Difficulty	Item Discrimination	α if deleted
Question 1							
Political Parties	1.88 %	82.81	22.84	-1.1	0.83 ↑	0.641	0.92
Congress	2.63 %	77.6	26.38	-0.84	0.78	0.643	0.92
Media	1.50 %	76.57	27.18	-0.76	0.77	0.52 ↑	0.92
Business	1.88 %	76.04	26.58	-0.76	0.76	0.47 ↑	0.92
Public Officials	2.00 %	67.12	27.71	-0.22	0.67 ↓	0.63	0.92
Taxes	3.13 %	68.95	29.46	-0.52	0.69	0.578	0.92
Elections	2.25 %	70.68	29.17	-0.58	0.71	0.676 ↓	0.92
Fed Govt	1.38 %	74.58	26.85	-0.62	0.75	0.707 ↓	0.92
Local Govt	2.00 %	63.62	27.92	-0.12	0.64 ↓	0.561	0.92
Question 2							
Political Corruption	0.00 %	75.06	22.2	-0.85	0.75	0.753 ↓	0.92
Corruption Election	0.00 %	71.9	24.75	-0.81	0.72	0.734 ↓	0.91
Money Election	0.00 %	84.76	19.2	-1.73	0.85 ↑	0.51 ↑	0.92
Question 3							
Personal Impact	1.50 %	44.88	31.22	0.31	0.45 ↓	0.53 ↑	0.92
Business Impact	1.63 %	79.89	24.51	-0.95	0.8	0.589	0.92
Political Impact	1.50 %	87.56	20.99	-1.63	0.88 ↑	0.62	0.92
Culture Impact	1.38 %	71.86	26.49	-0.53	0.72	0.612	0.92
Question 4							
Personal Severity	2.13 %	34.78	36.47	0.54	0.35 ↓	0.442	0.92
Business Severity	1.25 %	78.52	27.79	-0.84	0.79	0.54 ↑	0.92
Political Severity	1.13 %	86.84	23.83	-1.53	0.87 ↑	0.567	0.92
Culture Severity	1.88 %	67.03	32.46	-0.47	0.67 ↓	0.55 ↑	0.92

Note. Mean inter-item-correlation=0.384; Cronbach's α=0.921

The first two indicators are indicators at the "test-level" i.e. at the level of the entire survey. Mean IIC can be though to capture "item redundancy" i.e. the degree to which items in the scale have overlapping measurement target (Piedmont, 2014). Mean IIC is used to compute Cronbach's α, which can be interpreted as the (lower bound) of a measure of "reliability", defined as the correlation between the true and observed score (for concerns, see Sijtsma, 2009). The last two indicators provide item-level information. Item discrimination can be interpreted as a measure of how well an item identifies and differentiates between people with different levels of the underlying measure. Item location / item difficulty is a complementary measure that allows me to identify items with maximum differentiation capacity.

Cronbach's alpha Cronbach's alpha is the most commonly relied upon measure of scale reliability in psychometric research (Trizano-Hermosilla & Alvarado, 2016). In the current study, the Cronbach's $\alpha = 0.92$, which is usually considered to be a good indicator of internal consistency (or, reliability) of the items in the scale. However, as recent work has shown, α is neither an indicator of underlying dimensionality nor a measure of *scale validity* (Sijtsma, 2009). Thus, a high value of α *does not imply unidimensionality* and any examinations of the dimensionality or validity of a scale are most appropriately addressed through confirmatory factor analyses (see following sections as well as estimates of general factor saturation like *McDonald's* ω). This is because Cronbach's α is only meaningful under the *a priori* assumption of unidimensionality. As a recent review article on "scale reliability" concluded, α "assesses neither the reliability of a test nor the internal consistency of a test, *unless* the test items all represent just one factor" (p. 728, Revelle & Condon, 2018, emphasis in original). In fact, recent work suggests that if this assumption is invalid, it is better to substitute α for Structural Equation Modeling (SEM) based indicators of reliability (Cho & Kim, 2015).

Inter-Item Correlation (IIC) Given the measure of $\alpha = 0.92$, it may appear concerning that the mean *inter-item correlation* for the current set of measures was only 0.38. Although α and mean IIC tend to move together, the divergence here is not evidence of any error *per* *se*. Crucially, since Cronbach's α depends not only on the correlation of the items, but also the number of items in the scale, it may be best considered a measure of how sufficiently a scale is supplied with items. If enough items are present, it is possible to have a high α even if items are weakly correlated, which is what I see in the current data (Revelle & Condon, 2018).[3] This distinction is particularly relevant given the low average inter-item correlation for the current set of measures. In fact, as is discussed in the subsequent section, the low IIC is the first of a set preliminary indicators suggesting that unidimensionality is unlikely.

Item-Discrimination Index This index is thought to measure the degree to which an item differentiates correctly among participants on the variable that the survey is designed to measure. For each item, the value D ranges from 0-1, the higher the values the more the item discriminates between participants. As shown in Table 4, a *d-index* in the top quartile is associated with items focused on (in descending order) elections, political corruption, and corruption in the Federal Government, whereas items with the d-index in the bottom quartile were focused on the effects of corruption on personal life, money in elections,[4] business, and the media. Since this index measures the degree to which a given item can differentiate participants *on the target being measured by the scale (i.e. here, the target being an assumed unidimensional scale of perceived corruption),* an examination of the subset of items with

[3]This is because α is a function of two features: (i) the number of items in the scale; and (ii) the average inter-item correlation in the scale - thus, even in measures where items has low inter-item correlation, the α measure of "reliability of the total scale can be increased by merely adding items" (Revelle & Condon, 2018). As explained in a recent chapter in *The Wiley Handbook of Psychometric Testing*, "just six highly correlated items (r = .4) are needed to achieve an α = .8, which requires 16 items with more typical correlations (r = .2). Even with barely related items (e.g., r = .05), an α of .8 may be achieved with 76 items. α values of .90 require 14, 21, 36, and 81 for intercorrelations of .4, .3, .2, and .1 respectively." (pg.725, Revelle & Condon, 2018).

[4]There is good prior reason to believe that the question "How much does money influence the selection of presidential candidates?" is likely to be heavily influenced by whether your candidate won or lost based upon motivated reasoning. Losers are likely to believe that the loss was largely determined by "money in politics" whereas winners are likely to believe the win was due to the superiority of their preferred candidate. There is also evidence that this response was highly determined by political affiliation. The lack of grouping information about (a) candidate supported and (b) election outcome should make this item appear to be quite uninformative, which is why it scores poorly on both indicators of item discriminability.

the highest D-index suggests that imposing a unidimensional structure results in a scale that largely tracks perceptions of political corruption and is poorly tuned to perceptions of corruption in the business, social, or personal spheres.

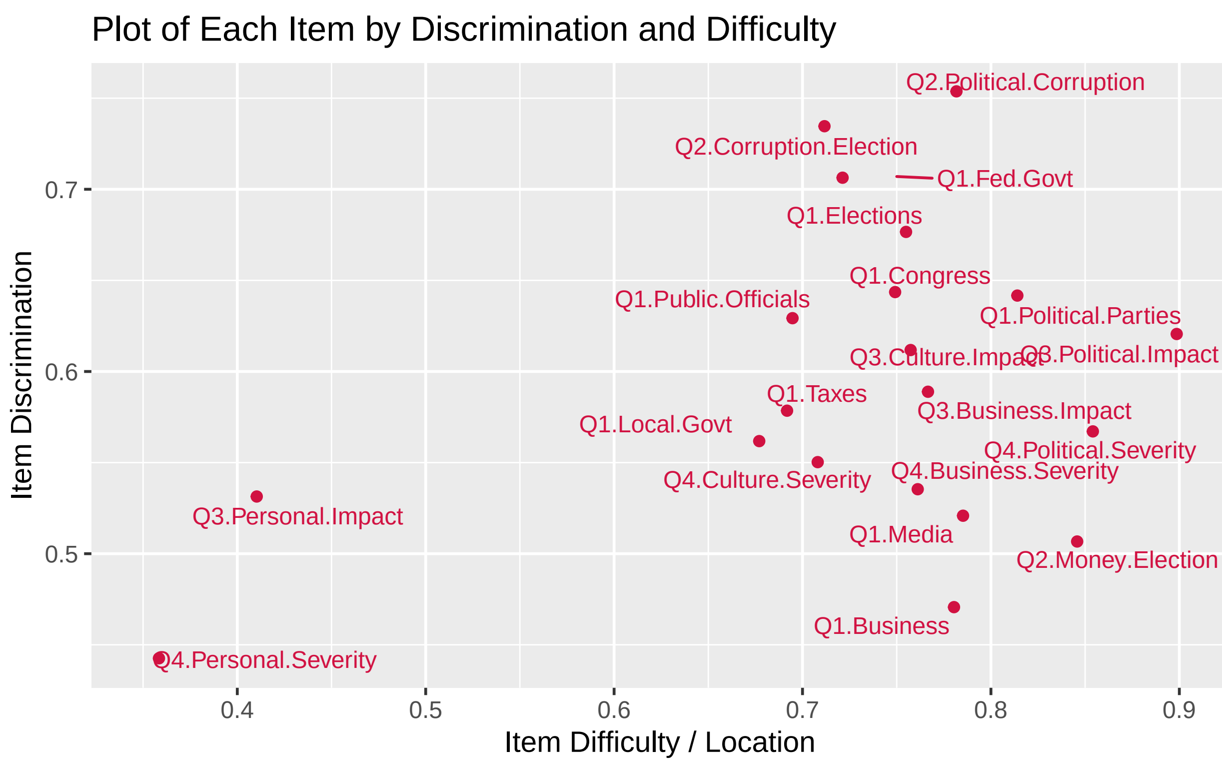


Figure 3.1: A plot of each item by its item discrimination and item difficulty scores shows the two items measuring personally-relevant corruption are clearly offset from other measures. Items dealing with political corruption, elections, and the Federal Government distinctly clustered as the items that were highest on item discrimination scores.

Item-Difficulty / Item-Location Index The item-difficulty index is a complementary measure that also aims to identify items with maximum differentiation capacity. However, this index is based upon the notion that items that are "too easy" or "too difficult" cannot contribute to the reliability of a scale i.e. as difficulty approaches 0 or 1, the item provides less and less differentiating information about the participants. Instead, items with difficulty values, p, closest to 0.5 are considered the most capable of differentiating between participants *in general.*

Unlike the item-discrimination index, item-difficulty does not measure the ability to differentiate based upon the assumed underlying measure. Using this index, I see that the most differentiating items focus on personally-relevant effects of corruption and local government (and, to a lesser degree public officials and taxes). Whereas items with the least discriminatory influence in general are focused on corruption in politics (including, political parties) and money in elections.[5]

It is worth noting that all the items with p-scores ≈ 0.5 involve judgments that are differentiable from political corruption and, in accordance with our hypothesis, involve local, less abstract judgments. The fact that these items are concomitantly associated with a low D-value also provides support for a multi-factor model. However, these four measures are indicative of reliability and discriminability, not dimension validity. Instead, these results serve to establish grounds for more targeted analyses of dimensional structure based upon exploratory factor analysis.

3.2.3 Using Exploratory Factor Analysis to Examine Response Integrity, Measure Validity and Dimensional Structure

Estimation Procedures and Rotation Methods Used for Factor Analysis

I used the Psych package (Revelle, 2017) in the R (Version 3.6.3; R Core Team, 2020) Statistical Environment to perform all factor analyses presented here. Since there are multiple estimation procedures that can be used, I computed the factor loadings using six common estimators (maximum likelihood, minimum residual, principle axes, minimize weighted chi-squares, weighted least squares, and generalized weighted least squares) and used the *Tucker Index of Factor Congruence* to compare the factor loadings across the different estimation procedures. In all cases, the factor estimates were robust to the choice of estimator with

[5]See previous footnote on this issue.

factor congruence between 0.99 and 1, suggesting that the estimated loadings only differed from each other by a scalar. For consistency, I used the minimum residual estimator and the varimax method for rotation, since it is known to maximize parsimony and to yield distinct loadings between individual indicators and the different latent factors (i.e. it minimizes the likelihood of that an observed variable has a large loading estimate for more than one factor) (Ertel, 2011; Jöreskog & Moustaki, 2001).

Preliminary Analysis: Correlation Structure in the Response Data

As a first step prior to exploratory factor analysis, I examined the correlation structure of the data. If the correlation between the indicators is weak, it is unlikely that they share common factors. As is self-evident by examining the histogram of correlation coefficients in Figure 3.2 as well as the correlation matrix in Table 3.2, there is a wide distribution in the correlation coefficients (ranging from 0.1 to 0.8). And, although there are many moderately correlated measures (marked in yellow on the table), there appear to be a handful of measures that are quite highly correlated with each other (marked in green on the table) and a handful of others that are very poorly correlated with most items in the survey (marked in red on the table). This suggests that the indicators are unlikely to arise from a single unidimensional latent factor. Instead, it appears likely that a multi-factor model would provide a better fit to the response data. Next, I present evidence from a simple principle-components factor model to address whether responses are granular enough to justify a multi-dimensional construct.

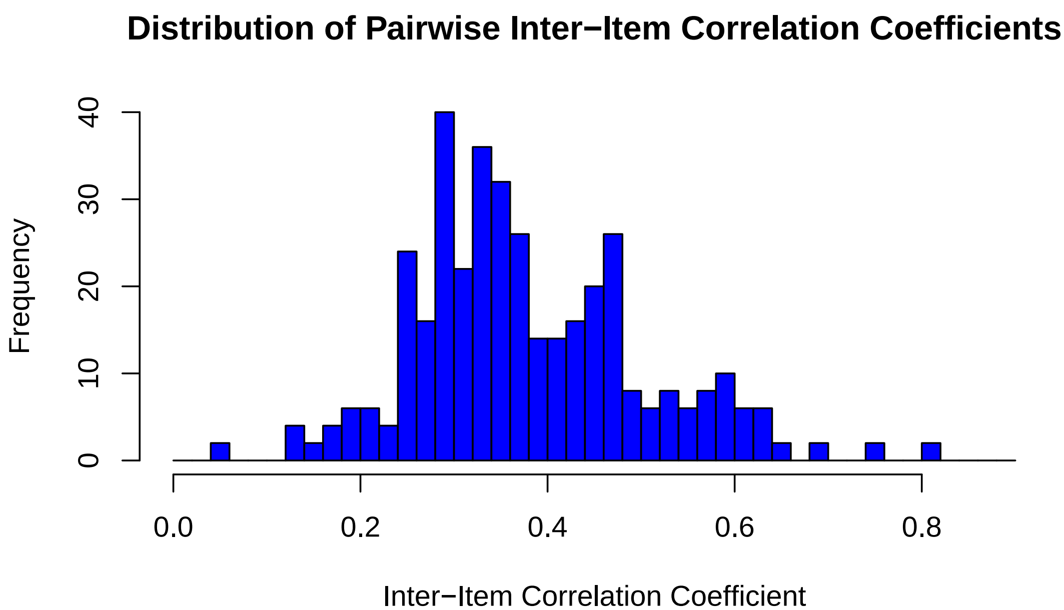


Figure 3.2: Histogram showing the distribution of correlation coefficients for all pairs of items in the perception of corruption survey. As can be seen from the distribution, there is a wide range of correlation coefficients (ranging from 0.1 to 0.8). Some items appear to correlate poorly with others, while others correlate very highly. The pairwise correlation coefficient matrix is presented in Table 3.2. Given the wide-range in distribution of correlation coefficients, it seems unlikely that the indicators arise from a single unidimensional latent factor —thereby supporting a multi-dimensional approach.

Table 3.2: Correlation across different items in the corruption survey.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Q1.Political.Parties																			
Q1.Congress	.6																		
Q1.Media	.4	.4																	
Q1.Business	.3	.4	.2																
Q1.Public.Officials	.4	.5	.4	.4															
Q1.Taxes	.4	.4	.3	.3	.5														
Q1.Elections	.6	.6	.5	.3	.5	.5													
Q1.Fed.Govt	.6	.6	.4	.3	.5	.5	.6												
Q1.Local.Govt	.4	.4	.3	.4	.6	.4	.4	.5											
Q2.Political.Corruption	.6	.6	.5	.3	.5	.5	.6	.6	.4										
Q2.Corruption.Election	.5	.5	.5	.3	.4	.5	.6	.6	.4	.8									
Q2.Money.Election	.5	.4	.3	.3	.3	.3	.4	.4	.3	.6	.6								
Q3.Personal.Impact	.2	.3	.2	.2	.4	.3	.4	.3	.3	.3	.4	.2							
Q3.Business.Impact	.4	.4	.3	.5	.4	.3	.3	.4	.3	.4	.4	.3	.4						
Q3.Political.Impact	.6	.4	.4	.2	.3	.3	.4	.5	.3	.6	.5	.5	.3	.5					
Q3.Culture.Impact	.3	.3	.4	.3	.4	.3	.4	.4	.3	.4	.5	.3	.5	.4	.5				
Q4.Personal.Severity	.1	.2	.2	.2	.3	.3	.2	.3	.3	.3	.3	.1	.8	.2	.1	.4			
Q4.Business.Severity	.3	.3	.2	.5	.3	.3	.3	.3	.3	.4	.4	.3	.3	.6	.4	.3	.3		
Q4.Political.Severity	.5	.4	.3	.3	.3	.3	.4	.4	.3	.5	.5	.4	.2	.4	.6	.4	.2	.4	
Q4.Culture.Severity	.2	.2	.3	.2	.4	.3	.3	.4	.3	.4	.4	.2	.4	.3	.3	.7	.5	.3	.4

Note. Computed correlation used Pearson-method with listwise-deletion. All correlation coefficients < 0.3 are shown in red, < 0.6 in yellow, and > 0.6 in green.

Principal Components: Decomposing the Observed Variance as a Means of Estimating Response Validity

The primary goal of the principal components analysis (PCA) for our purposes is to allay concerns that participants were either not capable or unwilling to respond to the corruption survey in a manner that is sensitive to the particularities of a given question. As such, I use the PCA estimates to group the variables and then assess whether the resulting "clusters" establish a *prima facie* case that the participant responses to different items are nuanced, responsive and meaningfully differentiable from each other. Although the normal approach to PCA is to start with a single factor and increment upwards while using a measure of fit as a stopping criterion, the current goal is better served by moving from the largest reasonable number of factors and decrementing.

One common approach to determining the optimal number of factors or components in a correlation matrix is to examine the "Scree" plot of the successive eigenvalues.[6] However, these graphical approaches can be particularly susceptible to research biases. Instead, I sought a non-graphical solution, relying especially on an alternative approach called "Parallel Analysis" which compares the scree of factors of the observed data with that of a randomly generated set of data matrices of the same size as the observed data. Using the *nFactors* package in R (Raiche, 2010), I computed multiple solutions to the Scree test —all of which supported the use of 5 or more components.[7]

So, I began by estimating a five-component model of the data and decrementing. The loadings show the correlation of each variable with the component. The resulting variables "clusters" are shown in the table below. The clusters were created by assigning each variable to the component with the highest loading. However, it should be noted, except for one indi-

[6] Graphical solutions rely upon a visual examination of the scree plot by looking for sharp breaks in the plot, which is used to predict the appropriate number of components or factors to extract.

[7] These solutions matched the estimates found using the *parallel analysis* function in the *Psych* package (bootstrapped using 10,000 simulations) (Revelle, 2017).

cator (Q1. Political Parties), all other indicators only loaded significantly on one component —with loadings being 0.5 or higher. Since the loadings reflect the correlation between each variable and the corresponding component, an average loading of 0.7 per component is quite reassuring.

PCA Loading Patterns: Five Component Model The results for the five-component model are shown in Table 3.3. Other than one exception —the question asking about *Political Parties* —all survey items only loaded significantly on only one component, each item showed a loading > 0.5. Thus, the loadings naturally produced "clusters" —created by assigning each variable to single component on which the variable loaded highest. The columns in Table 3.3 show which survey item (shown in the rows) loads on which component (shown in the columns). And, because of the loading pattern, each component can also be seen as a "cluster" that defines a group of related survey items. For each component-item pairs in Table 3.3, a loading value is shown. This loading value can be understood as the strength of the correlation between each variable and that component.

First, overall, the results are quite reassuring. We see an average loading of 0.7 per component —and, the PCA loadings reflect the correlation between each variable and the corresponding component —an average loading of 0.7 suggests very high correlation between different survey items and extracted components.

Second, even a quick glance at the variable clusters in Table 3.3 reveals a set of surprisingly meaningful clustering, with the first cluster holding items that targeted public institutions, the second cluster holding items that target the political process and elections in specific, the third clustering business related items, the fourth —self-relevant, and the fifth cluster holding items focused on society and culture (including an item asking about media).

In addition the PCA loadings themselves, two other factors support the above-mentioned results. First, many of the items that clustered together were often spread across multiple pages when presented to participants. Second, responses clustered together despite often being collected using different ordinal scales.

In total, even a cautious interpretation of these results suggest that they are initial evidence that participant responses were (i) well-structured and meaningful; (ii) the responses were sensitive to the particularities of different questions; and, (iii) they responses were meaningfully more similar for items within a cluster than for items across different clusters.

PCA Loading Patterns: Incrementally Decreasing the Number of Components from Five to One Subsequent analysis with four factors combines Cluster 4 and 5 from above and left the other clusters largely untouched. A reduction to three components combined all variables dealing with public institutions and politics (Cluster 1 and 2), while leaving the business-related items and the social/personal items distinct. Finally, a reduction to two components results in the combination of the Business and the Public Sector variables (Cluster 1, 2, 3) —leaving behind two clusters —one with measures of global / systemic corruption; and, the second with measures of socially and personally relevant judgments.

Structures in the Data: Stable, Meaningful, Hierarchical Clustering Patterns The specific clustering of variables seems relatively stable and seems to be independent of the statistical procedure used exploratory used for variable classification, including exploratory factor analysis (see Appendix A.5 for plots of 2, 3, and 4-factor models), hierarchical factor analysis (see Appendix A.6 for a hierarchical 2-factor model), and item clustering analysis (see dendrograms in Appendix A.7.1 and A.7.2 for two approaches to grouping survey items using hierarchical clustering). The final takeaway is that the response patterns seen in the current survey reveal a robust and meaningful structure —one that reappears across vastly different approaches and is not an artifact of any particular method for examining dimensionality. Taken together, these analyses support the notion that a nuanced, multi-dimensional measure of perceived corruption can be meaningfully fit to these data.

Table 3.3: Loading matrix for a 5 component PCA showed high and unique loading values for different components. In addition, the clustering of variables in the different components lent itself to easy interpretation —with each component naturally mapping to a self-evident conceptual analogue.

	Principle Components				
	PC: 1	**PC: 2**	**PC: 3**	**PC: 4**	**PC: 5**
Meaningful Analogues of Component Structure	*Public Institutions*	*Political / Electoral*	*Business Sector*	*Personally Relevant*	*Social Cultural*
Loadings					
Q1. Taxes	0.63				
Q1. Congress	0.64				
Q1. Elections	0.64				
Q1. Local Govt	0.66				
Q1. Fed Govt	0.68				
Q1. Public Officials	0.71				
Q1. Political Parties	0.52	0.60			
Q2. Political Corruption		0.66			
Q2. Corruption Election		0.67			
Q4. Political Severity		0.70			
Q2. Money Election		0.72			
Q3. Political Impact		0.74			
Q3. Business Impact			0.74		
Q1. Business			0.75		
Q4. Business Severity			0.76		
Q3. Personal Impact				0.85	
Q4. Personal Severity				0.88	
Q1. Media					0.55
Q3. Culture Impact					0.76
Q4. Culture Severity					0.76
Variance Explained					
SS loadings	3.89	3.76	2.17	1.96	1.90
Proportion Var	0.19	0.19	0.11	0.10	0.10
Cumulative Var	0.19	0.38	0.49	0.59	0.68
Proportion Explained	0.28	0.27	0.16	0.14	0.14
Cumulative Proportion	0.28	0.56	0.72	0.86	1.00

Note. PCA: 5 orthogonal components constructed using Study 1 data (n=718); Rotation: varimax; Test Hypothesis: 5 components are sufficient. The root mean square of the residuals (RMSR) = 0.05 with the empirical $\chi^2 = 688.96$ with prob <.001; Mean item complexity = 1.7 ; Fit based upon off diagonal values = 0.98

3.3 Confirmatory Factor Analysis and Structural Equation Modelling (SEM)

Having established *compliance* by addressing the reliability and integrity of the participant responses to the corruption survey, the analysis now can turn to the primary measurement model using Structural Equation Modeling (SEM) to describe the relationship between the proposed latent variables and the observed indicators.

The Structural Equation Modeling approach detailed here can be understood as composed of a (i) "measurement model," where a confirmatory factor analyses is used to combine the observed indicators into hypothesized latent variables, and a (ii) structural model, which describes the hypothesized relationship between latent variables.

3.3.1 Selection of Statistical Procedures for Estimation

There are numerous approaches to estimating loading of observed variables on the latent factors when Structural Equation Modeling (SEM) (MacCallum & Austin, 2000; Ullman & Bentler, 2003). While it is common to use Ordinary Least Squares (OLS) to find the maximum likelihood (ML) solution, numerous measures in the current study are ordinal and depart significantly from the normality assumption, which can lead to significant errors in estimation (Dreher et al., 2004; Rhemtulla et al., 2012).

Given the non-normality of the underlying data, the current analyses used a *diagonally weighted least squares* (DWLS) estimator in order to explicitly account for the ordinal nature of the observed response variables (Bollen & Maydeu-Olivares, 2007; Li, 2016b), since it has been shown to be superior to ML in estimates of factor loadings, inter-factor correlations, structural coefficients, robust standard errors, and chi-square statistics (Li, 2016a). The SEM models reported here were fit using the Lavaan package (Rosseel, 2011) in R (Version 3.6.3; R Core Team, 2020) and were visualized using the semPlot package (Epskamp, 2015).

3.3.2 Measurement Model: Specification and Fit

Specification of Measurement Model

The measurement model segregated the indicator variables into the four "spheres of life" explicitly specified in the GCB (*political environment*, business *environment*, *the culture and values in society*, and *personal and family life*) as well as an additional latent factor to capture "public institutions" as distinct from the "political environment" (Warren, 2017, 2004).

The specification of the measurement model is shown in Figure 3.3, where each latent variable was defined in terms of non-overlapping, distinct sets of response items. To see the exact wording of the survey items associated with each latent variable in the measurement model, see Appendix A.8.

On the right-hand side of Figure 3.3, each latent variable is depicted using an oval node. On the left-hand side of Figure 3.3, each survey items is shown as a rectangle —both the rectangles and the ovals are color-coded to help identify each item's relationship to its parent latent variable.

The causal relationship between the latent and the manifest variables is shown using directed arrows originating from the latent variable and flowing towards the manifest ones. The loading of the individual survey items on their corresponding latent variable is shown on the links connecting each pair. The bi-directional arrows between latent variables show the residual covariance and the self-directed arrows on each manifest variable (left) show the residual variance that is not captured by the latent variable.

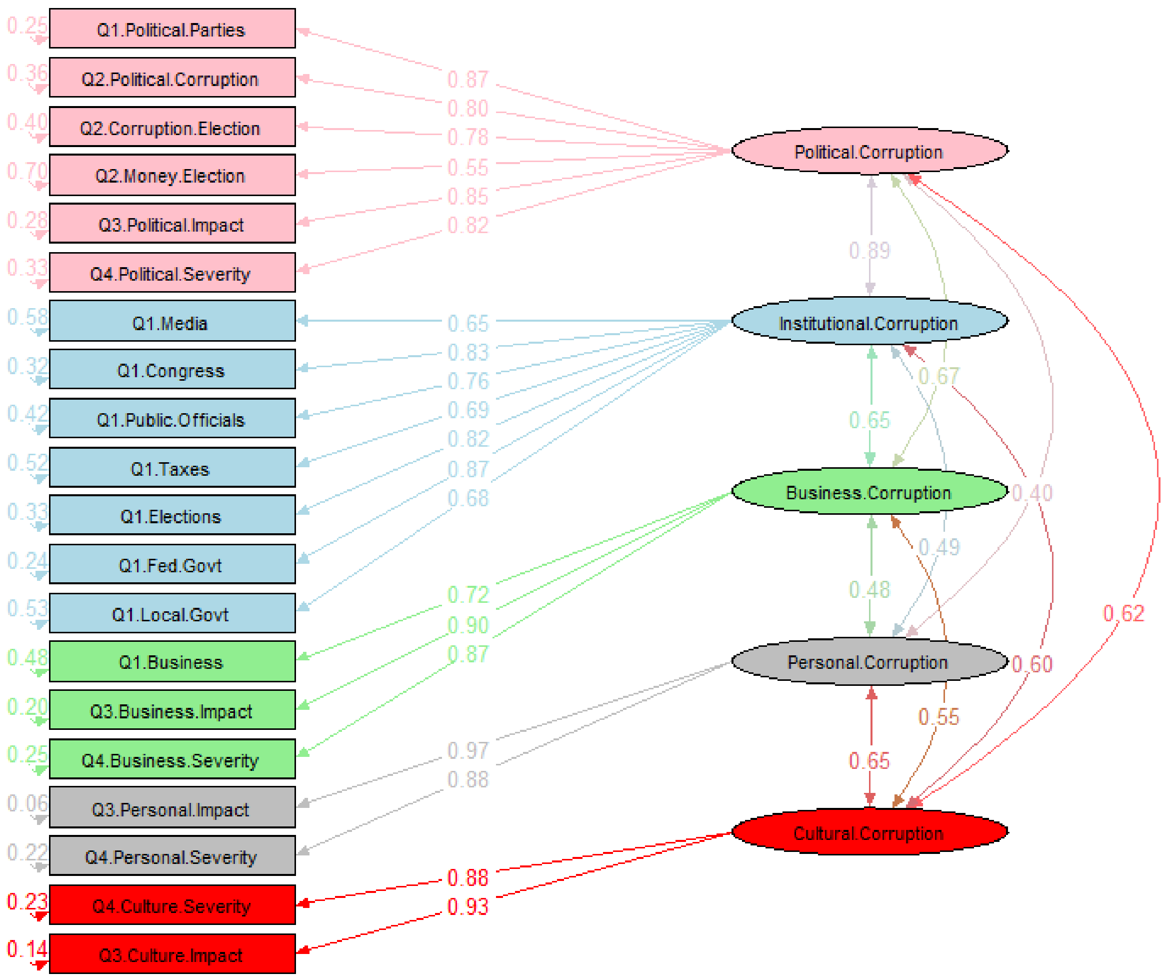


Figure 3.3: Measurement model using CFA —Constructing a five-factor model: Each latent variable was defined in terms of non-overlapping, distinct sets of survey items. The latent variables are shown as oval nodes on the right-side of the figure. Survey items are shown as rectangular boxes on the left-side of the figure. The rectangles are color-coded to capture the mapping to each survey question to one of the five latent variables. To see the exact wording of the survey items associated with each latent variable in the measurement model, see Appendix efMeasurementModelItemQDetails. The causal relationship between the latent and the manifest variables is shown using directed arrows The number of each arrow represents the standardized loadings. Bidirectional arrows between latent variables show the residual covariance. Self-directed arrows on manifest variables (left) show the residual variance that is not captured by the latent variable.

Evaluation of Fit for Measurement Model

This five-factor measurement model performed quite well. First, we see that the standardized loadings were all > 0.3, indicating a decent fit for all items. Second, based upon all the prominent and recommended measures of model fit, the five-factor measurement model showed evidence of a good-or-excellent level of fit to the data, including on (i) recommended measures of goodness-of-fit, like, *Comparative Fit Index (CFI): 0.99*;[8] and, (ii) recommended measures of badness-of-fit, like: *Standardized Root Mean Square Residual (SRMR), 0.059*;[9] and *Root Means Square Error of Approximation (RMSEA): 0.05 (90% CI: 0.05–0.06).*[10] Finally, on all the same measures of model fit, this five-factor model performed significantly better that the unidimensional model, which served as the baseline and showed evidence of mediocre-to-poor model fit (*CFI*: 0.938; *SRMR*: 0.125; *RMSEA*: 0.149; *ECVI*: 4.216).

3.3.3 Structural Model: Specification and Fit

Specification of the Structural Model

As the second step, the relationship between these latent variables was then specified as shown in Figure 3.4. In addition, taking advantage of the SEM approach, the model also specified the hypothesized relationship between perceived corruption and willingness to cheat, allowing us to estimate regression coefficients and test that hypothesis.

[8]For the Comparative Fit Index, cutoffs for acceptable values is 0.9, with values close to 1 indicating that the model and data match nearly identically.

[9]For badness of fit measures like SRMR and RMSEA, the cutoff for acceptable values is usually 0.1 with values closer to 0 being better, since these indices indicate a mismatch between the model and the data.

[10]For RMSEA, recent reviews have suggested cutoffs of 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively (MacCallum & Austin, 2000). For SRMR, a value less than .08 is generally considered a good fit (Hu & Bentler, 1999).

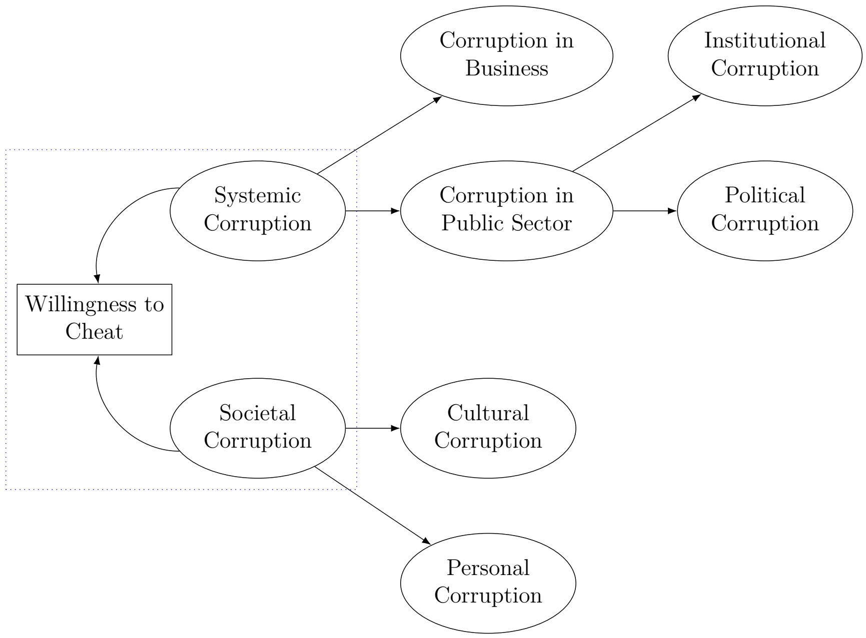


Figure 3.4: Structural Model: Mapping the hypothesized relationship between the various first and second-order latent variables. The diagram also shows the hypothesized relationship between perceived corruption and willingness to cheat (blue box), which could be tested using the SEM model.

Evaluation of Fit for Measurement Model

All standard measures of model fit suggested a good fit (CFI: 0.991; SRMR: 0.056; RMSEA: 0.046; RMSEA p-value < 0.98; Baseline RMSEA: 0.46; ECVI: 1.233). All loadings were significantly different from 0 with $p < 0.001$ as were the estimated covariances between electoral engagement and the perceived corruption variables. The loadings for the full model are presented in Table 3.4 —with loadings of individual survey items onto their corresponding latent variables; loading of intermediate latent variables on the higher-order ones; statistical tests for each loading as well as the residual variance are included in the accompanying table. Finally, the full structural model (measurement / structural components) can also be seen in the model diagram in Figure 3.5.

Estimates with the std. error for each estimate and test statistic. All estimates were significantly different from 0 at p < .01.

Latent Variable		Indicator	Std. Estimate	Std. Error	Z	p < |Z|
	→	Care Politics	0.65	0.05	12.57	0.000
Electoral Engagement	→	Care Election	0.82	0.05	15.15	0.000
	→	Candidate Support	0.80	0.06	13.29	0.000
Personal Corruption	→	Q3. Personal. Impact	0.98	0.03	34.52	0.000
	→	Q4. Personal. Severity	0.88	0.03	33.08	0.000
Cultural Corruption	→	Q4. Culture. Severity	0.87	0.02	44.23	0.000
	→	Q3. Culture. Impact	0.93	0.02	47.88	0.000
	→	Q1. Business	0.72	0.03	25.04	0.000
Business Corruption	→	Q3. Business. Impact	0.90	0.02	45.31	0.000
	→	Q4. Business. Severity	0.87	0.02	35.87	0.000
	→	Q1. Political. Parties	0.86	0.02	45.18	0.000
	→	Q2. Political. Corruption	0.81	0.02	42.50	0.000
Political Corruption	→	Q2. Corruption. Election	0.78	0.02	36.95	0.000
	→	Q2. Money. Election	0.55	0.03	19.48	0.000
	→	Q3. Political. Impact	0.85	0.02	39.28	0.000
	→	Q4. Political. Severity	0.81	0.03	29.96	0.000
	→	Q1. Media	0.65	0.03	21.87	0.000
	→	Q1. Public. Officials	0.74	0.02	32.99	0.000
	→	Q1. Taxes	0.70	0.02	28.09	0.000
Institutional Corruption	→	Q1. Elections	0.81	0.02	45.34	0.000
	→	Q1. Federal Government	0.87	0.02	56.08	0.000
	→	Q1. Local Government	0.67	0.03	26.38	0.000
	→	Q1. Congress	0.82	0.02	45.75	0.000
Public. Sector	→	Political. Corruption	0.95	0.02	53.90	0.000
	→	Institutional. Corruption	0.93	0.02	52.32	0.000
Systemic Corruption	→	Public. Sector	0.88	0.03	30.14	0.000
	→	Business. Corruption	0.80	0.03	26.54	0.000
Social Corruption	→	Personal. Corruption	0.70	0.04	19.26	0.000
	→	Cultural. Corruption	0.91	0.03	26.58	0.000
Electoral. Engagement	↔	Systemic. Corruption	0.16	0.05	3.11	0.002
Electoral. Engagement	↔	Personal.Social.Corruption	0.14	0.05	2.97	0.003
Social Corruption	→	Cheating Amount	0.38	0.12	3.24	0.001
Systemic Corruption	→	Cheating Amount	-0.39	0.12	-3.38	0.001

Table 3.4: Loadings for Latent and Manifest Variables in the full model. Estimates with the std. error for each estimate and test statistic. All estimates were significantly different from 0 at p < .01.

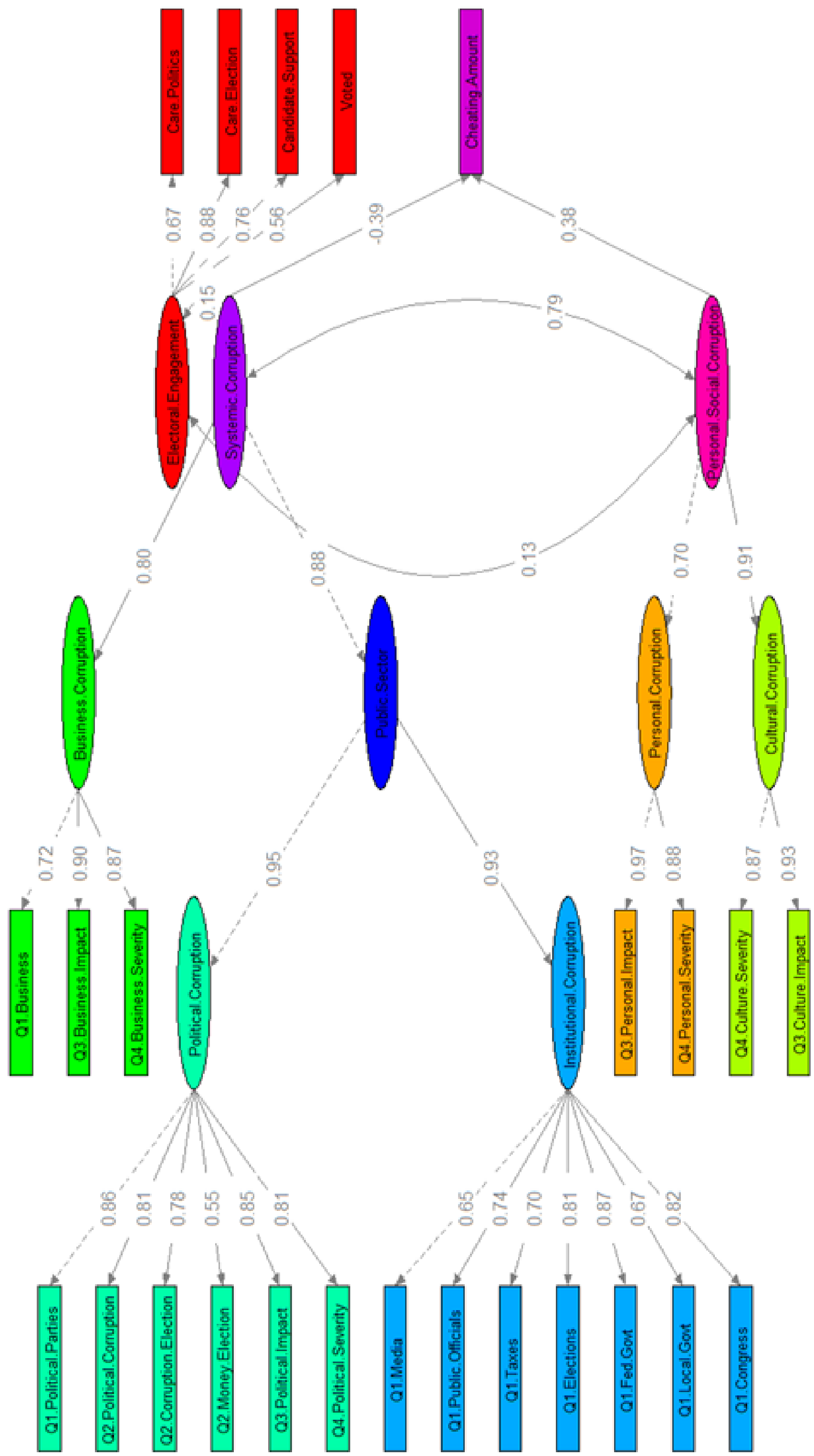


Figure 3.5: Complete Measurement and Structural Model

CHAPTER 4

RESULTS: STUDY 1 - DEMOCRATIC CALIFORNIA PRIMARY - JUNE 7TH 2016

The *California Primary* was widely discussed in the news at the time as the make-or-break moment for the candidacy of Senator Bernie Sanders i.e. it would determine who would be nominated as the presidential candidate for the Democratic Party. Due to the broad interest and clear implications for the outcome, the California Primary served as the first "naturalistic shock" to test our hypothesis.

4.1 Participants

The California Democratic Primary election took place on 7th June 2016. On 6th June (one day prior to the election), 398 US-based participants were recruited on Amazon's Mechanical Turk to complete the study for the pre-election sample. On 8th June (one after the election), once the results of the election had been widely disseminated online, an additional 401 US-based participants were recruited to complete the study as part of the post-election sample. These 799 participants constituted the between-subjects sample.

4.1.1 Participant Exclusion Criteria

Only minimal data exclusion criteria were applied, namely: (i) participants that did not complete all sections of the study were excluded; (ii) participants with multiple submissions were excluded —if the multiple responses started concurrently, both excluded, otherwise, first response retained; (iii) multiple responses from the same IP address were excluded.

4.1.2 Equivalence among Experimental Groups

To start, I confirmed that the samples were similar before and after the election. The pre-and-post election samples did not differ in terms of demographic characteristics ($p > 0.6$ on all tests), the number of participants reporting support for any of the major candidates ($p > 0.7$ on all tests), their political affiliation ($p > 0.9$ on all tests), or the degree of support for their preferred candidate ($t = -0.86,\ df = 710,\ p = 0.39$). However, there appeared to be a possible (albeit non-significant and minor) increase in the reported level of caring about the primaries after the election ($t = -1.7,\ df = 795.51,\ p = 0.08$) and an increase in the reported level of caring about politics in general ($t = -1.7,\ df = 795.51,\ p = 0.08$). This was accounted for when creating the composite measure of electoral engagement.

4.1.3 Demographic Characteristics

I examined the basic demographic composition of the participant pool with a special focus on participant characteristics known to influence judgments of perceived corruption. Of 799 participants, 342 or 43% were female, 455 or 57% were male, and 2 participants did not identify with either gender —these participants were excluded from any analyses involving gender due to small sample size in this category. The average age was 34.7 (SD = 11.5). Of the sample, 39% of participants reported having finished a bachelor's degree, 38% completed some college or an associate's degree, 12% completed high-school or less, and 10% completed some graduate education (Master's, Professional, or PhD).

In terms of income, 20% reported earning less than $10,000; 16% reported earning between $10–19,000; 15% between $30–39,000; 10% between $40–49,000; 10% between $50–59,000; 6% between $60–69,000; and the remainder were spread with less than <5% in any of the remaining bins. In terms of employment, 609 participants reported being employed, while 184 reported having no unemployment. Of there, 447 were employed full time, 187 part time, and the remainder either didn't reply or were not employed. Overall, 325 or 40%

of participants were employed in the private sector; 187 or 24% in the public sector, 53 or 7% run household or individual businesses, 32 or 4% have other employment, and the remaining are either unemployed or did not reply. In addition, 115 participants (14%) reported being currently enrolled as students.

4.1.4 Political Characteristics

Party Affiliation

As shown in Table 4.1, out of the 799 participants, 342 participants (43%) self-identified as Democrats, 260 (33%) as Independents, 139 (17%) as Republicans, 27 (3%) as Third-Party, and the remaining 31 (4%) were not sure. Most crucially, as the table below shows, the level of party affiliation did not differ across the pre-and-post election sample, $\chi^2(3,\ N = 799) = 0.48,\ p = .923,\ V = .02$ (with pre-post comparisons for Boolean dummy membership variables resulting in p < 0.99 for Democrat, p < 0.85 for Republicans, and p < 0.65 for Independents).

Table 4.1: Study 1 - Pre-vs-Post Election Sample: Consistent Distribution of Party Affiliation.

Political Affiliation	Pre-Election Sample	Post-Election Sample	Total
Democrat	170	172	342
Independent	133	127	260
Republican	68	71	139
Third Party	16	11	27
Unsure	7	10	17
Don't know / NA	4	10	14
Total	398	401	799

$\chi^2(3,\ N = 799) = 0.48,\ p = .923,\ V = .02$

Table 4.2: Pre-vs-Post Election Sample: Consistent Support for Hillary Clinton

Political Affiliation	Experimental Group: Pre-Election Sample	Experimental Group: Post-Election Sample	Total
Candidate Supported			
Voted: Other	351	359	710
Voted: Hillary	47	42	89
Total	398	401	799

Note. $\chi^2(1,\ N = 799) = 0.24,\ p = .626,\ V = .02$

Candidate Choice

From the entire sample of 799, 89 participants (26% of the Democrats) supported Hillary Clinton, with 47 in the pre-election sample and 42 in the post-election, $\chi^2(1,\ N = 89) = 0.28$, $p = .596,\ V = .06$, all 89 of whom identified as Democrats. An additional 239 participants (70% of the Democrats) supported Bernie Sanders with 116 in the pre-election sample and 123 in the post-election sample, $\chi^2(1,\ N = 239) = 0.21$, $p = .651,\ V = .03$, all 239 of whom identified as Democrats as well. Only 14 participants who identified as Democrats did not support one of these two candidates. On the Republican side, 67 out of 139 (48%) supported Donald Trump, with the support for other candidates spread across other candidates (with no candidate getting more than 15% of the remaining vote).

Representativeness of mTurk sample

This distribution of party affiliation and candidate choice in the study sample suggests that the online mTurk participant pool may not be quite representative of the national populace, with Hillary supporters being under-represented. I saw this pattern again in Study 2. In general, it appears that mTurk online samples tend to be more liberal than more nationally representative samples. In the Section 5.1.5, I address the issue of sample representativeness in greater detail —and, after reviewing the literature —conclude that the specific issues facing the representativeness of mTurk samples does not impact any of the inferences drawn in the current context.

Political and Electoral Engagement

Multiple measures were used to determine the level of political and electoral engagement. First, participants were asked about their general interest in politics, how much they care about the primaries, how strongly they supported their preferred candidate, and whether they voted in any primaries.

Voting Patterns A total of 448 (56%) of participants reported voting in the primaries, whereas 351 (44%) reported not voting. However, among self-identified Democrats, 230 or 67% reported voting and only 112 or 33% reported not voting. A similar pattern was seen among self-identified Republicans with 87 (63%) reporting that they voted in the primaries. On the other hand, the pattern was reversed for self-identified Independents, only 112 (42%) reported voting in any primary. Thus, I see divergent patterns of voting between Democrats and Independents, $\chi^2(1,\ N = 602) = 34.20,\ p < .001,\ V = .24$, between Republicans and Independents, $\chi^2(1,\ N = 399) = 13.03,\ p < .001,\ V = .18$, but not between Republicans and Democrats, $\chi^2(1,\ N = 481) = 0.76,\ p = .383,\ V = .04$, suggesting higher engagement among self-identified partisans.

Self-Reported Interest and Engagement On average, participants reported high levels of interest in politics, $M = 68.31$ ($SD = 26.52$), moderately high levels of caring about the primaries, $M = 72.17$ ($SD = 29.27$), high levels of support for their favored candidate, $M = 74.99$ ($SD = 25.51$), and remarkably high levels of awareness about the outcome of the primaries, $M = 80.34$ ($SD = 25.07$), especially when I consider that the post-election participants completed the study the very next day after the election.

4.2 Perception of Corruption - Effect of the Democratic California Primary

4.2.1 Effect of Election and Candidate Support on Perceived Corruption

I used the above-described SEM model to estimate values of the hypothesized latent variables for all participants. Only 637 participants answered all items in the survey; the remainder had one or more missing responses. Since imputation approaches like full-information ML (FIML) are not yet implemented for categorical data in Lavaan (Rosseel, 2011), I used "pairwise estimation." This meant all available data was used; however, each covariance is based on different Ns. The analyses were also performed using "listwise deletion" (i.e. restricting to only the 637 participants who answered all items). All statistical tests and subsequent conclusions remained substantively unchanged.

Two-Dimensional Index of Perceived Corruption: Systemic vs Social

In the analyses described below, I used the two-dimensional index of perceived corruption described above. As before, I tested the hypothesis that there should be an increase in perceived corruption before and after the primary, with the direction of the effect being determined by whether my preferred candidate won or lost the election. The table below shows the results of regressing perceived corruption on the *Experimental Group* (Pre-Post Election) x a *Dummy Hillary* (voted for Hillary or not). The regression was run for both *Perceived Systemic Corruption* and *Perceived Social Corruption.* For ease of interpretation, I rescaled both perceived corruption measures to range from 0 to 100. None of the results are impacted by this rescaling. For the sake of simplicity, the results presented here do not include estimates for demographic control variables. The inclusion of these demographic variables did not significantly alter any of the results reported here.

Systemic Corruption

The primary analysis examines whether there was an effect of election outcome (i.e. whether one's preferred candidate won or or lost) on perceived corruption. To that end, I regressed the measure of *Systemic Corruption* on $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Voted: Hillary Clinton}]\!] \times \underset{\{0,1\}}{\mathbb{1}} [\![\textit{Post-Election}]\!]$, where $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Voted: Hillary Clinton}]\!]$ is an indicator / dummy variable that equals 1 for participants that voted for Hillary Clinton and 0 for those participants that did not and $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Post-Election}]\!]$ is an indicator / dummy variable that equals 1 for the post-election participants and 0 for the pre-election participants. The full results of the regression are reported in Table 4.4 Column (1). The omnibus test for the multivariate regression was significant, $F(3, 707) = 11.29$, $p < .001$. However, when I examined the two contrasts of interest (see Table 4.3), it appears that the election effect is only significant for the non-Clinton supporters, for whom the results of the California primary increased their perceptions of *systemic corruption* $M = 66.48$, 95% CI [64.60, 68.35] to $M = 69.71$, 95% CI [67.84, 71.57] (election effect = 3.23, $t(707) = 2.40$, $p = .017$). For Hillary Clinton supporters, the results of the California primary did decrease their perceptions of systemic corruption from $M = 59.21$, 95% CI [54.40, 64.03] pre-election to $M = 56.71$, 95% CI [51.55, 61.86] post-election, however, the difference was not statistically significant at the 0.05 level (election effect = -2.51, $t(707) = -0.70$, $p = .486$). Overall, the pattern appears to support the hypothesis that the election outcomes impact perceptions of systemic corruption —increasing if one's preferred candidate loses an election and decreasing if one's candidate wins. The full results are shown in Figure 4.1.

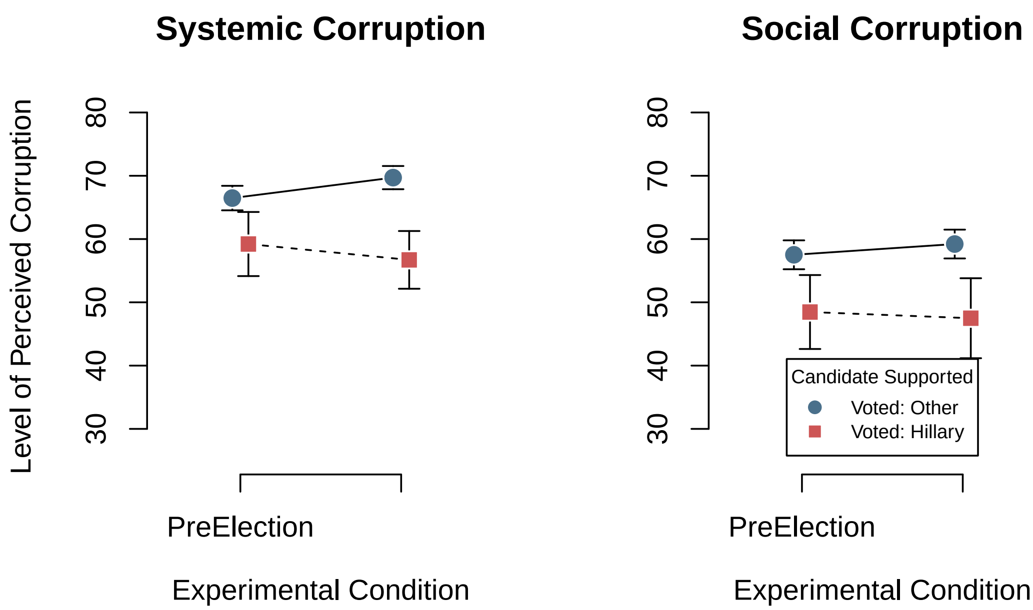


Figure 4.1: Diverging effect of election outcome on perceptions of corruption for supporters of Hillary Clinton compared to those who did not support her. For *Systemic Corruption*, a favorable election outcome was associated with a decrease in perceived corruption and an unfavorable outcome was associated with an increase. This election effect did not appear for perception of corruption for the *Social Corruption* dimension.

Social Corruption

As above, to examine whether there was an effect of election outcome (i.e. whether one's preferred candidate won or or lost) on perceived social corruption, I regressed the measure of *Social Corruption* on $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Hillary Clinton}]\!] \times \underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$. The full results of the regression are reported in Table 4.4 Column (1). Although the omnibus test for the regression was significant, $F(3, 707) = 7.00$, $p < .001$, both contrasts of interest were not significant and election outcomes did not seem have an impact ($p > 0.3$, see Table 4.3 for contrasts of interest). For Clinton Supporters, pre-vs-post election: $M = 48.47$, 95% CI [42.63, 54.31] vs $M = 47.49$, 95% CI [41.23, 53.74]. For non-Clinton Supporters, pre-vs-post election: $M = 57.51$, 95% CI [55.23, 59.79] vs $M = 59.21$, 95% CI [56.95, 61.47].

Table 4.3: Study 1: Contrasts of interest for Systemic Corruption and Social Corruption.

Candidate Supported	Contrast of Interest	Estimated Difference	t-Stat.	p-val.
Systemic Corruption				
Voted: Other	PostElection - PreElection	3.23	2.40	.017
Voted: Hillary	PostElection - PreElection	-2.51	-0.70	.486
Social Corruption				
Voted: Other	PostElection - PreElection	1.70	1.04	.299
Voted: Hillary	PostElection - PreElection	-0.98	-0.22	.822

Note. For both domains of perceived corruption, I compare the levels of perceived corruption in the pre-election sample to the post-election sample for both supporters of Hillary Clinton and for those participants that did not support her in the election. For systemic corruption, for the non-Clinton group, the election produced an increase in the perceived levels of corruption; for the Clinton group, there was a similar decrease in reported levels of corruption, however, this difference was not statistically significant. For social corruption measures, for both groups, the election effects were not statistically significant.

Table 4.4: Regressions showing the effect of the election as a function of the candidate supported on both dimensions of corruption: (i) systemic corruption; (ii) social corruption. There was a significant effect of the election on perceptions of systemic corruption, but no such effect was seen on perceptions of social corruption. The coefficient on the interaction term can be interpreted as capturing 'how much bigger was the before-afters' difference for the Clinton supporters relative to the non-Clinton supporters.

	Two Dimensions of Perceived Corruption	
	Systemic Corruption (0 extendash 100 Scale)	Social Corruption (0 extendash 100 Scale)
	(1)	(2)
(Intercept)	69.707^{***}	59.210^{***}
	(0.948)	(1.151)
I[Pre-Election]	-3.231^{*}	-1.699
	(1.347)	(1.635)
I[Voted: Clinton]	-13.000^{***}	-11.722^{***}
	(2.791)	(3.388)
I[Pre-Election] × I[Voted: Clinton]	5.737	2.679
	(3.836)	(4.656)
Observations	711	711
R^2	0.046	0.029
Adjusted R^2	0.042	0.025
Residual Std. Error (df = 707)	16.806	20.401
F Statistic (df = 3; 707)	11.289^{***}	7.000^{***}

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Std. Errors shown below each estimate.

Concerns regarding number of survey items

Although these above-mentioned results are in accordance with the proposed hypothesis that perceptions of *social corruption* should be more resilient to election-driven effects, there is one major concern that arises from the fact that the estimate of *perceived systemic corruption* is based upon 16 individual survey items whereas the estimate of *perceived social corruption* is based upon four individual survey items. Thus, the lack of an election effect for *social corruption* may simply be because estimation of the latent variable is impaired by measurement limitations in terms of reliability and precision.

Four-Item Systemic Corruption Index To address this concern, I run the same analyses as above, while restricting both *systemic* and *social corruption* to the same four-items. As such, the estimate of *systemic corruption* was based on responses to the two items that address the impact (Question 3) and the severity (Question 4) of corruption on *"the political environment"* and the *business environment* (*Item B* and *Item C* in both questions). The estimate of *social corruption* —exactly as above —was based upon responses to the two items that address the impact and the severity of corruption on *"your personal and family-life"* and the *"values and culture in society"* (*Item A* and *Item D* in both questions; see Appendix A.3 for details). The measures of model fit performed remarkably well, indicating an excellent fit using the restricted model (*CFI*: 0.996; *SRMR*: 0.048; *RMSEA*: 0.039; *RMSEA 90% CI*: 0.029-0.050; *RMSEA p-value* = 0.96), with all loadings > 0.3 and all loadings statistically significant at p < 0.01.

Fitting the multivariate regression models For the four-item index, I see the same pattern of results described for the full SEM model. For *systemic corruption*, I see a main effect of the election, $F(1, 795) = 7.323,\ p < 0.01;\ \beta = 3.57,\ p < 0.01$. Introducing the dummy variable for the individual's preferred candidate leads to a significant difference — model comparison via ANOVA, $F(2, 793) = 6.448,\ p < 0.002$ [*four-item index*]. As with the full model, I see the effect of the election is determined by whether an individual's preferred candidate won the California Democratic Primaries. For non-Hillary supporters, the results of the California primary increased their perceptions of *systemic corruption* ($\beta = 4.06$, $p = 0.004$) [*four-item index*]. And, as before, there was no statistically significant difference between the pre-and-post election measures for Hillary supporters. Thus, even with the restricted-model, the effect of the election is clearly visible on estimates of *systemic corruption.* On the other hand, as reported before, I cannot detect any such election effects for the four-item index of *social corruption.*

In conclusion, the difference in sensitivity to the electoral outcome for these two measures of perceived corruption *does not* arise because one index was estimated using more individual survey items. Instead, the same pattern of results exists —when we use a restrictive and matched four-item based model for *systemic corruption.*

Addressing concerns regarding participant compliance

When interpreting these findings about perceived corruption, one potential concern that had been previously-raised was that these corruption judgments may have been simply the only venue for participants to express dislike of the electoral outcome. Such an argument contends that participants' perceptions of corruption had not *in actuality* been influenced by the election. Instead, participants were upset about the outcome, they did not have any other suitable mechanism for expressing these feelings, so they strategically used the corruption survey as a vehicle for do so.

Three factors speak against this likelihood. First, the exploratory factor analysis showed that the corruption judgments seemed to cluster together in a manner that was theoretically meaningful. Second, the election effect was only found in judgments concerning *systemic corruption* and did not seem to show any measurable effect on judgments concerning *social corruption.* As was shown above, this asymmetry in the effect of the election cannot be explained away as a measurement issue. Instead, this pattern is consistent of the idea that electoral outcomes were impacting these "social perceptions of corruption" differently from "systemic perceptions of corruption." It is also supportive of the idea that the responses on the "systemic perceptions of corruption" were not an artifact resulting from participants seeking a venue to express "dislike of the electoral outcome."

As will be addressed in the discussion section, there are multiple findings in the current political science and development studies literature that follow a similar pattern —and, the most parsimonious explanation for all these "anomalous findings" requires that we account for the cognitive tendencies that render abstract judgments particularly susceptible to systematic biases. Finally, as will be shown in the subsequent sections, the cheating data from both studies provide additional, positive evidence that these two constructs are meaningfully different —not only because they show differing sensitivity to external events like elections —but also because they make reliably different predictions about participant behavior on measures such as willingness to cheating.

4.3 Cheating Behavior

4.3.1 Measuring Cheating

Based on previous work differentiating based upon levels of cheating, I categorized participants into three groups: (i) *non-cheaters*; (ii) *minimal cheaters* who cheat just a little, but "only to a justifiable amount" (Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008); and, (iii) *maximal cheaters:* who brazenly lie as much as necessary to maximize reward (Hilbig & Thielmann, 2017). Thus, for the analyses here, cheating is measured in the following way: (a) a binary measure of cheating (cheated / did not cheat), (b) an ordinal measure of cheating (non-cheaters, minimal-cheaters, maximal cheaters), (c) a continuous measure of cheating (amount of cheating treated as a continuous variable ranging from \$0 to \$1). None of the central results or conclusions reported here are substantively impacted by the choice of scale.

Control Variables

I controlled for gender based on previous work that has reported substantial gender differences in willingness to cheat for monetary gain (Dreber & Johannesson, 2008; Erat, 2013; Erat & Gneezy, 2011; Fosgaard et al., 2013), I also included participant gender as a control variable, since it is also know to influence reported perceptions of corruption (Swamy et al., 2001) and also played a major role in the 2016 election year. No results were fundamentally altered by the inclusion or exclusion of gender or any other demographic controls. Since gender was used as a control variable, the two participants that did not identify with either binary gender had to be excluded in these analyses.

Opportunity to cheat

On the initial demographics form, 568 or 71% of participants reported having 0 dependents; 94 or 12% reported having 1 dependent; 89 or 11% reported having 2 dependents; 29 or 4% reported having 3 dependents; 17 or 2% reported having 4 or more dependents; and, 2 participants did not reply. Participants that reported 4 or more dependents in the demographic form were simply presented a full bonus and not offered an opportunity to cheat. For participants with 0 dependents, the bonus payment system would have looked most "natural" since the bonus started with 1 or more dependents —as such, this subset was of particular interest.

4.3.2 Patterns of Cheating

For the 778 participant who had the opportunity to cheat, the distribution of cheating was as follows: 582 participants (75%) did not cheat and 196 participants (25%) cheated to some degree. Of these 196 participants, 41 participants (21%) cheated minimally by reporting only 1 extra dependent ($0.50 bonus), 15 participants (8%) cheated by reporting 2 extra dependents ($0.75 bonus), and 140 participants (71%) cheated maximally by reporting 3 extra dependents ($1.00 bonus). For the ordinal measure of cheating, I combined the participants who reported 1 or 2 extra participants; the distribution of cheating was as follows: *no cheating* —582 participants; *some cheating* —56 participants; *maximal cheating* —140 participants.

4.3.3 Effect of the Election on Willingness to Cheat

There was no main effect of the election on willingness to cheat, $\chi^2(1,\ N = 780) = 1.63$, $p = .202$, $V = .05$ (Pre-Election: $M = 0.27$ [0.23, 0.32] vs Post-Election: $M = 0.23$ [0.19, 0.28]). Similarly, there was no difference in the willingness to cheat and candidate supported, $\chi^2(1,\ N = 780) = 0.27$, $p = .600$, $V = .02$ (Supported Clinton: $M = 0.28$ [0.19, 0.38] vs Did Not Support Clinton: $M = 0.25$ [0.22, 0.28]).

Binary Measure: Effect of Election Outcome on Cheating

To test whether whether having one's preferred candidate win or lose impacted cheating (i.e. whether there was an effect of election outcome on cheating), I regressed a binary measure of cheating $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Cheated}]\!]$ on $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Hillary Clinton}]\!] \times \underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$, where $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Cheated}]\!]$ is an indicator / dummy variable that equals 1 for participants that cheated and 0 otherwise; $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Hillary Clinton}]\!]$ is an indicator / dummy variable that equals 1 for participants that voted for Hillary Clinton and 0 otherwise; and, $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$ is an indicator / dummy variable that equals 1 for the post-election participants and 0 for the pre-election participants. When I examined the two contrasts of interest, I see that the election outcome appears to decrease cheating among the losers, PostElection - PreElection: $\Delta M = -0.32$, 95% CI $[-0.67,\ 0.02]$, t(∞) = -1.82, $p = .068$ (Odds-Ratio: 0.73), although this contrast is only marginally significant.

On the other hand, the election outcome appears to increase cheating among the winners, PostElection - PreElection: $\Delta M = 0.49$, 95% CI $[-0.44, 1.42]$, t(∞) = 1.04, $p = .300$ (Odds-Ratio: 1.63). If I restrict the analysis to participants with only 0 dependents (i.e. participants that faced the most naturalistic cheating scenario), the same pattern is seen and the results are significant or almost significant at p = 0.05 for both groups. The election outcome appears to decrease cheating among the losers, PostElection - PreElection: $\Delta M = -0.42$, 95% CI $[-0.83, -0.01]$, t(∞) = -2.00, $p = .046$ (Odds-Ratio: 0.66) and appears to increase cheating among the winners, PostElection - PreElection: $\Delta M = 1.17$, 95% CI $[0.00, 2.33]$, t(∞) = 1.95, $p = .051$ (Odds-Ratio: 3.22). The more pronounced effect of election outcome on participants with 0 dependents (i.e. on participants with the most naturalistic cheating test) is also seen in Study 2 and is examined in greater detail at that point.

Continuous Measure: Effect of Election Outcome on Cheating

Rather than focusing solely on cheating as a binary measure, in the current analyses, I examine the effect of election outcome on *amount of cheating* (measured as a continuous variable in dollar amounts) by regressing *Amount Cheated* on $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Voted: Hillary Clinton}]\!] \times \underset{\{0,1\}}{\mathbb{1}} [\![\textit{Post-Election}]\!]$, where $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Voted: Hillary Clinton}]\!]$ and $\underset{\{0,1\}}{\mathbb{1}} [\![\textit{Post-Election}]\!]$ are indicator / dummy variables. Using a continuous measure, I see a much clearer effect of election on cheating amount. When I examined the two contrasts of interest, I see that the election outcome decreases cheating among the losers, PostElection - PreElection: $\Delta M = -0.06$, 95% CI $[-0.12, -0.01]$, $t(776) = -2.14$, $p = .033$. As before, the election outcome also appears to increase cheating among the winners, PostElection - PreElection: $\Delta M = 0.08$, 95% CI $[-0.09, 0.24]$, $t(776) = 0.94$, $p = .350$, however, this difference was not statistically significant.

As before, if I restrict the analysis to participants with only 0 dependents (i.e. participants that faced the most naturalistic cheating scenario), the same pattern is seen and the results are significant at p = 0.05 for both groups. The election outcome appears to decrease cheating among the losers from pre-election $M = 0.26$ ($SD = 0.43$) to post-election $M = 0.18$ ($SD = 0.37$), PostElection - PreElection: $\Delta M = -0.08$, 95% CI $[-0.15, -0.01]$, $t(564) = -2.27$, $p = .024$ and appears to increase cheating among the winners from pre-election $M = 0.14$ ($SD = 0.32$) to post-election $M = 0.35$ ($SD = 0.46$), PostElection - PreElection: $\Delta M = 0.21$, 95% CI [0.01, 0.41], $t(564) = 2.08$, $p = .038$. The more pronounced effect of election outcome on participants with 0 dependents (i.e. on participants with the most naturalistic cheating test) is also seen in Study 2 and is examined in greater detail at that point.

4.3.4 Effects of Perceived Corruption on Willingness to Cheat

Perception of corruption is often considered a significant concern in and of itself because it is widely assumed in the literature that perception of corruption influences people's willingness to engage in corruption themselves (e.g. Čábelková & Hanousek, 2004; DeBacker et al., 2015; Treisman, 2000). As such, I examined how perceived corruption influenced people's willingness to cheat. The following pattern of results was seen: (a) perceived systemic corruption resulted in decreased willingness to cheat; (b) perceived social corruption resulted in increased willingness to cheat. These effects were consistent for both pre-election and post-election samples and introducing the pre-post variable did not result in any significant interaction effects, nor did an ANOVA test reveal a significant effect, F = 1.04, p=0.37. The same pattern of results was found using a restricted four-item or minimal two-item scale for systemic and social corruption. The details of the analyses are described in the sections below.

	Odds of Cheating as a function of Perceived Corruption					
	OR	*p*	*OR*	*p*	*OR*	*p*
(Intercept)	0.83	.573	0.18	<.001	2.46	.014
Systemic Corruption	0.99	.004			0.87	<.001
Social Corruption			1.01	.013	1.12	<.001
Observations	778		778		778	

Table 4.5: Both dimensions of perceived corruption significantly predicted willingness to cheat. Increases in systemic corruption was linked to a decrease in cheating; with the opposite true for social corruption.

Corruption and the Decision to Cheat

I regressed willingness to cheat on the two latent variables *systemic* and *social corruption.* Since all the current results are robust to rescaling, as before, I rescaled the latent variables to range from 0 to 100 to ease interpretation.

As shown in Table 4.5, when regressed individually, each 1-point increase in perceived systemic corruption *decreases* the likelihood of cheating by 1% (OR: 0.99, p = 0.004) whereas each 1-point increase in perceived social corruption *increases* the likelihood of cheating by 1% (OR: 1.01, p = 0.013). A multivariate regression taking both types of corruption into account shows a 13% decrease in willingness to cheat for each 1-point increase in *systemic corruption*, holding all else equal;and a 12% increase in willingness to cheat for each 1-point increase in *social corruption*, holding all else equal. The Analysis of Deviance showed a significant effect for each term sequentially added to the model; $D_{Systemic\ Corruption} = 8.29,\ p < 0.01$; $D_{Social\ Corruption} = 126.3,\ p < 0.01$. The predicted marginal effect of both predictors is plotted in the Figure 4.2.

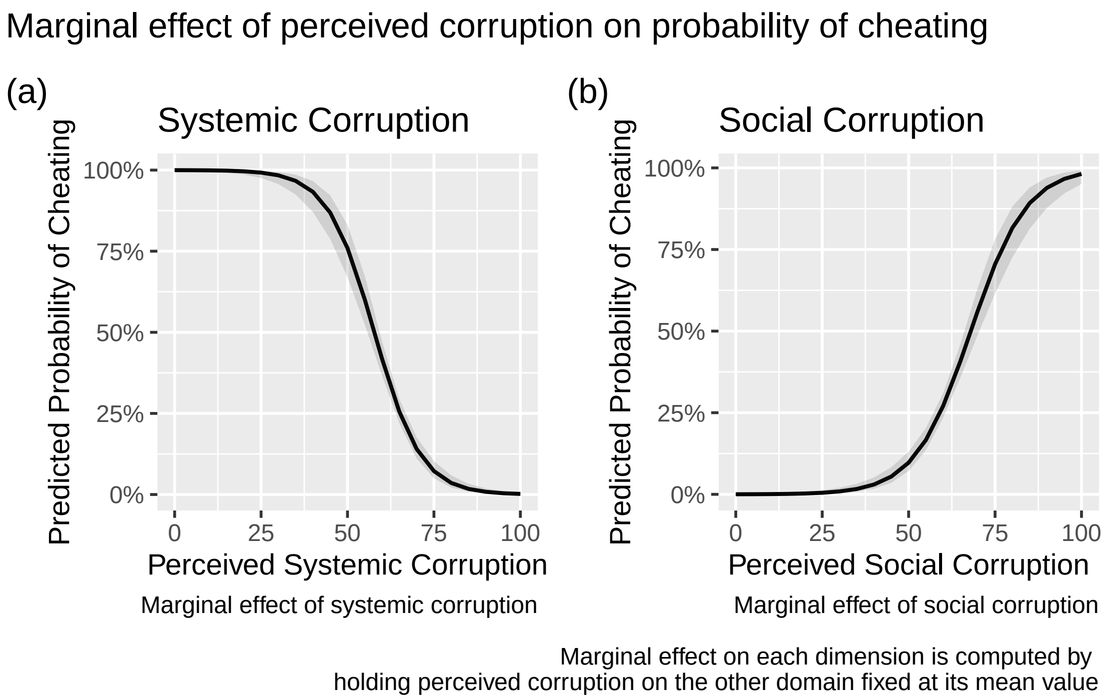


Figure 4.2: The figure shows the predicted marginal effect of both perceived systemic corruption (a) and social corruption (b) on the predicted probability of cheating. Perceived personally-relevant corruption is predicted to systematically increase the likelihood of cheating, whereas perceived systemic corruption is predicted to systematically act in the opposing direction by decreasing the likelihood of cheating.

Corruption and How Much to Cheat

I also examined the relationship between the perceived corruption variables and the amount of cheating by treating it as a continuous variable ranging from \$0–\$1.00. The multivariate regression (cheating regressed on systemic corruption + social corruption) was statistically significant compared to both the univariate regression involving only a measure of systemic corruption, $F(1, 775) = 129.55,\ p < 0.01$ and the univariate regression involving only of social corruption, $F(1, 775) = 134.73,\ p < 0.01$. As shown in Table 4.6, the estimates from the multivariate regression suggest that, holding all else constant, each 10-point increase in systemic corruption is predictive of a \$0.02 decrease ($\beta = -0.02,\ t = -11.61,\ p < 0.01$) in cheating while each 10-point increase in social corruption is associated with a \$0.02 increase in the cheating amount ($\beta = 0.002,\ t = 11.38,\ p < 0.01$).

	Coefficient estimates for cheating amount ($0-$1.00) as a function of perceived corruption					
	β	*p*	*β*	*p*	*β*	*P*
(Intercept)	0.390	**<.001**	0.133	**<.001**	0.57	**<.001**
Perceived Systemic Corruption	-0.003	**.002**			-0.02	**<.001**
Perceived Personal Corruption			0.002	**.024**	0.02	**<.001**
Observations	778		778		778	
F-statistic	$F(1,776) = 10$		$F(1,776) = 5$		$F(2,775) = 70$	

Table 4.6: Effect of perceived corruption on amount of cheating.

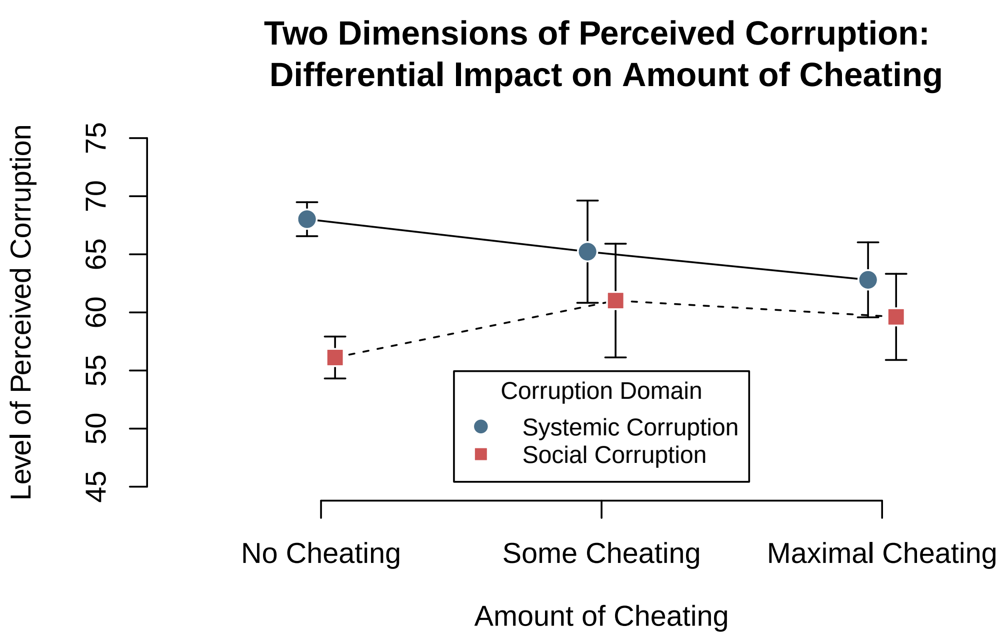


Figure 4.3: There was a significant difference between perceived personal and systemic corruption for non-cheater. However, as cheating increased, there was a steady decline in perceived systemic corruption and an increase in perceived personal/social corruption.

4.3.5 Cheating and Decision Time

I measured the amount of time taken by participants to respond to the cheating question. This afforded me some insight into the possible processes that may be underlying the decision to cheat. Overall, participants who cheated took longer on the cheating question than participants who chose not to cheat ($\beta = 9.56,\ t = 3.66,\ p < 0.001;\ M_{no\ cheating} =$ 26.7 secs, $SD_{no\ cheating} = 23.38$ secs; $M_{cheated} = 36.3$ secs, $SD_{cheated} = 48.7$ secs). When I examine the difference in decision time by distinguishing by amount of cheating, the omnibus test is significant, $F(2, 777) = 7.01,\ p < 0.001$. Comparisons within types reveals that *minimal cheaters* (those who cheated for either $0.50 or $0.75) took approximately 6.7 seconds longer than non-cheaters to make their decision, although this difference was not significant ($M_{minimal} = 33.5$ secs, $SD_{minimal} = 27.4$ secs, $t = 1.51, p = 0.13$), whereas *maximal cheaters* took 10.69 seconds longer than non-cheaters to make their decision, a significantly difference ($M_{maximal} = 37.4$ secs, $SD_{maximal} = 54.9$ secs, $t = 2.98,\ p < 0.001$).

CHAPTER 5

RESULTS: STUDY 2 - U.S. PRESIDENTIAL ELECTION - NOVEMBER 2016

Study 2 used the 2016 Presidential Election as a naturalistic shock that was expected to manipulate perceptions of corruption. The predicted effect of the presidential election would then allow me to replicate the model and findings from Study 1 and further evaluate whether and, if so, how and when do measures of perceived corruption predict behavior, specifically the willingness to comply with societal demands, either in the form of norms of good behavior or in the form of legal mandates. The heightened interest in the 2016 Presidential Election —a closely watched, nationally-relevant election – made it a particularly effective way to deliver a strong manipulation to a broad segment of the population.

5.1 Participants

The 2016 U.S. Presidential election took place on 8th November. A total of 1302 participants were recruited on Amazon's Mechanical Turk to complete the study, consisting of two between-subject samples. As part of the Pre-Election sample, 651 US-based participants were recruited on Election Day (08-Nov-2016) and completed the study prior to the closing of any polls in any U.S. State. The day after the Election, on 09-Nov-2016, after the results of the Presidential Election had been widely disseminated, an additional 651 US-based participants completed the study as part of the post-election sample.

5.1.1 Exclusion Criteria

Identical to Study 1, minimal data exclusion criteria were applied. Only participants that did not complete all sections of the study (for example, failing to provide crucial information like number of dependents in the demographic form, etc.); participants that submitted multiple responses;[1] multiple responses from the same IP address that were initiated within a short time-frame; participants that did not select a binary gender marker (due to very small sample size in "Gender: Non-Binary / Other" category).

5.1.2 Equivalence among Experimental Samples

Confirming that the samples were similar before and after the election, the pre-and-post election samples did not differ in terms of demographic characteristics like gender ($p > 0.2$), level of education ($p > 0.2$), age ($p > 0.9$), or income ($p > 0.6$). Similarly, in terms of political characteristics, there was no difference between the pre-vs-post election sample in terms of support for the major candidates ($p > 0.8$), party affiliation ($p > 0.5$), the degree of support for their preferred candidate ($p > 0.8$), the degree to which they care about the election ($p > 0.3$), care about politics ($p > 0.25$), how much they supported their preferred candidate ($p > 0.9$), or their voting behavior ($p > 0.7$).

5.1.3 Demographic Characteristics

Gender

Of 1302 participants, 615 or 47% were female, 686 or 53% were male, and 1 participants did not identify with either gender – this participants were excluded from any analyses involving gender due to small sample size in this category.

Age and Education

The average age was 36 (SD = 11.8). The average age was 34.7 (SD = 11.5). Of the sample, 37% of participants reported having finished a bachelor's degree, 39% completed

[1] If the multiple responses started concurrently, both responses were excluded, otherwise, the first response was retained and subsequent entries were excluded.

some college or an associate's degree, 11% completed high-school or less, and 13% completed some graduate education (Master's, Professional, or PhD).

Income and Employment

In terms of income, 17% reported earning less than \$10,000; 12% reported earning between \$10-19,000; 15% reported earning between \$20-\$29,000; 14% between \$30-39,000; 9% between \$40-49,000; 10% between \$50-59,000; 8% between \$60-69,000; 5% between \$70-79,000, and the remainder were spread with less than <5% in any of the remaining bins. In terms of employment, 1011 participants (78%) reported being employed, with 744 (57%) employed full-time and 234 (18%) part-time, while 291 (22%) reported having no unemployment. Overall, 456 participants (35%) were employed in the private sector; 296 (23%) in the public sector, 178 (14%) run household or individual businesses, 39 (3%) have other employment, and the remaining are either unemployed or did not reply. In addition, 170 participants (13%) reported being currently enrolled as students.

5.1.4 Political Characteristics

Party Affiliation

Out of the 1302 participants, 549 participants (43%) self-identified as Democrats, 271 (20%) as Republicans, and 379 (29%) as Independents, which is almost identical to the distribution of self-identified party affiliation seen in the Study 1 sample. Crucially, as Table 5.1 shows, the level of party affiliation did not differ across the pre-and-post election sample, $\chi^2(3,\ N = 1302) = 2.19,\ p = 0.5$.

Table 5.1: Study 2 - Pre-vs-Post Election Sample: Distribution of Political Affiliation

	Experimental Group		
	Pre-Election Sample	Post-Election Sample	Total
Political Affiliation			
Democrat	274	275	549
Independent	180	199	379
Republican	141	130	271
Unsure / Third-Party	56	47	103
Total	651	651	1,302

Note. The distribution of participants by self reported political affiliation remained constant for both the pre-election and post-election samples. $\chi^2(3,\ N = 1302) = 2.19,\ p = .534,\ V = .04$

Candidate Choice

From the entire sample of 1302 participants, 51% supported Hillary Clinton, 29% supported Donald Trump, 6% went for Gary Johnson, 4% for Jill Stein, and 10% did not support any candidate. This pattern was equivalent for both pre- and post-election samples, $\chi^2(4,\ N = 1302) = 1.5,\ p = 0.8$. It is especially noteworthy that levels of support for Donald Trump was essential identical before and after the results of the election were declared (p > 0.99) – suggesting that our samples did not exhibit the kind of underreporting of support for Donald Trump that had compromised some polls in 2016.

5.1.5 Concerns about representativeness of mTurk sample

Comparing the Study 2 mTurk sample to national trends

As with Study 1, the online mTurk sample was significantly less conservative than one would expect from a nationally representative sample. For example, in both the pre- and post-election sample, only 29% of participants supported Trump, even though he won 46%

of the national vote share in the election (MIT Election Data and Science Lab, 2018) and in online surveys conducted around the election (Morning Consult, 2016). There was also an over-representation of very liberal voters, resulting in 4% of the sample supporting Jill Stein, the Green Party Candidate who only received 1.5% of the national vote (MIT Election Data and Science Lab, 2018).

Liberal attitudes are over-represented in the mTurk sample

There appears to a broad consensus among Political Science researchers that mTurk samples differ significantly from nationally representative samples like American National Election Survey (ANES) or Cooperative Congressional Election Survey (CCES), both in terms of demographics (Huff & Tingley, 2015), partisanship (Levay et al., 2016) and political ideology (Clifford et al., 2015), with the liberal participants expressing more liberal attitudes that liberals in other nationally-representative samples.

Evidence addressing the reliability and validity of mTurk as a convenience sample for political science research

While there is evidence that mTurk samples diverge from national samples in terms of population characteristics for domains like health-care (Walters et al., 2018) or psychopathy (McCredie & Morey, 2018), this does not appear to be the case for the current domain of study. Recent studies evaluating the validity of mTurk as a convenience sample for political science research found: (a) *use of controls*: most differences in findings between mTurk and representative samples can be reduced or eliminated by controlling for visible, easily measurable sample features (Levay et al., 2016) —i.e. mTurk respondents do not differ fundamentally from population-based respondents; (b) *psychological concomitants of political ideology and partisanship*: although, mTurk participants are more liberal in attitudes and prefer more liberal candidates, in terms of psychological models of political ideology, mTurk samples replicate findings addressing the psychological concomitants of political ideology (Clifford et al., 2015). So, for psychological research on political ideology, there is evidence that mTurk is as reliable and valid as nationally-representative samples; (c) *effect size versus valid inferences*: mTurk samples provide valid inferences into how partisanship or ideology moderate treatment effects – the main concern being the external validity of *average treatment effect size* (Boas et al., 2018). Although, in practice, the difference in effect size appear to be small or negligible (Clifford et al., 2015).

For the purposes of the current study, I was largely interested in comparisons within partisan-groups. Since the current work is not interested in the predictive validity of the estimate average *treatment effect size*, the concerns raised about the ideological or partisan unrepresentativeness do not apply to the current context. As such, given the findings discussed above, it is reasonable to conclude that the unrepresentativeness of the mTurk sample should only impact the current work by reducing its ability to detect effects for groups that are under-sampled.

5.1.6 Political and Electoral Engagement

Self-Reported Caring and Engagement

Prior to completing the study, participants were first asked if they knew that Donald Trump had won the Presidential Elections – 99.5% indicated that they were aware. Then, as with Study 1, I relied on three measure of political and electoral engagement: how much participants care about politics; how strongly they support their preferred candidate; and, how much they care about the 2016 Presidential Election. Responses on all three measures indicated high levels of engagement with the elections, politics, and their candidate. Overall, participants reported moderately high levels of interest in politics (M = 70.1, SD = 25.9), moderately high levels of support for their favored candidate (M = 71.3, SD = 27.9), and high levels of caring about the 2016 Presidential Elections (M = 81.5, SD = 24.2). There was no difference between the pre- and post-election samples on any of these three measures.

Engagement and Candidate Supported

Engagement: Two Main Parties I examined whether there were any systematic differences in engagement for participants based upon the candidate they supported. To test for these differences, I conducted simultaneous tests for multiple comparisons of means (using *Tukey* contrasts) and report the adjusted p-values (adjusted using *single-step method*). The analyses were conducted using the R packages: *multcomp* (Hothorn et al., 2008) and *lsmeans* (Lenth, 2016). For the Hillary and Trump supporters, there were no difference in average levels of *support for their candidate* (p > 0.6) or the average levels of *caring about the 2016 Presidential Elections* (p > 0.5). However, Trump supporters did appear to report a slightly higher level of *caring about politics in general* compared to Hillary supporters ($M_{\textit{Voted: Trump}} = 76.9$ vs $M_{\textit{Voted: Hillary}} = 72.1;\ \Delta = 4.82,\ 95\%\ \text{CI}\ [2.0, 7.6],\ t(813.92) = 3.36,\ p < .001$)

Lower levels of Engagement for Third-Party Supporters However, I did see significantly lower levels of political and electoral engagement for participants who supported third-party candidates (10% of the sample) vs supporters of the two main parties (80% of the sample). For example, supporters of the Libertarian candidate, Gary Johnson, reported lower levels of interest in politics in general (7 points lower than Trump supporters, $t = -2.6, p = 0.056$), lower levels of support for their preferred candidate (12.4 points lower than Trump supporters; $M_{Trump}:\ 74.2;\ M_{Libertarian}:\ 61.9,\ t = -3.62,\ p < 0.003$); and lower levels of caring about the 2016 Presidential Elections (11 points lower than Trump supporters; $M_{Trump}:\ 87.7;\ M_{Libertarian}:\ 76.5,\ t = -4.26,\ p < 0.001$. I saw significantly lower-levels of engagement for Jill Stein supporters as well, although not all contrasts remain significant after correcting for multiple comparisons.

Electoral Engagement and the Efficacy of the Election Manipulation In addition, as one might expect, the 10% of participants who were not supporting any of the candidates also reported much lower interest in politics than any other group (e.g. 36 points lower than Trump supporters; M_{Trump} : 76.9; $M_{No\ Candidate}$: 40.9, $t = -14.9$, $p < 0.001$); and significantly lower concern for the presidential elections (e.g. 40.5 points lower than Trump supporters; M_{Trump} : 87.7; $M_{No\ Candidate}$: 47.1, $t = -19.1$, $p < 0.001$).

Since I rely upon the election as the primary manipulation, participants that have systematically lower levels of engagement with the 2016 Presidential Election are likely to show a diminished response to the manipulation. For this reason, for major findings, I will both results from the entire sample and results that are restricted to the participants who supported either the Republican or the Democratic nominee.

5.2 Perception of Corruption - Effect of the 2016 Presidential Election

5.2.1 Unidimensional Index of Corruption

Aggregating Using Simple Averages

As with Study 1, I constructed a unidimensional index of corruption. First, the response for rating-scale items (Questions 1, 3, and 4) were rescaled to fall between 0 and 100. Then, I used an unweighted average of responses to each of the 20 items in the corruption survey to produce a *perception of corruption score* for each participant. For the analyses here, I excluded the 134 participants that did not support any candidate – first, because it meant I could not reliable classify them with any participant subsample; second, as previously discussed, such participants were outliers in terms of engagement with the presidential election; and, most crucially, I would *not* expect a reliable and predictable effect of the election on participants that do not care about it at all. It should be noted that the inclusion of these participants does not substantively change the significance of any statistical tests; nor does it change any of the conclusions drawn from the pattern of results.

Effect of the Election Results on Perception of Corruption

Replicating the approach from Study 1, I first show that there is a change in perceived corruption after the presidential elections, and, that the direction of the effect depends upon whether one's preferred candidate won or lost the election. To that end, I regressed the measure of *Perceived Corruption* on $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Trump}]\!] \times \underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$, where $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Trump}]\!]$ is an indicator / dummy variable that equals 1 for participants that supported Donald Trump and 0 for those participants that did not and $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$ is an indicator / dummy variable that equals 1 for the post-election participants and 0 for the pre-election participants. The omnibus model test is statistically significant, $F(3, 1164) = 8.77$, $p < .001$. All the estimated coefficients are reported in Column 1 of Table 5.2.

Subsequent examination of the contrasts of interest revealed the following pattern of results: prior to the election, Trump supporters, reported higher levels of corruption by 7-points compared to non-Trump supporters, Trump: $M = 71.2$ ($SD = 16.0$) vs Non-Trump: $M = 64.2$ ($SD = 19.8$), $\Delta = 7.02$ (SE: 1.65), $t(1164) = 4.27$, $p < .001$. After the election, this gap between the two groups vanishes. First, I see an increase in perceived corruption by 5.1 points for non-Trump supporters to $M = 69.2$ ($SD = 18.6$), which was statistically significant, Election Effect$_{\textit{Non-Trump}} = 5.08$, $t(1, 164) = 3.89$, $p < .001$. If I exclude third-party supporters (i.e. limit the comparison to Trump and Clinton supporters), the pre-election gap between the two groups increases to a 9-point difference; and, the election effect strengthens: increasing perception of corruption by 7-points for Clinton supporters, $t(1030) = -4.7$, $p < 0.001$ (see Figure 5.1). For Trump supporters, although the report a slight decrease in perceived corruption as a result of the election, Post-Election: 71.2 vs 69.8, this decrease was not statistically significant, Election Effect$_{\textit{Trump}} = -1.40$, $t(1, 164) =$

-0.73, $p = .465$. This lack of effect on Trump supporters matches the findings from Study 1, where Hillary supporters showed no change in perceived corruption as result of the election.

Demographic Controls I also ran a regression with a full set of demographic controls, including participant gender, education, age, income, student status, employment status, and employment sector, which also showed a significant omnibus test, $F(29, 1135) = 3.29$, $p < .001$. An ANOVA comparing the baseline model with the one that included demographic controls showed a significant difference between the two, as would be expected based upon the prior findings identifying the micro-level determinants of perceived corruption, $F(26, 1135) = 2.63$, $p < .001$.

Crucially, however, the inclusion of these demographic control variables did not change the statistical significance on the coefficients of interest; nor did it significantly alter the estimated effect size or statistical significance for the contrasts of interest. The estimated coefficients for variables of interest are shown in Table 5.2 for both regressions. The coefficient estimates for the demographic variables are not reported, since these variables were not central to our hypotheses. The contrasts of interest were very similar to those reported above. Thus the demographic control variables do not appear to alter the election effect.

Table 5.2: Coefficient estimates from regressing the score of perceived corruption (unidimensional scale) on the election and its interaction with the candidate supported. In the pre-election sample, there was a significant difference between Trump and non-Trump voters ($p < 0.001$). After the election, Clinton supporters reported an increase in perceived corruption ($p < 0.001$). There was no change in the levels of corruption reported by Trump supporters.

	Election Effect on Perceived Corruption (Unidimensional)	
	Baseline Model *(no controls)*	Full Model $\left(\textit{all demographic controls}\right)$
	(1)	(2)
(Intercept)	64.16^{***}	65.83^{***}
	(0.92)	(2.77)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Post-Election}]\!]$	5.08^{***}	4.82^{***}
	(1.30)	(1.30)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Voted: Trump}]\!]$	7.02^{***}	7.19^{***}
	(1.65)	(1.65)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Post-Election}]\!] \times \underset{\{0,1\}}{\mathbb{I}} [\![\textit{Voted: Trump}]\!]$	-6.48^{**}	-6.38^{**}
	(2.32)	(2.31)
Excluded Participants with 'No Candidate'?	Y	Y
Demographic Controls?	N	Y
Observations	1,168	1,165
Adjusted R^2	0.02	0.05
Residual Std. Error	18.42 (df = 1164)	18.11 (df = 1135)
F Statistic	8.77^{***} (df = 3; 1164)	3.29^{***} (df = 29; 1135)
Note:		$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Std. Errors shown below each estimate.

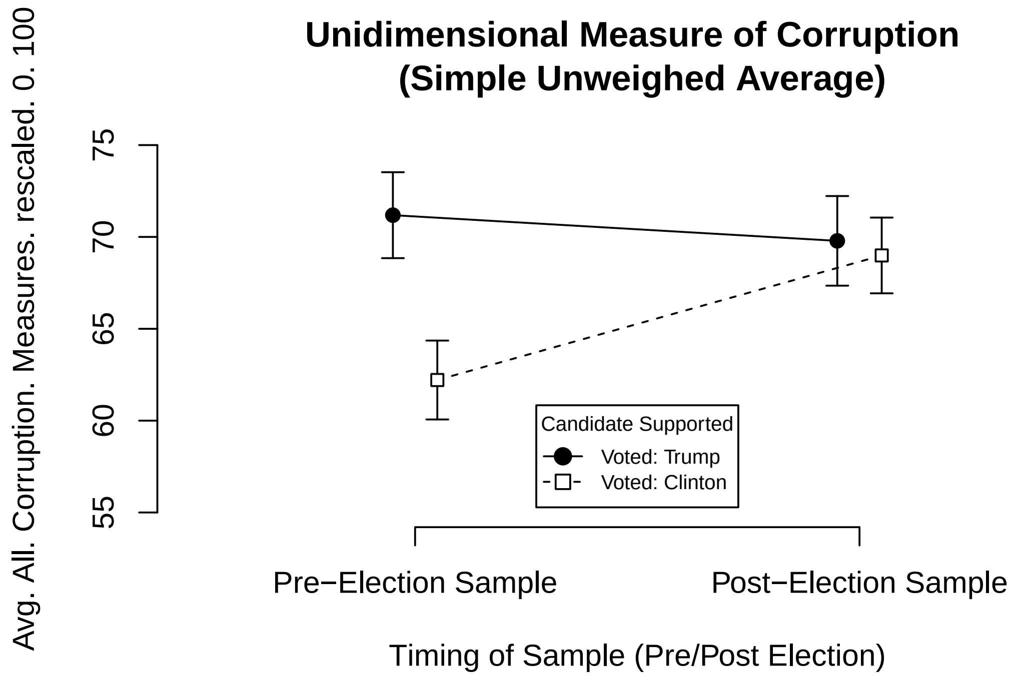


Figure 5.1: Diverging effect of election outcome on perceptions of corruption for supporters of Hillary Clinton compared to those who did not support her. For *Systemic Corruption*, a favorable election outcome was associated with a decrease in perceived corruption and an unfavorable outcome was associated with an increase. This election effect did not appear for perception of corruption for the *Social Corruption* dimension.

5.2.2 Bidimensional Index of Perceived Corruption

Validity and replication of SEM-based model: Using Study 1 estimates on Study 2

In order to validate and test the reliability of the findings from Study 1, I used the SEM model fit and loadings estimated from the Study 1 data to construct the latent measures of perceived corruption in the Study 2 sample. Taking such a stringent approach more than addresses any concerns about potentials for over-fit that have previously been raised as concerns in the application of SEM models (Bauldry, 2015; MacCallum & Austin, 2000; McArdle & Kadlec, 2013). Not only does Study 2 use the identical measurement model and structural model: by using the loadings estimated from the SEM fit on Study 1 data, the replication in Study 2 provides an assessment of the generalizability of findings – especially given the fact that the context for Study 1 (the California primary for the Democratic Party) and the context for Study 2 (the U.S. Presidential Election) are quite different from each other.

Finally, through the sheer vagaries of electoral chance, the winning candidate changed across the elections in Study 1 and Study 2. In Study 1, participants that supported Hillary Clinton won the election, whereas in Study 2, the Clinton supporters lost the election. This reversal in fortunes for Secretary Clinton meant that the same candidate did not won in both elections. So, with the replication of findings in Study 2, we can put aside concerns that the current findings were uniquely driven by endogenous characteristics of Hillary Clinton supporters or any other group characteristics.

Effect of Election on Perceptions of Corruption

After generating the fitted values for the two latent measures of perceived corruption, I test the effect of the election on both social and systemic corruption. As before, I tested the hypothesis that there should be a change in perceived corruption before and after the election, with the direction of the effect being determined by whether one's preferred candidate won or lost the election. For both *Perceived Systemic Corruption* and *Perceived Social Corruption*, I regressed the measure of perceived corruption on $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Clinton}]\!] \times \underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$, where $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Voted: Clinton}]\!]$ is an indicator / dummy variable that equals 1 for participants that supported Secretary Clinton and 0 for those participants that did not and, as before, $\underset{\{0,1\}}{\mathbb{1}}[\![\textit{Post-Election}]\!]$ marks the sample window (pre-post). In both cases, I ran the baseline model as well as the full model with all demographic controls. The details for each index of corruption are discussed in the subsections below.

Table 5.3: The Effect of Unfavorable Election Outcomes on Perceived Systemic Corruption.

	Perceived Systemic Corruption	
	Baseline Model *(no controls)*	Full Model $\left(\begin{array}{c}\textit{all demographic}\\ \textit{controls}\end{array}\right)$
	(1)	(2)
(Intercept)	67.91^{***}	68.44^{***}
	(1.31)	(3.01)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Post-Election}]\!]$	-1.03	-1.13
	(1.85)	(1.85)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Voted: Clinton}]\!]$	-7.70^{***}	-7.78^{***}
	(1.65)	(1.66)
$\underset{\{0,1\}}{\mathbb{I}} [\![\textit{Post-Election}]\!] \times \underset{\{0,1\}}{\mathbb{I}} [\![\textit{Voted: Clinton}]\!]$	7.84^{***}	7.62^{**}
	(2.32)	(2.32)
Trump-Clinton Only?	Y	Y
Demographic Controls?	N	Y
Observations	925	923
Adjusted R^2	0.03	0.07
Residual Std. Error	17.05 (df = 921)	16.75 (df = 893)
F Statistic	11.35^{***} (df = 3; 921)	3.29^{***} (df = 29; 893)
Note:	$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Std. Errors shown below each estimate.	

Table 5.4: Study 2 —Contrasts of Interest: Election effects on perceived corruption

Within Group	Contrast of Interest	$Est.\Delta$	$t(921)$	p
Candidate Supported				
Voted: Trump	PostElection - PreElection	-1.03	-0.56	.578
Voted: Hillary	PostElection - PreElection	6.81	4.83	$< .001$
Experimental Group				
PreElection	Voted: Hillary - Voted: Trump	-7.70	-4.67	$< .001$
PostElection	Voted: Hillary - Voted: Trump	0.14	0.09	.930

Note. The estimates for the contrast of interest showed a significant effect of the election outcome for perceptions of corruption for Hillary Clinton supporters. No effect was seen for Donald Trump supporters, which replicated the pattern seen in Study 1 —where no effect was detected for the winning group (Clinton supporters) after the California primaries.

Systemic Corruption As with the unidimensional model, the omnibus F-test suggested that there was a strong effect of the election outcome on people's perceptions of systemic corruption, and the effect depended upon whether one's preferred candidate had won or lost the election, $F(3, 921) = 11.35$, $p < .001$. The regression coefficient estimates for the variables of interest are included in Table 5.3 for both the baseline model and the full model with all demographic controls. As before, estimates for the demographic controls are not reported here since they were not relevant to the hypotheses under consideration.

Upon examination of the contrast of interest, we see that the pattern of results replicates exactly the pattern seen in Study 1. First, we see that there was no effect of the election outcome on perceived systemic corruption for Trump supporters, recreating the pattern seen for Clinton supporters in Study 1, $t(921) = -0.56,\ p < 0.6$. In the pre-election sample, Trump supporters reported significantly higher levels of corruption with almost an 8-point difference in reported scores across Clinton and Trump supporters, $M_{Trump} = 68$ vs $M_{Clinton} = 60,\ t(921) = 4.67,\ p < 0.001$. The difference between these groups was eliminated in the post-election sample, $p < 0.9$, after Clinton supporters showed an increase in reported levels of systemic corruption of 6.8 points after the election, $t(921) = 4.8,\ p < 0.001$.

The inclusion of demographic control variables neither changed the pattern of result, nor did it significantly impact the coefficient estimates.

5.3 Cheating Behavior

In Study 2, I used the identical demographics-based measure I used in Study 1 to determine each participant's willingness-to-cheat. In this section, I first show the main effect of the election on cheating behavior. As discussed in Study 1, I found that the main effect of the election on cheating was only statistically significant when I restricted the sample to participants with zero dependents. This was because participants who reported 0 dependents in the initial demographics form saw the most naturalistic version of the cheating task, where a bonus was provided to anyone who had 1 or more dependents. Participants with 1 or more dependents, on the other hand, saw a version that looked more like a measure that was obviously designed to measure participant honesty. As discussed before, it is proposed that this change in the "transparency" of the cheating measure may drive a shift in strategy among participants.

Having established the main effect of the election outcomes on cheating, I then replicate the findings from Study 1 and show that social and political corruption impact willingness-to-cheat in opposing directions. I then provide initial evidence that the effect of the election on willingness-to-cheat derives in part from the election-driven changes in perceived corruption.

5.3.1 Measures of Cheating

As with Study 1, I relied on the same three measures of cheating: (a) a binary measure of cheating (cheated / did not cheat), (b) an ordinal measure of cheating (non-cheaters, minimal-cheaters, maximal cheaters), and (c) a continuous measure of cheating (amount of cheating treated as a continuous variable ranging from \$0 to \$1).

Overall Patterns of Cheating in Study 2

For the 1267 participant who had the opportunity to cheat, the distribution of cheating was as follows: 927 participants (73%) did not cheat and 340 participants (27%) cheated to some degree. Of these 340 participants, 108 participants (31%) cheated minimally by reporting only 1 extra dependent ($0.50 bonus), 34 participants (10%) cheated by reporting 2 extra dependents ($0.75 bonus), and 198 participants (58%) cheated maximally by reporting 3 extra dependents ($1.00 bonus). For the ordinal measure of cheating, I combined the participants who reported only 1 or 2 extra participants to get the following three class distribution: 927 participants (73%) —*honest*, 142 participants (11%) —*minimal cheaters*, and 198 participants (15%) – *maximal cheaters.*

5.3.2 Differences in Transparency: Opportunity to Cheat vs Test of Honesty

On the initial demographics form, 809 or 62% of participants reported having 0 dependents; 208 or 16% reported having 1 dependent; 169 or 13% reported having 2 dependents; 82 or 6% reported having 3 dependents; 34 or 3% reported having 4 or more dependents. As with Study 1, participants that reported 4 or more dependents in the demographic form were simply presented the full bonus and not offered an opportunity to cheat.

Number of Dependents and Plausibility of the Cover Story

As may be recalled, the cheating measure in the current study was designed as a stylized abstraction of a means-based benefits program. The headline prompt asked participants – "Do you have any dependents (e.g. children or elderly parents that you financially support)" and then proceeded to tell them that they were receiving a bonus as a token of appreciation for completing the study with the amount of bonus being on a sliding scale based upon number of dependents. However, the adaptive feature ensured that the first threshold to receive a bonus was always one greater than the number of dependents reported in the demographic form. This meant that the plausibility / face-validity of the cover story changed as a function of the number of dependents.

Plausibility as an ideal disguise for measures of cheating For participants with 0 dependents, the bonus screens appear the most natural. To get a bonus, participants simply had to report having 1 or more dependents. If one were running a good-faith, means-based bonusing system in earnest, it would be structured as such. So, for these participants, the cheating opportunity is most likely to appear as though I were providing additional bonuses to anyone who has dependents on the principal of need-based assistance. Many of the participant comments (both positive and negative) suggest that this was at least true for some.

Implausibility renders a cheating opportunity into an honesty test On the other end of the scale, for participants who reported 3 dependents, the cheating opportunity would appear quite implausible. After all, for these participants, to get the smallest bonus they would have to report having 4 dependents, which seems unusual enough to draw their attention back to the fact they had completed a question about dependents earlier in the demographics form and that they had reported having 3 dependents. Once their attention has been drawn to this fact, for these participants, the bonus question would most immediately and transparently appear to be a "psychology study" style measure of honesty. As a result, they would know that they are participating in an honesty test and are likely to respond based upon that awareness.

Plausibility and Patterns of Cheating

This intuition is also borne out in the cheating data, where we see a strong spike in cheating for participants who reported 3 or more dependents. Treating *cheating* as a continuous measure, we see that the number of dependents had a significant impact, $F(3, 1264) = 3.85$, $p = .009$. As number of dependents reported increased —and, as result, the transparency of the cheating measure increased —so did the amount of cheating. The estimated regression coefficients are included in Table 5.5. Closer examination of the main effect reveals that it arises due to a combination of two factors: (a) as number of dependents increase, there is an increase in the proportion of cheaters; and (b) conditional upon cheating, as number of dependents increase, there is a decrease in the degree of cheating (measured in terms of dollar amounts).

Transparency and Binary Measures: To Cheat or Not to Cheat When Someone Is Looking The lowest rate of cheating occurs among participants who saw the 0 dependents version (22.6% cheated). Fewer people cheat when the task most plausibly resembles a needs-based benefits scheme (i.e. it plausibly resembles a good-faith attempt to provide additional compensation to people who support dependents). A greater proportion cheat in the case of 1 or 2 dependents (32% cheated in each), where – although still possibly ambiguous – the cheating opportunity increasingly starts to resemble an honesty test. And, the largest percentage cheat in the 3 dependents condition (42% cheated), where the possibility of getting a bonus requires participants to report 4, 5, or 6 dependents for a bonus of \$0.50; \$0.75; or \$1.00 bonus respectively —and, where the cheating measure most clearly resembles an honesty test. The Chi-Square "goodness of fit" test of the frequency distribution for Cheated (Yes/No) x Number of Dependents (0,1,2,3) was highly significant, $\chi^2(3) = 22.5$, $p < 0.001$; *Cramer's* $V = 0.134$, allowing me to reject the null-hypothesis that the distribution is uniform across different number of dependents.

Table 5.5: Cheating as a function of the number of dependents reported by the participant in the demographics form. Column 1 shows the Odds-Ratio for participants based upon number of dependents. Column 2 shows the average amount of cheating across all participants based upon number of dependents, whereas Column 3 shows estimates for amount of cheating conditional on cheating i.e. limiting to participants that cheated. The difference in pattern of results across Column 2 and Column 3 results from the fact that the odds of cheating increases significantly as the number of dependents increases (as shown in Column 1), which initially gives the impression that cheating amount increases as a function of number of dependents (Column 2), but, the amount of cheating —conditional upon having cheated —actually decreases as the number of dependents increases (as shown in Column 3).

	Odds of and Amount of Cheating		
	All Participants		Conditional on Cheating
	Odds Ratio Cheated	**Amt. Cheat (USD)**	**Amt. Cheat (USD)**
	logistic	*OLS*	*OLS*
	(1)	(2)	(3)
Number of Dependents Reported			
0 dependents (Reference)	0.29^{***}	0.20^{***}	0.88^{***}
	(0.02)	(0.01)	(0.02)
1 dependent	1.65^{**}	0.04	-0.14^{***}
	(0.28)	(0.03)	(0.03)
2 dependents	1.64^{**}	0.03	-0.18^{***}
	(0.30)	(0.03)	(0.03)
3 dependents	2.41^{***}	0.14^{**}	-0.05
	(0.58)	(0.04)	(0.04)
Observations	1,268	1,268	341

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Std. Errors shown below each estimate.

Transparency of Cheating Measure: How Much Does One Cheat on an Honesty Task? In the section above, I established evidence of substantive differences in people's willingness to engage in cheating as a function of the number of dependents. Then, conditional upon cheating, we see a divergence in the amount of cheating as a function of the number of dependents, $F[3, 337] = 12.81$, $p < .001$. Conditional upon cheating, participants in the 0 dependents condition cheat by the largest amount ($M_{\textit{0 dependents}}$ = \$0.88, 95% CI [0.85, 0.91]) – they cheated \$0.14 more that participants who saw the 1 dependents version, $t(337) = 4.39$, $p < .001$ ($M_{\textit{1 dependent}}$ = \$0.74, 95% CI [0.69, 0.79]); and they cheated \$0.18 more than participants who saw the 2 dependents version, $t(337) = 5.33$, $p < .001$ ($M_{\textit{2 dependents}}$ = \$0.70, 95% CI [0.64, 0.76]); but only cheated \$0.05 more than participants that saw the 3 dependents version, which was not statistically significant —possibly due to small sample size, $t(337) = 1.34$, $p = .182$ ($M_{\textit{3 dependents}}$ = \$0.82, 95% CI [0.75, 0.90]). The estimated regression coefficients are included in Column 3 of Table 5.5. It is worth noting at this point that the pattern of results described here matches the general pattern seen in Study 1 as well – for example, for both studies regressing Cheating (in dollar amounts) on Number of Dependents (as a continuous variable) estimates an average \$0.05 decrease in cheating for each additional dependent, *Study 1:* $\beta = -0.05$, 95% CI [−0.08, −0.02], $t[195] = -2.89$, $p = .004$; *Study 2:* $\beta = -0.05$, 95% CI [−0.07, −0.02], $t[339] = -4.04$, $p < .001$.

The goal of this analysis *is not* to present a specific theory on the relationship between transparency of an honesty measure and people's willingness to cheat – nor do I argue that the empirical relationship between these concepts will be consistent in magnitude or direction across contexts. Instead, I propose the more modest claim that the inputs and evaluation processes that people use when deciding whether to cheat on an "honesty test" are likely to be quite different from the inputs, evaluation schemas and cognitive processes deployed when deciding whether to take advantage of an opportunity to cheat. In the current study, I see evidence of this possibility in the fact that main effect of the election outcomes on cheating measures are divergent for participants who saw the 0-dependents version of the cheating compared to the 3-dependents version.

Given the pattern of findings covered here, it seems both justified and reasonable to use the *demographics – number of dependents variable* to more accurately detect and estimate the effect of the election outcomes on cheating. To begin, I simply restrict the sample to participants who saw the most naturalistic version of the cheating measure (i.e. participants who reported 0 dependents).

5.3.3 Effect of Election Outcomes on Willingness-to-cheat

As a baseline, I first establish that there are no differences in the frequency of cheating for the pre-election vs post-election samples ($p < 0.9$); nor is there evidence of an overall difference in the frequency of cheating between Trump and Clinton supporters ($p < 0.3$).

Main Effect of Election on Cheating

When I test for an effect of election outcome on cheating – the contrasts of interest are not significant when all participants are excluded. However, as we saw in Study 1, if I restrict the sample to participants with 0 dependents, i.e. if I restrict the sample to participants who saw the most naturalist version of the cheating measure, there is a clear and statistically significant effect of election outcomes on willingness to cheat – with the direction of the effect determined by whether the electoral outcome was in the participant's favor.

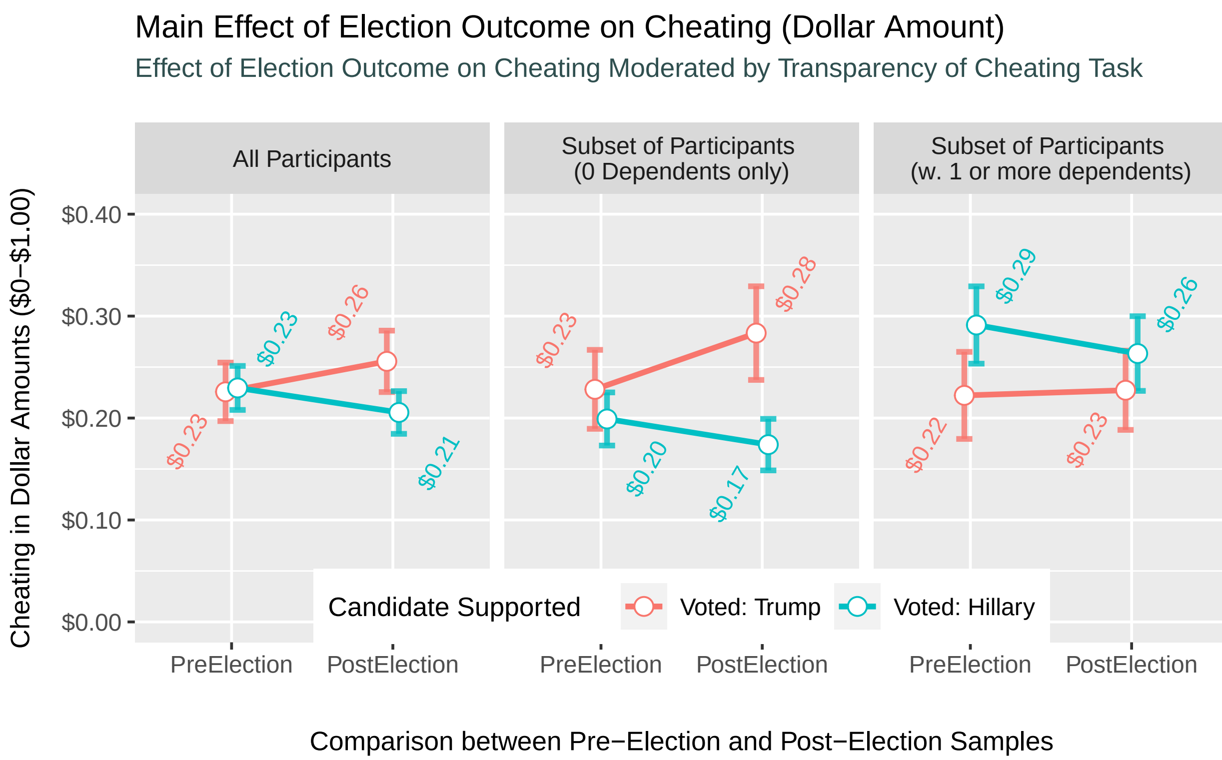


Figure 5.2: Effect of the election on cheating behavior was determined by the favorability of the outcome —however, this election effect was largely restricted to the 0-dependents case, where the cheating opportunity appears more naturalistic, in comparison, participants with one or more dependents saw a version that looked more obviously designed to measure honesty, which I propose results in a different strategy for determining willingness-to-cheat.

As can be seen in Figure 5.2 above, once I restrict the sample to participants with 0 dependents, there is a clear and visible effect of the election. For Trump supporters, the election increases the mean marginal cheating amount by \$0.05 (Trump: $M_{Pre\text{-}Election}$ = \$0.23 vs $M_{Post\text{-}Election}$ = \$0.28). On the other hand, for Hillary supporters, the unfavorable election outcome results in an estimated \$0.03 decrease in cheating (Hillary: $M_{Pre\text{-}Election}$ = \$0.20 vs $M_{Post\text{-}Election}$ = \$0.17). As a result, the difference in estimated mean cheating for the two groups jumps from \$0.05 before the election, t(617) = 1.08, p = 0.29, to a \$0.11 difference in estimated mean cheating for the two groups after the election, t(617) = 2.12, p = 0.03.

5.3.4 Effects of Perceived Corruption on Willingness to Cheat

The pattern of findings linking perceived corruption to willingness-to-cheat in Study 1 were unexpected —specifically, I had not predicted that the two dimensions of perceived corruption would have different relations to cheating. As such, I relied upon the Study 2 sample to test the pattern of results through a direct replication. In the sections below, first, I examined how perceived corruption relates to people's willingness to cheat —measured as a binary choice: cheat or not. Then, I examined how perceived corruption relates to the degree of cheating —measured as a continuous variable in dollar amounts. Remarkably, the pattern of results in Study 2 exactly replicated the findings from Study 1. To foreground the findings, for both Study 1 and Study 2, the bi-dimensional index of perceived corruption predicted both the propensity to engage in cheating and the amount of cheating in dollar terms. The results are discussed in greater detail below.

Corruption and the Decision to Cheat

To examine the relationship between perceived corruption and willingness-to-cheat, I regressed a binary measure of cheating (cheated / honest) on the two latent variables *perceived systemic corruption* and *perceived social corruption*. Since all the results were robust to rescaling, as before, the latent variables were rescaled to range from 0 to 100 in order to ease interpretation. The results of the logistic regressions are shown in Table 5.6. Overall, the following pattern of results was seen: (a) increases in perceived *systemic corruption* was associated with a decrease in willingness-to-cheat; whereas, (b) increases in perceived *social corruption* had the opposite effect and was associated with an increase in willingness-to-cheat. These effects were consistent for both pre-election and post-election samples and introducing the pre-post variable did not result in any significant interaction effects, nor did an ANOVA test reveal a significant effect of including the pre-post variable, F = 1.04, p=0.37.

Table 5.6: Both dimensions of perceived corruption significantly predicted cheating. In Column 1, we see cheating as a function of *systemic corruption*, where each unit increase in *systemic corruption* was associated with a 1% *decrease* in the odds of cheating. Column 2 does the same for *social corruption*. However, for each unit increase in *social corruption*, the model predict a 1% *increase* in the odds of cheating. Column 3 shows cheating as a function of both dimensions. Here, I find, holding all else equal, a 1-pt increase in *social corruption* predicted an 11% increase in the odds of cheating, whereas, a 1-pt increase in *systemic corruption* predicted an 11% *decrease* instead. Column 4 shows Model 3 with controls variables. The inclusion of controls did not meaningfully alter the results. These findings precisely replicated the patterns seen in Study 1.

	Odds of Cheating			
	Univariate Model		**Basic Multivariate**	**Full Model**
	(1)	(2)	(3)	(4)
	Systemic Corruption	Social Corruption	Systemic + Social *(no controls)*	Systemic + Social *(full controls)*
(Intercept)	.60^{+}	.15***	.85	.22
	(.10)	(.05)	(.12)	(13.10)
Systemic Corruption	.99^{+}		.89***	.92***
	(0.00)		(.01)	(.01)
Social Corruption		1.01***	1.11***	1.09***
		(0.00)	(.01)	(.01)
Clinton or Trump Supporters Only?	*N*	*N*	*N*	*Y*
Controls Incl.'?	*N*	*N*	*N*	*Y*
Observations	1,267	1,267	1,267	1,002

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Model 4 —Controls: *Demographic*: gender; education; number of dependents; age; income; student; employment status; if employed: sector of employment; if employed: part-time or full-time. *Political Affiliation*: supported Clinton or Trump. *Sample*: pre-or-post election; the interaction between candidate supported and sample period

When regressed individually, each 1-point increase in perceived systemic corruption *decreased* the odds of cheating by 1% (Model 1: $OR_{systemic\ corruption} = 0.99$, $p = 0.06$). Conversely, each 1-point increase in perceived social corruption *increased* the odds of cheating by 1% (Model 2: $OR_{social\ corruption} = 1.01$, $p < 0.0001$). In Model 3, which takes both dimensions of perceived corruption into account, this pattern of findings seems to strengthen. Based upon this multivariate model, one would predict that: (i) holding the level of *social corruption* fixed, each 1-point increase in perceived *systemic corruption* would result in an 11% decrease in the odds of cheating (Model 3: $OR_{systemic\ corruption} = 0.89$, $p < 0.0001$); and, (ii) holding the level of *systemic corruption* fixed, each 1-point increase in perceived *social corruption* would result in an 11% increase in the odds of cheating (Model 3: $OR_{social\ corruption} = 1.11$, $p < 0.0001$).

In Model 4, I run the same multivariate regression from Model 3 —however, I now included all demographic, political affiliation, and sample-identifying variables as controls (specifically, *demographic:* gender, education level, number of dependents, age, income, student status, employment status, if employed: sector of employment —private vs public vs business owner, if employed: part-time or full-time; *political affiliation:* whether the participant supported Hillary Clinton or Donald Trump, *sample-identifier:* whether they participated as part of the pre-election or the post-election sample; and, finally, the interaction term between candidate supported and election sample). As the coefficients shown in Column 4 reveal, the inclusion of these variables as controls did little to alter the relationship between the two dimensions of perceived corruption and people's willingness to cheat, with a one-point increase in perceived *systemic corruption* predicted to result in an 8% decrease in the odds of cheating (Model 4: $OR_{systemic\ corruption} = 0.92$, $p < 0.0001$); and, conversely, a one-point increase in perceived *social corruption* predicted to result in a 9% increase in the odds of cheating (Model 4: $OR_{social\ corruption} = 1.09$, $p < 0.0001$). The robustness of this pattern to the inclusion of control variables is reassuring —suggesting that the two dimensions of perceived corruption have a reliable relationship with people's willingness to cheat and the effect size remains largely unchanged by individual-level factors like demographics.

In Columns 1 and 2 of Table 5.6, I present the results of a simple logistic regression evaluating the proposition that the likelihood of cheating is related to each dimension of corruption independently: *perceived systemic corruption* (reported in Column 1) and to *perceived social corruption* (reported in Column 2). Finally, in Column 3 of Table 5.6, I report the results from a multivariate regression predicting the relationship between likelihood of cheating as a function of both dimensions of perceived corruption: perceived systemic corruption and perceived social corruption. An Analysis of Deviance using a Chi-Square Likelihood Ratio Test showed a significant effect for each term added sequentially to the model: $D_{Systemic\ Corruption} = 10.40$; $D_{Social\ Corruption} = 148.54$, $p < 0.01$. Crucially, adding the sample identifier (Pre-Election vs Post-Election) was not statistically significant, $D_{Exp\ Group} = 0.76$, $p > 0.8$, which suggests that the patterns reported below are consistent and stable across the pre-election and post-election samples.

Overall, the results from multivariate model match the findings we saw in Study 1 (Democratic Party Primaries —California) —where, holding all else constant, (i) an increase in the levels of perceived *systemic corruption* was associated with a *decrease* in the odds of cheating and (ii) an increase in *social corruption* was associated in an *increase* in the odds of cheating. Given the initially surprising nature of this relationship, it is reassuring to see these results replicate across two different samples.

Corruption and How Much to Cheat

As with Study 1, we also examined the relationship between the perceived corruption variables and the amount of cheating by treating the cheating measure as a continuous variable ranging from $0-$1.00. As before, I ran two simple, univariate regressions predicting cheating as a function of perceived social corruption and perceived systemic corruption individually.

Then, I ran a multivariate regression predicting cheating as a function of both dimensions of perceived corruption. The results of both univariate regressions were significant: the univariate regression involving only *systemic corruption*, $F(1, 1265) = 5.31$, $p = .021$; the univariate regression involving only *social corruption*, $F(1, 1265) = 11.90$, $p = .001$; as was the multivariate regression involving both dimensions of corruption, $F(2, 1264) = 88.33$, $p < .001$. I also ran a full model that included all demographic variables as well as variables indicating the participant's candidate and party preferences, which was also significant, $F(35, 966) = 6.70$, $p < .001$. The coefficient estimates for variables of interest from all these regression are shown in Table 5.7. The inclusion of a full battery of demographic and political control variables did not significantly alter the estimated coefficients on the two corruption measures and both remained statistically significant. The control variables included were the same as the ones described in the previous section (see Table 5.6) The estimated regression coefficients for the control variables are not reported in the table, since they were not relevant to the primary hypotheses under consideration. However, it is worth noting that the coefficients on the election variables were no longer statistically significant once the systemic and social corruption variables were included as part of the regression.

Table 5.7: Perceived corruption and cheating in Study 2: As the regression coefficients below show, both dimensions of perceived corruption significantly predicted cheating. As with the logistic regression, increases in systemic corruption was linked to a decrease in amount of cheating; with the opposite true for social corruption. These findings exactly replicated the patterns seen in Study 1.

	Amount Cheated in USD			
	Univariate Models		**Multivariate Models**	
	Systemic Corruption	Social Corruption	Systemic + Social Corruption *(no controls)*	Systemic + Social Corruption *(full controls)*
	(1)	(2)	(3)	(4)
(Intercept)	.323***	.091*	.405***	.525***
	(.046)	(.039)	(.044)	(.086)
Systemic Corruption	−.001*		−.017***	−.016***
	(.001)		(.001)	(.001)
Social Corruption		.002***	.015***	.016***
		(.001)	(.001)	(.001)
$\mathbb{I}[\![$*Voted: Clinton*$]\!]$				−.045
				(.034)
$\mathbb{I}[\![$*Post-Election*$]\!]$				−.013
				(.028)
$\mathbb{I}[\![$*Post-Election*$]\!]$ × $\mathbb{I}[\![$*Voted: Clinton*$]\!]$				.055
				(.048)
Clinton / Trump Only?	N	N	N	Y
Controls Incl.'?	N	N	N	Y
Observations	1,267	1,267	1,267	1,002
R^2	.004	.009	.123	.195
Adjusted R^2	.003	.009	.121	.166
Residual Std. Error	.380	.379	.357	.352
F Statistic	5.310*	11.896***	88.334***	6.696***

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Note. Model 4 —Controls: See note in Table 5.6 for details.

As shown in Table 5.7, the estimates from the multivariate regression suggest that, holding all else constant, each 1-point increase in systemic corruption is associated with a \$0.02 *decrease* in cheating; $b = -0.02$, 95% CI $[-0.02, -0.01]$, $t(1264) = -12.78$, $p < .001$. On the other hand, holding all else constant, each 1-point increase in social corruption is associated with a \$0.02 *increase* in the cheating amount; $b = 0.02$, 95% CI $[0.01, 0.02]$, $t(1264) = 13.06$, $p < .001$. The findings here confirm the consistent pattern seen across all the analyses in the current work —namely, that increases in perceived systemic corruption and increases in perceived social corruption appear to have equal but opposite effects on people's ethical choices —both in terms of their willingness to engage in cheating and also in terms of the amount of cheating that they are willing to engage in. Finally, it is worth noting —like with previous examples —although demographic variables like gender, age, education and income are predictive of cheating behavior in previous work and also in our sample —the inclusion of these and other demographic controls and the inclusion of candidate preference, party alignment and election outcome variables do not appear to change the size or significance of the coefficients on either of the two corruption measures. Thus, we have reason to believe that the relationship between perceived corruption and cheating behavior may be both robust and direct.

CHAPTER 6

DISCUSSION

The current work examined the effects of two election in the 2016 Presidential cycle —the California Democratic Primary in 2016 and the November General Election —on perceptions of corruption and willingness-to-cheat. Perceptions of corruption were measured using an adapted version of Transparency International's Global Corruption Barometer (as detailed in Chapter 3 with exact questions reproduced in Appendix A.3). Willingness-to-cheat was measured by providing participants the opportunity to receive a bonus based upon the number of dependents they self-reported – this measure was designed to serve as a stylized abstraction of a means-based welfare program (as detailed in Section 2.3.4). For each election, we collected two data samples, (i) a pre-election sample: one the day before the election (6 June for the California Democratic Primary; and Nov 8 for General Election); and, (b) post-election sample collected right after the election results had already been confirmed and widely publicized (8 June in Study 1; and 9 November in Study 2). Both studies used US-based mTurk participants for both the pre-election and post-election samples.

To analyze the data, we relied upon a structural equation modeling (SEM) approach to construct both the measurement model (i.e. a model for aggregating individual survey items into first-order latent constructs) and to specify the relationship between hypothesized latent variables (both to construct higher-order latent variables and to specify the relationship between these latent constructs and exogenous variables of interest). Both measures of goodness of fit and badness of fit were used to evaluate and confirm that the specified model was well-suited to capture the observed patterns in the data. The model-driven estimates of the hypothesized latent variables were then used to test the central hypotheses. The results from Study 1 revealed multiple distinct results, each of which was replicated in Study 2 (reported in Chapter 4 and Chapter 5 respectively). Moreover, in order to avoid concerns about over-fit that can sometimes accompany a SEM-based approach, I used the model structure and fitted loadings from Study 1 to generate the latent variables in Study 2. This allowed me to use Study 2 to validate the SEM-based approach and helped to ensure the generalizability and robustness of the model presented here.

6.1 Overview of the Main Findings

6.1.1 Multi-dimensionality of Perceived Corruption

First, evidence from both exploratory factor analysis (EFA) as well as from confirmatory structural models supported the idea that perceived corruption may be profitably construed as a multi-dimensional construct; (1.a) *measures of fit*: all measures of fit supported the idea that a multi-dimensional construct was better able to capture the participant responses; (1.b) *predictive value of multi-dimensional approach*: since measures of fit do not constitute confirmatory evidence of a multi-dimensional construct, we also relied upon subsequent findings which showed that this multi-dimensional construct was better able to capture relationships between perceived corruption and other measures, such as willingness to cheat; (1.c) *unsupervised classification*: item clustering and individual loadings on the factors seemed to suggest that participant were compliant with the demands of the survey and their responses were sensitive to the particularities of different questions on the survey. In fact, the use of unsupervised hierarchical clustering algorithms produced classification schemas that were almost fully congruent with the constructs found in theoretical work on political trust and legitimacy. This astounding congruence between theory and data-driven classification shows the incredible degree of structure and semantic content that exists in the survey data. This pattern should allay concerns that participants were simply using the corruption survey to express an unrelated political opinion. In fact, it should serve to reassure me that participants in the study were capable and willing to make fine-grained and sophisticated distinction in their political judgments – at least, in the aggregate.

6.1.2 Elections as Exogenous Shocks

Second, comparisons of pre-and-post election judgments revealed that electoral outcomes can behave like exogenous shocks that significantly alter people's perception of corruption; (2.a) subsequent analyses suggests that the size and direction of electoral shocks depends upon both the candidate supported and the degree of support for that candidate; (2.b) these electoral shocks do not impact all dimensions of perceived corruption equally – perceptions of systemic corruption are more malleable than perceptions of social corruption.

6.1.3 Election Outcomes Influence People's Willingness to Cheat

Third, electoral outcomes also appeared to influence people's willingness to cheat on a stylized version of a means-based welfare program; (3.a) these effects appear to be largely driven by changes in perceptions of corruption that resulted from either the victory or the loss of one's preferred candidate; (3.b) these results were robust to the inclusion of gender as a control variable, which – in accordance with previous findings – also significantly impacted cheating behavior; (3.c) the effect of the election on people's willingness to cheat was sensitive to the transparency of the cheating measure used in this study – if the cheating measure appeared as an obvious "test of honesty" rather than an unexpected "opportunity to benefit from cheating," it seems to modify the processes underlying people's decision to cheat and, in doing so, undermines the link between electoral loss/victory and wiliness to cheat for small monetary gains.

6.1.4 Measures of Perceived Corruption Systematically Predict Willingness to Cheat

Fourth, the two hypothesized dimensions of perceived corruption have systematic and opposing effects on people's willingness to cheat as well as the amount of cheating; (4.a) the diverging effects of these two dimensions of perceived corruption adds support to the argument that perceived corruption is best not understood as a unidimensional construct; (4.b) increases in perceived systemic corruption was associated with decreased willingness to cheat, whereas increases in perceived social corruption was associated with decreased willingness to cheat ; (4.c) the effect size of systemic and social corruption appeared to be about equal for each unit change, such that the effects could be feasibly modeled as arising from the difference between the two perceptions.

6.1.5 Decisions to Cheat Are Effortful and Sensitive to Context

Fifth, increased willingness to cheat was associated with increased decision time in both studies, suggesting that cheating was not driven by impulsive behavior, but instead it was driven by effortful justification or rationalization. While the data is consistent with such a hypothesis, I lack any direct measures of the decision-making process that could provide supporting evidence. However, I did find that – people's decision to engage in dishonest behavior was quite sensitive to change in the plausibility of the cover story used to disguise or cheating measure. One implication of this finding is that people may have multiple and diverging approaches to the decision to cheat – thus, any attempt to model the relationship between willingness to cheat and political judgments must be sensitive to the schema used by people in a given scenario.

6.1.6 The Links Between Perceptions of Corruption and Willingness to Cheat are Robust

In Study 2, we had the opportunity to test the robustness and replicability of both the underlying structural model as well as the pattern of results described above. Since Study 2 employed the same exact structure, but relied upon the November 8 U.S. Presidential election as the exogenous shock, it provided a significantly novel context to test the patterns found in

Study 1. In addition, since Hillary Clinton won the June 7th California Democratic Primary and lost the November 8th US Presidential elections, we see different populations receive a positive / negative shock across the two studies. The replication of the pattern of results across both studies, despite the great difference in electoral contexts, provides evidence that the empirical connection between perceived corruption and willingness-to-cheat is likely to be endogenous to the structural position of the election winner rather than the ideological or demographic characteristics of supporters of any given candidate or political party.

6.2 Concerns for Future Research

6.2.1 Systemic Corruption and Cheating: Less Cheating in a More Corrupt World?

One of the unusual findings from both studies was that perceptions of systemic corruption were reliably associated with decreases in willingness-to-cheat. The theoretical literature on perceived corruption, legitimacy and trust almost universally agrees that these perceptions should increase dishonest behavior —as such the negative relationship between perceived systemic corruption and willingness to cheat was both unexpected and perplexing.

Distinguishing between Political and Administrative Institutions

I propose that this unexpected finding can be explained by examining the first-order subfactors that constitute the higher-order construct of "systemic corruption." In the current study, given the sample of survey items used to construct the corruption index, systemic corruption can be thought of as largely capturing people's perception of corruption in the political realm (I.e. most items that contribute to the measures of systemic corruption focus on political aspects of systemic corruption). To test this idea, I ran initial analyses examining the relationship between willingness-to-cheat and the subfactors of systemic corruption.

In both Study 1 and Study 2, I saw initial, preliminary evidence that —when I controlled for participant's opinions of political institutions —increases in perceived corruption in administrative or legal institutions was actually associated with an increase in cheating, as would be predicted by theories on legitimacy and voluntary compliance. These patterns suggest an important avenue for future research.

In terms of the negative relationship between willingness-to-cheat and perceived corruption in the political realm, one plausible hypothesis is that it derives from participants' desires to distance themselves from the dishonest behavior of the political class. Put another way, after having condemned the political class for corruption, participants may be more willing to forgo an opportunity to cheat as a way of creating distance between themselves and those they have just condemned as immoral. Thus, by avoiding the chance to take advantage of an opportunity to cheat, they can signal to themselves and others that they are not like the corrupt politicians (who, in the current case, are also conveniently from the opposite party and, consequently, are even more aversive than a generic corrupt politician). I tested for this possibility in a third study, which was conducted over the 2018 midterm elections. I used the split outcome in the Senate and House to experimentally manipulate participants into believing that their party of choice had won the midterm elections overall (in 2018, Republicans gained seats in the Senate, while Democrats gained seats and won the majority in the House, leaving the "overall outcome" of the midterms unclear and open to framing). This allowed me to experimentally allocate electoral victory or loss at random and test the hypotheses discussed above, thereby establishing the causal links implied by the data in Study 1 and Study 2. One of the significant changes in Study 3 had been the inclusion of more survey items that addressed administrative institutions, which were included in order to test whether there was robust evidence in favor of the hypothesized divergence in the effects of perceived corruption on willingness-to-cheat for judgments targeting political institutions vs judgments targeting administrative or judicial institutions. Unfortunately, a loss of data on the cheating measure for Study 3 prevented me from being able to test this hypothesis. Additional studies conducted around the 2020 Presidential Election will be required before any conclusive statements can be made on the topic.

In addition, in Study 3, I also systematically measured participants' opinions on politicians and asked them to provide similarity ratings between politicians and "people like them." These measures allowed me to test whether the distancing hypothesis can explain the inverse relationship between cheating and perceived corruption in the political realm. The loss of the cheating data precluded a direct and conclusive test of this hypothesis, however, there appeared to be initial evidence supporting this line of inquiry. Further discussion on these topics would benefit from a closer analysis of Study 3 data, integration of those findings with the analyses reported here, and follow-up studies surrounding the 2020 Presidential Election.

CHAPTER 7

CONCLUSION

Overall, the pattern of findings from Study 1 replicated remarkably well in Study 2, despite the significant differences in the nature and context of the two elections. I found a reliable effect of elections on perceptions of corruption that appears to be driven by attribute-substitution as well as a reliable link between perceptions of corruption and cheating, which, while hypothesized in the literature, had not, as of yet, been empirically established. I also found an unexpected negative relationship between perceptions of corruption in the political sphere and people's willingness-to-cheat. This finding was unexpected and would not not be predicted by the current literature. The fact that this pattern emerged in both Study 1 and Study 2 suggests that the negative relationship is robust and captures a genuine consequence of participants perceiving an increase in political corruption. Additional studies, as outlined in the section above, may be useful and may allow this finding to make a meaningful contribution to the current literature, which almost universally assumes a unidirectional effect between perceived corruption and one's willingness to engage in dishonest behavior. Finally, the reliable and bidirectional relationship between perceptions of corruption and cheating provided further evidence that a multi-dimensional conceptualization of corruption is likely to be profitable and will be better able to capture people's judgments in this domain.

Part II

Election Outcomes and Tax Compliance: An Empirical Test of Legitimacy-Based Theories of Voluntary Compliance with the Law

CHAPTER 8

INTRODUCTION

"We can start by recognizing that compliance with the law begins not with the fear of arrest or even of incarceration, but with respect for the institutions that guide our democracy."[1]

– Attorney General Eric Holder, *27 October 2014*

Keynote Address; Annual Conference

International Association of Chiefs of Police

8.1 Overview

Why do people comply with the law? Although compliance with the law derives in part from the threat of reprisal (both in turns of magnitude of punishment and likelihood of being apprehended), it is clear that a significant proportion of law-abidance cannot be explained simply as a response to the deterrent powers of the state (Tyler et al., 2015). Instead, there is fair amount of evidence that people choose to voluntarily comply with the law.

Since there are both ethical and practical reasons to minimize the coercive power of the state, the field of criminology and policing has long been interested in the antecedents of voluntary compliance with the law. In recent years, the idea that voluntary compliance is heavily influenced by the "perceived legitimacy" of legal authorities has been gaining currency among practitioners and theoreticians alike (Levi et al., 2012; Tyler, 2004). The essence of such theories is that "if [citizens] regard legal authorities as more legitimate, they are less likely to break any laws, for they believe that they ought to follow them, regardless of potential for punishment" (Nagin & Telep, 2017) (hereinafter, I refer to these theories as

[1] Holder, 2014. Remarks "as Prepared for Delivery" can be accessed from the Department of Justice website at: https://www.justice.gov/opa/speech/remarks-attorney-general-holder-international-association-chiefs-police-annual-conference

“the legitimacy hypothesis”).

Evidence in support of the legitimacy hypothesis is currently limited and draws primarily from analyses of large-scale (national or international) survey data —including panels with multiple measurements of the same individual or entity over time (often, annual longitudinal panels)(Beetham, 1991a; Levi et al., 2009; Nagin & Telep, 2017; Tankebe & Gutierrez-Gomez, 2017; Tyler, 2006). Based upon such data, for example, multiple researchers have found evidence that *“perceptions of procedural justice”* are closely tied to *“perceptions of legitimacy”*; and, both *“perception of procedural justice”* and *“perceptions of legitimacy”* appear to be closely associated with (self-reported) willingness to comply with the law (Jackson et al., 2012; Levi et al., 2012; Tyler, 2006, 2004). However, as Nagin & Telep (2017) note in a recent review of the legitimacy literature: although *perceived legitimacy* and *voluntary compliance* may co-occur, there is almost no evidence establishing a causal relationship between the two constructs, a concern that is conceded even by the most prominent scholar and advocate of the legitimacy hypothesis (Tyler, 2017a, 2017b).

The most stringent critiques of the legitimacy hypothesis hinge on the claim that “perceptions of … legitimacy cannot be directly manipulated” (Nagin & Telep, 2017). Instead, most field interventions manipulated the manner in which legal authorities interact with the public, which influences many factors above and beyond the perception of legitimacy (Nagin & Telep, 2017). Thus, most current evidence in support of the legitimacy hypothesis relies upon self-reported intentions to comply with the law, which can diverge from behavioral measures (Dang et al., 2020) and may not be reliable predictors of real-world behavior due to a variety of well-established biases, including social desirability concerns (Abeler et al., 2019; Bertrand & Mullainathan, 2001), demand effects (Levitt & List, 2007; Zizzo, 2010) and self-presentation goals (Dufwenberg & Dufwenberg, 2018; Frank et al., 2017; Hao & Houser, 2017). And, to make matters worse, when studies have attempted to implement naturalistic, legitimacy interventions in real-world contexts, there appears to be little evidence that they have produced any meaningful changes in voluntary compliance (Nagin & Telep, 2017).

Such critiques, however, may be too pessimistic. Although any reasonable assessment of the above-mentioned critique must concede its core argument: it is incredibly hard to experimentally manipulate "perceptions of legitimacy" in naturalistic settings —especially when trying to balance (i) the need for realism and external validity with (ii) the need for fine-grained control over the constructs being manipulated.

Nonetheless, as I argue in the following section (see Theoretical Framework: Constructing an Empirical Test of the Legitimacy Hypothesis), three central features of the current American political system make it possible to treat election outcomes as naturalistic shocks to the perceived legitimacy of the Federal government. First, the Presidency has come to stand-in for "the Government" for most of the US populace. As such, opinions about the President easily transfer to opinions about the government at-large (see, Current Dissertation —Part 1: Perceived Corruption). Second, in a de-facto two-party system (see Note on 3rd Party Voting)[2], Presidential Elections can appear to be a zero-sum game —ones that facilitate strictly binary interpretations: either "my team won" or "my team lost" —no intermediate interpretations are easily accessible. Third, wide-spread hyper-partisanship and polarization leads to a broad indictment of the government-at-large when it is controlled by the opposing party (Abramowitz & Webster, 2018). As such, election outcomes can drive significant shifts in the perceived competence, trustworthiness and fairness of the government —all of which are central inputs to the perceived legitimacy of the government (Tyler, 2006).

Taken together, these three features provide the basis for claiming that US Presidential

[2]The elections of 1992 and 1996 - when Ross Perot ran an unusually strong 3rd party challenge - are unusual in modern Presidential politics for the sheer prevalence of 3rd party voting (with 19.6% in 1992 and 10.1% in 1996 of the vote going to 3rd Party candidates respectively). Since 1996, most presidential elections can be seen as adhering almost strictly to a two-party voting pattern. Even the most unusual election had relatively minimal support for 3rd parties - with only 5.7% of the populace voting for the 3rd party in 2016 despite having two of the most unpopular Presidential candidates topping the ballot. In other elections, the 3rd party vote remained between 1-4% (3.7% in 2000; 1% in 2004; 1.4% in 2008; and, 1.7% in 2012). The unique situation in 1992 and 1996 (as well as 2016 to some degree) is addressed in subsequent sections, when choosing between differing methods for constructing party vote share (See Section B.1 Classification Based upon Simple Majority vs Absolute Majority)

Elections can be treated as naturalistic shocks to the perceived legitimacy of the US government (with the size of this effect determined by the strength and importance of partisan identity). The details of this position, along with other theoretical and empirical support, is presented in Section 8.2 below (Theoretical Framework: Constructing an Empirical Test of the Legitimacy Hypothesis).

8.1.1 Tax Compliance as an Ideal Case Study

Having identified a potential naturalistic shock to perceived legitimacy, we must address the second portion of the critique: previous attempts did not sufficiently isolate the effects on the legitimacy intervention. In fact, most of them relied upon making significant changes to both the nature and the quality of the interactions between law enforcement authorities and the public —making it nearly impossible to disentangle the effect of perceived legitimacy from other causes (Nagin & Telep, 2017). To address this concern, I rely upon tax compliance as the case study for "compliance with the law" because it allows for an identification strategy that can isolate election-based shocks to perceptions of legitimacy from election-based shocks to the legislative or administrative components of enforcement actions (see section, Section 12.1 Identification Strategy). As such, tax compliance can serve as an ideal case study for the present purposes —one that can leverage election-based shocks to construct an empirical test of the causal connection between perceived legitimacy and voluntary compliance with the law.

As detailed in Section 8.3 below (Empirical Framework), there are many characteristics of tax compliance that make it an ideal case study, including, inter alia: (a) *data availability*: there exist open-access, longitudinal data on actual tax payments (from which, I could construct proxy measures for compliance); (b) *relevance*: tax non-compliance is both prevalent and can be a significant source of threat to good governance; (c) *potential gradations in compliance*: unlike some other forms of law abiding behavior, with tax compliance, there is a potential for degree of non-compliance —and, such gradations in compliance suggest that it is a domain where it might be easier to detect behavior caused by factors other than the coercive threat of law-enforcement actions; and, (d) *importance of voluntary compliance*: finally, there is substantial evidence that current levels of tax compliance cannot be sufficiently explained with enforcement threat alone —with many researchers thereby suggesting that voluntary compliance plays a significant role (Alm, McClelland, et al., 1992; Bernheim & Rangel, 2005a; Congdon et al., 2009a; Cummings et al., 2009; Feld & Frey, 2002, 2004; Hofmann et al., 2008; Kirchler et al., 2008; Luttmer & Singhal, 2014; Pickhardt & Prinz, 2014; Plumley, 2002).

In addition to these features, a series of quirks in the American election system allows me to construct an identification strategy that can leverage election-based effects while isolating the effects of the legitimacy shocks from other changes in material factors. Specifically, the crux of the strategy depends upon limiting the analyses to the taxes paid for "election tax years" (hereinafter, I used election tax years to the taxes paid for economic activity conducted between Jan-Dec of an election year: 1992, 1996, 2000, 2004, 2008, 2012, and 2016). The reason for focusing on these years is three fold: (a) *real economic activity is insulated from election results* —the election results are only known in November, by when most of the tax year has already passed and thus is untouched by the election outcome. As for the final two months of the election year: according to recent work by Mian et al. (2018), there does not appear to be measurable shifts in economic behavior in the two months immediately following an election (i.e. Nov-Dec of these years); (b) *legislative and regulatory conditions are insulated from election results*: the election results are known in November, but the new administration does not take office until January —thus, the President-elect cannot impact any of the legal or administrative conditions governing those tax years; (c) *lame-duck sessions do not produce significant legislation*: due to norms governing the "lame-duck" session (~ 1 month), the out-going President is severely hampered in the range of activities that are possible during the lame-duck session —and, significant changes to the legal or administrative conditions are not customary (e.g. an important exception is 2008, where the financial crisis resulted in extraordinary actions —these are addressed in greater detail in Section 10.3.2 Concerns with Standardizing using Number of Returns: The Economic Stimulus Act of 2008 and Section 12.3 Testing 2007 Imputed Data; (d) *tax-filing season coincides with outcome-salience*: the vast majority of tax returns are filed between January-April of the following year, in the exact window when the outcome of the election is most salient (starting with the inauguration on 20th January and going through the "first 100-day window" when the newly elected President attempts to push major agenda items).

In sum, by focusing on election tax years, I am limiting the analyses to the specific case where taxes are accrued on economic activity largely conducted prior to the election, the tax liability is governed by laws and administrative rules that are unaffected by the election, but the actual taxes are filed in a window when the election outcomes are most salient to the tax payer. By comparing relative changes in tax compliance for these years, it is possible to examine the potential effects of changes in perceived legitimacy while holding constant the material conditions governing tax compliance and enforcement actions (Kirchler et al., 2007; Plumley, 1996).

The remainder of the paper is organized as follows. In Section 8.2 Theoretical Framework: Constructing an Empirical Test of the Legitimacy Hypothesis, I briefly review the literature focused on the legitimacy hypothesis, including the hypothesized antecedents to perceived legitimacy. Based upon theoretical and empirical evidence, I outline: (i) why one may expect partisan identity to serve as a mediator of perceived legitimacy; and, consequently, (ii) why one may expect presidential elections to serve as naturalistic, exogenous shocks to perceived legitimacy. In Section 8.3 Empirical Framework: Voluntary Tax Compliance as a Case Study, I detail why tax compliance serves as a particularly beneficial case study for empirically testing these claims. Most crucially, in selecting tax compliance as the case-study, I present an identification strategy that allows me to construct an empirical test of the causal relationship between perceived legitimacy and voluntary compliance with the law.

In Chapter 9 Main Hypotheses: Voluntary Compliance with the Law, I use the theoretical framework to generate the primary hypothesis and corollaries that arise from these claims. I then consider alternative hypotheses. In Chapter 10 Data, Measures and Descriptive Statistics, I cover the basic characteristics of the three datasets used for the empirical tests, an overview of the primary trends in the main datasets, and a summary of the data-cleaning and exclusion strategies. In Chapter 11 Graphical Representations of Hypothesized Effects, I present the main hypothesis as time-series graphs using the dependent variables of interest. The hypotheses graphs are intended to serve as a reference guide for the reader as they evaluate the graphical analyses presented in the next chapter. In Chapter 12 Graphical Analyses, I present the main analyses and results in graphical format —with the goal of allowing the reader to gain familiarity with the data. In Chapter 13 Statistical Analyses, I select the most promising trends based upon the graphical analyses and present regression models to provide tests of statistical significance with references to supplementary analyses included in Appendices as needed. Finally, in Chapter 14 Conclusion, I review the main findings, compare the current work with similar work that was completed contemporaneously by Cullen et al. (2018), and present some thoughts on future directions.

8.2 Theoretical Framework: Constructing an Empirical Test of the Legitimacy Hypothesis

Legitimacy-based theories of government have a long history within Western political philosophy (Habermas, 1975; Rousseau, 2018; Weber, 2009) and are intuitively-compelling in nature, which may explain why —despite the lack of causal evidence —legitimacy-based theories of government are widely accepted —in one form or another —by both practitioners and the political establishment. For example, in the field of criminology, policy derived from legitimacy-based theories are being vocally and actively promoted across the country and the world (Hough et al., 2010; Jackson et al., 2012; Jackson, 2018), including most famously in President Obama's Task Force on Policing —which identified the need to "nurture legitimacy" as the first of six pillars for police reform (President's Task Force on 21st Century Policing, 2015). This attitude was best captured in Attorney General Eric Holder's speech to the Association of Chiefs of Police, where he said: "A substantial body of research tells us that —when those who come into contact with the police feel that they are treated fairly —they are more likely to accept decisions by the authorities, obey the law, and cooperate with law enforcement in the future —even if they disagree with specific outcomes." Given the increasing policy emphasis on legitimacy-based approaches, it is of both theoretical and practical importance that we examine the causal relationship between perceived legitimacy and voluntary compliance with the law.

8.2.1 Challenge of Testing

Political Identity as Primary Mediator of Perceived Legitimacy

The main challenge in constructing an empirical test of these theories lies in the belief that "perceptions [of legitimacy] cannot be directly manipulated by experiment or policy" (Nagin & Telep, 2017). The current work challenges this idea by proposing that political identity serves as a primary mediator of perceived legitimacy under a (de-facto) two-party system, especially during periods of heightened polarization. By relying upon the relationship between perceived legitimacy, political identity and polarization, the current work proposes that electoral outcomes can be treated as semi-exogenous shocks that can be used to construct an empirical test of the causal relationship between perceived legitimacy and voluntary compliance with the law.

8.2.2 Three Primary Characteristics of US Election System

Three central features of the current American political system make it possible to treat election outcomes as naturalistic shocks to the perceived legitimacy of the Federal government. They are: (a) the tendency among the populace to have the President serve as the stand-in for "the government" at-large; (b) the effect that a (de-facto) two-party system has on promoting a zero-sum interpretation of election outcomes; (c) the role of hyper-partisanship and negative partisanship in transforming electoral outcomes into existential indictments of the government. Each of these issues are discussed in detail below.

The Imperial Presidency: How the President is "the Government."

First and foremost, as seen in Part 1 of the Dissertation, in the current American political system, the Public strongly associates the US Federal Government almost solely with the occupant of the White House —for example, as seen in economic expectations (Mian et al., 2018). And, despite the significant amount of administrative and political power at the State and Local level —especially in matters of everyday importance (e.g. schools, policing, and criminal courts) —the US Federal Government serves as the stand-in and default point of reference for sentiments about "the Government." Thus, within the American system, opinions about a particular President can dramatically influence opinions about not just the Federal government and but also of "the government" at-large.

Two-Party Systems Produce Binary Zero-Sum Frames: Either You Win or You Lose

Second, due to the two-party Presidential system —and, in contrast to a multi-party Parliamentary system (Dahlberg & Linde, 2016; Hooghe & Dassonneville, 2018; Zmerli & Hooghe, 2013) —the results of a US Presidential Election naturally lend themselves to discrete, binary interpretations of the outcomes in zero-sum terms —for most voters, either your team won or it lost —there are no intermediate outcomes.

This tendency towards binary interpretation of election outcomes is exacerbated by the Public's almost total discounting of Congressional election results during a Presidential Election cycle. For example, although many Presidential Elections produce gains in Congress that favor the President's party[3], there are also cases where the President's party suffers losses in Congressional races. When we consider these "mixed outcomes" —the Congressional gains do not appear to temper the sense of defeat: the fact that —in 2016 —Democrats gained 2 seats in the Senate and 8 seats in the House did very little to tamp the belief among Democrats that "the government has fallen into the wrong hands." Similarly, a gain of 4 Senate seats by Democrats in 2000 did little to lessen the sense of defeat for Democrats after the Bush v. Gore election.

Finally, due to the winner-takes-all nature of the Presidential contest, there is no gradation of losses —e.g. the fact that Bush was declared the President in 2000 on the basis of 571 votes in Florida is no source of comfort for Democrats —and, in fact, close results may work to further exacerbate the negative response, since, the counter-factual is more easily imagined in close-class (Kahneman et al., 1982), as was found in studies of perceived electoral integrity among the losing side (Sances & Stewart, 2015).

In sum, due to the two-party, winner-takes-all, Presidential system of politics, every 4 years, after the Presidential elections, approximately one half of the country feels a sense of total defeat while the other feels a sense of total victory.

[3]So, e.g. 1992: 0D in the Senate / -9D in the House of Representative; 1996: -2D in the Senate / +2D in the House of Representative; 2000: +4D in the Senate / +1D in the House of Representative; 2004: +4R in the Senate / +3R in the House of Representative; 2008: +8D in the Senate / +21D in the House of Representative; 2012: +2D in the Senate / +8D in the House of Representative

Hyper-Partisanship Transforms Binary Electoral Outcomes into Existential Indictments.

The effects of a two-party system are exacerbated when combined with widespread polarizations and the adoption of hyper-partisan political identities. It is this feature of the American political landscape that transforms the binary, zero-sum interpretation of Presidential Election outcomes into an indictment of the entire government. For example, perceptions of the current economic situation as well as future economic expectations appear to shift dramatically after election losses and wins (Mian et al., 2018). As shown in Part 1 of the Dissertation, perceptions of corruption also follow a similar pattern. Much of the reason why electoral outcomes appear to result in a broad, sweeping indictment of the entire system can be attributed to the rise of negative partisanship.

Negative partisanship refers to the fact that —not only do people support issues associated with their party and prefer members of their own party —they also increasingly hold derogatory and delegitimizing beliefs about members of the other party —see Citrin & Stoker (2018) for a recent review. This results in a tendency to both question the competence and ethical integrity of the other party and to delegitimize the sincerity of their concerns (Pew Research Center, 2019b). It is this tendency —when combined with a zero-sum view of electoral politics —that results in delegitimizing views of the entire government, starting with doubts about the most central source of legitimacy in a democracy: electoral integrity (Flesken Anaïd & Hartl Jakob, 2017; Sances & Stewart, 2015). This argument is further examined within the theoretical framework of the legitimacy hypothesis (see Section 8.2.3 Perceived Legitimacy: The Threat from Hyper-Partisanship).

8.2.3 Using Elections to Construct an Empirical Test of the Legitimacy Hypothesis

The use of partisanship as an identifying variable is supported both by empirical evidence and theoretical prediction. For example, in Part 1 of the Dissertation, we saw that electoral outcomes produced a significant change in the perceived levels of corruption in government. Perceived corruption is a primary input into perceptions of legitimacy (Hough et al., 2010; Tyler et al., 2015; Tyler & Jackson, 2014) and can significantly influence how citizens interpret the governments actions: both in terms of government decisions regarding the distribution of resources (Stevens, 2016) and in terms of law-enforcement actions or efforts at institutional reform (Börzel & Hüllen, 2014). As Warren (2004) argues —no matter whether corruption occurs in the executive branch, the legislative branch, the judiciary, or the public sphere —within each domain, corrupt actions —by their very nature —undermine the legitimacy of State. For example, within the executive branch, corruption undermines legitimacy because (i) corrupt actions do not "abide by the goals and rules that have been legitimately decided within the more political domains" (Warren 2004, p. 335) and (ii) because such actions diverge "from legitimately decided norms of office for private gain" (Warren 2004, p. 336). Within the legislative branch, corruption undermines legitimacy because corrupt legislative actions are no longer tethered to the public justifications and open deliberative processes that give law-making its legitimacy. Within the judiciary, corruption undermines legitimacy because it severs the link between the actions taken by lawyers, prosecutors, jurors, or judges and the essential "truth-finding and equity/fairness-seeking goals of the process" (Warren 2004, p. 337). And, finally, for the public sphere, corruption undermines legitimacy because it undermines the discursive forces[4] which should be "the constituting force —of public spheres" and prevents them from being able to properly "guide, limit, cor-

[4] Where "discursive forces" are constituted by the "opportunities and spaces to argue and persuade" that gives people their capacity "to move others through normative and factual claims" and allows people "to be moved in turn" (Warren 2004, p. 338).

rect, and legitimate spheres within which administrative power (the state) or money (the market) are dominant" (p. 338). Although the arguments above (Tyler et al., 2015; Warren, 2017, 2004) all draw from Liberal, democratic-theory account of legitimacy, the "realist view of political theory" also agree on this point. As a recent review from the "realist perspective" states, "rather than being a residual question for politics, wherever corruption is at all extensive it threatens the institutions and legitimacy of the status quo, and it reopens the question of the principles on the basis of which the existing order claims people's allegiance" (Philp & Dávid-Barrett, 2015, pp. 388–389).

Given the intimate relationship between perception of corruption and perception of legitimacy, the empirical findings reported in Part 1 of the current dissertation already serve as initial evidence that election outcomes should impact perceived legitimacy in ways that lead to increases in perceived legitimacy following a favorable election outcome and decreases in perceived legitimacy after an election loss. This intuition is further supported by data from recent national surveys, which provide robust evidence in favor of the hypothesized effect of elections on perceived legitimacy (Pew Research Center, 2017a, 2014, 2019c). However, independent of these empirical findings, the hypothesized link between elections and perceived legitimacy also naturally arises from an examination of theories of legitimacy within a partisan context. To this end, in the following sections, I examine how theories of perceived legitimacy and theories of partisanship intersect to provide independent support for the hypothesized "election effect" —and, do so above and beyond any empirical patterns reported here or in the broader literature.

Precursors of Perceived Legitimacy: Shared Values and Frameworks of Belief

Current theories of law abiding behavior propose that *voluntary compliance* —especially with costly demands —relies on the perception that the government has a "legitimate authority" to demand obedience (Levi et al., 2012). Across a range of disciplines, scholars have argued that the central antecedent to legitimate authority is “a common framework of belief” (Tankebe, 2013; Tankebe et al., 2016; Tankebe & Gutierrez-Gomez, 2017), “a shared sense of values” (Tyler, 2006) or “the acceptance of widespread consensual schemas” (Johnson et al., 2006). In the absence of these antecedents, “the powerful can enjoy no moral authority for the exercise of their power, whatever its legal validity; and their requirements cannot be normatively binding, though they may be successfully enforced” (Beetham, 1991b). As such, for the effective functioning of a nation-state, it is essential the populace widely accepts the basic tenets of a common governing ideology (Tyler, 2017a).

Perceived Legitimacy: The Threat from Hyper-Partisanship

Unfortunately, as recent work has shown (Citrin & Stoker, 2018), there is good reason to believe that increasing partisanship has dramatically compromised the wide-spread acceptance of “the basic tenets of a common governing ideology” across party lines. For example, a recent survey from Pew Research found that differences along party lines now “dwarf all other differences by demographics or other factors” —with an average partisan gap of 39 points across the 30 political values measured, including a 35-point gap on questions regarding the proper role of government (Pew Research Center, 2019c), a trend which is consistent with earlier findings (Pew Research Center, 2017d). These trends are also consistent with related findings about the effect of partisanship on perceived corruption (Blais et al., 2017; Sances & Stewart, 2015), as well as many current theories on the identity-based formation of political beliefs (Gerber et al., 2010; Van Bavel & Pereira, 2018). From this perspective, in an era of scorched-earth politics, political partisanship erodes broadly-shared “consensual schemas” and undermine our sense that we collectively share “a common framework” or “core values” with members of the opposite party.

In addition to undermining a sense of shared values, political partisanship also produces significant in-group favoritism (Oc et al., 2018). For example, political homophily in online dating markets shows an effect size "comparable to that of educational homophily and half as large as racial homophily" (Huber & Malhotra, 2017). Not only does partisanship shape our preferences in a broad range of domains, it also shapes and accentuates the perceived gulf between us and members of the opposing party (Pew Research Center, 2020; Westfall et al., 2015), leading people to overestimate the inter-party differences and underestimate the intra-party disagreements.

Finally, as recent work on negative partisanship has shown, increasingly this in-group favoritism is associated with out-group hostility (Iyengar & Krupenkin, 2018) —producing derogatory perceptions of members of the opposing party as immoral, dishonest, and unintelligent (Pew Research Center, 2016, 2019b). As a result, hostility towards the opposing party has become the primary stated motive for supporting a given candidate, choosing to vote, and engaging in other forms of political action (Abramowitz & Webster, 2018; Iyengar & Krupenkin, 2018). This effect has grown dramatically over the period considered in this paper, with negative perceptions of opposing presidential candidates on a 0-100 (worst-best) scale steadily decreasing from around 40 in 1992 to 12 by 2016, such that —by 2016 — dislike for the opposing candidate was almost twice as important in predicting party loyalty as feelings towards one's own candidate (Abramowitz & Webster, 2018).

In sum, highly polarized environments make it more psychologically costly to grant political legitimacy to a government formed by the opposing party without undermining one's deeply-held ideological commitments. If there is no middle ground that is consistent with my values, then compromise and joint-gains become unpalatable. Under such conditions, election outcomes are likely to serve as discrete and potent shocks to people's perception of their government-at-large —with loss of legitimacy occurring whenever the government is not run by one's own party.

8.3 Empirical Framework: Voluntary Tax Compliance as a Case Study

8.3.1 Overview

For reasons discussed in the following sections, "tax compliance" serves as the ideal case study to test the relationship between perceived legitimacy and voluntary compliance with the law.

First, by using tax compliance as the empirical case-study, the current work empirically tests the relationship between perceived legitimacy and voluntary compliance in a large, representative data sample. Most crucially, the key measures track actual tax payments (as opposed to self-reported intentions to act). Moving beyond self-reported measures is critical since it can deviate from actual tax compliance in unexpected ways. For example, previous work using audit data from Danish tax payers showed —unsurprisingly —that self-report levels of tax evasion were very low among actual tax evaders (as verified by audits) (Hessing et al., 1988). In addition, somewhat unexpectedly, they also found that among the fully-compliant subset, some individuals appeared to be falsely reporting evasion that never occurred (Hessing et al., 1988). Finally, the current work allows for a test of the legitimacy-compliance relationship under conditions where there are severe legal consequences associated with non-compliance, and significant financial costs associated with compliance.

Moreover, under the current identification strategy, it should be possible to quantify the proportion of tax compliance that can be directly attributed to changes in perceived legitimacy (or, its antecedents), while holding constant the material factors that may influence tax compliance like the regulatory environment and background economic conditions (Bloomquist, 2003) as well enforcement factors like real detection risk, real probability of enforcement and real penalties associated with enforcement action (Kirchler et al., 2008; Plumley, 1996, 2002).

The current approach cannot control for changes in perceived risk of enforcement —which could be influenced by election outcomes and are known to be a significant factor in deciding whether to comply with one's tax obligations (Plumley, 1996). However, as argued in Hypothesis 3: Perceived Enforcement Risk, election-based effects on perceived enforcement risk should drive compliance in a direction contrary to the one predicted by the legitimacy hypothesis and thus can be distinguished on that basis.

In addition, as a methodological contribution, by comparing the findings from Part 1 of this dissertation with the current findings, I can examine whether artifactual cheating paradigms used in lab studies are informative of real-world compliance decisions. Part 1 examined the effect of election outcomes on lab-based measures of cheating. Part 2 examines the effect of election outcomes on changes in actual tax compliance. A comparison of the two findings can provide some insight into the generalizability of lab-based measures of cheating to real-world measures of non-compliance.

In the following sub-sections, first, I provide some historical context about tax-compliance in the US; second, I briefly examine the literature on tax compliance —with a focus on summarizing the dominant economic ("rational choice") model of tax compliance as well as the more-recent, behaviorally-motivated work in this domain. Among the behavioral inputs, I examine the literature on "tax morale" (i.e. the people's attitudes towards taxation); evaluate the current state of evidence linking attitudes towards taxation to rates of compliance; and, situate this conversation within the broader framework of perceived legitimacy and voluntary compliance with the law. Finally, I present the identification strategy used in the current analyses to test the hypotheses under consideration —specifically, by relying upon the theoretical and empirical justification for using electoral outcomes as pseudo-experiments (see Section 8.2 Theoretical Framework) and by examining the effects of electoral outcomes on tax payments —one can test whether perceptions of tax legitimacy causally determine voluntary tax compliance.

8.3.2 Tax Compliance: Background

The IRS produces an annual report called the "Data Book" summarizing the critical variables associated with tax payments.[5] In its more recent Data Book (2018), covering Tax Year 2017, the IRS reported that it processed more than 250 million tax returns and related documents; and, it collected nearly $3.5 trillion in tax payments, of which $464 billion was returned in tax refunds (IRS, 2019a).

In terms of tax compliance, in its latest report, entitled "Reducing the Federal Tax Gap and Improving Voluntary Compliance," the IRS estimates that 19% of tax payers may be under-reporting their income or under-paying their taxes (81% voluntary compliance rate) —and has estimated that the under-reporting of income may cost the US treasure $458 billion, with $291 billion resulting from individual income tax (*Tax Gap Estimates for Tax Years 20082010*, 2016). These compliance rates seem to match the findings of the 2018 Comprehensive Taxpayer Attitude Survey (CTAS), 85 percent said it is "not at all acceptable to cheat on their income taxes," while 95% said "it is a civic duty to pay their fair share of taxes" (IRS, 2020b)

In order to decrease non-compliance, the IRS report on "Reducing the Federal Tax Gap" —the latest of which focuses on "Improving Voluntary Compliance" (*Update on reducing the federal tax gap and improving voluntary compliance*, 2009) —highlights four principal approaches: (a) increased enforcement; (b) simplification of the tax code; (c) increased assistance in filing; and, (d) the targeting sources of non-compliance with specificity.

Based on the literature on perceived legitimacy and perceived corruption, the current work identifies one potential source of non-compliance: changes in perceived legitimacy as a consequence of election outcomes (and, more broadly, the cost to effective governance as a result of heightened polarization of the electorate).

[5] https://www.irs.gov/statistics/soi-tax-stats-irs-data-book

Unique Benefits of Using Tax Compliance as a Case Study for Testing Theories of Legitimacy-Based Compliance

There are numerous reasons why tax compliance would be a particularly profitable case study to test the legitimacy hypothesis under consideration. Five primary reasons are discussed in greater detail, namely: (a) role of voluntary compliance in taxes; (b) applicability of tax compliance to a broad segment of the population; (c) highly salient cost of tax compliance; (d) absence of salient harm to others as a result of non-compliance; and, (e) practical relevance for public policy.

First and foremost, a significant majority of the research on tax compliance suggests that the current rates of tax compliance cannot be explained in terms of deterrence alone. As the IRS itself notes, "the US tax system is built on voluntary compliance" (IRS, 2013). If we only take into account the deterrent effects of enforcement efforts by the IRS (in terms of the risk of detection and punishment consequences), tax compliance rates in the US are entirely too high —even after assuming unusually high rates of risk aversion —a problem known as the "compliance puzzle" in the tax literature (Plumley, 2002). Since law enforcement efforts by tax authorities appears to be insufficient to explain the current rates of tax compliance, most researchers have concluded that voluntary compliance must be a major factor in the tax-related decision processes (Alm & Yunus, 2009; Gentry & Kahn, 2009). Second, it is worth noting that approximately ~40-50% of the American population is required to file taxes annually. For example, in 2019, any single person making more than $12-13,000 is required to file taxes with the IRS (IRS, 2020a). As a result, millions of Americans must annually decide whether to comply with tax law or not —making tax compliance a domain of voluntary compliance that is broad, representative of the population. Third, tax compliance requires upfront costs —i.e. in order to obey tax law, I must actively pay money (as opposed to other domains like theft, where compliance results in forgoing illicit gains). Fourth, tax avoidance does not require one to commit a salient, direct harm on others. Since causing direct harm is particularly aversive (Greene, 2016), compliance in those cases may be driven by other psychological mechanisms beyond the legitimacy of the law.[6] And, finally, as a practical matter, tax compliance is a major issue facing governments worldwide. Since it can be cost-prohibitive for governments to gain compliance solely through enforcement, there is important practical value in examining the behavioral factors that determine compliance (Chetty et al., 2009; Congdon et al., 2009b; Hallsworth et al., 2014; Rees-Jones & Taubinsky, 2016).

8.3.3 Models of Tax Compliance

The utility-theory model (hereinafter, the "standard model") of tax compliance treats the compliance decision at the individual level as a choice between a sure loss (full compliance) and gamble (evading taxes). The sure loss is the money lost by choosing to be fully compliant (i.e. paying taxes due on all true income). The gamble results from the decision to evade taxes, where the pay-off from the gamble is determined by (a) the amount saved in taxes as determined by the marginal tax rate (Alm et al., 1990); (b) the probability of detection; and (c) the potential penalties if detected (allingham1972income; Gentry & Kahn, 2009). Yet, in most countries, the level of compliance is too high to be explained purely in terms of deterrence, thus suggesting that a purely economic approach cannot explain the pattern of data without assuming unrealistic levels of risk aversion (Alm et al., 2010; Alm, Jackson, et al., 1992; Dulleck et al., 2016). The standard economic model also produces some unexpected predictions —e.g., that the effect of an increase in marginal tax rate should result in a decrease in tax evasion —have also failed to meet the standards of intuition or empirical evidence. Such descriptive gaps in the standard model led most researchers to conclude that a more behaviorally-informed approach was required (Bernheim & Rangel, 2005b; Congdon et al., 2009c; Luttmer & Singhal, 2014). And, recent work has shown that incorporating

[6] After all, one could be a committed Marxist that does not support the legitimacy of laws governing the primacy of private property and yet be averse to robbing rich, old ladies.

behavioral insights, e.g. insights from cumulative prospect theory (Kahneman, 1992), can correct such descriptive failures (Bernasconi et al., 2014; Bernasconi & Zanardi, 2004).

Behavioral Models

Numerous examinations of tax data —both audit studies and group measures —provide clear evidence that tax evasion is very highly sensitive to the existence of opportunities to cheat — a finding that is both sensible and hardly unique to the United States (Bloomquist, 2003). In fact, a clear body of work confirms that deterrence (Almunia & Lopez-Rodriguez, 2018) and deterrence messaging promotes tax compliance (Blumenthal et al., 1998; Brockmeyer et al., 2019; Castro & Scartascini, 2015; Doerrenberg & Schmitz, 2017; Kleven et al., 2011), and can do so even indirectly through "spillover effects" across social networks (Boning et al., 2018; Lopez-Luzuriaga & Scartascini, 2019). Even more behavioral approaches have confirmed such a relationship, showing that when taxpayers have limited attention, interventions that increase the salience of penalties (either fines or legal action) can increase tax compliance (Bernheim & Rangel, 2009, 2007).

However, evidence that tax payers are extremely sensitive to the opportunity to cheat does not imply that a rational choice model is sufficient and nor does it clarify the non-material antecedents of tax compliance. At this point, the "rational model" of tax compliance serves as the straw man (Alm, Jackson, et al., 1992; Kirchler et al., 2007), as many scholars of tax evasion move to adopt and improve upon behaviorally-inspired expected utility models like Prospect Theory (Dhami & al-Nowaihi, 2010, 2007).

As shown in the section below, a review of the current research on behavioral aspects of tax compliances reveals the following: (1) non-material factors that were traditionally ignored by the standard model can play a substantial role in the decision to comply or evade taxes; (2) there is clear evidence that standard biases from the Judgment and Decision Making (hereinafter, JDM) literature are present in tax compliance decisions; (3) factors like political identity and political ideology are especially relevant when considering the influence of non-material factors on tax compliance; and, 4) there is some evidence that election effects can influence the antecedents to perceived legitimacy of taxes, for example, the perceived fairness of taxes or the perceived wastefulness / efficiency of government expenditure of tax monies.

Non-Material Factors and Other Biases in Tax Compliance Decisions As with many other domains of economic research, recent findings have highlighted the fact that behavioral factors can often be among the central determinants of of tax compliance (Castro & Scartascini, 2019; Congdon et al., 2009c; Pickhardt & Prinz, 2014), work which has resulted in changes to the standard utility-theory model. For example, tax audit data from Sweden showed evidence of loss aversion in tax compliance decisions —in a manner consistent with Prospect Theory (Kahneman, 1992) —with a measurable increase in the amount of deductions claimed once people learn about an impending tax balance (Engström et al., 2015). An investigation of individual level tax data from the US also confirmed this exact tendency (Rees-Jones, 2018). Statistical analyses that examined the shape of the density function of total deductions revealed a "missing mass" around $0, leading the researchers to conclude that taxpayers facing a payment on tax-day reduce their tax liability by $34 more than taxpayers who were owed a refund (Rees-Jones, 2018).

In addition, tax morale (i.e. people's feelings and opinions regarding their tax obligations) may have the most substantial effects on tax compliance above-and-beyond perceived audit risk or the penalties associated with tax evasion (Cummings et al., 2009; Luttmer & Singhal, 2014), including the use of tax evasion as a tool for political signaling (Braithwaite, 2009). This work on tax morale fits nicely in support of a broader framework that emphasizes perceptions of legitimacy as being crucial for voluntary compliance.

Role of Political Ideology in the Perceived Legitimacy of Taxes

Tax Morale: Role of the Perceived Quality and Competence of Government Outputs As mentioned above, "tax morale" refers to people's opinions, attitudes and feelings about taxes. Attitudes towards taxes are significantly impacted by people's beliefs about both how governments spend tax funds and who benefits from government expenditure (Sussman & White, 2018). These two judgments are considered absolutely central to the perceived competence of a regime and, thereby, central to the perceived legitimacy of the authorities in charge (Levi et al., 2009). Perceived legitimacy increases when there is clear information and evidence about public goods and services provided, the goods and services are deemed worthwhile, and they are only provided to "deserving recipients" (Levi et al., 2009, 2012).

The finding that "information about public goods and services" is a central antecedent to perceived legitimacy has a parallel in the literature on tax attitudes and tax morale. In the tax morale literature, it is held that people's opinions about the manner in which tax dollars are spent can significantly impact their tax attitudes (Sussman & Olivola, 2011). In fact, individuals with anti-tax beliefs can be made to adopt more neutral positions when provided information about the positive ways in which tax dollars are spent (Sussman & Olivola, 2011), and, in some cases, exposure to such information can reduce or erase the differences between liberals and conservatives on the perceived legitimacy of taxation (Duhaime & Apfelbaum, 2017). Overall, comparing these sets of findings suggests that the research on "tax morale" and the research of "perceived legitimacy" are tracking the same underlying set of political cognitions (Djawadi & Fahr, 2013; Duhaime & Apfelbaum, 2017; Lamberton et al., 2014).

Political Ideology and Partisan Lenses

Theoretical Reasons to Expect Elections Effects on Tax Legitimacy It is a stable finding in public opinion research that conservatives tend to hold more negative opinions about taxation than their progressive counterparts (Sussman & Olivola, 2011). However, rather than being stable preferences that reflect well-founded, deeply-held policy opinions *per se* —most voters develop preferences over tax policy largely as markers of their political identities and see their tax attitudes as opportunities to express those identities (Fernbach et al., 2013). As a result, we see that tax attitudes are relatively malleable —depending upon how one feels about the current government. For example, opinions about the wastefulness and inefficiency of government use of tax money or the perceived fairness of the current tax system shifts dramatically depending upon whether one's preferred political party is in charge of the White House (Pew Research Center, 2019a) —with changes in perceived fairness or competence taking place immediately after elections (and, even before the new administration has taken power). In addition, as discussed above, perceived legitimacy is also impacted by perceived outputs of the system (Holmberg et al., 2009; Rothstein, 2009), including types of spending (Doerrenberg, 2015; Doerrenberg & Schmitz, 2017), and selective awareness of these outputs (Duhaime & Apfelbaum, 2017). All these factors are also known to be influenced by the degree of perceived political alignment with the government. Taken together, we thus have strong reason to believe that political alignment with the government should impact the perceived legitimacy of the tax authorities —both through direct effects on the election on the perceived fairness of the tax code and through indirect effects on the perceived competence of the authorities.

Tax Morale: Impact on Tax Compliance. There is some evidence that tax attitude are related to people's willingness to comply with their tax obligations. For example, a recent study showed that a "single one-unit increase in institutional trust leads to a 15 percentage point increase in the willingness to pay more taxes to help the needy" or, to support other public causes like "public health care and education" (Habibov et al., 2018), a finding that is in line with the general thrust of the tax morale literature (Sussman & White, 2018). However, as with the broader legitimacy literature, the link between tax morale or tax attitudes more broadly and tax compliance is almost entirely based upon either: (a) self-reported measures of compliance; or (b) self-reported willingness to comply; or (c) correlational data linking low-levels of tax compliance to negative attitudes towards taxes. Given this gap, for the current purposes, tax compliance is a particularly well suited for our purposes —both as a relevant case study for testing the legitimacy hypothesis and for the potential contribution of such analyses to the "tax morale" literature.

CHAPTER 9

MAIN HYPOTHESES: VOLUNTARY COMPLIANCE WITH THE LAW

The current work relies on the heightened polarization of the electorate as the basis for treating electoral outcomes as a exogenous shock to perceived legitimacy. Within this context, I predict that suffering an electoral loss would drive highly partisan individuals to experience a sudden and salient drop in the perceived legitimacy of the broader government. And, if the legitimacy hypothesis correct, in the absence of "legitimate authority," voluntary compliance with costly obligations like taxes would come to rest almost solely on enforcement effort —as such, we should see a decrease in legal compliance following electoral loss (Section 9.1 Primary hypothesis: Perceived Legitimacy Matters).

Moreover, as partisanship increases, there is an increase in the perceived gulf between "competing sets of values", which should directly increase the effect (see Hypothesis 1.a). And, this perceived lack of "shared values" should become more salient during elections and political transitions, when political parties spend significant efforts attempting to differentiate each other (Hypothesis 1.c) - a process that is only exacerbated in highly contentious elections (see Hypothesis 1.b).

However, it is quite possible that the election outcomes impact voluntary compliance through two alternative non-material pathways: moral licensing (see Hypothesis 2: Moral Licensing and a Winner Effect) and changes in risk perception (see Hypothesis 3: Perceived Enforcement Risk), both of which are described in Section 9.2 Alternative Hypotheses: Victory Not Legitimacy is What Matters. Crucially, both these alternative pathways would predict a change in compliance that is contrary to the one predicted by the legitimacy hypothesis. As a result, the current approach can distinguish between the legitimacy hypothesis and these alternatives (for details on distinguishing between the moral licensing and the risk perception hypotheses, see Section 9.2.3 Distinguishing Between Alternatives).

9.1 Primary hypothesis: Perceived Legitimacy Matters

9.1.1 Hypothesis 1: Perceived Legitimacy Drives Voluntary Compliance

People are less likely to voluntarily comply with legal obligations if their preferred party loses the election. The primary reason for a decrease in compliance is due to a drop in perceived trustworthiness and competence of the government —which is a central input into judgments of perceived legitimacy. This proposed mechanism also leads to the following three corollaries described below. These hypotheses are also shown as graphs of predicted effects using simulated data in Chapter 11 (Graphical Representations of Hypothesized Effects).

Hypothesis 1(a) - Partisanship

The effect of the election on compliance should be determined by the level of partisanship. As partisanship increases, there is an increase in the perceived gulf between "competing sets of values", which should directly increase the effect —with the effect size monotonically linked to degree of partisanship. Since partisanship has increased over time (Mian et al., 2018), we should see an increase in the effect size across the time sample studied here.

Hypothesis 1(b) - Salience

The perceived lack of "shared values" will be most salient immediately after an electoral loss, which suggests that the effect of the election on compliance should be largest immediately following the presidential election (1992, 1996, 2000, 2004, 2008, 2012, and 2016) and this effect should slowly decrease in the following years. In addition, the salience of an electoral outcomes should be large following elections that result in a change in the ruling party (1992, 2000, 2008, 2016) as compared to elections where the ruling party retained power (1996, 2004, 2012).

Hypothesis 1(c) - Contentiousness

The salience-driven effects of the election will be exacerbated in highly contentious elections, a hypothesis similar in approach to the recent work by Chen & Rohla (2018). Thus, the effect of the election on voluntary compliance should be determined by the contentiousness of the election and the degree of perceived difference between the winning party and the preferred party i.e. the size of the gulf in "shared values." This perceived gulf should be especially large after elections where the political parties have spent significant efforts trying to differentiate themselves from their opponents and attempting to persuade the electorate that there exists a vast gap between us and them.

9.2 Alternative Hypotheses: Victory Not Legitimacy is What Matters

9.2.1 Hypothesis 2: Moral Licensing and a Winner Effect

In contrast to the legitimacy hypothesis, the moral licensing hypothesis would propose the opposite effect: winning elections would result in a decrease in voluntary compliance. The premise of such a hypothesis is that there is a "winner effect" such that the perception of victory produces a temporary reprieve from the factors that drive voluntary compliance — primarily by promoting the self-serving belief that "the rules don't apply to winners." There is evidence consistent with such a hypothesis from the results of Part 1 of this dissertation where participants who identified with the winning party were more likely to lie in order to obtain a larger payment. There is also evidence from the cheating literature that has found that winning a competition predicted dishonest behavior (Gentry & Kahn, 2009; Jacobsen & Piovesan, 2016; Lammers et al., 2010; Schurr & Ritov, 2016). The corollaries from Hypothesis 1 should also apply here, since perceived electoral victory is central to the winner effect.

9.2.2 Hypothesis 3: Perceived Enforcement Risk

Like the moral licensing hypothesis, Hypothesis 3 would predict an *increase in compliance* after losing an election. However, it proposes an alternative explanation for how elections could impact compliance: specifically, *perceived enforcement risk*. After losing an election, members and supporters of the losing party may come to believe that they are more likely to be targeted with enforcement actions by the new government and more likely to increase compliance as a result.

Under such a hypothesis, compliance would decrease (increase) after electoral victory (loss) due to a change in the perceived enforcement risk —with winners assuming that the elected government will not direct its enforcement efforts towards its supporters and losers assuming that they will be scrutinized more heavily as a punitive action. Although such shifts in perceived enforcement risk are less likely to impact criminal actions that come under the purview of local law enforcement, for the current case study —compliance with federal tax obligations —it seems more plausible given the connection between the Presidency and the Federal government, despite the IRS's best efforts to be seen as apolitical and non-partisan.

Some anecdotal evidence for a change in perceived enforcement risk within the tax compliance domain could be seen under President Obama in 2013 when conservative media and activist groups raised the specter of being targeted by the IRS as an act of political retribution (Overby, 2017) —a belief that gained significant currency in conservative circles at the time (Taranto, 2013), with Congressman Hal Rogers, R-Ky., on the House Appropriations Committee, even telling Fox News that the IRS had an "enemies list out of the White House" ("Fox News - Interview Hal Rogers - IRS," 2013). Such actions were not uncommon in the pre-Nixon years (Andrew, 2002). And, the sentiment appears to be prevalent among tax payers as well. According to the IRS's Comprehensive Taxpayer Attitude Survey (CTAS), in 2014 only 60% of American agreed with the statement: "I trust the IRS to fairly enforce the tax laws as enacted by Congress and the President", with ~40% disagreeing slightly or strongly (IRS, 2020b).

9.2.3 Distinguishing Between Alternatives

It is difficult to distinguish between Hypothesis 2 and 3, however, it should be noted that Hypothesis 3 (perceived enforcement risk) could not explain the findings from Part 1 of the current dissertation or, the reported findings from Schurr & Ritov (2016) and would be considered a parallel (or, perhaps, domain-specific) explanation to Hypothesis 2 (the moral licensing effect) described above. One point of distinction would be that Hypothesis 3 effects should only occur when there is a change in the ruling party. There should be negligible effect when the ruling party is re-elected.

9.3 Psychological Model / Hypothesized Mechanism

Thus far, I have presented numerous theoretical and empirical grounds that would support the hypothesized link between perceived legitimacy and voluntary compliance with the law —specifically, with one's tax obligations. I would like to now supplement those formal arguments with a more informal description of how I believe such a decision process would be likely to unfold at an individual level —with the aim of both providing an intuitive sense of why this hypothesis could be plausible as well as providing some intuitive guidance as to which measures would be most likely to demonstrate the predicted effect.

9.3.1 Don't mess with the IRS

"I'm crazy enough to take on Batman, but the IRS? Nooo, thank you!"

– The Joker, *The New Batman Adventures, "Joker's Millions"*[1]

[1]Altieri et al., 1998 —Video accessed online at: https://www.youtube.com/watch?v=QIju16ngAR8

At first glance, it could seem quite counter-intuitive that the election of a new president in November could impact people's compliance decisions in March or April as they sit and file the 1040s. After all, everyone knows that the IRS got Al Capone on tax evasion: *"You know what they say, 'Dont mess with the IRS.'"*[2] There is certainly some truth to this sentiment. For some people, the penalties and stigmas associated with tax evasion may render them fully compliant no matter what. However, I would suspect that such numbers are rather small —and, largely limited to people with no opportunity for non-compliance in the first place (i.e. people whose entire income is reported via W–2s where taxes are withheld in advance).[3] Outside of these "total compliers," there is a wide range of compliance possibilities.

9.3.2 Compliance with all tax obligations can be very time consuming

As the IRS's own models show, compliance is around 80% on average —and, the rate of compliance varies greatly with opportunity. Also, to be fully compliant can be actually quite time consuming and cumbersome (Benzarti, 2015a, 2015b). Some may remember, for example, under President Bill Clinton's first term, his very first nomination for the post of US Attorney General, i.e. his first appointee to be the chief lawyer of the Federal Government of the United States, was scuttled when it was discovered that she and her husband had failed to fail to pay the "Nanny Tax" (Schedule H) for their domestic workers (aka, "Nannygate"[4]). A few years later, President George W. Bush's nominations for the head of the Department of Homeland Security and the Department of Labor suffered the same fate(Bloomquist & An, 2005). Finally, one of the more robust recent findings about the effects of audits shows that direct audits from the IRS not only *increased compliance* among the audited party,

[2]Or, its more "erudite" version that is often attributed to Ben Franklin: "Nothing is certain except for death and taxes."

[3]In future work, I take the opportunity to comply into account when constructing variables as well.

[4]https://www.nytimes.com/1993/01/14/us/clinton-s-choice-for-justice-dept-hired-illegal-aliens-for-household.html

but also among their immediate network —suggesting that there was always some room for additional compliance (Hallsworth, 2014). As such, when we move away from the binary question of “should I comply or not” —to the more realistic question: “how much do I need to (or, want to) comply” —it is possible to imagine multiple places where households make decisions about tax compliance in a manner that are susceptible to psychological factors, including the perceptions about the incoming Presidency or the perceived legitimacy of the government-at-large.

9.3.3 There are many ways to reduce your tax burden

There are multiple ways that one could cheat on one’s taxes.[5] Some of the most obvious involve: (a) deciding not to file; (b) under-reporting income that was filed (especially, income from secondary sources or non-reported streams like cash tips); and, (c) exaggerating the offsets to income and to tax claimed. Although, my original hypothesis —at the time of initiating this project —was limited to testing whether legitimacy determines voluntary compliance, having gained familiarity with this domain, I suspect that the nature and direction of change in government will vary according to the type of compliance.

Compliance in terms of filing rate

I suspect that the decision of whether to file or not file is largely determined by perceived threat of enforcement to such a degree that other factors are unlikely to be relevant. One potential reason for the over-sized importance of enforcement when considering the filing rate may be that non-filing, unlike, adjustment to income or offset reporting, appears unquestionably like “tax evasion” and leaves no room for justification or alibis to oneself or others.

[5] Other forms of non-compliance highlighted by the IRS including late filing, late payments, and failure to pay taxes owed. For the current purposes, we lack any data on filing dates and on payments (whether timely or delinquient) - thus, such forms of non-compliance are not examinable.

Compliance in terms of income reported

Here we should separate under-reporting of income into two categories. In the first category of tax filer, the majority of their income comes from "reportable streams" such as a full-time or part-time salary. These filers may also have other streams of income (e.g. for restaurant workers, bar tenders or delivery drivers, there may be cash tips; others may have side income from small, informal businesses —e.g. having a baking, catering, gardening, moving service, etc), renting portions of their house, or participating in the gig economy etc. Recently, with the rise and widespread adoption of "side gigs" —according to a 2017 online poll, between 30–40% of individuals with incomes between $150–300K make money on the side and approximately 27–30% of them do not declare it to the IRS[6] (this rate of reporting 'side income' matches results from my own surveys in 2018 of Amazon Mechanical Turk workers, who also chose not to report the income made on the platform). The IRS's own study of this problem suggested that 13% of gig economy workers with income above $400 (the threshold required to file) failed to report any of this side income. The IRS has opened a "gig center"[7] targeting this segment and considers it a source of growing concern in terms of compliance.

For those with income not reported to the IRS, how might perceived legitimacy enter the picture? Lets take the example of restaurant workers, many of whom earn a substantial portion of the income in the form of cash tips, especially during the earlier years of the current time sample under consideration. The restaurant servers know that they must disclose some of this tip income to the IRS. However, what is the correct number to disclose? They must pick a number above $0 and (assuming they are not fully honest) less that their true tip income. It seems highly unlikely that they are using some formal system for balancing trade-offs in order to find an optimal (or even a semi-optimal) solution that minimizes their tax burden after accounting for the risk of detection and punishment in a manner that aligns

[6] https://www.finder.com/side-hustle

[7] https://www.irs.gov/businesses/small-businesses-self-employed/manage-taxes-for-your-gig-work

with their personal risk tolerance. It is much more likely that they check with their peers and generally do something similar to what they did last year with some adjustment. It is during ambiguous judgments like this that I anticipate the effect of extraneous factors like perceived legitimacy can push individuals to disclose marginally more or less of their income.

In addition, income disclosure is only one portion of the puzzle. Prior to computing Adjusted Gross Income (AGI), individuals are also allowed to deduct business expenses. The same problem arises for anyone who chooses to deviate from full honesty: how much should I deduct? Consider the case of car mileage for a salesperson —there are many ambiguous trips that could reasonably count or not. How many miles should they consider to deductions? What about when computing square footage of your home office? In all these cases, I expect the main trends are explained by "behavioral factors" like anchoring on past behavior and trying to align with social norms. In fact, some evidence for such a hypothesis comes from the finding that having recently heard about a tax audit increases compliance across social networks (Alm & Yunus, 2009). Thus, it is very likely that when trying to choosing how much to disclose or how much to claim as a business expense, a huge host of extraneous factors are consciously and inadvertently incorporated. And, among those factors, in the margins, I argue that people's perceptions of government may also influence their decisions (either consciously or unconsciously) as they perform these highly ambiguous computations. While these effects may not be large, I believe them to be large enough that —as they aggregate across populations —they should be detectable using a substantially fine-grained longitudinal dataset with a broad array of tax-relevant control variables.

In addition to the model that emphasizes changes in the decision process, there may also be changes in motivation. For example, there is evidence that —due to the burden of tax filing —individuals tend to leave some income on the table by forgoing itemized deductions (Benzarti, 2015a, 2015b). Rees-Jones (2018) has also similar effect, with households more willing to take action when they owe a small tax burden on Tax Day than they are to secure an equivalent tax refund. Similarly, one can imagine a person that finds the current government lacking in legitimacy being more motivated to both itemize deductions and —if itemizing —to seek out a broader range of deductions; both factors that would impact their final tax burden. Although the current work does not assess the election effects at such a fine granularity, such analyses will follow in future work. Instead, the goal here is to give the reader an idea of the kinds of psychological and behavioral pathways that may facilitate the link between perceived legitimacy and tax compliance.

Finally, when talking about tax compliance, I believe it is worth briefly mentioning the distinction between tax avoidance and tax evasion. Tax avoidance is the process of using all legal means ("loopholes") to minimize your tax burden to the extent possible. Tax evasion involves incorrectly reporting information to the IRS in a manner that allows one to decrease their tax burden outside of the law. Tax avoidance is legal; tax evasion is not. In studies of tax compliance using aggregate (non-audit) data, it can be very hard to distinguish between these variables. There are some arguments to be made that tax avoidance should not significantly impact AGI for individual tax returns, and is more likely to be visible when considering total tax burden. For example, itemized deductions are computed downstream of the AGI computation. That said, for my current purposes, given the preliminary nature of these analyses, I remain agnostic on whether tax avoidance is occurring or tax evasion and, in the current work, I use the term tax compliance as the umbrella term to cover both tax avoidance and tax evasion. Whether there is a legitimacy-based increase in tax avoidance or tax evasion, in both cases, the government loses revenue and thus experiences a diminished ability to provide core public services. As such, it is not necessary to distinguish between the two when trying to determine whether and to what degree do changes in perceived legitimacy pose a direct risk to effective governance.

CHAPTER 10

DATA, MEASURES AND DESCRIPTIVE STATISTICS

The current chapter is organized as follows: (a) first, I provide a general overview of the datasets used to test the current hypothesis linking perceived legitimacy to voluntary tax compliance; (b) second, I present each of the three primary datasets used: namely, voting data, tax data, and income data at the county level; (c) for each dataset, (c.i) I examine the specific measures available; (c.ii) I present the primary measure of interest constructed from the available data; (c.iii) I present some basic descriptive statistics; and, (c.iv) I summarize any data-exclusion or data-cleaning strategies that were performed. For the voting data, my goal is to describe how raw vote tallies can be used to construct a measure of partisanship at the county-level for each of the 7 elections in the sample and as well as an aggregate measure of partisanship for each county across the entire sample of 27 years. For the tax and income data, my goal is to present my current approach for constructing a proxy-measure of tax compliance by cross-referencing tax data from the Internal Revenue Service (hereinafter, IRS) with income estimates from the Bureau of Economic Affairs (hereinafter BEA) or income estimates from the Census Bureau.

10.1 Overview of Data

The ideal statistical model for the current project would establish the relationship between:

$$\Delta Compliance \sim \Delta Legitimacy$$

However, since we cannot directly measure either of these variables, we must rely upon proxy measures that reasonably approximate these constructs. For measures of tax compliance, we rely upon longitudinal tax data gathered from the IRS as well as estimates of Personal Income produced by the BEA. For measures of legitimacy, we rely upon indicators of political preference (namely, voting patterns). Since all of these variables are difficult to access at an individual level, all analyses rely upon aggregated data.

The present work uses the county as its fundamental unit of analysis. This allows us to leverage the vast array of county-level data to develop a rich data-profile for each unit across time. I examine data for all 3,142 US counties (and, county-equivalents) from 1990-2017 to determine if political outcomes impact the level of voluntary tax compliance.

The analyses rely upon the merger of three separate datasets: (a) taxes: the first contains county-level tax data; (b) voting data: the second contains county-level vote tallies for all presidential elections; and, (c) income estimates: the third contains annual county-level estimates of population, personal income, and median household income. For tax data, I rely upon the IRS's annual release of the Statistics of Income (SOI) County tax data series, which are available for the years 1990-2017 at the county level. These tax data are then merged with voting data and the measures of political partisanship constructed for each county. Finally, these two components are supplemented by county-level estimates produced by the BEA, specifically: (i) estimates of total personal income for each county; and, (ii) annual, intercensal population estimates in order to standardize income variables on a per-capita basis. These data are also merged with another alternate estimate of income: median household income estimated by the Census Bureau. As a result, for each county I procure two distinct and independent estimates of the true income of a county for a given year.

By examining the ratio of reported income to estimated income at the county level, I can construct proxy measures for tax compliance for each county as described in Section 10.3.3 AGI over PI: Constructing Compliance Measures using the BEAs Income Estimates below. The cross-referenced and merged panel data across these three domains (taxes reported, political preference, and income estimates) provide the foundation for the analyses presented in Chapter 12.4 and Chapter 13.

Additional control variables at the county level are introduced by matching the primary panel data with unemployment data series from the Bureau of Labor Statistics (BLS), demographic information from the Current Population Survey (CPS) and, information about patterns of small business ownership from the County Business Patterns (CBP) survey by the U.S. Census Bureau.

10.1.1 Note on County Harmonization

The total number of counties (and, county-equivalents) in America currently stands at 3,142 (from an initial figure of 3121 in 1989), of which 3098 have existed under the current geographic boundaries for the entire period. Out of these 3098 counties, only 3060 counties have data for all important attributes for the entire 28 years (1990-2017) —these form the master dataset. Due to substantial changes in county structure across time in Alaska, it was necessary to exclude all counties from that state, so as not to impede matching across datasets.

Census Counties were harmonized with the BEA's County Identification systems (resulting in some counties being recombined with an adjacent county to account for changes in county boundaries over time, (see entry on BEA County Equivalents in the Glossary for details). These corrections ensure that perfectly valid county data was not excluded when merging IRS, Voting, Census and BEA data. After the data cleaning procedures, the final dataset contains a total of 77084 observations covering 2753 counties over 28 years each —resulting in a fully balanced panel.

10.2 Voting Data

In an ideal world, we would have survey instruments that directly assessed people's perceptions of legitimacy. However, in the absence of such a measure, I rely upon the partisanship status for each county as the proxy measure for perceived legitimacy. In order to capture the partisanship status of a given county, we rely upon voting data which allows us to estimate the *degree to which a county supports a given candidate*. Ideally, these data would allow me to determine (a) the degree to which the county is in or out of sync with the election winner; and, (b) the degree to which the county is engaged in the electoral process (i.e. cares about being in or out of sync).

For the current analyses, the voting data were compiled using David Leip's Atlas of US Presidential Results (Leip, 2019) as well as the election data sourced from the County Presidential Election Returns 2000-2016 dataset shared by MIT Election Data and Science Lab (MIT Election Data and Science Lab, 2019; van der Wal, 2018). Voting Data for all counties spans a total of 7 Presidential Elections (1992, 1996, 2000, 2004, 2008, 2012, 2016), which involved 4 elections where the White House changed parties (1992, 2000, 2008, 2016) and 3 elections where the incumbent president was re-elected (1996, 2004, 2012). In two cases of turnover, the White House switched from Republican to Democratic control (1992; 2008) and in two cases, it switched the other way from Democratic control to Republican (2000; 2016).

For each election year, these datasets provide the raw vote tallies received by each candidate / party within each county. As shown in Table 10.1, the total number of votes cast in an election have grown from 101 million in 1992 to 132 million in 2016, which constitutes between 35-42% of the overall population in any given Presidential Election year.

Table 10.1: Total Votes and Participation Across 7 Elections (1992-2016)

Election Year	Total Votes Cast	National Population	Approx. Participation Rate
1992	98.52M	242.39M	41%
1996	90.78M	254.49M	36%
2000	99.29M	266.42M	37%
2004	115.13M	276.21M	42%
2008	123.50M	286.71M	43%
2012	121.37M	295.82M	41%
2016	128.79M	304.51M	42%

Note. The participation rate is an approximate estimate. In order to compute actual participation rates, we would need to know the Voter Eligible Population for each election year.'

10.2.1 Using Voting Data to Construct a Measure of Partisanship

At a higher-order level, the task of assigning counties a partisanship status across the 28 year sample requires two steps: (i) quantify the partisanship for a given county in a given election year —either as a continuous measure or a binary one; (ii) aggregate the partisanship measure computed in step 1 for all 7 elections in in a manner that allows us to assign a partisanship status to a county for the entire sample. Each of these steps are addressed individually in the sections below.

Standardizing Election Data

As mentioned above, the first step is to create a common (i.e. standardized) measure of partisanship status from the raw voting data for each county for each of the 7 election years. Counties in America are incredibly diverse in terms of population, land area, and other demographic characteristics. This heterogeneity is also visible in the voting data. The raw voting data varies highly across counties —for example, in 1992, total votes ranged from 442 votes cast by the smallest county to a total of 2.75 million votes cast by the largest one. By 2016, this disparity had grown by half a million, with 434 votes cast by the smallest county and 3.43 million cast by the largest one.

In order to accommodate the heterogeneity across counties and time, for each county —in any given election year —it is necessary to construct a standardized measure of partisanship from the raw voting tallies. When standardizing the voting data, we can either construct a binary variable at the county-election level or a continuous measure.

Binary Measures of Partisanship When election results are presented, the most common approach is to assign counties a binary status either identifying them as Democrat or as Republican. Within any given election, for a particular county, we can produce a binary classification based upon its voting pattern by relying on the two most prominent standards in electoral politics: plurality and majority. It is my contention that —for our purposes —within any given election, both these measures can be considered alternate measures of the "true" classification of a county along party lines.

When classifying counties according to a plurality standard, the county is classified as supporting a given party (Dem. / Rep. / Ind.) if the candidate for that party won the most votes. When classifying counties according to the majority standard, the county is classified as supporting a given party if the candidate for that party won at least 50% of the vote share. To clarify, consider the example of Maricopa County (AZ) in 2016. During this election, residents of Maricopa cast 1.6 million votes, with 703K ballots for Hillary Clinton, 747K for Donald J. Trump, and another 118K for 3rd party candidates. Under the plurality standard, Maricopa would be classified as a Republican County whereas under the majority standard, the county would be considered as having supported neither party. As this example makes clear, the absolute majority standard is stricter in its classification threshold. For a detailed discussion of the relative merits of both classification systems and their comparison, see Appendix B.1 Classification Based upon Simple Majority vs Absolute Majority.

Continuous Measures of Partisanship In terms of continuous measures, the most natural approach to summarizing the partisanship of a county in a given election would be to examine the proportion of voters that supported a Democratic or Republican candidate for President i.e. Party Vote Share. Another alternative represents vote share not in terms of party, but in terms of margin of victory. Both are discussed in the sections below.

Party Vote Share Party vote share can be constructed either by examining the Democratic or Republican Vote Share (DVS; RVS hereinafter). For the current sample, Democratic Vote Share is my preferred measure, since it maximizes intra-election consistency.

$$\textit{Democratic Vote Share} = \frac{\textit{Votes for Democrat Presidential Candidate}}{\textit{Total Votes Cast in County}}$$

The Democratic Vote Share (DVS) produces a more stable intra-election measure than Republican Vote Share (RVS). For example, the correlation in Democratic Vote Share between 2016 and 1992 is 0.54, whereas the correlation in Republican Vote Share between those years drops to 0.43. A similar pattern can be seen for other years in the sample. The additional intra-election stability in Democratic Vote Share derives from the fact that Democratic Presidential candidates have not faced dramatic 3rd party challenges during our sample window (unlike, Republican candidates, whose support was significantly cannibalized by the candidacy of Ross Perot in 1992 and 1996).[1]

The correlation in Democratic Vote Share across all elections can be seen in Table 10.2. Overall, the correlation in vote shares across elections for all counties in the dataset seems very high for adjacent years, but the strength of the association diminishes significantly with time.

[1]For similar reasons, I also did not choose a commonly-used alternative (two-party vote share), which standardize the vote share by the total number of votes cast for the two major parties only (as opposed to all votes). As discussed in Appendix B.1 Classification Based upon Simple Majority vs Absolute Majority, since I include the 1992, 1996, and 2016 elections (all three of which featured significant number of 3rd party votes: 19%, 10%, and 6% respectively), standardizing by the two-party-vote-share produces artificial discontinuities for measures of partisanship of a county across the 7 elections in our sample.

Table 10.2: Correlation in Dem.Vote Share Across Elections

	1	2	3	4	5	6	*M*	*SD*
1. 1992	-						0.40	0.11
2. 1996	.94***	-					0.44	0.11
3. 2000	.87***	.94***	-				0.40	0.12
4. 2004	.79***	.88***	.95***	-			0.39	0.12
5. 2008	.61***	.74***	.84***	.93***	-		0.41	0.14
6. 2012	.60***	.74***	.84***	.92***	.98***	-	0.38	0.15
7. 2016	.54***	.67***	.77***	.85***	.91***	.95***	0.31	0.15

Note. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Includes all counties in sample for elections between 1992-2016

Margin of Victory (Signed and Unsigned) Since, ceteris paribus, the degree of partisan affiliation with the winning or losing party should be central in determining the degree to which an electoral outcome influences compliance rates, I consider "margin of victory" as a particularly important measure of the "partisanship" of a county in any election year. The most basic version can be written as follows and would range from 0 to 1:

$$\textit{Margin of Victory (M)} = \frac{(\textit{Votes}_{\textit{Most Preferred Candidate}} - \textit{Votes}_{\textit{Runner Up}})}{\textit{Votes}_{\textit{Total Cast by County}}}$$

This measure captures the relative difference between the top two contenders, however, it does not allow us to distinguish between counties that supported the winner of the Presidential election from counties that backed the loser. To address this, I modified the above measure to construct a "signed version" which —in a single variable —captures both (i) the county's relationship to the national outcome; (did the county support the President-elect) as well as (ii) the degree of support/opposition in a single variable:

$$\textit{Signed Margin of Victory} = \frac{(\textit{Votes}\ _{\textit{Winner of Presidential Election}} - \textit{Votes}\ _{\textit{Runner Up}})}{\textit{Votes}\ _{\textit{Total Cast by County}}}$$

This measure ranges from +1 to -1. A Signed Margin of Victory of +1 would indicate that the county cast 100% of their ballots in favor of the winner of the presidential election. Conversely, a Signed Margin of Victory of -1 would indicate that 100% of county ballots were cast in favor of the losing candidate. In practice, the Signed Margin of Victory ranged from -0.6 to +0.7 in the 1990s and -0.9 to +0.9 in 2012 and 2016. The increase in margin of victory is the clearest sign of an increasingly polarized nation and supports the hypothesis that any election-based effects should increase over time.

10.2.2 Aggregating Partisanship Status Across Elections

Based upon the theory, I would predict that each county's response to an election outcome is likely to depend upon the county's party affiliation and the extremity of partisanship. These factors have significantly changed over the 27 years in the sample (often for a variety of reasons including: general increase in partisanship, migration, changes in demographic factors, and changes in the political stances of the two parties). This makes the aggregation of voting data across the sample very difficult. For example, as partisanship has increased, an aggregate measure (e.g. using average democratic vote share) would over-estimate partisanship earlier in the sample and underestimate it later in the sample.

Despite these concerns, for the purposes of simplifying data visualization and statistical analyses, it is helpful to assign each county a single, aggregated measure of partisanship based upon their voting patterns across the 7 elections. As before, the aggregated partisanship status can either be binary (Dem. / Rep.) or continuous. The approach to constructing both types of measures is described below.

Continuous Aggregate Measures of Partisanship

To aggregate across all 7 elections using a continuous measure of partisanship (e.g. Democratic Vote Share) for each county-election year, there are 3 natural choices: average vote share across the 7 elections, the median vote share, or the vote share for a reference year (here, 2004 —the midpoint of the 7 elections). For the majority of counties, the choice between these 3 methods does not result in a significant difference.

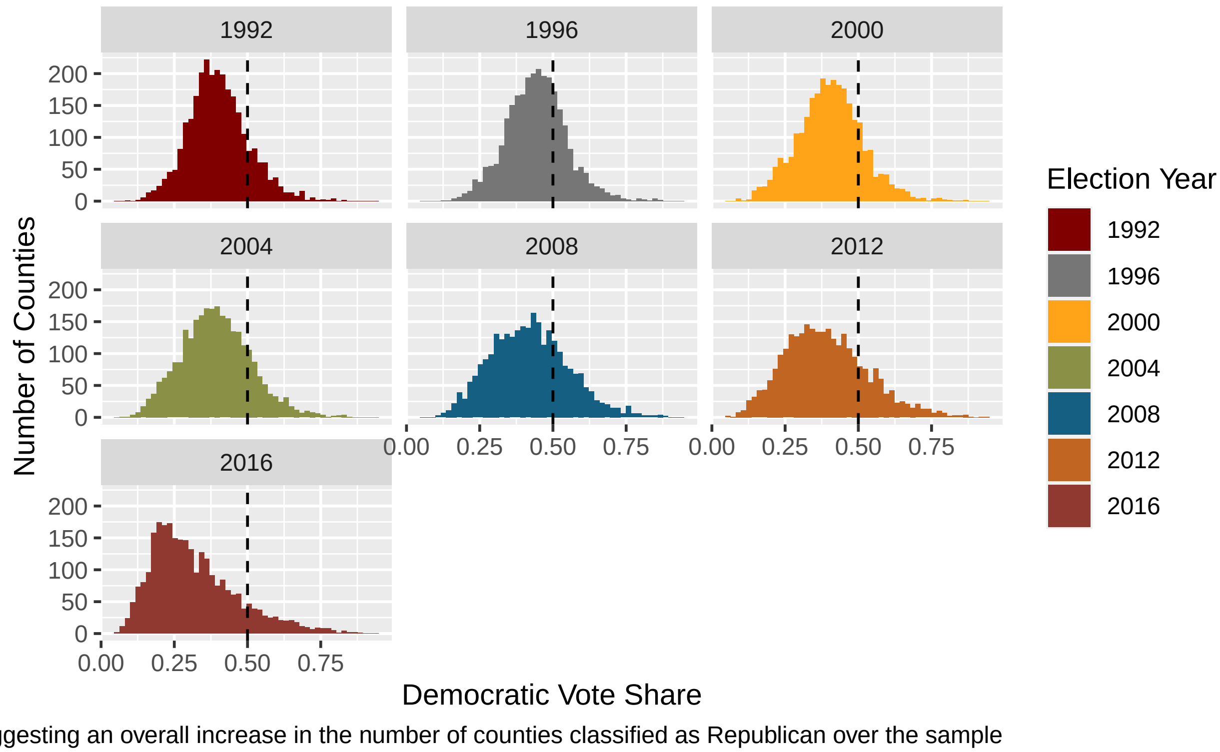


Figure 10.1: Distribution of Democratic Vote Share Across 7 Elections (1992-2016)

For 99% of the counties, the difference between the average and the median is less than 5%; for 87% of the counties, it is less than 2.5%; and for 51% of the counties, the difference is less than 1%. When comparing the median vote share to the vote share received in the reference year (2004), a similar pattern emerged: for 97% of the counties, the difference between the Median Democratic Vote Share and the 2004 Democratic Vote Share is less than 5%; for 82% of the counties, it is less than 2.5%; and for 60% of counties, the difference is less than 1%. I choose to use the median simply because the overall distribution is less skewed and for any given county, unusual voting patterns in 1992 or 1996 are less likely to distort the aggregate measure. The Democratic Vote Share for all elections is presented in Figure

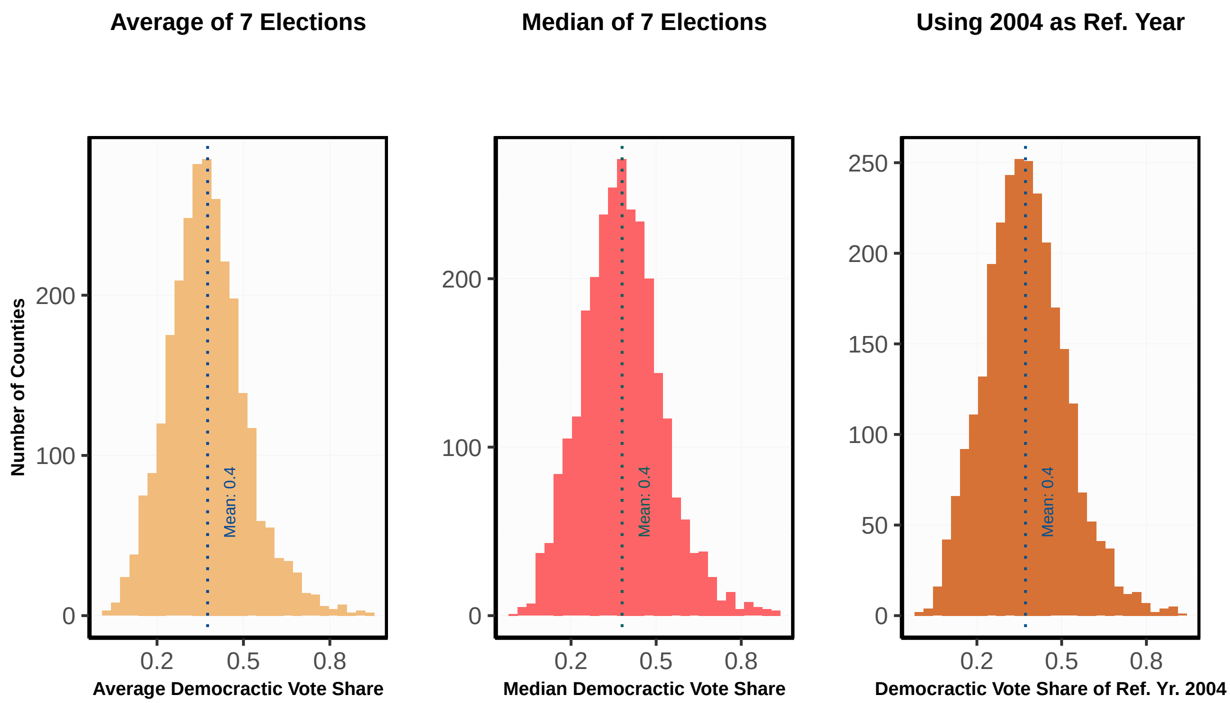


Figure 10.2: Comparing Three Different Approaches to Aggregating Voting Data across All 7 Elections (1992-2016)

10.1. The distributions of aggregated Democratic Vote Share using the three approaches: the median, average, and 2004 reference year are shown in Figure 10.2.

Binary Aggregate Measures of Partisanship

There are multiple ways to construct a binary aggregate measure of partisanship at the county level. I used two broad approaches. First, I used count ratios to classify counties on the basis of the election-level binary measures described above in Section 10.2.1 Binary Measures of Partisanship. Second, I used cutoff thresholds to classify counties on the basis of the aggregated county-level continuous measures described above in Section 10.2.1 Continuous Aggregate Measures of Partisanship. Both approaches are described below in greater detail.

Some approaches considered here are: (i) select all counties where over 50% of the voting population supported a given party for all 7 elections (or, for less stringent standard, for the majority of elections i.e. 4/7 elections); (ii) select all counties where a plurality supported a given party for all 7 elections (or, as above, for a less stringent standard, 4/7 elections);

Perfect Loyalty (Party Vote Share > 50% For All 7 Elections) The perfect method for classifying counties by partisanship would select counties where over 50% (an absolute majority) of voters supported either the Democratic or the Republican candidate for president for all 7 elections. However, classifying using this stringent standard results in an very small sample with only 166 counties consistently voting with an absolute majority in favor of the Democrats; 275 counties in favor of Republicans; and the remaining 2312 counties not meeting the threshold in favor of either party for all 7 elections. As such, we move towards other classification approaches that trade-off between classification accuracy and sample size.

If we relax the standard to allow counties where they voted for one party for the majority of elections, we see that the sample size increases to a more reasonable subset of the data. Table 10.3 below shows the frequency count for the number of election where over 50% of the votes cast in a given county went towards Democrats or Republicans.

As Table 10.3 makes clear, there are large number of counties where over 50% votes were cast for the Republican party in 5 out 7 elections. This issue was detailed in Appendix B.1 Classification Based upon Simple Majority vs Absolute Majority —where it was shown that the presence of Ross Perot in 1992 and 1996 creates a sharp discontinuity in classification attempts between those two elections and the 5 elections that follow. As such, in the trade-off between classification accuracy and sample size, it may be preferable to use 5/7 elections as the appropriate threshold. Such a threshold would classify 1734 counties as Republican, 366 counties as Democrat, and leave only 653 counties as unclassified.

Consistent Plurality: Using Counts of Simple Majority

Table 10.3: Patterns of Victory for Democrats — Absolute Majority Standard (1992–2016)

	Majority Supported Dems.	Majority Supported Reps.
For 0 Elections	1678	377
1/7 Elections	255	138
2/7 Elections	181	124
3/7 Elections	152	151
4/7 Elections	121	229
5/7 Elections	95	991
6/7 Elections	105	468
All 7 Elections	166	275

Note. In the table above, we examine county voting patterns by examining the level of party loyalty under an absolute majority standard (i.e. number of times over 50% of the county voted for a given party). We can classify counties as 'Perfectly Loyal' Republican counties if — for all elections — an absolute majority (over 50%) voted for Republican Presidential Candidates. Similarly, we can classify counties as 'Perfectly Loyal Democrat' counties if — for all elections — an absolutely majority supported Democratic Presidential candidates in all 7 elections. Unfortunately, such a stringent classification scheme dramatically shrinks the sample size, since causes most counties to fail to be classified in either direction. However, we also see a large number of counties where an absolute majority supported the Republican party in 5 out 7 elections. As the Appendix 'Classification Based upon Simple Majority vs Absolute Majority' makes clear, this can largely be attributed to the presence of uncharacteristically large 3rd party voting in 1992 and 1996. As such, it may be reasonable to relax the standard from perfect loyalty (7/7) to significant loyalty (5/7) to increase the sample of counties that are classified with a partisan identity.

A different approach to balancing the trade-off between classification accuracy and sample size would rely upon the simple majority standard at the election level. Such an approach would classify counties as being loyal to a certain party if that party gained a plurality of votes in all 7 elections (i.e. that party won a simple majority in every election in the sample).

Classifying counties using this standard increases the subset of counties that get classified —with the number of unclassified counties dropping to 1309 under the "Consistent Plurality Standard" standard (compared to 2312 under the "Perfect Loyalty Standard", which required that a majority of the county supported the same party for all 7 elections). However, most of this gain goes towards increasing the Republican sample from 275 under the "Perfect Loyalty Standard" to 1143 under the "Consistent Plurality Standard." The Democrat sample also increases (albeit more modestly) from 166 to 301. Table 10.4 shows the frequency count for the pattern of election outcomes classified under a Simple Majority Standard.

Table 10.4: Patterns of Victory by a Party - Simple Majority Standard (1992-2016)

Pattern of Election Outcome	Number of Counties	Percent of Counties
All 7 Elections: Won By Republicans	1143	42%
6/7 Elections: Won by Republicans	247	9%
5/7 Elections: Won by Republicans	500	18%
4/7 Elections: Won by Republicans	211	8%
4/7 Elections: Won by Democrats	124	5%
5/7 Elections: Won by Democrats	100	4%
6/7 Elections: Won by Democrats	127	5%
All 7 Elections: Won By Democrats	301	11%

Note. In the table above, we examine county voting patterns by examining the level of party loyalty (i.e. number of times a plurality of the county voted for a given party). We classify counties based on their party loyalty, identifying 'High Republican counties' as those that counties that supported Republicans in all 7 elections (42% of counties) and similarly 'High Democrat counties' as those that supported Democrats in all 7 elections (11% of counties). However, such a classification may cause some concern regarding unbalanced samples. In addition, it does not take into 'degree of support.'

Classification: Binning Continuous Measures I also classified counties by aggregating continuous measures of partisanship (as described in the Section on Continuous Aggregate Measures of Partisanship) and then use a threshold value to assign binary classification. The selection of thresholds relies upon conventional cutoffs, namely: (i) median split; (ii) selecting the top and bottom quintile; or, (iii) using 50% (the conventional threshold for absolute majority). In the latter case, if a county's average Democratic Vote Share across 7 elections was greater 50%, I would classify that county as a Democratic county. Conversely, if a county's average Republican Vote Share across 7 elections $> 50\%$, I would classify that county as a Republican county (counties that fail to meet these cutoffs are marked as unclassified). This last approach produces the following distribution of county classifications: Democrat —430 counties; Republican —1839 counties; and, Unclassified —484 counties.

10.3 Tax Data

In an ideal world, we would want to measure tax compliance directly. However, given the importance of maintaining privacy, any IRS data products addressing tax compliance (e.g. outcomes of tax audits) are highly restricted in access. In addition, due to an IRS policy intended to maintain the agency's status as a non-partisan, apolitical arm of the Federal Government, the IRS is not allowed to conduct research that uses political affiliation of tax filers as a variable of interest. This policy also extends to all researchers who are provided privileged access to the anonymized IRS tax data (e.g. see note in Cullen et al., 2018). As such, we rely upon the IRS's publically distributed data products —specifically, their Statistics of Income (SOI) for Individual Income tax.

10.3.1 IRS's Statistics of Income Data at the County Level

The IRS receives 200-254 million filings for each tax year in the sample, with approximately 254 million filings for Tax Year 2017. All filings received in a given year form the IRS Master

File, which serves as the entire population of tax data for any given year. Based upon the Master File, the IRS produces an annual series of data products, collectively known as the Statistics of Income (SOI), that capture the tax-related activities of businesses, non-profits, charitable entities, and individuals. These data are aggregated at the county level and cover all tax returns filed from 1989 onwards. However, due to significant data quality concerns with the 1989 filings, the current work is restricted to the 28 year window from 1990-2017.

The IRS's data release procedure for the county-level SOI is particularly profitable for the current approach. When compiling the annual data at the county-level, the IRS's data do not reflect any changes resulting from a subsequently-filed amendment triggered after an IRS automatic notification; nor does it reflect changes resulting from an IRS audit. These data only reflect the income as it was voluntarily disclosed in the initial filing and is not tainted by any modifications that resulted from enforcement actions taken by the IRS. As such, they provide an accurate snapshot of voluntary compliance at the aggregate level for any given year.

Tax Income Variables Available as Part of SOI County Data

As part of the annual Statistics of Income, the IRS provides data for all tax filings at the county-level for 1990 to 2017. The county level tax data prior to 2010 only contain the following six variables of interest. The data descriptions provided are adapted from the "County Income Data Users Guide and Record Layouts" provided along with the dataset for each year:

1. **Number of returns:** includes a count of all individual taxes filed using the standard tax forms Forms 1040, 1040A, 1040EZ, 1040PC filing in a given year;

2. **Total Number of Exemptions:** reflects the number of individuals covered on the tax returns. This includes the person filing and any other person who they claimed as a dependent.

3. **Adjusted Gross Income:** includes the taxable income from all sources, less the adjustments to income, such as IRS deductions, alimony etc. This is the equivalent of line 33 of Form 1040 for Tax Year 2001.

4. **Salary and Wages:** includes all income from wages, salaries, tips, etc. (line 7 on IRS form 1040).

5. **Taxable Interest:** includes all taxable interest income and the non-taxable interest income (lines 8a and 8b on the IRS form 1040).

6. **Returns on Dividends:** includes taxable distributions of money, stock, or other property, excluding non-taxable distributions (line 9 on IRS form 1040).

It is important to note that the IRS changed its data inclusion policy after 2010. For the first 19 years of the sample (1990-2009), the IRS included only the data for tax returns filed between January and September of a given year. According to the IRS, this accounted for 97% of all returns filed in a year. After 2010, in addition to including an enlarged set of data variables at the county level, the IRS also decided to include all tax returns filed between January and December 31 of the filing year. This change in data inclusion procedures is addressed in Section 10.3.2 Standardizing Using Per-Return Data.

As with the voting data, there is significant heterogeneity across counties for the core tax variables mentioned above. For example, for the first year in the sample, Tax Year 1990, the lowest Adjusted Gross Income (hereinafter, AGI) reported was $4.28 million and the highest AGI reported was $120.44 billion. In 2017, these numbers ranged from $15.72 million to $364.56 billion. The first thing to note here is that the IRS data is reported in unadjusted USD. Thus, to effectively compare across years, we must account for inflation (see Section on Inflation-Adjustment using BLS CPI-Urban).

In addition, some of the heterogeneity in tax variables arises from the difference in population and the difference in number of returns filed. For example, in 1990, the number of returns filed in a county ranged from 227.00 to 3.20 million and in 2017 from 320.00 to 4.70 million. As these numbers make clear, there is significant heterogeneity across counties and the tax variables must be standardized in order to facilitate inter-county and inter-year (intra-county) comparisons.

Two obvious approaches to standardization would be to compute per-capita and per-return versions of the tax variables (see Section on Standardizing Using Per-Return Data for reasons why the latter is the preferred method). In terms of per-return, the range in AGI varies from \$12.64K (Starr County, TX) to \$56.19K (Fairfield County, CT) in 1990 and \$28.16K (Baca County, CO) to \$185.64K (New York County) in 2017.

As we can see, standardizing by number of returns still leaves significant heterogeneity in place. Unfortunately, a similar problem exists even if we were to standardize by the population of each county instead, with the numbers range from \$3.17K to \$24.11K per-capita in 1990 and \$8.49K to \$93.89K per-capita in 2017.

As these number show, despite standardizing by return or by population, there remains incredible levels of heterogeneity in the economic and tax variables across counties. As a result, it is ideal to represent these variables in terms of within-county differences (i.e. either raw annual change or percent annual change) as described in the Section 10.3.2 Representing Variables as Difference and Percent Change Variants of the Original Levels, which can serve to remove the county-level fixed effects for the measures of interest.

10.3.2 Data Transformation of the SOI's Taxable Income Measures

For reasons described above, for the each of the tax measures of interest, the following transformations were applied to facilitate inter-year comparisons, to remove extraneous county-level fixed-effects, and to reduce or eliminate serial-correlation of measures across time.

First, all monetary measures (e.g. reported or estimated income) are inflation adjusted using the BLS Consumer Price Index —Urban (CPI-U) and presented in 2012 real USD.

After adjusting for inflation, as shown below, I standardize tax variables by accounting for the number of returns filed in that county for that tax year (see Section on Standardizing Using Per-Return Data).

$$\textit{Standardized value: } \frac{\textit{(Variable of interest)}_t}{\textit{(Num of returns)}_t}$$

When comparing variables across years, for ratios of income variables (e.g. AGI over PI), I utilize the raw annual change version of the measures; for most other measures, I rely upon a percent annual change (i.e. I scale the annual year-on-year change by the starting year's value). The reasoning behind the choice of such data transformations are presented in greater detail in the sub-sections below.

$$\textit{For income ratios, I use raw annual change: } \Delta = X_t - X_{t-1}$$

$$\textit{For income variables, I use percent annual change: } \Delta_{pct} = \frac{(X_t - X_{t-1})}{X_{t-1}}$$

Inflation-Adjustment using BLS CPI-Urban

There are two primary reasons for inflation adjustment: interpretive and statistical. On the interpretative front, the current sample covers 28 years, during which time inflation significantly changes the effective purchasing power of the dollar: for example, \$100 in 1990 is worth approximately \$188 in 2017, representing a cumulative rate of inflation of 88% over the 28 years covered in the sample. Thus, to facilitate appropriate interpretations, it is important to convert all (nominal) income data into inflation-adjusted (hereinafter, real) income data. In addition, there are important statistical benefits of removing the inflation rate from the measures of tax payment —namely, doing so reduces the autocorrelation among variables from Lag 1–5 to Lag 1–2 for the log-transformed variable. The remaining autocorrelation can be removed by representing the measure as a log-difference or percent change version of itself.

In order to adjust for inflation, I rely upon the Bureau of Labor Statistics' (BLS) Consumer Price Index for Urban consumers (CPI-U) —a commonly used measure of inflation —to produce the annual adjustment factor. The decision to rely upon the CPI-U is also based upon the common use of this index by the IRS, which uses CPI-U inflation-adjusted "real" dollars in any data products that make longitudinal comparisons across multiple years (e.g., see IRS, 2019b).

Standardizing Using Per-Return Data

There are many ways one may consider standardizing the tax variables, each with different implications for interpretability, robustness, sensitivity to different macro and microeconomic contextual trends, changes in tax regimes. They each also differ in their relative advantages and disadvantages in terms of addressing the limitations of the available datasets. For my purposes, there are two natural approaches to standardizing the tax data —both of which diminish the effects of the enormous heterogeneity in county size and thus facilitate cross-county comparisons. The first is to produce a per-capita measure. The second relies upon a per-return measure. Other approaches to standardizing tax variables involve representing them in terms of annual change and percent change, which is discussed in Section 10.3.2 Representing Variables as Difference and Percent Change Variants of the Original Levels. When choosing between per-capita and per-return measures, I argue in favor of the latter for reasons having to do with the way in which the tax burden is distributed.

The vast majority of reported AGI is driven by the top 50% of tax filers, which suggests that the use of AGI brackets to select the subpopulation of tax filers in the upper 50th percentile of reported income will be effective in both isolating the subpopulation that contributes most significantly to the underlying data itself. In doing so, the standardized tax variables (e.g. AGI / return) are more likely to accurately represent the decisions about voluntary compliance —since, choosing between compliance and non-compliance fundamentally depends upon (a) having a tax-burden in the first place; (b) having some choice in the level of income to disclose; (c) having some choice in the amount and types of exemptions and deductions that could be utilized to minimize tax burdens.

Concerns with Standardizing using Number of Returns: The Economic Stimulus Act of 2008 For the county data, the use of number of returns as a base for standardization is compromised by two factors: the 2008 economic stimulus and the change in sampling as a result of 2010 IRS decision to extending sampling from Jan-Sep to Jan-Dec of the filing year. The latter problem should not be of as much concern, since the other tax variables also increase (although, late filers tend to be slightly different —often filing more complicated taxes with higher AGI reported; see Pierce, 2015).[2]

The first problem is more troubling: the economic stimulus of 2008 resulted in a massive spike in the number of tax returns filed, which was a prerequisite to receive the stimulus income. For our purposes, this is particularly troubling since this spike largely represents people who were previously choosing not to file —presumably because they did not meet the required threshold. Normally, single individuals under 65 were required to file taxes in 2007 if they earned over $8,750.[3] However, after the passage of the Economic Stimulus Act of 2008, individuals who earned at least $3,000 of qualifying income could receive a stimulus check if they filed a federal income tax return. Thus, the spike in number of returns filed was a result of filings by previously unrepresented, poorer sections of the American populace (e.g. see, Ramnath & Tong, 2017, who estimate that this represents about 9.1 million workers who

[2]The other reason why the change in data inclusion should not be a major concern is that it occurs exactly 2 years prior to the election (2012), thus allowing us to maintain the same data inclusion for the 2010-2014 election window. This issue is discussed in greater detail in the Analyses.

[3]The filing requirements are more complicated, but this figure represents the lower threshold of filing requirements. For single individuals over 65, this threshold rises to $10,050. For head of household, the threshold was $11,250 for under 65 and $12,550 for over 65. For more details, please see IRS publication "Individual Income Tax Returns - Section 1 - Introduction and Change in Law" which can be accessed at https://www.irs.gov/pub/irs-soi/07insec1.pdf.

were now eligible to receive a stimulus check, but would not have normally met the threshold requiring them to file as they had no income-tax liability). The inclusion of these new filer in the data pool increased the number of returns without a commensurate increase in the other income tax variables. Thus, standardizing AGI by number of returns in 2007 results in an artificially low figure for measures of taxable income in a county.

Thus, the first problem cannot be resolved without recovering the number of filers that would have filed if 2007 had not taken place. In order to address this, we take the following approach: treat number of tax returns data for tax year 2007 as missing for all counties and then impute the data by averaging the number of returns from the two nearest years (2006 and 2008). To assess how well this imputation strategy does, I applied the same approach to data from 1992-2006 and compared the imputed and reported values. In general, the simple averaging strategy did remarkably well: across all county-years in the validation sample (~65K observations): the mean absolute percent error was only 1.19% for all previous years. Given the low overall average error using such a strategy, the reported number of returns for 2007 was replaced with the imputed values and these imputed values were used to construct the standardized measures for that year.

Representing Variables as Difference and Percent Change Variants of the Original Levels

As was mentioned earlier, income variables were adjusted for inflation and standardized by the number of returns. Despite these steps, we see that —due to the heterogeneity across counties and the persistent characteristics of each county's data across the sample — even these adjusted and standardized variables show a significant degree of auto-correlation (i.e. correlation between two or more years of a given county's data). To remove these auto-correlations and reduce some of the county-specific fixed effects, one effective and intuitive approach involves the first-difference of the variable of interest such that it is represented as either an annual change variable (which I use when dealing with log-transformed variable or income ratios) or as a percentage change variable (which I use when dealing with the inflation-adjusted "real" per-return variables).

As such, the transformation to percent change allows me to avoid violations of the assumption of stationarity and assumptions of independence that arise in cases with significant serial-correlation / time-dependence. For example, for inflation-adjusted AGI per Return, the simple log-transformed variable shows evidence of auto-correlation for lag 1,2, and 3. However, the percent change variable (i.e. annual change in real AGI / return) shows no statistical evidence of time dependencies with only 0.7% of counties showing an inter-year correlation outside the bounds expected by chance at a 95% standard.[4] In addition, for some variables like the compliance measure AGI over PI (described in the Section below on Constructing Proxy Measures of Compliance, representing the measure in terms of annual change is also crucial to facilitate appropriate interpretation and inter-county comparisons. And, as with the raw income variable, it also reduces time-dependencies —with only weak evidence of auto-correlation (lag 1) and only 3.5% counties showing inter-year correlation outside of bounds expected by chance.

In the current datasets, transformation to percentage change representations also corrected the significant skew in distribution that is common among income variables without having to rely solely on log-transformations. Finally, on a practical note, when examining and comparing means for variables with repeated measures, any significant inter-year correlation requires the construction of adjusted standard errors —a process that becomes increasingly hard if the correlation extends across multiple lags, (e.g. Cousineau, 2017; Morey, 2008). Thus, representing variables in terms of mean percent change or mean annual change can significantly reduce the statistical corrections required.

[4]The annual change in log real AGI / return also shows the same lack of auto-correlation, but is not preferred due to the less intuitive nature of representing variables in terms of log-differences.

10.3.3 Constructing Proxy Measures of Compliance

Since I lack a direct measure of tax compliance, I rely upon background knowledge of tax evasion trends as well as parallel sources of information (e.g. income estimates from the BEA) to construct proxy measures that may track the incidence of tax evasion —or, more specifically, the changes in the rate of tax evasion. One approach is to use independent estimates of income from the BEA, as discussed below.

AGI over PI: Constructing Compliance Measures using the BEAs Income Estimates

There is significant value that can be gained by matching data across different data sources — especially when the alternate data sources give you a chance to "peek behind the curtain" and estimate a shrouded attribute. In the current case, the Bureau of Economic Analysis (BEA) produces an annual estimate of Personal Income (PI) at the county level. They describe personal income as the "income that is received by all persons from all sources", consisting of "income that persons receive in return for their provision of labor, land, and capital used in current production as well as other income, such as personal current transfer receipts." (see definition of Personal Income in Glossary: Definitions of Common Terms). These income estimates are the most comprehensive and technically-accurate US Government figures on personal income at the county level. Using the BEA's Personal Income measures, I construct the following variable, hereinafter referred to as AGI over PI:

$$AGI\ over\ PI = \frac{Adjusted\ Gross\ Income_{IRS\ Data}}{Personal\ Income_{BEA\ Data}}$$

Since the definition of Personal Income (PI) for the BEA is distinct from the definition of Adjusted Gross Income (AGI) for the IRS, the two measures can differ for a variety of reasons, including many legitimate reasons as well as reasons that could arise from tax-evasion. Thus, even if all income was transparently disclosed to both agencies, a ratio of these income concepts would not necessarily equal 1. In addition, since counties differ in terms of inhabitants and the types of economic activities that they pursue, there should be natural heterogeneity in the way these two concepts diverge for a given county. However, assuming: (i) no change in the definition of these two incomes concept; and, (ii) no dramatic change in the economic profile of a county, the year-on-year change in the ratio of AGI / PI should be approximately 0.

Thus, if a significant change in this ratio occurs, we can conclude either: (i) the economic profile of the county dramatically changed; (ii) the definition of AGI substantially changed; or (iii) the definition / estimation accuracy of PI substantially changed; or, (iv) the disclosure patterns of income by residents to the IRS substantially changed. The definition of PI has not changed in the years covered by the current sample. There is also no *a priori* reason to assume that the estimation accuracy of personal income has changed either. Even if there was a year-on-year variation in estimation accuracy, it is not reasonable to believe that the accuracy of the BEA's estimate of Personal Income differentially change for a county based upon the outcome of the Presidential Election. Similarly, it is not reasonable to believe that changes in the definition of AGI would impact winners and losers of elections differentially across the entire sample. Thus, significant and differential changes in the ratio of AGI/PI should lead us to conclude that either: (i) the economic profile of the county dramatically changed; or (ii) the disclosure patterns of income to the IRS by residents of a county substantially changed.

On its own, annual changes in the ratio AGI/PI would be difficult to interpret —since we would not be able to know whether there had been a significant change in the economic profile of a county. However, since Presidential elections occur late in the Year, we can be relatively certain that election outcomes cannot retroactively and selectively change the economic conditions for counties that supported the winner or loser of an election. Thus, if we see systematic changes in the ratio of the AGI/PI that *diverge based on whether a county's preferred candidate won or lost*, we can reasonably infer that there was a change in the pattern of income disclosure.

Finally, although it is possible to hold some reservations about interpreting the raw AGI-over-PI ratios as measures of compliance, comparisons of differential changes in these ratios for election winners and election losers can be more reasonably interpreted as tracking changes in tax compliance. In addition, notwithstanding any concerns in the use of the AGI-over-PI ratio, it is reassuring that this ratio of AGI to PI has also been used by internal IRS documents when modeling individual income tax compliance. The use of this ratio by the IRS —which is arguably the best authority on tax compliance in the US —lends this measure a high degree of credibility.

It should be noted that —unlike other measures —AGI over PI is computed using nominal (i.e. unadjusted) income data from both the BEA and the IRS for the given year. Since inflation adjustment would be done to both the numerator and denominator of the ratio, assuming a consistent adjustment factor, the ratio should independent of inflation adjustments. This has the advantage of removing the need to select an inflation factor or to adopt the assumptions necessary in computing a common price index that can apply equally well to all counties across all years.

CHAPTER 11

GRAPHICAL REPRESENTATIONS OF HYPOTHESIZED EFFECTS

Since graphical analyses of the data are particularly central to my approach herein, it may be helpful to illustrate the hypotheses described above by showing graphical representations of the predicted effects under the legitimacy hypothesis. The purpose of doing so is to provide of comparison with the graphs of the actual data. In order to construct the following graphs, I began by sampling the 1990 Adjusted Gross Income for all counties in the sample. In order to highlight the predicted effect, for each subsequent year, I applied a consistent 3% growth rate to produce a simulated tax dataset for all counties for 27 years. The counties were then assigned a partisan classification, such that counties in the top quartile of Avg. Democratic Vote Share were classified as "Democratic Counties" and those in the bottom quartile were classified as "Republican Counties." The remaining counties (in the 2nd and 3rd quartiles) were marked as "Unclassified."

In line with the Legitimacy Hypothesis, the simulated data was constructed such that counties with a partisan affiliation responded to the election outcome by either increasing or decreasing their income disclosure (as captured by the AGI). Counties that supported the winner of the Presidential Election increased their AGI by 8% and counties that supported the loser of the election decreased their AGI by 12%. The simulated effect size of 8-12% is exaggerated and the predicted effect is likely to be much smaller (on the order of 1-3%). The difference in the effect of victory and loss represents the belief that (a) people have a natural tendency to want to minimize income disclosure; (b) in line with the win-loss asymmetry, I believe that losing an election feels worse than winning one. Counties that were marked as "unclassified" show no effect of the election and thus represent the null hypothesis.

11.1 AGI per Return

For the purposes of this illustration, all predicted effects are shown in terms of AGI per return or annual percent change in AGI per return. These graphs could have also been constructed using one of the proxy measures of compliance (e.g. AGI over PI). However, under the simplifying assumptions used to simulate the data (namely, consistent 5% growth rate and no exogenous events), there is no difference in the shape and character of the predicted effects —whether represented in terms of compliance or in terms of changes in AGI / return, we see the same pattern of findings (for reference, see Appendix B.2 Supplementary Hypotheses Graphs for the same graphs constructed using a compliance measure like AGI over PI).

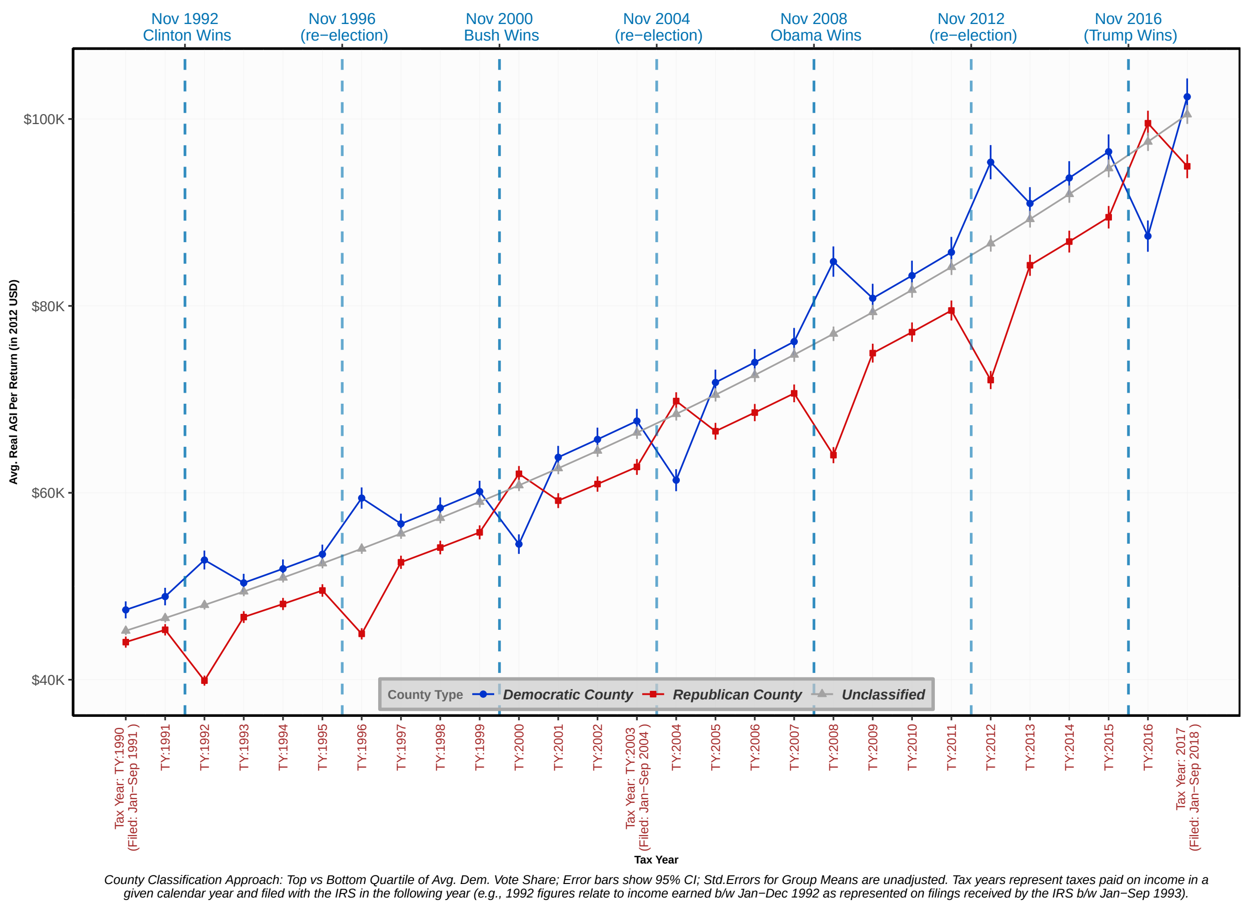


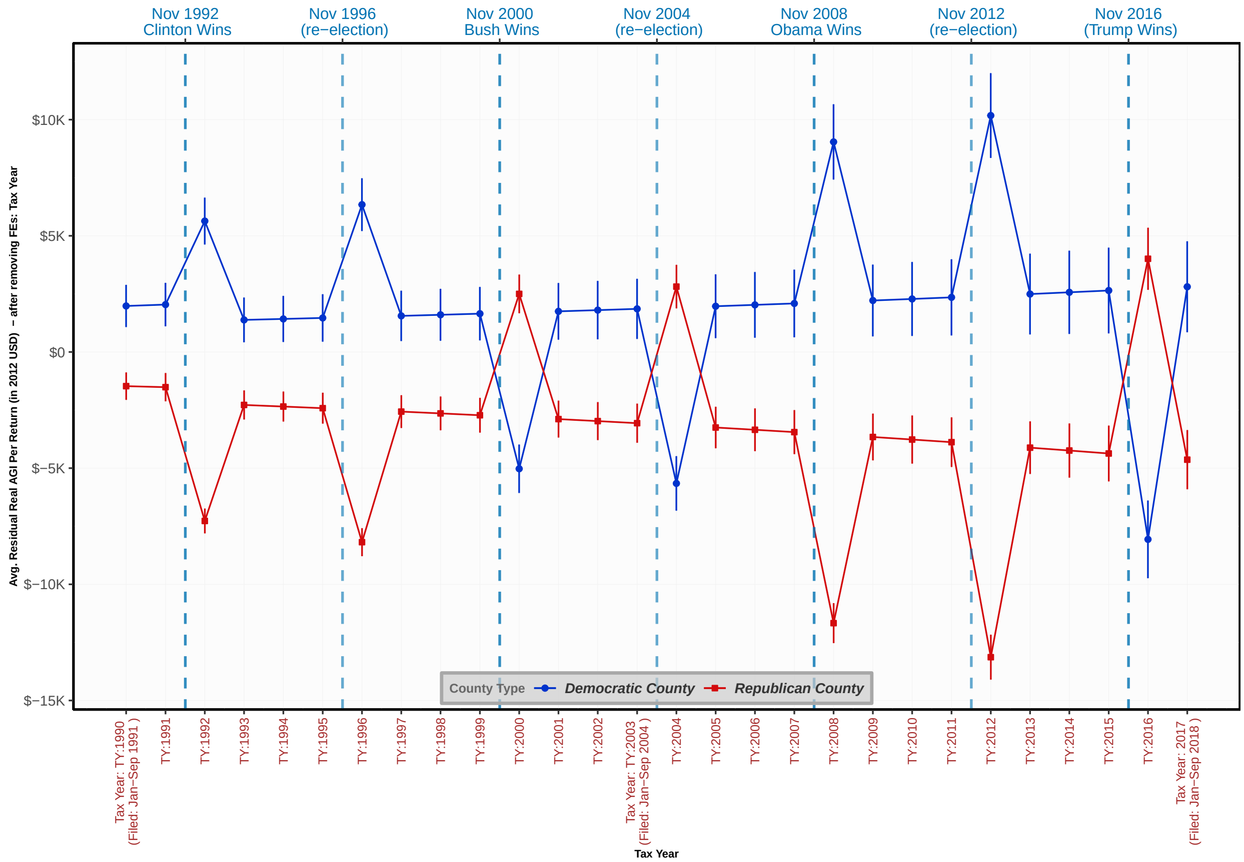
Figure 11.1: Legitimacy Hypothesis: Effects of Elections on AGI per Return over Time

11.1.1 Simulation of Legitimacy Hypothesis Under Consistent Growth Rate

To begin, we start with a graph of the simulated data across the entire 28 year sample. To understand the effects shown in Figure 11.1, let us take the 1992 Clinton election as an example: before the election, in 1991, the Democratic counties —shown in blue —had an average AGI per return of approximately \$48.9K; Republican counties —shown in red —had an average of AGI / return of approximately \$45.3K; and unclassified counties —shown in grey —had an average AGI per return of approximately \$46.6K.

After the 1992 election of Bill Clinton, as per the Legitimacy Hypothesis, we see the Democratic Counties increase disclosure of income such that their average AGI per return increases to \$52.8K. Under the null hypothesis, if they had not responded to the election, they would have only increased by 3% to \$50.4K per return. Thus, under the simulation, the effects of winning the election would be estimated to be: \$2.4K / return. On the Republican side, as per the Legitimacy Hypothesis, we see a decrease in income disclosure such that the average AGI per return decreases to \$39.9K. Under the null hypothesis, it should have increased by 3% to \$46.7K per return. Thus, under the simulation, the effects of losing the election would be estimated to be: \$-6.8K / return. For every other election in the sample, we see the identical pattern of predicted effects. Although the effect sizes presented here stay constant across time, in reality, we should expect the pattern of effects to increase in size as a function of the growing increase in political partisanship over time.

11.1.2 Simulation - Controlling for Year Fixed Effects



County Classification Approach: Top vs Bottom Quartile of Avg. Dem. Vote Share; Error bars show 95% CI; Std.Errors for Group Means are unadjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure 11.2: Legitimacy Hypothesis: Effects of Elections on AGI per Return over Time (After Removing Year Fixed Effects)

Another way of representing the data showing in Figure 11.1 is to control for the effect of time by removing the Year fixed effect (i.e. removing the idiosyncratic effects of time on AGI per return that are independent of the partisanship of a county). In the current simulated dataset, the idiosyncratic effect of time is constant —it is the 3% annual growth rate. In the real dataset, these effect of time may be more complicated, but we can still use the same procedure to exclude the effect of systemic, nation-level trends that are common to all U.S. counties. In the current graph in Figure 11.2, each value is the average residual AGI per Return after removing Year fixed effects. One way to interpret these values is to see them as the group distance from the overall mean. Thus, in 1991, the blue line representing Democratic counties shows a value of ~$2K, which indicates that the Average

AGI per return for Democratic Counties was ~2K greater than the average AGI per return for all U.S. Counties. Similarly, the red line showing a value of $ -$1.5K $ in 1991 indicates that the Average AGI per return for Republican counties was ~1.5K less than the overall national average. The use of year fixed effects is particularly helpful in the current analyses in order to control for extraneous economic trends that may occlude the election effects of interest.

11.1.3 Simulation - Varying the Time Dynamics of Election Effects

When considering the Legitimacy Hypothesis, another factor that may be worth considering is the time dynamics of the hypothesized effects. To facilitate the illustration of time dynamics, in the subsequent graphs, I have removed any effects resulting from the constant 3% annual growth rate. Instead, I vary the number of years that the election effect persists.

In the first graph, Figure 11.3 (a), we see the Legitimacy Effect diminish after just 1 year and income disclosure returns back to pre-election baseline for both parties. Let us again consider the example of the 1992 Clinton Election, where the Republican counties' AGI per return goes from a baseline value of $44.0K per return for Tax Year 1991 to $38.7K per return for Tax Year 1992 (filed in April 1993, i.e. filed following the election loss in Nov 1992) before returning back to the baseline of $40.5K for Tax Years 1993-1995 (filed Apr 1994-1996). As the graph shows, under this model, the departure from baseline as a result of the election only lasts for the year immediately following the election i.e. there is only an immediate and temporary effect of the election on tax compliance.

However, such a model may seem unlikely since there is some stickiness to judgments like perceived legitimacy and behaviors like tax compliance. To capture this stickiness and tendency to anchor on the recent past, in Figure 11.3 (b), the effect of the election systematically diminishes over 3 years before returning back to baseline. To take the same example of the 1992 Clinton Election, the Republican counties' AGI/return goes from the $44.0K baseline value for Tax Year 1991 to \$38.7K following the election loss in 1992 before slowly starting to return back to the baseline with a value of \$40.5K for Tax Year 1993; \$42.3K for Tax Year 1994; and, a complete return back to baseline of \$44.0K by filings for Tax Year 1995.

There are of course multiple other time dynamics that could be considered. One salient possibility is that the effect of the election peaks during the first year of the presidency (rather than peaking right after the election) e.g. in the 1992 Clinton case, the effect is detectable for taxes paid in April 1993 on earnings for Tax Year 1992, but is stronger for taxes paid in April 1994 on earnings for Tax Year 1993 (i.e. for earning under the first year of the Clinton Presidency). There are multiple reasons why such a time course could be feasible. For one, the segment of the population with the greatest latitude in income disclosure (small business owners filing Schedule C tax returns) are required to file quarterly, which means, not only are their 1993 filings constrained to some small degree by the estimated Q1-Q3 quarterly filings which were made prior to the November 1992 election. In addition, when they file in April 1993, they must submit new quarterly tax estimates. Thus, their filings in April 1993 impact both the taxes paid for Jan-Dec 1992 and these decision propagate into their estimated filings for the next year as well. Under such a situation, just mechanistically, we would expect the effect on income disclosure to be larger for Tax Year 1993. Figure 11.3 (c) shows exactly such a time dynamic. The AGI / return for Republican counties decreases from \$44.0K for Tax Year 1991 to \$40.5K for Tax Year 1992 before dropping to the lowest point of \$38.7K for Tax Year 1993 and then slowly returning back to baseline with \$42.3K declared for Tax Year 1994 and a full return to baseline of \$44.0K for Tax Year 1995.

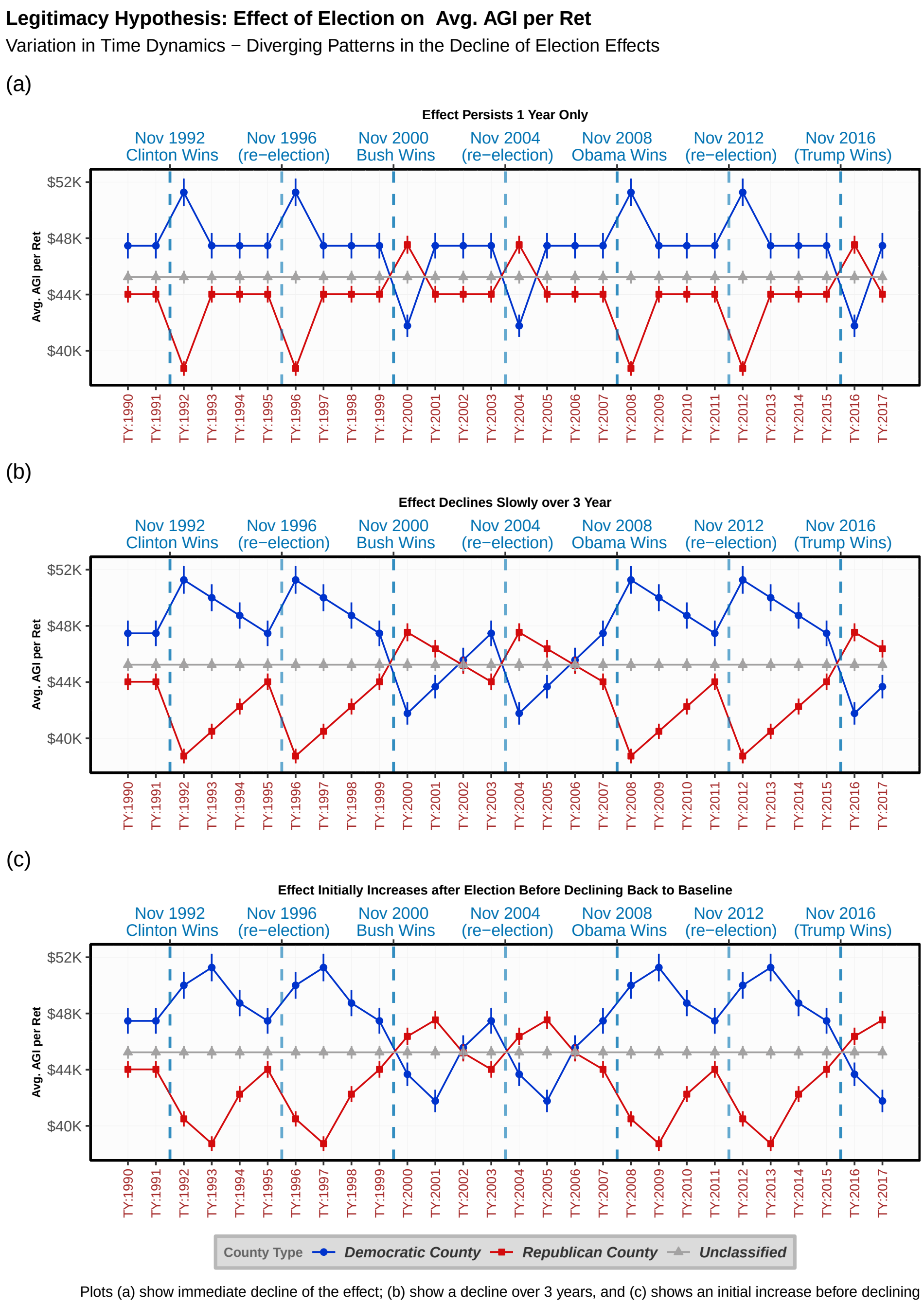


Figure 11.3: Legitimacy Hypothesis: Effects of Elections on AGI per Return over the Time Sample —Election Effects Shown with Varying Time Dynamics

11.2 Percent Change AGI per Return

11.2.1 Varying Time Dynamics - Election Effects Persist for 1-3 Years

As was detailed in Section 10.3.2 Representing Variables as Difference and Percent Change Variants of the Original Levels, for a variety of reasons, the analyses of election effects on tax compliance are best supported by an examination of year-on-year changes in tax behavior. Since the previously shown hypotheses dynamics are not trivial to translate into annual change variables, Figure 11.4 shows the same predicted effects shown in Figure 11.3 but represented in terms of Annual Percentage Change in AGI per return.

To facilitate comprehension of Figure 11.4, let us again consider the example of the Clinton 1992 election. First thing to notice is that —since we are lacking data for Tax Year 1989 —the first data point for a percentage change variable begins with Tax Year 1991 (which captures the change from Tax Year 1990 to Tax Year 1991). In Figure 11.4 (a), we see the Legitimacy Hypothesis represented with the assumption that the effect diminishes immediately following the election. To begin, the annual percentage change for the Democratic and Republican Counties is 0.0% for Tax Year 1991, which is the baseline year-on-year change by construction (note: the 3% annual growth shown in Figures 11.1 and 11.2 has been removed for ease of illustration). Immediately following the election, we see the effect of electoral loss represented by a (12.0%) change in AGI / return for Republican Counties. The return to baseline AGI per return in the year following the election (i.e. Tax Year 1993) translates into an 13.6% change in AGI / return (i.e. Tax Year 1993), followed by a return to the baseline of 0.0% annual change starting in Tax Year 1994. Figure 11.4 (b) represents the Legitimacy Hypothesis under the assumption that a return to baseline in income disclosure can take up to 3 years after the initial shock. In terms of annual percentage change, Republican Counties go from the 0.0% baseline in Tax Year 1991 to (12.0%) in Tax Year 1992 to 4.5% from Tax Years 1993-1995, which represents the steady return to baseline at the rate of 4.5% per year.

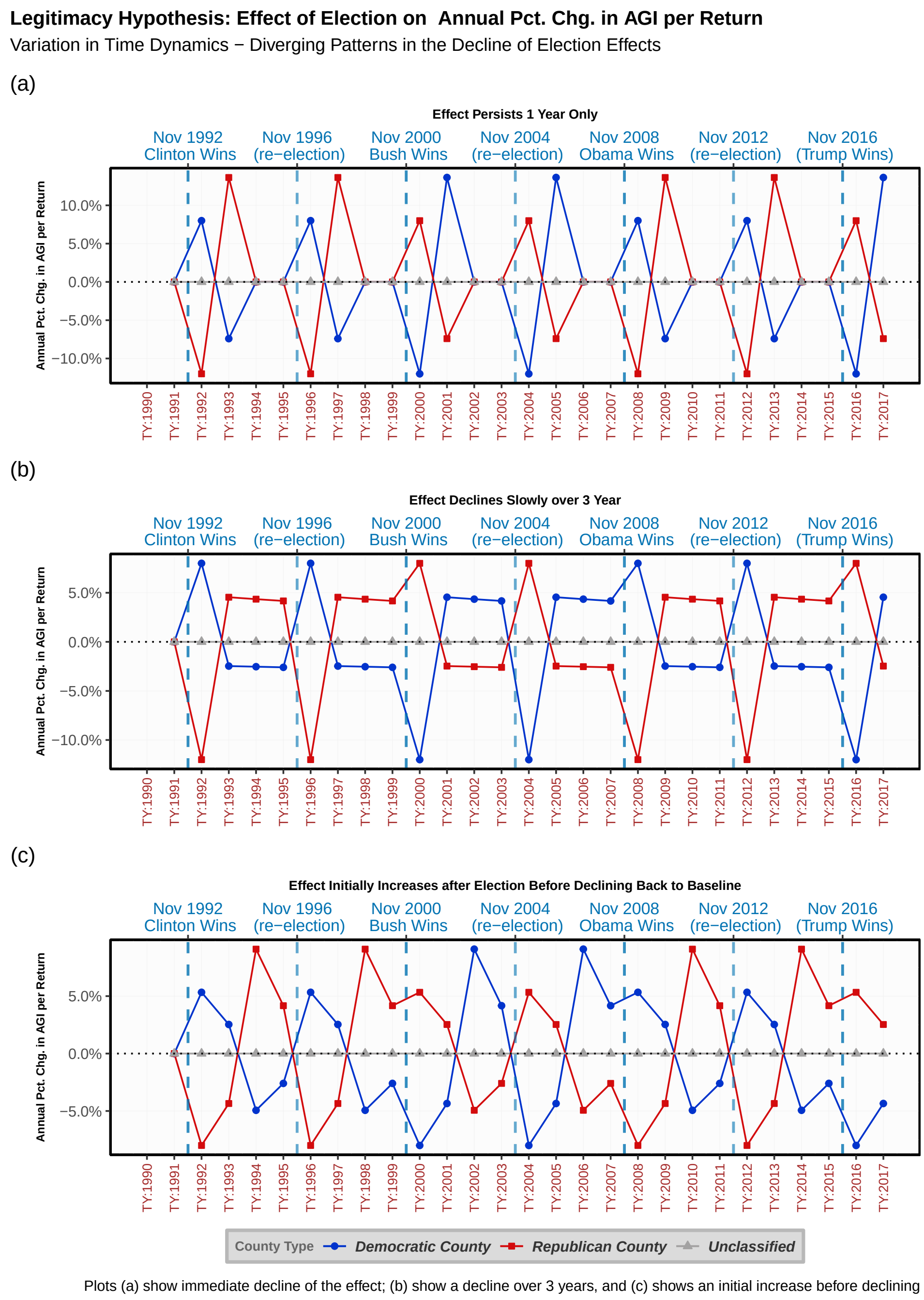


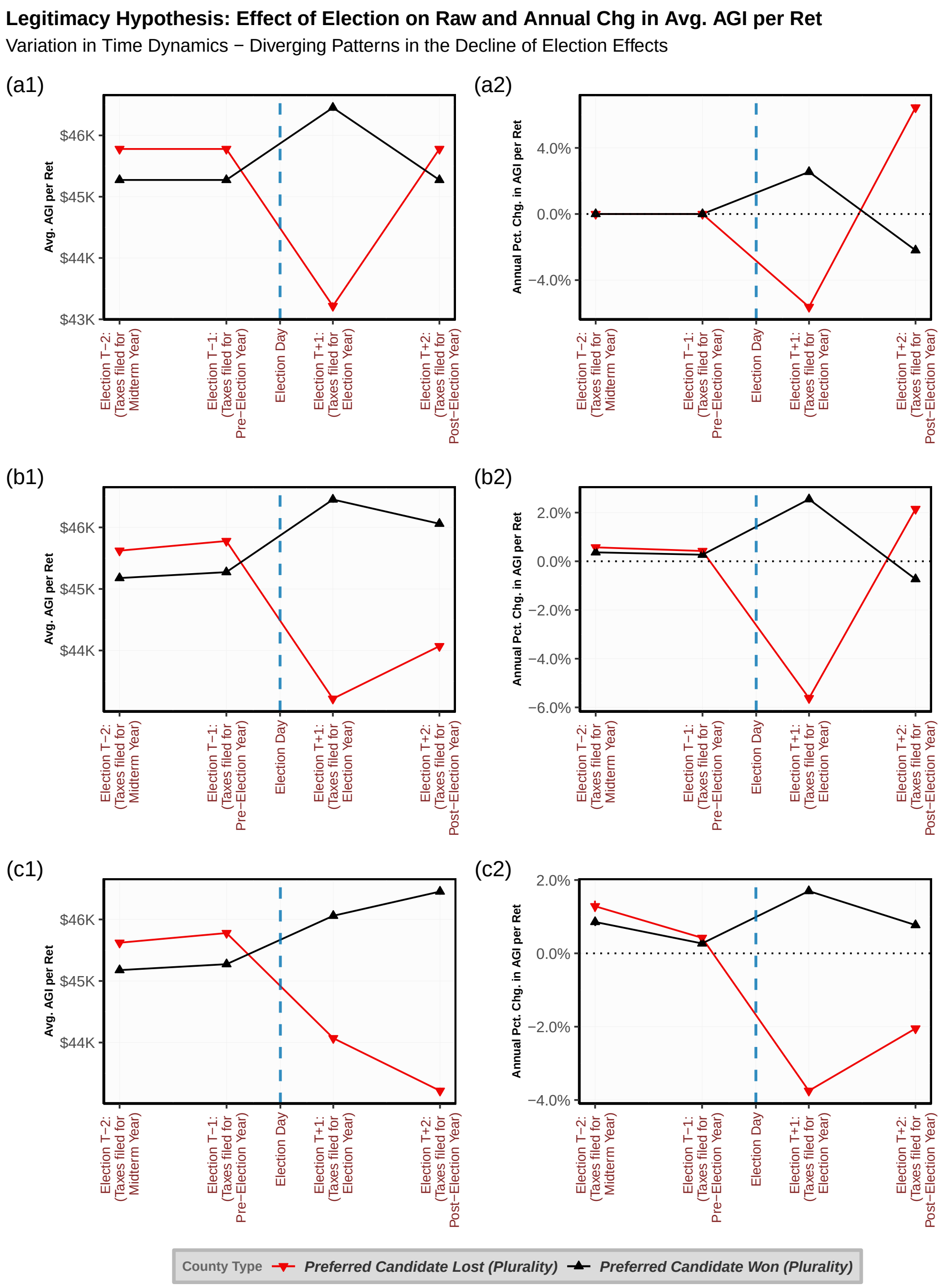
Figure 11.4: Legitimacy Hypothesis: Effects of Elections on Annual Percent Change in AGI per Return over the Time Sample —Election Effects Shown with Varying Time Dynamics

Finally, as before, Figure 11.4 (c) represents the Legitimacy Hypothesis under the assumption that the election effect continues to increase for the year following the election before returning to baseline. In terms of of annual percentage change, we see Republican Counties go from the 0.0% baseline in Tax Year 1991 to (8.0%) in Tax Year 1992 followed by an additional (4.3%) change in Tax Year 1993 (which is when the decline in raw income disclosure peaks under this model). The diminishing of the election effect in Tax Year 1994 is represented by the 9.1% annual change in AGI / return for the same year —a process that continues with a 4.2% change for Tax Year 1995.

11.3 Effect of Election on Raw and Pct Chg in AGI per Return - Collapsed Across Winners and Losers

In the next set of graphs, we re-examine the predicted effects of the Legitimacy Hypothesis by collapsing across election cycles and reviewing trends in terms of the two years before and after an election. In these set of graphs, counties are no longer coded by the party affiliation, and instead were recoded to reflect their win or loss status for any given election. Counties were coded as winners or losers based upon whether their preferred presidential candidate won the White House.

In Figure 11.5, I present six plots broken into three rows and two columns. In each of the plots, the red line (marked by downward pointing arrows) represents the counties that supported the losing candidate and the black line (marked by upward pointing arrows) represents the counties that supported the winning candidate. Each line shows four points representing the two years prior to the election and the two years after the election. The election is marked by the blue, dashed vertical line which bisects each plot. On the x-axis, each point is labeled as either T-2; T-1, T+1, and T+2, which represent their position relative to the election. The data shown at Election T-1 represents the taxes due in April of an election year for the earnings of the previous tax year and, similarly, Election T+1 represents taxes due in the April immediately following the Presidential Election. If we take the 2000 election as an example, T-1 represents taxes paid in April 2000 for Tax Year 1999 and T+1 represents taxes paid in April 2001 (immediately after the election and inauguration of Bush Jr) for Tax Year 2000.



Plots (a) show immediate decline of the effect; (b) show a decline over 2 years, and (c) shows an initial increase before declining

Figure 11.5: Legitimacy Hypothesis: Effect of Election on Raw and Annual Chg in AGI per Return by Winners and Losers of Elections —Election Effects Shown with Varying Time Dynamics

The first column of plots —Figures 11.5 (a1-c1) —show the effects of election outcomes in terms of average AGI per return and the second column of plots 11.5 (a2-c2) —show the effects in terms of annual percentage change in AGI per return. These are the two variables that were previously considered in —Figures 11.3 and 11.4 respectively.

Each of the rows shows the effects of victory or loss under the three different time dynamics models that we have previously considered. The first row, Figure 11.5(a) shows the Legitimacy Hypothesis under the modeling assumption that the election effect persists for only one year. The second row, Figure 11.5(b) shows the Legitimacy Hypothesis under the modeling assumption that the election effect slowly dissipates over the course of 3 years. And, the final row, Figure 11.5(c) shows the Legitimacy Hypothesis under the modeling assumption that the election effect initially increases after the election before dissipating over the course of next few years. These three rows represent the same variation in time dynamics shown in the previous two figures, Figures 11.3 and 11.4.

To begin, let us start by examining Figure 11.5(a1–c1) which represent trends in terms of Avg AGI per Return. The first thing to notice is that —given the absence of any time fixed effects[1], the lines for winning and losing counties approach the election marker in parallel before diverging after the election. This pattern remains constant across all three variations in time dynamics. In fact, surprisingly, when all extraneous noise is removed, the basic shape of the predicted effect looks remarkably consistent whether we examine AGI per return (a1–c1) or Annual Percentage Change in AGI per Return (a2–c2). As was mentioned earlier, this consistency is also found when we examine the same patterns represented in terms of AGI over PI or Annual Change in AGI over PI (see Section B.2 Supplementary Hypotheses Graphs). In some sense, this consistency should be reassuring that (a) idiosyncratic differences in the time dynamics of the predicted effects may not be

[1]Since there is no annual growth rate or any exogenous shocks in the simulated data, there graphs do not remove time fixed effects. However, in the graphs of real data, such trends will normally be presented with fixed effect removed.

particularly influential once we aggregate across all 7 election cycles; (b) choice of dependent variable does not impact the main pattern and thus choices can be made solely by taking into account the ability of the variable to minimize extraneous noise from fixed effects (time or regional) and to minimize time dependencies in a manner that best supports inferential statistics. This also suggests that if —proxy variable like AGI over PI and AGI per return / Median Household Income —that are constructed from different data sources[2] are tracking the variable of interest, namely, tax compliance, we should expect them to produce similar looking patterns once data have been aggregated across election cycles.

[2] AGI over PI uses BEA data; AGI per Return / Median Houshold Income uses Census survey data

CHAPTER 12

GRAPHICAL ANALYSES

I am examining tax and voting data at the county level. The data spans 1990-2017. For tax data, I have one data point per county for each year for each of the six core tax variables (See Section 10.3.1 Tax Income Variables Available as Part of SOI County Data). For voting data, I have the voting pattern for Presidential election cycles (every 4 years) for all counties. In addition, I have population estimates and income estimates for each county for each year from the BEA. All analyses were conducted using the R statistical language (R Core Team, 2020) along with a series open-sources packages shared in R. For a full list of packages and citations, see Appendix B.6 Statistical Software and Packages.

The basic premises assumed for these analyses —as established in the sections covering the theoretical framework and the empirical framework are as follows: (i) election outcomes significantly alter the perceived legitimacy of the broader government; (ii) elections can be treated as a natural quasi-experiment, with the county being the unit-of-analysis and the election outcome serving as the exogenous treatment. The central hypothesis is that election outcomes change tax compliance at the county level by altering perceived legitimacy. As such, I am looking at whether election outcomes change the amount of taxes paid by a county immediately following an election relative to its tax burden in the year leading up to the election —with the main comparison being the relative difference in the annual difference in tax payment for winning counties vs losing counties (difference in difference analysis).

12.1 Identification Strategy

The current identification strategy derives from the fact that my central analyses focus on changes in tax compliance for the same tax year as the presidential election (e.g. examining the effects of the 2000 presidential election on the tax payments for Jan-Dec 2000). Since elections take place in early November, over 80% of the time covered by the tax year (Jan-Nov) has already passed before the outcome of the election is known, which effectively constrains the effect that election outcomes could have on the true tax burden via any established economic, social, political, or demographic pathways. And, with less than 2 months remaining in a tax year after the election outcome is determined and no legal or administrative power transferred to the winning party until Inauguration Day (January 20) in the following year, I argue that there is no probable pathway[1] for the outcome of an election to substantially alter the true tax burden (i.e. reduce the true income or increase the true amount of deductions) in those final 2 months —especially not in a manner that would differentially target counties based upon whether they voted for the winning or losing candidate. This intuition is also supported by analyses conducted by Mian et al. (2018), where they find no evidence for a change in economic activity in the two months immediately following an election. Finally, since the tax burden concerns the past year, the incoming government cannot retroactively impact the tax burden through shifts in tax policy. Thus, there are no reasonable pathways for policy decisions taken by the incoming administration to impact the true earnings or true tax burden. As such, this arrangement of statistical analysis largely obviates the need for an exhaustive list of control variables to justify preliminary causal inferences.

One point of potential concern may arise from the fact that losing or winning an election could change the motivations of residents —for example, there is evidence that winning or losing an election can dramatically alter one's stated future economic expectations. Although there is no logical or analytic method for controlling for the effects of changes in economic expectation, a recent paper that examined the effect of elections on economic expectations found that there was no evidence for a corresponding change in actual economic activity (Mian et al., 2018). Thus, we can conclude that it is reasonable to assume the election

[1] I fully acknowledge the caveat that there always exists some unforeseen pathway for the improbable to occur

outcome did not differentially impact the actual economic activity of winning or losing counties in the final 2 months of the tax year. Having ruled out any plausible pathway from election outcome to shifts in actual tax obligations, I argue that any significant effect of the elections can only occur by changing the tax compliance behavior directly.[2]

12.2 Comparison of Core Variables across Counties

12.2.1 Four Primary Classification Methods

As was described in Section 10.2.2 Aggregating Partisanship Status Across Elections, counties were classified as either Republican or Democratic using multiple classification approaches. From the set of multiple possible classification approaches, the following four primary classifications were selected as the most profitable and most intuitive options: (a) *County —Loyal Partisans*: Counties that voted for the same party in all 7 elections (Simple Majority Standard); (b) *Median Split*: Classify counties using a median split on their average Democratic Vote Share across 7 elections; (c) *Quartile Split*: same as median split, except we simply select the top and bottom quartiles; (d) *Majority (i.e. 50%) support*: classify a county as belonging to a party if it received over 50% of the average vote share across 7 elections.

Figure 12.1 shows the 4 county classification approaches below. Republican counties are marked in red and Democrat counties are marked in blue. The counties shown in white were not classified as belonging to either party. Map 3 shows the "median split" approach to classification. As one would expect, using median split produces the fewest number of unclassified counties. Of all the four maps, Map 3 also shows the greatest number of counties shaded in blue. This high number of Democratic counties largely reflects the fact that the "median split" approach results in the highest rate of misclassification (i.e. counties that

[2]By directly, I mean that elections change tax compliance behavior not through an indirect pathway, such as a changes in the legal, financial, or economic constraints or incentives.

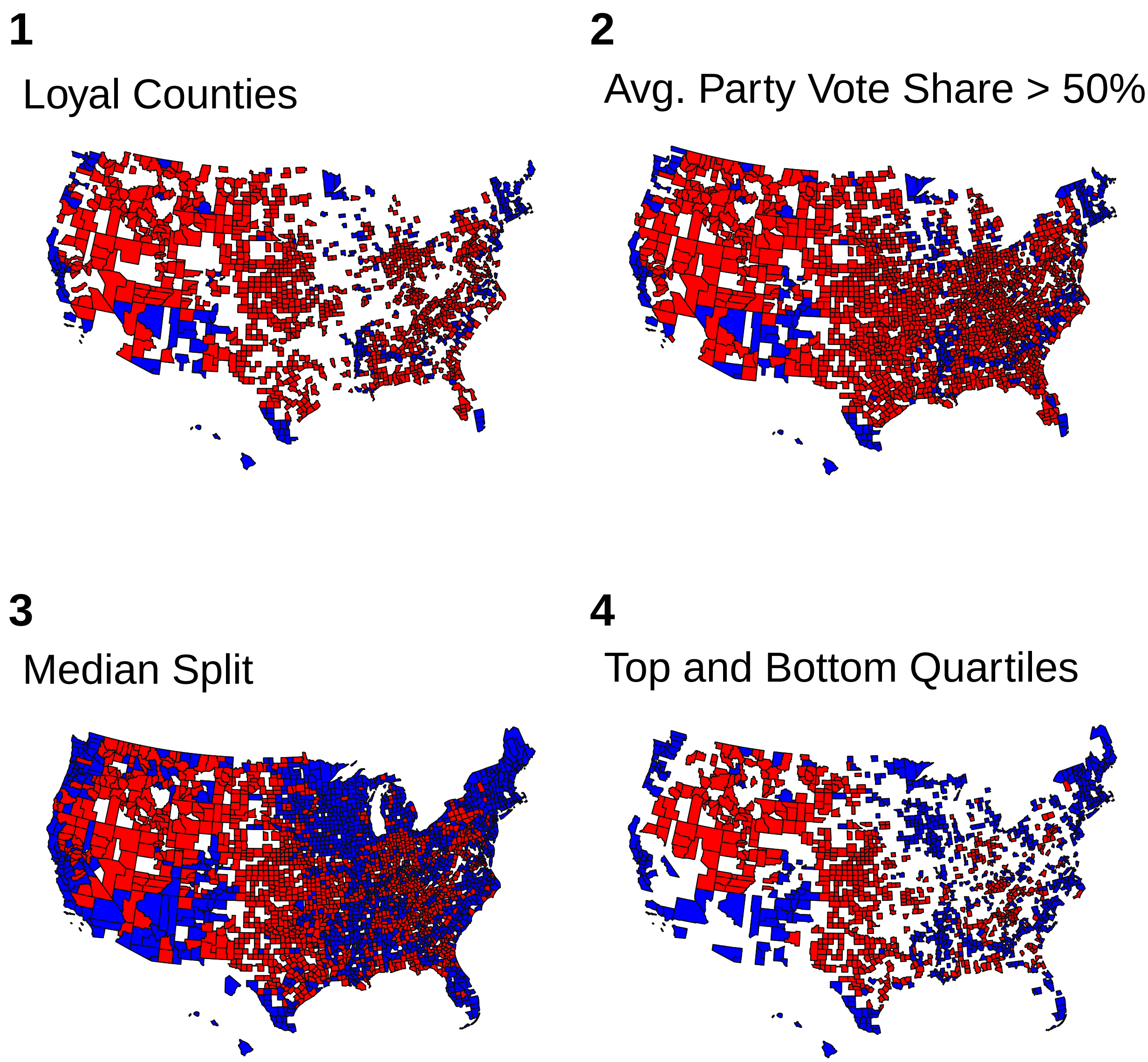


Figure 12.1: Four Approaches to County Classification by Partisanship

almost always voted Republican can still end up classified as Democrat due to the nature of the distribution of counties across the US).

Map 1 and Map 4 reveal that the "Loyal Counties" and "Top and Bottom Quartiles" classification approach leaves the largest number of counties unclassified. For the *Loyal Counties* approach, 1314 (approximately, 48%) counties are unclassified. The *Top and Bot-*

tom Quartile approach yield approximately the same result, with 1377 unclassified counties. Although, the two approaches produce an approximately equal number of classifications, they differ significantly in the distribution of counties classified as Democrat vs Republican. In the *Quartile* approach, by design, there are an equal number of counties assigned to each party (688 each), whereas in the *Loyal County* approach, 1138 (41%) are assigned as Republican and 301 (11%) are assigned as Democrat. The distribution of economic and tax variables across the four classification systems is examined next in the tables below.

Comparing Summary Statistics for Variables of Interest Across Classification Approaches

Table 12.1 shows information about basic tax variables, demographics, the constructed tax variables, and voting measures for counties. Its main purpose is to facilitate easy comparison of the four main classification techniques according to these primary variables. The first column shows the list of measures: (a) *Basic Tax Variables*: Population, Number of Returns, Real AGI; (b) *Standardized Tax Variables*: Real AGI per Return; Real Unmatchable Portion of AGI per Return; (c) *Constructed Proxy Measures*: Ratio of AGI to PI; Ratio of AGI per Return to Median Household Income; (d) *Voting Measures*: Democratic Vote Share; Margin of Victory; Proportion of Elections won by Democratic Presidential Candidates. For each measure, the average is presented with the standard deviation in parentheses below. Finally, in the bottom section, we see information about the range of years and number of counties covered in the panel.

Data Type	Variable	Statistic	All Counties	Median Split on Avg. Democrat Vote Share		Top vs Bottom Quartile of Avg. Dem. Vote Share			Majority - Avg. Party Vote Share > 50%for either Ds/Rs			Loyal - Supported Same Party for All Elections		
				Democratic County	*Republican County*	*Democratic County*	*Republican County*	*Unclassified*	*Democratic County*	*Republican County*	*Unclassified*	*Democratic County*	*Republican County*	*Unclassified*
Basic Tax Variables	**Population (Number of persons)**	Avg.	99K	154K	45K	224K	34K	70K	264K	53K	130K	327K	61K	81K
		Std. Dev.	(315K)	(433K)	(72K)	(563K)	(52K)	(168K)	(639K)	(134K)	(317K)	(741K)	(140K)	(228K)
	Number of Returns	Avg.	41K	64K	18K	94K	14K	28K	111K	21K	54K	138K	25K	33K
		Std. Dev.	(130K)	(179K)	(30K)	(233K)	(21K)	(69K)	(264K)	(56K)	(132K)	(304K)	(58K)	(96K)
	Real AGI	Avg.	$2B	$4B	$982M	$6B	$747M	$2B	$7B	$1B	$3B	$9B	$1B	$2B
		Std. Dev.	($9B)	($12B)	($2B)	($16B)	($2B)	($5B)	($18B)	($4B)	($9B)	($20B)	($4B)	($7B)
Standardized Variables	**Real AGI Per Return**	Avg.	$49K	$50K	$48K	$50K	$47K	$48K	$51K	$48K	$51K	$52K	$49K	$47K
		Std. Dev.	($12K)	($13K)	($10K)	($15K)	($10K)	($10K)	($16K)	($10K)	($12K)	($18K)	($11K)	($10K)
	Real Unmatchable Portion of AGI Per Return	Avg.	$11K	$11K	$11K	$11K	$11K	$11K	$11K	$11K	$11K	$11K	$11K	$10K
		Std. Dev.	($5K)	($5K)	($5K)	($5K)	($5K)	($4K)	($6K)	($5K)	($5K)	($6K)	($5K)	($4K)
Constructed Variables	**Ratio of AGI to BEA's Estimate of Personal Income**	Avg.	58%	59%	58%	59%	57%	59%	58%	58%	60%	58%	59%	58%
		Std. Dev.	(8%)	(8%)	(8%)	(8%)	(8%)	(8%)	(8%)	(8%)	(7%)	(8%)	(8%)	(8%)
	Ratio of AGI per Return / Median Household Income	Avg.	107%	108%	105%	110%	105%	106%	110%	106%	107%	112%	105%	107%
		Std. Dev.	(11%)	(12%)	(11%)	(12%)	(11%)	(11%)	(13%)	(11%)	(10%)	(13%)	(11%)	(11%)
Voting Measures	**Democrat Vote Share**	Avg.	40%	48%	31%	54%	26%	39%	58%	34%	46%	60%	31%	43%
		Std. Dev.	(13%)	(10%)	(9%)	(10%)	(7%)	(8%)	(10%)	(9%)	(7%)	(9%)	(8%)	(10%)
	Margin of Victory	Avg.	24%	16%	32%	18%	41%	19%	23%	28%	10%	27%	32%	17%
		Std. Dev.	(18%)	(13%)	(19%)	(15%)	(18%)	(14%)	(16%)	(19%)	(10%)	(16%)	(19%)	(15%)
	Proportion of Elections Dem. Candidate Won County	Avg.	30%	54%	6%	81%	2%	18%	93%	10%	50%	100%	0%	40%
		Std. Dev.	(34%)	(33%)	(10%)	(21%)	(6%)	(17%)	(12%)	(13%)	(19%)	0%	0%	(22%)
Panel Data Coverage	**Num. Unique Observations in Panel**		77084	38556	38528	19264	19264	38556	12040	51492	13552	8428	31864	36792
	Num. Unique Counties in Panel		2753	1377	1376	688	688	1377	430	1839	484	301	1138	1314
	Years Covered by Panel		1990-2017											
	Num. Unique Tax Years Covered by Panel		28											

Detailed description of each of the following variables can be found in the glossary at the rear of the document. Please note all dollar amounts reported here are inflation adjusted to 2012 USD.

Table 12.1: Comparing Tax, Income, and Demographic Variables Across the Four Primary Approaches to Classification of Counties

The top column headers list the four main approaches to classifying counties according to their partisanship: (1) median split; (2) top versus bottom quartile; (3) using majority approach; and, (4) loyalty standard. Under each of these columns headers, we see the counties split by either Democrat, Republican, or Unclassified (other than Median Split, which leaves no counties unclassified).

Reading along the bottom row in bold, we see the number of unique counties in each of the categories. The first thing to notice is that the table is organized from less to more restrictive as we move from left to right along classification approaches (i.e in terms of restrictiveness : Median Split < Quartile < Majority < Loyalty). As a result, we see under the first approach (Median Split) 1377 counties classified as Democrats; under the second approach (Quartile), 688 counties are classified as Democrat; under the third approach (Majority), this number drops to 430, and under the most restrictive approach, only 301 counties get classified as Democrat.

In exchange for this increasing restrictiveness, we see better segregation of Democratic and Republican counties. For example, reading along the row showing "Proportion of Elections that Democratic Candidates Won", we see that a county classified as Democrat under the first approach (Median Split) only supported the Democratic candidate for President in 54% of the county-election observations. In other words, for each of the ~1400 counties, we have observations for 7 elections —thus, giving a total of ~9800 observations. Out of these ~9800 observations, for 54%, the plurality of the county supported a Democratic candidate. For 46% of the observed county-elections, the plurality supported either a Republican or a third party candidate. This suggests a relatively poor classification of counties under the median split approach. In contrast, the second approach (Quartile) shows improved segregation such that a county classified as Democrat supported the Democratic candidate for 81% of the county-elections. The third approach (Majority) is even better at segregation: counties classified as Democrats supported the Democratic candidate for President 93% of the time. The final and fourth approach (Loyalty) produces a perfect segregation (by construction) with counties classified as Democrat supporting the Democratic candidate for 100% of the county-elections observed over the 28 year sample. This improved segregation is also reflected in the vote share, which reading along the "Democratic Vote Share" row shows an increase from 48% (Median Split) to 54% (Quartile) to 58% (Majority approach) to 60% (Loyalty approach). The improved segregation can also be seen when we examine the average "Margin of Victory", which increases from 16% for Democratic Counties under the first approach (Median Split) to 27% average margin of victory for the final approach (Loyalty).

Having oriented to the organization of the table, if we examine the first row (Population), we see that the average population of a county is around 99K for the entire country. As we read along the first row, it becomes clear that as we draw increasingly strict segregation between Democratic and Republican Counties, the sample of Democratic Counties increasingly represents more and more populous counties, going from an average of 154K when classified using Median Split to an average of 327K when classified using the Loyalty approach. This increase is also reflected in the average number of returns filed in a county, which increases from 64K under the Median classification to around 138K by the time we limit Democratic counties to the purely loyal ones. As one would expect, the increase in population and the increase in number of tax returns also reflects in the average total Adjusted Gross Income (AGI) reported in a county. The difference between the average total AGI between Democratic and Republican counties moves from $3B under a median split to $8B when comparing the loyal counties.

However, reassuringly, when we examine the standardized variables —for example, AGI per Return, the average AGI per Return is quite similar across Democratic and Republican counties, no matter which classification approach is used. The same is true when we examine the Average Unmatchable Portion of AGI per Return across counties, which is $11K/return for Democratic and Republican counties across all 4 classification systems. The similarity in measures across Democratic and Republican counties persists even when we examine the constructed proxy variables (i.e. AGI over PI or AGI per return over Median Household Income), a pattern which holds for all four classification approaches. In summary, for all the variables of concern, Democratic and Republican counties are very similar to each other and this similarity persists independent of the method used for classification.

12.3 Testing 2007 Imputed Data

The careful reader may recall that number of returns for the year 2007 were imputed by averaging the two neighboring years (2006 and 2008) (see Section 10.3.2 Concerns with Standardizing using Number of Returns: The Economic Stimulus Act of 2008). Since a majority of the initial analyses rely upon the standardization of AGI by number of returns, I first present some descriptive graphs to examine how well the imputed data capture the actual events in that time period.

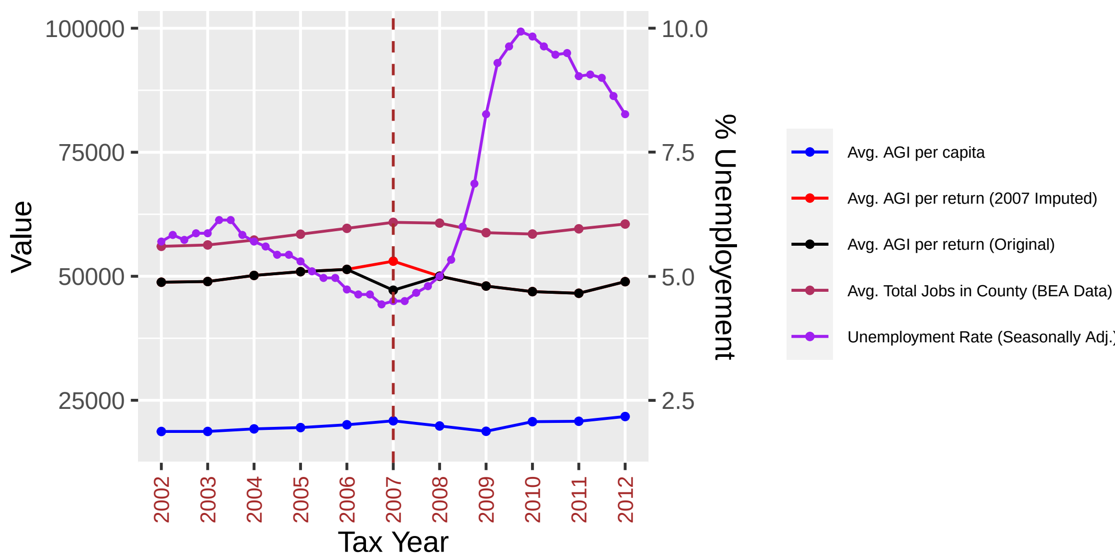


Figure 12.2: Comparison of Multiple Measures of Economic Activity Surrounding 2008 Financial Crisis

In Figure 12.2, I plot time course of economic activity around the 2008 financial crisis so that one can evaluate the appropriateness and success of imputing the 2007 number of returns using the simple strategy of averaging data from the 2 neighboring years (2006 and 2008) for each county. The black line shows the average inflation-adjusted AGI which was standardized using the unadjusted number of returns. The red line shows the average inflation-adjusted AGI standardized using the imputed 2007 data for number of returns. The red line only diverges from the black line for the year 2007, since that was the only year when I imputed data. The blue, purple and maroon lines provide three independent standards for comparison. The blue line shows the average (inflation-adjusted) AGI per capita. The purple line shows the quarterly unemployment rate (seasonally-adjusted) as compiled by the BLS. The maroon line shows the annual estimate for the average total number of jobs in a county as compiled by the BEA. The purple and maroon lines serve as markers for the primary trends in the economy and, thus, provide measures for the effects of the 2008 economic crisis on the earning / income of counties. The vertical dotted line marks 2007, the year of interest for our current purposes.

As one can see, the unadjusted data (i.e. the black line) dips right at 2007 and then spikes up for 2008 before showing a steady decline until a period of recovery around 2011. This would suggest that the most dramatic effects of the financial crisis were felt in 2007 and that they were followed by a rebound the next year. In fact, the major effects of the 2008 crises were not felt until September of 2008, when Lehman Brothers declared bankruptcy. As was explained above, this 2007 dip in standardized AGI seen in the unadjusted data (black line) are an artifactual consequence of the surge in tax filings in 2007. This surge largely consists of a previously unrepresented segment of the population that was making between \$3000 (the threshold of income required to receive a stimulus) and ~ \$7000+ (the minimum threshold of income that required a person to file under the IRS guidelines). Thus, there was an artificial drop in the AGI per return.

In fact, if one examines the AGI per capita figures in blue and the jobs numbers in maroon, it is clear that the economy was still doing well for 2007. Additionally, the purple line representing unemployment rate also confirms this reading. Until the very end of 2007, one can see a steady decline in the overall unemployment rate across the country. As the unemployment rate begins to rise between 2007 and 2008, the red line correctly depicts a corresponding decrease in the average earnings as captured by the AGI per return measure (in contrast to the black line, which incorrectly depicts the opposite trend). In sum, it is reassuring, that the time course shown in red matches the time course represented by the three other lines. In other words, the AGI figures standardized by the imputed 2007 data show the same general pattern and trends as the AGI per capita, the average total jobs in a county, and the country-wide unemployment rate, all three of which correctly track the time course of the financial crises.

12.4 Graphical Analyses of Legitimacy Hypothesis

To evaluate the legitimacy hypothesis, I first examine primary tax variables like AGI per return. Finally, I examine the proxy measures of tax compliance that were created as a ratio of reported taxable income and estimated personal income (AGI / PI). For the analyses presented here, I selected the "Perfect Loyalty" approach (see, Section 10.2.2 for details), which restricts the assignment of party affiliation to only those counties that voted for the same party by plurality in all 7 elections in the sample. This approach to classification is the least likely to suffer from endogeneity concerns, since the counties do not change their party support over the entire sample. However, for the sake of robustness, I also complete the analyses using other methods of classification and present these alternate graphs in the Appendices (for an easy comparison of the main classification approaches, see Table 12.1).

12.4.1 Trends in AGI per Return

In order to examine the effect of elections on tax compliance, a natural place to start might be to examine how tax payments change over time. First, let us begin by examining the distribution of AGI per return across all counties and all years. The four panels in Figure 12.3 show the distribution of AGI per return; log AGI per return; percentage change in AGI per return; and a box-and-whisker plot which shows the distribution of percentage change in AGI per return for each of the 28 years. Please note that, hereinafter, whenever I use the term AGI, I am referring to the AGI in Real USD (i.e. inflation-adjusted). And, whenever I mention the number of returns, it includes the imputed data for 2007.

Distribution of Data

Distribution of AGI per Return In Figure 12.3, we see the distribution of AGI per return for the raw, log, and annual percent change versions of the variable. Figure 12.3(a) shows the raw AGI/return values which range from \$18K to \$174K. As is common with income variables, values of AGI per return across counties are highly skewed with a long, rightward tail which closely approximates a textbook log-normal distribution. The mean value is \$49K and the median is \$46K. Figure 12.3(b) shows the log-transformed variable, which shows a more balance distribution with similar mean and median values of 10.77 and 10.75 respectively. Due to the highly skewed nature of the raw variable, for the time series graphs, I will normally default to this log-transformed version. Figure 12.3(c) shows the annual percentage change in AGI per return across all counties. The first thing to notice is the presence of extreme outlier values with the maximum positive annual percentage change equaling 224% and the maximum negative annual change equaling -70%. However, 99% of the data actually fall within the much narrower and more reasonable range of -15% to 17% and 95% only spans -9% to 10%. A review of the extreme outliers reveals some notable discrepancies which should encourage us to remove them for the main analyses. For example, one of the more extreme values comes from Yazoo County, MS, which has a normal range of overall inflation-adjusted AGI from $289M in 1993 to $373M in 2013. However, in the year 2010, we see the overall AGI spike (for just that single year) from this consistent range to $1.30B resulting in the reported 224% annual percentage change. Basic searches for events in 2010 only reveal the occurrence of a tornado, which should hardly result in a $1 billion jump in AGI. This case is exemplary of a number of similar situations where divergence from normal reporting patterns represents either reporting errors or transient and extraneous factor or a combination, none of which represents the changes of interest in tax reporting behavior for the normal population of the county. As such, the inclusion of such data would be inappropriate for any analyses that rely upon voting patterns / partisanship status of a county to identify election-based changes.

The presence of extreme outliers can be most clearly seen in Figure 12.3(d), which shows the distribution of Annual Percent Change in AGI per Return for every year in the sample as a box-and-whiskers plot. The dots represents values that are outside the 1.5*IQR threshold that is normally used. The dotted and dashed lines running horizontally across the box plot capture the winsorizing strategy. In order to prevent the spurious influence of extreme outliers and bad data, I implement a simple two-stage winsorizing strategy that first excludes all values outside the lower and upper bounds of 50.0%–50.0% and then computes the 0.25% and 99.75% percentiles. The first stage cutoffs are shown by the bright red, dashed lines. The second-stage, percentile-based, computed cutoffs are shown by the muted red, dotted lines. All values outside the dotted lines were set to equal these percentile-based thresholds, resulting in winsorized values which ranged from -21.71% to 25.71%. The histogram in 12.3 on the bottom-right shows the overall distribution of the winsorized variable. The dashed and dotted lines represent the same thresholds they did in the 12.3(d) —namely, the dashed lines are the first-stage *a priori* thresholds and the dotted lines are the second-stage, percentile-based winsorizing thresholds. Right at the vertical dotted lines, small red bars are visible in the histogram, which show the quantity of additional (previously outlying) observations that accumulate in the bins at the thresholds as a result of the winsorizing procedure.

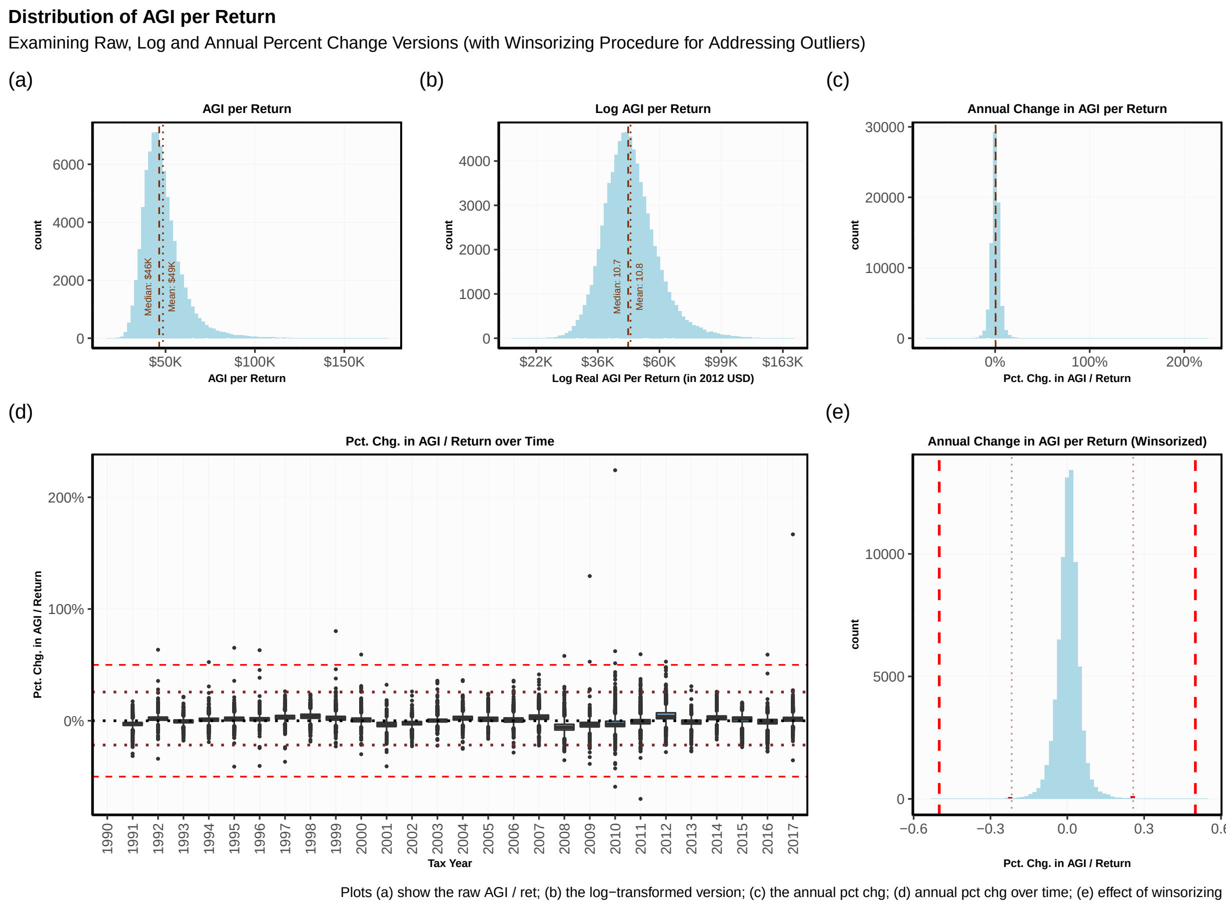


Figure 12.3: Distribution of Adjusted Gross Income per Return for all U.S. Counties across the 28 Year Sample

Distribution of AGI per Return - Split by County Having reviewed the overall distribution, in Figures 12.4(a-c), I present density plots capturing the distribution of raw, log, and annual percent change variants of AGI per return separated by county type. The first thing to notice is that —unlike the Republican and Unclassified counties —the distribution of raw and log AGI per return for Democratic counties 12.4(a,b) are not well-behaved and are clearly bi-modal. In the Appendix B.3 Section B.3.1, I present the same distributions for all classification approaches (Figure B.5: Median Split; B.6: Top vs Bottom Quartile, B.7: Majority Threshold, and B.8: Loyalty). In comparing the density plots across the different classifications, there is a clearly discernible pattern for the distribution of Democratic counties: as the classification approach gets better at identifying Democratic counties, the distribution becomes correspondingly more and more bimodal —with Median Split being the least bimodal and Loyalty-based classification being the most so. This pattern suggests that the Democratic party has two distinct constituencies in terms of income distribution, with one set showing an average AGI per return of $35K and another with an average AGI per return of $57K. As shown in 12.4(c), this feature of the Democratic base is less troubling for the current purposes when we examine the distribution of the Annual Percent Change in AGI per Return, which shows almost identical distributions for Democratic, Republican, and Unclassified counties.

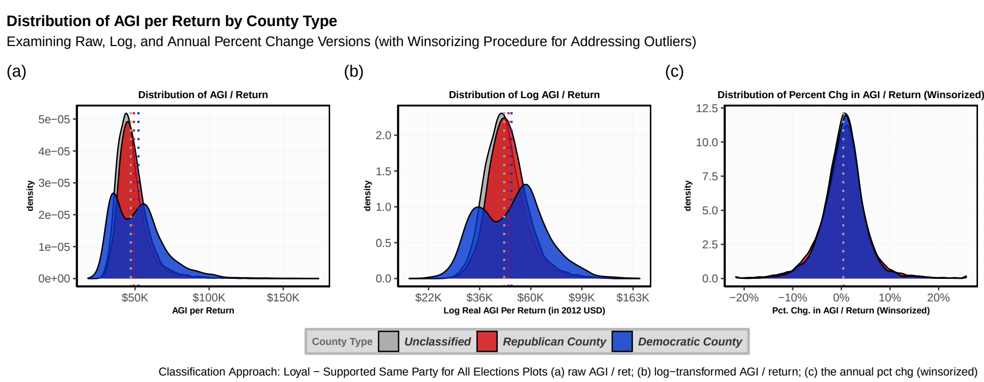


Figure 12.4: Distribution of Adjusted Gross Income per Return for all U.S. Counties across the 28 Year Sample Classified According to Partisanship

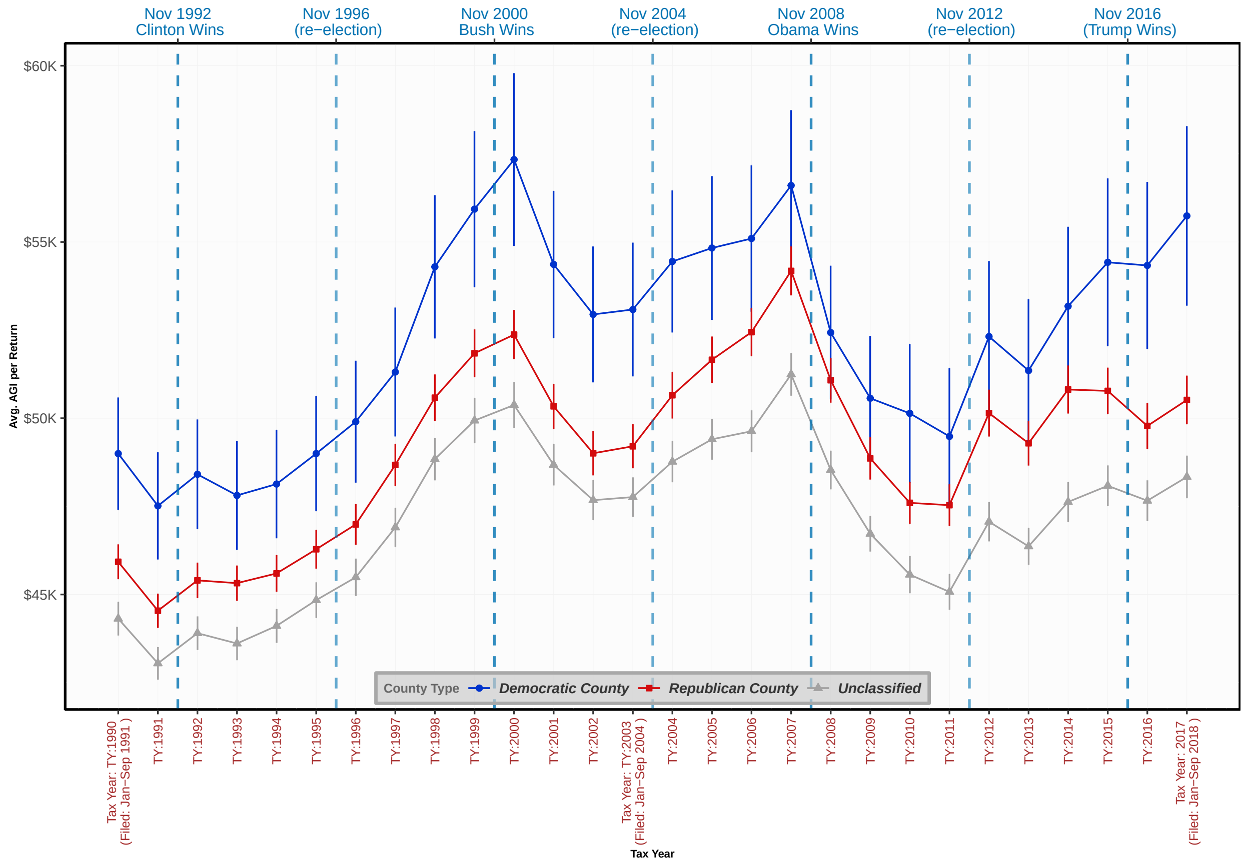
Full Time Series

I use the full time series graph to examine the broader trend of a variable across the entire sample and to develop a sense of any major, over-arching patterns in the measures' temporal evolution. I begin by examining the real (inflation adjusted) adjusted gross income (AGI) per return. After examining the full time series for the raw variable, we move to the annual percentage change version of the same.

Real AGI per Return In Figure 12.5, the red line (with square markers) represent the Republican counties, the blue line (with round markers) represent the Democratic counties, and the gray line (with triangle markers)represent the Unclassified counties. This color and shape palette is used consistently across this chapter and will always represent the same groups. The vertical, dashed blue lines mark the November Presidential Elections —with each election outcome marked at the secondary x-axis on the top of the graph. The primary x-axis shows the tax years across the sample, starting with Tax Year 1990 till Tax Year 2017. It is worth highlighting again that the value for any given tax year represents the AGI per return based upon income earned during Jan-Dec of a given year. However, taxes are filed and must be paid to the IRS by April of the following year. Thus, for example, the figures for Tax Year 1992 relate to income earned between January-December of 1992 as represented on filings received by the IRS between January-September of 1993.

As we can see in Figure 12.5, the red line representing the Republican counties starts off with about $3K less in AGI per return than the blue line representing the Democratic counties. Over the course of the 1990s and early 2000s, the average AGI per return for Democratic, Republican, and the Unclassified counties grows at about the same rate —with the difference between Democratic and Republican counties remaining between $2K–$5K. However, from the period 2004-2008, we see a steady and significant increase in the growth of AGI per return for Republican counties until the average AGI per return for both types of counties are statistically indistinguishable from each other. In the following years, AGI per return for both types of counties remain comparable —with the difference between them hovering between \$0–1.5K —until 2015. Starting with 2015, we see a dip in the AGI / return for Republican counties. The difference between the two types of counties continues to grow after the 2016 Trump election —resulting in a difference of \$5K by 2017, the final year in the sample.

In terms of the hypothesis under consideration, there are no obvious trends that are visible at this level of detail. Overall, the average AGI per return does not appear to vary in any discernible manner as a function of the presidential election outcomes. Log transforming the data does not appear help, as shown in Figure B.9 of the Supplementary Figures Section B.3. When year fixed effects are removed, as shown in Figure B.10, it is clear that variability around the elections is too large to be able to draw any meaningful inferences. This conclusion is also supported by Figure B.11 which shows the time series in terms of the difference between Democrat and Republican counties. As a result, it not considered valuable to pursue any further inquiry in terms of raw AGI per return. Instead, in the next section, I examine the same time series represented in terms of annual percentage change in AGI per return.



County Classification Approach: Loyal − Supported Same Party for All Elections; Error bars show 95% CI; Std.Errors for Group Means are unadjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure 12.5: AGI per Return for Democrat and Republican Counties over Time (inflation-adjusted 2012 USD)

Annual Pct. Change in AGI per Return In Figure 12.6, we can see the annual percentage change in AGI per return plotted over time. Overall, the lines representing the Democratic, Republican and Unclassified counties tend to travel together. While it is possible to detect some segregation in the three types of counties in the years surrounding the presidential election, there is entirely too much year-on-year variability to be able to detect any meaningful overall trends at this level of detail.

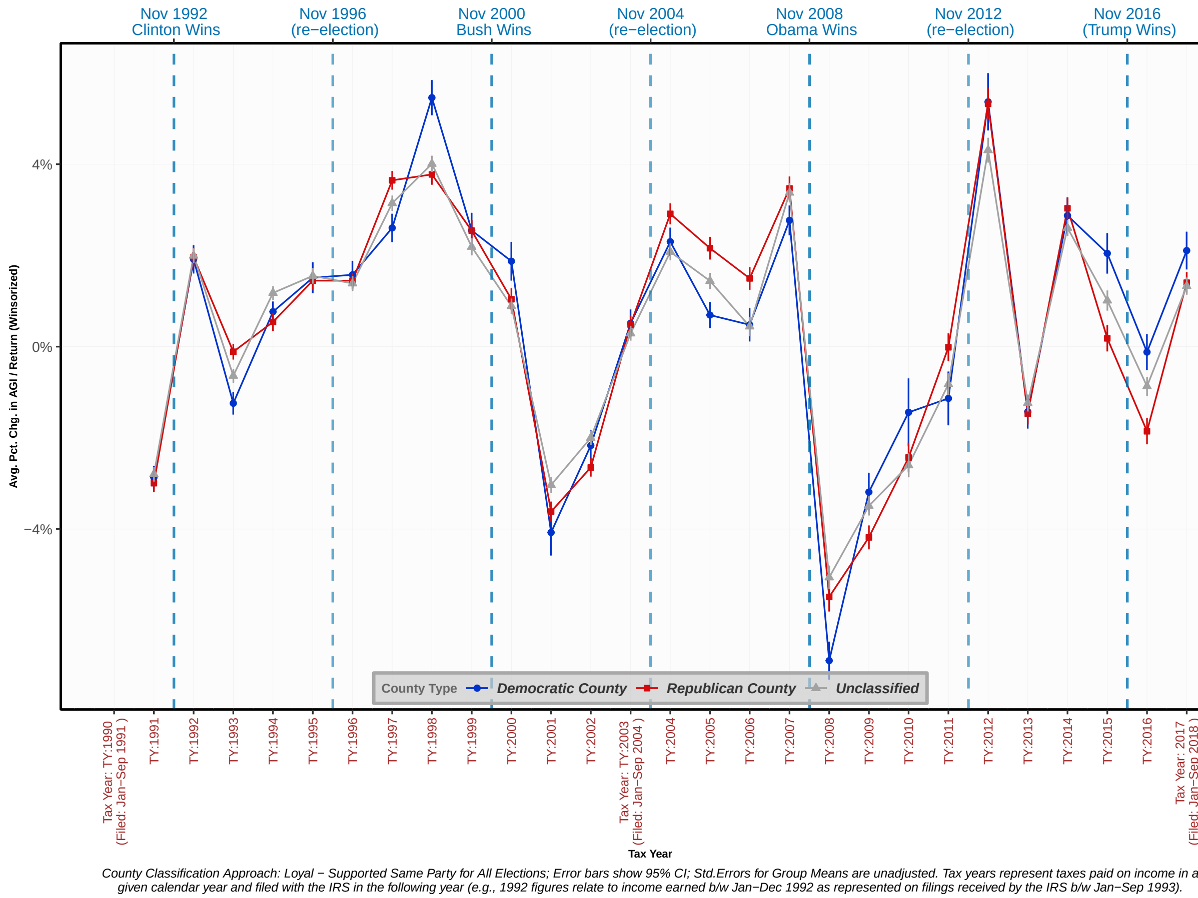
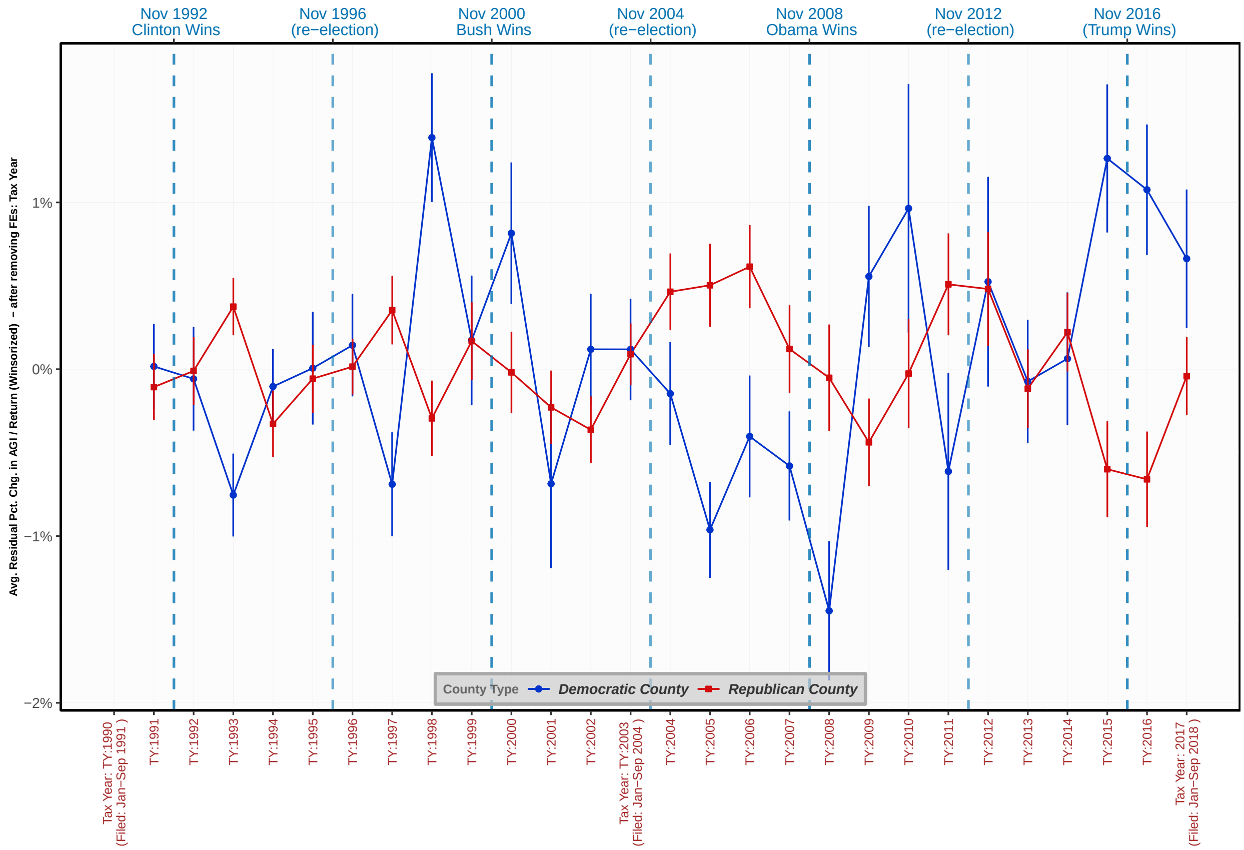


Figure 12.6: Annual Percent Change in AGI per Return for Democrat and Republican Counties over Time (inflation-adjusted 2012 USD)

Annual Pct. Change in AGI per Return - Removing Time Fixed Effects To clarify underlying trends, in Figure 12.7, I present the same time series with year fixed effects removed. To declutter the graph, data for unclassified counties (i.e. counties that did not meet the threshold to be assigned a partisan status) is no longer shown. With the idiosyncratic effects of time removed, we can begin to see evidence of a clear divergence in the lines representing Democratic and Republican counties, but no overall trend is apparent.

The lack of an obvious pattern in these full sample time-series is a common characteristic across most variables, so after the initial few, I generally refrain from presenting such a figure for other variables. Instead, in the following figures, I segregate the data by election cycle.



County Classification Approach: Loyal − Supported Same Party for All Elections; Error bars show 95% CI; Std.Errors for Group Means are unadjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure 12.7: Annual Percent Change in AGI per Return for Democrat and Republican Counties over Time —After Removing Year Fixed Effects

Time Series Split by Election Cycle - Annual Pct Change in AGI per Return

In order to examine the elections trends, in Figure 12.8, I present the annual percentage change in AGI per return for Democratic and Republican counties split by election cycle. To facilitate analysis, in Appendix B.3, Figure B.12 shows the identical graph with the time fixed effects removed and Section B.3.3 contains the corresponding analyses for each election. I have also presented the identical graph in terms of difference in means for Democratic and Republican counties in Figure B.13 of the same appendix. Finally, in Appendix B.3.4 [Panel by Election - Comparing Classification Approaches], I present the graph for each individual election cycle according to all four classification approaches side-by-side for comparison. This

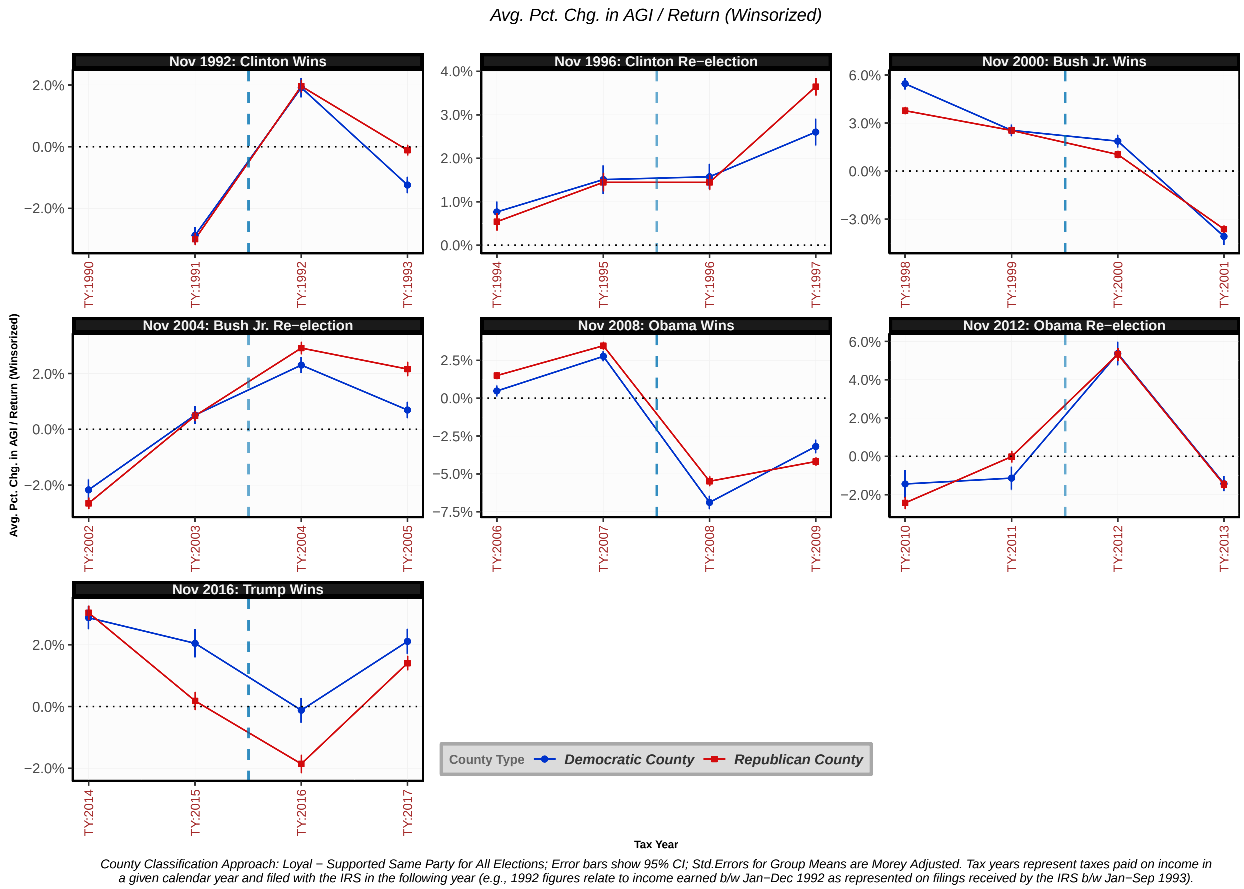


Figure 12.8: Annual Percentage Change in AGI per Return for Democrat and Republican Counties Separated by Elections Cycle

allows the reader to examine the effect of different classification approaches on the trends seen here (for 1992 Election Cycle, see Figure B.14; for 1996 Election Cycle, see Figure B.15; for 2000 Election Cycle, see Figure B.16; for 2004 Election Cycle, see Figure B.17; for 2008 Election Cycle, see Figure B.18; for 2012 Election Cycle, see Figure B.19; for 2016 Election Cycle, see Figure B.20).

I begin with the top, left-most graph showing the 1992 Election Cycle, in which, Bill Clinton, a Democrat won. In this cycle, the rate of annual growth in AGI per return for both Democratic and Republican counties hovers at -3% for Tax Year (TY) 1991, then jointly moves to positive 2% in the election year (TY 1992). Post-election, these trends diverge with a 2% drop for Republican Counties (from 2% to 0%) and a 3.1% drop for Democratic

Counties (from 2% to -1%). This additional drop in the growth rate for Democratic Counties is consistent with the alternative hypotheses, where losing an election is accompanied by a decrease in cheating and winning is accompanied by an increase (i.e. either Hypothesis 2: Moral Licensing and a Winner Effect or Hypothesis 3: Perceived Enforcement Risk).

The identical pattern is seen in the 1996 election cycle. For both types of counties, the growth rate moves jointly from 1% to 1.5% and stays at 1.5% the following year as well (TY 1994–1996). Post-election, Republican counties show a 2% increase in growth rate (from 1.5% to 3.5% for TY 1996–1997) while Democratic counties only show a 1% increase (ending with 2.5% growth for TY 1997)—i.e. following Clinton's re-election, there is an additional 1% increase in growth for Republicans. As with 1996, this pattern is consistent with the alternative hypotheses. It should be noted that the election effect takes place for the first full year under the presidency (i.e. Tax Year 1993; 1997) and no effect is discernible for the election year itself (i.e. Tax Year 1992; 1996 do not show an election effect).

In the 2000 Election Cycle, the pattern appears to switch in support of the Legitimacy Hypothesis. Prior to the election, both Democratic and Republican counties show the same growth rate of 2.5% for TY 1999. The election does not seem to impact the steady decline in the growth rate for Republican counties, while it appears to arrest the rate of decline for Democratic ones. Post-election, however, we see much sharper 6.0% decline for Democratic Counties (from 1.9% to -4.1% for TY 2000-2001) in contrast to a slower 4.7% decline for Republican counties (from 1.1% to -3.6% for TY 2000-2001). Unlike the previous two election cycles, the additional 1.3% decline for Democratic counties post-election is consistent with the Legitimacy Hypothesis, which predicts that losers of the election would decrease income disclosure, which should decrease the AGI / return.[3]

For the 2004 election cycle, the pattern also supports the legitimacy hypothesis, albeit less strongly. For the 3 first years in the election window, the growth rate for Republican

[3]This pattern of legitimacy-hypothesis consistent trends can be clarified by examining Supplementary Figure B.13 showing the differences in means for Democratic and Republican counties.

counties steadily increased by ~2.5% year-on-year (from -3% to 0.5% to 3% in TY 2002–2004). Democratic counties showed a similar ~2.5% increase in growth rate for the years prior to the election (from -2% to 0.5% in TY 2002–2003). However, unlike their Republican counterparts, this rate of change slows after the election year and then drops by 1.6% between the election (TY 2004) and post-election years (TY 2005) —dropping from 2.3% to 0.7% in TY 2004–2005. This rate of decline post-election is more than twice the rate seen for Republican counties, which dropped by ~0.7% in the same window (from 2.9% to 2.2% in TY 2004–2005).[4] Like the previous election, this pattern of a slowdown and sharper drop in growth for Democratic counties post-election would be in line with the Legitimacy Hypothesis.[5]

The same pattern is seen in the 2008 election cycle, where —after moving in parallel for the first three years in the cycle (TY 2006–2008) —the change in growth rate for Democratic counties spikes upwards with a 4% year-on-year change in growth (from -7% to -3% in TY 2008–2009) while Republican counties show a much smaller upward trend of ~1.5% (from -5.5% to -3% in TY 2008–2009) —exactly what one would expect under the legitimacy hypothesis.

The 2012 Obama re-election cycle also appears to show a pattern consistent with the legitimacy hypothesis, with the growth rate in Democratic counties increasing by 6.5% (from -1% to +5.5% in TY 2011-2012) while Republican counties increased by a smaller 5.5% (from 0% to 5.5% in TY 2011-2012). The larger increase in growth rate for Democratic counties is consistent with the pattern one would expect if income disclosure increase for these counties after the re-election of Barack Obama in 2012. The change from election year to post-election year (TY 2012-2013) was identical for both types of parties and did not appear to show any

[4] This feature can also be seen clearly when examining Supplementary Figure B.12, which shows that - after removing time fixed effects - after the election, Republican Counties showed a positive (approx. +0.5%) residual growth rate for and Democratic counties showed a negative (approx. -1%) residual growth rate.

[5] However, a review of the Supplementary Graphs in Figure B.13 showing the differences in means suggests that the election effect for the 2004 re-election cycle is minimal.

divergence as a consequence of the election.

Like with previous elections, the pattern for the 2016 election cycle also appears to be in line with the Legitimacy Hypothesis. From the pre-election (TY 2015) to election year (TY 2016), both Democratic and Republican counties appear to be moving in parallel, shifting by ~2% year-on-year (Democratic Counties: from 2% to 0%; Republican counties from 0% to -2%). However, in the year following the election, we see a 3.5% spike in Republican counties (from -2% to +1.5% in TY 2016-2017). In contrast, we only see a 2% spike for Democratic counties (from 0% to 2.1% in TY 2016-2017).[6] Overall, this pattern of relative increase for Republican counties is what we would expect if winning the election increased income disclosure for the winning party i.e. a "legitimacy effect."

Overall, we see a mixed pattern of findings. The first two elections in the sample 1992 and 1996 appear to support the alternative hypotheses, whereas the subsequent elections 2000, 2004, 2008, 2012, and 2016 appear to support the legitimacy hypothesis. The election effects consistent with the legitimacy-hypothesis appear to be smaller for re-election cycles (2004, 2012) compared to the turnover elections (2000, 2008, 2016) —a feature I examine in subsequent sections. Overall, there appears to be a differential election effect between winners and losers of 1.2% to 2.3%.

Time Series for Elections Separated by Republican and Democrat Victory

Having examined the trends in each individual election cycle in detail, the following set of figures examine annual percent change in AGI per return collapsed across election cycles. In Figure 12.9(a), I present data for election cycles where the Democratic Candidate won (1992, 1996, 2008, and 2012) and in Figure 12.9(b), I present data for the election cycles where the Republican Candidate won (2000, 2004, and 2016). As with prior figures, the dashed, blue,

[6] This effect can also be seen clearly in the Supplementary Graphs in Figure B.13 showing the differences in means where we see a drop in the difference between Democratic and Republican counties between the election and post-election years from 1.8% in TY 2016 to 0.7% in TY 2017.

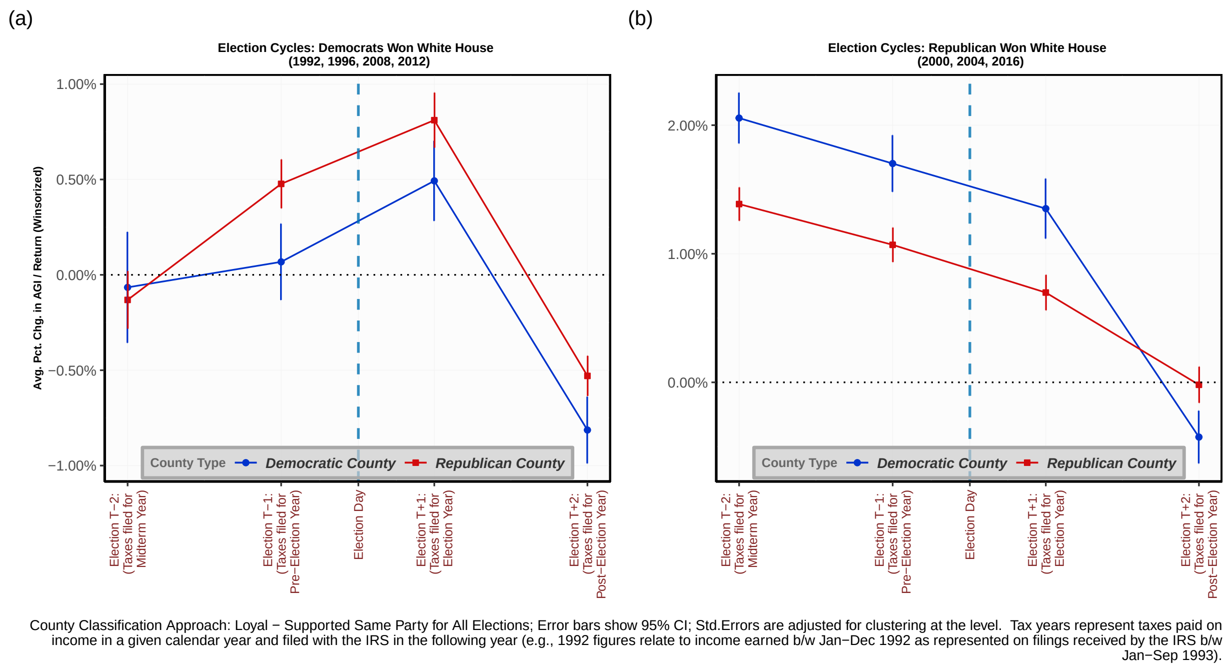


Figure 12.9: Annual Percentage Change in AGI per Return (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans

vertical line marks the November Presidential election. To the immediate left of the dashed line, at Election T-1, the values represent the annual percent change in AGI per return for the pre-election tax year (e.g. for the 2000 election, T-1 represents taxes paid in April 2000 for Tax Year 1999). To the immediate right of the dashed line, at Election T+1, the values represent the tax data for the election year (e.g. for the 2000 election, T+1 represents taxes paid in April 2001 for Tax Year 2000). At Election T+2, the values represent the tax data for the first post-election year (e.g. for the 2000 election, T+2 represents taxes paid in April 2002 for Tax Year 2001, the first year under the Bush presidency). The standard errors shown in the graph are cluster-robust, as recommended by Cameron & Miller (2015), and computed using the “arellano” method which allows for both heteroskedasticity and serial (cross-sectional) correlation (Zeileis, 2004) using the “HC1” estimator (MacKinnon & White, 1985); this approach to panel regression matches what is implemented in prominent econometrics software like STATA. The clustering of standard-errors was at the county-level, as recommended by Cameron et al. (2011); Esarey & Menger (2019); Bertrand et al. (2004). Cluster-robust standard errors were computed using the Sandwich package v. 2.5-1 (Berger et al., 2017; Zeileis, 2006) in R (Version 3.6.3; R Core Team, 2020).

To begin, let us examine the trends for election cycles where Republican candidates won, i.e. 12.9(b). For the first three years in the election window, the trend lines for both Republican and Democratic counties travel in parallel. Democratic counties steadily declining from 2.1% at T-2 to 1.7% at T-1 to 1.35% at T+1 (an approximately ~0.3% decline per year). Republican counties also steadily decline in the same period from 1.4% at T-2 to 1.1% at T-1 to 0.7% at T+1. For T+2, Republican counties drop by 0.7% to a -0.0% annual change (i.e. there is almost no change in AGI per return for Republican counties in the first year after the election). For T+2, Democratic counties, on the other hand, show a significantly larger decline of 1.8% (or, approximately 2.41735 x the decline seen for Republican county). Thus, after losing the election, Democratic counties show a decrease in the reported AGI / return, with a negative annual percentage change of -0.4%. This pattern of findings is exactly what one would expect according to the Legitimacy Hypothesis.

For election cycles where Democratic candidates won, 12.9(a), there does not appear to be a significant election effect. The trend lines for Republican and Democratic county remain parallel for both the T-1 to T+1 transition and the T+1 to T+2 transition. This lack of election effect is not surprising since these data include the 1992 and 1996 elections, which showed support for the alternative hypothesis. In Figure 12.10, I present the same data with 1992 and 1996 elections omitted from Panel (a), which now only represents 2008 and 2012, which are the remaining two election cycles where Democrats won the White House. Having removed 1992 and 1996, the pattern for elections where Democrats won the White House looks remarkably similar to the pattern shown in Figure 12.10(b), except now it is the Republican Counties which show a dramatic drop in annual percentage change between T+1 and T+2. Democratic counties continue an uninterrupted decline after the election with a

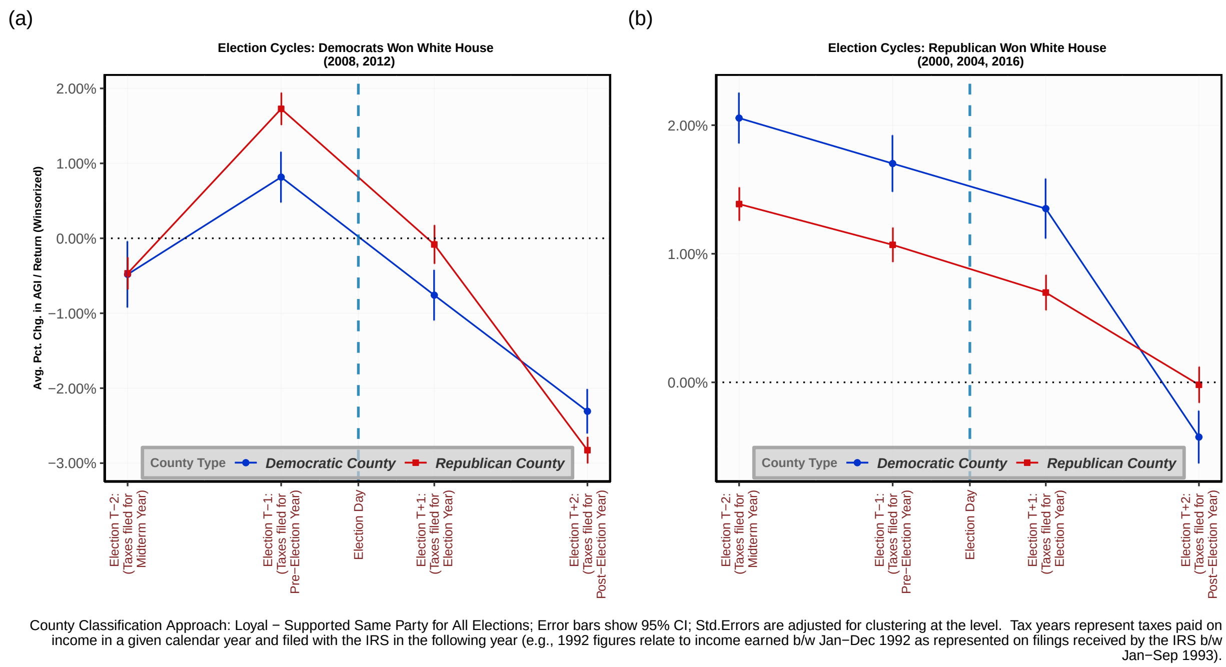


Figure 12.10: Annual Percentage Change in AGI per Return (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans with 1992 and 1996 excluded (i.e. excluding cycles with unusually large third party vote share).

1.6% drop in annual percentage change in AGI / return from -0.7% at T+1 to -2.3% at T+2. Republican counties, on the other hand, show a sharper rate of decline with a 2.7% drop in annual percentage change from -0.05% at T+1 to -2.8% at T+2. As with the elections won by Republicans, the pattern shown here also supports the Legitimacy Hypothesis.

In order to support the analyses presented here and to examine the effect of the choice in classification approach, in Appendix B.3, Figure B.21 shows the identical graphs for Democratic and Republican counties classified using all the four approaches to classification of partisan identity. As with the current figures, Figure B.21 shows the data segregated by elections where the Democrats won the White House and elections where the Republicans won the White House. Figure B.22 does the same, but as with Figure 12.10, the data for 1992 and 1996 are omitted.

Comparing Effect of Election for Winners and Losers

As the final graph in the sequence, in Figure 12.11(a), I present the effect of the election on annual percent change in AGI / return collapsed across all elections. In the accompanying Figure 12.11(b), I also present the same pattern restricted to the post-2000 election samples. The counties are classified into winning (the county's preferred candidate won the White House) and losing counties (the preferred candidate lost). Winning counties are shown using the black lines with upward pointing triangles; losing counties are shown using the red lines with downward point triangles. The x-axis (T-2; T-1; T+1; and T+2) and the blue vertical marker are the same as the previous section. As before, the standard errors are cluster-robust and bias-adjusted for heteroskedasticity.

The reader may recall that county support for a presidential candidate can be measured by using the Plurality Standard or the Absolute Majority Standard (see Appendix B.1 for a discussion). Figure 12.11 uses the plurality standard, while Supplementary Figure B.24 presents a point of comparison with across both methods of classification. It is worth noting that using the Absolute Majority Standard produces a figure that is identical in the trend lines to Figure 12.11(b), where 1992 and 1996 are excluded. This is likely due to unusual strength of the 3rd party candidate in these elections —a feature which may also explain the unexpected trends seen in these elections.

In the lead up to the election, the average percent change for counties that supported the winning candidate drops from 1.1% to 0.7%, stays stable across the election, and then drops by 0.8% to -0.1%. For losing counties, the percent change initially rises from 0.1% to 0.8%, also stays stable across the election and then drop by 1.5% to -0.8% —which is approximately a 0.8% drop larger than the one seen for winning counties. This larger negative drop for losing counties is consistent with what we would expect according to the legitimacy hypothesis.

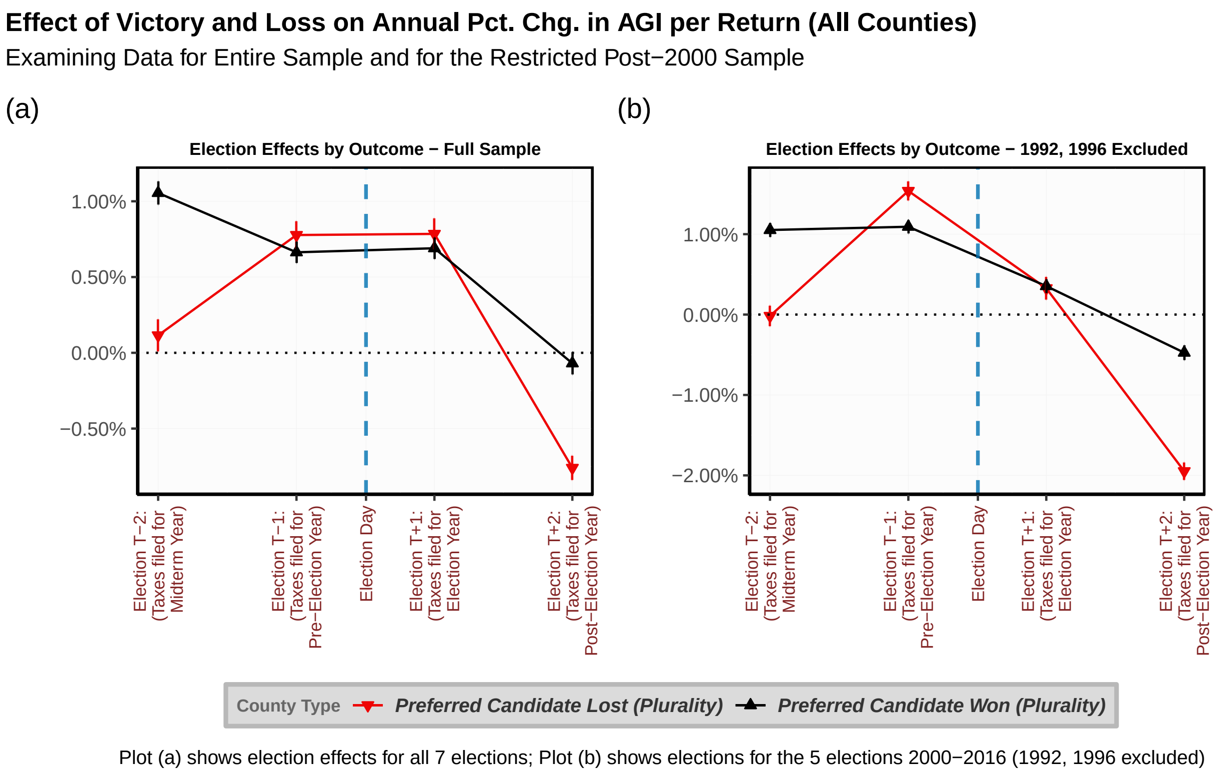


Figure 12.11: Annual Percentage Change in AGI per Return (Winsorized) for Counties that Supported the Winning Candidate vs those that Supported the Losing Candidate —Shown for the Full Sample and the Post-2000 Restricted Sample

The overall trends in the restricted sample tell a similar story, although the post-election movements are larger in size. In the lead up to the election, the average percent change for winning counties remains stable at 1.1% before declining steadily to 0.4% for the election year (T+1) and —after a drop of 0.8% —ending at -0.5% (T+2). For losing counties, as with the full sample, the percent change initially rises from -0.0% to 1.5% before dropping to 0.3% in the election year and then dropping by 2.3% to end at -1.9% at T+2. This drop is almost 2.7 times the one seen for winning counties for the post-election year, with an additional decline of 1.4% compared to the winning counties. Again, this pattern —with a larger negative drop for losing counties —is consistent with the legitimacy hypothesis.

Summary of Findings for AGI per Return

In order to examine the potential effects of election outcomes on Adjusted Gross Income per return, I primarily focused on the annual percentage change variant of the variable. Having looked at the percent change in AGI per return, there was clear indication of patterns that are consistent with the legitimacy hypothesis. For almost all elections (with the exception of 1992 and 1996), losing an election was accompanied with a sharper drop in the annual percent change in AGI per return when compared to the winning counties, albeit in the first tax year after the election. The effect did not appear to change based upon whether the Republican counties lost or the Democratic counties lost. The consistency of effects was reassuring as was the consistency of the effect size —which was approximately 1-1.5 additional percentage points drop for the losing counties.

The 1992 and 1996 elections were a potential source of concern, since they showed an effect in the opposite direction. One potential reason for this difference may be the unusually large role played by the 3rd party candidate for President, who won 19% of the vote in 1992 and 8% of the vote in 1996. As a result of this out-sized role, the use of the plurality standard (as opposed to the absolute majority standard), which was applied to classify counties for the loyalty standard, could have resulted in numerous potential misclassifications for 1992 and 1996 (this issue is examined in greater detail in Appendix B.1, specifically, see Table B.3, which shows that 2/3 of counties failed to reach a majority in favor of either candidate in 1992 and 1/3 of counties failed to do so in 1996). To understand why this could result in a reversal of the election effect, consider a simplified example. Imagine the case of a county where 40% voted for Bill Clinton, 35% voted for George W. Bush Sr., and 25% voted for Ross Perot), under the plurality standard, this county would be classified as having supported the Democratic candidate (since Bill Clinton received the most votes). Now, assume that everyone in the county would increase income disclosure by one dollar if their preferred candidate won and decrease disclosure by one dollar if their preferred candidate lost. Then, after Bill Clinton won, 40% of the people should increase their disclosure by \$1 and 60% of the people should decrease their disclosure by \$1, resulting in an overall decrease of \$20. Thus, this ostensibly "Democratic" county would end up decreasing their overall disclosure by \$20 in the aggregate, thus appearing to support the licensing hypothesis, even though it is driven by legitimacy consistent behavior. It should be noted: although this pattern is plausible, it requires further investigation at this stage.

Finally, regarding the main trends seen in this section, although the graphical analysis showed patterns that are consistent with the Legitimacy Hypothesis, the reader may have noticed that the election effect was only detectable for the post-election year (marked as T+2 on the aggregate graphs). This means I only found evidence of an election effect for the tax year that covered the first year under the new presidency. Although it was expected that the election effect could get stronger in the post-election year (see Section 11.1.3 Simulation - Varying the Time Dynamics of Election Effects), the identification strategy central to the approach here relied upon the ability to detect a smaller (but directionally consistent) effect in the election year itself (i.e. for the tax year during which the election took place). There does not appear to be evidence for such an election year effect —only the post-election year effect. The delayed timing poses a significant problem for the identification strategy used here, since there are potentially alternative explanations for the observed pattern —namely, the incoming president would have had one entire year to enact out policies that punish counties that voted against him. One could imagine many potential pathways for doing so, including, the withholding or delay of federal projects in certain districts or enacting of policies that inhibit growth in those areas (or, transfer growth to politically aligned areas).

The primary reason this timing issue threatens the identification strategy arises from the fact that the measure considered here, percent change in AGI per return cannot distinguish between changes in income disclosure and actual changes in income growth. Thus, given the fact that the election effect occurs after the first full year of the new Presidency, in order to make any conclusions (even preliminary ones) about the causal effect of election outcomes on income disclosure (as opposed to income earned), I would need to implement a strategy of controlling for an impressive array of economic variables that could reveal true changes in economic conditions on the ground. The entire identification strategy of the current approach was aimed at avoiding this pitfall by focusing on the election year as opposed to the post-election year, and thus was an attempt to circumvent the need for an exhaustive array of social economic and demographic controls.

To overcome this issue, in the next section, I turn to a variable that serves as a better proxy for tax compliance, namely, AGI over PI. If election based effects are driven by the economic consequence of partisan reallocation of Federal monies, then, this election-based economic punishment towards the unaligned counties or economic favor towards the aligned counties should also show up in the BEA's measurement of Personal Income (PI) for those counties. As a result, the ratio of AGI over PI should not be impacted. As the subsequent sections, I proceed with the same structure of analyses, although, I skip some of the initial graphs shown here, which can be found in the Supplementary Graphs in Appendix B.4 instead.

12.4.2 Trends in AGI over PI

Distribution of AGI over PI

In Figure 12.12, we see the distribution of AGI over PI represented both as a box-plot (where, the distribution can be seen for each year) and as a histogram (distribution collapsed across all years). A quick glance at both plots reveals some self-evident outliers, which show AGI over PI values that should not be possible (e.g. values of AGI over PI > 1). A review of these outliers reveals some notable discrepancies which should encourage us to remove them for the main analyses. For example, if we look at Washington County (KS), which is a prototypical case of such outliers, we see the number of returns jump from ~2600 in 2008 to ~11K in 2009 and ~12.5K in 2010 before dropping back to within its usual range of ~2400–2900. This two year surge in number of returns is accompanied by a quadrupling of the AGI from 100K in 2008 to $435K in 2009 to $475K in 2010 before dropping back to its usual range. There is, however, no similar jump in the BEA's estimates of income, which stay consistent and within the broader trend. The jump in reporting appears to be driven by temporary workers who were participating in the construction of the Keystone-Cushing pipeline which traversed the county from North to South. This case is exemplary of a number of similar situations where divergence from normal reporting patterns represents either reporting errors or transient and extraneous factors that do not represent on-the-ground changes in tax reporting behavior for the usual population of the county. The inclusion of such data would be inappropriate for any analyses that rely upon voting patterns / partisanship status of a county to identify election-based changes.

In order to prevent the spurious influence of such outliers, I implement a simple two-stage winsorizing strategy that first excludes all values outside the lower and upper bounds of 0–1 and then computes the 0.25% and 99.75% percentiles. The first stage cutoffs are shown by the bright red, dashed lines. The second-stage, computed percentiles are shown by the muted red, dotted lines. All values outside the dotted lines were set to equal the percentile thresholds, resulting in winsorized values which ranged from 0.3 to 0.9. The histogram on the right (plot b) shows the overall distribution. The small red bars near the vertical dotted lines show the quantity of observations that are produced at the thresholds as a result of the winsorizing procedure. The same procedure is applied for the Annual Change in AGI over PI, which is shown in Figure 12.13. After winsorizing, the Annual Change in AGI over PI values ranged from -0.16 to 0.17. The distribution of the winsorized variables is shown in Appendix B.4 Supplementary Graphical Analyses: AGI over PI, Figure B.25.

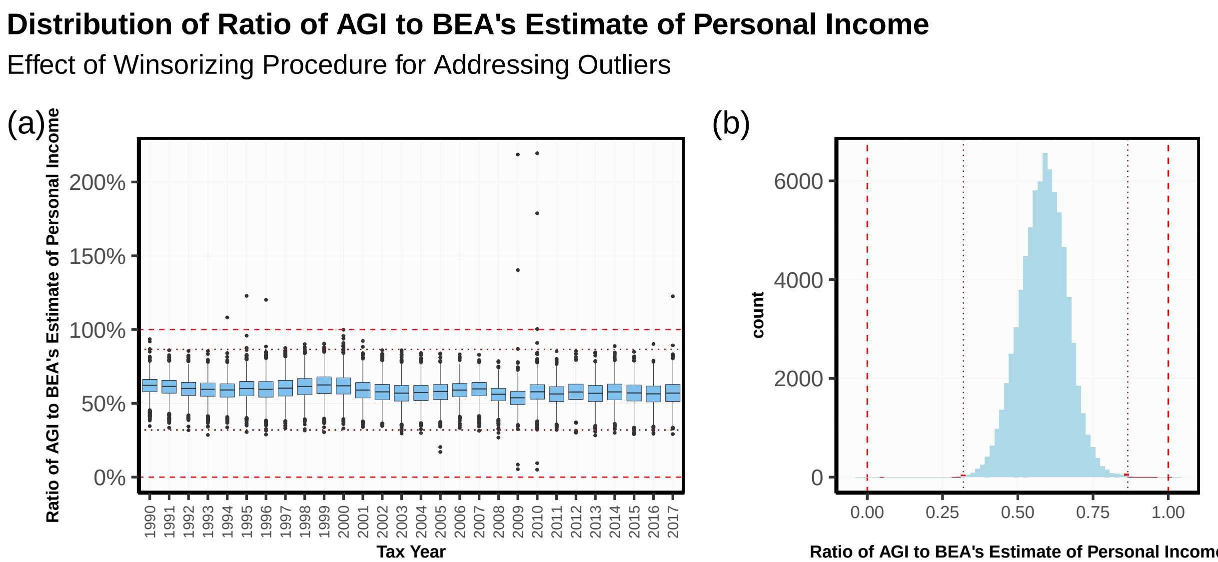


Figure 12.12: Distribution of Adjusted Gross Income / BEA's Estimate of Personal Income (AGI / PI) —Boxplot and Histogram Showing the Effect of Winsorizing Procedure for Addressing Outliers

Finally, in Figure 12.14, we see the distribution of AGI over PI and the Annual Change in AGI over PI compared after grouping counties according to the partisanship status. Overall, the distribution of AGI over PI seems to have similar properties across the county classification: Democratic Counties have a mean AGI over PI of 58% ($SD = 8$); Republican Counties have a mean AGI over PI of 59% ($SD = 8$); and unclassified counties have a mean AGI over PI of 58% ($SD = 7$). The median values are also similar, ranging from 58% to 59%. The central tendency in terms of modal frequency does appear to differ between Republican and Democratic counties, with Democratic counties most frequently having an AGI over PI near 63% and Republican counties most frequently having an AGI over PI near 59%. As we saw with AGI per return, the annual change variable is much better behaved and shows almost no differences across partisan classifications, with the mean annual change in AGI over PI

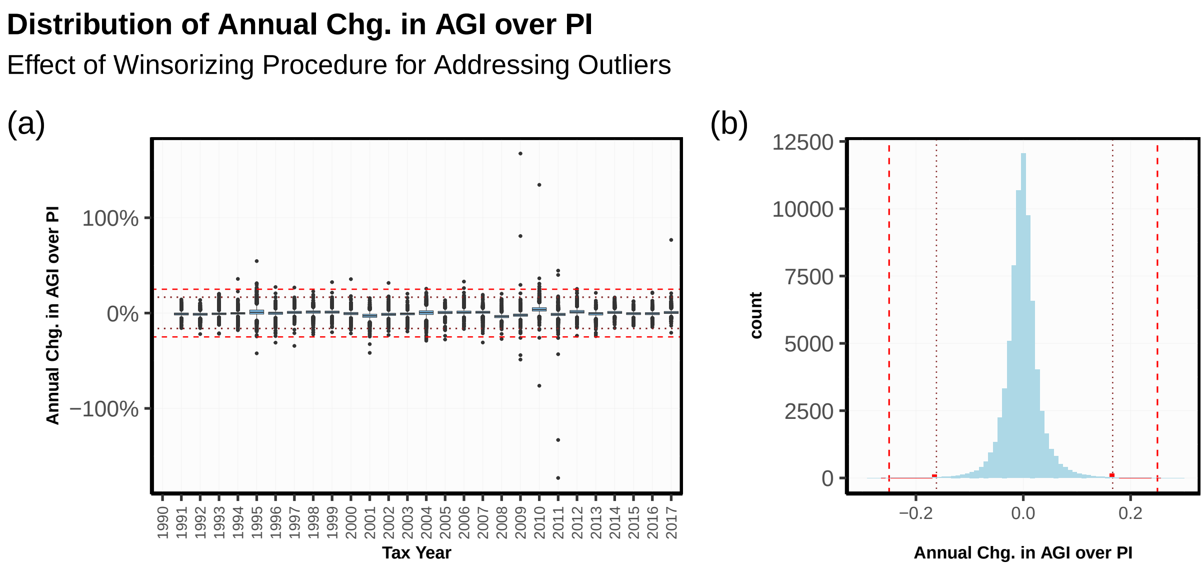


Figure 12.13: Distribution of Annual Change in Adjusted Gross Income / BEA's Estimate of Personal Income (AGI / PI) extemdash Boxplot and Histogram Showing the Effect of Winsorizing Procedure for Addressing Outliers

for each group ranging from -0.19% to -0.17%. As before, in Appendix B.4 Supplementary Graphical Analyses: AGI over PI, Figure B.26, I present the same distributions for each of the four classification approaches.

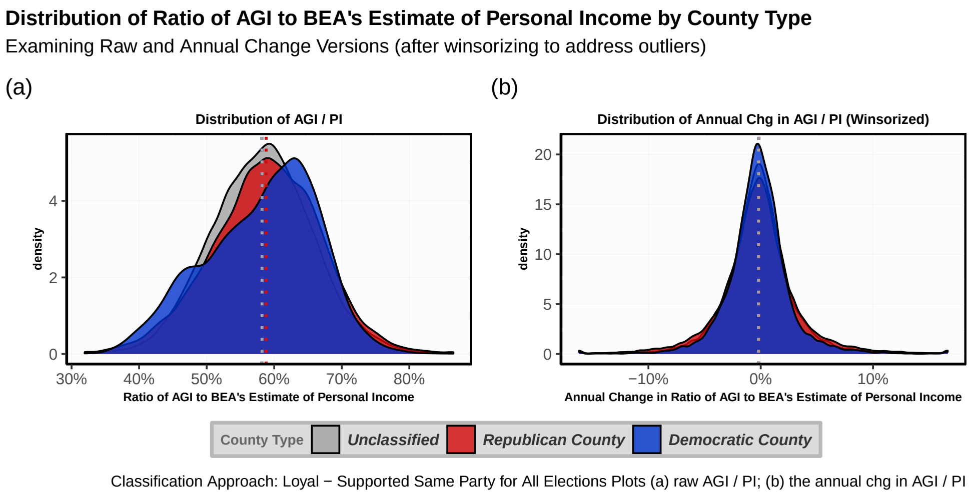


Figure 12.14: Distribution of Adjusted Gross Income / BEA's Estimate of Personal Income (AGI / PI) for all U.S. Counties Classified According to Partisanship

Full time series: AGI over PI

To begin, in Figure 12.15, I present AGI over PI for all counties classified by partisanship across the full 28 years in the sample. In general, it is clear that counties show similar patterns across classifications and there are unlikely to be any differential effects of election outcomes discernible at this level of detail. However, some overarching trends are clear: (a) AGI / PI systematically declines from 1990–1994 and then systematically increases from 1995–2001; (b) after 2001, there is a steep and steady decline in AGI over PI until 2003, after which it remains largely within a stable range —only showing a momentary departure from that range during the 2008–2009 financial crises. Although, the causes of these moments are difficult to pinpoint, the timing does align with major changes in tax law that impacted the IRS's definition of Adjusted Gross Income relative to the BEA's definition of Personal Income. Fortunately, for the current purposes, these changes seem to impact the ratio of AGI over PI in a similar manner independent of the partisan affiliation of a given county.

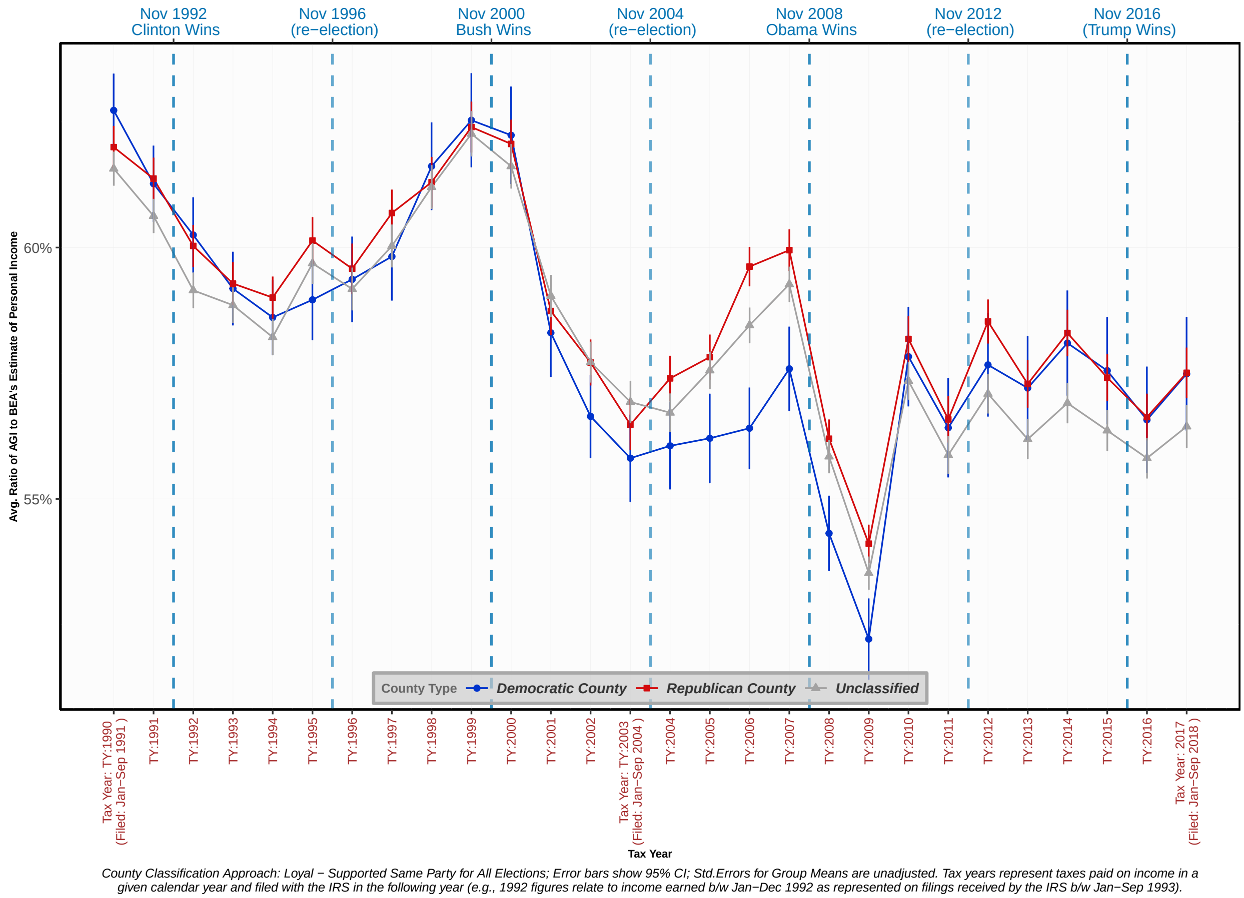


Figure 12.15: AGI over PI (Winsorized) for Democrat and Republican Counties over Time

Next, in Figure 12.16, I present the annual change in AGI over PI for each county type across the full 28 years in the sample (in the Supplementary Graphs in Appendix B.4, Figure B.27, I present the same time series with time fixed effects removed). It should be noted that the variable Annual Change in AGI over PI represents a raw difference between two years. This is different from the previous section where changes in AGI / return were expressed as percent changes and not as raw annual changes. However, since AGI over PI is already expressed in percentage terms, Annual Changes in AGI / PI are expressed in terms of raw changes as percentage points:

$$\textit{Annual Change in AGI / PI} = \Delta_{(AGI/PI)_t} = (AGI/PI)_t \; - \; (AGI/PI)_{t-1}$$

Although in Figure 12.16, no immediate overall trends are apparent at this level of detail, as expected, this variant of the AGI over PI variable seems to show more potential in terms of detecting differential effects of election outcomes across partisanship status. The reasons for this were discussed in greater detail in Section 10.3.2 Representing Variables as Difference and Percent Change Variants of the Original Levels. As a result, from this point forward, I focus solely on the annual change in AGI over PI, starting with the next figure.

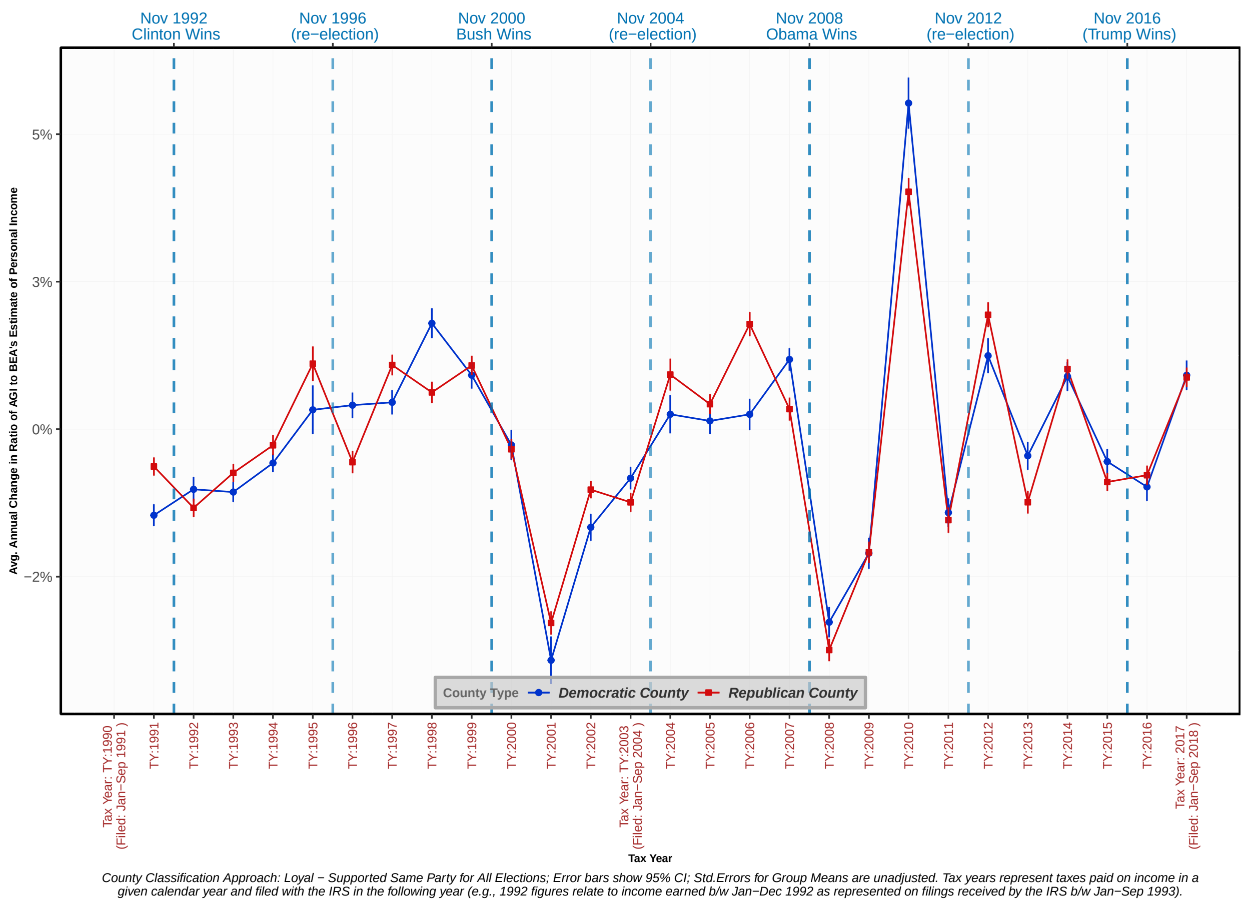


Figure 12.16: Annual Change in AGI over PI (Winsorized) for Democrat and Republican Counties over Time

Time Series for Elections Seperated by Republican and Democrat Victory - Annual Change in AGI over PI

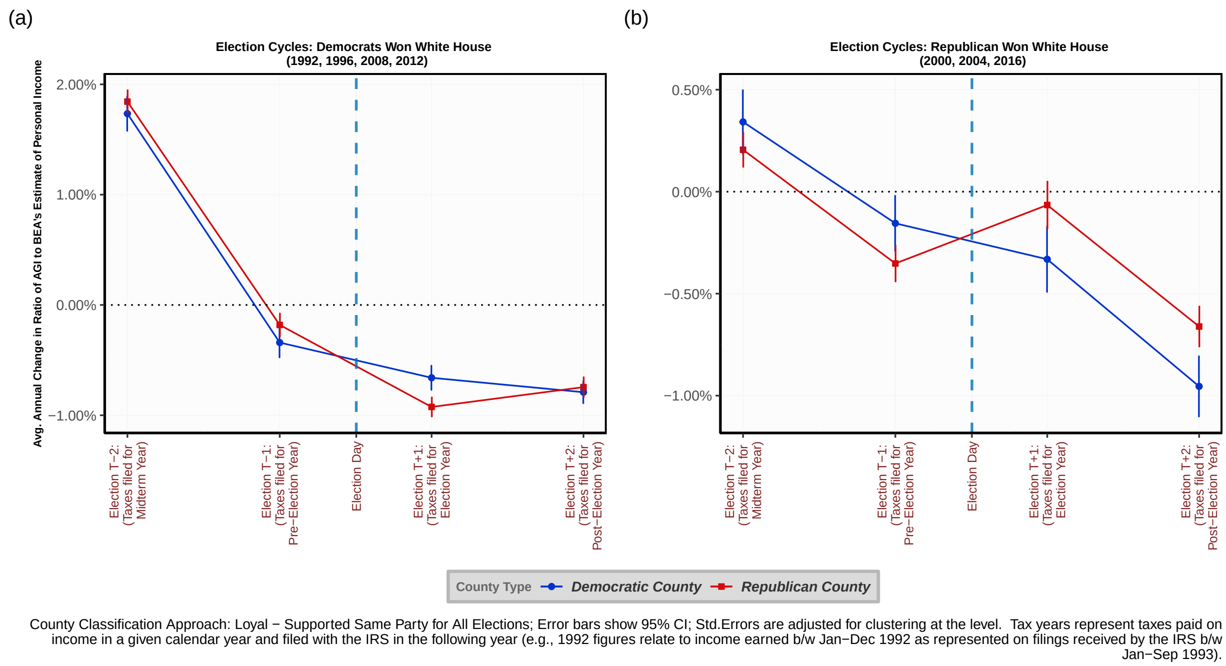


Figure 12.17: Annual Change in AGI over PI (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans

When comparing the effect of election outcomes on a cycle-by-cycle basis, we see a mixed series of findings, with most elections appearing to support the Legitimacy Hypothesis and a few appearing to support the alternative hypotheses. Rather than review each election here, I have presented the cycle-by-cycle graphs in Appendix B.4 Supplementary Graphical Analyses: AGI over PI, where, in Figure B.28, I present the Annual Change in AGI over PI for Democratic and Republican counties split by election cycle and in Figure B.29, I present the same with time fixed effects removed.[7]

[7]As before, in Appendix B.4.5 Difference in AGI over PI - Comparing Classification Approaches for each Election Cycle, I present the graph for each individual election cycle according to all four classification approaches side-by-side for comparison. This allows the reader to examine the effect of different classification approaches (for 1992 Election Cycle, see Figure B.30; for 1996 Election Cycle, see Figure B.31; for 2000 Election Cycle, see Figure B.32; for 2004 Election Cycle, see Figure B.33; for 2008 Election Cycle, see Figure B.34; for 2012 Election Cycle, see Figure B.35; for 2016 Election Cycle, see Figure B.36).

Instead, here, in Figure 12.17, I focus on examining the effects of election outcomes on Annual Change in AGI over PI collapsed across election cycles —split by election cycles where Democrats won the White House and election cycles where Republicans won the White House. As with prior figures, the dashed, blue, vertical line marks the November Presidential election. To the immediate left of the dashed line, at Election T-1, the values represent the annual change in AGI over PI for the pre-election tax year; to the immediate right of the dashed line, at Election T+1, the values represent the tax data for the election year; at Election T+2, the values represent the tax data for the first post-election year. The standard errors used to compute the confidence intervals shown in the graph are cluster-robust standard-errors computed using the "arellano method" with HC1-type of bias adjustment[8] for heteroskedasticity and clustering at the county-level.[9]

In Figure 12.17(a), I begin by examining data for the election cycles where the Democratic Candidate won the White House (1992, 1996, 2008, and 2012). At T-2, the annual change in AGI over PI for both Republican and Democratic Counties starts at approximately 1.84 and 1.73 percentage points, respectively. It then drops in parallel to -0.18 and -0.34 respectively for both types of counties. At that point, we see a strong election effect with a crossover and —in the election year —Democratic counties show an additional decrease in AGI over PI by -0.66 percentage points, while Republican counties show an additional decrease of 0.26 to end with -0.92 percentage points annual change in AGI over PI. Finally, in the post-election, both types of counties rejoin and show an annual change in AGI over PI of -0.8 percentage points. The steeper decline for Republican counties in the election year following a loss is consistent with the Legitimacy Hypothesis.

Second, in Figure 12.17(b), I present data for the election cycles where the Republican

[8] As recommended by Cameron & Miller (2015) and as implemented in prominent statistical software like STATA

[9] As recommended by Cameron et al. (2011); Esarey & Menger (2019); Bertrand et al. (2004)]. Cluster-robust standard errors were computed using the Sandwich package v. 2.5-1 (Berger et al., 2017; Zeileis, 2006) in R (Version 3.6.3; R Core Team, 2020)

Candidate won the White House (2000, 2004, and 2016). As before, at T-2, Democratic and Republican counties start at an annual change in AGI over PI of 0.34 and 0.21 respectively, before dropping in parallel to -0.16 and -0.35 respectively. As before, for the election year, we see a similar crossover pattern, with Republican counties showing a lower annual change in AGI over PI of only -0.07 and Democratic Counties continuing on a steeper negative trajectory with an annual change in AGI over PI of -0.33. As before, the steeper decline for Democrats after losing an election is consistent with the Legitimacy Hypothesis —however, the size of the standard errors make such conclusions tentative. In Supplementary Figure B.37, for other approaches to classification that result in larger samples for Democratic Counties, we see the same trend and smaller standard errors for the election year values. Finally, in the post-election year, trends for both types of counties travel in parallel —dropping by 0.7 each to a value of -0.7 and -1 for Republican and Democratic counties, respectively. In summary, in both the election cycles where Democrats won and the election cycles where Republicans won, the pattern of results appear to be consistent with the Legitimacy Hypothesis, where losing an election is accompanied by a steeper decline in the amount of income that is revealed to the IRS as a proportion of the overall income earned (as estimated by the BEA).

Finally, in order to examine the effect of classification approach, in Appendix B.4 Supplementary Figure B.37, I present the annual change in AGI over PI for Democratic and Republican counties classified according to all four approaches —segregated by elections where the Democrats won the White House and election where the Republican won the White House.

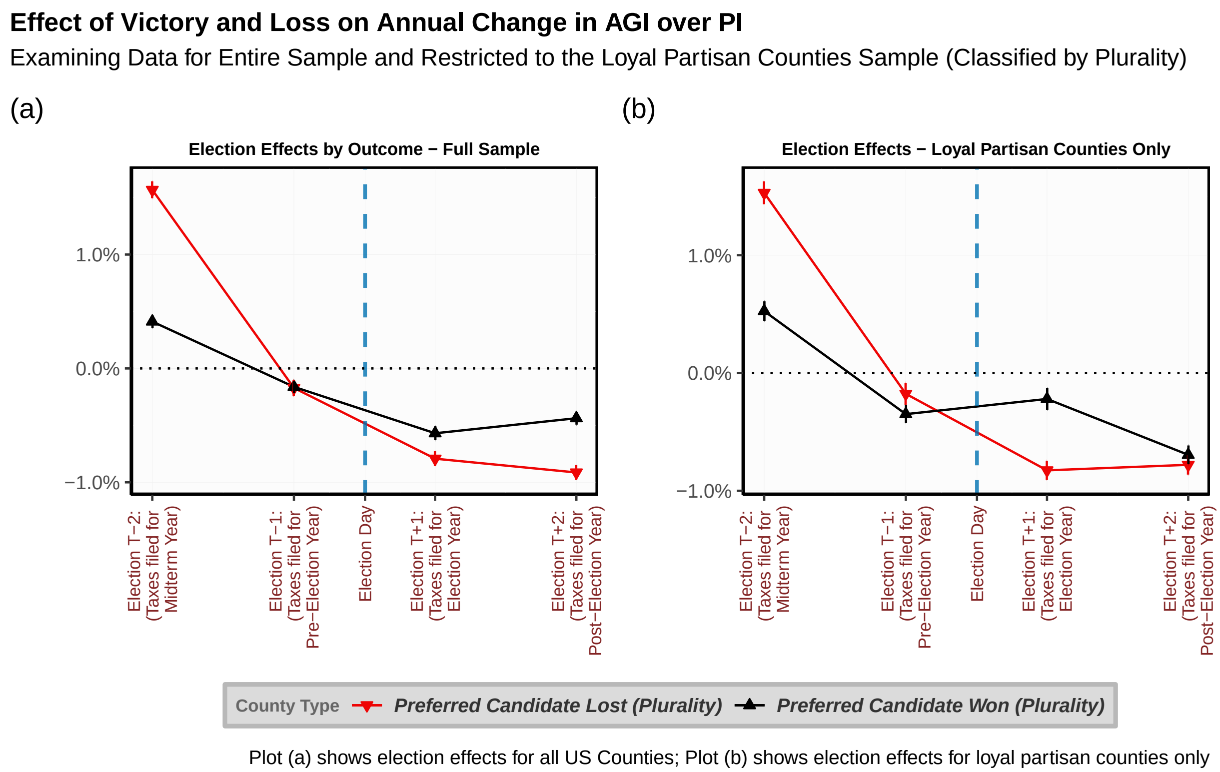


Figure 12.18: Annual Change in AGI over PI (Winsorized) for Counties that Supported the Winning Candidate vs those that Supported the Losing Candidate —Shown for the Full Sample and the Loyal Partisan Sample —Counties Classified Using Plurality

Comparing Effect of Election for Winners and Losers - Annual Change in AGI over PI

As the final graphs in the sequence, in Figure 12.18(a), I present the effect of the election on Annual Change in AGI over PI collapsed across all elections. In this figure, all U.S. counties are classified either as winning (the county's preferred candidate won the White House) or losing (the preferred candidate lost). In Figure 12.18(b), the same is shown —but the sample is restricted to only those counties which were loyal partisans (i.e. counties that voted for the same party in all 7 elections). Winning counties are shown using the black lines with upward pointing triangles; losing counties are shown using the red lines with downward point triangles. The x-axis (T–2; T–1; T+1; and T+2) and the blue vertical marker are the same as the previous section. As before, the standard errors are cluster-robust and bias-adjusted for heteroskedasticity.

The reader may recall that county support for a presidential candidate can be measured by using the Plurality Standard or the Absolute Majority Standard (see Appendix B.1 for a discussion). Figure 12.18 uses the plurality standard, while Supplementary Figure B.38 allows for a comparison of current trends classified using both the plurality and the absolute majority standards for classifying electoral outcomes.

In the lead up to the election, the average change in AGI over PI for counties that supported the winning candidate drops from 0.4% to -0.2% and then drops by an additional 0.4% to -0.6% in the election year, before increasing by 0.1% to -0.4%. For losing counties, the average change in AGI over PI also dropped from 1.6% to meet the level seen for winning counties in the pre-election year, before dropping by an additional 0.6% to end at -0.8% in the election year before continuing to decrease further to end at -0.9% post-election. The difference in election year drop for winning and losing counties is approximately 0.2%. By the post-election year, the winning counties had dropped by 0.3% relative to the pre-election measure, whereas the losing counties had dropped by 0.7% relative to the pre-election measure, which is approximately 2.7 times larger than the drop seen for winning counties. This larger negative drop for losing counties is consistent with what we would expect according to the legitimacy hypothesis.

In Figure 12.18(b), the sample is restricted to loyal partisan counties only. Even in this more restrictive sample, the overall trends tell a similar story, although the post-election movements are larger in size and more clearly defined. From the pre-election year (T-1) to the election year (T+1), the average annual change in AGI over PI for winning counties moves from -0.3% to -0.2%, a shift upward of 0.1%. For losing counties, during the same period, we see a shift from -0.2% to -0.8%, a shift downward of 0.6% —resulting in a overall difference in election effect of 0.8%. In the post-election period, we see both winning and

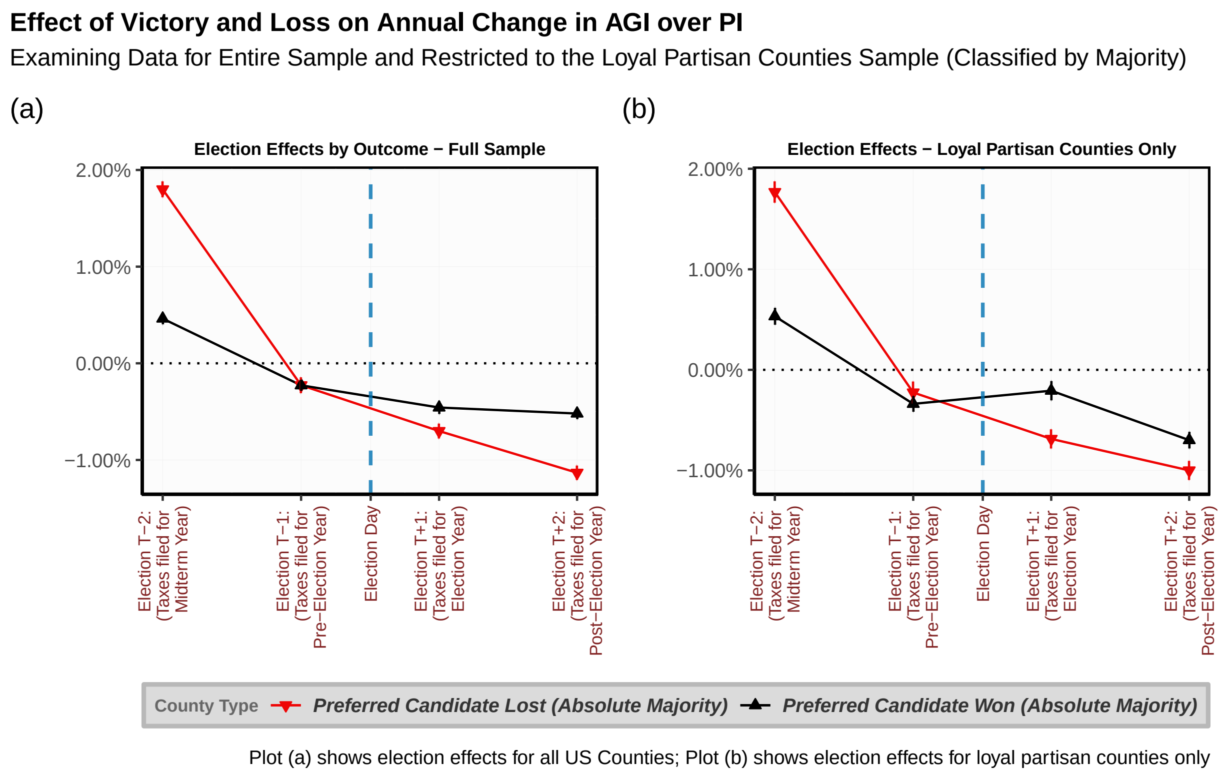


Figure 12.19: Annual Change in AGI over PI (Winsorized) for Counties that Supported the Winning Candidate vs those that Supported the Losing Candidate —Shown for the Full Sample and the Loyal Partisan Sample —Counties Classified Using Absolute Majority

losing counties return back to a similar annual change in AGI over PI of -0.7% for winning counties and -0.8% for losing counties. Again, this pattern —with a negative drop for losing counties and a minor position movement for winning counties —is consistent with the legitimacy hypothesis. The same analyses were conducted using a more stringent criterion for determining support —namely, an absolute majority of the county had to support one of the candidates before the county was classified as having won or lost. These analyses are presented in Figure 12.19. Having shown patterns that are consistent with the legitimacy hypothesis using the exploratory graphical analyses, in the subsequent section, I turn towards a more formal test of the Legitimacy Hypothesis using a simple linear model.

CHAPTER 13

STATISTICAL ANALYSES

Having examined the trends in the data using an exploratory, graphical approach in Chapter 12.4, I now present some basic, preliminary statistical analyses to test the main hypothesis under consideration. One of the primary trends seen in the previous section was the effect of election outcomes on the Annual Change in AGI over PI in the election year: losing the election was accompanied by a decrease in AGI over PI relative to the winning side. In the current section, I present statistical tests to examine whether these changes are statistically significant. The statistical tests were implemented as panel regression models after controlling for two-way year and state / county fixed effects. The test statistics were computed using adjusted standard errors that are robust to the effects of heteroskedasticity, serial (cross-sectional correlation) and appropriate for data clustered at the county level. These statistical tests were computed in R (R Core Team, 2020) using the LFE package (Gaure et al., 2019). All standard errors reported here were computed using the "arellano" method which allows for both heteroskedasticity and serial (cross-sectional) correlation (Zeileis, 2004) using the "HC1" estimator (MacKinnon & White, 1985) and were adjusted to account for clustering using procedures recommended by Cameron & Miller (2015). Unless noted otherwise, by default, the clustering of standard-errors was at the county-level, as recommended by Cameron et al. (2011); Esarey & Menger (2019); Bertrand et al. (2004).

In Section 13.1, I examine election effects using binary measures of election outcomes and in Section 13.2 I further examine whether the strength of the election effects varies as function of the margin of victory. To foreground the results, in all cases, the findings were statistically significant at or beyond the 0.05 level of significance: election outcomes appear to causally impact changes in AGI over PI in a manner consistent with the Legitimacy Hypothesis and these apparent election effects increase in strength as the margin of victory increases.

13.1 Binary Measures of Election Outcomes

13.1.1 Examining Differences By Election Cycles with Republican vs Democratic Victory

Table 13.1 compares the difference in mean Annual Change in AGI / PI for Democratic and Republican Counties across two sets of election cycles: elections in which Democrats won the White House (1992, 1996, 2008, 2012), which is shown in Column 1; and, elections in which Republicans won the White House (2000, 2004, 2016), which is shown in Column 2.

Table 13.1: Difference between Democratic and Republican Counties: Annual Change in AGI over PI

	Annual Change in AGI over PI	
	Elections where Dems Won	Elections where Reps Won
	(1)	(2)
Republican County	-0.311^{***}	0.241^{*}
	(0.082)	(0.115)
Observations	5,756	4,317
R^2	0.314	0.070
Adjusted R^2	0.307	0.059
Residual Std. Error	0.030 (df = 5703)	0.033 (df = 4265)
Note:	$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Data from loyal partisan counties only; Model includes fixed effects for State and Year; Column 1: Democrats won includes TY 1992, 96, 08, 12; Column 2: Republicans won includes TY 2000, 04, 16; SEs: heteroskedasticity-robust, clustered at County.	

In Column 1, we see that —after controlling for state and year fixed effects —for election years in cycles where Democrats won, Republican counties had an Annual Change in AGI /

PI that was -0.31 percentage points *lower* than that seen for Democratic counties ($\hat{\beta}_{Rep.Cty} = -0.311, \textit{clustered s.e.} = 0.082, p < 0.001$).

In Column 2, we see that —after controlling for state and year fixed effects —for election years in cycles where Republicans won, Republican counties had an Annual Change in AGI / PI that was +0.24 percentage point *higher* than that seen for Democratic counties ($\hat{\beta}_{Rep.Cty} = +0.241, \textit{clustered s.e.} = 0.115, p < 0.05$).

In other words, when Democrats won the White House, the *AGI / PI* for Republican counties *decreased* from pre-election to election year by an additional 0.3 percentage points relative to the change seen across the same period in Democratic counties. But, when Republicans won the White House, the *AGI / PI* for Republican counties *increased* from pre-election to election year by an additional 0.24 percentage points relative to the change seen across the same period in Democratic counties. This pattern of greater decrease after a loss and a greater increase after a victory is exactly what we would expect to see under the Legitimacy Hypothesis.

13.1.2 Examining Differences Between Victory and Loss Across All Election Cycles

Table 13.2 presents the same analyses collapsed across all 7 elections in order to test whether there is a statistically significant difference in mean Annual Change in AGI / PI for counties whose preferred candidate won vs lost the national election: Column (1) shows data for all counties; Column (2) shows data for loyal partisan counties only (i.e. counties that always voted for the same party in all 7 elections); Column (3) shows the results after amending Model 2 to include control variables.

Model 3 includes three controls variables. The first control variable is: the raw level of AGI / PI in the election year —this can be thought to capture features of a county's economic characteristics, specifically heterogeneity in the sources of income for residents in the county.

The second control variable is: annual percentage change in the AGI / return —which can be thought to capture changes in the economic performance of the county and changes in household incomes. The third control variable is a lagged version of the dependent variable.

The inclusion of a lag term allows us to address potential concerns regarding the temporal dependencies across observations for a county and to remove potential serial correlation in the resulting residuals (Beck & Katz, 1995, 2011; Wilkins, 2018; and, see Achen, 2000 for a cautionary note). However, prior to inclusion of a lagged dependent variable, it is necessary to verify that the dependent variable is stationary (Wilkins, 2018). As recommended, using the Dickey-Fuller test to check for stochastic trends in the dependent variable: Annual Change in AGI over PI, I reject the null hypothesis that the series has a unit root (i.e. non-stationary) —allowing us to confirm that the series are stationary, Dickey-Fuller = -52.2, Lag order = 1-3, p-value < 0.01. As a result, it is reasonable to use lagged dependent variables in the regressions shown here to address potential serial correlation in the model residuals. For the sake of completeness, in the Appendix in Section B.5.1 in Table B.4, the regressions shown in Column 1 and 2 were also conducted with a lag term —for both (i) a lag of 1 year and (ii) a lag of 1 and 2 years. Overall, the results shown here are robust to the inclusion of the lagged dependent variables and did not substantively impact the estimated coefficient or its statistical significance. It should also be noted that the estimator used default throughout this paper to compute the clustered, heteroskedasticity-robust standard errors in all regressions has also been shown to be consistent even in the presence of serial correlation in the residuals.

In Column 1, we see that —after controlling for county and year fixed effects —counties whose preferred candidate lost the Presidential Election had an Annual Change in AGI / PI that was -0.11 percentage points *lower* than counties whose preferred candidate won ($\hat{\beta}_{Losing.Cty} = -0.112, \textit{clustered s.e.} = 0.05, p < 0.05$). In column 2, we see that, for loyal counties, losing an election has an even larger effect, with an Annual Change in AGI /

Table 13.2: Difference between Win and Loss: Annual Change in AGI / PI

	Annual Change in AGI over PI		
	Full Sample	Loyal Partisans	
	(1)	(2)	(3)
Preferred Candidate Lost	-0.112^{*}	-0.265^{***}	-0.409^{***}
	(0.050)	(0.064)	(0.055)
AGI / PI			0.210^{***}
			(0.011)
Pct.Chg AGI/Ret			0.324^{***}
			(0.011)
Lag Chg. AGI/PI			-0.217^{***}
			(0.015)
Observations	19,271	10,073	10,073
R^2	0.298	0.319	0.608
Adjusted R^2	0.181	0.205	0.542
Residual Std. Error	3.086 (df = 16511)	3.171 (df = 8627)	2.407 (df = 8624)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model fixed effects: State, County and Year;
SEs: heteroskedasticity-robust, clustered at County;
Estimated coefficients shown in percentage points.

PI that was -0.27 percentage points *lower* for losing counties than for loyal counties whose preferred candidate won ($\hat{\beta}_{Losing.Cty} = -0.265, \textit{clustered s.e.} = 0.064, p < 0.001$). In other words, after an election, the *AGI / PI* for counties whose preferred candidate lost decreased by an additional 0.26 percentage points relative to the change seen across the same period for counties whose preferred candidate won the Presidential Election.

Finally, in column 3, we see that, after including control variables that (i) capture features of a county's economic characteristics (raw AGI / PI); (ii) capture changes in a county's household income (annual percent change in AGI / return); (iii) condition on change in AGI / PI in the previous year (lag annual change in AGI / PI), the estimated effect of losing an election appears to be even larger. Holding all else equal, counties that supported the losing candidate were predicted to have an Annual Change in AGI / PI that was -0.41 percentage points *lower* than counties that supported the winning candidate ($\hat{\beta}_{Losing.Cty} = -0.409, \textit{clustered s.e.} = 0.055, p < 0.001$). Overall, the pattern found in all three models —namely, a larger decreases in AGI/PI after a loss —is exactly what we would expect to see under the Legitimacy Hypothesis.

Although the regression analyses confirm that the difference in mean Annual Change in AGI / PI is statistically different after victory or loss, it can be helpful to examine the distribution and not just its central tendency (i.e. its mean). In Figure 13.1, I present the distributions of Annual Change in AGI / PI for winning and losing counties overlapped.

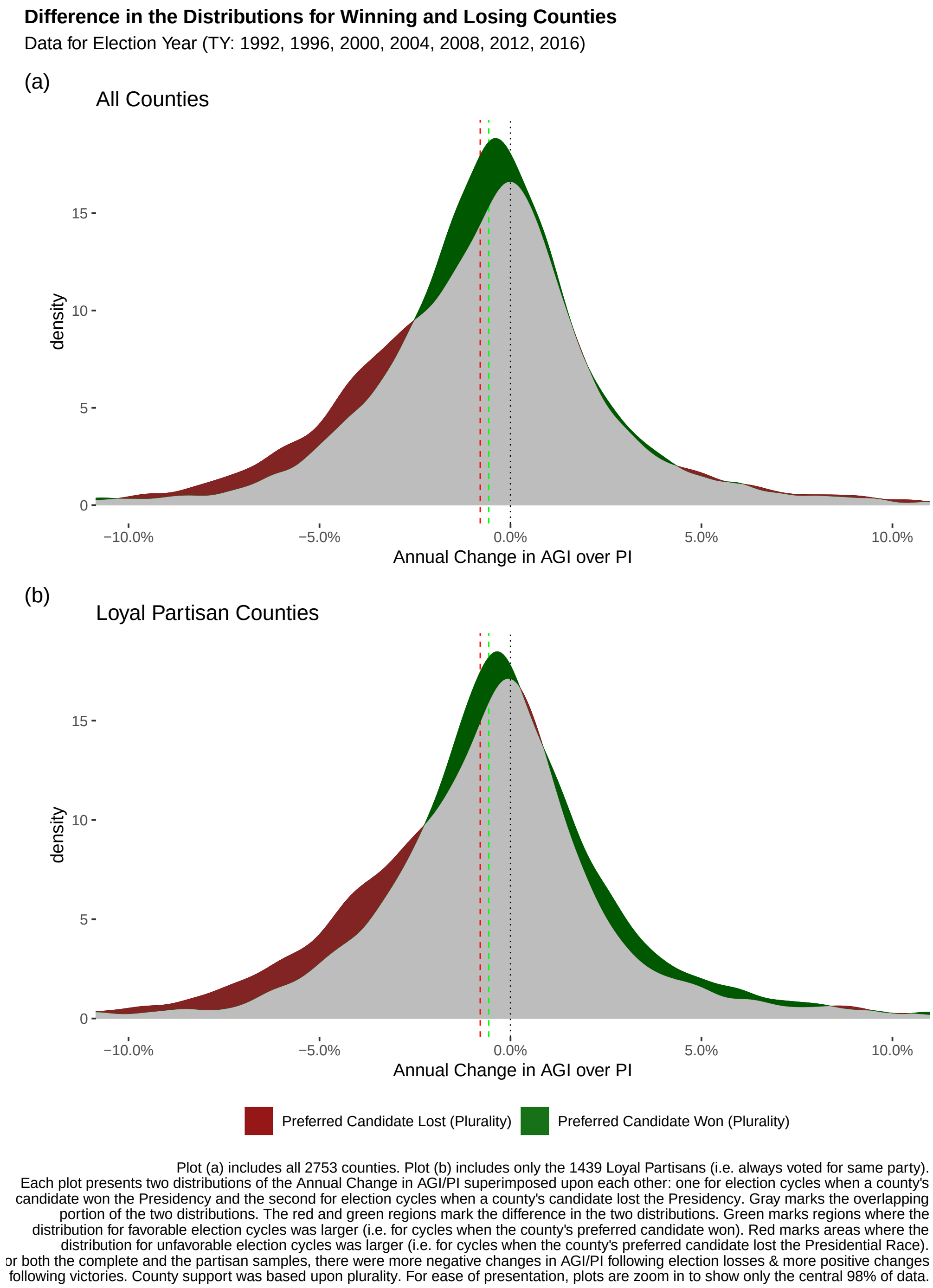


Figure 13.1: Difference in Distribution of Annual Change in AGI over PI for Winning and Losing Counties —Shown for Both the Full Sample of Counties and for the Loyal Partisan Sample —Victory and Loss Classified Using Plurality Threshold

Figure 13.1 (a) shows data for all 2753 counties in the sample, and Figure 13.1 (b) shows data for the 1439 loyal partisan counties that have been the primary focus of analysis. In both figures, the gray section represents the common portions of the two distribution; the red section show surplus mass in the distribution for losing counties; the green section shows surplus mass in the distribution for winning counties. The green and red dashed lines show the mean Annual Change in AGI over PI for winning and losing counties respectively, while the dotted black line separates the negative change from positive change (i.e. marks 0).

By examining the differences in the two distributions, it is clear that the surplus mass for losing counties is largely concentrated towards more negative values, whereas the surplus mass for winning counties is largely concentrated towards less negative values and positive values. This pattern is even more pronounced when we limit the analysis to loyal counties only, as shown in Figure 13.1 (b). In both Figures 13.1 (a) and (b), it is clear that the difference between winning and losing counties, as represented in Table 13.2, arises from a distinct increase in the number of losing counties with a more negative change in AGI over PI across the pre-election to post-election period (red areas are concentrated towards negative values) and a corresponding increase in the number of winning counties with either a positive change or a less negative change in AGI / PI during the same period (green areas are concentrated towards less negative or positive values). By examining the differences in the distribution, it is possible to verify that the election effect is not simply driven by a handful of outliers, but instead represents a broader shift across a wide range of the data.

13.2 Continuous Measures of Election Outcomes

In the previous section, we saw evidence of an election effect consistent with the Legitimacy Hypothesis such that winning counties showed a relative increase in tax compliance —as inferred by the positive change in AGI over PI from the pre-election to the post-election year for winning counties relative to ones that lost. However, if election-based changes in legitimacy are driving tax compliance, then —as reasoned in Section 9.1.1 Hypothesis 1(a) —one should expect the effect of the election on tax compliance to be determined by the level of partisanship. At the county-level, this would imply that the effect of the election on AGI over PI should be determined by the margin of victory or loss (which, serves as proxy measure for the aggregate partisanship of the county). To understand why, consider the stylized example of a county with 100 people where each person will either increase their income disclosure by $1 if their candidate wins or decrease their income disclosure by $1 if their candidate loses. Now consider the case of a county with 51 Democrats and 49 Republicans. When a Democrat wins the White House, the 51 Democrats would increase their income disclosure by $1 and the 49 Republican would decrease their income disclosure by $1, resulting in an increase in aggregate income disclosure of $2 ($51 − $49). Similarly, for a county which has 75 Democrats and 25 Republicans, under a Democratic victory, the aggregate income disclosure would increase by $75 − $25, or $50. Thus, under this aggregation framework, as margin of victory increased by 1 percentage point, the income disclosure would increase by $2 (and, vice versa in case of loss).

The current Section 13.2 Continuous Measures of Election Outcomes examines this corollary. The organization of this section remains the same as the previous section. First, in Table 13.3, I combine data across all 7 elections and examine whether the election effect observed in Table 13.2 varies as a function of the signed margin of victory using a simple univariate model. Then, in Table 13.4, I separately examine the effect of margin of victory for winning and losing counties after controlling for the main effect of election outcome. I also extend this model by including other control variables, including the base level of Adjusted Gross Income and the annual percentage change in AGI per return. To foreground the results, in all cases, the findings were statistically significant at the 0.05 level or lower. Overall, as predicted, the magnitude of the election effect varied as a function of the margin of victory and the direction of the effect was consistent with the Legitimacy Hypothesis.

13.2.1 Signed Margin of Victory - By Win Loss

In Table 13.3, I present the results of fitting a linear model regressing Annual Change in AGI over PI onto a continuous measure of election outcome —controlling for County and Year fixed effects. As before, significance tests were computed in R (R Core Team, 2020) using the LFE package (Gaure et al., 2019), which implements the clustering procedures for standard errors described in Cameron & Miller (2015).[1]

$$\Delta_{AGI/PI_{cy}} = \underbrace{\alpha_c * D_{County} + \underset{[Y=92,96,...,12,16]}{\gamma_y * D_{Year}}}_{\text{Fixed Effect}} + \underbrace{\beta_1 * \textit{Signed Margin of Victory}}_{\text{Election Effect}} + \epsilon_{cy}$$

The terms D_{County} and D_{Year} refer to dummy variables, where α_C and γ_Y capture the county and year fixed effects. The result of the election is represented in terms of Signed Margin of Victory. As was described in Section 10.2.1, the variable "Signed Margin of Victory" captures both: (i) whether the county's preferred candidate won the White House (sign); and (ii) the degree of support/opposition (margin of victory). For ease of reference, the formula is reproduced below:

$$\textit{Signed MoV} = \frac{(\textit{Votes}_{\textit{National Winner}} - \textit{Votes}_{\textit{Runner Up}})}{\textit{Votes}_{\textit{Total Cast by County}}} \begin{cases} \textit{Winning Counties:} & 0 \text{ to } +1 \\ \textit{Losing Counties:} & 0 \text{ to } -1 \end{cases}$$

This variable could be thought to capture the degree to which the county is aligned with

[1] The test statistics were computed using adjusted standard errors that are robust to the effects of heteroskedasticity, serial (cross-sectional) correlation) and appropriate for clustered data. All standard errors reported here were computed using the "arellano" method which allows for both heteroskedasticity and serial (cross-sectional) correlation (Zeileis, 2004) using the "HC1" estimator (MacKinnon & White, 1985) and were adjusted to account for clustering using procedures recommended by Cameron & Miller (2015). Unless noted otherwise, by default, the clustering of standard-errors was at the county-level, as recommended by Cameron et al. (2011); Esarey & Menger (2019); Bertrand et al. (2004).

and supports the winner of the Presidential Election. This measure ranges from +1 to -1. A Signed Margin of Victory of +1 would indicate that the county cast 100% of their ballots in favor of the winner of the presidential election. Conversely, a Signed Margin of Victory of -1 would indicate that 100% of county ballots were cast in favor of the candidate that lost the national election. In practice, the Signed Margin of Victory ranged from -0.6 to +0.7 in the 1990s and -0.9 to +0.9 in 2012 and 2016.

Table 13.3: Signed Margin of Victory (Plurality) and Annual Change in AGI over PI

	Annual Change in AGI over PI		
	Full Sample	Restricted Sample	Loyal Partisans
	(1)	(2)	(3)
Signed Margin of Victory	0.334***	0.419***	0.361***
	(0.092)	(0.084)	(0.103)
Observations	19,271	18,494	10,073
R^2	0.299	0.349	0.319
Adjusted R^2	0.182	0.235	0.205
Residual Std. Error	0.031 (df = 16511)	0.025 (df = 15735)	0.032 (df = 8627)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model includes fixed effects for County and Year;
SEs: heteroskedasticity-robust, clustered at County-Level;
Sample: central 98% of the data for both IV and DV;
Estimated coefficients shown in percentage points.

Table 13.3 shows the same model fit over three samples in increasing order of restrictiveness. In Column 1, the regression is fit on data from all 2753 counties. In Column 2, the regression is fit on the central 98% of the data (i.e. the highest and lowest 1% of annual change in AGI over PI values are excluded as are the top and bottom 1% of Signed Margin of Victory) to control against the possibility that outliers or unusual values are casting undue influence on the measured relationship between the two variables. And, finally, in Column 3, the regression is fit on the 1439 "loyal partisan" counties that always voted for the same party. In all three cases, the overall model is significant ($R^2 \neq 0$) and the estimated coefficient on Signed Margin of Victory is both statistically significant and is consistent with the Legitimacy Hypothesis. In the current set of analyses, victory of a candidate at the county level is decided using the plurality threshold. However, in Appendix B.5.2, the same analyses are conducted using the majority threshold for determining the winning candidate —the results remain unchanged.

Starting with the sample of all counties, we see that the overall significance for the "projected model" (i.e. the model fit after removing the effects of county and year fixed effects) was statistically significant, $F(1, 2752) = 13.11; p < 0.001; N_{Obs} : 19,271; FEs : N_{Counties} = 2753; N_{Years} = 7$. The effect of Signed Margin of Victory is positive and significant ($\hat{\beta}_{Sign.MoV} = 0.334$, $clustered\ s.e. = 0.092, p < 0.001$). If we examine the county with the widest margin of victory in the sample (i.e. Signed Margin of Victory = 87.5%), the model would predict that if the preferred candidate had lost the Presidency —the county's AGI over PI would have decreased by 0.58 percentage points relative to the case when the preferred candidate won. This represents the largest effect size we would predict for observations seen in the data sample. Now, let us consider the difference between win and loss for a county with an average margin of victory / loss. The average margin of victory in our sample was 24.1%. Thus, the predicted Annual Change in AGI over PI for this average county if their preferred candidate won would be 0.16 percentage points higher than it would be if their preferred candidate lost the Presidency.

In Model 2, the same linear model was fit on a more restricted set of observations. Since there was minimal data-exclusion conducted in the current analyses, this restriction in sample was conducted to guard against the potential effects of extreme values and outliers. In the sample for Model 2, the top and bottom 1% of observations for both the response variable and independent variable were excluded before fitting the model. As such, only observations with an Annual Change in AGI over PI between -9.9% and 10.0% and a Signed Margin of Victory between -63.2% and 69.1% were retained. The model fit on this restricted sample remained statistically significant, $F(1, 2751) = 24.94; p < 0.001; N_{Obs} : 18,494; FEs : N_{Counties} = 2752; N_{Years} = 7$, as did the estimated beta coefficient, ($\hat{\beta}_{Sign.MoV} = 0.419$, *clustered s.e.* $= 0.084, p < 0.001$) —suggesting that extreme values are not driving the relationship between the two variables. In fact, the predicted election effect appears to increase after outliers were excluded. For the most partisan counties in the restricted sample, which had a Signed Margin of Victory of 69.1%, the AGI / PI would be predicted to be 0.58 percentage points higher under a favorable election outcome than under an unfavorable one.

Finally, in Model 3, the analyses are restricted to the "loyal partisan" counties only. Since these counties always voted for the same party, for this sample, the election outcome satisfies the exogeneity requirement by construction (i.e., since these counties never alter their party preference, the favorability of the election outcomes for these counties in any given election is determined by residents of counties external to the loyal partisan sample). In this loyal partisan sample, the overall model remained statistically significant, $F(1, 1438) = 12.42; p < 0.001; N_{Obs} : 10,073; FEs : N_{Counties} = 1439; N_{Years} = 7$, as did the estimated beta coefficient, ($\hat{\beta}_{Sign.MoV} = 0.361$, *clustered s.e.* $= 0.103, p = 0.001$). The predicted effect of election outcome also seems very similar to what was seen for Model 1, with the predicted difference between winning and losing for the most partisan county (i.e. Signed Margin of Victory = 87.5%) is 0.63 percentage points.

The linear relationship between these variables is shown in Figure 13.2 as a scatter-plot with Signed Margin of Victory on the X-axis and Annual Change in AGI over PI on the Y-axis. The red dashed line demarcates 0, separating observations where the preferred candidate lost from observations where the preferred candidate won. The solid blue line shows the fit estimated by Model 1 shown in Table 13.3. In order to facilitate presentation, the plot is zoomed in and only shows the central 95% of values for Annual Change in AGI over PI. As the blue fit line shows, when the county's preferred candidate wins an election

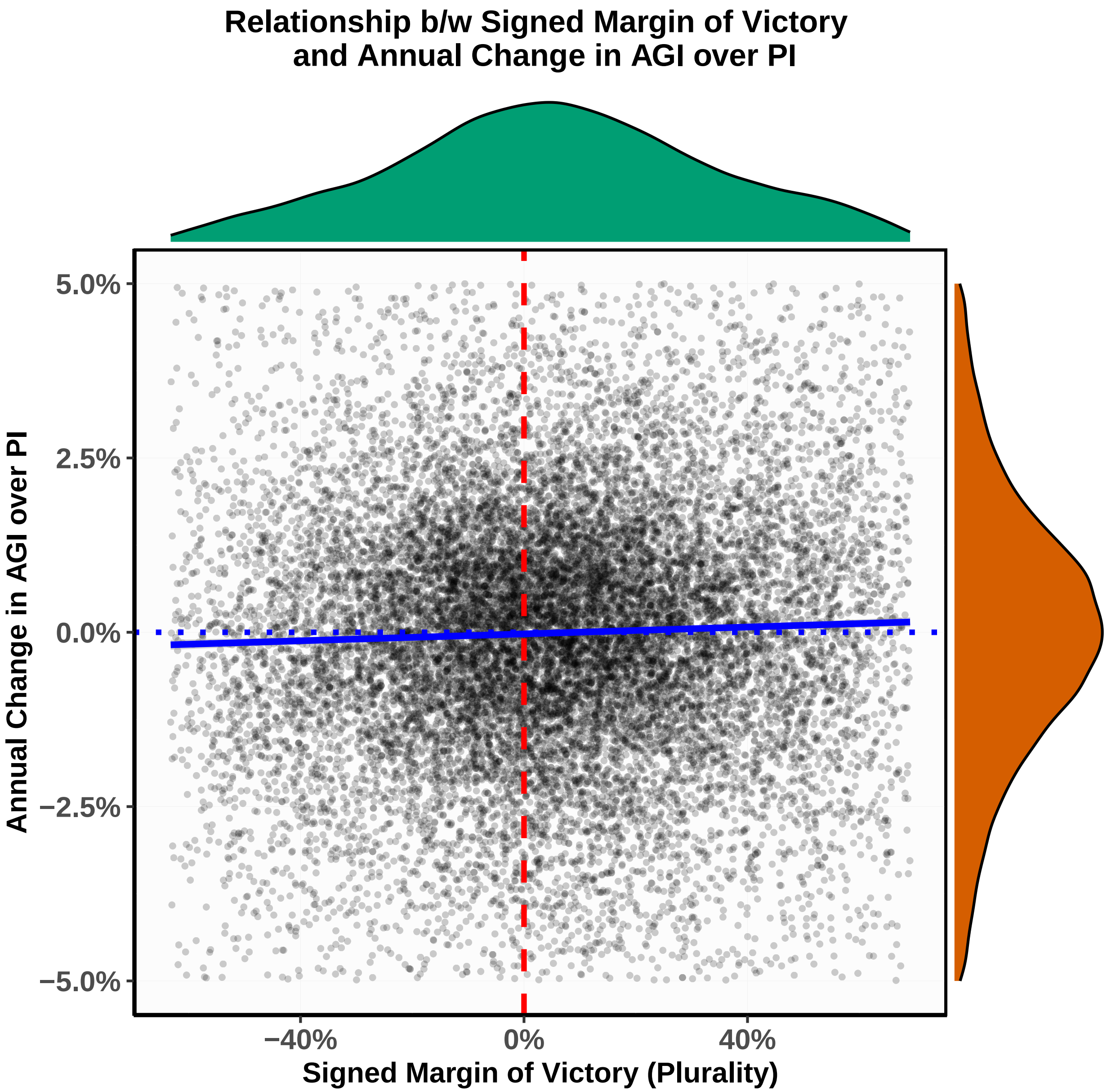


Figure 13.2: Relationship between Signed Margin of Victory and Annual Change in AGI over PI for Election Year Data

—increasing margin of support is associated with increases in AGI over PI after the election relative to the previous year; conversely, when the county's preferred candidate loses —increasing margin of support is associated with decreases in AGI over PI relative to the previous year.

13.2.2 Margin of Victory - By Win Loss

The linear model estimated in Table 13.3 using a simple univariate model provided initial evidence linking margin of victory to changes in AGI / PI pre-election vs post-election. However, the univariate structure imposed the artificial constraint that the effect of victory and loss must be equal and opposite. It also did not separate and control for the main effect of victory or loss as possibly distinct from the effect of margin of victory. To correct for these shortcomings of the univariate regression, in this section, I present the results from fitting the following model:

$$\Delta_{AGI/PI_{cy}} = \underbrace{\alpha_c * D_{County} + \underset{[Y=92,96,00,04,08,12,16]}{\gamma_y * D_{Year}}}_{\text{Fixed Effect}} +$$

$$\underbrace{\beta_1 * I\,(Lost) + \beta_2 * MoV + \beta_3 * I\,(Lost) * MoV}_{\text{Election Effect}} + \epsilon_{cy}$$

As before, the terms D_{Year} and D_{County} refer to indicator variables and α_C and γ_Y are used to capture the year and county fixed effects. The term $I\,(Lost)$ represents an indicator variable such that $I\,(Lost) = 1$ when the county's preferred candidate (by plurality) lost the national election and 0 otherwise. The term MoV represents the margin of victory and is equal to the difference between the top vote-getter in the county and the runner-up in the county —thus, it can range from 0 to +1. When the preferred candidate wins the national election, β_2 captures the effect of victory as a function of the strength of support for the candidate. When the preferred candidate loses the national election, $\beta_2 + \beta_3$ captures the effect of the margin of loss —while β_1 captures the main effect of losing the election. Thus, the coefficient β_3 indicates how the effect of MoV differs for a county when their preferred candidate wins vs loses the Presidential Election.

For the main hypothesis under consideration, it is necessary to examine whether $\beta_3 \neq$

0, since that would imply that there is a difference in the effect of margin of victory for counties when their preferred candidate wins the election vs when the candidate loses. More specifically, under the Legitimacy Hypothesis, one should expect to see $\beta_3 < 0$. The intuition here is that if $\beta_3 < 0$, ignoring any main effect, for any given margin of victory, when the preferred candidate loses there is either a smaller positive change ($|\beta_3| < |\beta_2|$) or a negative change in AGI / PI ($|\beta_3| > |\beta_2|$) compared to when the preferred candidate wins. As a result, $\beta_3 < 0$ can be seen as consistent with the Legitimacy Hypothesis.

Table 13.4: Annual Change in AGI / PI: Election Outcome x Margin of Victory

	Annual Change in AGI over PI			
	Full Sample			
	(1)	(2)	(3)	(4)
I(Lost) ($\widehat{\beta_1}$)	0.131	0.196^{*}	0.344^{***}	0.326^{***}
	(0.081)	(0.079)	(0.067)	(0.065)
MoV ($\widehat{\beta_2}$)	1.211^{***}	1.135^{***}	1.812^{***}	1.710^{***}
	(0.241)	(0.248)	(0.199)	(0.198)
AGI / PI		0.274^{***}	0.163^{***}	0.211^{***}
		(0.009)	(0.008)	(0.009)
Pct.Chg AGI/Ret			0.374^{***}	0.337^{***}
			(0.008)	(0.008)
Lag Chg. AGI/PI				-0.215^{***}
				(0.011)
I(Lost) x MoV ($\widehat{\beta_3}$)	-0.792^{**}	-1.582^{***}	-2.478^{***}	-2.266^{***}
	(0.305)	(0.290)	(0.255)	(0.247)
Observations	19,271	19,271	19,271	19,271
R^2	0.299	0.387	0.567	0.595
Adjusted R^2	0.182	0.284	0.494	0.528
Residual Std. Error	3.084 (df = 16509)	2.884 (df = 16508)	2.425 (df = 16507)	2.344 (df = 16506)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model includes fixed effects for County and Year;
SEs: heteroskedasticity-robust, clustered at County-Level;
Models shown least-to-most restrictive vis-a-vis controls;
Estimated coefficients shown in percentage points.

In Table 13.4, I show the results of fitting four models to data from the full sample of counties —with each model adding additional control variables. For the sake of completeness, in Appendix B.5.2 Table B.6, the same four models are fit on data limited just to the 1439 loyal partisan counties. The pattern of results are largely unchanged and the conclusions substantively identical to the results presented here. In all models, we find support for the Legitimacy Hypothesis: winning an election is associated with increases in AGI / PI relative to losing and the size of the effect is determined by the degree of support for the candidate.

In both Table 13.4 and Supplementary Table B.6, the table columns are organized as follows. Column 1 shows the estimated coefficients from the "basic model" specified in the equation above. Column 2 shows the results when the basic model is amended to include the raw AGI/PI in the election year as a control variable —which allows the model to capture capture aspects of a county's economic characteristics, specifically capturing heterogeneity in the sources of income for across different counties. Column 3 shows the results of amending Model 2 by adding the *Annual Percent Change in AGI per Return* for the election year —which allows the model to control for changes in the economic fortunes of the county, especially changes in household incomes. Finally, Column 4 shows the results of amending Model 3 to control for the lag term (i.e. controlling for the *Annual Change in AGI over PI* in the pre-election year). The inclusion of the lagged version of the dependent variable allows the model fit to address potential concerns regarding the temporal dependencies across observations for a county and to remove potential serial correlation in the resulting residuals (Beck & Katz, 1995, 2011; Wilkins, 2018). In all four models , the estimated coefficient on the variable of interest (i.e. $\widehat{\beta_3}$ i.e. $\widehat{\beta}_{I(Lost)\ x\ MoV}$) was negative and remained statistically significant ($p < 0.01$). Thus, the results suggest that the national election outcome determines how support for a candidate relates to tax compliance immediately following the election —with the direction of the estimated effects remaining consistent with the Legitimacy Hypothesis.

For the basic model in Column 1, the overall model fit for the "projected model" (i.e. the model fit for the election effect variables after removing the county and year fixed effects) was statistically significant, $F(3, 2752) = 9.83; p < 0.001; N_{Obs} : 19, 271; FEs : N_{Counties} = 2753; N_{Years} = 7$. Along the first row of Table 13.4, the coefficient estimates for the "main effect" of election loss ($\widehat{\beta_1}$) can be found for each of the four models. For the basic model (Column 1), we see little statistical evidence of a main effect of election loss, ($\hat{\beta}_{I(Loss)} = 0.13$, *clustered s.e.* $= 0.08, p = 0.12$). However, as additional control variables are added, the coefficient estimate becomes statistically significant starting in Model 2, ($\hat{\beta}_{I(Loss)} = 0.2$, *clustered s.e.* $= 0.08, p = 0.01$), and the estimated effect doubles in magnitude for Model 3 and Model 4, ($\hat{\beta}_{I(Loss)} = 0.33$, *clustered s.e.* $= 0.06, p < 0.001$).

Along the second row of Table 13.4, the coefficient estimates for the effect of Margin of Victory ($\widehat{\beta_2}$) are presented. This coefficient captures the relationship between Margin of Victory and the Annual Change in AGI over PI for counties when their preferred candidate won the Presidency. For all four models, the estimated coefficient is positive and statistically significant, starting with Model 1, ($\hat{\beta}_{MoV} = 1.21$, *clustered s.e.* $= 0.24, p < 0.001$) and ending with Model 4, ($\hat{\beta}_{MoV} = 1.71$, *clustered s.e.* $= 0.2, p < 0.001$).

Finally, along the bottom-most row of Table 13.4, coefficient estimates for the "interaction effect" between election loss and margin of victory ($\widehat{\beta_3}$) can be read. This coefficient captures the difference in the effect of margin of victory for a county when its preferred candidate wins the White House relative to when its preferred candidate loses the White House. For all model fits, the estimated coefficient is negative and statistically significant, starting with Model 1, ($\hat{\beta}_{MoV} = -0.79$, *clustered s.e.* $= 0.31, p = 0.005$) and ending with Model 4, ($\hat{\beta}_{MoV} = -2.27$, *clustered s.e.* $= 0.25, p < 0.001$). The consistently negative sign on this coefficient suggests that all four model fits agree with the Legitimacy Hypothesis.

For counties where the preferred candidate *won* the Presidential Election, these estimated models would predict a positive increase in the Annual Change in AGI / PI as a function of the margin of victory relative to counties that lost. For the most partisan counties in the sample, which had a margin of victory of 87.4%, based upon the estimates from Model 4, we would predict an increase in the Annual Change in AGI / PI of 1.49 percentage points. If the preferred candidate had lost the national election, instead of the increase, the model would predict predict a decrease in the Annual Change in AGI / PI of -0.16 percentage points.[2] Thus, according to this model, the effect of having your preferred candidate win the national election would be an increase of 1.65 in the Annual Change in AGI / PI relative to the change expected if the preferred candidate had lost. To put this effect size into context, the average magnitude of change in AGI over PI is 2.29 percentage points ($SD = 2.43$). Thus, the effect of the election outcome for the highly partisan counties is approximately equivalent to a change of 0.68 SD.

Using the same estimated coefficients from Model 4, let us now make predictions for the average county, which had a margin of victory of 25.6%. For this average county, if the preferred candidate won the White House, the model would predict an additional increase in AGI / PI of 0.44 percentage points. If the preferred candidate failed to win the White House, we would instead predict a smaller increase of 0.18 percentage points. Thus, for the average county, the difference in AGI / PI between winning vs losing the Presidential Election is 0.26 percentage points. In terms of standard deviations, the election effect for moderately partisan counties is approximately 0.11 SD.

In comparing the two predicted results above, we see a significant difference in the magnitude of the predicted election effect for counties as a function of the size of the margins of victory: highly partisan counties show a more dramatic effect of election outcomes compared to less partisan counties (strong partisan: 1.65 vs average partisan: 0.26 percentage point additional increase in AGI over PI for winning counties vs losing counties, resulting in an election effect that is approximately 6 times as large for the highly partisan counties). As described in Section 9.1.1 Hypothesis 1(a) - Partisanship, this pattern of results

[2] Computed Using Coefficients from Model 4 as follows: 0.326 + 1.710 * 0.87 + -2.266 * 0.87

—demonstrating the moderating role of partisanship in the relationship between election outcome and tax compliance —is exactly what one would expect if election-based legitimacy effects were impacting levels of tax compliance.

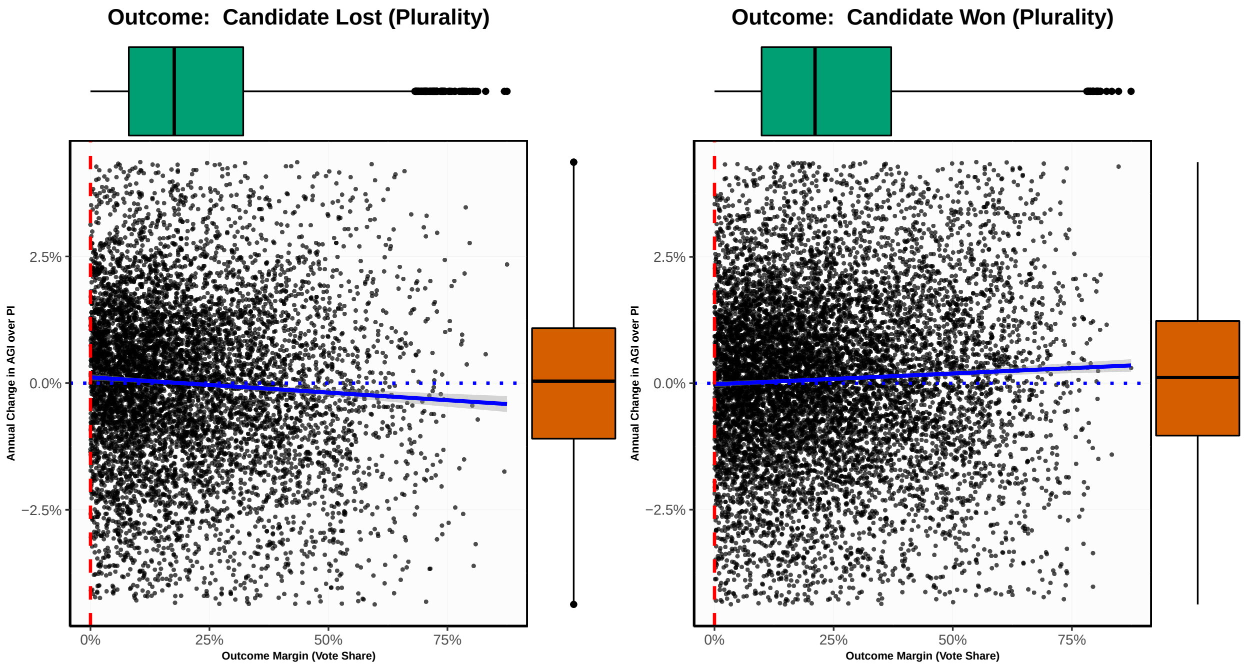


Figure 13.3: Relationship between Margin of Victory and Annual Change in AGI over PI for Election Year Data for Winning and Losing Counties — Shown for the Full Sample of Counties — Victory and Loss Classified Using Plurality Threshold

The moderating effects of *Margin of Victory / Loss* on the *Annual Change in AGI over PI* is captured in Figure 13.3 as a scatter-plot with Margin of Victory / Loss on the X-axis and Annual Change in AGI over PI on the Y-axis.[3] The panel on the left shows data for all counties from the election cycles when their preferred candidate lost the White House, whereas, the panel on the right shows data for these same counties from the election cycles when their preferred candidate won the White House. The solid blue line shows the fit estimated by the linear model. In order to facilitate presentation, both plots are zoomed in to highlight the central 90% of values for Annual Change in AGI over PI. A comparison of

[3]Annual Change in AGI over PI has been de-meaned two-ways to remove county and year fixed effects and thus more accurately reflects the linear relationship reported in the regression table.

the two plots shows the diverging impact of the margin of support depending on whether the preferred candidate won the Presidency. When the preferred candidate lost (left panel), larger margins of support for the candidate are associated with larger decreases in AGI / PI relative to the previous year; however, when the preferred candidate won, larger margins of support for the candidate were associated with larger increases in AGI / PI relative to the previous year.

CHAPTER 14

CONCLUSION

14.1 Overview

14.1.1 Legitimacy and the American State

Legitimacy —as a guiding concept —has long occupied a central place in the Western political tradition, shaping the development of political theory, the design of political institutions and the relationship between citizens and the State. Starting with Thomas Hobbes, all major political theorists of the Enlightenment —including luminaries like John Locke, Jeremy Bentham, John Stuart Mill, and Jean-Jacques Rousseau —have wrestled with the concept of political legitimacy and its relationship to the central problem of government: what is the source of political authority i.e. how does the State gain the right to demand obedience and to expect compliance with its edicts? While there were substantive disagreements among these theorists on the origins and characteristics of political legitimacy, almost all agreed that political legitimacy was necessary to produce any "moral or societal obligation" to obey the political authorities and to comply with their commands and laws (Peter, 2017). These Enlightenment thinkers (especially John Locke and Rousseau) were a central source of inspiration for the American Revolution. Their writings were highly influential among the Founders and their theories provided the philosophical framework for the creation of the American state. As a result, since the founding of America, the notion of political legitimacy —as the source of political authority via the "consent of the governed" —has been central in the design of American political institutions, manifest in its political traditions and has continued to figure prominently in both elite and everyday political discourse to this day (Bristow, 2017).

14.1.2 Evidentiary Gap: Testing the Emprical Relation between Legitimacy and Compliance

Although political legitimacy has been central to the discourse in political science and political philosophy, the psychological basis and the behavioral consequences of perceived legitimacy have only been examined more recently (Levi et al., 2012; Tyler, 2004). In keeping with the Enlightenment models, these psychological models of perceived legitimacy propose that compliance with the law in a liberal Democratic society is largely voluntary and this voluntary compliance is grounded in people's perceptions of legal authorities as legitimate (the Legitimacy Hypothesis). However, as a recent review by Nagin & Telep (2017) points out, there is almost no empirical evidence establishing a causal relationship between *perceived legitimacy* and *voluntary compliance with the law*, a concern that is conceded even by the most prominent advocate of the legitimacy hypothesis (Tyler, 2017a, 2017b). The primary obstacle to testing this causal relationship has been the difficulty in designing interventions that can target perceived legitimacy without changing enforcement capacities or other material conditions that might impact compliance with the law.

To overcome this obstacle, the current work argued that Presidential Elections can serve as natural quasi-experiments that produce as an exogenous shock to the perceived legitimacy of the Federal Government. The idea is that —in a highly polarized society with widespread and strongly held negative views about members and leaders of the other party (i.e. widespread negative partisanship) —losing an election results in a decrease in the perceived legitimacy of the State writ large. Similarly, winning an election results in an increase in perceptions of political legitimacy (see Section 8.2 Theoretical Framework: Constructing an Empirical Test of the Legitimacy Hypothesis). Since there are strong empirical and theoretical reasons to believe that election outcomes significantly impact people's perceptions of legitimacy and election outcomes are largely exogenous to the actions of any given individual, presidential elections can be treated as quasi-random, exogenous shocks to perceptions of legitimacy and can thus be used to examine the causal relationship between perceived legitimacy and voluntary compliance with the law.

Using these election-based shocks, the current work focused upon tax compliance as the domain to test whether election-based changes in perceived legitimacy were accompanied by a corresponding change in voluntary compliance. There are numerous reasons why tax compliance would be a particularly profitable case study to test the legitimacy hypothesis under consideration. First, prior research in the domain suggests that compliance rates are too high to be explained by enforcement efforts (the "compliance puzzle"), thus voluntary compliance appears to be a major factor in the tax compliance (Alm & Yunus, 2009; Gentry & Kahn, 2009). Second, tax compliance is a domain of voluntary compliance that is broad and representative of the population —approximately ~40-50% of the American population is required to file taxes annually and must decide whether to comply or not. Third, there are severe legal consequences associated with non-compliance, and salient "up front" significant financial costs associated with compliance —making it a consequential decision with difficult trade-offs. Finally, as a practical matter, tax compliance is a major issue facing governments worldwide and there is important practical value in examining the behavioral factors that influence tax compliance (Chetty et al., 2009; Congdon et al., 2009b; Hallsworth et al., 2014; Rees-Jones & Taubinsky, 2016) (see Section 8.3 Empirical Framework: Voluntary Tax Compliance as a Case Study for additional details).

The current analyses were conducted at the county level —covering data from 2753 U.S. counties across 28 years covering 7 Presidential Elections. To examine changes in tax compliance, the current work relied upon a variable constructed by cross-referencing estimated Personal Income (PI) from the Bureau of Economic Analysis (BEA) with the Adjusted Gross Income (AGI) reported to the Internal Revenue Service (IRS). The ratio AGI over PI serves as a measure of the self-reported taxable income as a proportion of total personal income. Since the definition of income for the BEA and IRS are not identical, even under perfect honesty, we would not expect this ratio to equal 1. And, due to differences in sources of income across counties, we should expect significant heterogeneity in AGI over PI across counties and years. As a result the raw values of AGI over PI are not particularly informative about the levels of tax compliance. However, within a given county, this ratio should remain relatively stable across consecutive years. Thus, by examining changes in AGI over PI across tax years —specifically, by comparing the difference in the change in AGI over PI for counties that won (i.e. the preferred candidate won the White House) with the change in AGI over PI for counties that lost —it is possible to estimate how elections impact the relative levels of tax compliance as a function of the election outcome. It is the comparison of the difference-in-differences (i.e. comparison of difference in changes between winners and losers) that allows for us to test the Legitimacy Hypothesis. In terms of the validity of this measure, the argument in its favor is detailed in Section 10.3.3 AGI over PI: Constructing Compliance Measures using the BEAs Income Estimates. However, it is worth noting that the use of similar ratios by the IRS —which is arguably the best authority on tax compliance in the US —lends this measure a high degree of credibility, especially when drawing conclusions based upon differential changes across two groups in response to an exogenous shock.

A key insight of the current analysis is that a series of quirks in the American election system allows us to construct an identification strategy that can isolate election-based shocks to non-material or psychological factors from election-based changes in material conditions. Specifically, the crux of the strategy is to focus upon the "election tax years" (i.e. taxes paid for economic activity conducted between Jan-Dec of an election year: 1992, 1996, 2000, 2004, 2008, 2012, and 2016) and to compare changes in AGI over PI from the pre-election year to the election year. The reason for focusing on election years is three fold. First, real economic activity is insulated from election results: the election results are only known in November, by when most of the tax year has already passed and thus is untouched by the election outcome. As for potential changes in economic behavior in the final two months, I rely upon recent work by Mian et al. (2018) —which specifically compared the six months immediately following the election to the six months prior to the election and found no measurable change in economic activity. Second, although election results are known in November, the new administration does not take office until January —thus, the President-elect cannot impact any of the legal or administrative conditions governing the election tax years. Third, tax-filing season coincides with the window when election outcomes are most salient: the vast majority of tax returns are filed between January-April of the following year, in the exact window when the outcome of the election is most salient (starting with the inauguration on 20th January and going through the "first 100-day window" when the newly elected President attempts to push major agenda items). Thus, by focusing on the election year, I can examine taxes paid on economic activity largely conducted prior to the election, where tax liability is governed by laws and administrative rules that are unaffected by the election, and the material conditions are insulated from the legislative or administrative influences of the incoming administration. However, while material conditions are insulated from election outcomes, the actual taxes are filed in a window when the election outcomes are most salient to the tax payer —creating ideal conditions for changes in perceptions about the government to influence decisions about tax compliance. Thus, the comparison of interest focus on differential changes between the pre-election (e.g. 1999) and election tax years (e.g. 2000) in AGI over PI for counties whose candidate won relative to changes over the same period for counties whose candidate lost. By focusing on changes in tax behavior for election years, it is possible to isolate election-based changes in perceived legitimacy from the material conditions governing tax compliance and enforcement actions (which remain unaffected by the elections). And, if we can isolate election-effects on perceived legitimacy from election-effects on material conditions, then —by comparing the tax filings of winners and losers of the election (i.e. by examining the effect of election outcome on tax filings) —it becomes possible to isolate and estimate the effects of changes in perceived legitimacy on change in tax compliance.[1]

Finally, it is important to highlight that —while the results are often presented in terms of comparisons between counties that won and those that lost —since the sample included 7 elections, every county had multiple elections cycles when their preferred candidate lost and multiple election cycles when their preferred candidate won. And, since —after removing county and year fixed effects —data was pooled across multiple election cycles, it is possible to think of each county as serving as a control for itself —with data from election cycles when its preferred candidate won the Presidency being compared with data from election cycles when its preferred candidate lost. This is one central feature of the sample that reduces the need to have a reference "synthetic" or "matched" control group and allows for preliminary causal inference, despite the lack of true random-assignment at the county-year level.

14.1.3 Overview of Current Analyses and Findings

The current work approached the data analyses by first using an exploratory approach using graphical analyses of two variables: AGI per return and AGI over PI. Having determined AGI over PI to be more appropriate for testing the Legitimacy Hypotheses, I proceeded with subsequent statistical tests to examine the significance of election-based differential changes in AGI / PI for election winners and losers. The results are summarized here.

[1]The current approach cannot control for changes in perceived risk of enforcement - which could be influenced by election outcomes, with losers assuming that they may be the target of extra scrutiny and winners assuming they will be given protection by their president. Perceived risk of enforcement is known to be a significant factor in deciding whether to comply with one's tax obligations (Plumley, 1996). Thus, this could be a significant concern.

However, as argued in Section 9.2.2 Hypothesis 3: Perceived Enforcement Risk, election-based effects on perceived enforcement risk should drive compliance in a direction contrary to the one predicted by the Legitimacy Hypothesis and thus can be distinguished on that basis.

Graphical Analyses

In the first set of analyses, in Section 12.4.1 Trends in AGI per Return, I examined the pattern of changes across the 27 years in the sample for the real (inflation adjusted) Adjusted Gross Income (AGI) per return (represented as an annual percentage change variable). AGI per return has some intuitive appeals, since it is both easily interpretable and a natural place where one should expect to see large changes in income disclosure. Unfortunately, while the graphical analysis showed patterns that were consistent with the Legitimacy Hypothesis, the election effect was only detectable for the post-election year (marked as T+2 on the aggregate graphs). This means I only found evidence of an election effect for the tax year that covered the first year under the new presidency (e.g. TY 2001; 2005; 2009 etc, rather TY 2000; 2004; 2008). Although it was expected that the election effect should both persist and likely get stronger in the post-election year (see Section 11.2.1 [Varying Time Dynamics —Election Effects Persist for 1-3 Years]), the strategy for isolating the non-material election effect relied upon detecting an effect in the election year itself (i.e. for the tax year during which the election took place, e.g. TY 2000; 2004; 2008).

Although AGI per return is an appealing variable due to its ease of interpretation, the delayed timing of the election effect poses a significant problem for the interpretation of changes in AGI per return. For example, if the AGI per return for a county goes up by $1, there are two possible causes: (i) everyone in the county made an additional $1 in income over the previous year or (ii) everyone decided to disclose an additional $1 in income to the IRS. So, both an increase in earning and an increase in disclosure would produce an increase of $1 in AGI per return. And, since the current analyses only reveal an election effect in the post-election year (i.e. after the elected President has had one year to implement policy), the additional growth in AGI per return for counties that supported the winner could be seen as either (i) evidence of a patronage network whereby Presidents reward their supporters or (ii) evidence of a Legitimacy Effect whereby constituents choose to increase tax compliance to support the government led by their preferred candidate.

Rather than attempting to address this problem by relying upon an exhausting battery of economic control variables, it was deemed preferable to use a different variable for the second set of analyses: namely, AGI over PI. As was discussed above, AGI over PI is a ratio constructed by standardizing the reported AGI for a county in a given year by the BEA's annual estimate of aggregate personal income for that county. Although the raw AGI over PI value is not sufficiently informative to infer conclusions about tax compliance, changes in AGI over PI across time can reasonable serve as a proxy measure for tax compliance. Or, more specifically, patterns of *differential change* in AGI over PI in response to an exogenous shock (winning vs losing an election) can serve as a reasonably strong proxy for *differential changes* in patterns of income disclosure across those two groups. Thus, the second set of graphical analyses focused on the difference in the changes in AGI over PI for counties that supported the winner and those that supported the loser.

The pattern of data revealed by the graphical analyses in Section 12.4.2 Trends in AGI over PI showed significant support for the Legitimacy Hypothesis. In both the election cycles where the Democratic Candidate won the White House (1992, 1996, 2008, and 2012) and cycles where the Republican Candidate won the White House (2000, 2004, and 2016), relative to counties that won, counties whose candidate lost the Presidential Election showed a steeper decline in AGI / PI from the pre-election to the election year. This pattern held up when data was collapsed across all 7 elections —with a larger decrease for losing counties than winning counties, a pattern that is consistent with the Legitimacy Hypothesis.

Statistical Analyses

Following the graphical analyses, some basic tests were conducted in Chapter 13 to statistically verify the patterns seen in the graphical examination of the data. The statistical tests were conducted by fitting panel regression models. County and year fixed effects were included in all statistical models to eliminate the risk of any biases due to omitted factors that either vary across counties (but are constant over time) or omitted factors that vary across time (but are constant over counties).[2] All models were aimed at testing the central hypothesis: do changes in AGI over PI systematically diverge for counties whose preferred candidate won the national election when compared with counties whose preferred candidate lost. In the first set of tests, election outcomes were measured as binary variables. In the second set of tests, election outcomes were measured as a continuous variable —allowing us to test whether the elections effects increased in magnitude as a function of the strength of support for the winning or losing candidate. Across all variations tested, the support for the Legitimacy Hypothesis was robust and consistent.

For example, when we compare Republican counties and Democratic counties (parties that voted for either the Republican or the Democrat candidate in all 7 elections), for election cycles when Democrats won the White House, the AGI / PI for Republican counties decreased from the pre-election to election tax year by an additional 0.3 percentage points relative to the change seen across the same period in Democratic counties. However, for election cycles when Republicans won the White House, the same comparison revealed that the AGI / PI for Republican counties increased from the pre-election to election tax year by an additional 0.24 percentage points relative to the change seen in Democratic counties. Collapsed across all 7 election cycles, the broader pattern holds: for election cycles where a county's preferred candidate wins the White House, the county's Annual Change in AGI / PI for the pre-election to election tax year was 0.27 percentage points higher than for election cycles where a county's preferred candidate lost. After including control variables, this effect further increases in magnitude —holding all else equal, having your candidate win the Presidential Race was a change in AGI / PI that was 0.41 percentage points higher than it was when your candidate lost.

[2]Since fixed effects can often introduce serial correlation in the residuals, all standard errors were adjusted to be heteroskedasticity and autocorrelation-consistent (HAC) standard errors clustered at the county-level.

Moreover, as predicted, the magnitude of these effects was directly moderated by the strength of the support for the candidate. For example, analyses using signed margin of victory —a simple univariate representation of election outcomes that ranged from +1 (100% of the county supported the candidate that won the Presidency) to -1 (100% of the county supported the candidate that lost) —the data suggested that a county that fully supported the winning candidate would show an Annual Change in AGI / PI (pre-election to election year) that was 0.66 percentage points higher than a county that fully supported the losing candidate. Once control variables were included, these analyses suggested a difference of 1.73 percentage points between counties that fully supported the winner and those that fully supported the loser. In both cases, these effects increased as the margin of support increased. An alternate specification allowed the effect of partisan support for a candidate on annual change in AGI / PI to vary across counties whose candidates won and lost —in effect, allowing the effect of victory to be distinct in magnitude from the effect on loss. With this alternate specification, after controlling for economic characteristics, suggested that the predicted difference between a county that fully supported the winner and a county that supported the loser was closer to 2 percentages point.

Interpreting the Findings

Based on the graphical and statistical analyses presented here, there is compelling evidence to suggest that election outcomes result in differential changes in AGI over PI for counties whose candidate won relative to those whose candidate lost. How should this be interpreted? As discussed before, changes in the ratio of AGI / PI can be caused by one the following: (a) changes in the estimation accuracy of personal income by the BEA; (b) changes in the relative proportion of income sources in a county; or, (c) changes in income disclosure. A priori, it does not seem reasonable that the accuracy of income estimation by the BEA changes differentially as a function of who wins or loses an election. Similarly, a priori, it does not seem reasonable that —across 7 elections —counties experience a systemic change in sources of income in a manner that produces a decrease in AGI over PI for counties that lose and increases for the counties that win. This interpretations could be read as a version of the "third variable problem" addressed below.

Finally, we can also address the following concerns before drawing any causal inferences from the data: (i) reverse causality i.e., the concern that counties that experience economic changes that increase AGI over PI end up supporting the winner and counties that experience economic changes that decrease AGI over PI end up supporting the loser (e.g. by using federal spending to buy votes, Groseclose & Snyder, 1996; Levitt & Snyder, 1997, 1995) (ii) third variable problem i.e., some unobserved variable both changes AGI over PI and produces alignment of a county with either the winning or losing candidate. Two factors guard against these concerns.

The first factor that guards against these concerns is the fact that patterns consistent with the Legitimacy Hypothesis remain significant (or, get stronger) even when the data sample was restricted to just the loyal partisan counties (i.e. the results appear even when we restrict the sample to counties that always vote for the same party). Since the loyal partisan counties —by definition —do not switch votes and they cannot be selecting into the winner and loser groups. Since their voting preferences are fixed, we can both rule out the reverse causality concern and the third variable concern.

The second factor guarding against these concerns has to do with the timing of the election. Election results are not known till November. And, since the analyses focus on changes in the election year, the "third variable" would have to either produce changes retroactively (after the results are revealed) or be able to predict the election outcome months in advance of November. Since we can rule out retroactive changes, it is worth considering the issue of predictability of Presidential elections. The pattern of results seen here is strongest in "changeover elections" (i.e. 1992, 2000, 2008, 2016). These same Presidential elections (especially the ones without incumbents) appear to be a lot closer to a coin-toss than is often acknowledged. For example, it is worth recalling that the last 3 changeover elections were decided by highly contingent factors (the 2000 election was decided by 537 votes in Florida and the Supreme Court; the 2008 election was dramatically shaped by the improbable and sudden collapse of Lehman Brother in mid-September 2008; and the 2016 election was quite possibly influenced —and, possibly determined by[3] FBI Director James Comey's decision to send a letter to Congress 2 weeks before election day announcing the reopening of the "email investigation"; the 2016 election was also decided by 107K votes across 3 states or 0.09% of the 120 million ballots cast). In light of the highly contingent nature of these changeover elections, it appears highly improbable that there exists a systemic variable that can both (i) influence the economic characteristics of a wide-swathe of US counties to alter AGI over PI in a consistent direct; (ii) predict electoral outcomes; and (iii) sufficiently alter voting patterns to produce alignment of counties with the Presidential Election winner / loser based upon these changes in economic characteristics.

Having ruled out these alternative interpretations, the reasonable conclusions from the pattern of data seen in the current analyses are: (1) election-outcomes are driving changes in the pattern of AGI over PI (i.e. causal direction flows from election outcomes to changes in AGI / PI and not the other way); (2) the changes in AGI over PI are most naturally interpreted as evidence of changes in patterns of income disclosure. Thus, the results presented here can reasonably be interpreted as providing evidence that counties that supported the winning candidate were more likely to increase income disclosure on the tax return relative counties whose candidate ended up losing the election —and, the magnitude of the change increases as a function of the level of partisanship in the county.

To get a rough sense of the election effect predicted by the final models, I use the regression models from Table 13.4 to predict the Annual Change in AGI over PI for each county

[3]Silver (2017)

assuming its preferred candidate won. Then, using the AGI over PI for the pre-election year, I recover the predicted AGI over PI for the election year, then by multiplying this with the BEA's estimate of Personal Income for the election year, I construct the predicted AGI for that county in that election year. The same process is then repeated for the county under the assumption that its preferred candidate lost. The difference in these two predicted values of AGI gives us the estimated election effect for the county in that election year and the sum of election effects across all counties gives us the estimated election effect for each election year.

$$\mathbb{E}[\mathrm{AGI}_{cy}|\,Won] = \left(\mathbb{E}[\Delta_{AGI/PI_{cy}}|\ \ Won\ ] + \frac{AGI_{c(y-1)}}{PI_{c(y-1)}}\right) * PI_{cy}$$

$$\text{Election Effect}_y = \sum_{c \in C} \mathbb{E}[\mathrm{AGI}_{cy}|\,Won] - \mathbb{E}[\mathrm{AGI}_{cy}|\,Lost]$$

Using the coefficients from Model 4 from Table 13.4, for Tax Year 2016, holding all else equal, the model would estimate that the predicted election effect across all counties would be \$45.75B. In 2016, taxpayers reported earning \$10.9 trillion in Adjusted Gross Income and paid \$1.6 trillion in individual income taxes —producing an average tax liability of 14.7% on AGI. Using this figure, an additional \$45.75B in AGI would result in an additional \$6.72B in Federal Income Taxes paid. Since the estimated tax gap in 2016 was approximately \$447B, this partisanship-based evasion would account for approximately 1.5% of the overall tax gap. This estimate is likely to be a lower bound —since it is derived from aggregate tax receipts, which include many households with W-2 salary income, which provides no real opportunity to modify tax burden. Additional analyses would benefit from examining tax receipts from Schedule C (Sole Proprietorship) submissions, where opportunity and prevalence of tax evasion is much higher.

Finally, this model also predicts a steadily increasing election effect for each election cycle

—driven by the steadily increasing levels of partisanship across the country. Compared to the AGI Gap of \$45.75B in 2016, the figure for 2012 was only \$29.68B, for 2008 only \$23.62B, and by 1992 it is a meager \$194.23M. This increasing effect of election outcomes on tax compliance aligns with other findings showing that the effect of election loss on economic perceptions have also grown over time (Mian et al., 2018) as have effects of election loss on perceptions of electoral integrity (Sances & Stewart, 2015). In this context, if the current model's predictions are to be taken at face value, one would have to conclude that the country is suffering an ever increasing cost to governance as a result of rising levels of partisanship.

14.2 Similarities and Differences with other current work

When this project was started and preliminary analyses were completed, there did not exist any other work in the literature that had examined the effects of US Presidential Election outcomes on tax compliance behavior. However, it appears that Cullen et al. (2018) were working contemporaneously on a very closely related study. In their work, Cullen et al. (2018) treat presidential turnover elections as quasi-experimental shocks that manipulate tax morale i.e. shocks that alter people's opinions (approval-disapproval) about "government tax and spending policy." They then use county-level data to examine whether election outcomes impact tax compliance —which they argue is a test of the causal relationship between tax attitudes and tax compliance, allowing them to ask whether "less negative attitudes toward taxation and spending translate into a higher willingness to comply with taxation."

The work by Cullen et al. (2018) is impressive in scope and execution. It also overlaps substantially with the empirical strategies used herein. Specifically, both papers use election outcomes as quasi-random, exogenous shocks to political perceptions as method of testing whether those perceptions influence tax compliance. Cullen et al. (2018) have access to individual-level income tax data under an agreement with the IRS through the Compliance Data Warehouse. This dataset gives them access to line-level tax data and allow them to conduct analyses on a rich set of tax variables, including Schedule C & E filings (which are the tax schedules associated with the bulk of tax evasion), Earned Income Tax Credit (EITC) filings and audit rates. Overall, their findings are robust and agree with the main findings here, namely, there appears to be strong evidence for election-effects on rates of tax compliance.

Despite these similarities, there are substantial procedural, empirical and theoretical differences that differentiate the two studies, including the conclusions that can be drawn and the predictions that would arise from the central hypotheses. In the following sections, I highlight some of the crucial differences in the analytic approaches adopted by the two papers —with specific attention towards features of the analytic approach used by Cullen et al. (2018) that result in critical obstacles to causal inference.

14.2.1 Differences in the Target DVs

Despite the breadth of data, Cullen et al. (2018) restrict their analyses to just the 2000 and 2008 election, giving each county only one election cycle when it is aligned with the government and one when it is not. In contrast, the current work draws upon a broader sample that covers 7 elections (4 crossover, 3 reelection). In addition, they limit their measures of tax compliance to measures of income reported to the IRS. By restricting their analyses to measures of income reported to the IRS, they face a problem interpreting the cause of changes —after all, an increase of $1 in AGI/return could result from either an increase in income or an increase in income disclosure. They address this problem by (i) relying upon control variables; and, (ii) by examining alternate variables like audit rate.

In contrast to these approaches, the current work takes the strategy of examining changes in the ratio of reported income (IRS data) and estimated income (BEA data). Not only does this reduce the need to rely upon economic controls, but by relying upon a variable that is constructed by cross-referencing two different data sources, the current analyses benefit from a significant increase in the ability to detect changes in tax compliance. For example, when I restricted my analyses to reported AGI per return, I was only able to detect an election effect in the post-election year (i.e. in the tax year that covers the first year of the incoming presidency). However, when I used the cross-referenced measure of AGI over PI, I was able to detect this effect in the election year itself —thereby obviating the need for control variables.

14.2.2 Differences in Target Year: Election Year vs Post-Election Year

Interpretative Concerns with Post-Election Data

Another feature of Cullen et al.'s (2018) empirical strategy increases the reliance upon control variables: their work relies focuses upon the year following the election to estimate the tax-consequences caused by the election i.e. they search for evidence of the exogenous shock delivered by the elections on taxes paid for first tax year of the presidency. In their analyses, they decide that the election year is ambiguous since earnings were made under one president and tax filings occurred under another. In the current approach, it is exactly this bifurcation between the "earning window" and the "filing window" that allows tests of changes in the election year to provide a much stronger claim in support of causal interpretations.

As Cullen et al. (2018) acknowledge in their paper, by examining the first post-election year (i.e. the first year under the new president), there are at least two established pathways for the election outcome to impact actual earnings (since they rely upon income reported to the IRS as the primary DVs, changes to actual earnings is a competing interpretation of the findings). The first pathway arises from changes in economic expectations and changes in economic optimism among residents of the pro-administration counties (Gerber & Huber, 2010); the second pathway arises from changes in transfers from the federal government to districts aligned with the President's party (Berry et al., 2010; Dynes & Huber, 2015). Cullen et al. (2018) address the matter by including a broad set of control variables and by corroborating their findings using non-income related measures like audit rates. However, by relying upon the post-election year as their measure, their claims to causal inference rest heavily upon the robustness of their control variables. In contrast, the current work, by relying upon the election year, does not face the same requirements and is largely free of the burden of needing to control alternative causal pathways that could link election outcomes to changes in reported income.

An additional concern with the reliance on post-election data —one that is not addressed by Cullen et al. (2018) —is that incoming presidents often pass omnibus tax legislation right after taking power that can significantly alter tax obligations through legislative and administrative changes. For example, President Bill Clinton significantly altered tax law with the "Omnibus Budget Reconciliation Act of 1993"; President George W. Bush did the same with the "Economic Growth and Tax Relief Reconciliation Act of 2001 (EGTRRA)"; as did President Barack Obama with "American Recovery and Reinvestment Act of 2009"; and, President Donald Trump with the "Tax Cuts and Jobs Act of 2017." In all these cases, these tax bills were passed in the first year of the incoming administration and changed the tax obligations for the first Tax Year of the new presidency i.e. exactly the target year used by Cullen et al. (2018). They try to address some of these concerns by restricting their sample to "policy constant" filers (i.e. individuals who would have had to file determined by 1996 tax law).

However, the restriction to "policy constant" filers only addresses the impact of such laws on filing thresholds. These legislative actions are often broad acting and their effects would be hard to control against. It seems likely that these laws —written by House and Senate majorities aligned with the President's party —differentially reduce the tax burdens on their co-partisans, thus potentially differentially decreasing the tax burden on tax payers aligned with the President, which would impact the tax variables considered by Cullen et al. (2018). Moreover, since tax policy —most notably marginal tax rates —are significant drivers of tax compliance (Plumley, 1996), these legislative actions could also drive changes in more direct measures of compliance used by Cullen et al. (2018) such as audit rates. In contrast, since the tax filings covering the election year remain insulated any potential actions by the incoming administration, the analyses presented here do not suffer from the same concerns.

14.2.3 Time Dynamics and Source of Election-Driven Effects

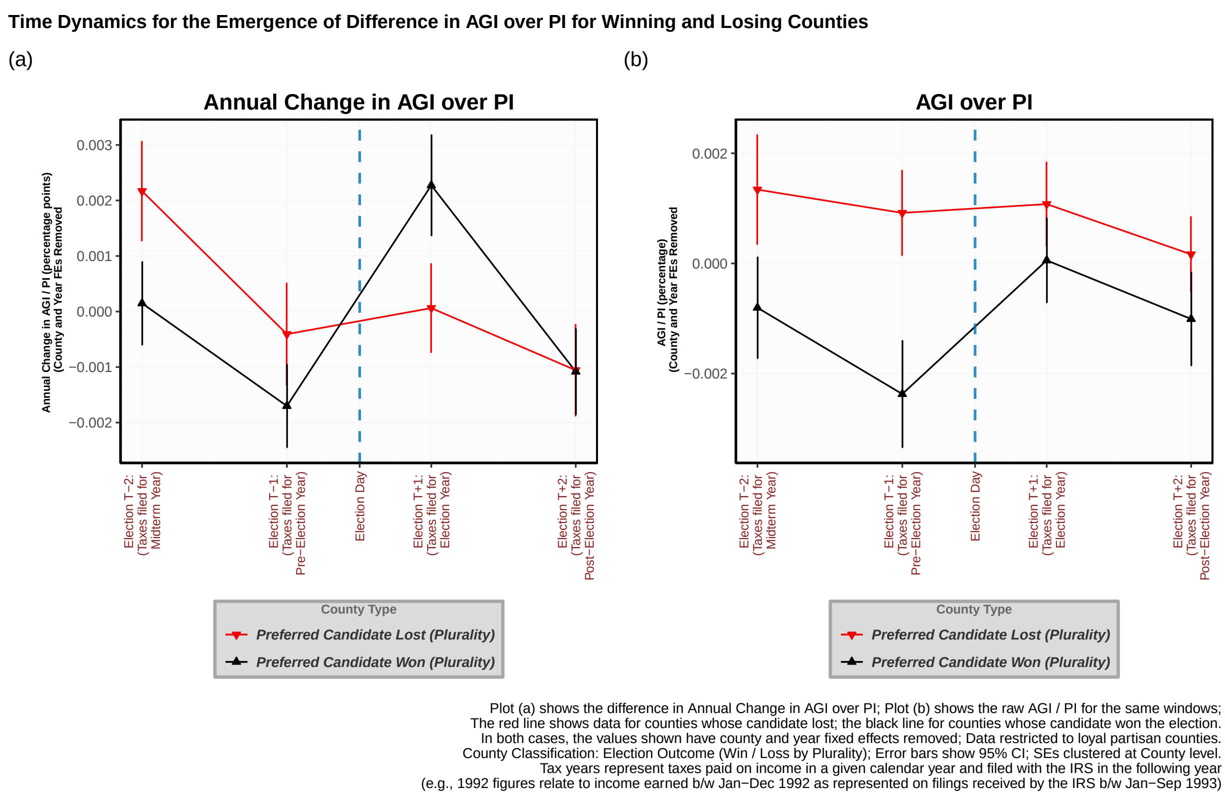


Figure 14.1: Time Dynamics of the Election Effect —Comparing Annual Change in AGI / PI and raw AGI /PI for Winning and Losing Counties —Loyal Partisan Counties Full Sample of Counties —Victory and Loss Classified Using Plurality Threshold

An examination of the time dynamics of changes in AGI over PI suggest that —while the election effect continues to strengthen in the first year of incoming presidency —the major portion of the election effect occurs in the income disclosure of the election year itself, which is what one would expect if the salience of election outcomes during the filing period is what is most important. After all, at no point is the change in power more salient than in the first hundred days of the new administrations i.e. during the filing window for the election year tax returns. This pattern is shown in Figure 14.1 —where in both the Annual Change in AGI / PI (left panel) and the corresponding raw AGI / PI (right panel), it is clear that the major election effect occurs between the pre-election and election tax years. As shown in Supplementary Figure B.39 in Section B.5.3, this time dynamic persists even when we examine election cycles where Democrats won vs cycles where Republicans won. Finally, as the AGI / PI graph in Figure 14.1 (right panel) shows: the difference between the winning and losing counties between the election and post-election tax years remains remarkably stable (as represented by the parallel lines). Thus, it appears that —rather than being strategic behavior that accumulates over the entire year —a significant portion of the partisanship-driven non-compliance appears to be determined by decisions taken during the filing window at the moment of filing itself.

Moreover, if we fit a standard difference-in-difference model for the event study time frame (i.e. for the 4 years surrounding the election), it is possible to estimate the predicted difference between winning and losing counties. If there is no selection taking place, the differences between the two groups should be equal and stable prior to the election and only shift in the last two years following the election. As shown in Figure 14.2, this is exactly what we see when we plot the estimated difference between the two groups across all periods in the election window: the difference in AGI / PI between winning and losing counties only appears in the post-election years.

Three models are represented in Figure 14.2. The lines in black show a basic model with no control variables; the lines in blue show the same model fit after including a lag AGI / PI as a control; and the lines in magenta show the same model fit after including a control for Annual Change in AGI / return (as a measure of economic change) as well as a lag AGI / PI term. Following the exact practice of Cullen et al. (2018), in the current model specification used to generate these estimates, I also included county-by-election-cycle and state-by-year fixed effects. The inclusion of the county-by-election fixed effects restricts the comparison of a county's post-election AGI/PI to only its pre-election AGI/PI from the same election. The inclusion of these additional fixed effects does not significantly alter the time dynamics across the four points in election window. The full regression fit is reported in Supplementary Table B.7 in Section B.5.4. All coefficients are estimated relative to the pre-election (T-1) period as the reference level.

All three models lead to the same conclusion: the difference in AGI / PI between winning and losing counties only appears in the post-election years and it appears starting with the election tax year (T+1). Notably, the difference between the winning and losing counties remains stable between the election (T+1) and post-election (T+2) years. Thus, not only does the emphasis on the election year facilitate a stronger claim to causal interpretation of the findings, it also more accurately captures the time dynamics of the election effects: election effect appear as soon as the "filing window" overlaps with the new administration.

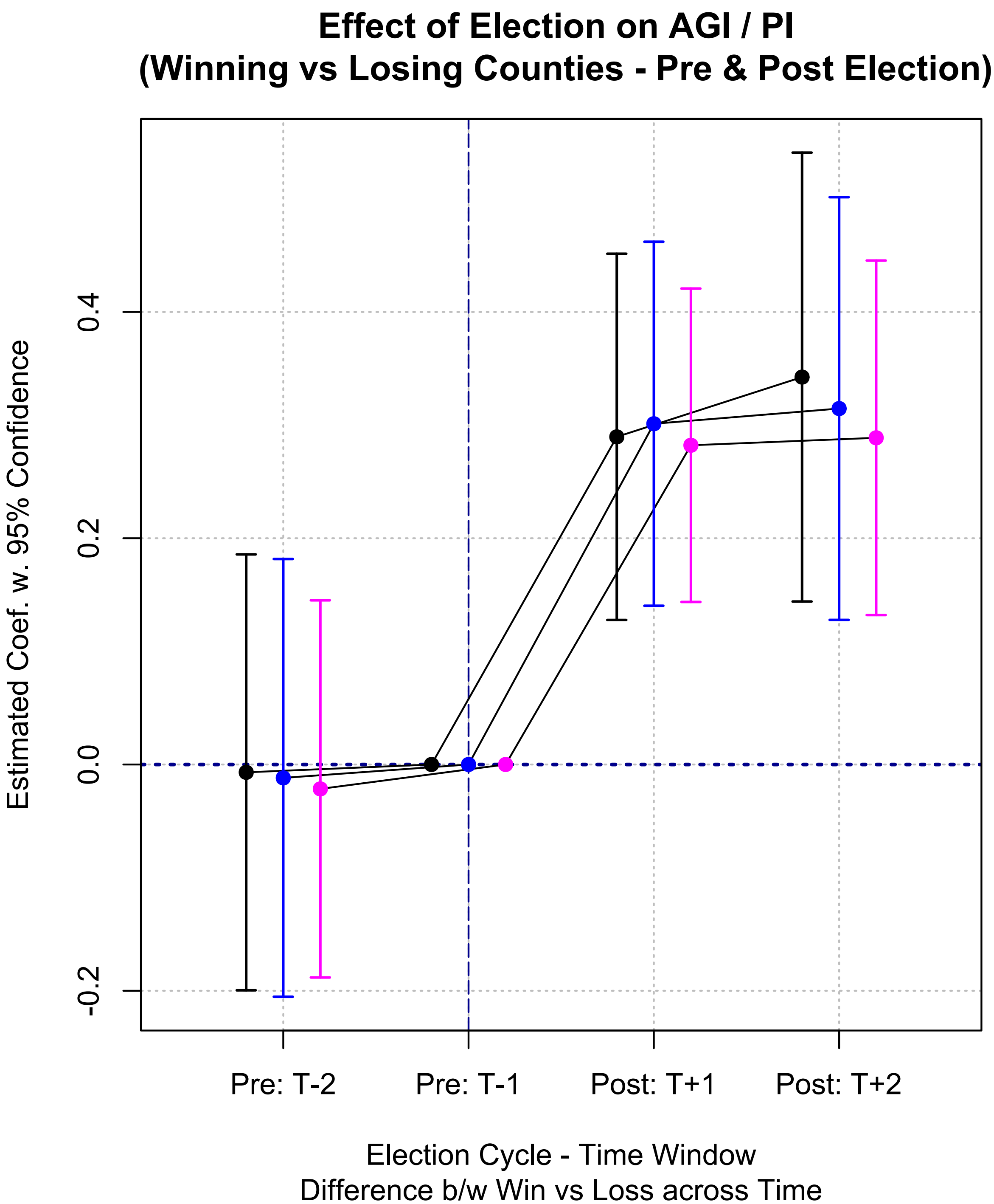


Figure 14.2: Time Dynamics of the Election Effect —Estimated Coefficients of Difference in AGI / PI between Winning and Losing Counties in All Years of Election Window (Pre vs Post) —Loyal Partisan Counties —Victory and Loss Classified Using Plurality Threshold —Three models fits: Basic (Black); Controlling for Lag DV (Blue); Controlling for Lag DV and Change in AGI / Return (Magenta)

14.2.4 Differences in Theoretical Framework and Predictions

Finally, while the current work significantly overlaps with that of Cullen et al. (2018) in empirical strategy (specifically, the use of election outcomes as a means to examine the effect of political perceptions on tax compliance), the work by Cullen et al. (2018) differs in its theoretical motivation —which frames the study as testing the effects of a "shift in tax attitudes" or "tax morale" on tax compliance. The current work thinks of the election effect as impacting broad perceptions of perceived legitimacy of the government writ large. These differences in theoretical framework impact both the interpretation of findings and also predictions that arise from them.

14.3 Conclusions and Future Work

Using data that covers 2753 U.S. counties across 28 years covering 7 Presidential Elections, I present initial evidence of a causal relationship between election outcomes and tax compliance. Counties that supported the winning candidate showed a relative increase in inferred income disclosure relative to counties that supported the losing candidate. These results confirm the findings reported by a contemporaneous study (Cullen et al., 2018), while also introducing novel analytic strategies that overcome some of the obstacles to causal inference in that work.

As a part of this research program, currently ongoing work is attempting to extend and replicate the findings reported here by taking advantage of the IRS's release of the same data universe aggregated at the ZIP-code level. This has the advantage of both providing an opportunity for pseudo-replication (since the same universe of tax filings are aggregated into different units) and for a test of the findings at a greater level of spatial granularity (~27,000 ZIP-Codes). Since voting data is not available at the ZIP-Code level, I have relied upon donation data to classify the partisanship status of each ZIP-Code (see Bonica, 2019 for tests of the validity of using donation data for such purposes). Preliminary results from

ZIP-code level analyses confirm the pattern reported here.

While the current set of results fit within the literature linking "tax morale" to "tax compliance," the theoretical framework deployed here would interpret the election effects demonstrated here as reflecting fundamental shifts in perceptions of legitimacy. Although the current data do not permit one to distinguish between these interpretations, if the interpretation in terms of perceived legitimacy is correct, then one would expect a broad-spectrum of downstream consequences extending to domains beyond tax compliance.

For example, the current theoretical framework would predict that adherence with public health orders like mask wearing would be highly responsive to perceptions of legitimacy. A recent study provides evidence of exactly such a pattern, finding that political affiliation is "more predictive than factors directly connected to the disease, including age, county infections per capita, and state public health policies" and that "the effectiveness of [mandatory mask] policies — and compliance with them —is mediated by political affiliation" (Makridis & Rothwell, 2020).

Similarly, the current framework would predict that prominent cases of police brutality that heighten awareness of systemic inequity, corruption or illegality —if they produce a salient decrease in the perceived legitimacy of the justice system —should produce a corresponding surge in crime in its aftermath. Although such arguments have been recently considered by criminologists (e.g. Lafree, 2018) and there is some evidence of increases in robbery and violent crime rates after prominent police killings —for example, after the killing of Michael Brown in Ferguson[4] (Pyrooz et al., 2016), of Freddie Gray in Baltimore (Morgan & Pally, 2020) and more recently after the killing of George Floyd in Minneapolis (Asher & Horwitz, 2020). Although the crime surges were documented, identifying the causes for these crime surges is difficult and likely to be multiply determined by factors like strategic

[4]Albeit, this increase in crime rates was restricted to "cities with historically high levels of violence, a large composition of black residents, and socioeconomic disadvantage" (Pyrooz et al., 2016) and the evidence suggested that the increase was significantly smaller than reported in the media (Bialik, 2015).

"under-policing" behavior by the police officers (Devi & Fryer, 2020) or changes in police perceptions (Nix & Pickett, 2017). The methodological approach presented here provides a different approach to examining the link between perceived legitimacy and crime rates while controlling for "under-policing" or other changes in the enforcement environment. For example, the same empirical framework could be applied to using the "Uniform Crime Reporting Program" Data Series produced by the National Archive of Criminal Justice Data (NACJD), which produces monthly data for arrests by local police forces aggregated at the county level.

If there is indeed an intimate link between perceived legitimacy and broad voluntary compliance with government authorities, the implications would be profound and have the potential to significantly reshape how we approach policy design and implementation.

Appendix

APPENDIX A

SUPPLEMENTARY MATERIALS AND ANALYSES FOR PART I: PERCEPTIONS OF CORRUPTION AND THEIR CONSEQUENCES

A.1 Prominent Indicators of Good Governance

Corruption Index or Indicator	**Institutions**	**Sources**	**Information Type**
CC *(Control of Corruption, World Governance Indicators)*	World Bank	Expert & Household & Firm	Perception
CPI *(Corruption Perceptions Index)*	Transparency International	Expert & Firm	Perception
ICRG *(International Country Risk Guide)*	Political Risk Services Group	Expert	Perception
IEF *(Index of Economic Freedom)*	Heritage Foundation	Expert	Perception
CI- EIU *Corruption Index*	Economic Intelligence Unit	Expert	Perception
CPIA *(Country Policy and Institutional Assessment)*	World Bank	Expert	Perception
BPI *(Bribe Payers Index)*	Transparency International	Firm	Perception
GCB *Global Corruption Barometer*	Transparency International	Household	Perception & Experience
AFBM *Afrobarometer Survey*	CDD-Ghana, IDASA, MSU	Household	Perception & Experience
EOS *(Executive Opinion Survey)*	World Economic Forum	Firm	Perception
WBES *(World Business Environment Survey)*	World Bank	Firm	Perception

SOURCE: Razafindrakoto & Roubaud (2010)

A.2 Measuring Political and Electoral Characteristics

1. **MEASURING ELECTION OUTCOME: AWARENESS**

   In the last 24 hours, various major news outlets have reported that Hillary Clinton has "clinched the nomination" as the Democratic presidential candidate (see below). *

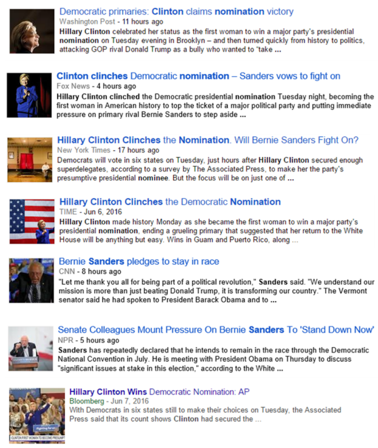


Figure A.1: Replica of full-page news prime

   How familiar were you with these events before taking this study? †

2. **MEASURING POLITICAL INTEREST**

   How much do you care about politics? †

3. **PARTY AFFILIATION**

   What is your political affiliation?

   (a) I am a Republican;
   (b) I am a Democrat;

*In both Appendix A.2 and Appendix A.3, dark blue ink is used to identify sample text and stimuli that are verbatim reproductions of the survey instruments used in both studies. The image of the news headlines occupied the entire screen and is only miniaturized here for ease of presentation (see Fig. 2.2).

†All questions about familiarity, degree of support, caring, emotional response, certainty judgments and other similar, *continuous* measures of opinion or sentiment were all measured on a 0-100 scale, with larger numbers signifying greater intensity on the dimension being measured.

(c) I am a Independent;
(d) I am a a third-party affiliate (Libertarian, Green Party, Other)
(e) I am a Unsure / Still deciding
(f) I am a Don't know / NA

4. **ELECTION ENGAGEMENT**[‡]

EMOTIONAL INVESTM.:	*Study 1:*	How much [do/did] you care about the primaries this year? †
	Study 2:	How much [do/did] you care about the presidential election? †

5. **CANDIDATE SUPPORT**

CHOICE OF CANDIDATE:	*Study 1:*	Who [do/did] you support in the primary?
	Study 2:	Who [do/did] you support in 2016 presidential election?
STRENGTH OF SUPPORT:	*Study 1 & 2:*	How strongly [do/did] you support your selected candidate? †

6. **VOTING BEHAVIOR**

ACTUAL VOTING:	*Study 1:*	Did you vote in the primaries?
	Study 2:	[Did you / Will you] vote in the 2016 presidential election?
VOTER ELIGIBILITY:	*Study 2:*	[Are you / Were you] eligible to vote in the 2016 presidential election?

7. **EXPECTATION**

[‡]The next set of questions include a square bracket of the format: [PRE/POST] on certain questions. This indicates that the question was measured for both the pre-election and post-election samples with the modifications in wording as shown in the square parentheses (e.g. '[do/did]') shown)

[†]Same as previous page.

Predicted Outcome:	*Study 2:*	[Who do you think will / Who did you expect would] win the 2016 presidential election?
Pred. Certainty:	*Study 2:*	[Before you found out the outcome of the election], how likely did you think it [is/was] that [expected candidate] would win the 2016 presidential election? †

8. **FEELINGS ABOUT ELECTION OUTCOME (STUDY 2)**

Neg. Emotion (Loss):	*Pre-Election:*	How upset would you be if [favored candidate] did not win? †
	Post-Election:	How upset are you that Hillary Clinton lost the election? †
Pos. Emotion (Win):	*Post-Election:*	How happy are you that Donald Trump won the election? †

A.3 Measuring Perceived Corruption

Q1. To what extent do you see the following affected by corruption in the US? *

	Not at all (0)	To a small extent (1)	To a moderate extent (2)	To a large extent (3)	Don't know / NA (NA)
Political parties (1)	◯	◯	◯	◯	◯
U.S. Congress (2)	◯	◯	◯	◯	◯
Media (3)	◯	◯	◯	◯	◯
Business / Private sector (4)	◯	◯	◯	◯	◯
Public officials / Civil servants (5)	◯	◯	◯	◯	◯
Tax (6)	◯	◯	◯	◯	◯
Elections (7)	◯	◯	◯	◯	◯
Federal government (8)	◯	◯	◯	◯	◯
Local government (9)	◯	◯	◯	◯	◯

Note: In Question 1, some of these items (Political parties; U.S. Congress; Media; Business / Private sector) were used for the entire period that the Global Corruption Barometer addressed corruption in institutions (2004-2013). Other questions were only present for half the samples, e.g. Tax (2004-2007) and Public Officials / Civil Servants (2009-2013). I did not use three items (Military, Police, & Judiciary), which I replaced with (a) elections; (b) Federal Government (more distant, harder to imagine); (c) Local Government (closer, easier to query about opinion).

Q2. Please respond to the following items using the slider scale ranging from "not at all" to "a lot".

(1) How much of a problem is political corruption in the US?

(2) How much does corruption influence the selection of presidential candidates?

(3) How much does money influence the selection of presidential candidates?

Note: Q2 was measured on a 0-100 scale. These questions were constructed solely for my specific purposes of examining an election effect.

*In both Appendix A.2 and Appendix A.3, dark blue ink identifies text and stimuli that are verbatim reproductions of the survey instruments used in both studies. Each question (Q1-Q4) took up the entire screen.

Q3. In your view, does corruption affect ...

	Not at all (0)	To a small extent (1)	To a moderate extent (2)	To a large extent (3)	Don't know / NA (NA)
Your personal and family life (Item A)	○	○	○	○	○
The business environment (Item B)	○	○	○	○	○
The political environment (Item C)	○	○	○	○	○
The culture and values in society (Item D)	○	○	○	○	○

Note: Q3 was only deployed in 2004, 2005, and 2006 (w/o the "culture and values" questions).

Q4. How seriously do you believe corruption affects ...

	Not significantly (1)	Somewhat significantly (2)	Very significantly (3)	Don't know / NA (NA)
Your personal and family life (Item A)	○	○	○	○
The business environment (Item B)	○	○	○	○
The political environment (Item C)	○	○	○	○
The culture and values in society (Item D)	○	○	○	○

Note: Q4 was only deployed in 2003. For Q4, since there was no "Not at all" option – scores ranged from 1-3. For Q3, the scores ranged from 0-3, with 0 being "Not at all."

A.4 Questions from the GCB that were Excluded

All survey items shown here in dark blue font have been reproduced from the 2013 GCB verbatim.

1. Questions Focused on Changes in Level of Corruption:

   GCB Q1. Over the past two years, how has the level of corruption in this country changed?

   GCB Q2. To what extent do you believe corruption is a problem in the public sector in your country? By public sector we mean all institutions and services which are owned and/or run by the government. Please answer on a scale of 1 to 5, where 1 is "not a problem at all" and 5 is "a very serious problem".

   GCB Q4. To what extent is this country's government run by a few big entities acting in their own best interest?

2. Questions Focused on Personal Experience with Corruption:

   GCB Q3.In your dealings with the public sector, how important are personal contacts and/or relationships to get things done?

   GCB Q7 (a). In the past 12 months, have you or anyone living in your household had a contact or contacts with one of the following [INSERT CATEGORY NAME 1–8]?

   GCB Q7 (b). If yes to Q7A, in your contact or contacts have you or anyone living in your household paid a bribe in any form in the past 12 months? Yes/no

   GCB Q8. What was the most common reason for paying the bribe/bribes? Please give only one answer.

   GCB Q12 (a). Have you ever been asked to pay a bribe?

   GCB Q12 (b). If yes, have you ever refused to pay a bribe?

3. Questions Focused on Fight Against Corruption

GCB Q5. How effective do you think your government's actions are in the fight against corruption?

GCB Q9. Do you agree or disagree with the following statement? "Ordinary people can make a difference in the fight against corruption."

GCB Q10. There are different things people could do to fight corruption and I am now going to ask whether you would be willing to do any of the following: Please answer "yes" or "no".

GCB Q11 (a). If yes to Q10F, to whom would you report it?

GCB Q11 (b). If no to Q10F, why not (report an incident of corruption)?

NOTE: *For further details about the GCB (questions or methodology), see Transparency International's website http://www.transparency.org/research/gcb/overview*

A.5 Two, Three, and Four-Factor Models Using Factor Analysis: WLS

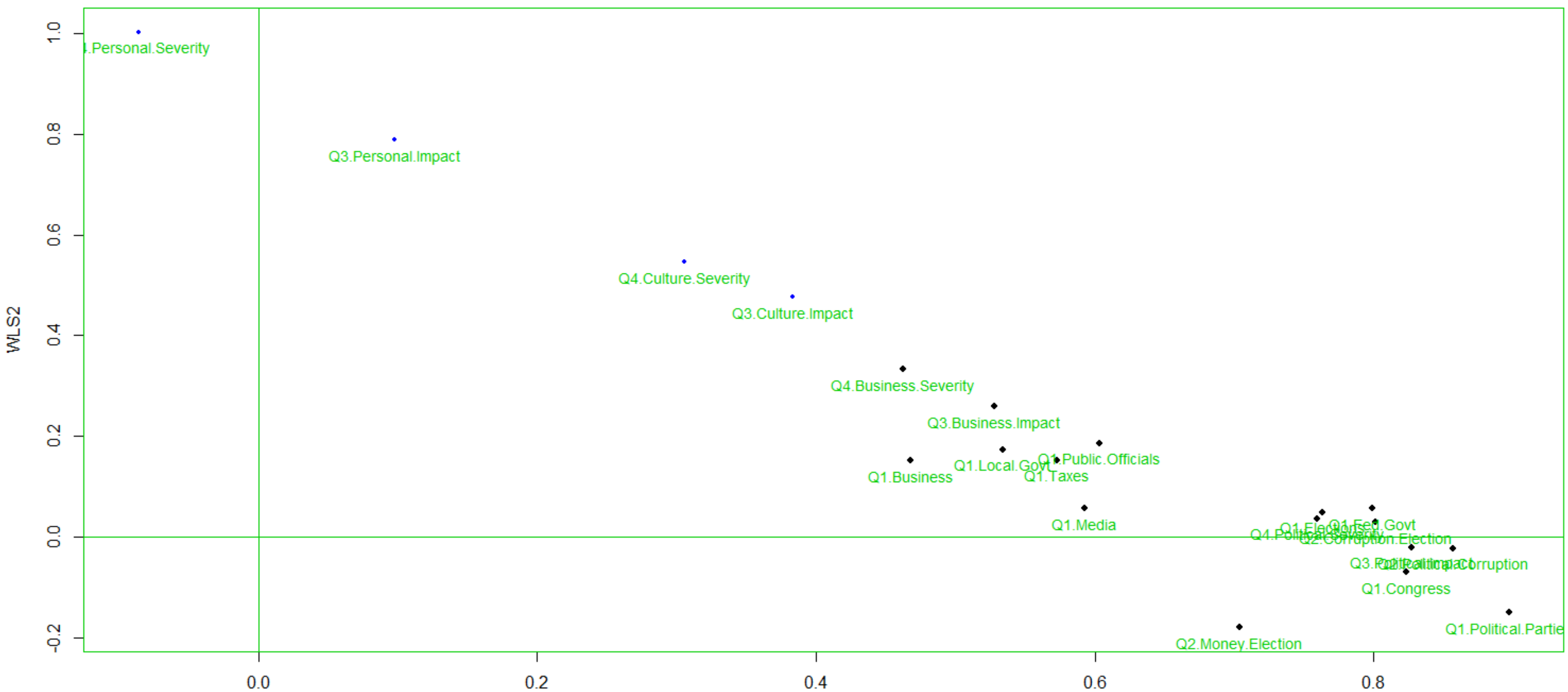


Figure A.2: Exploratory Factor Analysis —Simple Two Factor Model —Plot of Each Corruption Item's Standardized Loadings on Two Dimensions

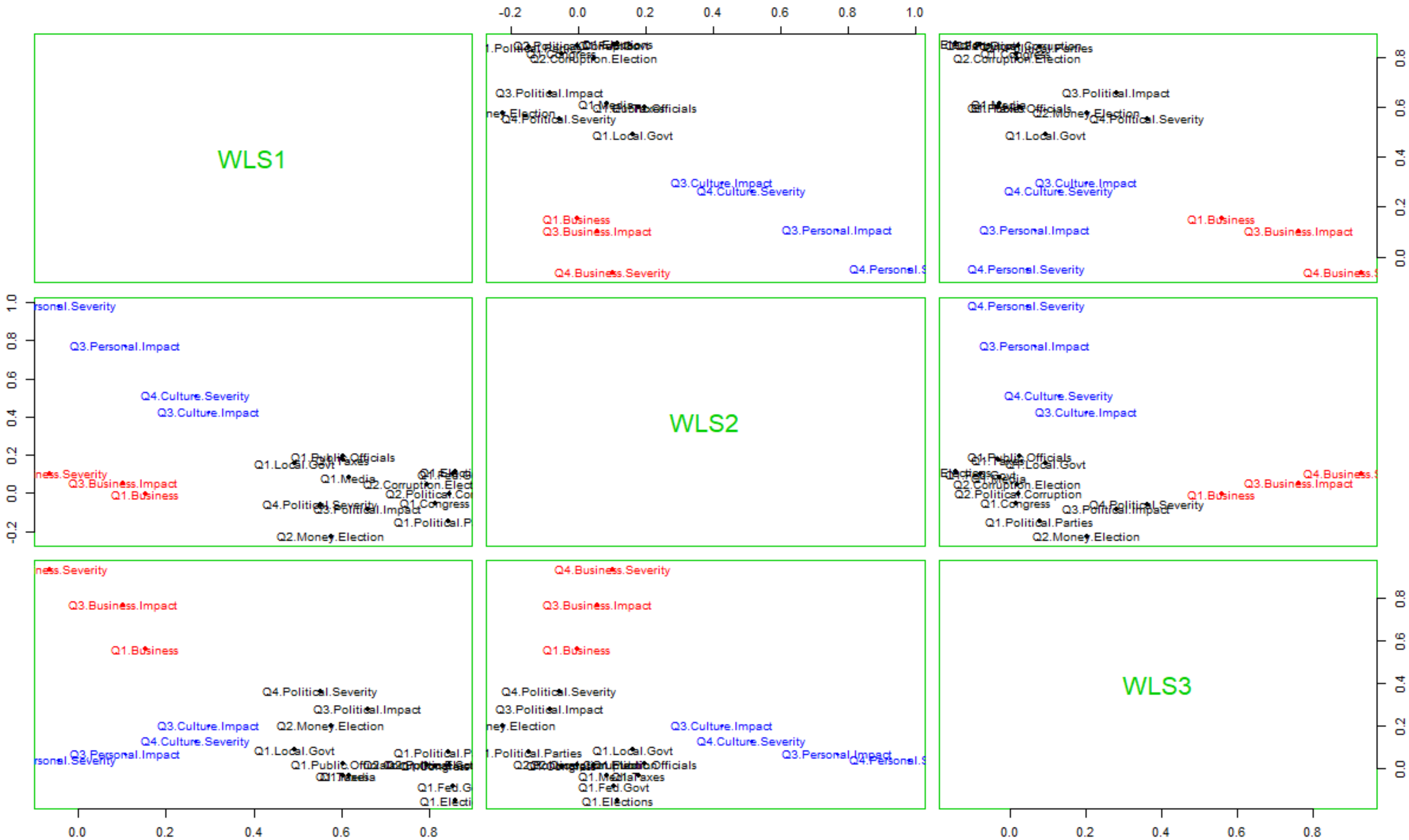


Figure A.3: Exploratory Factor Analysis —Three Factor Model —Plot of Each Corruption Item's Standardized Loadings on Three Factor Dimensions

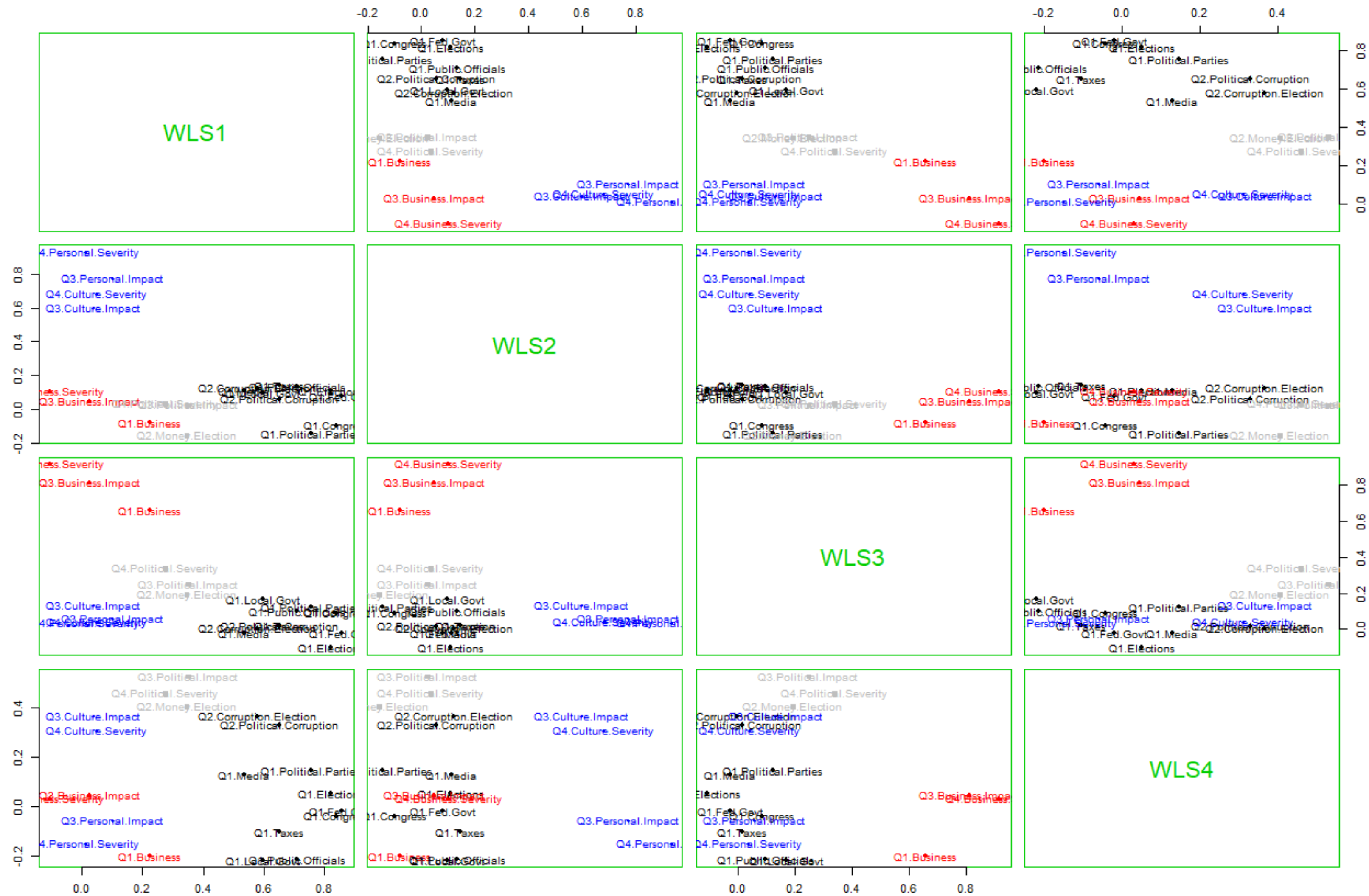


Figure A.4: Exploratory Factor Analysis —Four Factor Model —Plot of Each Corruption Item's Standardized Loadings on Four Factor Dimensions

A.6 Hierarchical Two-Factor Model

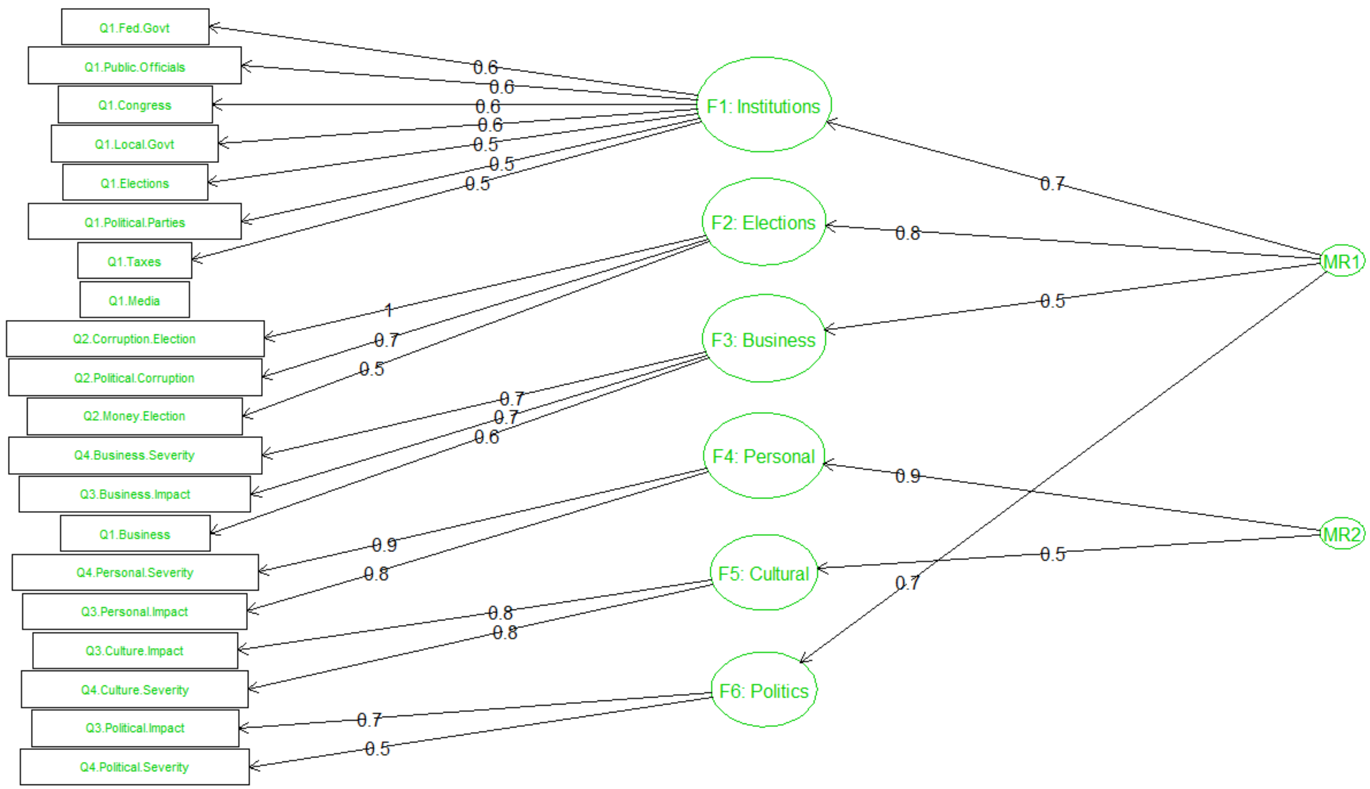


Figure A.5: Hierarchical Clustering Using WLS —Clustering Items into a Six-Factor to a Two-Factor Model

A.7 Agglomerative Hierarchical Clustering

A.7.1 Complete-Linkage Method

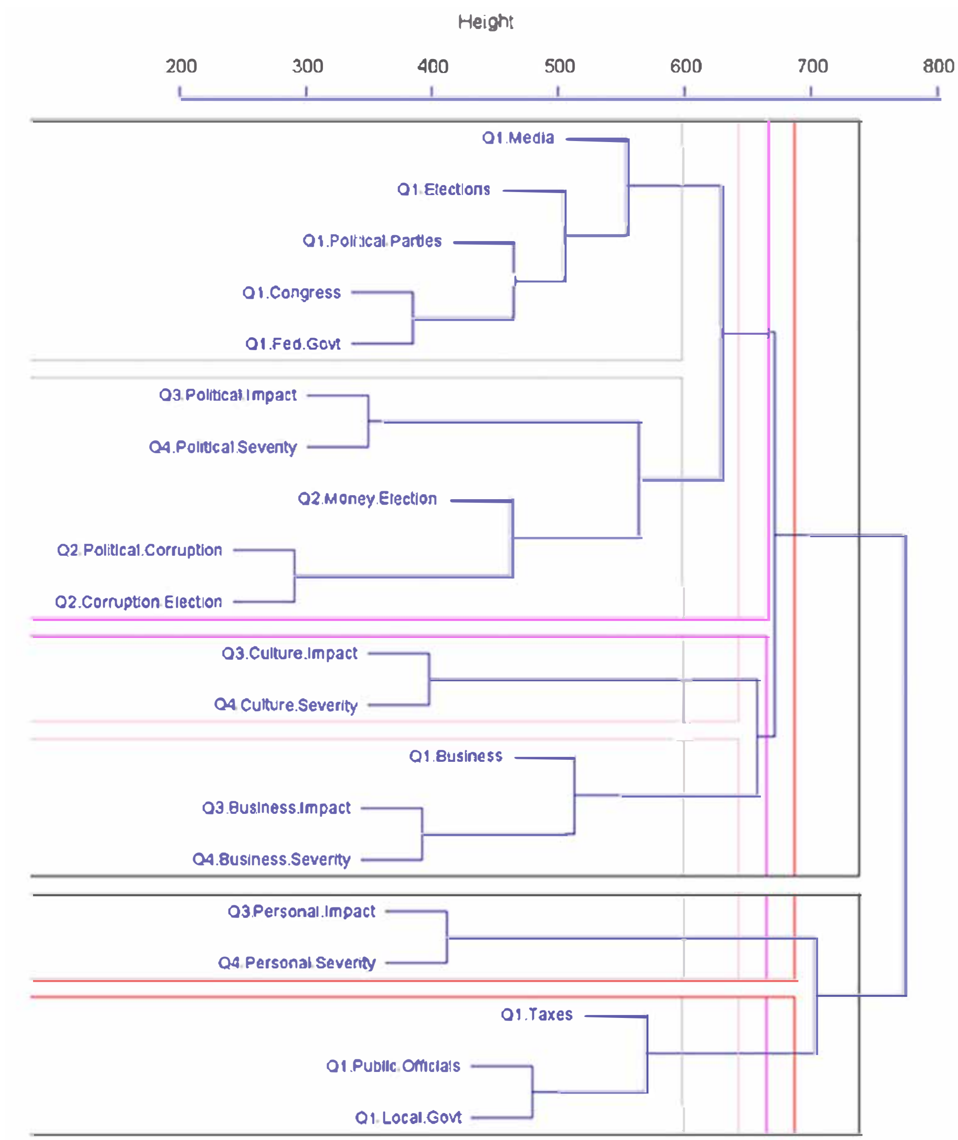


Figure A.6: Results of bottom-up, hierarchical clustering using the Complete-Linkage method (distance = Manhattan). The nested boxes show the clusters for different number of clusters (light-grey shows classification with 6 clusters; pink: 5 ; purple: 4; red: 3; and black: 2). This clustering method distinguished between measures of 'systemic corruption' (perceptions of overall, system-wide corruption, e.g. in US Congress) and measures of 'self-relevant corruption' (corruption perceived to be directly-relevant to life e.g. corruption that impacts local government, public officials, taxes, culture, or family).

A.7.2 Unweighted Average Method

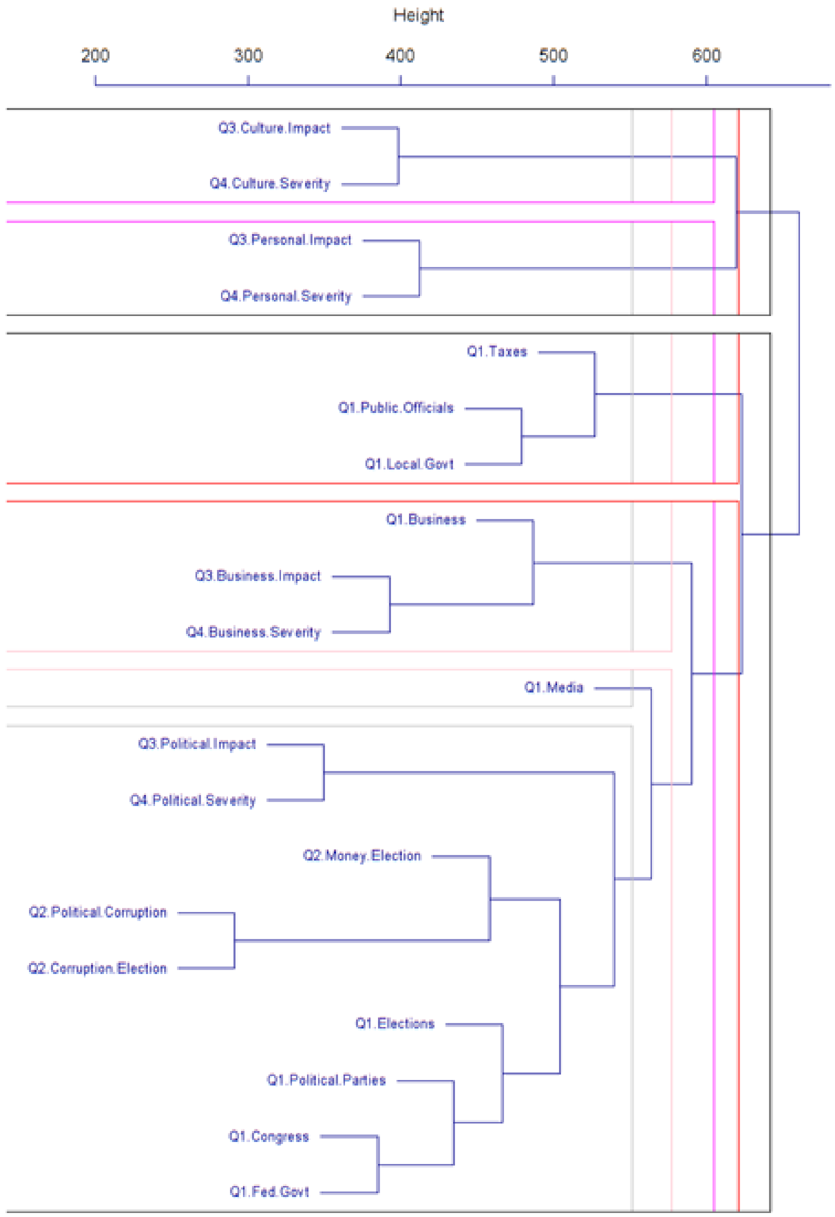


Figure A.7: Results of bottom-up, hierarchical clustering using the UPGMA method (distance = Manhattan). The nested boxes show the clusters for different number of clusters (light-grey shows classification with 6 clusters; pink: 5 ; purple: 4; red: 3; and black: 2). This clustering method naturally distinguished measures of 'systemic corruption' from measures of 'social corruption.'

A.8 Measurement Model - Linking Latent Variables to Survey Items

(A). PERCEIVED CORRUPTION IN THE POLITICAL ENVIRONMENT

Q1. Political Parties: To what extent do you see [Political Parties] affected by corruption in the US?

Q2. Political Corruption: How much of a problem is political corruption in the US?

Q2. Corruption Election: How much does corruption influence the selection of presidential candidates?

Q2. Money in Elections: How much does money influence the selection of presidential candidates?

Q3. Political Impact: In your view, does corruption affect the political environment?

Q4. Political Severity: How seriously do you believe corruption affects the political environment?

(B). PERCEIVED INSTITUTIONAL CORRUPTION

Q1. Media: To what extent do you see [the Media] affected by corruption in the US?

Q1. Public Officials: To what extent do you see [Public Officials / Civil servants] affected by corruption in the US?

Q1. Taxes: To what extent do you see [Taxes] affected by corruption in the US?

Q1. Elections: To what extent do you see [Elections] affected by corruption in the US?

Q1. Fed. Govt: To what extent do you see [Federal Government] affected by corruption in the US?

Q1. Local Govt: To what extent do you see [Local Government] affected by corruption in the US?

Q1. Congress: To what extent do you see [Congress] affected by corruption in the US?

(C). PERCEIVED CORRUPTION IN THE BUSINESS ENVIRONMENT

Q1. Business: To what extent do you see [Business / Private Sector] affected by corruption in the US?

Q3. Business Impact: In your view, does corruption affect [the business environment]?

Q4. Business Severity: How seriously do you believe corruption affects [the business environment]

(D). PERCEIVED CORRUPTION: PERSONAL AND FAMILY LIFE:

Q3. Personal Impact: In your view, does corruption affect [your personal and family life]?

Q4. Personal Severity: How seriously do you believe corruption affects [your personal and family life]?

(E). PERCEIVED CORRUPTION: CULTURE AND VALUES IN SOCIETY:

Q3. Culture Impact: In your view, does corruption affect [the culture and values in society]? '

Q4. Culture Severity: How seriously do you believe corruption affects [the culture and values in society]?

APPENDIX B

SUPPLEMENTARY MATERIALS AND ANALYSES FOR PART II: ELECTION OUTCOMES AND LEGITIMACY-BASED THEORIES OF VOLUNTARY COMPLIANCE

B.1 Classification Based upon Simple Majority vs Absolute Majority

When determining who won in an election, the American political system tends to evaluate voting results using the simple majority standard (i.e. the candidate with the most votes is declared the winner, independent of whether they won a majority). As a result, the simple majority procedure is most frequently used to classify election outcomes at the county-level as either Democrat or Republican (for example, see "Presidential Election Results," 2017).

In the current project, however, a simple-majority based rubric is inappropriate for classifying election outcomes at the county level. In order to understand why, we must recall two central assumptions: (a) counties are treated as individual units of analysis; (b) our primary interest is in the county's perceptions and affective evaluations of the winning presidential candidate. In an ideal world, we would have survey instruments that directly assessed these measures. However, in the absence of such data, we rely upon voting patterns as a proxy measure for the "degree to which a county supports a given candidate."

Given that measurement goal, when producing binary classifications of counties, it is more appropriate to rely upon an *absolute majority* based standard because it allows us to ensure that for counties classified as "Democrat" the majority of residents will react to the election outcome of a Democrat as a victory and the election of a Republican as a loss, and vice-versa.

To understand why the *absolute majority* standard is more appropriate, consider the case of a county where 45% supported the Democrat presidential candidate, 40% supported the Republican presidential candidate, and the remaining 15% supported third-party candidates. If a Democrat won the Presidency, under *simple majority* classification, the county would be seen as a "Democrat county" and would be classified as having "supported the winning candidate" —which would be inaccurate since a majority of the county did not support the candidate (40% Republican + 15% Other). Such patterns of voting were seen in elections

Table B.1: Counties Classified Using Simple Majority

	Democrat	Republican	Third Party
1992	1,383	1,361	9
1996	1,392	1,361	0
2000	609	2,144	0
2004	517	2,236	0
2008	793	1,960	0
2012	627	2,126	0
2016	424	2,329	0

where third party candidates garnered unusually strong levels of support (e.g. 1992, 1996, and 2016)

Other contemporaneous work has chosen to exclude third party voting and treat the two-party vote share as the entire voter base (Mian et al., 2018). Although this eliminates any difference between the *simple majority* and *absolute majority* rubrics for binary classification, it posses a problem for the current work since the tax data includes data from all residents, not just the residents that support the two main parties. And, as the example above shows, application of the incorrect standard can produce misclassifications.

The distinction and use of the *absolute majority* standard is especially important in the case of the Presidential Elections in 1992 and 1996, where Ross Perot was an anomalous 3rd party candidate that dramatically split the Republican vote and even won 13 counties outright (see Table B.1).

As we can see in the tables below, the choice of standard between *simple majority* and *absolute majority* has a significant impact on the number of counties that are classified as either Democrat or Republican, especially in the election years 1992 and 1996.

As we see in Table B.1, for 1992 and 1996, the number of counties classified as Democrat is approximately 1400. However, starting in the 2000s, this number drops to between 400–800. The reason for this shift can be seen in Table B.2: the number of counties that failed to produce a majority in favor of either of the two main parties in 1992 is 2021 (i.e. approxi-

Table B.2: Counties Classified Using Absolute Majority

	Democrat	Republican	No Majority
1992	434	298	2021
1996	758	765	1230
2000	498	1993	262
2004	487	2207	59
2008	730	1894	129
2012	577	2075	101
2016	340	2211	202

Table B.3: Counties Misclassified When Using a Simple Majority Standard

	1992	1996	2000	2004	2008	2012	2016
Dem. Counties Misclassified	949	633	111	30	63	50	84
Rep. Counties Misclassified	1062	596	151	29	66	51	118

mately 2/3 of all US counties). This number falls to 1230 in 1996 (i.e. approximately 1/3 of US counties) and then hovers between 100–300 (i.e. between 3–10% of US counties) for the remaining elections.

Table B.3 below contrasts the two classification approaches and demonstrate the incidence of misclassification that occurs when a simple majority approach is used as compared with an *absolute majority* approach. In all the cases show in in table B.3, due to the presence of a 3rd party candidates, the candidate only secured a simple majority. The first row shows the number of counties classified as Democratic despite the fact that the Democratic Presidential Candidate failed secure a majority (i.e. received less than 50% of the votes). The second row shows the number of counties that would be classified as Republican, even though the Republican candidate failed to receive a majority of support.

As we see, in 1992, 949 counties (approximately 68% of the Democratic Counties) would be misclassified: i.e. they are classified as Democrat even though less than 50% of votes were class in support of Bill Clinton; similarly 1062 (78%) would be misclassified: i.e. they are classified as Republican even though less than 50% supported George H. W. Bush. A similar (though less pronounced) trend can be seen in the 1996 election results.

The majority of misclassifications are limited to the elections in 1992 and 1996. This occurrence of misclassification in the first two elections should guide the way in which we aggregate the binary election-level partisanship classifications to generate the county-level partisanship status. Specifically, it may be advisable to adjust the aggregation approach in a manner that relies mostly on the 5 elections between 2000 and 2016.

B.2 Supplementary Hypotheses Graphs

B.2.1 Recreating Primary Hypotheses Graphs: Using AGI over PI

Full Time Series

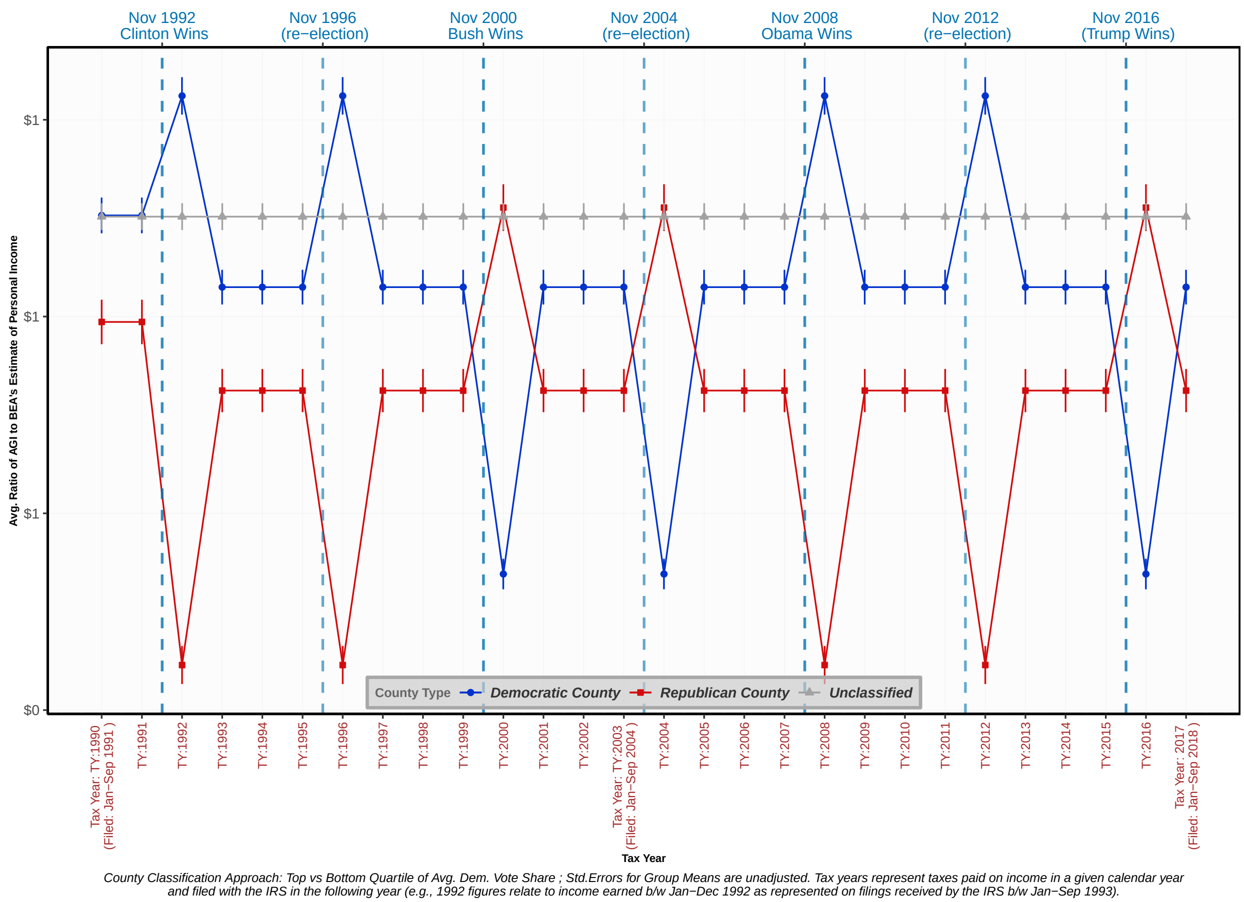


Figure B.1: Legitimacy Hypothesis: Effects of Elections on AGI over PI over Time

Varying Time Dynamics - Election Effects Persist for 1-3 Years

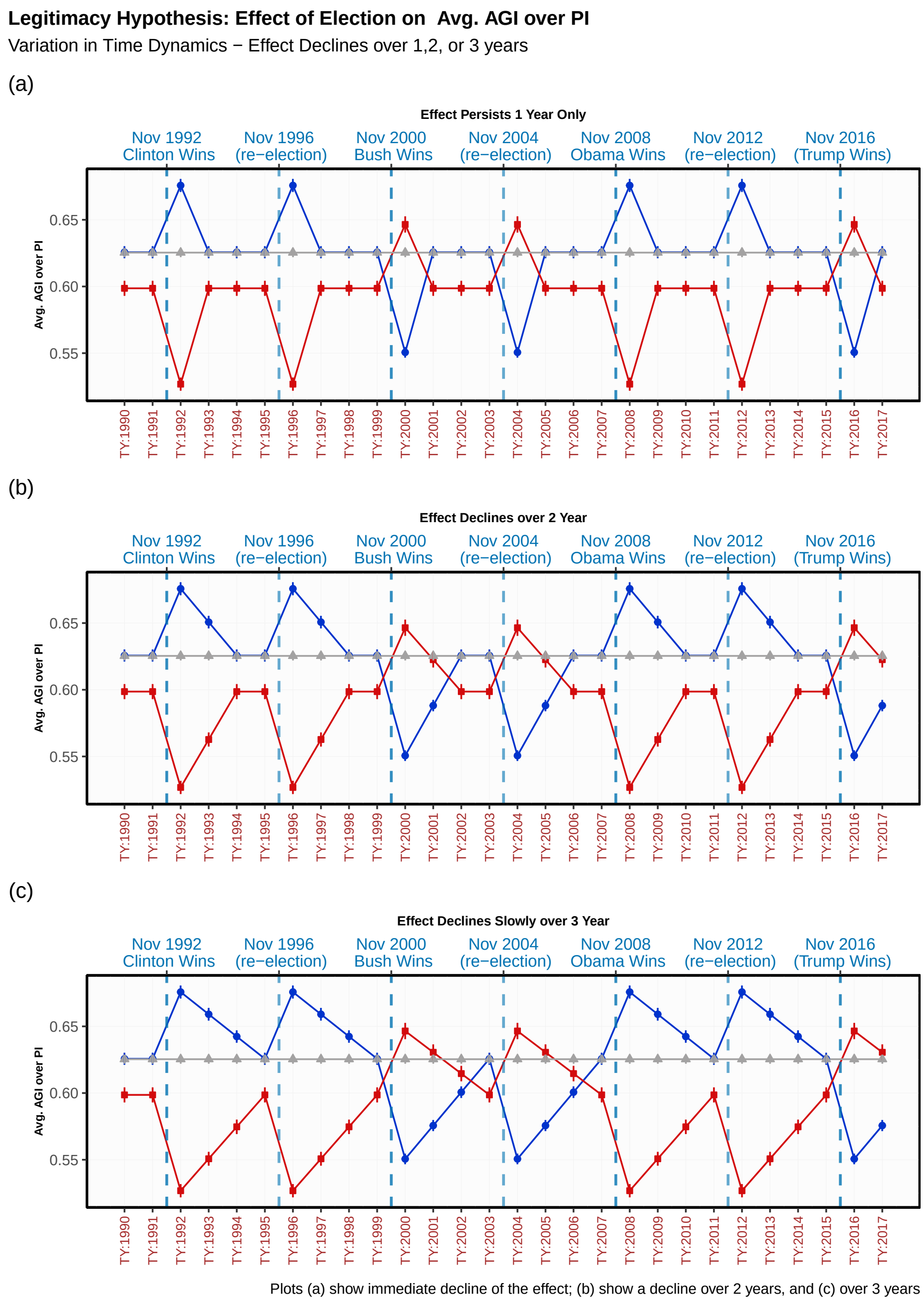


Figure B.2: Legitimacy Hypothesis: Effects of Elections on AGI over PI over the Time Sample —Election Effects Shown with Varying Time Dynamics of 1, 2, or 3 Years

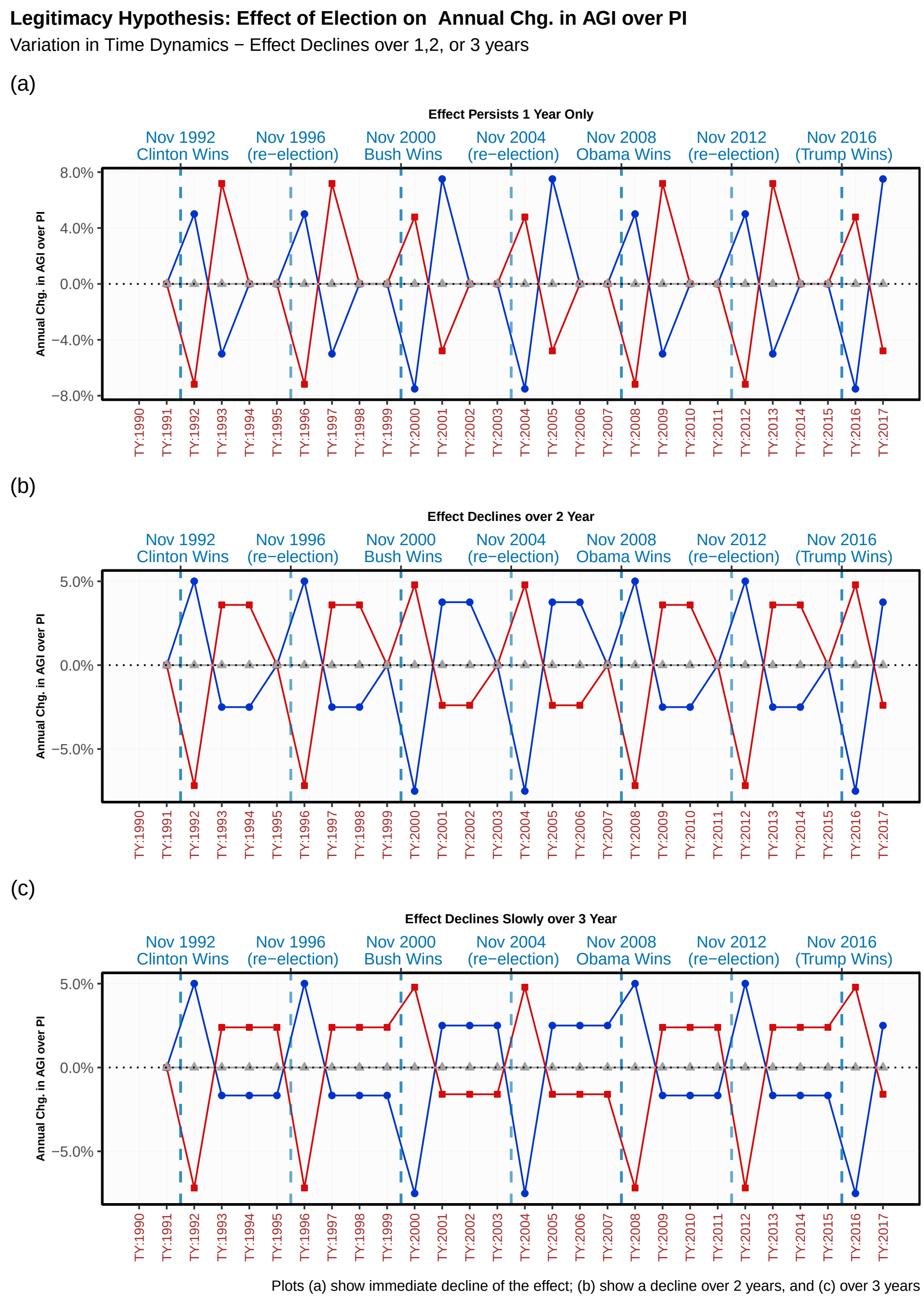


Figure B.3: Legitimacy Hypothesis: Effects of Elections on Annual Change in AGI over PI over the Time Sample —Election Effects Shown with Varying Time Dynamics of 1, 2, or 3 Years

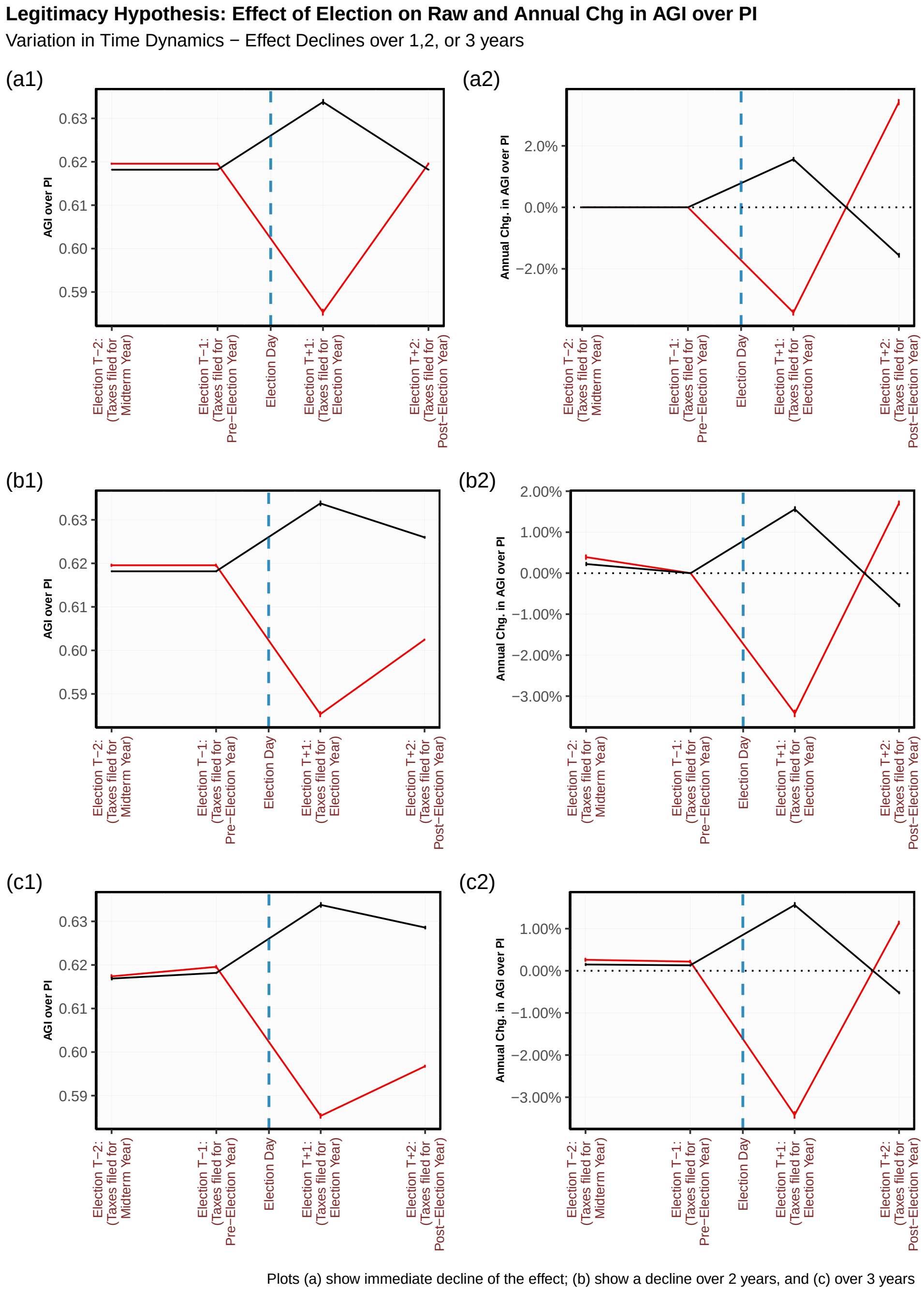


Figure B.4: Legitimacy Hypothesis: Effect of Election on Raw and Annual Chg in AGI over PI by Winners and Losers of Elections —Election Effects Shown with Varying Time Dynamics of 1, 2, or 3 Years

B.3 Supplementary Graphical Analyses: AGI per Return

B.3.1 Data Distribution

Distribution by County Classification

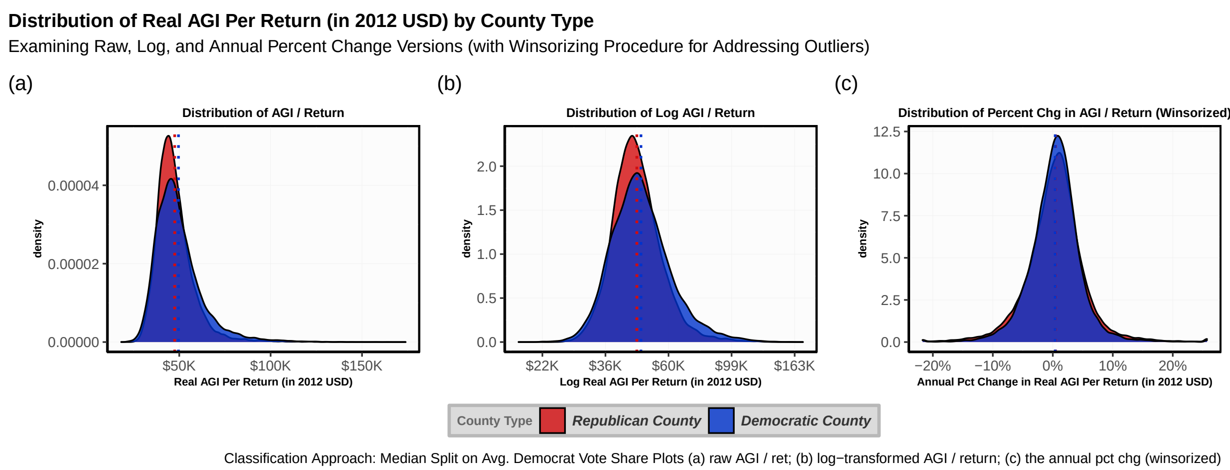


Figure B.5: Distribution of Adjusted Gross Income (AGI) per return for all U.S. counties across the 28 year sample after classifying according to political affiliation. The classification was performed by averaging the Democratic Vote Share across all 7 elections for each county. Then, the median was used to classify counties as either 'Democratic' and 'Republican'. Section 10.2.2 Aggregating Partisanship Status Across Elections describes in greater detail the main approaches used in this paper for classifying counties according to partisanship.

Classification by Median Split

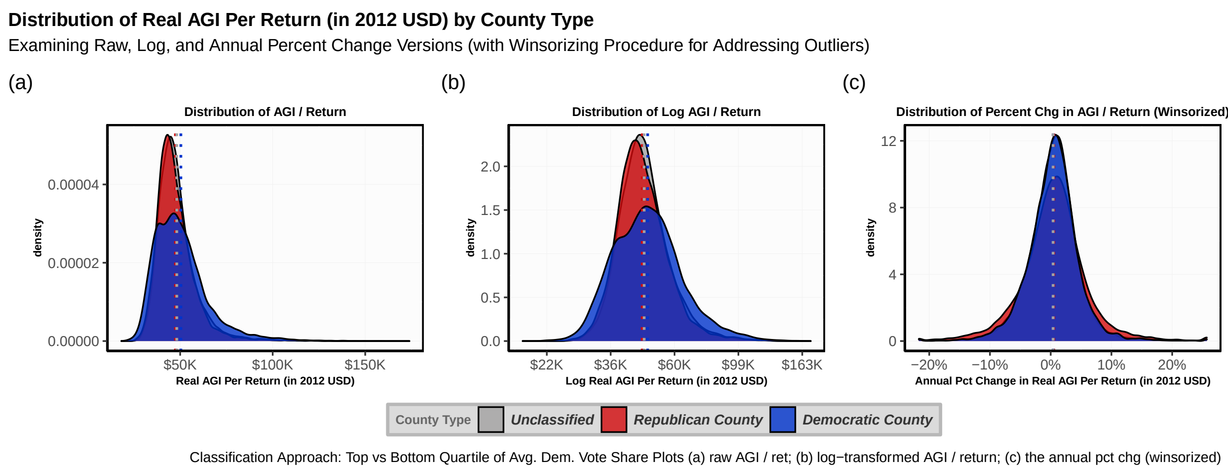


Figure B.6: Distribution of Adjusted Gross Income (AGI) per return for all U.S. counties across the 28 year sample after classifying according to political affiliation. The classification was performed by averaging the Democratic Vote Share across all 7 elections for each county. Then, the top and bottom quartile of all counties were classified as 'Democratic' and 'Republican' counties respectively. Section 10.2.2 Aggregating Partisanship Status Across Elections describes in greater detail the main approaches used in this paper for classifying counties according to partisanship.

Classification by Top vs Bottom Quartile

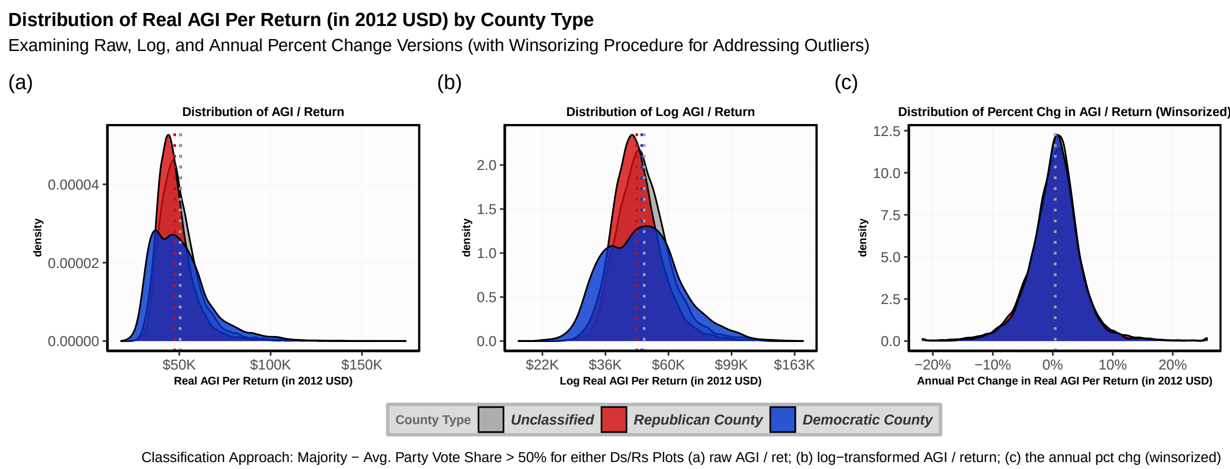


Figure B.7: Distribution of Adjusted Gross Income (AGI) per return for all U.S. counties across the 28 year sample after classifying according to political affiliation. The classification was performed by averaging the Democratic and Republican Vote Share across all 7 elections for each county. The county was classified as belonging to a party if the average vote share for the party across 7 elections was greater than 50% (i.e. on average, the party had majority support in the county). Section 10.2.2 Aggregating Partisanship Status Across Elections describes in greater detail this approach to classifying counties according to partisanship.

Classification by Majority Threshold - Avg. Party Vote Share $> 50\%$

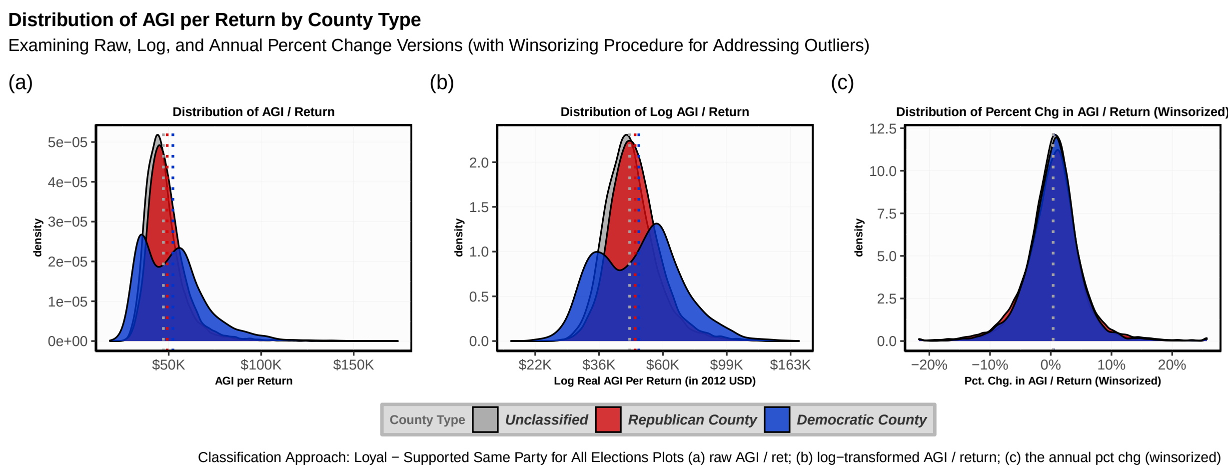


Figure B.8: Distribution of Adjusted Gross Income (AGI) per return for all U.S. counties across the 28 year sample after classifying by loyal partisanship. The classification examined voting patterns across all 7 elections. If a plurality of voters in a county supported the same party in every election, the county was classified as a loyal partisan. All counties that flipped between parties were left unclassified. Section 10.2.2 Aggregating Partisanship Status Across Elections describes this approach in greater detail.

Classification by Loyalty - Voted for Same Party in all 7 Elections

B.3.2 Full Time Series

The current section presents full time series graphs. The graphs are presented with the primary purpose of allowing the reader to examine the broad trends across the entire 27 year sample. The time series are presented with counties grouped according to their party affiliation (Democrat or Republican) or as "unclassified" (for counties with no clearly discernible party affiliation). This should allow the reader to both review overall trends in the data, to compare the overall trends for Democratic, Republican and Unclassified counties, and to develop an intuition for the major, over-arching patterns in a variable's evolution over time. These graphs are primarily intended as a tool for transparency and to reassure the reader as they examine the graphical analyses in the main text, which often collapse across years and present data for election cycles only.

Log AGI per Return

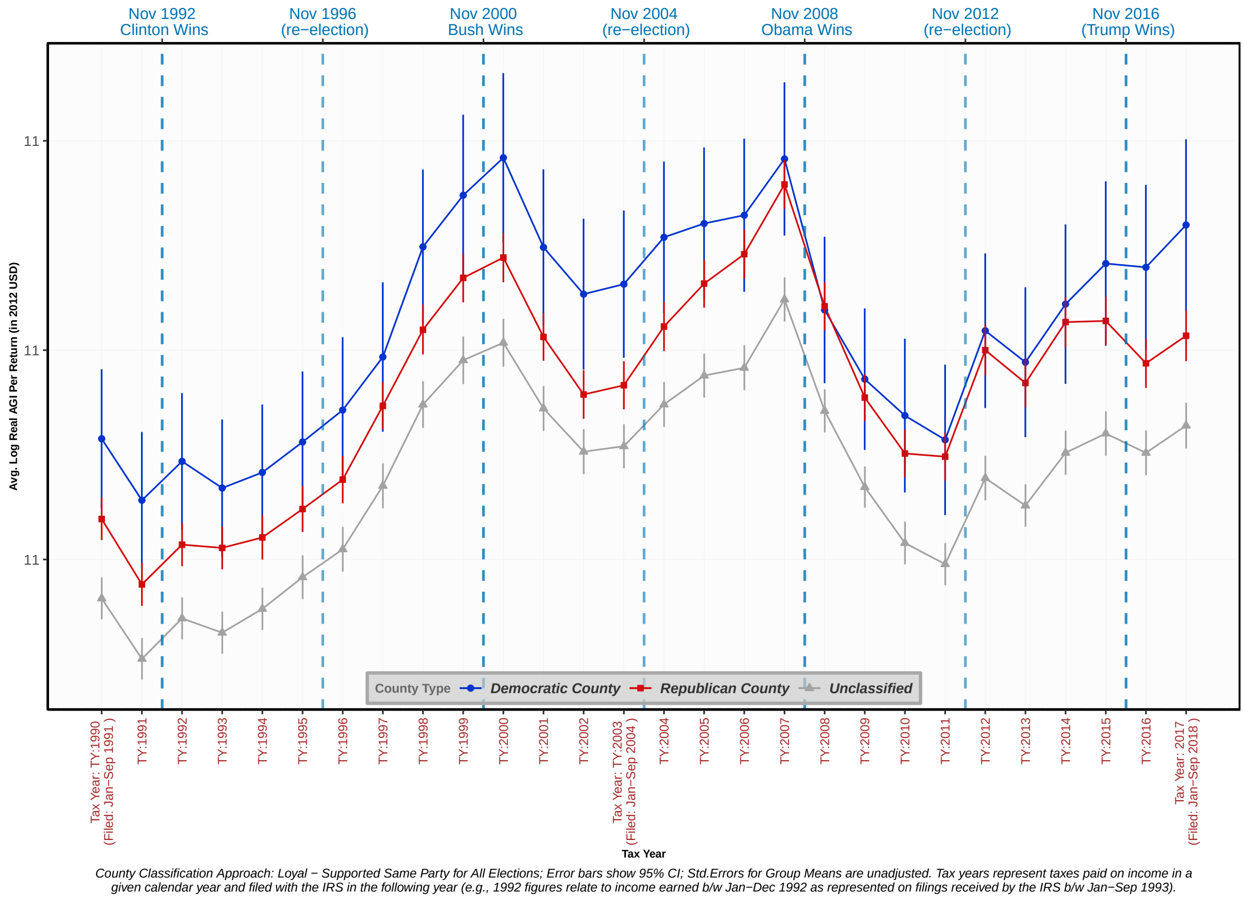


Figure B.9: (log) AGI per Return for Democrat and Republican Counties over Time (inflation-adjusted 2012 USD)

AGI per Return - Removing Year Fixed Effects

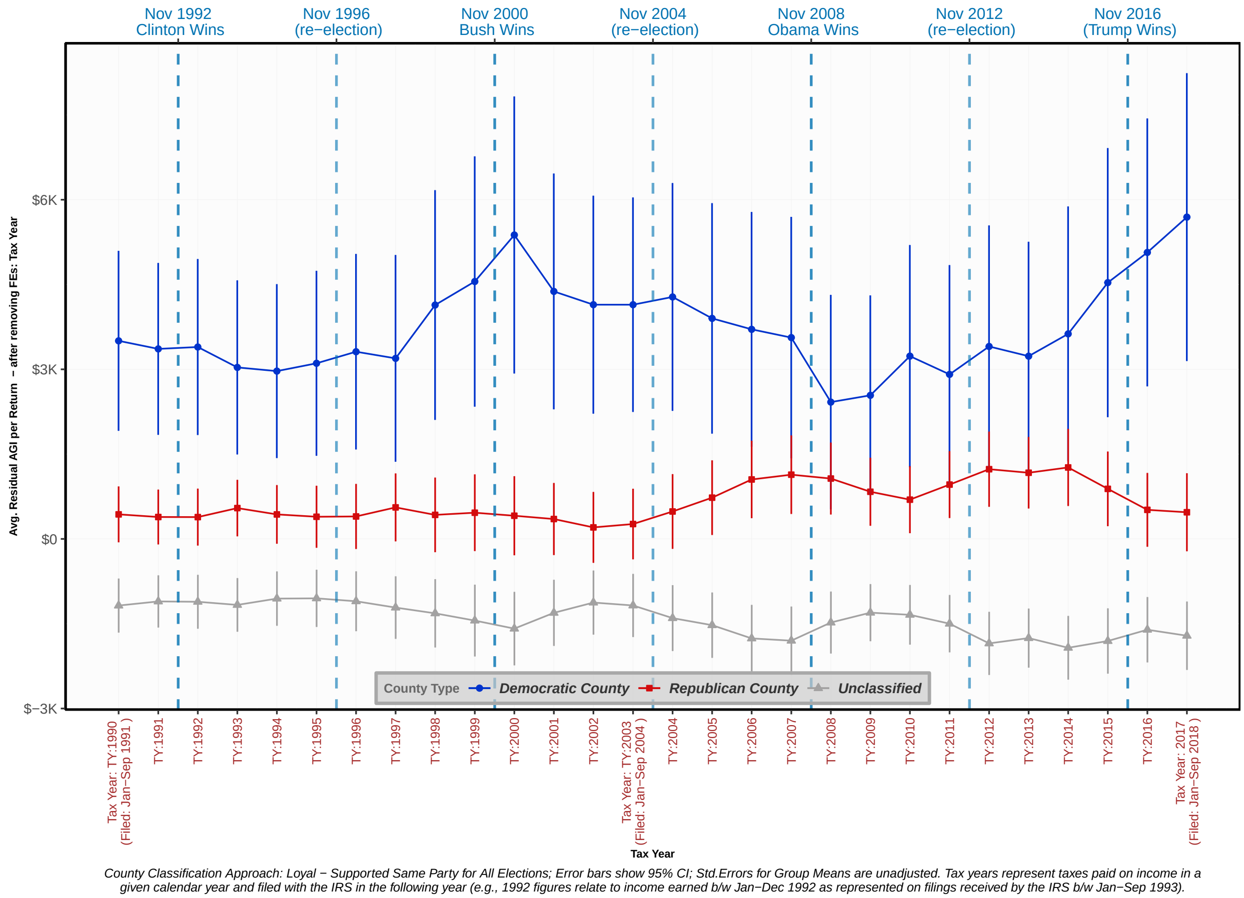


Figure B.10: AGI per Return for Democrat and Republican Counties over Time After Removing Year Fixed Effects

AGI per Return - Shown as Difference Variable

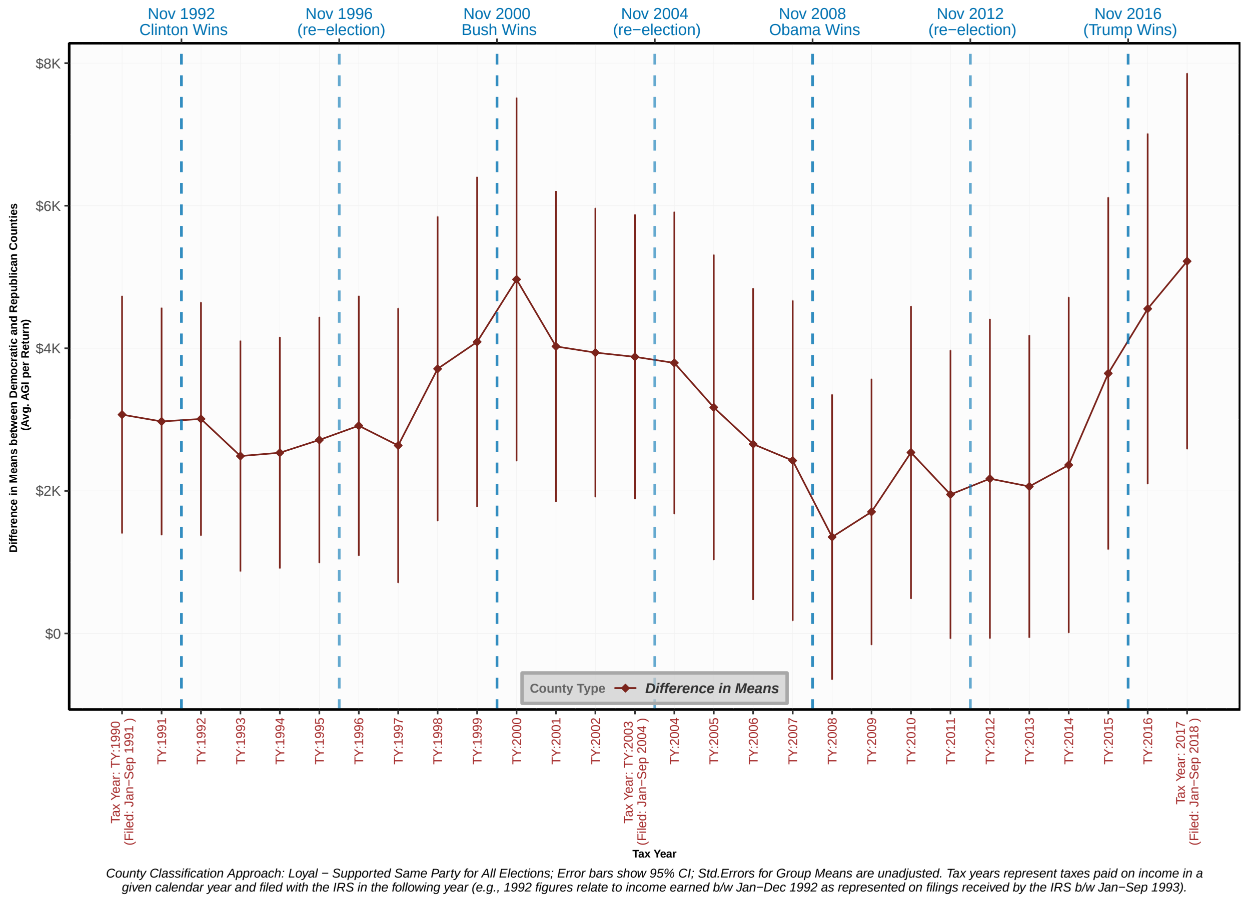


Figure B.11: Difference in AGI per Return between Democrat and Republican Counties over Time (inflation-adjusted 2012 USD)

Annual Pct. Change in AGI per Return - Removing Year Fixed Effects

B.3.3 Panel by Election

Annual Pct Change in AGI per Return After Removing Year Fixed Effects - Spearated by Election -

In order to examine the elections trends, in Figure B.12, I present the annual percentage change in AGI per return for Democratic and Republican counties split by election cycle with all time fixed effects removed. This graph supplements Figure 12.8 in the main text.

I begin with the top, left-most graph showing the 1992 Election Cycle. The residual annual change for both Democratic and Republican counties hovers at 0% after removing time fixed effects for both Tax Year 1991 and 1992 before diverging for Tax Year 1993 with a positive residual growth rate for Republican Counties and a negative residual growth rate for Democratic Counties. The same identical pattern is seen for the 1996 election cycle. In both cases, this positive residual growth rate for Republican counties would be read as supporting the alternative hypotheses, where losing an election is accompanied by a decrease in cheating and winning is accompanied by an increase (i.e. either Hypothesis 2: Moral Licensing and a Winner Effect or Hypothesis 3: Perceived Enforcement Risk). It should be noted that the election effect takes place for the first full year under the presidency (i.e. Tax Year 1993; 1997) and no effect is discernible for the election year itself (i.e. Tax Year 1992; 1996 do not show an election effect).

In the 2000 Election Cycle, prior to the election, in Tax Year 1999, both types of counties show a residual 0% growth rate. The election year produces a positive 0.7% residual growth rate for Democratic counties followed by a sharp negative -0.6% residual growth post-election. Republican counties show a 0% residual growth rate for both years. Thus, we see an additional 1.3% drop in the growth rate for Democratic counties, which is a pattern in line with the Legitimacy Hypothesis, where losing an election is accompanied by a decrease in tax disclosure and winning is accompanied by an increase. Like 1992 and 1996, the timing of the election effect appears between the election-year (TY 2000) and the post-election year (TY 2001).

For the 2004 election cycle, again, prior to the election, we see approximately 0% residual growth rate for both types of counties, with a divergence in Tax Year 2005 —where we see a positive (approx. +0.5%) residual growth rate for Republican Counties and negative (approx. −1%) residual growth rate for Democratic counties. This pattern —like the prior election —would be in line with the Legitimacy Hypothesis.

The same pattern is seen in the following election of 2008, where —following a dip in Tax

Year 2008 —we see a spike towards positive residual growth among Democratic counties and a corresponding drop towards negative residual growth among Republican counties —as one would expect under the legitimacy hypothesis. The 2012 Obama re-election cycle does not appear to show any clear patterns. The 2016 election pattern also appears to be in line with the legitimacy hypothesis. Prior to the election, we see Democratic Counties with a positive residual growth rate of (approx. 1%) and Republican Counties with a negative residual growth rate of (approx. -0.5%). Following the election, we see a decline in the residual growth rate for Democratic counties and a corresponding increase in residual growth rate for Republican counties.

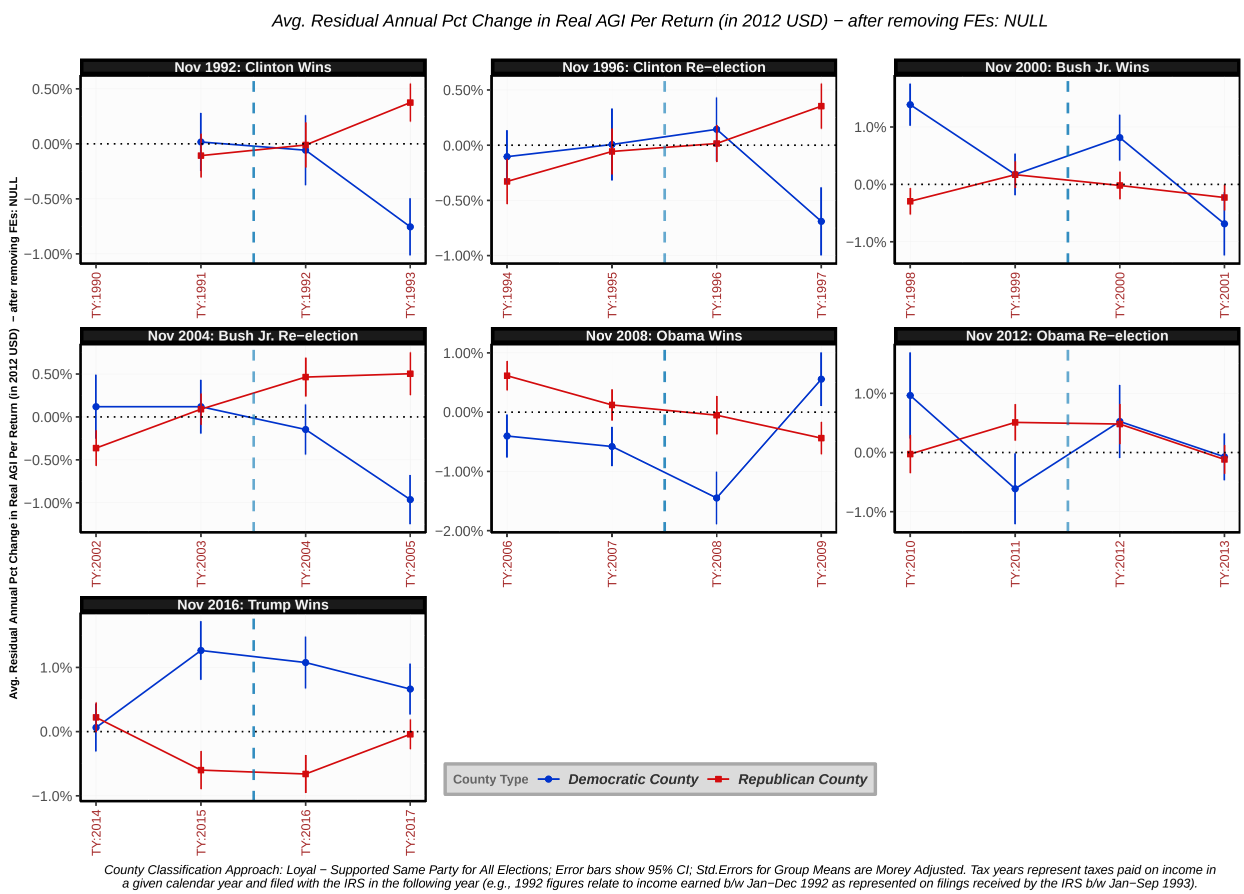


Figure B.12: Annual Percent Change in AGI per Return by Political Affiliation —Full Time Series Segregated by Election (Winsorized Values)

Pct Chg in AGI per Return By Election Cycle - Difference in Means

In order to examine the elections trends, in Figure B.13, I present the annual percentage change in AGI per return represented as a difference in means for Democratic and Republican counties. Each panel below captures the trends for an individual election cycle. This graph supplements Figure 12.8 in the main text. The values in the graphs can be read as follows: a value of +1% would mean that Democratic counties showed on average 1% greater growth in AGI per return than Republican counties; conversely, a value of -1% would mean that growth rate for Democratic counties was 1% lower on average than for Republican counties.

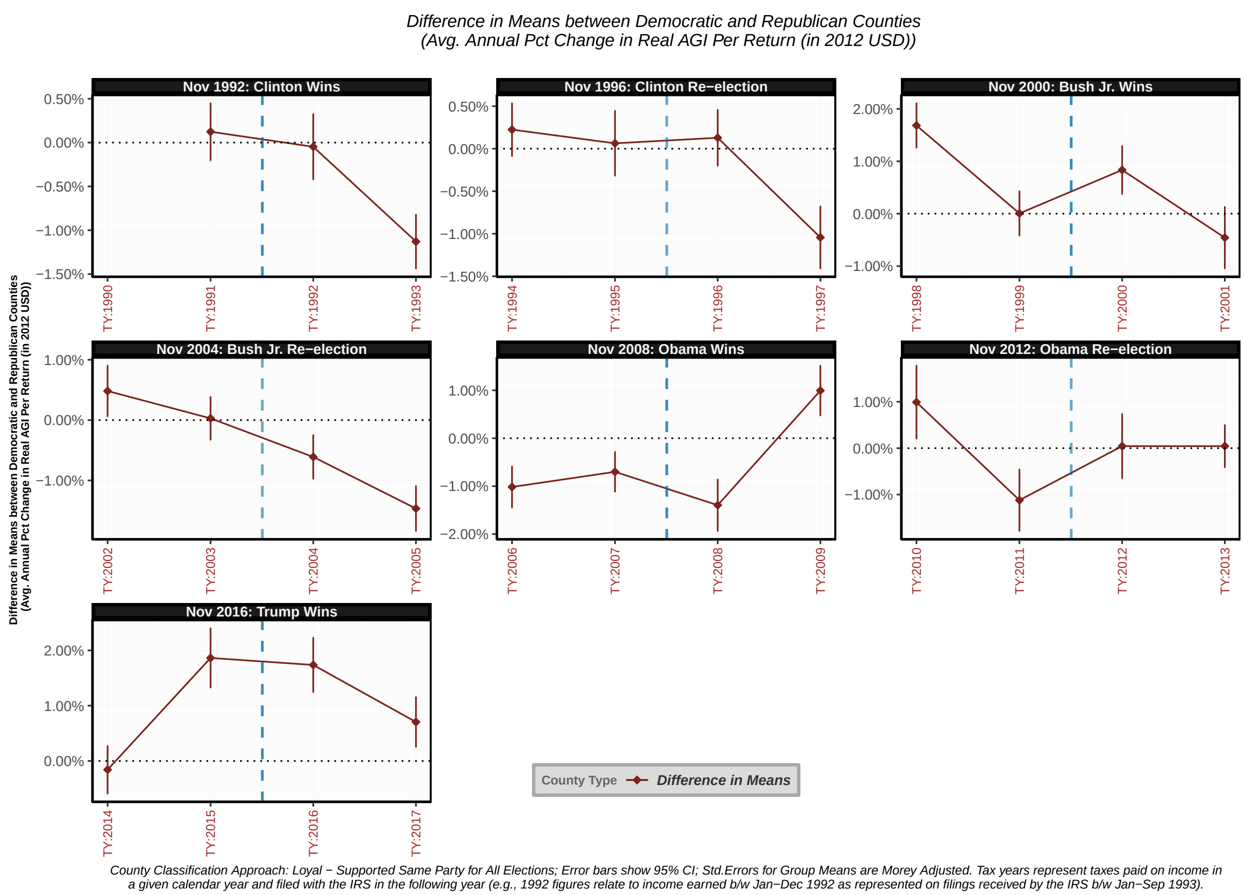
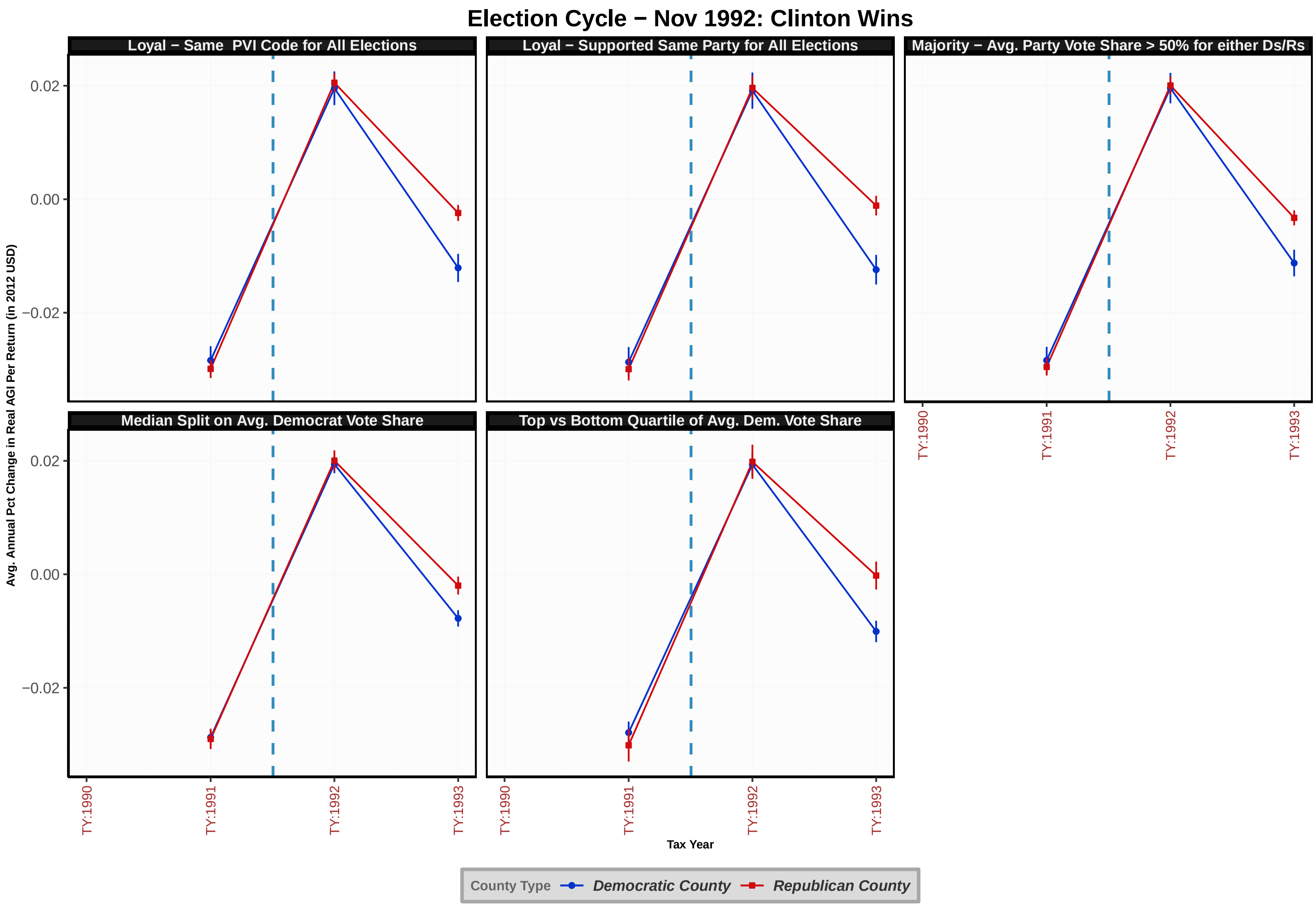


Figure B.13: Annual Percent Change in AGI per Return —Difference in Democratic and Republican Counties —Segregated by Election (Winsorized Values)

B.3.4 Comparing Classification Approaches: By Election Cycle

In the current section, I present each election graph according to all classification approaches side-by-side for comparison. This allows the reader to examine the effect of classification for the trends seen here for all elections individually. The collected graphs are presented in the main analyses section: Trends in AGI per Return.



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.14: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 1992

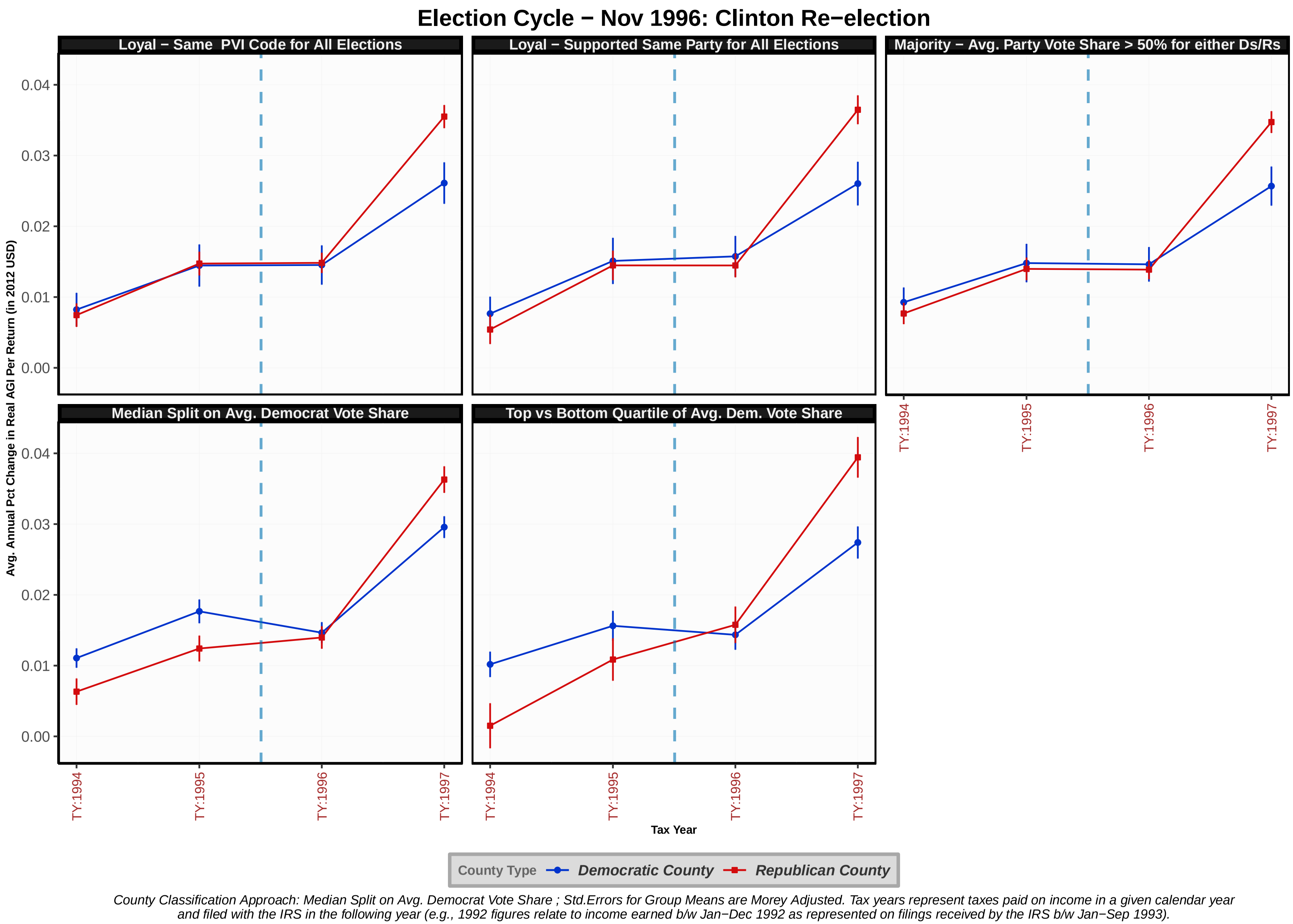
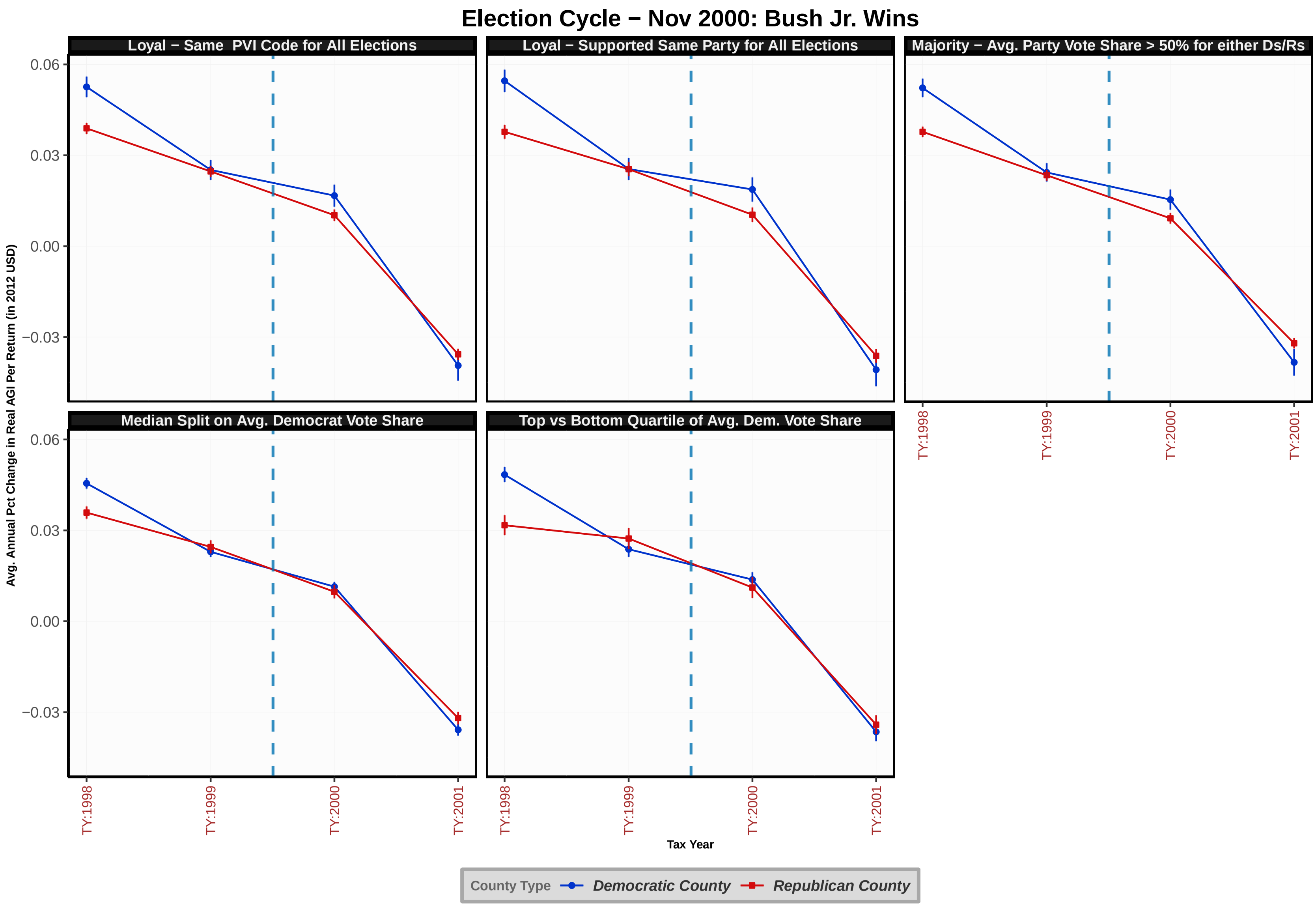
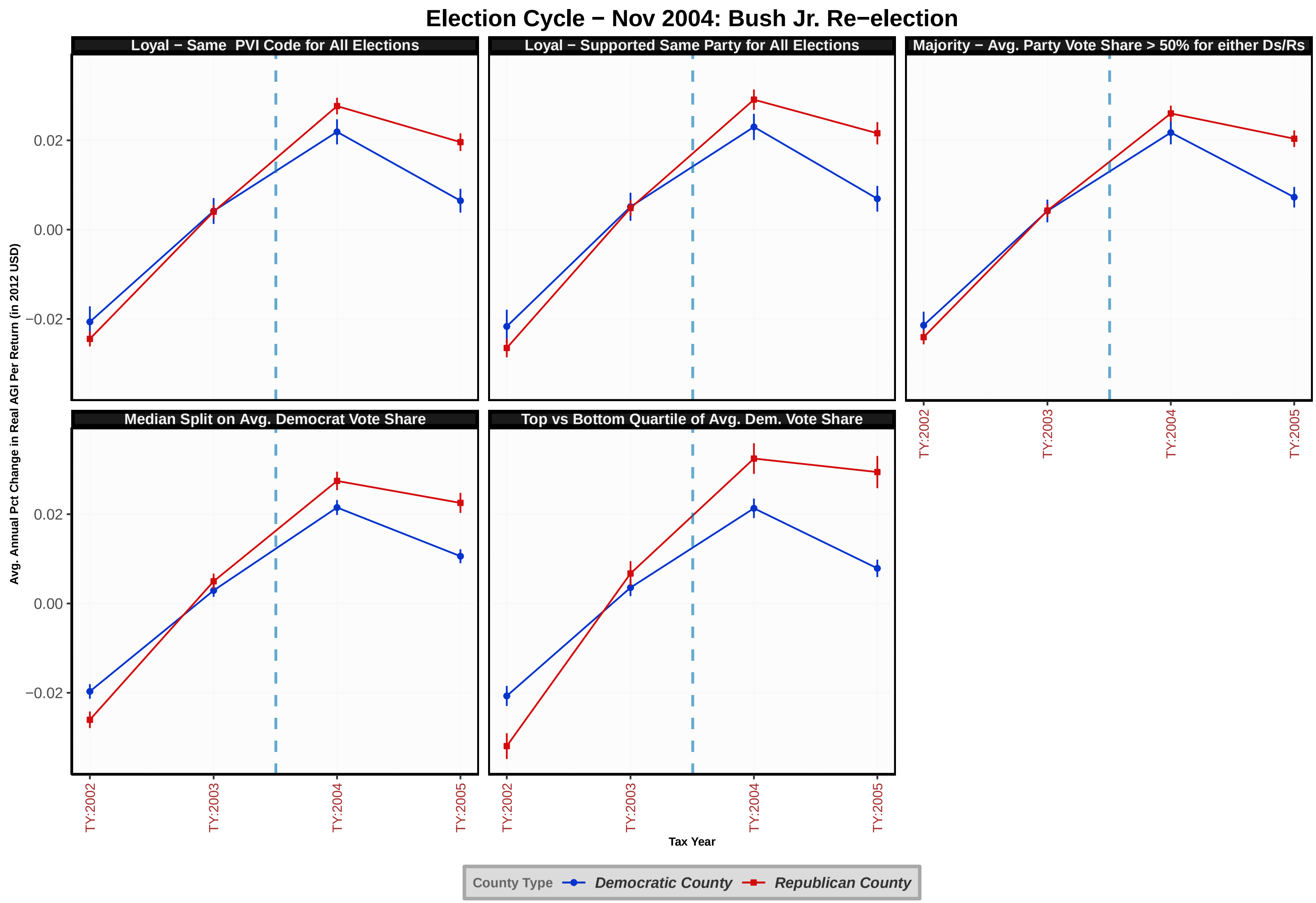


Figure B.15: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 1996



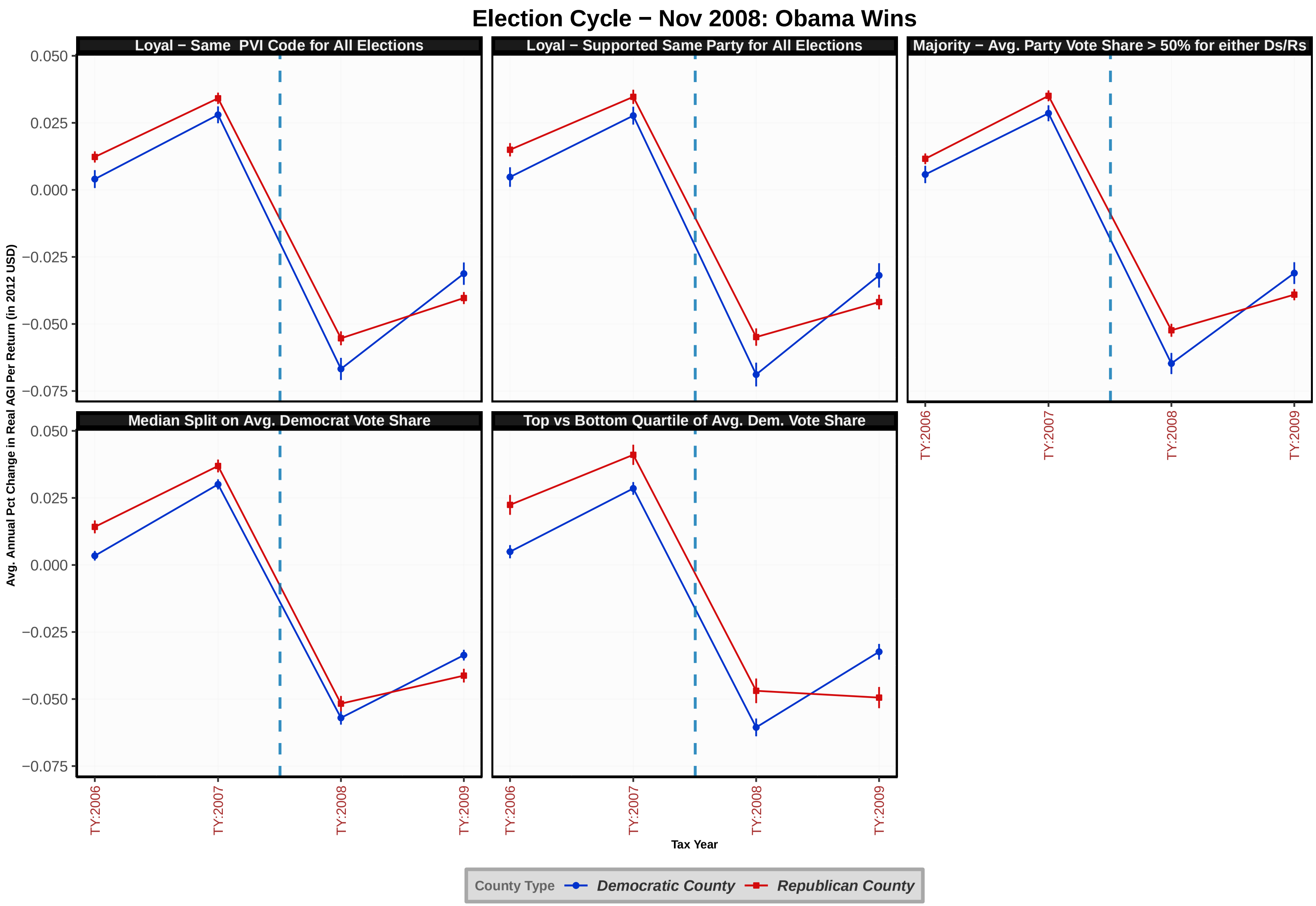
County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.16: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 2000



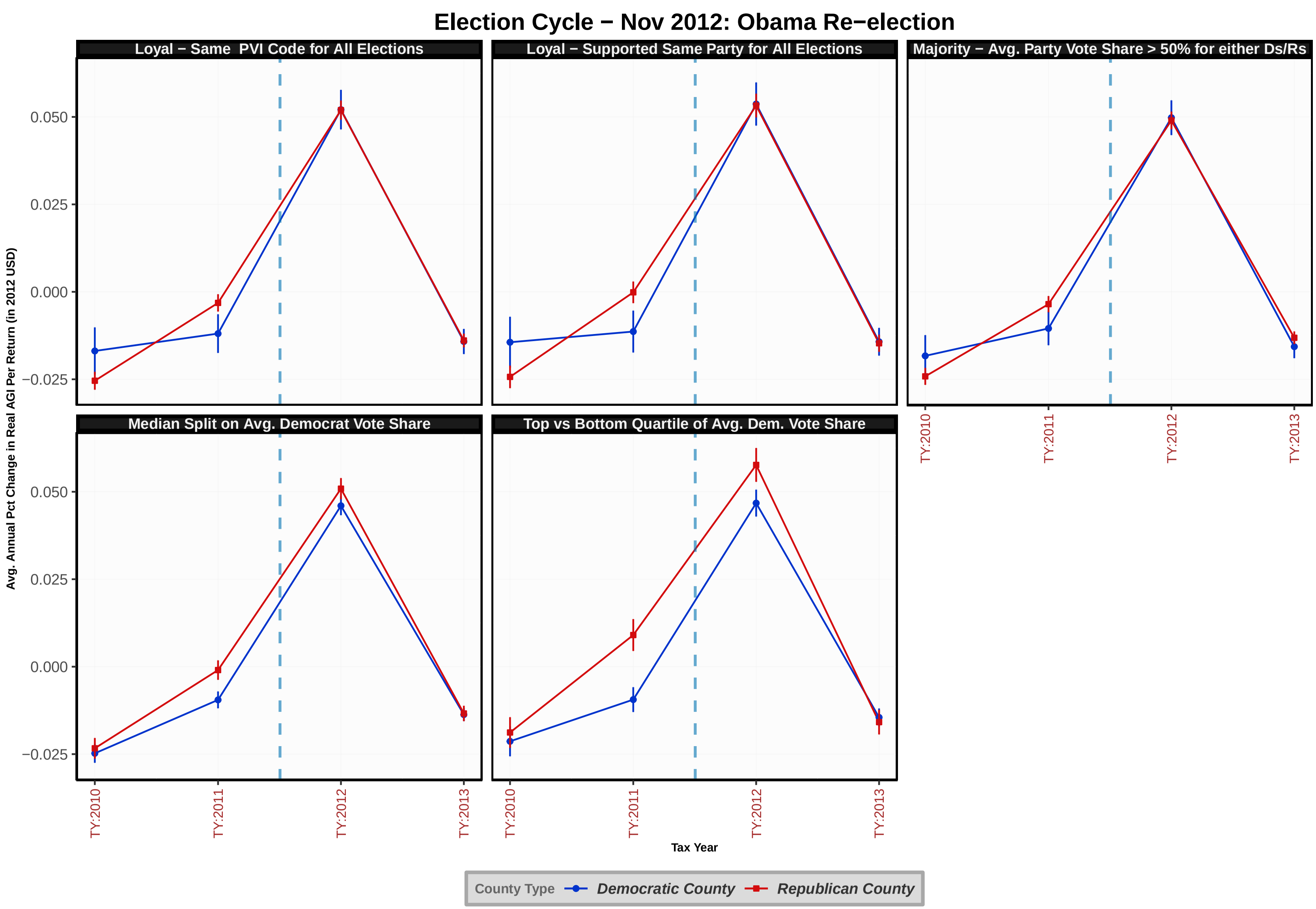
County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.17: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 2004



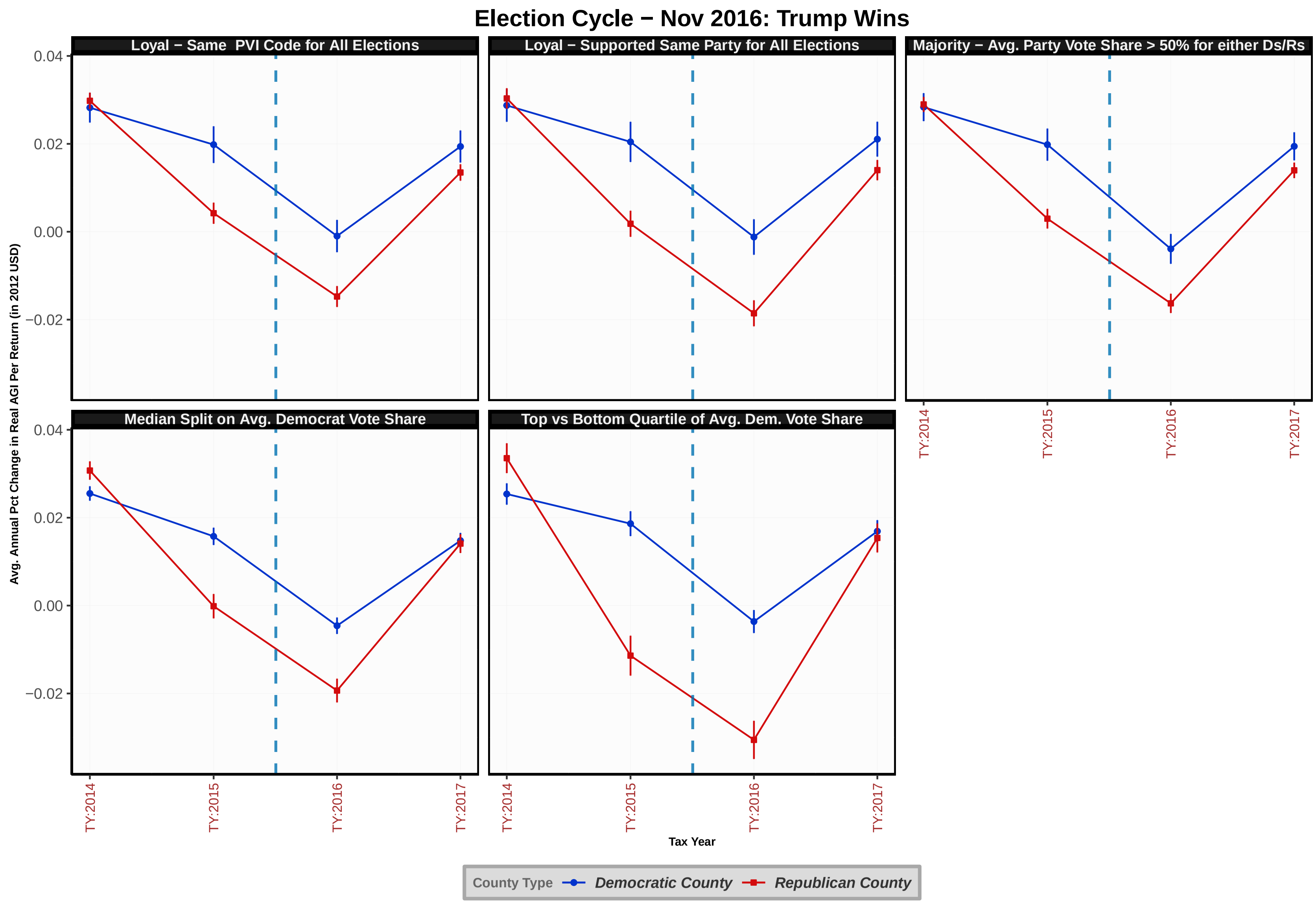
County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.18: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 2008



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.19: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 2012



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.20: Comparing Classification Approaches —Pct Chg in AGI per Return —Election Cycle: 2016

B.3.5 Comparing Classification Approaches: Aggregating by Election Winner (Democrat vs Republican)

Annual Percent Change in AGI per Return: All 7 Elections (1992-2016)

In order to examine the effect of classification approach, in Figure B.21, I present the annual percentage change in AGI per return for Democratic and Republican counties classified according to all four approaches —segregated by (a) elections where the Democrats won the White House and (b) where the Republican won the White House.

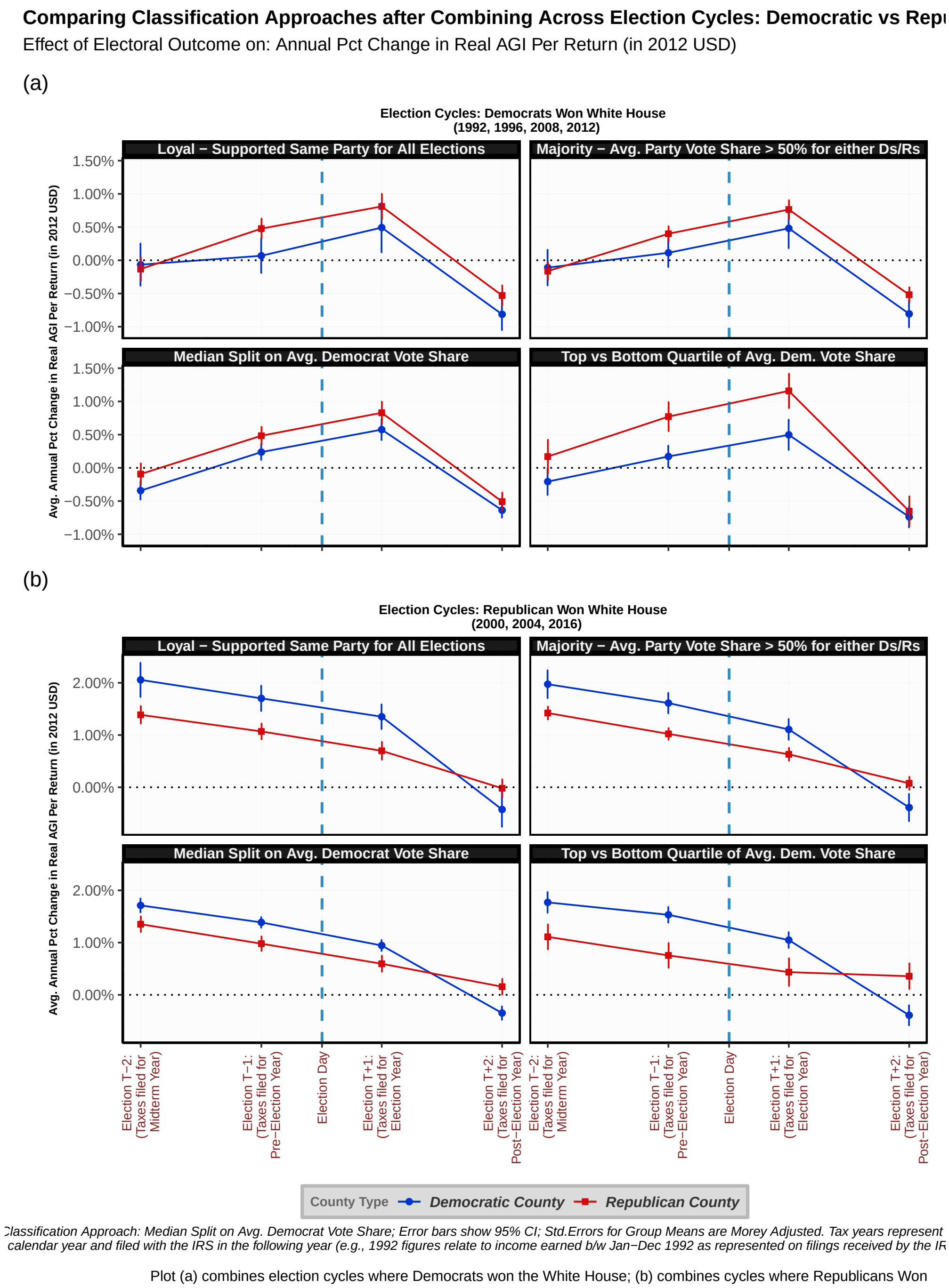


Figure B.21: Annual Percentage Change in AGI per Return (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans —Shown for All Classification Methods - All Election Data Included (1992-2016)

Annual Percent Change in AGI per Return: Excluding 1992 and 1996

The 1992 and 1996 election cycles were unusual in two respects: (a) due to substantial 3rd party success, in many counties the winner did not actually secure a majority of support and may have had more people against the candidate than for him, leading to a potential misclassification; (b) we saw support for the alternative hypotheses, suggesting that there may be some reason to worry about the problem of misclassification. It is not my intention to pursue this point further. Rather than pursue this further, in Figure B.22, I simply present Figure B.21 after excluding data from the 1992 and 1996 election cycles.

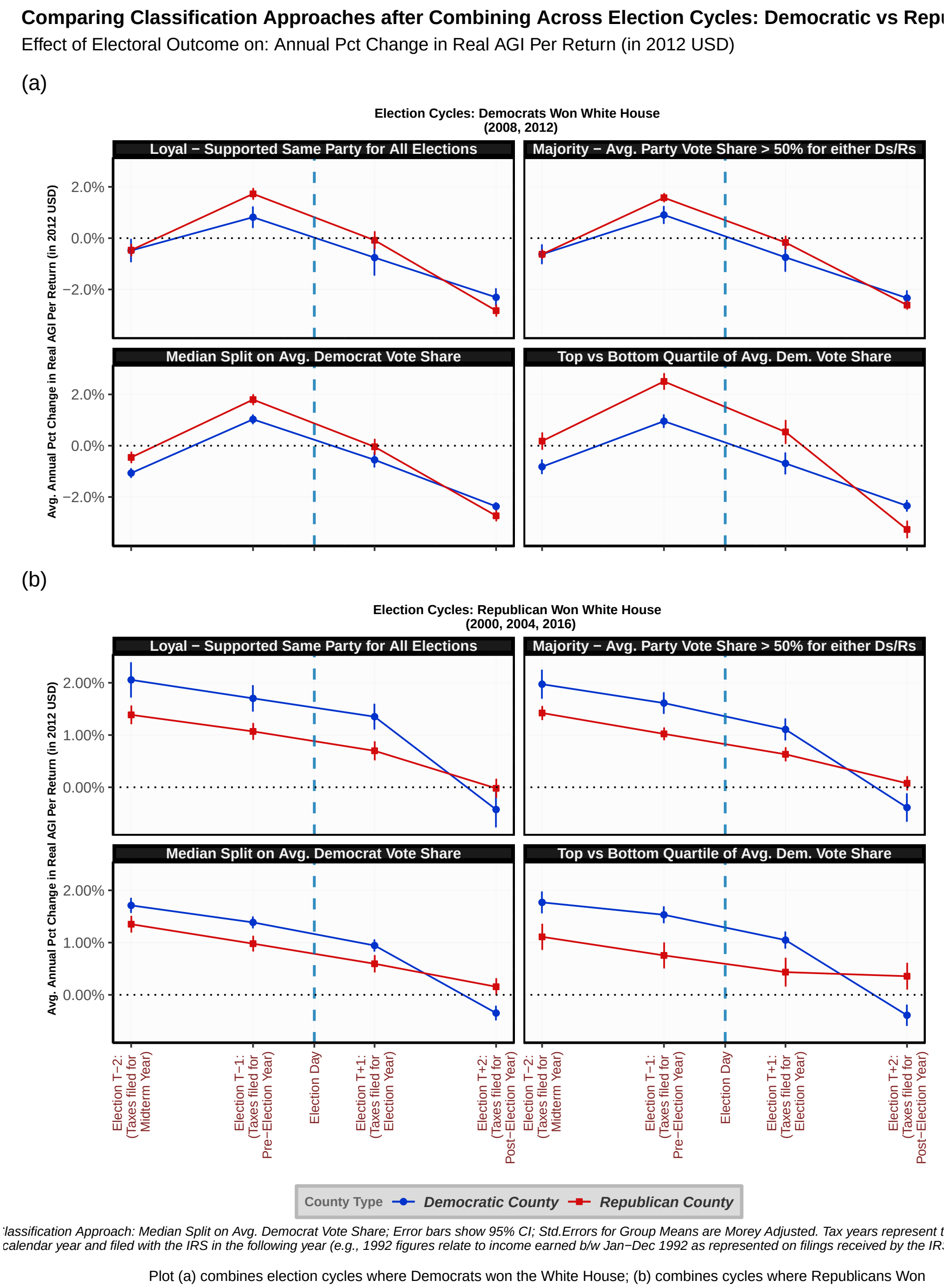


Figure B.22: Annual Percentage Change in AGI per Return (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans —Shown for All Classification Methods —After Removing 1992 and 1996 Election Cycles

B.3.6 Comparing Effects Separated by Election and Re-Election

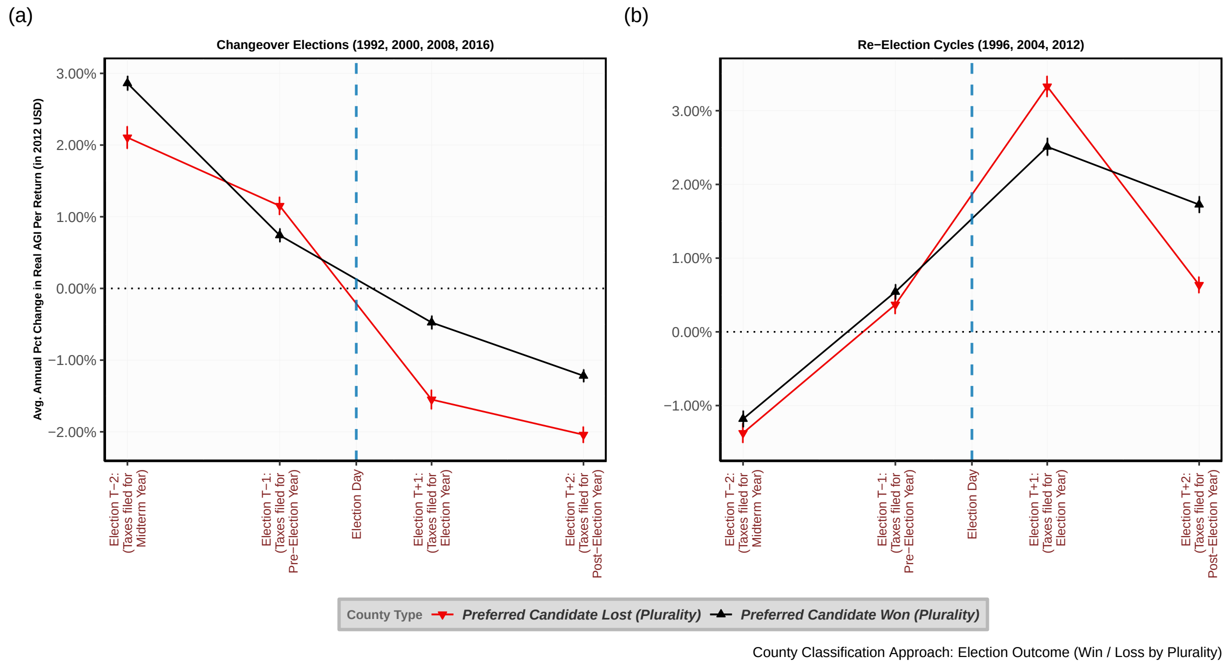


Figure B.23: Annual Percentage Change in AGI per Return for Winning Counties and Losing Counties —Separated by Elections with Change in Party vs Re-Election of Incumbent Party

B.3.7 Panel by Win or Loss - Comparing Classification Approaches

In Figure B.24, I present Annual Percentage Change in AGI per Return for counties classified as winners and losers under the plurality standard and under the absolute majority standard. In the first set of Figures B.24 (a1-a2), I present the entire sample of data. In Figures B.24 (b1-b2), I present the restricted dataset which excludes the 1992 and 1996 election cycles, which were unusual in that they showed support for the alternative hypotheses.

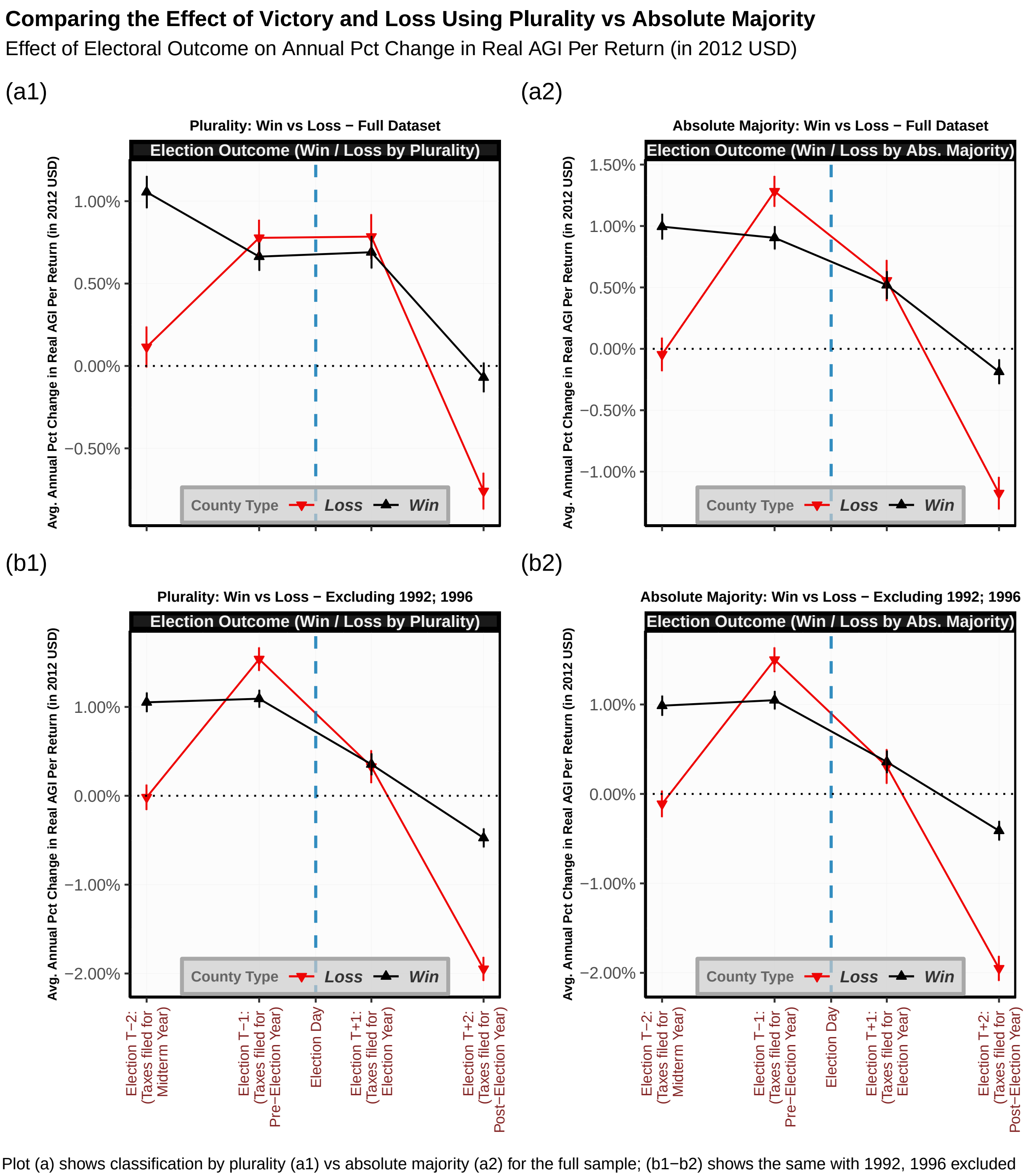


Figure B.24: Annual Percentage Change in AGI per Return (Winsorized) for Counties that Supported the Winning Candidate vs those that Supported the Losing Candidate —Comparing Classification Standards for Winners and Losers

B.4 Supplementary Graphical Analyses: AGI over PI

B.4.1 Distribution of Winsorized Variables

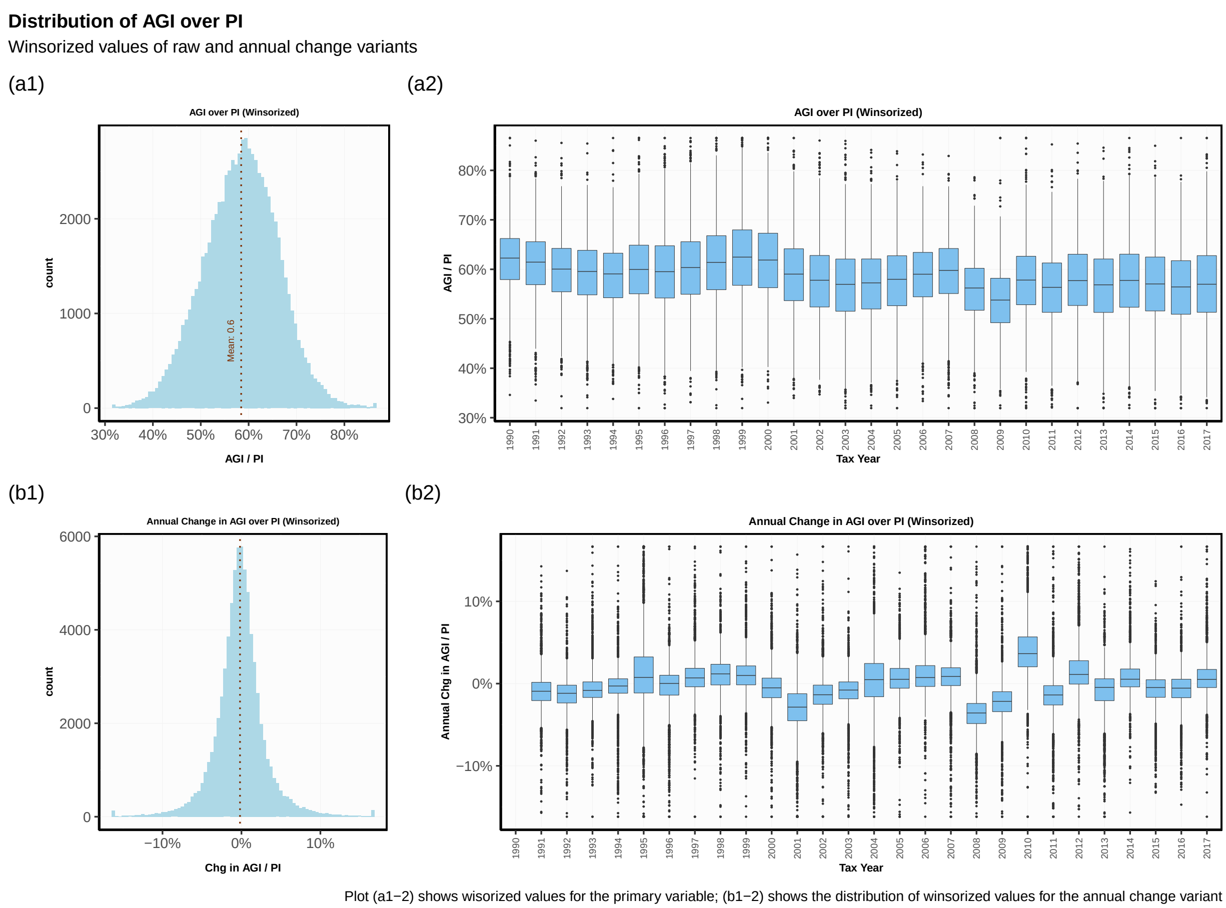


Figure B.25: Distribution of AGI over PI After Winsorizing Procedure for Addressing Outliers —Raw and Annual Change Variant

B.4.2 Distribution Across Classification Approaches

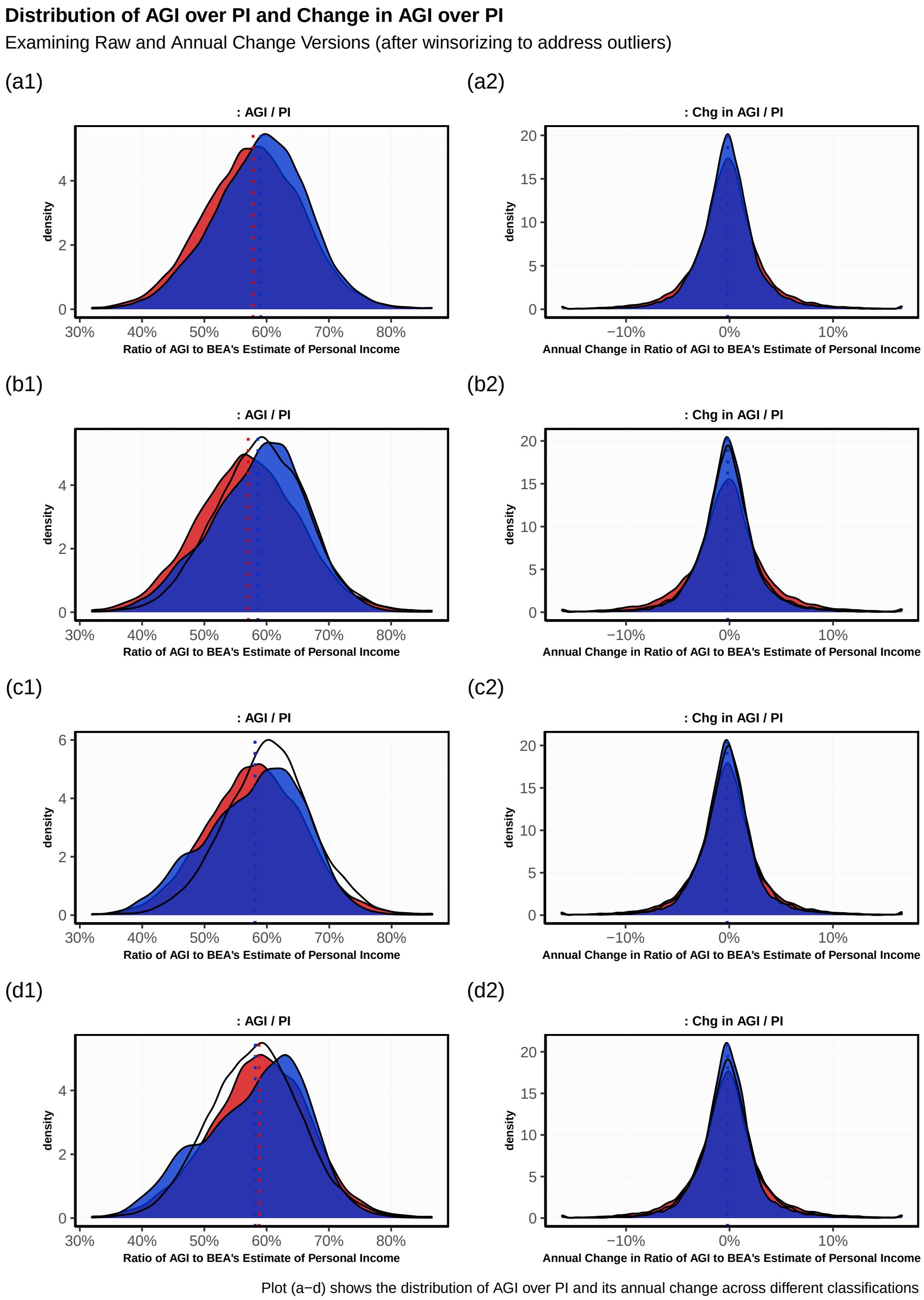


Figure B.26: Distribution of AGI / PI for all U.S. Counties Classified According to Partisanship —Comparing All Four Classification Approaches

B.4.3 Full Time Series: Difference in AGI over PI with Time Fixed Effects Removed

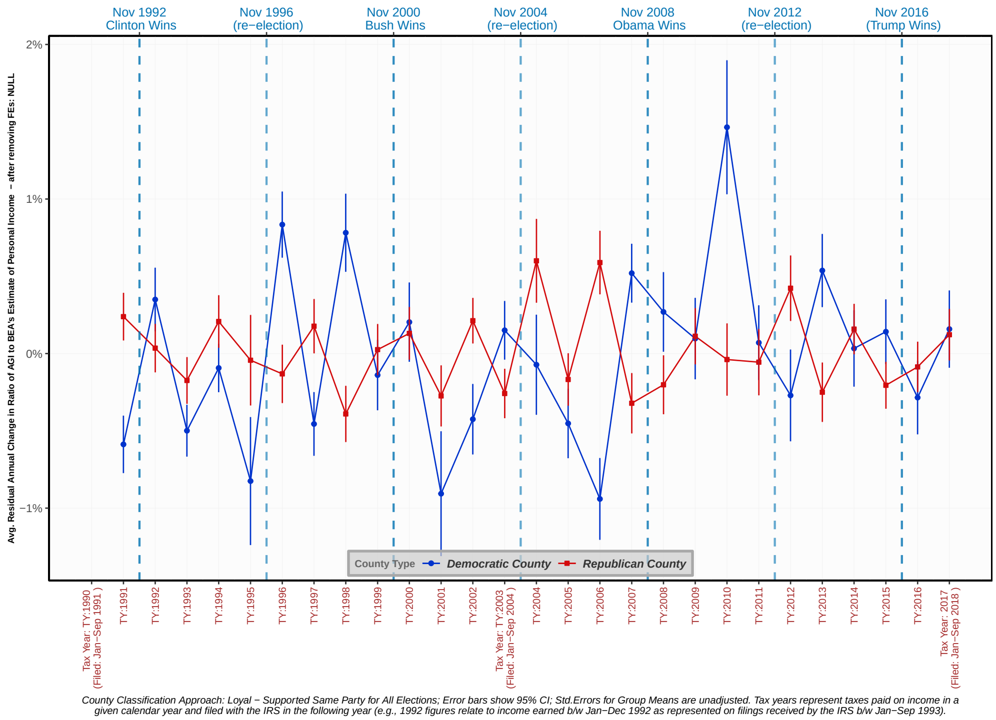


Figure B.27: AGI over PI (Winsorized) for Democrat and Republican Counties over Time

B.4.4 Panel by Election: Difference in AGI over PI

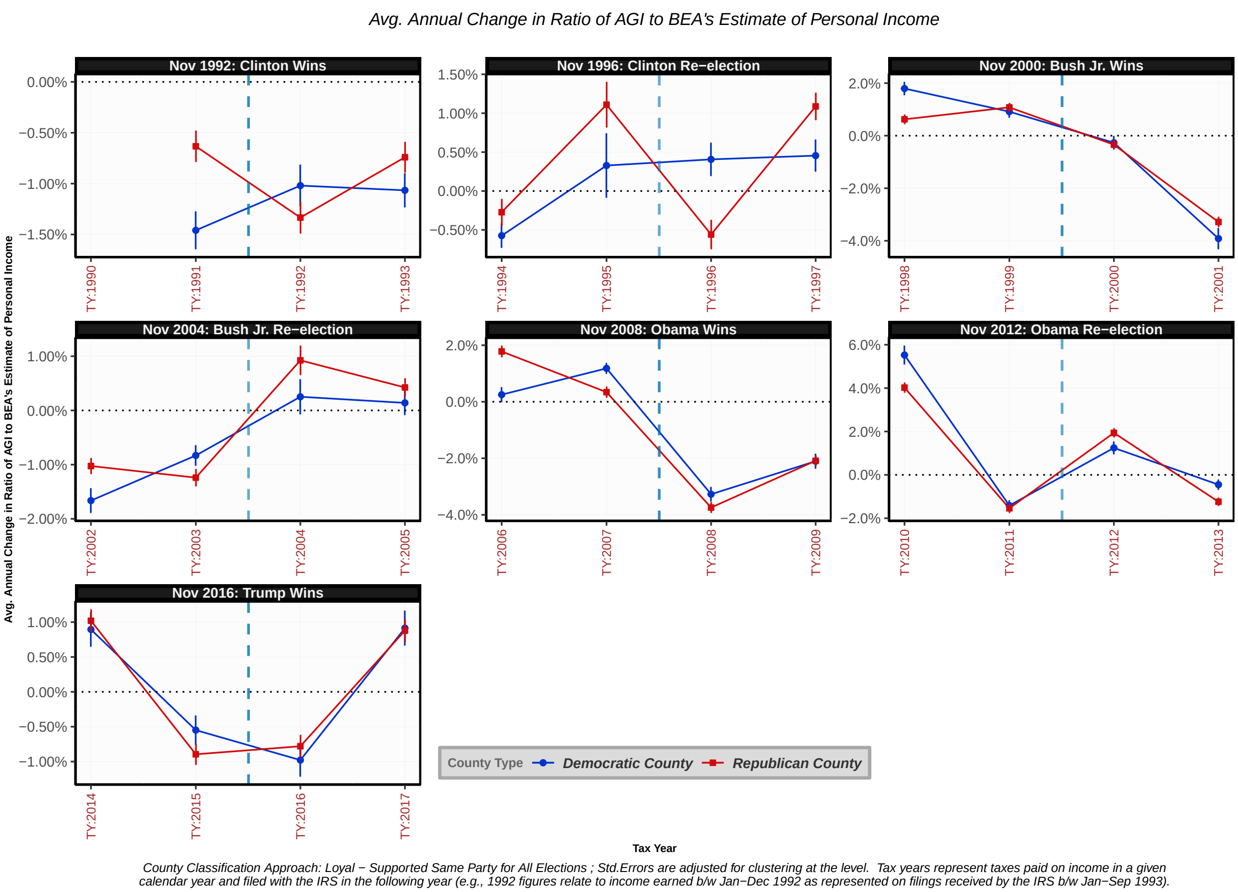
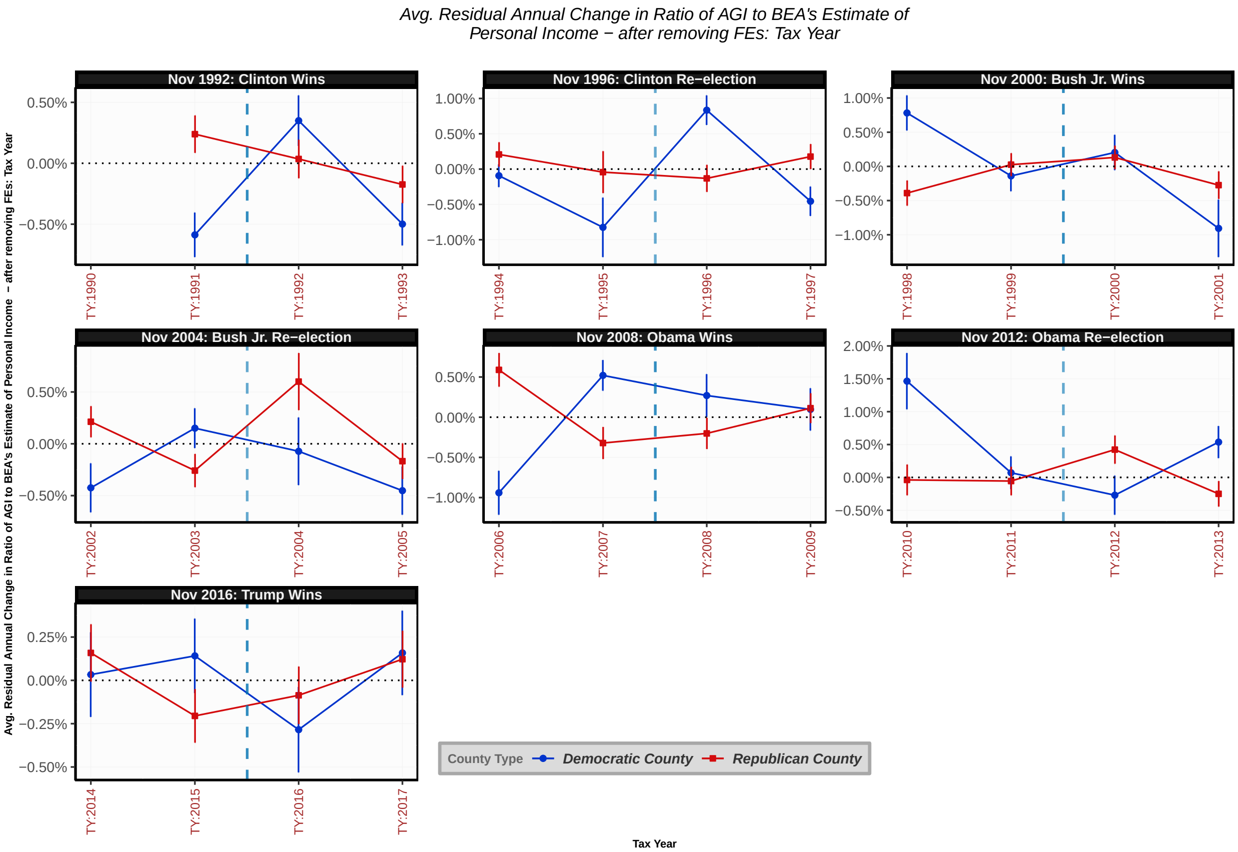


Figure B.28: Annual Change in AGI over PI for Democrat and Republican Counties Separated by Elections Cycle

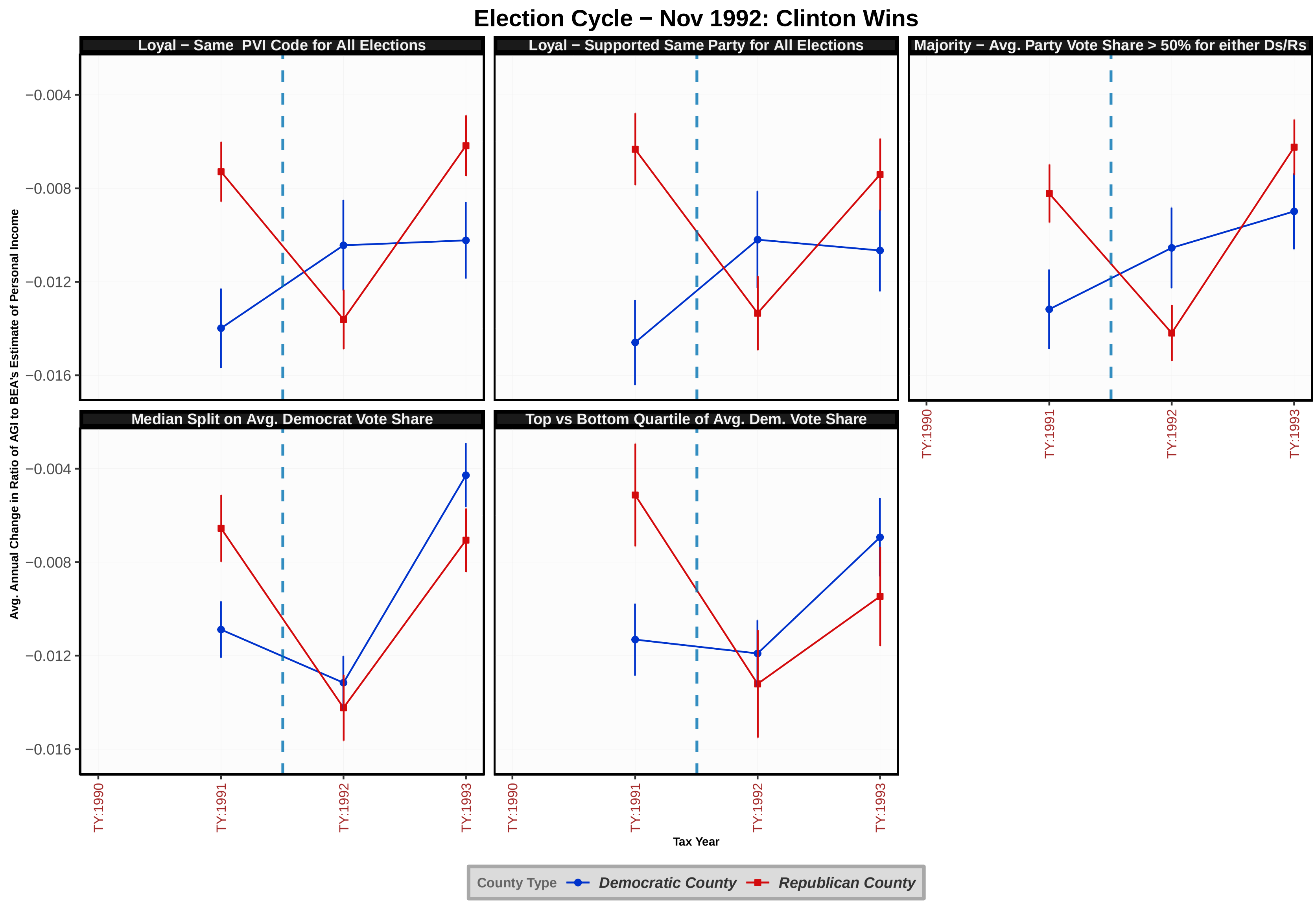


County Classification Approach: Loyal − Supported Same Party for All Elections; Error bars show 95% CI; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.29: Annual Change in AGI over PI for Democrat and Republican Counties Separated by Elections Cycle - Year Fixed Effects Removed

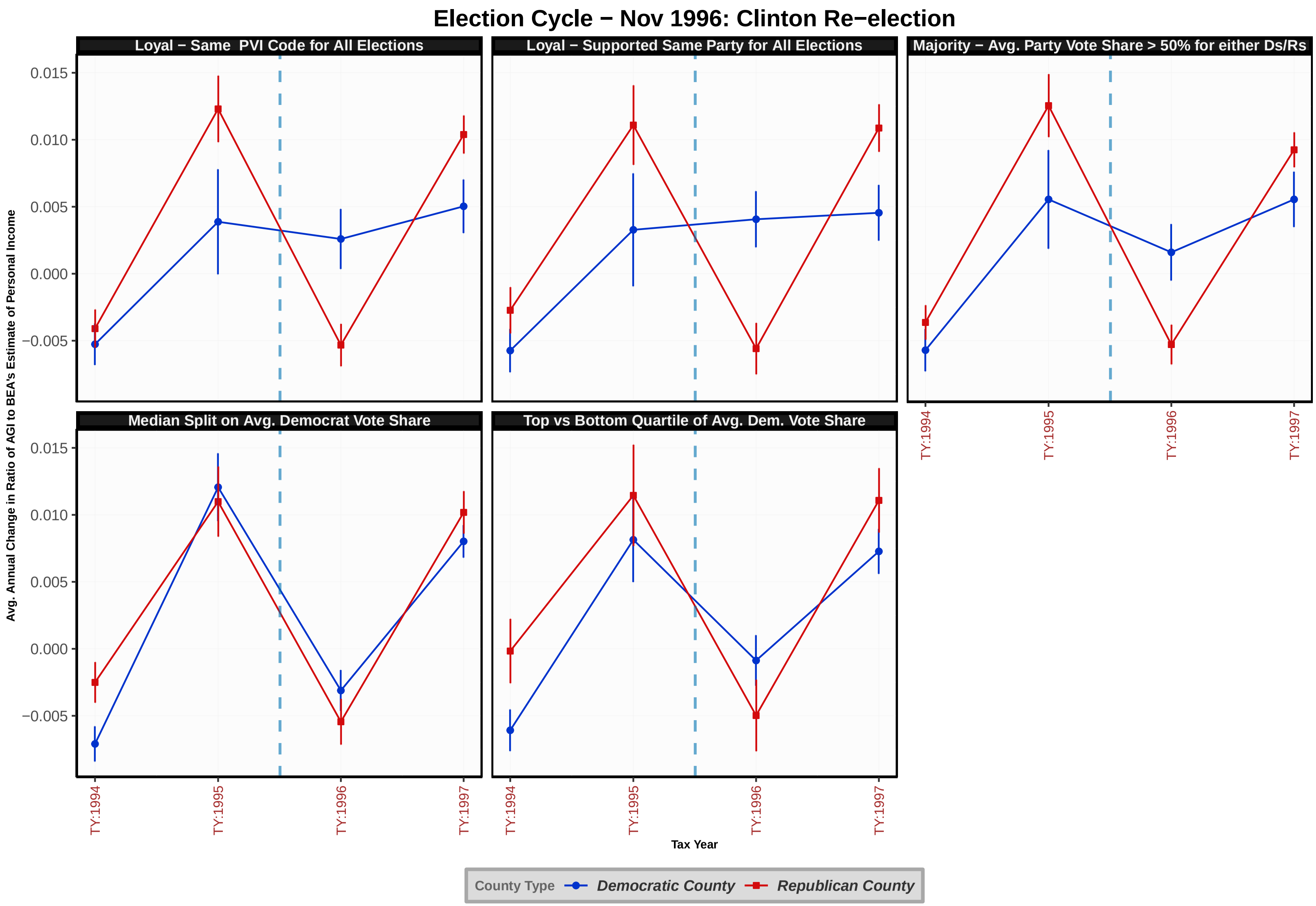
B.4.5 Difference in AGI over PI - Comparing Classification Approaches for each Election Cycle

In the current section, I present each election graph according to all classification approaches side-by-side for comparison. This allows the reader to examine the effect of classification for the trends seen here for all elections individually. The collected graphs for all election are presented in section Panel by Election: Difference in AGI over PI.



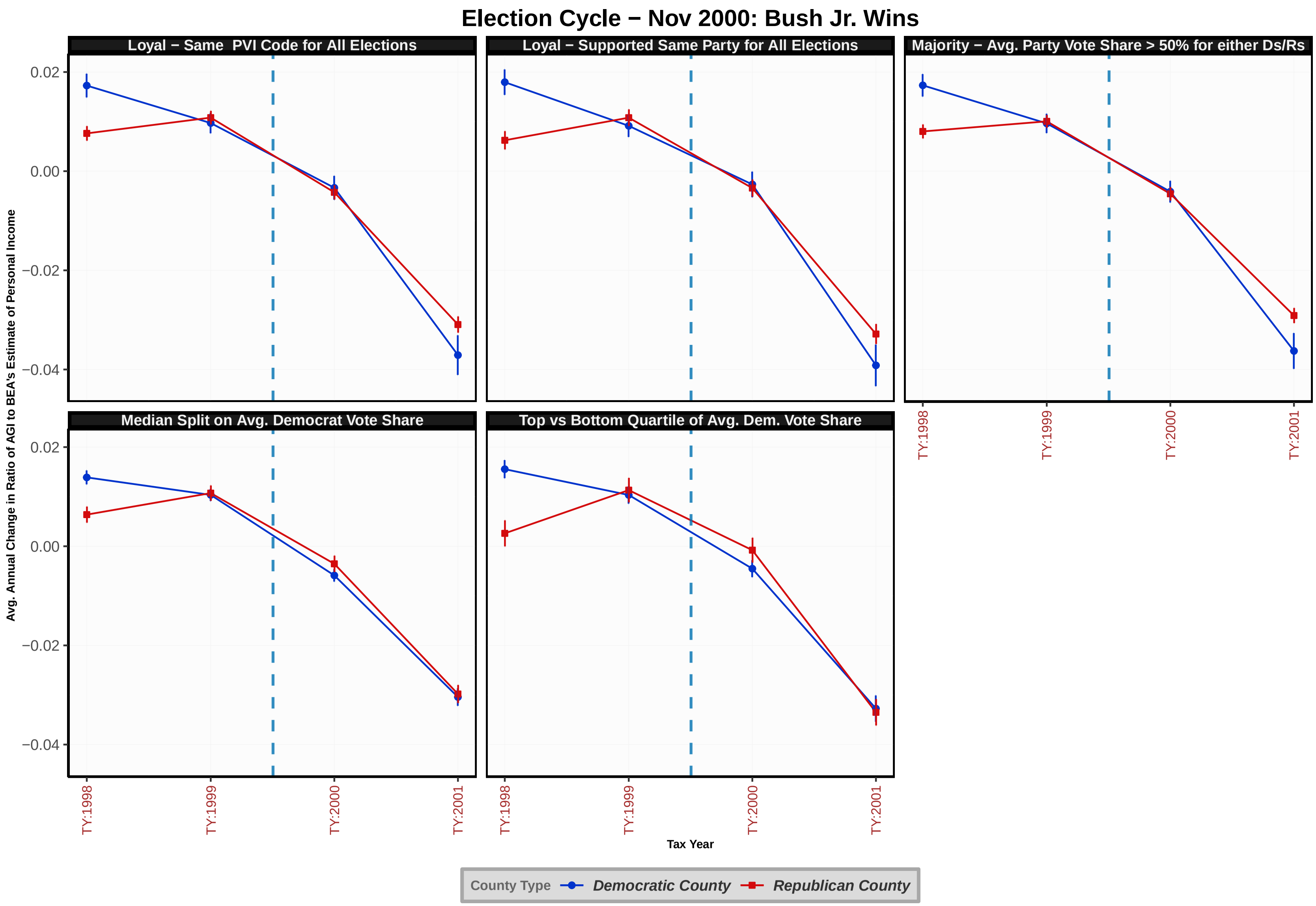
County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.30: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 1992



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.31: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 1996



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.32: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 2000

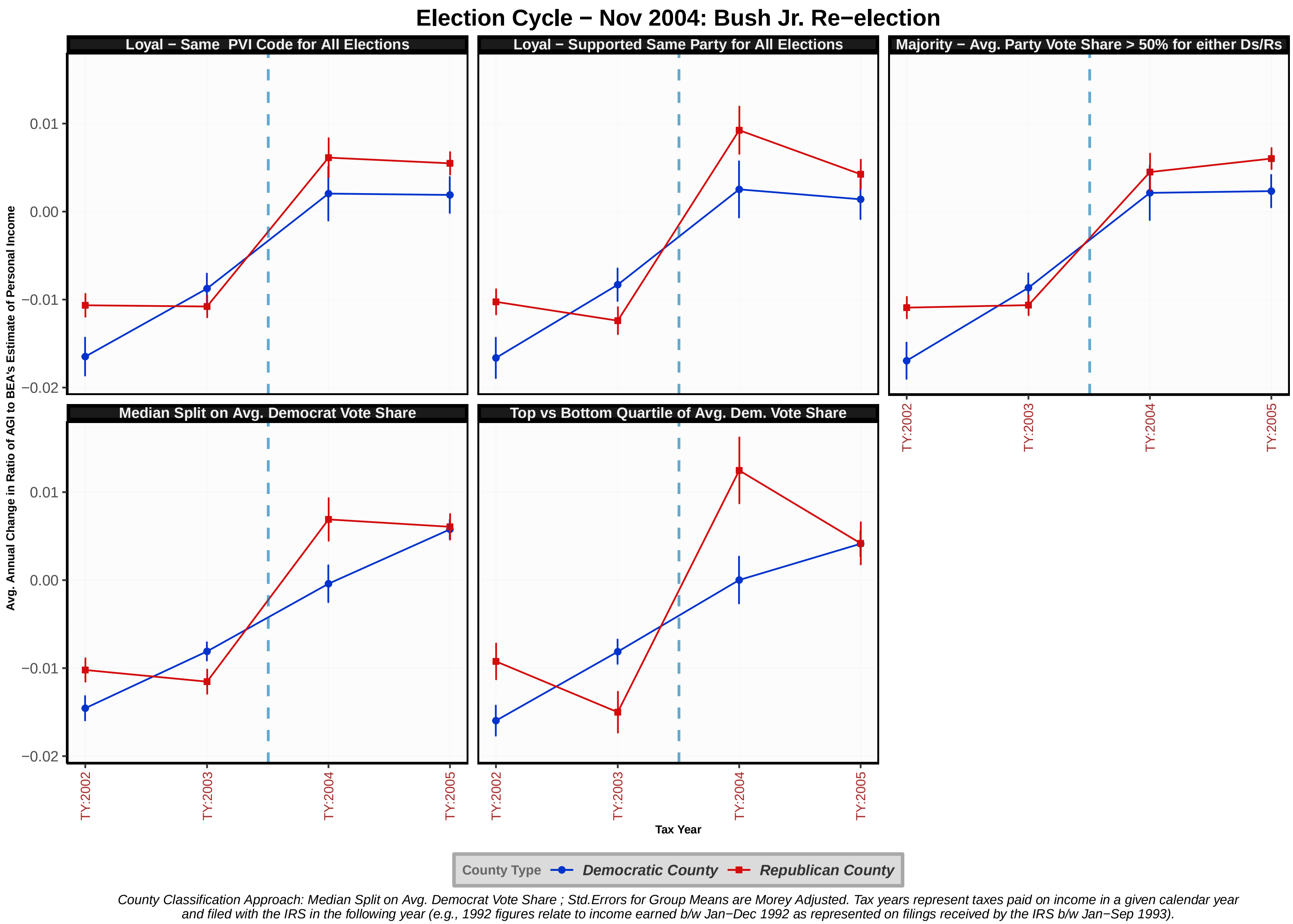
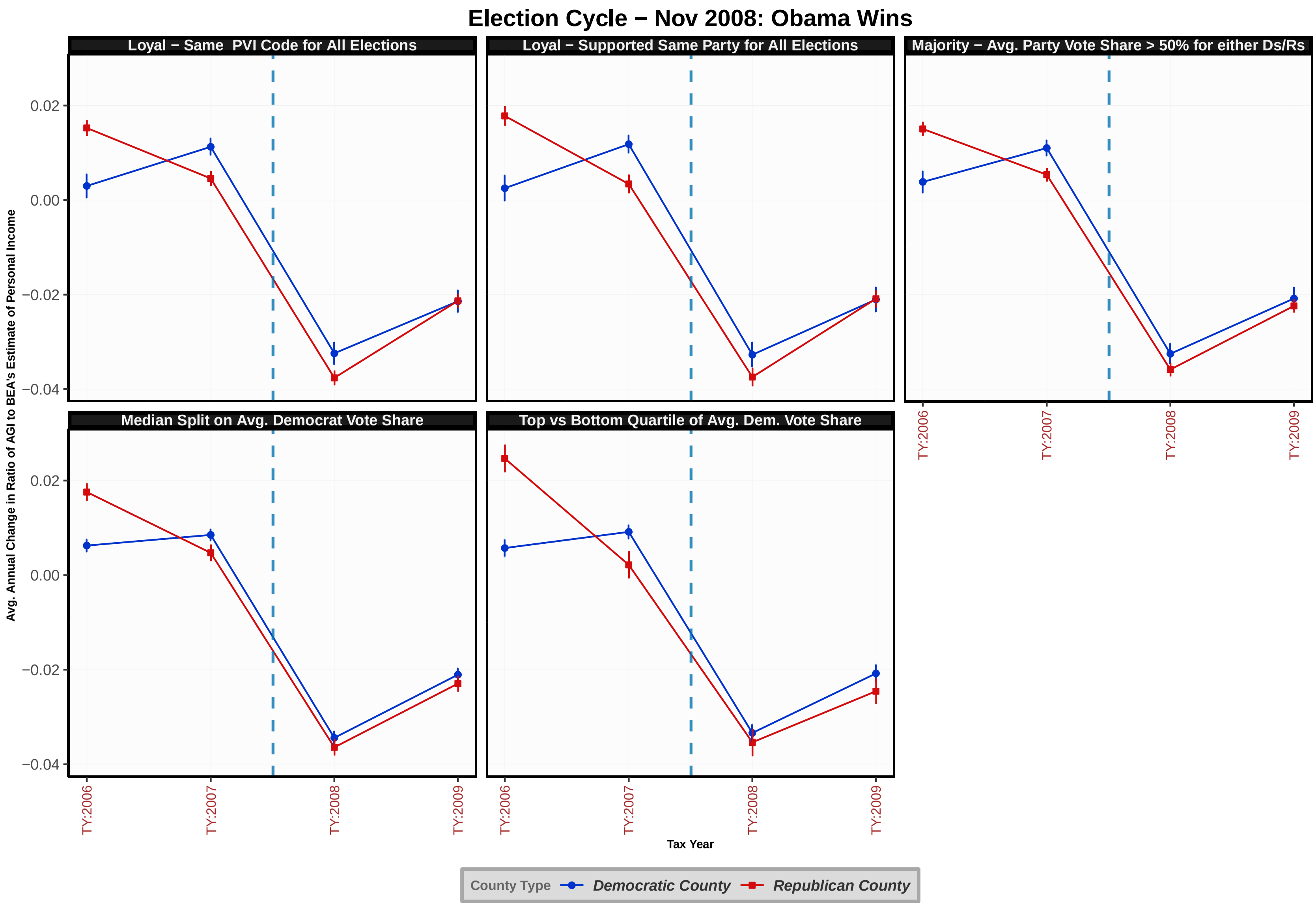
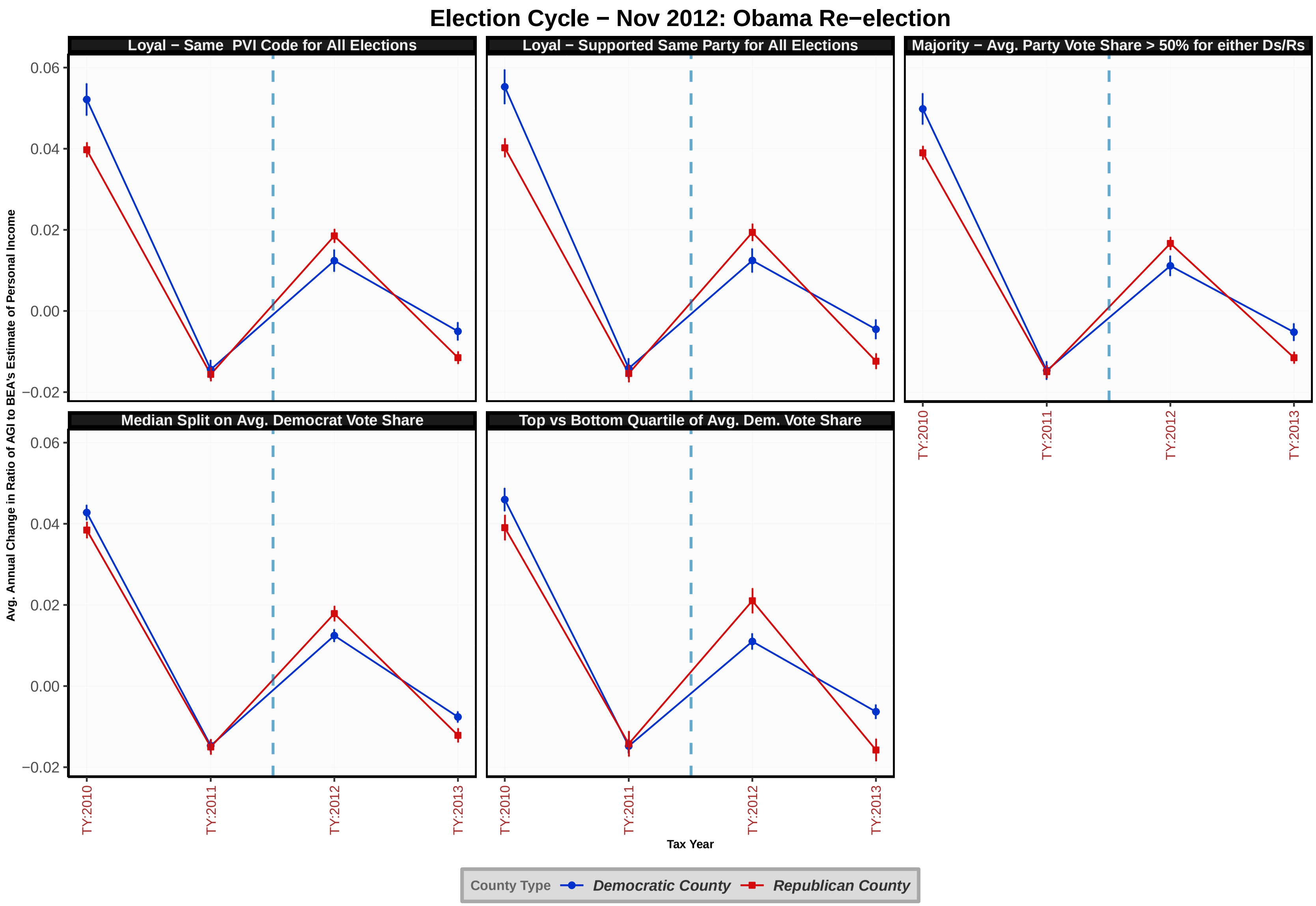


Figure B.33: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 2004



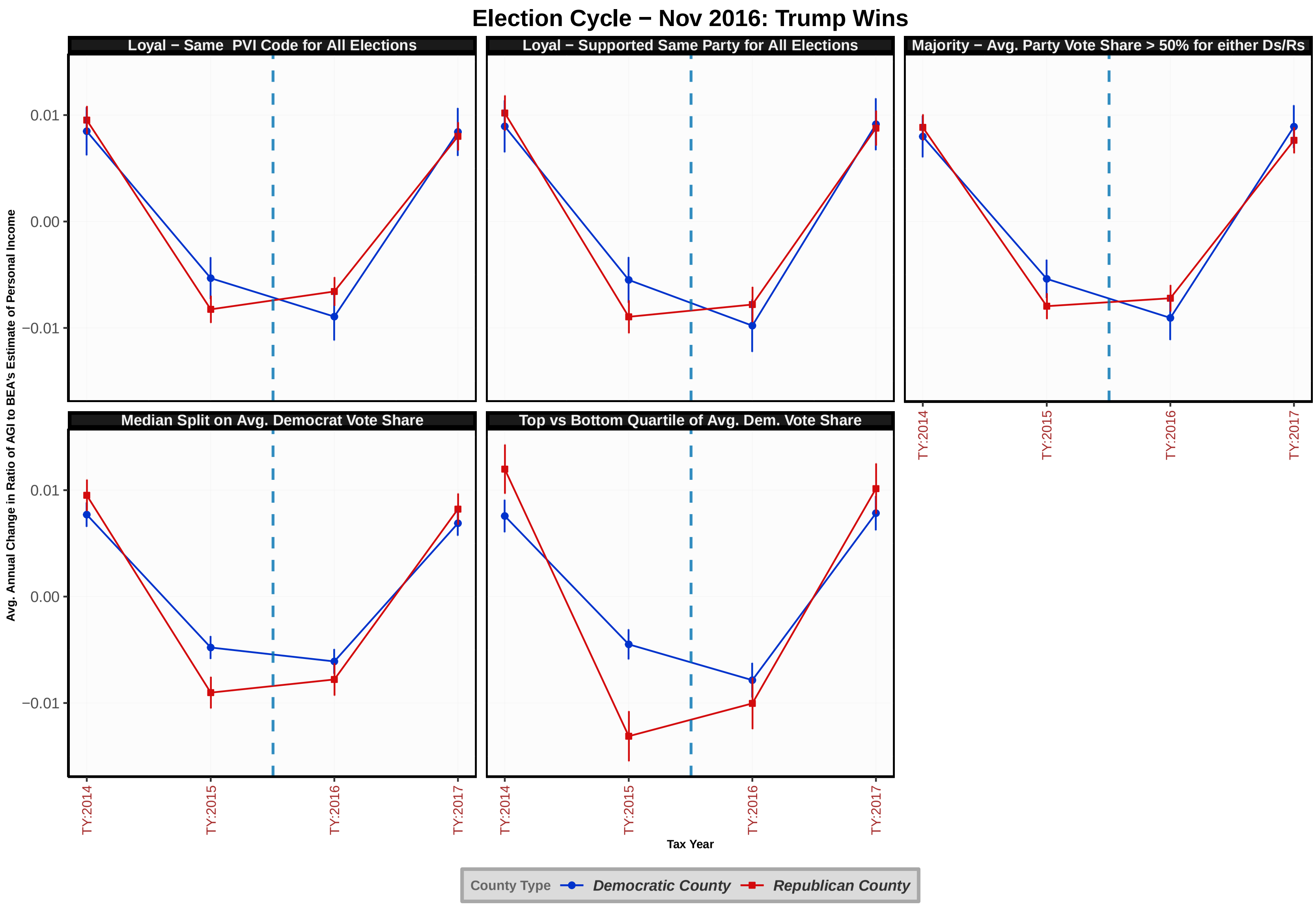
County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.34: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 2008



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.35: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 2012



County Classification Approach: Median Split on Avg. Democrat Vote Share ; Std.Errors for Group Means are Morey Adjusted. Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year (e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993).

Figure B.36: Comparing Classification Approaches —Change in AGI over PI —Election Cycle: 2016

B.4.6 Panel by Dem. vs Rep. Victory - Comparing Classification Approaches

In order to examine the effect of classification approach, in Figure B.37, I present the annual change in AGI over PI for Democratic and Republican counties classified according to all four approaches —segregated by (a) elections where the Democrats won the White House and (b) where the Republican won the White House.

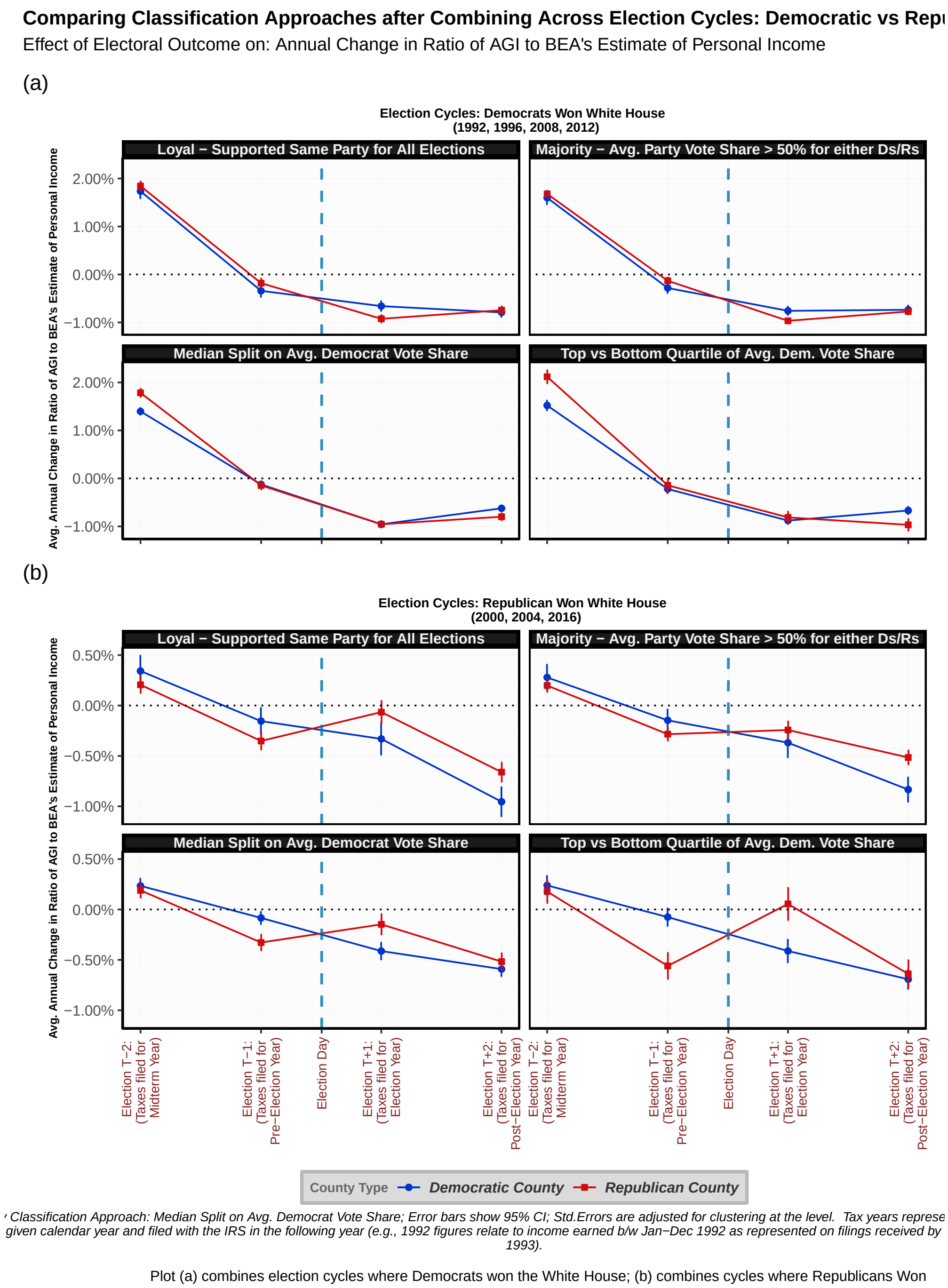


Figure B.37: Annual Change in AGI over PI (Winsorized) for Democrat and Republican Counties Separated by Elections Won by Democrats and those Won by Republicans — Shown for All Classification Methods

B.4.7 Panel by Win or Loss - Comparing Classification Approaches

In Figure B.38, I present Annual Change in AGI over PI for counties classified as winners and losers under the plurality standard and under the absolute majority standard. In the first set of Figures B.38 (a1-a2), I present the entire sample of data. In Figures B.38 (b1-b2), I present the restricted dataset which excludes the 1992 and 1996 election cycles, which were previously seen to be unusual for the AGI / return variable. With the current variable, as can be seen, these election cycles no longer appear to significantly influence the trends seen here, which remain clearly in favor of the legitimacy hypothesis.

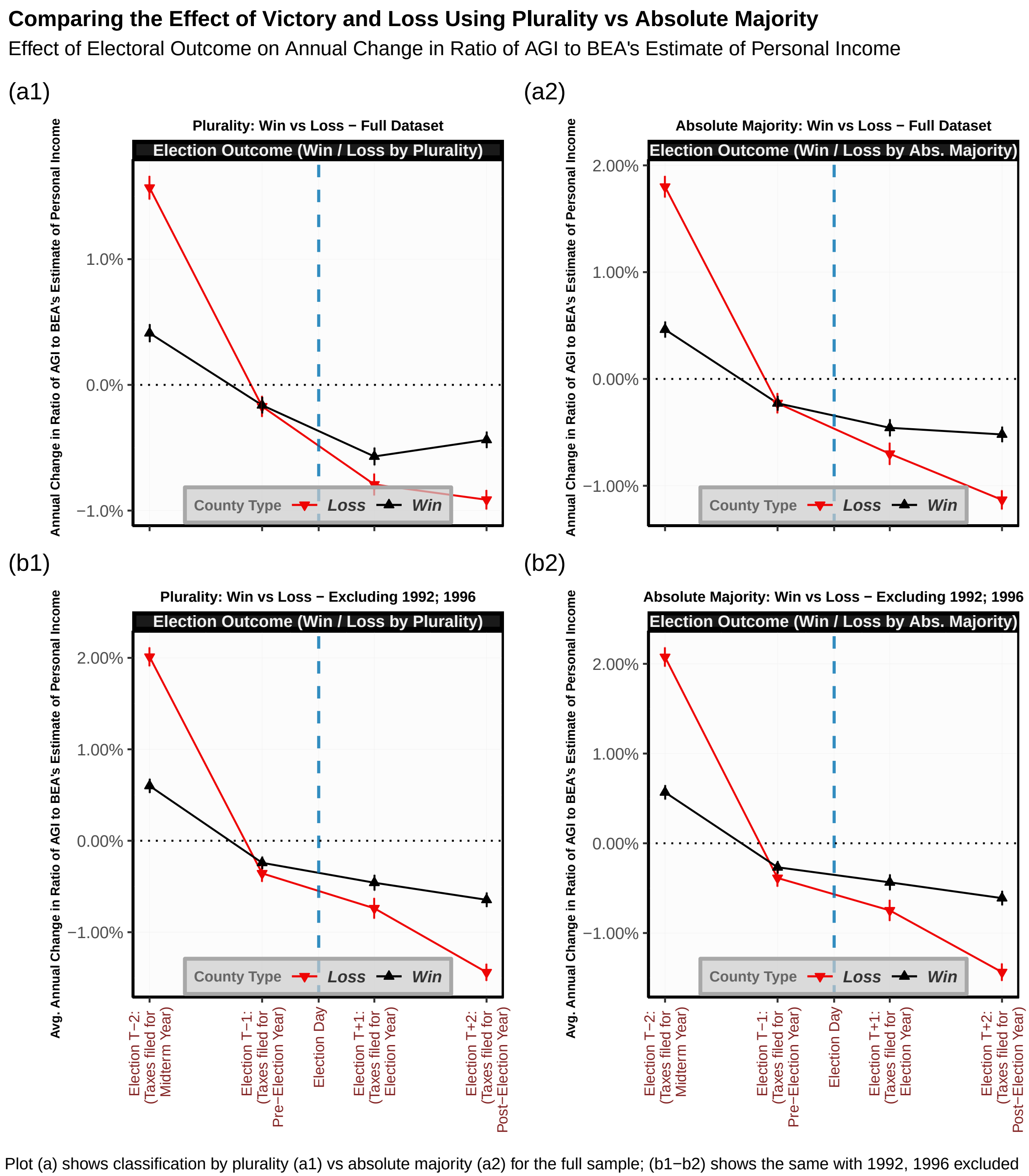


Figure B.38: Annual Change in AGI over PI (Winsorized) for Counties that Supported the Winning Candidate vs those that Supported the Losing Candidate —Comparing Classification Standards for Winners and Losers

B.5 Supplementary Statistical Analyses: AGI over PI

B.5.1 Supplementary Tables for Binary Measures of Election Outcomes

Effect of Election Outcome on AGI over PI in the Election Year - By Win Loss with Lag Terms

Table B.4: Difference between Win and Loss: Annual Change in AGI / PI

	Annual Change in AGI / PI					
	Full Sample			Loyal Partisans Only		
	(1)	(2)	(3)	(4)	(5)	(6)
Preferred Candidate Lost (Plurality)	-0.112^{*}	-0.078	-0.090^{+}	-0.265^{***}	-0.224^{***}	-0.232^{**}
	(0.050)	(0.050)	(0.055)	(0.064)	(0.066)	(0.072)
Lag 1 Yr: An. Chg. AGI/PI		-0.231^{***}	-0.226^{***}		-0.234^{***}	-0.236^{***}
		(0.014)	(0.015)		(0.018)	(0.021)
Lag 2 Yr: An. Chg. AGI/PI			-0.055^{***}			-0.073^{***}
			(0.014)			(0.018)
Observations	19,271	19,271	16,518	10,073	10,073	8,634
R^2	0.298	0.336	0.350	0.319	0.356	0.373
Adjusted R^2	0.181	0.225	0.220	0.205	0.248	0.247
Residual Std. Error	3.086 (df = 16511)	3.003 (df = 16510)	3.125 (df = 13757)	3.171 (df = 8627)	3.084 (df = 8626)	3.198 (df = 7187)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model includes fixed effects for County and Year;
SEs: heteroskedasticity-robust, clustered at County-Level.

B.5.2 Supplementary Tables for Continuous Measures of Election Outcome

Signed Margin of Victory Absolute Majority - By Win Loss

Table B.5: Signed Margin of Victory (Abs. Majority) and Change in AGI / PI

	Annual Change in AGI / PI		
	Full Sample	Restricted Sample	Loyal Partisans
	(1)	(2)	(3)
Signed Margin of Victory	0.300**	0.369***	0.350***
	(0.096)	(0.086)	(0.105)
Observations	15,267	14,631	8,565
R^2	0.332	0.386	0.346
Adjusted R^2	0.185	0.244	0.213
Residual Std. Error	0.032 (df = 12507)	0.026 (df = 11871)	0.032 (df = 7119)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model includes fixed effects for State and Year;
SEs: heteroskedasticity-robust, clustered at County-Level;
Sample: central 98% of the data for both IV and DV;
Estimated coefficients shown in percentage points.

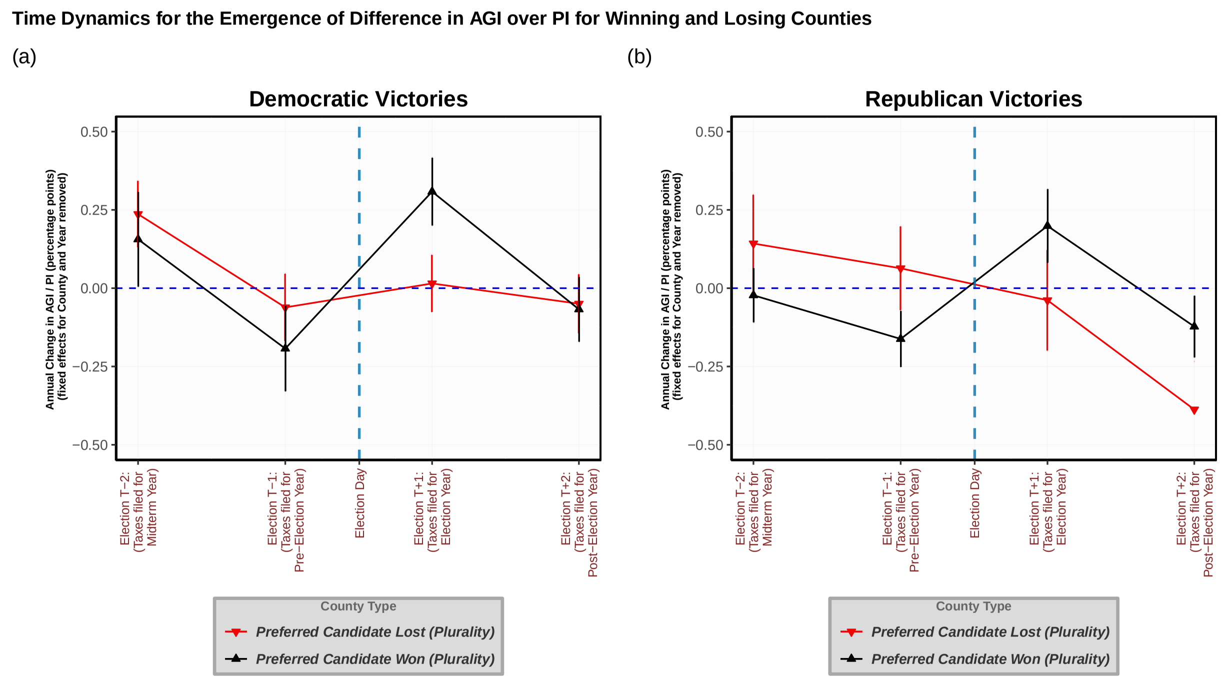
Margin of Victory - By Win Loss - Loyal Partisans Only

Table B.6: Change in AGI / PI: Election Outcome x Margin of Victory

	Annual Change in AGI over PI			
	Loyal Partisans Only			
	(1)	(2)	(3)	(4)
I(Lost) ($\widehat{\beta_1}$)	−0.040	0.187	0.455***	0.459***
	(0.131)	(0.128)	(0.113)	(0.114)
MoV ($\widehat{\beta_2}$)	1.189**	1.228**	2.273***	2.202***
	(0.392)	(0.413)	(0.341)	(0.345)
AGI / PI		0.279***	0.168***	0.213***
		(0.011)	(0.010)	(0.011)
Pct.Chg AGI/Ret			0.362***	0.327***
			(0.011)	(0.011)
Lag Chg. AGI/PI				−0.216***
				(0.015)
I(Lost) x MoV ($\widehat{\beta_3}$)	−0.431	−1.496***	−2.793***	−2.643***
	(0.397)	(0.377)	(0.346)	(0.342)
Observations	10,073	10,073	10,073	10,073
R^2	0.320	0.409	0.583	0.611
Adjusted R^2	0.205	0.310	0.512	0.546
Residual Std. Error	3.170 (df = 8625)	2.955 (df = 8624)	2.483 (df = 8623)	2.397 (df = 8622)

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Model includes fixed effects for County and Year;
SEs: heteroskedasticity-robust, clustered at County-Level;
Models shown least-to-most restrictive vis-a-vis controls;
Estimated coefficients shown in percentage points.

B.5.3 Comparing Time Dynamics: Democrat versus Republican Victories



Plot (a) shows the difference in Annual Change in AGI over PI for election cycles where Democrats won and Plot (b) for cycles where Republican won;
The red line shows data for counties whose candidate lost; the black line for counties whose candidate won the election.
In both cases, the values shown have county and year fixed effects removed; Data restricted to loyal partisan counties.
County Classification: Election Outcome (Win / Loss by Plurality); Error bars show 95% CI; SEs clustered at County level.
Tax years represent taxes paid on income in a given calendar year and filed with the IRS in the following year
(e.g., 1992 figures relate to income earned b/w Jan−Dec 1992 as represented on filings received by the IRS b/w Jan−Sep 1993)

Figure B.39: Time Dynamics of the Election Effect —Comparing Annual Change in AGI / PI for Election Cycles where Democrats Won and those where Republicans Won —Loyal Partisan Counties Full Sample of Counties —Victory and Loss Classified Using Plurality Threshold

B.5.4 Time Dynamics - Supplementary Table Showing Election Effect across Event Window

Table B.7: Difference in AGI over PI between Winning and Losing Counties (Time Dynamics)

Dependent Variable:	AGI / PI		
	Basic	Lag Control	All Controls
Model:	(1)	(2)	(3)
Variables			
Won $\times T$ - 2 $\times$Window = Pre-Election	-0.007	-0.012	-0.022
	(0.098)	(0.099)	(0.085)
Won $\times T$ + 1 $\times$Window = Post-Election	0.290***	0.301***	0.282***
	(0.083)	(0.082)	(0.071)
Won $\times T$ + 2 $\times$Window = Post-Election	0.342***	0.315***	0.289***
	(0.101)	(0.095)	(0.080)
Lag AGI/PI		0.136***	0.330***
		(0.015)	(0.015)
Pct.Chg AGI/Ret			0.298***
			(0.009)
Fixed-effects			
County	Yes	Yes	Yes
County $\times$Election Cycle	Yes	Yes	Yes
State $\times$Year	Yes	Yes	Yes
Number of Observations			
N_{Total}	38,853	38,853	38,853
N_{County}	1,439	1,439	1,439
$N_{County \times ElectionCycle}$	10,073	10,073	10,073
$N_{State \times Year}$	1,323	1,323	1,323
Fit statistics			
Squared Correlation	0.936	0.937	0.952
Pseudo R^2	0.39128	0.394	0.43395
BIC	301,692	300,962	290,075

One-way (County) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

B.6 Statistical Software and Packages

For all computational portions of this project, including including all data manipulation, cleaning, analyses, graphing, and write-up for presentation, I used R (Version 3.6.3; R Core Team, 2020) and the R-packages *apastats* (Version 0.3; Chetverikov, 2019), *apaTables* (Version 2.0.5; Stanley, 2018), *assertr* (Version 2.7; Fischetti, 2020), *assertthat* (Version 0.2.1; Wickham, 2019a), *blscrapeR* (Version 3.2.0; Eberwein, 2019), *bookdown* (Version 0.20; Xie, 2020a), *broom* (Version 0.7.0; Robinson et al., 2020), *car* (Version 3.0.9; J. Fox et al., 2020a, 2020b), *carData* (Version 3.0.4; J. Fox et al., 2020b), *checkr* (Version 0.5.0; Thorley, 2019), *citr* (Version 0.3.2; Aust, 2019), *conflicted* (Version 1.0.4; Wickham, 2019b), *corrplot* (Version 0.84; Wei & Simko, 2017), *cowplot* (Version 1.0.0; Wilke, 2019), *data.table* (Version 1.13.0; Dowle & Srinivasan, 2020), *dplyr* (Version 1.0.1; Wickham, François, et al., 2020), *dygraphs* (Version 1.1.1.6; Vanderkam et al., 2018), *elections* (Version 1.0; van der Wal, 2018), *emmeans* (Version 1.4.8; Lenth, 2020), *estimatr* (Version 0.22.0; Blair et al., 2020), *fixest* (Version 0.6.0; Berge, 2020), *foreach* (Version 1.5.0; Microsoft & Weston, 2020), *formattable* (Version 0.2.0.1; Ren & Russell, 2016), *Formula* (Version 1.2.3; Zeileis & Croissant, 2018), *formula.tools* (Version 1.7.1; Brown, 2018), *ggdark* (Version 0.2.1; Grantham, 2019), *ggeffects* (Version 0.15.1; Lüdecke, 2020a), *ggplot2* (Version 3.3.2; Wickham, Chang, et al., 2020), *ggpubr* (Version 0.4.0; Kassambara, 2020), *ggrepel* (Version 0.8.2; Slowikowski, 2020), *ggthemes* (Version 4.2.0; Arnold, 2019), *gridExtra* (Version 2.3; Auguie, 2017), *Hmisc* (Version 4.4.1; Harrell Jr et al., 2020), *htmlTable* (Version 2.0.1; Gordon et al., 2020), *htmlwidgets* (Version 1.5.1; Vaidyanathan et al., 2019), *installr* (Version 0.22.0; Galili, 2019), *issuer* (Version 0.1.0; Doane, 2019), *knitr* (Version 1.29; Xie, 2020b), *labelled* (Version 2.5.0; Larmarange, 2020; Lüdecke, 2020b), *lattice* (Version 0.20.41; Sarkar, 2020), *lavaan* (Version 0.6.7; Rosseel et al., 2020), *leaflet* (Version 2.0.3; Cheng et al., 2019), *lfe* (Version 2.8.5.1; Gaure, 2020), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.51.6; Ripley, 2020), *Matrix* (Version 1.2.18; Bates & Maechler, 2019), *multcomp* (Version 1.4.13; Hothorn et al., 2020; Graves et al., 2019), *multcompView* (Version 0.1.8; Graves et al., 2019), *mvtnorm* (Version 1.1.1; Genz et al., 2020), *officer* (Version 0.3.12; Gohel, 2020), *papaja* (Version 0.1.0.9842; Aust & Barth, 2019), *patchwork* (Version 1.0.1; Pedersen, 2020), *plm* (Version 2.2.3; Croissant et al., 2020), *plyr* (Version 1.8.6; Wickham, François, et al., 2020; Wickham, 2020), *psych* (Version 2.0.7; Revelle, 2020), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version 1.3.1; Wickham et al., 2018), *readxl* (Version 1.3.1; Wickham & Bryan, 2019), *RefManageR* (Version 1.2.12; McLean, 2019), *remedy* (Version 0.1.0; Fay et al., 2020), *report* (Version 0.1.0; Makowski et al., 2020), *retractcheck* (Version 1.0.0; Hartgerink & Aust, 2019), *rmarkdown* (Version 2.3; Allaire et al., 2020), *scales* (Version 1.1.1; Wickham & Seidel, 2020), *semPlot* (Version 1.1.2; Epskamp, 2019), *semTools* (Version 0.5.3; Jorgensen et al., 2020), *servr* (Version 0.18; Xie, 2020c), *sjlabelled* (Version 1.1.6; Lüdecke, 2020b), *sjmisc* (Version 2.8.5; Lüdecke, 2020c), *sjPlot* (Version 2.8.4; Lüdecke, 2020d), *sjstats* (Version 0.18.0; Lüdecke, 2020e), *skimr* (Version 2.1.2; Waring et al., 2020), *stargazer* (Version 5.2.2; Hlavac, 2018), *statcheck* (Version 1.3.0; Michele B. Nuijten <m.b.nuijten@uvt.nl>, 2018), *stringr* (Version 1.4.0; Wickham, 2019c), *survival* (Version 3.2.3; Therneau, 2020), *TH.data* (Version 1.0.10; Hothorn, 2019), *tidyr* (Version 1.1.1; Wickham & Henry, 2020), *tigris* (Version 1.0; Walker, 2020), *tinytex* (Version 0.25; Xie, 2020d), *webshot* (Version 0.5.2; Chang, 2019), *wesanderson* (Version 0.3.6; Ram & Wickham, 2018), *yardstick* (Version 0.0.7; Kuhn & Vaughan, 2020), and *zipcode* (Version 1.0; Breen, 2012). All R packages used are open-source statistical tools built with R, a statistical programming language, in order to facilitate specialized tasks like graphing or printing high quality tables.

Glossary

DEFINITION OF VARIABLES

Democratic Vote Share

$$Democratic\ Vote\ Share = \frac{Votes\ for\ Democrat\ Presidential\ Candidate}{Total\ Votes\ Cast\ in\ County}$$

Margin of Victory

$$Margin\ of\ Victory\ (M)\ = \frac{(Votes_{Most\ Preferred\ Candidate}\ -\ Votes_{Runner\ Up})}{Votes_{Total\ Cast\ by\ County}}$$

Also, see Signed Margin of Victory

Republican Vote Share

$$Republican\ Vote\ Share = \frac{Votes\ for\ Republican\ Presidential\ Candidate}{Total\ Votes\ Cast\ in\ County}$$

Signed Margin of Victory

$$Signed\ Margin\ of\ Victory\ = \frac{(Votes\ _{Winner\ of\ Presidential\ Election}\ -\ Votes\ _{Runner\ Up})}{Votes\ _{Total\ Cast\ by\ County}}$$

AGI over PI

$$AGI\ over\ PI = \frac{Adjusted\ Gross\ Income\ \ _{IRS\ Data}}{Personal\ Income\ \ _{BEA\ Data}}$$

Also, see Adjusted Gross Income and Personal Income in the Definitions of Common Terms

Unmatchable AGI

$$AGI_{Unmatchable} = AGI_{Total}\ -\ Wage\ -\ Interest\ -\ Dividend\ Incomes$$

ABBREVIATIONS

AGI

Adjusted Gross Income. See entry in Definitions of Common Terms for more details on how Personal Income is defined.

AGI-over-PI

Ratio of Adjusted Gross Income as reported by the IRS data / Personal Income as estimated by the BEA. For the current work, this serves as a proxy measure of compliance.

BEA

Bureau of Economic Analysis. As per their website, the "BEA is an independent, principal federal statistical agency that promotes a better understanding of the U.S. economy by providing timely, relevant, and accurate economic accounts data in an objective and cost-effective manner."

BLS

Bureau of Labor Statistics.

CBP

County Business Patterns survey

CPI-U

Consumer Price Index —Urban (Bureau of Labor Statistics)

CPS

Current Population Survey

DVS

Democratic Vote Share i.e. the proportion of a county's votes cast in support of the Democratic Presidential Candidate in an election

FEC

The Federal Election Commission (FEC) is a Federal regulatory agency responsible for administering and enforcing campaign finance laws at the Federal Level. It is supposed to be independent and serve as a watchdog —attempting to secure the integrity of the election rules vis-a-vis campaigns.

GCB

Global Corruption Barometer, an annual global corruption survey by Transparency International

IRS

The Internal Revenue Service is responsible for tax collections for the United States federal government. It is a bureau of the Department of the Treasury.

PI

Personal Income as defined by the BEA. See entry in Definitions of Common Terms for more details on how Personal Income is defined.

SOI

Statistics of Income (SOI) for Individual Income is a publically distributed data product compiled by the IRS. These data serve as the primary source of tax data for the analyses included in the current document.

TI

Transparency International —a non-profit organization responsible for the largest global corruption survey: the Global Corruption Barometer, which formed the basis of the corruption survey used in Part I of the dissertation.

DEFINITIONS OF COMMON TERMS

A

Absolute Majority Standard The absolute majority standard is one of the two most common thresholds used to determine electoral outcomes (also see Simple Majority Standard). It requires that one of the candidates received 50% or more of the votes in order to be declared in a winner. In countries that use such a standard (e.g. France, Brazil), if none of the candidates receive 50% of the votes, a "run-off" election is held between the top two vote getters. In contrast, the United States usually relies upon a simple majority standard to decide elections (and, most reports of presidential election outcomes at the county level use this standard for classification of counties).

Adjusted Gross Income Adjusted Gross Income consists of the taxable income from all sources, less the adjustments to income, such as IRS deductions, alimony etc. This is the equivalent of line 33 of Form 1040 for Tax Year 2001.

B

BEA County Equivalents As per the BEA Website[1]:

BEA counties and county equivalents are identified by five-digit Federal Information Processing Standard (FIPS) codes. The first two digits are the FIPS state code and the last three digits are the county code within the state. BEA has created modified FIPS codes for the following conditions:

- Kalawao County, Hawaii is combined with Maui County and combined area is designated 15901.

[1] https://apps.bea.gov/regional/docs/statelist.cfm

- The independent cities of Virginia with populations of less than 100,000 have been combined with an adjacent county and given codes beginning with 51901. In the name of the combined area, the county name appears first and is followed by the city name(s).
- Menominee County, Wisconsin is combined with Shawano County for 1969-1988 as 55901. Separate estimates for Menominee and Shawano Counties begin in 1989.

A detailed list of BEA modifications to FIPS codes is available in Excel formats.

C

Consistent Plurality Standard In aggregating county election outcomes across the 7 elections, I defined the "Consistent Plurality Standard" as an approach to categorizing counties. Under this standard, I classify a county as being loyal to a certain party if that party gained a plurality of votes in all 7 elections (i.e. that party won a simple majority in every election in the sample). For more details, see the section on Consistent Plurality: Using Counts of Simple Majority.

D

Demeaned Variable —Two-Way Demeaning a variable can be used to remove fixed effects in a panel regression and is theoretically equivalent to the inclusion of "dummy variables" in a simple linear regression. Two-way demeaning is useful in panel data to account for both time and entity fixed effects.

E

Exemptions Exemptions —also known as personal exemptions —reflect a minimum deduction prior to the AGI available to all tax filers and their dependents. The total Number of Exemptions reflects the number of individuals covered on a tax returns i.e. the person filing and any other person who they claimed as a dependent.

N

Nominal USD In order to facilitate comparisons across years, it is often necessary to adjust income data for inflation. In the context of inflation adjusted (real USD), the year-specific numerical amounts are referred to as nominal (i.e. unadjusted for inflation). In the current work, I rely upon the Bureau of Labor Statistics' (BLS) Consumer Price Index for Urban consumers (CPI-U) to convert nominal USD into inflation-adjusted "real" USD. Also, see Real USD.

P

Personal Income As per the BEA's Definition of Regional Terms: 'Personal Income is the income that is received by all persons from all sources. It is calculated as the sum of wages and salaries, supplements to wages and salaries, proprietors' income with inventory valuation and capital consumption adjustments, rental income of persons with capital consumption adjustment, personal dividend income, personal interest income, and personal current transfer receipts, less contributions for government social insurance.

Personal income- Consists of the income that persons receive in return for their provision of labor, land, and capital used in current production as well as other income, such as personal current transfer receipts. In the state and local personal income accounts the personal income of an area represents the income received by or on behalf of the persons residing in that area. It is calculated as the sum of wages and salaries, supplements to wages and salaries, proprietors' income with inventory valuation (IVA) and capital consumption adjustments (CCAdj), rental income of persons with capital consumption adjustment (CCAdj), personal dividend income, personal interest income, and personal current transfer receipts, less contributions for government social insurance plus the adjustment for residence.

Perfect Loyalty Standard In aggregating county election outcomes across the 7 elections, I defined the "Perfect Loyalty Standard" as an approach to categorizing counties. Under this standard, I classify a county as being loyal to a certain party if that party gained an absolute majority (i.e. > 50% of all ballots cast) in all 7 elections. For more details, see the section on Perfect Loyalty (Party Vote Share > 50% For All 7 Elections)

Plurality See Simple Majority Standard.

PVI A method for characterizing the partisanship of a district or subdivision, based upon the metric introduced by the Cook Political Report. As per their website, the PVI or CPVI "measures how each district performs at the presidential level compared to the nation as a whole." Since 1996, following every election, new PVI scores have been released by the Cook Political Report —"each time taking into account the prior two presidential elections."

As they state in an example, in 2016, "a Partisan Voting Index score of D+2, for example, means that in the 2012 and 2016 presidential elections, that district performed an average of two points more Democratic than the nation did as a whole, while an R+4 means the district performed four points more Republican than the national average. If a district performed within half a point of the national average in either direction, we assign it a score of EVEN."

Note: Recent work by Nate Silver described on fivethirtyeight.com, has shown that the PVI index can be significantly improved by the inclusion of a single measure of socio-economic status: namely, percentage of households with income below $25,000. This —in part — motivates our use of SAIPE Poverty Rate measures as a crucial conrol variable.

R

Real USD In order to facilitate comparisons across years, it is often necessary to adjust income data for inflation. In the current work, I rely upon the Bureau of Labor Statistics' (BLS) Consumer Price Index for Urban consumers (CPI-U) to convert nominal USD into inflation-adjusted figures. These inflation-adjusted income measures are referred to as being in "real" USD. Also, see Nominal USD.

S

SAIPE The Census Bureau website describes their SAIPE data series as follows: "The U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program provides annual estimates of income and poverty statistics for all school districts, counties, and states. The main objective of this program is to provide estimates of income and poverty for the administration of federal programs and the allocation of federal funds to local jurisdictions. In addition to these federal programs, state and local programs use the income and poverty estimates for distributing funds and managing programs."

Simple Majority Standard The simple majority standard is one of the two most common thresholds used to determine electoral outcomes (also see Absolute Majority Standard). Under such a standard, the candidate that receives the most votes is declared a winner, whether or not they received over 50% of the votes. If 3rd Party candidates are allowed, this can produce the paradoxical outcome where the winner of the election may have received more votes against them than in their support. For example, if Candidate A received 40%, Candidate B received 35%, and Candidate C received 25%, Candidate A can be declared the winner, even though 60% of voters did not vote for that candidate.

Statistics of Income The IRS received approximately 254 million filings in Financial Year 2018 (relating to earnings in the previous year Jan-Dec 2017).[1] These filings forms the IRS Master File, which serves as the entire population of tax data for any given year.

Based upon the Master File, the IRS produces an annual series of data products, collectively known as the Statistics of Income (SOI), that capture the tax-related activities of businesses, non-profits, charitable entities, and individuals. For the purposes of this chapter, we focus solely on the SOI for Individual Income tax aggregated at the County Level.

These data series and studies provide statistics on "income, deductions, tax, and credits reported on individual Form 1040 income tax returns and associated schedules are available in this area." see SOI page for additional information and examples

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, *87*(4), 1115–1153. https://doi.org/10.3982/ECTA14673

Abramo, C. W. (2008). How Much Do Perceptions of Corruption Really Tell Us? *Economics: The Open-Access, Open-Assessment E-Journal*, *2*(2008-3), 1. https://doi.org/10.5018/economics-ejournal.ja.2008-3

Abramowitz, A. I., & Webster, S. W. (2018). Negative Partisanship: Why Americans Dislike Parties But Behave Like Rabid Partisans. *Political Psychology*, *39*(S1), 119–135. https://doi.org/10.1111/pops.12479

Achen, C. H. (2000). Why lagged dependent variables can suppress the explanatory power of other independent variables. *Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA*, *20*, 07–2000.

Ades, A., & Di Tella, R. (1999). Rents, Competition, and Corruption. *American Economic Review*, *89*(4), 982–993. https://doi.org/10.1257/aer.89.4.982

Alemann, U. V. (2004). The unknown depths of political theory: The case for a multi-dimensional concept of corruption. *Crime, Law and Social Change*, *42*(1), 25–34. https://doi.org/10.1023/B:CRIS.0000041035.21045.1d

Alexander, A. C., & Welzel, C. (2017). The Myth of Deconsolidation: Rising Liberalism and the Populist Reaction. *Journal of Democracy*, 27.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *Rmarkdown: Dynamic documents for r*. https://CRAN.R-project.org/package=rmarkdown

Alm, J., Bahl, R., & Murray, M. N. (1990). Tax Structure and Tax Compliance. *The Review of Economics and Statistics*, *72*(4), 603–613. https://doi.org/10.2307/2109600

Alm, J., Jackson, B., & McKee, M. (1992). Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal*, *45*(1), 107–114.

Alm, J., Martinez-Vazquez, J., & Torgler, B. (2010). *Developing Alternative Frameworks for Explaining Tax Compliance*. Routledge.

Alm, J., McClelland, G. H., & Schulze, W. D. (1992). Why do people pay taxes? *Journal of Public Economics*, *48*(1), 21–38. https://doi.org/10.1016/0047-2727(92)90040-M

Alm, J., & Yunus, M. (2009). Spatiality and Persistence in U.S. Individual Income Tax Compliance. *National Tax Journal*, *62*(1), 101–124. https://doi.org/10.17310/ntj.2009.1.05

Almunia, M., & Lopez-Rodriguez, D. (2018). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*, *10*(1), 1–38. https://doi.org/10.1257/pol.20160229

Altieri, K., Kirkland, B., & Paur, F. (1998). *Joker's Millions: Vols. Season 4, Episode 7*. Warner Bros. Entertainment.

Andersson, S. (2017). Beyond Unidimensional Measurement of Corruption. *Public Integrity*, *19*(1), 58–76. https://doi.org/10.1080/10999922.2016.1200408

Andersson, S., & Heywood, P. M. (2009). The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption. *Political Studies*, *57*(4), 746–767. https://doi.org/10.1111/j.1467-9248.2008.00758.x

Andrew, J. A. (2002). *Power to destroy: The political uses of the IRS from Kennedy to Nixon*. Ivan R Dee.

Arndt, C. (2008). The Politics of Governance Ratings. *International Public Management Journal*, *11*(3), 275–297. https://doi.org/10.1080/10967490802301278

Arndt, C., & Oman, C. P. (2006). *Uses and Abuses of Governance Indicators*. OECD Publishing.

Arnold, J. B. (2019). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. https://CRAN.R-project.org/package=ggthemes

Asher, J., & Horwitz, B. (2020). It's Been "Such a Weird Year." That's Also Reflected in Crime Statistics. *The New York Times*.

Atkeson, L. R., Alvarez, R. M., & Hall, T. E. (2015). Voter Confidence: How to Measure It and How It Differs from Government Support. *Election Law Journal: Rules, Politics, and Policy*, *14*(3), 207–219. https://doi.org/10.1089/elj.2014.0293

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. https://CRAN.R-project.org/package=gridExtra

Aust, F. (2019). *Citr: RStudio add-in to insert markdown citations*. https://CRAN.R-project.org/package=citr

Aust, F., & Barth, M. (2019). *Papaja: Prepare reproducible apa journal articles with r markdown*. https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r.* https://CRAN.R-project.org/package=magrittr

Bai, B.-y., Liu, X.-x., & Kou, Y. (2016). Belief in a just world lowers bribery intention: Belief in a just world and bribery intention. *Asian Journal of Social Psychology*, *19*(1), 66–75. https://doi.org/10.1111/ajsp.12108

Banuri, S., & Eckel, C. (2012). *Experiments in Culture and Corruption: A Review* (Policy Research Working Paper Nos. WPS6064; p. 29). Development Research Group, The World Bank. https://doi.org/10.1596/1813-9450-6064

Barr, A., & Serra, D. (2010). Corruption and culture: An experimental analysis. *Journal of Public Economics*, *94*(1112), 862–869. https://doi.org/10.1016/j.jpubeco.2010.07.006

Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods.* https://CRAN.R-project.org/package=Matrix

Bauer, P. C., & Freitag, M. (2018). Measuring Trust. *The Oxford Handbook of Social and Political Trust.* https://doi.org/10.1093/oxfordhb/9780190274801.013.1

Bauldry, S. (2015). Structural Equation Modeling. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition, pp. 615–620). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.44055-9

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with Time-Series Cross-Section Data. *The American Political Science Review*, *89*(3), 634–647. https://doi.org/10.2307/2082979

Beck, N., & Katz, J. N. (2011). Modeling Dynamics in Time-SeriesCross-Section Political Economy Data. *Annual Review of Political Science*, *14*(1), 331–352. https://doi.org/10.1146/annurev-polisci-071510-103222

Beetham, D. (1991a). Legitimacy in Political Science and Political Philosophy. In *The Legitimation of Power* (pp. 243–250). Macmillan Education UK. https://doi.org/10.1007/978-1-349-21599-7_8

Beetham, D. (1991b). *The Legitimation of Power.* Macmillan Education UK. https://doi.org/10.1007/978-1-349-21599-7

Belousova, V., Goel, R. K., & Korhonen, I. (2016). Corruption perceptions versus corruption incidence: Competition for rents across Russian regions. *Journal of Economics and Finance*, *40*(1), 172–187. https://doi.org/10.1007/s12197-014-9298-y

Benzarti, Y. (2015a). *How Taxing Is Tax Filing? Leaving Money on the Table Because of Hassle Costs* (SSRN Scholarly Paper ID 2412703). Social Science Research Network.

Benzarti, Y. (2015b). How Taxing Is Tax Filing? Leaving Money on the Table Because of Compliance Costs. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *108*, 1–79.

Berge, L. (2020). *Fixest: Fast fixed-effects estimations*. https://CRAN.R-project.org/package=fixest

Berger, S., Graham, N., & Zeileis, A. (2017). *Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R* [Working Paper]. Faculty of Economics and Statistics, University of Innsbruck.

Bernasconi, M., Corazzini, L., & Seri, R. (2014). Reference dependent preferences, hedonic adaptation and tax evasion: Does the tax burden matter? *Journal of Economic Psychology*, *40*, 103–118. https://doi.org/10.1016/j.joep.2013.01.005

Bernasconi, M., & Zanardi, A. (2004). Tax Evasion, Tax Rates, and Reference Dependence. *FinanzArchiv / Public Finance Analysis*, *60*(3), 422–445.

Bernheim, B. D., & Rangel, A. (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *The Quarterly Journal of Economics*, *124*(1), 51–104. https://doi.org/10.1162/qjec.2009.124.1.51

Bernheim, B. D., & Rangel, A. (2007). Toward Choice-Theoretic Foundations for Behavioral Welfare Economics. *American Economic Review*, *97*(2), 464–470. https://doi.org/10.1257/aer.97.2.464

Bernheim, B. D., & Rangel, A. (2005a). *Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers* (Working Paper Nos. 11518). National Bureau of Economic Research. https://doi.org/10.3386/w11518

Bernheim, B. D., & Rangel, A. (2005b). *Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers* (Working Paper Nos. 11518). National Bureau of Economic Research.

Berry, C. R., Burden, B. C., & Howell, W. G. (2010). The President and the Distribution of Federal Spending. *American Political Science Review*, *104*(4), 783–799. https://doi.org/10.1017/S0003055410000377

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. https:

//doi.org/10.1162/003355304772839588

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *The American Economic Review*, *91*(2), 67–72.

Bialik, C. (2015). Scare headlines exaggerated the US crime wave. *FiveThirtyEight. Retrieved from Https://Fivethirtyeight. Com/Features/Scare-Headlines-Exaggerated-the-Us-Crime-Wave*.

Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Sonnet, L. (2020). *Estimatr: Fast estimators for design-based inference*. https://CRAN.R-project.org/package=estimatr

Blais, A., Gidengil, E., & Kilibarda, A. (2017). Partisanship, Information, and Perceptions of Government Corruption. *International Journal of Public Opinion Research*, *29*(1), 95–110. https://doi.org/10.1093/ijpor/edv037

Bloomquist, K. M. (2003). Tax Evasion, Income Inequality and Opportunity Costs of Compliance. *Proceedings of the Annual Conference of the National Tax Association*, *96*, 91–104.

Bloomquist, K. M., & An, Z. (2005). Geographic variation in schedule h filing rates: Why should location influence the decision to report“ Nanny” taxes? *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *98*, 26–39.

Blumenthal, M., Christian, C., & Slemrod, J. (1998). *The Determinants of Income Tax Compliance: Evidence from a Controlled Experiment in Minnesota* (Working Paper Nos. 6575; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w6575

Boas, T. C., Christenson, D. P., & Glick, D. M. (2018). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods*, 1–19. https://doi.org/10.1017/psrm.2018.28

Bollen, K. A., & Maydeu-Olivares, A. (2007). A Polychoric Instrumental Variable (PIV) Estimator for Structural Equation Models with Categorical Variables. *Psychometrika*, *72*(3), 309–326. https://doi.org/10.1007/s11336-007-9006-3

Bonica, A. (2019). Are Donation-Based Measures of Ideology Valid Predictors of Individual-Level Policy Preferences? *The Journal of Politics*, *81*(1), 327–333. https://doi.org/10.1086/700722

Boning, W. C., Guyton, J., Hodge, I., Ronald H, Slemrod, J., & Troiano, U. (2018). *Heard it Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment*

(Working Paper Nos. 24305). National Bureau of Economic Research. https://doi.org/10.3386/w24305

Börzel, T. A., & Hüllen, V. van. (2014). State-Building and the European Union's Fight against Corruption in the Southern Caucasus: Why Legitimacy Matters. *Governance*, *27*(4), 613–634. https://doi.org/10.1111/gove.12068

Braithwaite, V. A. (2009). *Defiance in taxation and governance: Resisting and dismissing authority in a democracy*. Edward Elgar.

Brand, A., & Bradley, M. T. (2012). More voodoo correlations: When average-based measures inflate correlations. *The Journal of General Psychology*, *139*(4), 260–272.

Breen, J. (2012). *Zipcode: U.S. ZIP code database for geocoding*. https://CRAN.R-project.org/package=zipcode

Bristow, W. (2017). Enlightenment. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.

Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, *11*(3), 55–87. https://doi.org/10.1257/pol.20160589

Brown, C. (2018). *Formula.tools: Programmatic utilities for manipulating formulas, expressions, calls, assignments and other r objects*. https://CRAN.R-project.org/package=formula.tools

Buckley v. Valeo (Nos. 75-436; Vol. 424, p. 1). (1976). (Vol. 424, p. 1).

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore. *Journal of Public Economics*, *93*(78), 843–851. https://doi.org/10.1016/j.jpubeco.2009.03.004

Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization*, *93*, 258–265. https://doi.org/10.1016/j.jebo.2013.03.037

Castro, E., & Scartascini, C. (2019). *Imperfect Attention in Public Policy: A Field Experiment during a Tax Amnesty in Argentina*. Inter-American Development Bank. https://doi.org/10.18235/0001661

Castro, L., & Scartascini, C. (2015). Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization*, *116*, 65–82. https://doi.org/10.1016/j.jebo.2015.04.002

Chairman Hal Rogers Talks IRS Targeting and Spending. (2013). In *America's Newsroom*. Fox News.

Chang, W. (2019). *Webshot: Take screenshots of web pages*. https://CRAN.R-project.org/package=webshot

Charron, N. (2016). Do corruption measures have a perception problem? Assessing the relationship between experiences and perceptions of corruption among citizens and experts. *European Political Science Review*, *8*(01), 147–171. https://doi.org/10.1017/S1755773914000447

Chen, M. K., & Rohla, R. (2018). The effect of partisanship and political advertising on close family ties. *Science (New York, N.Y.)*, *360*(6392), 1020–1024. https://doi.org/10.1126/science.aaq1433

Cheng, J., Karambelkar, B., & Xie, Y. (2019). *Leaflet: Create interactive web maps with the javascript 'leaflet' library*. https://CRAN.R-project.org/package=leaflet

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, *99*(4), 1145–1177. https://doi.org/10.1257/aer.99.4.1145

Chetverikov, A. (2019). *Apastats: APA-style results*.

Cho, E., & Kim, S. (2015). Cronbach's Coefficient Alpha: Well Known but Poorly Understood. *Organizational Research Methods*, *18*(2), 207–230. https://doi.org/10.1177/1094428114555994

Citizens United v. Federal Election Com'n (Nos. 08-205; Vol. 558, p. 310). (2010). (Vol. 558, p. 310).

Citrin, J., & Stoker, L. (2018). Political Trust in a Cynical Age. *Annual Review of Political Science*, *21*(1), 49–70. https://doi.org/10.1146/annurev-polisci-050316-092550

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical

Turk valid for research on political ideology? *Research & Politics*, *2*(4). https://doi.org/10.1177/2053168015622072

Congdon, W., Kling, J. R., & Mullainathan, S. (2009a). *Behavioral Economics and Tax Policy* (Working Paper Nos. 15328). National Bureau of Economic Research. https://doi.org/10.3386/w15328

Congdon, W., Kling, J. R., & Mullainathan, S. (2009b). *Behavioral Economics and Tax Policy* (Working Paper Nos. 15328). National Bureau of Economic Research. https://doi.org/10.3386/w15328

Congdon, W., Kling, J. R., & Mullainathan, S. (2009c). *Behavioral economics and Tax policy*. National Bureau of Economic Research.

Corrado, G., Corrado, L., De Michele, G., & Salustri, F. (2017). *Are Perceptions of Corruption Matching Reality? Theory and Evidence from Microdata* (SSRN Scholarly Paper ID 3086711). Social Science Research Network. https://doi.org/10.2139/ssrn.3086711

Council of the European Union. (2011). *Action to combat economic fraud and corruption: Vol. TA* (Nos. 0388).

Cousineau, D. (2017). Varieties of Confidence Intervals. *Advances in Cognitive Psychology*, *13*(2), 140–155. https://doi.org/10.5709/acp-0214-z

Croissant, Y., Millo, G., & Tappe, K. (2020). *Plm: Linear models for panel data*. https://CRAN.R-project.org/package=plm

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Cullen, J. B., Turner, N., & Washington, E. L. (2018). *Political Alignment, Attitudes Toward Government and Tax Evasion* (Working Paper Nos. 24323). National Bureau of Economic Research. https://doi.org/10.3386/w24323

Cummings, R. G., Martinez-Vazquez, J., McKee, M., & Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior & Organization*, *70*(3), 447–457. https://doi.org/10.1016/j.jebo.2008.02.010

Čábelková, I., & Hanousek, J. (2004). The power of negative thinking: Corruption, perception and willingness to bribe in Ukraine. *Applied Economics*, *36*(4), 383–397. https://doi.org/10.1080/00036840410001674303

Dahlberg, S., & Linde, J. (2016). Losing Happily? The Mitigating Effect of Democracy and Quality of Government on the WinnerLoser Gap in Political Support. *International Journal of Public Administration*, *39*(9), 652–664. https://doi.org/10.1080/01900692.2016.1177831

Dancey, L. (2012). The Consequences of Political Cynicism: How Cynicism Shapes Citizens' Reactions to Political Scandals. *Political Behavior*, *34*(3), 411–423. https://doi.org/10.1007/s11109-011-9163-z

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

DeBacker, J., Heim, B. T., & Tran, A. (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States. *Journal of Financial Economics*, *117*(1), 122–138. https://doi.org/10.1016/j.jfineco.2012.11.009

Devi, T., & Fryer, J., Roland G. (2020). *Policing the police: The impact of "Pattern-or-Practice" investigations on crime* (Working Paper Nos. 27324; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w27324

Dhami, S., & al-Nowaihi, A. (2010). Optimal taxation in the presence of tax evasion: Expected utility versus prospect theory. *Journal of Economic Behavior & Organization*, *75*(2), 313–337. https://doi.org/10.1016/j.jebo.2010.03.004

Dhami, S., & al-Nowaihi, A. (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior & Organization*, *64*(1), 171–192. https://doi.org/10.1016/j.jebo.2006.08.006

Djawadi, B., & Fahr, R. (2013). *The Impact of Tax Knowledge and Budget Spending Influence on Tax Compliance.*

Doane, W. (2019). *Issuer: Issue tracker using the local file system.* https://github.com/WilDoane/issuer

Doerrenberg, P. (2015). Does the use of tax revenue matter for tax compliance behavior? *Economics Letters*, *128*, 30–34. https://doi.org/10.1016/j.econlet.2015.01.005

Doerrenberg, P., & Schmitz, J. (2017). *Tax Compliance and Information Provision - A Field Experiment with Small Firms* (SSRN Scholarly Paper ID 2941474). Social Science Research Network.

Donchev, D., & Ujhelyi, G. (2014). What Do Corruption Indices Measure? *Economics & Politics*, *26*(2), 309–331. https://doi.org/10.1111/ecpo.12037

Dong, B., & Torgler, B. (2009). Corruption and Political Interest: Empirical Evidence at the Micro Level. *Journal of Interdisciplinary Economics*, *21*(3), 295–325. https://doi.org/10.1177/02601079X09002100304

Dowle, M., & Srinivasan, A. (2020). *Data.table: Extension of 'data.frame'*. https://CRAN.R-project.org/package=data.table

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197–199. https://doi.org/10.1016/j.econlet.2007.06.027

Dreher, A., Kotsogiannis, C., & McCorriston, S. (2004). Corruption Around the World: Evidence from a Structural Model. *Social Science Research Network*, 31.

Dufwenberg, M., & Dufwenberg, M. A. (2018). Lies in disguise A theoretical analysis of cheating. *Journal of Economic Theory*, *175*, 248–264. https://doi.org/10.1016/j.jet.2018.01.013

Duhaime, E. P., & Apfelbaum, E. P. (2017). Can Information Decrease Political Polarization? Evidence From the U.S. Taxpayer Receipt. *Social Psychological and Personality Science*, 194855061668712. https://doi.org/10.1177/1948550616687126

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M., & Torgler, B. (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics*, *134*, 9–18. https://doi.org/10.1016/j.jpubeco.2015.12.007

Dynes, A. M., & Huber, G. A. (2015). Partisanship and the Allocation of Federal Spending: Do Same-Party Legislators or Voters Benefit from Shared Party Affiliation with the President and House Majority? *American Political Science Review*, *109*(1), 172–186. https://doi.org/10.1017/S000305541400063X

Eberwein, K. (2019). *BlscrapeR: An api wrapper for the bureau of labor statistics (bls)*. https://CRAN.R-project.org/package=blscrapeR

Engström, P., Nordblom, K., Ohlsson, H., & Persson, A. (2015). Tax Compliance and Loss Aversion. *American Economic Journal: Economic Policy*, *7*(4), 132–164. https://doi.org/10.1257/pol.20130134

Epskamp, S. (2015). semPlot: Unified Visualizations of Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 474–483. https://doi.org/10.1080/10705511.2014.937847

Epskamp, S. (2019). *SemPlot: Path diagrams and visual analysis of various sem packages'* *output*. https://CRAN.R-project.org/package=semPlot

Erat, S. (2013). Avoiding lying: The case of delegated deception. *Journal of Economic Behavior & Organization*, *93*, 273–278. https://doi.org/10.1016/j.jebo.2013.03.035

Erat, S., & Gneezy, U. (2011). White Lies. *Management Science*, *58*(4), 723–733. https://doi.org/10.1287/mnsc.1110.1449

Ertel, S. (2011). Exploratory factor analysis revealing complex structure. *Personality and Individual Differences*, *50*(2), 196–200. https://doi.org/10.1016/j.paid.2010.09.026

Esarey, J., & Menger, A. (2019). Practical and Effective Approaches to Dealing With Clustered Data. *Political Science Research and Methods*, *7*(3), 541–559. https://doi.org/10.1017/psrm.2017.42

Escresa, L., & Picci, L. (2015). A New Cross-National Measure of Corruption. *The World Bank Economic Review*, lhv031. https://doi.org/10.1093/wber/lhv031

Escresa, L., & Picci, L. (2016). Trends in Corruptions Around the World. *European Journal on Criminal Policy and Research*, *22*(3), 543–564. https://doi.org/10.1007/s10610-016-9317-y

Fay, C., Sidi, J., & Smith, L. (2020). *Remedy: RStudio addins to simplify 'markdown' writing*. https://github.com/ThinkR-open/remedy

Fazekas, M., & Kocsis, G. (2017a). *Replication Data for: Uncovering high-level corruption: Cross-national objective corruption risk indicators using public procurement data*. https://doi.org/10.7910/DVN/6XYZOD

Fazekas, M., & Kocsis, G. (2017b). Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. *British Journal of Political Science*, 1–10. https://doi.org/10.1017/S0007123417000461

Fazekas, M., & Tóth, B. (2017). Proxy indicators for the corrupt misuse of corporations. *U4 Brief*, *2017:6*.

Feld, L. P., & Frey, B. S. (2002). Trust breeds trust: How taxpayers are treated. *Economics of Governance*, *3*(2), 87–99. https://doi.org/10.1007/s101010100032

Feld, L. P., & Frey, B. S. (2004). *Illegal, immoral, fattening or what?: How deterrence and responsive regulation shape tax morale*. Marburger volkswirtschaftliche Beiträge.

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political Extremism Is

Supported by an Illusion of Understanding. *Psychological Science*, *24*(6), 939–946. https://doi.org/10.1177/0956797612464058

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in DisguiseAn Experimental Study on Cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/jeea.12014

Fischetti, T. (2020). *Assertr: Assertive programming for r analysis pipelines*. https://CRAN.R-project.org/package=assertr

Fisman, R., & Miguel, E. (2007). Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets. *Journal of Political Economy*, *115*(6), 1020–1048. https://doi.org/10.1086/527495

Fisman, R., & Miguel, E. (2014). *Replication data for: Corruption, Norms and Legal Enforcement: Evidence from Diplomatic Parking Tickets*. https://doi.org/10.7910/DVN/28059

Flesken Anaïd, & Hartl Jakob. (2017). Party Support, Values, and Perceptions of Electoral Integrity. *Political Psychology*, *0*(0). https://doi.org/10.1111/pops.12431

Foa, R. S., & Mounk, Y. (2017a). The End of the Consolidation Paradigm: A Response to Our Critics. *Journal of Democracy*, Web Exchange.

Foa, R. S., & Mounk, Y. (2016). The Democratic Disconnect. *Journal of Democracy*, *27*(3), 5–17. https://doi.org/10.1353/jod.2016.0049

Foa, R. S., & Mounk, Y. (2017b). The signs of deconsolidation. *Journal of Democracy*, *28*(1), 5–15.

Fosgaard, T. R., Hansen, L. G., & Piovesan, M. (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization*, *93*, 279–284. https://doi.org/10.1016/j.jebo.2013.03.027

Fox, J., Weisberg, S., & Price, B. (2020a). *Car: Companion to applied regression*. https://CRAN.R-project.org/package=car

Fox, J., Weisberg, S., & Price, B. (2020b). *CarData: Companion to applied regression data sets*. https://CRAN.R-project.org/package=carData

Frank, M. R., Cebrian, M., Pickard, G., & Rahwan, I. (2017). Validating Bayesian truth serum in large-scale online human experiments. *PLOS ONE*, *12*(5), e0177385. https://doi.org/10.1371/journal.pone.0177385

Galili, T. (2019). *Installr: Using r to install stuff on windows os (such as: R, 'rtools', rstudio, 'git', and more!)*. https://CRAN.R-project.org/package=installr

Gallup. (2015). *75% in U.S. See Widespread Government Corruption* (News Briefing 2014 World Poll Results). Gallup.

Gaure, S. (2020). *Lfe: Linear group fixed effects*. https://CRAN.R-project.org/package=lfe

Gaure, S., McDermott, G., Werner, K. D., Stigler, M., & Lüdecke, D. (2019). *Lfe: Linear Group Fixed Effects*.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*.

Gentry, W. M., & Kahn, M. E. (2009). Understanding Spatial Variation in Tax Sheltering: The Role of Demographics, Ideology, and Taxes. *International Regional Science Review*, *32*(3), 400–423. https://doi.org/10.1177/0160017609336536

Genz, A., Bretz, F., Miwa, T., Mi, X., & Hothorn, T. (2020). *Mvtnorm: Multivariate normal and t distributions*. https://CRAN.R-project.org/package=mvtnorm

Gerber, A. S., & Huber, G. A. (2010). Partisanship, Political Control, and Economic Assessments. *American Journal of Political Science*, *54*(1), 153–173. https://doi.org/10.1111/j.1540-5907.2009.00424.x

Gerber, A. S., Huber, G. A., & Washington, E. (2010). Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment. *American Political Science Review*, *104*(4), 720–744. https://doi.org/10.1017/S0003055410000407

Gohel, D. (2020). *Officer: Manipulation of microsoft word and powerpoint documents*. https://CRAN.R-project.org/package=officer

Golden, M. A., & Picci, L. (2005). Proposal for a New Measure of Corruption, Illustrated with Italian Data. *Economics & Politics*, *17*(1), 37–75. https://doi.org/10.1111/j.1468-0343.2005.00146.x

Goldsmith, A. A. (1999). Slapping the Grasping Hand. *American Journal of Economics and Sociology*, *58*(4), 865–883. https://doi.org/10.1111/j.1536-7150.1999.tb03398.x

Gordon, M., Gragg, S., & Konings, P. (2020). *HtmlTable: Advanced tables for markdown/html*. https://CRAN.R-project.org/package=htmlTable

Grantham, N. (2019). *Ggdark: Dark mode for 'ggplot2' themes*. https://CRAN.R-project.org/package=ggdark

Graves, S., Piepho, H.-P., & Sundar Dorai-Raj, L. S. with help from. (2019). *MultcompView: Visualizations of paired comparisons*. https://CRAN.R-project.org/package=multcompView

Greene, J. (2016). Solving the trolley problem. *A Companion to Experimental Philosophy*, 175–178.

Groseclose, T., & Snyder, J. M. (1996). Buying Supermajorities. *The American Political Science Review*, *90*(2), 303–315. https://doi.org/10.2307/2082886

Gutmann, J., Padovano, F., & Voigt, S. (2015). *Perception vs. Experience: Explaining Differences in Corruption Measures Using Microdata* (SSRN Scholarly Paper ID 2659349). Social Science Research Network.

Gutmann, J., Padovano, F., & Voigt, S. (2020). Perception vs. Experience: Explaining differences in corruption measures using microdata. *European Journal of Political Economy*, *65*, 101925. https://doi.org/10.1016/j.ejpoleco.2020.101925

Habermas, J. (1975). *Legitimation crisis* (Vol. 519). Beacon Press.

Habibov, N., Cheung, A., & Auchynnikava, A. (2018). Does Institutional Trust Increase Willingness to Pay More Taxes to Support the Welfare State? *Sociological Spectrum*, *38*(1), 51–68. https://doi.org/10.1080/02732173.2017.1409146

Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, *30*(4), 658–679. https://doi.org/10.1093/oxrep/gru034

Hallsworth, M., List, J., Metcalfe, R., & Vlaev, I. (2014). *The behavioralist as tax collector: Using natural field experiments to enhance tax compliance*. National Bureau of Economic Research.

Hao, L., & Houser, D. (2017). Perceptions, intentions, and cheating. *Journal of Economic Behavior & Organization*, *133*, 52–73. https://doi.org/10.1016/j.jebo.2016.10.010

Harrell Jr, F. E., Charles Dupont, & others. (2020). *Hmisc: Harrell miscellaneous*. https://CRAN.R-project.org/package=Hmisc

Hartgerink, C., & Aust, F. (2019). *Retractcheck: Retraction scanner*. https://CRAN.R-project.org/package=retractcheck

Hawken, A., & Munck, G. L. (2008). Measuring Corruption. A Critical Assessment and a Proposal. In A. Rajivan & R. Gampat (Eds.), *Perspectives on Corruption and Human Development* (Vol. 1, p. 38). Macmillan India.

Hawken, A., & Munck, G. L. (2011). *Does the Evaluator Make a Difference? Measurement Validity in Corruption Research* (SSRN Scholarly Paper ID 2480857). Social Science Research Network.

Heath, A. F., Richards, L., & de Graaf, N. D. (2016). Explaining Corruption in the Developed World: The Potential of Sociological Approaches. *Annual Review of Sociology*, *42*(1), 51–79. https://doi.org/10.1146/annurev-soc-081715-074213

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. https://CRAN.R-project.org/package=purrr

Hessing, D. J., Elffers, H., & Weigel, R. H. (1988). Exploring the limits of self-reports and reasoned action: An investigation of the psychology of tax evasion behavior. *Journal of Personality and Social Psychology*, *54*(3), 405–413. https://doi.org/10.1037/0022-3514.54.3.405

Heywood, P. M., & Rose, J. (2014). “Close but no Cigar”: The measurement of corruption. *Journal of Public Policy*, *34*(3), 507–529. https://doi.org/10.1017/S0143814X14000099

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition*, *163*, 15–25. https://doi.org/10.1016/j.cognition.2017.02.011

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables*. https://CRAN.R-project.org/package=stargazer

Hofmann, E., Hoelzl, E., & Kirchler, E. (2008). Preconditions of Voluntary Tax Compliance: Knowledge and Evaluation of Taxation, Norms, Fairness, and Motivation to Cooperate. *Zeitschrift Fur Psychologie*, *216*(4), 209–217. https://doi.org/10.1027/0044-3409.216.4.209

Holder, E. H., Jr. (2014). *Remarks by Attorney General Holder at the International Association of Chiefs of Police Annual Conference* [Keynote Address].

Holmberg, S., Rothstein, B., & Nasiritousi, N. (2009). Quality of Government: What You Get. *Annual Review of Political Science*, *12*(1), 135–161. https://doi.org/10.1146/annurev-polisci-100608-104510

Hooghe, M., & Dassonneville, R. (2018). A Spiral of Distrust: A Panel Study on the Relation between Political Distrust and Protest Voting in Belgium. *Government and Opposition*, *53*(1), 104–130. https://doi.org/10.1017/gov.2016.18

Hothorn, T. (2019). *TH.data: TH's data archive*. https://CRAN.R-project.org/package=TH.data

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Hothorn, T., Bretz, F., & Westfall, P. (2020). *Multcomp: Simultaneous inference in general parametric models*. https://CRAN.R-project.org/package=multcomp

Hough, M., Jackson, J., Bradford, B., Myhill, A., & Quinton, P. (2010). Procedural Justice, Trust, and Institutional Legitimacy. *Policing: A Journal of Policy and Practice*, *4*(3), 203–210. https://doi.org/10.1093/police/paq027

Howe, P. (2017). Eroding Norms and Democratic Deconsolidation. *Journal of Democracy*, *28*(4), 15–29. https://doi.org/10.1353/jod.2017.0061

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, G. A., & Malhotra, N. (2017). Political Homophily in Social Relationships: Evidence from Online Dating Behavior. *The Journal of Politics*, *79*(1), 269–283. https://doi.org/10.1086/687533

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 2053168015604648. https://doi.org/10.1177/2053168015604648

Inglehart, R. F. (2016). How Much Should We Worry? *Journal of Democracy*, *27*(3), 18–23. https://doi.org/10.1353/jod.2016.0053

IRS. (2020a). *Dependents, Standard Deduction, and Filing Information for Filing in 2019* (Nos. 501). Internal Revenue Service.

IRS. (2020b). *Comprehensive Taxpayer Attitude Survey* (Executive Report 5296 (Rev. 32020); Catalog Number 71353Y). Internal Revenue Service.

IRS. (2019a). *Internal Revenue Service, Department of the Treasury - Data Book, 2018* (Government Report Publication 55B (Rev. 52019); p. 86). Internal Revenue Service.

IRS. (2019b). *Statistics of Income - 2017* (Nos. 1304). Internal Revenue Service.

IRS. (2013). *2013 Annual Report to Congress* (p. 76) [Executive Summary]. Internal Revenue Service.

Iyengar, S., & Krupenkin, M. (2018). The Strengthening of Partisan Affect. *Political Psychology*, *39*(S1), 201–218. https://doi.org/10.1111/pops.12487

Jackson, J. (2018). Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives. *Annual Review of Law and Social Science*, *14*(1), 145–165. https://doi.org/10.1146/annurev-lawsocsci-110316-113734

Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., & Tyler, T. (2012). Why do People Comply with the Law? Legitimacy and the Influence of Legal Institutions. *The British Journal of Criminology*, *52*(6), 1051–1071. https://doi.org/10.1093/bjc/azs032

Jacobsen, C., & Piovesan, M. (2016). Tax me if you can: An artifactual field experiment on dishonesty. *Journal of Economic Behavior & Organization*, *124*, 7–14. https://doi.org/10.1016/j.jebo.2015.09.009

Johnson, C., Dowd, T. J., & Ridgeway, C. L. (2006). Legitimacy as a Social Process. *Annual Review of Sociology*, *32*(1), 53–78. https://doi.org/10.1146/annurev.soc.32.061604.123101

Johnson, R. A., & Sharma, S. (2004). About Corruption. In R. A. Johnson (Ed.), *The Struggle Against Corruption: A Comparative Study* (pp. 1–19). Palgrave Macmillan US. https://doi.org/10.1057/9781403981004_1

Jong-sung, Y., & Khagram, S. (2005a). A Comparative Study of Inequality and Corruption. *American Sociological Review*, *70*(1), 136–157. https://doi.org/10.1177/000312240507000107

Jong-sung, Y., & Khagram, S. (2005b). A Comparative Study of Inequality and Corruption (Correction). *American Sociological Review*, *70*(3), 539–539. https://doi.org/10.1177/000312240507000309

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). *SemTools: Useful tools for structural equation modeling*. https://CRAN.R-project.org/package=semTools

Jöreskog, K. G., & Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, *36*(3), 347–387. https://doi.org/10.1207/S15327906347-387

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational*

Behavior and Human Decision Processes, *51*(2), 296–312.

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (pp. 49–81). Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kassambara, A. (2020). *Ggpubr: 'Ggplot2' based publication ready plots*. https://CRAN.R-project.org/package=ggpubr

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2004). Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002. *The World Bank Economic Review*, *18*(2), 253–287. https://doi.org/10.1093/wber/lhh041

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The Worldwide Governance Indicators: Methodology and Analytical Issues* (SSRN Scholarly Paper ID 1682130). Social Science Research Network.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*, *3*(2), 220–246. https://doi.org/10.1017/S1876404511200046

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, *29*(2), 210–225. https://doi.org/10.1016/j.joep.2007.05.004

Kirchler, E., Muehlbacher, S., Kastlunger, B., & Wahl, I. (2007). *Why Pay Taxes? A Review of Tax Compliance Decisions* (Nos. paper0730). International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, *79*(3), 651–692. https://doi.org/10.3982/ECTA9113

Ko, K., & Samajdar, A. (2010). Evaluation of international corruption indexes: Should we believe them or not? *The Social Science Journal*, *47*(3), 508–540. https://doi.org/10.1016/j.soscij.2010.03.001

Kuhn, M., & Vaughan, D. (2020). *Yardstick: Tidy characterizations of model performance*. https://CRAN.R-project.org/package=yardstick

Lafree, G. (2018). *Losing Legitimacy: Street Crime And The Decline Of Social Institutions In America*. Routledge.

Lamberton, C. P., De Neve, J.-E., & Norton, M. I. (2014). *Eliciting taxpayer preferences increases tax compliance*.

Lambsdorff, J. G. (2006). Measuring corruptionthe validity and precision of subjective indicators (CPI). *Measuring Corruption*, *81*, 81.

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, *21*(5), 737–744. https://doi.org/10.1177/0956797610368810

Langbein, L., & Knack, S. (2010). The Worldwide Governance Indicators: Six, One, or None? *The Journal of Development Studies*, *46*(2), 350–370. https://doi.org/10.1080/00220380902952399

Larmarange, J. (2020). *Labelled: Manipulating labelled data*. https://CRAN.R-project.org/package=labelled

Lederman, D., Loayza, N. V., & Soares, R. R. (2005). Accountability and corruption: Political institutions matter. *Economics & Politics*, *17*(1), 35. https://doi.org/10.1111/j.1468-0343.2005.00145.x

Leip, D. (2019). Dave Leip's Atlas of U.S. Presidential Elections. In *Dave Leip's Atlas of U.S. Presidential Elections*.

Lenth, R. (2020). *Emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The Demographic and Political Composition of Mechanical Turk Samples. *SAGE Open*, *6*(1), 2158244016636433. https://doi.org/10.1177/2158244016636433

Levi, M., Sacks, A., & Tyler, T. (2009). Conceptualizing Legitimacy, Measuring Legitimating Beliefs. *American Behavioral Scientist*, *53*(3), 354–375. https://doi.org/10.1177/0002764209338797

Levi, M., Tyler, T., & Sacks, A. (2012). The Reasons for Compliance with Law. In R. Goodman, D. Jinks, & A. K. Woods (Eds.), *Understanding Social Action, Promoting Human*

Rights (pp. 70–91). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195371895.003.0004

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, *21*(2), 153–174. https://doi.org/10.1257/jep.21.2.153

Levitt, S. D., & Snyder, J. M. (1997). The Impact of Federal Spending on House Election Outcomes. *Journal of Political Economy*, *105*(1), 30–53. https://doi.org/10.1086/262064

Levitt, S. D., & Snyder, J. M. (1995). Political Parties and the Distribution of Federal Outlays. *American Journal of Political Science*, *39*(4), 958–980. https://doi.org/10.2307/2111665

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/met0000093

Liu, Y., Magnus, B., O'Connor, H., & Thissen, D. (2018). Multidimensional Item Response Theory. In *The Wiley Handbook of Psychometric Testing* (pp. 445–493). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489772.ch16

Lopez-Luzuriaga, A., & Scartascini, C. (2019). Compliance spillovers across taxes: The role of penalties and detection. *Journal of Economic Behavior & Organization*, *164*, 518–534. https://doi.org/10.1016/j.jebo.2019.06.015

Luttmer, E. F. P., & Singhal, M. (2014). Tax Morale. *Journal of Economic Perspectives*, *28*(4), 149–168. https://doi.org/10.1257/jep.28.4.149

Lüdecke, D. (2020a). *Ggeffects: Create tidy data frames of marginal effects for 'ggplot' from model outputs*. https://CRAN.R-project.org/package=ggeffects

Lüdecke, D. (2020b). *Sjlabelled: Labelled data utility functions*. https://CRAN.R-project.org/package=sjlabelled

Lüdecke, D. (2020c). *Sjmisc: Data and variable transformation functions*. https://CRAN.R-project.org/package=sjmisc

Lüdecke, D. (2020d). *SjPlot: Data visualization for statistics in social science*. https:

//CRAN.R-project.org/package=sjPlot

Lüdecke, D. (2020e). *Sjstats: Collection of convenient functions for common statistical computations*. https://CRAN.R-project.org/package=sjstats

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, *51*(1), 201–226. https://doi.org/10.1146/annurev.psych.51.1.201

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Makowski, D., Lüdecke, D., & Ben-Shachar, M. S. (2020). *Report: Automated reporting of statistical models in r*. https://github.com/easystats/report

Makridis, C., & Rothwell, J. T. (2020). *The Real Cost of Political Polarization: Evidence from the COVID-19 Pandemic* (SSRN Scholarly Paper ID 3638373). Social Science Research Network.

Malito, D. V. (2014). *Measuring Corruption Indicators and Indices* (Working Papers Nos. 13; p. 34). Robert Schuman Centre for Advanced Studies (RSCAS).

Matheus, A., & Paulo, F. L. (2016). Narrow Replication of Fisman and Miguel's (2007a) "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets". *Journal of Applied Econometrics*, *32*(4), 919–922. https://doi.org/10.1002/jae.2540

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*, *110*(3), 681–712. https://doi.org/10.2307/2946696

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

McArdle, J. J., & Kadlec, K. M. (2013). Structural Equation Models. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (First, Vol. 2). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199934898.013.0015

McConnell v. Federal Election Comm'n (Nos. 02-1674; Vol. 540, p. 93). (2003). (Vol. 540, p. 93).

McCredie, M. N., & Morey, L. C. (2018). Who Are the Turkers? A Characterization of MTurk Workers Using the Personality Assessment Inventory. *Assessment*, 1073191118760709.

https://doi.org/10.1177/1073191118760709

McLean, M. W. (2019). *RefManageR: Straightforward 'bibtex' and 'biblatex' bibliography management*. https://CRAN.R-project.org/package=RefManageR

Meijer, R. R., & Tendeiro, J. N. (2018). Unidimensional Item Response Theory. In *The Wiley Handbook of Psychometric Testing* (pp. 413–443). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489772.ch15

Melgar, N., Rossi, M., & Smith, T. W. (2010). The Perception of Corruption. *International Journal of Public Opinion Research*, *22*(1), 120–131. https://doi.org/10.1093/ijpor/edp058

Mendes Fialho, F. (2017). *Institutional Trust in Cross-National Research: A Measurement Invariance Approach* [Masters of Science in Statistics]. University of California, Los Angeles (UCLA).

Mian, A. R., Sufi, A., & Khoshkhou, N. (2018). *Partisan Bias, Economic Expectations, and Household Spending* (SSRN Scholarly Paper ID 2620828). Social Science Research Network. https://doi.org/10.2139/ssrn.2620828

Michele B. Nuijten <m.b.nuijten@uvt.nl>, S. E. &. (2018). *Statcheck: Extract statistics from articles and recompute p values*. https://CRAN.R-project.org/package=statcheck

Microsoft, & Weston, S. (2020). *Foreach: Provides foreach looping construct*. https://CRAN.R-project.org/package=foreach

MIT Election Data and Science Lab. (2018). *U.S. President 19762016*. https://doi.org/10.7910/DVN/42MVDX

MIT Election Data and Science Lab. (2019). *County Presidential Election Returns 2000-2016*. https://doi.org/10.7910/DVN/VOQCHQ

Mocan, N. (2004). *What Determines Corruption? International Evidence from Micro Data* (Working Paper Nos. 10460; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w10460

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morgan, S. L., & Pally, J. A. (2020). *Ferguson, Gray, and Davis: An Analysis of Recorded Crime Incidents and Arrests in Baltimore City, March 2010 through December 2015* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/nshme

Morning Consult. (2016). *Yes, There Are Shy Trump Voters. No, They Won't Swing the Election.* [Political Survey]. Morning Consult.

Morris, S. D., & Klesner, J. L. (2010). Corruption and trust: Theoretical considerations and evidence from Mexico. *Comparative Political Studies*, *43*(10), 1258–1285.

Nagin, D. S., & Telep, C. W. (2017). Procedural Justice and Legal Compliance. *Annual Review of Law and Social Science*, *13*(1), 5–28. https://doi.org/10.1146/annurev-lawsocsci-110316-113310

Neumann, R., & Graeff, P. (2010). A Multitrait-Multimethod approach to pinpoint the validity of aggregated governance indicators. *Quality & Quantity*, *44*(5), 849–864. https://doi.org/10.1007/s11135-009-9238-7

Nix, J., & Pickett, J. T. (2017). Third-person perceptions, hostile media effects, and policing: Developing a theoretical framework for assessing the Ferguson effect. *Journal of Criminal Justice*, *51*, 24–33. https://doi.org/10.1016/j.jcrimjus.2017.05.016

Nixon v. Shrink Missouri Government PAC (Nos. 98-963; Vol. 528, p. 377). (2000). (Vol. 528, p. 377).

Norris, P. (2017). Is Western Democracy Backsliding? Diagnosing the Risks. *Journal of Democracy*, Web Exchange.

Oc, B., Moore, C., & Bashshur, M. R. (2018). When the tables are turned: The effects of the 2016 U.S. Presidential election on in-group favoritism and out-group hostility. *PLOS ONE*, *13*(5), e0197848. https://doi.org/10.1371/journal.pone.0197848

OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide.* OECD publishing.

Olken, B. A. (2009). Corruption perceptions vs. Corruption reality. *Journal of Public Economics*, *93*(78), 950–964. https://doi.org/10.1016/j.jpubeco.2009.03.001

Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*, *4*(1), 479–509. https://doi.org/10.1146/annurev-economics-080511-110917

Oman, C. P., & Arndt, C. (2006). Uses and abuses of governance indicators. *OECD iLibrary.*

Overby, P. (2017). As IRS Targeted Tea Party Groups, It Went After Progressives Too. *NPR.org.*

Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Veszteg, R., Gil-Gómez de Liaño, B., Gunia, B., Weichselbaumer, D., Hilken, K., Antinyan, A., Delnoij, J., Proestakis, A., Tira, M. D., Pratomo, Y., Jaber-López, T., & Brañas-Garza, P. (2015). Context-dependent cheating: Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization*, *116*, 379–386. https://doi.org/10.1016/j.jebo.2015.04.020

Pázmándy, M. (2011). Socio-Economic Influences on Corruption Perception - Empirical Evidence from 27 European Countries. *Hamburg Review of Social Sciences*, *6*(2), 52–80.

Pedersen, T. L. (2020). *Patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

Peter, F. (2017). Political Legitimacy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University.

Pew Research Center. (2020). Far more Americans see "very strong" partisan conflicts now than in the last two presidential election years. In *Pew Research Center*.

Pew Research Center. (2019a). *Growing Partisan Divide Over Fairness of the Nation's Tax System* (p. 18).

Pew Research Center. (2017a). *Public Trust in Government Remains Near Historic Lows as Partisan Attitudes Shift* (pp. 1–16). Pew Research Center.

Pew Research Center. (2015). 2. Russian Public Opinion: Putin Praised, West Panned. In *Pew Research Center's Global Attitudes Project*.

Pew Research Center. (2016). *Partisanship and Political Animosity in 2016*. Pew Research Center.

Pew Research Center. (2017b). 1. President Putin: The Russian perspective. In *Pew Research Center's Global Attitudes Project*.

Pew Research Center. (2017c). 2. Russians' take on the economy and problems at home. In *Pew Research Center's Global Attitudes Project*.

Pew Research Center. (2017d). *The Partisan Divide on Political Values Grows Even Wider*. Pew Research Center.

Pew Research Center. (2019b). *Partisan Antipathy: More Intense, More Personal*. Pew Research Center.

Pew Research Center. (2014). Global worries about corruption are on the rise. In *Pew*

Research Center.

Pew Research Center. (2019c). *In a Politically Polarized Era, Sharp Divides in Both Partisan Coalitions* (p. 132). Pew Research Center.

Philp, M., & Dávid-Barrett, E. (2015). Realism About Political Corruption. *Annual Review of Political Science*, *18*(1), 387–402. https://doi.org/10.1146/annurev-polisci-092012-134421

Pickhardt, M., & Prinz, A. (2014). Behavioral dynamics of tax evasion A survey. *Journal of Economic Psychology*, *40*, 1–19. https://doi.org/10.1016/j.joep.2013.08.006

Piedmont, R. L. (2014). Inter-item Correlations. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493

Pierce, K. (2015). SOI migration data: A new approach: Methodological improvements for SOI's united states population migration data, calendar years 2011-2012. *Statistics of Income. SOI Bulletin*, *35*(1).

Plumley, A. H. (1996). *The determinants of individual income tax compliance: Estimating the impacts of tax policy, enforcement, and IRS responsiveness.* [Ph.D.]. Harvard University.

Plumley, A. H. (2002). The impact of the IRS on voluntary tax compliance: Preliminary empirical results. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *95*, 355–363.

Presidential Election Results: Donald J. Trump Wins. (2017). *The New York Times*.

President's Task Force on 21st Century Policing. (2015). *Final Report of the President's Task Force on 21st Century Policing. :* Office of Community Oriented Police Service, US Department of Justice.

Pyrooz, D. C., Decker, S. H., Wolfe, S. E., & Shjarback, J. A. (2016). Was there a Ferguson Effect on crime rates in large U.S. Cities? *Journal of Criminal Justice*, *46*, 1–8. https://doi.org/10.1016/j.jcrimjus.2016.01.001

Raiche, G. (2010). *An R package for parallel analysis and non graphical solutions to the Cattell scree test*.

Ram, K., & Wickham, H. (2018). *Wesanderson: A wes anderson palette generator*. https://CRAN.R-project.org/package=wesanderson

Ramnath, S. P., & Tong, P. K. (2017). The Persistent Reduction in Poverty from Filing a

Tax Return. *American Economic Journal: Economic Policy*, *9*(4), 367–394. https://doi.org/10.1257/pol.20150486

Razafindrakoto, M., & Roubaud, F. (2010). Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa. *World Development*, *38*(8), 1057–1069. https://doi.org/10.1016/j.worlddev.2010.02.004

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rees-Jones, A. (2018). Quantifying Loss-Averse Tax Manipulation. *The Review of Economic Studies*, *85*(2), 1251–1278. https://doi.org/10.1093/restud/rdx038

Rees-Jones, A., & Taubinsky, D. (2016). *Heuristic Perceptions of the Income Tax: Evidence and Implications for Debiasing* (Working Paper Nos. 22884). National Bureau of Economic Research. https://doi.org/10.3386/w22884

Reinhart, R. J. (2020). Faith in Elections in Relatively Short Supply in U.S. *Gallup.com*.

Ren, K., & Russell, K. (2016). *Formattable: Create 'formattable' data structures*. https://CRAN.R-project.org/package=formattable

Revelle, W. (2017). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University.

Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research*. https://CRAN.R-project.org/package=psych

Revelle, W., & Condon, D. M. (2018). Reliability. In *The Wiley Handbook of Psychometric Testing* (pp. 709–749). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489772.ch23

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Ripley, B. (2020). *MASS: Support functions and datasets for venables and ripley's mass*. https://CRAN.R-project.org/package=MASS

Rivas, M. F. (2013). An Experiment on Corruption and Gender. *Bulletin of Economic Research*, *65*(1), 10–42. https://doi.org/10.1111/j.1467-8586.2012.00450.x

Robinson, D., Hayes, A., & Couch, S. (2020). *Broom: Convert statistical objects into tidy tibbles*. https://CRAN.R-project.org/package=broom

Rose, R., & Mishler, W. (2010). Experience versus perception of corruption: Russia as a test case. *Global Crime*, *11*(2), 145–163. https://doi.org/10.1080/17440571003669175

Rosseel, Y. (2011). Lavaan: An R Package for Structural Equation Modeling. *J Stat Softw*, *48*. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., Jorgensen, T. D., & Rockwood, N. (2020). *Lavaan: Latent variable analysis*. https://CRAN.R-project.org/package=lavaan

Rothstein, B. (2009). Creating Political Legitimacy: Electoral Democracy Versus Quality of Government. *American Behavioral Scientist*, *53*(3), 311–330. https://doi.org/10.1177/0002764209338795

Rothstein, B., & Uslaner, E. M. (2005). All for All: Equality, Corruption, and Social Trust. *World Politics*, *58*(1), 41–72. https://doi.org/10.1353/wp.2006.0022

Rousseau, J.-J. (2018). *Rousseau: The Social Contract and other later political writings*. Cambridge University Press.

Saltelli, A., Giovannini, E., Tarantola, S., Saisana, M., Hoffman, A., & Nardo, M. (2005). *Handbook on Constructing Composite Indicators* (OECD Statistics Working Papers 2005/03). https://doi.org/10.1787/533411815016

Sances, M. W., & Stewart, C. (2015). Partisanship and confidence in the vote count: Evidence from U.S. National elections since 2000. *Electoral Studies*, *40*, 176–188. https://doi.org/10.1016/j.electstud.2015.08.004

Sarkar, D. (2020). *Lattice: Trellis graphics for r*. https://CRAN.R-project.org/package=lattice

Schneider, I. (2017). Can We Trust Measures of Political Trust? Assessing Measurement Equivalence in Diverse Regime Types. *Social Indicators Research*, *133*(3), 963–984. https://doi.org/10.1007/s11205-016-1400-8

Schurr, A., & Ritov, I. (2016). Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences*, *113*(7), 1754–1759. https://doi.org/10.1073/pnas.1515102113

Sheth, S. (2019). *America's Top Fears 2019* (Survey Results Wave 6 (2019)). Chapman

University.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Silver, N. (2017). The Comey Letter Probably Cost Clinton The Election. In *FiveThirtyEight*.

Slowikowski, K. (2020). *Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'*. https://CRAN.R-project.org/package=ggrepel

Soares, R. R. (2004). Crime Reporting as a Measure of Institutional Development. *Economic Development and Cultural Change*, *52*(4), 851–871. https://doi.org/10.1086/420900

Stanley, D. (2018). *ApaTables: Create american psychological association (apa) style tables*. https://CRAN.R-project.org/package=apaTables

Stevens, A. (2016). Configurations of corruption: A cross-national qualitative comparative analysis of levels of perceived corruption. *International Journal of Comparative Sociology*.

Sunshine, J., & Tyler, T. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review*, *37*(3), 513–548. https://doi.org/10.1111/1540-5893.3703002

Sussman, A. B., & Olivola, C. Y. (2011). Axe the tax: Taxes are disliked more than equivalent costs. *Journal of Marketing Research*, *48*(SPL), S91–S101.

Sussman, A. B., & White, S. M. (2018). Negative Responses to Taxes: Causes and Mitigation. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 224–231. https://doi.org/10.1177/2372732218790008

Svensson, J. (2005). Eight Questions about Corruption. *Journal of Economic Perspectives*, *19*(3), 19–42. https://doi.org/10.1257/089533005774357860

Swamy, A., Knack, S., Lee, Y., & Azfar, O. (2001). Gender and corruption. *Journal of Development Economics*, *64*(1), 25–55. https://doi.org/10.1016/S0304-3878(00)00123-1

Tankebe, J. (2013). Viewing Things Differently: The Dimensions of Public Perceptions of Police Legitimacy. *Criminology*, *51*(1), 103–135. https://doi.org/10.1111/j.1745-9125.2012.00291.x

Tankebe, J., & Gutierrez-Gomez, L. (2017). Procedural Justice. *The Wiley-Blackwell Encyclopedia of Social Theory*, 1–3.

Tankebe, J., & Gutierrez-Gomez, L. (2018). Procedural Justice. In B. S. Turner (Ed.), *The Wiley encyclopedia of social theory* (pp. 1–3). Wiley-Blackwell. https://doi.org/10.1002/9781118430873.est0293

Tankebe, J., Reisig, M. D., & Wang, X. (2016). A multidimensional model of police legitimacy: A cross-cultural assessment. *Law and Human Behavior*, *40*(1), 11–22. https://doi.org/10.1037/lhb0000153

Taranto, J. (2013). President Asterisk. *Wall Street Journal*.

Tax Gap Estimates for Tax Years 20082010. (2016). Interal Revenue Service.

Therneau, T. M. (2020). *Survival: Survival analysis*. https://CRAN.R-project.org/package=survival

Thomas, M. A. (2010). What do the worldwide governance indicators measure? *The European Journal of Development Research*, *22*(1), 31–54.

Thorley, J. (2019). *Checkr: Check the properties of common r objects*. https://CRAN.R-project.org/package=checkr

Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics*, *76*(3), 399–457. https://doi.org/10.1016/S0047-2727(99)00092-4

Treisman, D. (2007). What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research? *Annual Review of Political Science*, *10*(1), 211–244. https://doi.org/10.1146/annurev.polisci.10.081205.095418

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00769

Tyler, T. (2006). Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology*, *57*(1), 375–400. https://doi.org/10.1146/annurev.psych.57.102904.190038

Tyler, T. (2017a). *Value-Driven Behavior and the Law* (F. Parisi, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199684267.013.030

Tyler, T. (2004). Enhancing Police Legitimacy. *The Annals of the American Academy of Political and Social Science*, *593*, 84–99.

Tyler, T. (2017b). Procedural Justice and Policing: A Rush to Judgment? *Annual Review of Law and Social Science*, *13*(1), 29–53. https://doi.org/10.1146/annurev-lawsocsci-110316-113318

Tyler, T., Goff, P. A., & MacCoun, R. J. (2015). The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement. *Psychological Science in the Public Interest*, *16*(3), 75–109. https://doi.org/10.1177/1529100615617791

Tyler, T., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law*, *20*(1), 78–95. https://doi.org/10.1037/a0034514

Ullman, J. B., & Bentler, P. M. (2003). Structural Equation Modeling. *Handbook of Psychology*. https://doi.org/10.1002/0471264385.wei0224

Ulman, S.-R. (2014). The Impact of the National Competitiveness on the Perception of Corruption. *Procedia Economics and Finance*, *15*, 1002–1009. https://doi.org/10.1016/S2212-5671(14)00660-1

United Nations Development Programme, & Global Anti-corruption Initiative. (2015). *A Users' Guide to Measuring Corruption and Anti-corruption* (2nd Edition). UNDP Global Anti-corruption Initiative (GAIN).

Update on reducing the federal tax gap and improving voluntary compliance. (2009). Interal Revenue Service.

Vaidyanathan, R., Xie, Y., Allaire, J., Cheng, J., & Russell, K. (2019). *Htmlwidgets: HTML widgets for r*. https://CRAN.R-project.org/package=htmlwidgets

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Vanderkam, D., Allaire, J., Owen, J., Gromer, D., & Thieurmel, B. (2018). *Dygraphs: Interface to 'dygraphs' interactive time series charting library*. https://CRAN.R-project.org/package=dygraphs

van der Wal, W. M. (2018). *Elections: USA presidential elections data*. https://CRAN.R-project.org/package=elections

Voeten, E. (2017). Are people really turning away from democracy? *Journal of Democracy*, Web Exchange.

Walker, K. (2020). *Tigris: Load census tiger/line shapefiles*. https://CRAN.R-project.org/package=tigris

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLoS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0198835

Wang, Y., & Dickson, B. (2018). *How Corruption Investigations Undermine Regime Support: Evidence from China* (SSRN Scholarly Paper ID 3086286). Social Science Research Network.

Waring, E., Quinn, M., McNamara, A., Arino de la Rubia, E., Zhu, H., & Ellis, S. (2020). *Skimr: Compact and flexible summaries of data*. https://CRAN.R-project.org/package=skimr

Warren, M. (2017). Trust and Democracy. In E. M. Uslaner (Ed.), *The Oxford Handbook of Social and Political Trust* (2nd Edition, Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190274801.013.5

Warren, M. E. (2004). What Does Corruption Mean in a Democracy? *American Journal of Political Science*, *48*(2), 328–343. https://doi.org/10.1111/j.0092-5853.2004.00073.x

Warren, M. E. (2006). Democracy and Deceit: Regulating Appearances of Corruption. *American Journal of Political Science*, *50*(1), 160–174.

Weber, M. (2009). *The theory of social and economic organization*. Simon and Schuster.

Wei, S.-J. (1999). Corruption in economic development: Beneficial grease, minor annoyance, or major obstacle? *World Bank Policy Research Working Paper*, *2048*.

Wei, T., & Simko, V. (2017). *Corrplot: Visualization of a correlation matrix*. https://CRAN.R-project.org/package=corrplot

Westfall, J., Boven, L. V., Chambers, J. R., & Judd, C. M. (2015). Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691615569849

Wickham, H. (2019a). *Assertthat: Easy pre and post assertions*. https://CRAN.R-project.org/package=assertthat

Wickham, H. (2019b). *Conflicted: An alternative conflict resolution strategy*. https://CRAN.R-project.org/package=conflicted

Wickham, H. (2019c). *Stringr: Simple, consistent wrappers for common string operations.* https://CRAN.R-project.org/package=stringr

Wickham, H. (2020). *Plyr: Tools for splitting, applying and combining data.* https://CRAN.R-project.org/package=plyr

Wickham, H., & Bryan, J. (2019). *Readxl: Read excel files.* https://CRAN.R-project.org/package=readxl

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2020). *Ggplot2: Create elegant data visualisations using the grammar of graphics.* https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation.* https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2020). *Tidyr: Tidy messy data.* https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data.* https://CRAN.R-project.org/package=readr

Wickham, H., & Seidel, D. (2020). *Scales: Scale functions for visualization.* https://CRAN.R-project.org/package=scales

Wilke, C. O. (2019). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.* https://CRAN.R-project.org/package=cowplot

Wilkins, A. S. (2018). To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis*. *Political Science Research and Methods*, *6*(2), 393–411. https://doi.org/10.1017/psrm.2017.4

Wu, B., & Muñana, C. (2018). Kaiser Health Tracking Poll Late Summer 2018: The Election, Pre-Existing Conditions, and Surprises on Medical Bills. In *The Henry J. Kaiser Family Foundation.*

Xie, Y. (2020a). *Bookdown: Authoring books and technical documents with r markdown.* https://CRAN.R-project.org/package=bookdown

Xie, Y. (2020b). *Knitr: A general-purpose package for dynamic report generation in r.* https://CRAN.R-project.org/package=knitr

Xie, Y. (2020c). *Servr: A simple http server to serve static files or dynamic documents.* https://CRAN.R-project.org/package=servr

Xie, Y. (2020d). *Tinytex: Helper functions to install and maintain tex live, and compile latex documents*. https://CRAN.R-project.org/package=tinytex

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, *11*(10). https://doi.org/10.18637/jss.v011.i10

Zeileis, A. (2006). Object-Oriented Computation of Sandwich Estimators. *Journal of Statistical Software*, *16*(9). https://doi.org/10.18637/jss.v016.i09

Zeileis, A., & Croissant, Y. (2018). *Formula: Extended model formulas*. https://CRAN.R-project.org/package=Formula

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Zmerli, S., & Hooghe, M. (2013). *Political Trust: Why Context Matters*. ECPR Press.

Zmerli, S., & Meer (Eds.). (2017). *Handbook on Political Trust*. Edward Elgar Publishing.