Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 19-40
October 2019
https://doi.org/10.21799/frbp.wp.2019.40

Policy Inertia, Election Uncertainty and Incumbency Disadvantage of Political Parties

Satyajit Chatterjee
Federal Reserve Bank of Philadelphia Research Department

Burcu Eyigungor
Federal Reserve Bank of Philadelphia Research Department

ISSN: 1962-5361

Policy Inertia, Election Uncertainty and Incumbency Disadvantage of Political Parties

Satyajit Chatterjee and Burcu Eyigungor

Federal Reserve Bank of Philadelphia

October 2019

Abstract

We document that postwar U.S. elections show a strong pattern of "incumbency disadvantage": If a party has held the presidency of the country or the governorship of a state for some time, that party tends to lose popularity in the subsequent election. We show that this fact can be explained by a combination of policy inertia and unpredictability in election outcomes. A quantitative analysis shows that the observed magnitude of incumbency disadvantage can arise in several different models of policy inertia. Normative and positive implications of policy inertia leading to incumbency disadvantage are explored.

Keywords: rational partisan model, policy inertia, incumbency disadvantage, election uncertainty, prospective voting

JEL Codes: D72 H50

Satyajit.Chatterjee@phil.frb.org. The authors thank Marina Azzimonti, Hülya Eraslan, Razvan Vlaicu, Ivan Werning and Pierre Yared for helpful suggestions on earlier drafts of this paper. Comments from conference participants at 2017 Society for Economic Dynamics Meetings, the 2017 Stony Brook Summer Workshop in Political Economy, 2017 Midwest Macro Meetings at Pittsburgh, the 2018 LAGV Conference at Aix-en-Provence, and the 2018 NA Summer Econometric Society Meeting and from seminar participants at FRB Chicago, Rice University, Ohio State University and York University are gratefully acknowledged. The authors thank Matt Grossman for sharing the data for the chart in his October 4, 2018 *FiveThirtyEight* blog post.

1 Introduction

This paper is motivated by the following observation about U.S. politics. Since 1954 there have been eight presidential elections in which the presidency had been held continuously by either a Democrat or a Republican for the eight preceding years or more (two or more terms). In seven of those elections, the incumbent president's party could not hold on to the presidency. This fact suggests that there is an *incumbency disadvantage* in U.S. politics: When a party has held the presidency for two or more terms, the popularity of the party with voters is strongly diminished.[1]

We make four contributions. First, we verify that the suggestion of incumbency disadvantage noted above is in fact present in U.S. national and gubernatorial elections during the postwar era. Specifically, we show that the Democratic vote share in the House is strongly affected by how long the two parties have held the presidency going into each election: If the Democratic (Republican) party has held the presidency for six or more years going into an election, the Democratic vote share of the House declines (increases) by 2.4 percentage points, on average.[2] We also study state gubernatorial elections, which allows us to expand the number of elections we can examine. We find that if a party has held the governor's office for six or more years, that party's candidate for governorship garners fewer votes in the subsequent election.[3]

Second, we show that incumbency disadvantage is implied by the Markov perfect equilibrium of an Alesina and Tabellini (1990) style model of partisan politics, if the model environment is extended in two very natural ways. First, political turnover occurs via elections in which outcomes depend on the anticipated policy choices of the two parties as well as on transient voter preference shocks, and, second, there is policy inertia stemming from the costs of (or constraints on) changing policies quickly. These features, combined with diminishing marginal utility, imply incumbency disadvantage of parties.

[1] After their first term in office, most presidents get reelected (this fact may reflect the personal appeal of a president once the public gets to know him and is not addressed in this paper). After two terms in office, a president cannot run for a third term, so the identity of the next presidential party mostly depends on the appeal of party platforms.

[2] We focus on the House because every seat in the House is generally contested in every national election (held every two years) and the scope of the electorate to express approval or disapproval of current policies is the greatest in House elections.

[3] In a political system such as the U.S., the president's and governor's party gets to set the policy agenda at the national and state levels, respectively. So, during national and state elections, we expect voters to vote against (or for) the members of the president's or governor's party if they disapprove (or approve) of the party's current policies.

Third, we show that the magnitude of the incumbency disadvantage effect estimated for U.S. data can arise in a quantitative version of the model for a range of parameter values. Furthermore, the degree of policy inertia required to account for observed incumbency disadvantage has important implications. First, policy inertia may cause both parties to target extreme policies, i.e., policies that are more conservative (liberal) than a conservative (liberal) individual would ideally want. Such extremism entails a welfare loss for all individuals as extremism increases policy volatility as parties cycle into and out of power. Second, going the other way, policy inertia also enlarges the set of environments in which political competition between office-motivated politicians favors centrist policies. Without being able to commit to policies before an election, or being able to employ history-dependent strategies, political parties can improve their chances of election — and of enjoying the benefits of office — by moving policies toward the center.

Fourth, we make a methodological contribution to the partisan political economy framework. Because of forward-looking voting behavior, the probability of a given party winning the election depends on the expected voter lifetime utilities delivered by the parties which, in turn, depend on the parties' anticipated policy choice conditional on winning in current and future elections. To the best of our knowledge, the issue of existence of a Markov equilibrium for a partisan political economy model with this feature has not been settled. Since an existence result is an extremely useful aid to computation, we take a step toward filling this gap by establishing the existence of a stationary pure strategy MPE for a class of models that includes the model of this paper as a special case.

The paper is organized as follows. Section 2 gives a brief review of the literature that bears on this paper. Section 3 presents our main empirical findings on incumbency disadvantage of political parties. Section 4 presents a two-period model to explain how incumbency disadvantage of political parties can arise as a result of policy inertia and voter preference shocks. Section 5 extends the two-period model to an infinite horizon setup that is suitable for computation. Section 6 presents our main quantitative findings. Section 7 discusses some other (potential) explanations of incumbency disadvantage. Section 8 concludes and three appendices collect additional findings, proof of existence and some computational details.

2 Literature Review

On the empirical side, our findings echo previous findings in the politics literature. In an early study, Stokes and Iverson (1962) observed that over the 24 presidential elections between 1868 and 1960, neither the Republican nor the Democratic party succeeded in winning more than 15 percent beyond an equal share of presidential or congressional votes. They saw this as evidence of "restoring forces" that elevate the popularity of the party that has been less popular in the past. More recently, Bartels and Zaller (2001) and Fair (2009) study a large set of empirical presidential vote models and identify an"incumbent fatigue" effect (Bartels and Zaller) and "duration" effect (Fair), wherein the percent of the two-party vote for the party of the incumbent president is negatively affected by how long that party has held the presidency. Our findings for gubernatorial elections are similar to theirs. Relatedly, Erikson (1988) and Alesina and Rosenthal (1989) noted that the president's party loses seats in midterm elections (the so-called "midterm cycle"). We show that there is a long-term incumbency disadvantage not related to midterm cycles.[4]

In the area of macro political economy, we contribute to the growing literature on quantitative-theoretic partisan political economy models. In terms of model structure and quantitative focus the closest is Azzimonti (2011) who also applies probabilistic voting (Lindbeck and Weibull (1987)) to a dynamic setup.[5] Prospective voting and endogenous reelection probabilities have featured in quantitative models of sovereign debt and default (Scholl (2017) and Chatterjee and Eyigungor (2017)) but incumbency disadvantage is not a (necessary) feature of these models. Other recent examples of quantitative political economy models include Mateos-Planas (2012) and Song, Storesletten, and Zilibotti (2012).

In the area of political theory, our work relates to Downs's (1957) celebrated insight that competition for elected office between politicians that care only about winning leads to convergence to centrist policies. Calvert (1985), building on Wittman (1983), showed that if politicians care also about policies then competition will again lead to close-to-centrist policies, provided politicians care sufficiently about winning. Alesina (1988) pointed out that Calvert's and Wittman's results depend crucially on politicians being able to commit to policies before the election. Without commitment,

[4] Alesina and Rosenthal's (1989) explanation of the midterm cycle centered around nonpolarized voters attempting to get closer to median policies by counterbalancing a partisan president's power. In our model, citizens are as polarized as parties but for random reasons do not always vote along party lines. Coupled with the policy drift resulting from inertia, our model predicts an electoral disadvantage that grows with incumbency.

[5] While reelection probability is endogenous in her model, her assumptions regarding preferences implied that the probability of reelection is state independent and thus constant over time.

the Markov perfect equilibrium outcome is for the winning politician to implement her *preferred* (not centrist) policy *regardless* of how much she cared about winning. We show that policy inertia changes this last result: Even without pre-election commitment, a politician's desire to win office will matter and it will be a force in favor of centrist policies.

There is a literature on political business cycles that is based on information frictions. In Rogoff (1990), the competency of the incumbent leader is not directly observable and cycles in fiscal policy arise from the competent type choosing policies that separate her from the incompetent type. In Ales, Maziero, and Yared (2014), the government's private information about its budgetary resources implies an optimal contract with the feature that after a sequence of bad outcomes, the contract is terminated and a new contract is entered into with a new government. Replacement of the government is, thus, endogenous and the logic is closely tied to that in models of political control (Barro (1973) and Ferejohn (1986)). There is also a literature on cycles in expenditures tied to exogenous shifts in the party or coalitions in power (Dixit, Grossman, and Gul (2000), Acemoglu, Golosov, and Tsyvinski (2011), Aguiar and Amador (2011), Battaglini and Coate (2008), and Azzimonti, Battaglini, and Coate (2016)). None of these studies focus on or attempt to explain incumbency disadvantage.

Our results on the positive and normative implications of the costs of adjustments bear a resemblance to equilibrium outcomes in legislative bargaining models with an endogenous status quo (Bowen, Chen, and Eraslan (2014), Piguillem and Riboni (2015) and Dziuda and Loeper (2016)). For instance, Bowen, Chen, and Eraslan (2014) show that mandated spending improves the bargaining position of politicians out of power and leads to equilibrium outcomes closer to the first best.[6] A different strand of the literature (Aghion and Bolton (1990), Milesi-Ferretti and Spolare (1994), Besley and Coate (1998), Hassler, Mora, Storesletten, and Zilibotti (2003)) finds, as we do, the possibility of inferior policies being adopted by incumbent governments (or superior policies ignored) for strategic electoral reasons.

Finally, regarding our methodological contribution, the closest related prior work is Duggan and Kalandrakis (2012). These authors study a general legislative bargaining game where the probability of a legislator coming into power (i.e., getting to be the one to propose and hence

[6] The dynamic link created by mandated spending is analogous to the dynamic link created by the costs of changing inherited policies in our model. However, an important reason why this dynamic link matters in our model is that it determines reelection probabilities, conditional on the state. In contrast, in Bowen, Chen, and Eraslan's (2014) environment (and in environments of many legislative bargaining models) the probability of being a proposer is exogenously given.

choose policies) is under the control of the legislator but *independent* of the policies the legislator is expected to propose (and choose), if she gets into power. In our model, this independence does not hold since a party's anticipated post-election policy choice is a key determinant of the party's probability of winning the election. Thus, a new element of endogeneity is present and our existence proof takes account of it.

3 Incumbency Disadvantage in U.S. Elections: 1954-2016

In this section we show that the incumbency disadvantage effect mentioned in the opening paragraph is a feature of postwar U.S. election outcomes. As noted in our introduction, in empirical models designed to predict presidential election outcomes, duration of the presidential party's incumbency negatively affects the party's reelection probability. In the first subsection, we evaluate party performance using U.S. House election outcomes, which doubles the number of elections we can consider and also has the advantage that the election outcome is less affected by the personal appeal of the two (presidential) candidates. In the second subsection, we extend the analysis to state gubernatorial elections, which allows us to further expand the number of elections we can examine. Our results here parallel the findings for presidential elections: If a party holds the governor's office for a long time, that party's candidate for governor garners fewer votes.

3.1 Congressional Elections

Letting DV denote the percentage two-party share of votes garnered by House Democrats in each national election, our main empirical specification is:[7]

$$DV_t = \beta_0 + \beta_1 SIX_t^+ + \beta_2 (TWO_t^+ \cdot RDPIGR_t) + \beta_3 (TWO_t^+ \cdot MIDTERM_t).$$

Here, SIX_t^+ is a trichotomous variable that takes a value of $+1$ if, at the time of the election, the presidency has been held by a Democrat for 6 or more years, takes a value of -1 if the presidency has been held by a Republican for 6 or more years, and takes the value 0 otherwise. If there is an incumbency disadvantage of political parties, we expect β_1 to be negative. A negative coefficient

[7] The data on House vote share and House vote seats are compiled from the official website of the U.S. House of Representatives at $https://history.house.gov/Institution/Party-Divisions/Party-Divisions/$. The data for presidential party incumbency are from $https://en.wikipedia.org/wiki/President\_of\_the\_United\_States$ and the data for real personal disposable income are from the U.S. Bureau of Economic Analysis, Real Disposable Personal Income: Per Capita [A229RX0], retrieved from FRED, Federal Reserve Bank of St. Louis; $https://fred.stlouisfed.org/series/A229RX0$.

implies that after 6 years of a Democratic presidency the Democratic vote share falls, and after 6 years of a Republican presidency, the Democratic vote share rises.

$TWO_t^+ \cdot RDPIGR_t$ is an interaction variable, where TWO_t^+ is a binary variable that takes a value of 1 if the presidency was held by a Democrat in the preceding two or more years and takes the value -1 if the presidency was held by a Republican in the preceding two or more years. And $RDPIGR_t$ is the deviation from the sample mean of the growth rate of real disposable per-capita income from the third quarter of the previous year to the third quarter of the election year. This interaction term takes into account that above-average economic performance in the preceding year may be attributed to the success of policies of the presidential party and so the presidential party gains more votes. If so, we expect β_2 to be positive.

$TWO_t^+ \cdot MIDTERM_t$ is also an interaction variable, where $MIDTERM_t$ is a dummy variable that is 1 for midterm elections and zero otherwise. The variable takes into account the tendency of voters to balance the power of a newly elected or newly reelected partisan president by electing representatives from the nonpresidential party in greater numbers (Erikson (1988), Alesina and Rosenthal (1989)) and so β_3 is predicted to be negative.

Table 1 presents our estimation results.[8] The first column reports the results of the main regression. The β_1 coefficient is estimated to be negative and statistically significant at the 1 percent level. Evidently, a long presidential incumbency is costly for the party: On average, a party's vote share declines by 2.37 percentage points in elections in which the party held the presidency for 6 or more years. The β_2 coefficient is estimated to be positive and statistically significant at the 5 percent level, which accords with the common finding (see, for instance, Bartels and Zaller (2001) and Fair (2009)) that good economic performance boosts the popularity of the president's party.

The β_3 midterm coefficient has the predicted sign, but the coefficient is not estimated to be statistically significant. Given the well-known midterm cycle effect, this finding might seem surprising. Note that our regression is a level regression and, as such, the β_3 coefficient may be consistent with a larger midterm cycle effect in terms of a *change* in vote shares between presidential and midterm elections. There is more discussion of the relationship between midterm cycle effect and our theory of incumbency disadvantage of political parties in Section 7.1.

[8] In this estimation, we are following the common practice (see, for instance, Lewis-Beck and Rice (1984) and Campbell (1997) among others) of ignoring that vote-shares are bounded variables.

The second column checks for robustness with respect to the measure of economic performance by using per-capita growth in real GDP instead of per-capita growth in real disposable income. The magnitudes of the estimated coefficients remain very similar.

While our main regression specification follows the literature in ignoring the bounded nature of the vote share variable, we confirm in the third column that the relationships hold when DV is replaced with $\ln[DV/(1-DV)]$, which is an unbounded variable.

Finally, in Appendix A, we report the estimation results of our main regression using the Democratic share of House seats as the dependent variable. The coefficient on the SIX^+ variable remains significant at the 1 percent level.

Table 1:
Presidential Incumbency of a Party and Democratic Share of House Votes

DEP. VAR.	DV	DV	$\ln(DV/(1-DV)$
CONSTANT	51.75^{***} (0.66)	51.68^{***} (0.65)	0.07^{**} (0.03)
SIX^+	-2.37^{***} (0.63)	-2.49^{***} (0.63)	-0.10^{***} (0.03)
$PDPIGR * TWO^+$	0.40^{**} (0.20)		0.02^{**} (0.01)
$PGDPGR * TWO^+$		0.31^{*} (0.15)	
$MIDTERM * TWO^+$	-1.01 (0.89)	-0.93 (0.89)	-0.04 (0.04)
SD(DEP VAR)	3.22	3.22	0.13
NO. OF OBS.	32	32	32
R^2	0.51	0.49	0.51
ADJ. R^2	0.45	0.44	0.45

How robust is the finding of incumbency disadvantage to the length of incumbency? Table 2 reports estimates of β_1 as the length of incumbency of the presidential party is varied from two or more years to eight or more years. In all cases, the β_1 coefficient is estimated to be negative and it is strongly statistically significant (at the 1 percent level) when the length of incumbency is four or more years, six or more years (the main regression) and eight or more years. Furthermore, the estimated β_1 coefficient is similar in magnitude for incumbency length of six or more years and eight or more years and the adjusted R^2 is highest for the main regression. The coefficients on the two control variables (β_2 and β_3) are always estimated to have the predicted signs but their statistical significance varies.

Table 2:
Presidential Incumbency of a Party and Dem. Share of House Votes
Other Specifications

DEP. VAR.		*DV*		
CONSTANT	51.90***	51.83***	51.750***	51.79***
	(0.50)	(0.67)	(0.66)	(0.65)
TWO^+	−1.08			
	(0.69)			
$FOUR^+$		−1.30***		
		(0.42)		
SIX^+			−2.37***	
			(0.63)	
$EIGHT^+$				−2.21***
				(0.50)
$PDPIGR * TWO^+$	0.41	0.44*	0.40**	0.36
	(0.25)	(0.22)	(0.205)	(0.23)
$MIDTERM * TWO^+$	−1.10	−1.53**	−1.01	−2.07***
	(0.97)	(0.73)	(0.89)	(0.71)
SD DEP. VAR.	3.22	3.22	3.22	3.22
NO. OF OBS.	32	32	32	32
R^2	0.36	0.40	0.51	0.43
ADJ. R^2	0.29	0.34	0.45	0.37

The message we take from Tables 1 and 2 is that there is robust evidence that a party that has held the presidency for some length of time suffers a loss in its popularity.

3.2 Gubernatorial Elections

In this section we investigate whether the incumbency disadvantage effect is also operative for elections of state governors. The idea here parallels the extant investigations done for U.S. presidential elections. Within a state, the governor holds agenda-setting powers that can be used to further the policy program of his or her party. Our goal is to determine whether the popularity of gubernatorial candidates is lessened if the candidates' party has held the governorship for a long time.

For this investigation, we compiled the two-party share of the gubernatorial vote in all elections for the 50 states between 1954 and 2016 subject to data availability.[9] To use this data, we need to address two related challenges. In Congressional elections, the two parties have always been competitive in the postwar era and, so, vote shares are "stationary" meaning that they fluctuate around 50 percent. While the two parties have been competitive at the national level, it does

[9]The data on governor races are compiled from *https : //www.ourcampaigns.com/*. Elections in which the top two parties in the election outcome are not Democrats and Republicans are ignored.

not follow that the two parties were also competitive in every state. This is especially true in the early part of our sample when the Democratic party was very dominant in the Southern states. If one party is dominant, disagreement regarding policies will show up as support for a different candidate in the same party, rather than support for a different party altogether. This means that disagreement with policies enacted will not register in the two-party vote share. To try to address this, we limit consideration to those elections in which the two parties are competitive. We do this by excluding the initial years in a state until both parties have won at least one election in the preceding 20 years. From that year on, we include all gubernatorial elections in that state.[10] In addition, we ignore elections in which the top two parties in the election outcome are not Democrats and Republicans. Even with this adjustment, however, there might still be trends in the state level data (not reflected in the national data) associated with the realignment of parties at the state level. To address this, we include a quadratic time trend that is separate for each state.[11]

The main regression specification is:

$$\begin{gathered} GDV_t = \beta_1 SIX_t^+ \cdot TL_t + \beta_2 SIX_t^+ \cdot [1 - TL_t] + \beta_3 INC_t \cdot TWO_t^+ + \dots \\ \text{year } FE + \text{ state } FE + \text{state-level } QUADRATIC\ TIME\ TREND. \end{gathered}$$

Here, SIX_t^+ has the same interpretation as in the Congressional elections: It takes the value $+1$ if Democrats have held the governor's office for 6 or more years, -1 if the Republicans have held the office for 6 or more years and 0 otherwise. TL_t is a dichotomous variable that takes the value 1 if the incumbent governor faces a term limit (and so cannot run for election).[12] To take into account the well known incumbency advantage of an individual politician we also have as a regressor INC_t, which takes the value 1 if the incumbent governor is running for reelection and 0 otherwise (notice that INC_t can be 1 only when $TL_t = 0$).[13] When the term limit binds and the incumbent governor cannot run for reelection, we can evaluate the popularity or unpopularity of the incumbent governor's party by the interaction of SIX_t^+ and $TL_t = 1$. Finally, the specification includes year

[10] The starting year in the sample for the affected states that were historically heavily Democratic are: Alabama (1990), Arkansas (1968), Florida (1970), Georgia (2006), Mississippi (1995), North Carolina (1976), Oklahoma (1966), Tennessee (1974), Texas (1982) and Virginia (1973). The starting year in the sample for the affected states that were historically heavily Republican are: Minnesota (1956), New Hampshire (1964) and Vermont (1964).

[11] Adding a further cubic term does not increase adjusted R^2.

[12] Whether the incumbent governor faced a term limit was determined on a case-by-case basis from various websites.

[13] Incumbency advantage of an incumbent politician running for reelection is widely established for the U.S. (see, for instance, Erikson (1971), Gelman and King (1990), Ansolabehere and Snyder Jr. (2002), Mayhew (2008) and Jacobson (2015)).

and state fixed effects and a separate quadratic time trend for each state. In our (panel) regression, standard errors are clustered by state, which corrects for both heteroscedasticity and autocorrelation of the error terms.

Table 3 reports the estimation results. The first column of results reports the estimates from the main regression. Both SIX^+ coefficients are negative and strongly statistically significant. The second column of results reports estimates of the regression where the quadratic time trends are dropped. In this case, when a term limit binds ($TL_t = 1$), the coefficient on SIX^+ is negative and strongly statistically significant but when it does not bind (which means the incumbent may or may not run), the coefficient on SIX^+ is slightly negative and not significant. In Section 7.2, after we have introduced our model, we discuss how long-term trends affect our regression results. The coefficient on the INC term is positive and strongly statistically significant in both regressions, indicating the presence of a strong individual incumbency effect.

Table 3:
Gubernatorial Incumbency of a Party and Its Share of Gubernatorial Votes

DEP. VAR.	*GDV*	*GDV*
$SIX^+ \cdot TL_t$	-4.52^{***} (0.73)	-3.06^{***} (0.76)
$SIX^+ \cdot [1 - TL_t]$	-1.99^{***} (0.64)	-0.19 (0.57)
INC	6.31^{***} (0.56)	6.40^{***} (0.52)
YEAR FE	YES	YES
STATE FE	YES	YES
STATE QUADRATIC TIME TREND	YES	NO
SD DEP. VAR.	9.89	9.89
NO. OF OBS	782	782
R^2	0.61	0.49
ADJ. R^2	0.47	0.41

3.3 Do Policy Choices Matter in Elections?

As noted in the introduction, our explanation of incumbency disadvantage works through policy choices: The ruling party implements policies that conform to its ideal policies, which then reduces the party's support among members of the other (opposition) party. For this mechanism to be active, it is necessary that the ideal policies of the two parties be different and that voters care about these differences sufficiently for the choice of policies to matter for election outcomes.

Regarding differences in the ideal policies of the two parties, Poole and Rosenthal (1985) showed that Congressional roll-call votes of individual legislators can be explained in terms of each member having some ideal point on a single liberal-conservative line, with Democrats occupying the liberal end of the line and Republicans the conservative end. Many studies that followed in the wake of their path-breaking work have confirmed this finding. Since liberalism and conservatism lead to different policy choices, the first requirement of different policy agendas across the two parties seems to be satisfied.

Regarding the second requirement — that voters care about policy choices of the two parties sufficiently for these choices to matter for election outcomes — there is also compelling, if less extensive, evidence. The difficulty here is that Congress passes hundreds of laws each year and it is challenging to determine how all this Congressional activity affects election outcomes. However, in another oft-cited work, Erikson, Mackuen, and Stimson (2002) compiled a list of significant legislation passed by Congress each year and coded each such legislation as conservative or liberal or neither in character. Matt Grossman[14] compiled an updated version of this list that covers 1954-2014. The correlation between the net number of liberal laws passed in the preceding two years and the change in the House Democratic vote share from the previous election is -0.47.[15] At a broad level, this finding is consistent with our claim that as policies move toward the ideal policies of a party, the party's popularity in elections is diminished.

More direct evidence of loss of party popularity from enacting policies that are strongly opposed by the other party is given in recent studies that sought to understand why Democrats lost a much higher than predicted number of House seats in the 2010 midterm elections. The 111th Congress (Obama's first term) saw the Democratic party enjoying majorities in both chambers. The party pursued an ambitious policy agenda, but that agenda had no support from the Republicans. Brady, Fiorina, and Wilkins (2011, Table 2, p. 248) examined the relationship between the vote share of Democratic candidates running for reelection in 2010 and the candidates' support for the Affordable Care Act (which did not receive a single Republican vote). They find that a candidate's support for ACA reduced his or her vote share significantly and the reduction was more severe in districts in which fewer people had voted for Obama. Kroger and Lebo (2012, Figure 5, p. 942), using a more sophisticated empirical approach, report similar results for the ACA. These findings support the

[14] "Voters Like a Political Party Until It Passes Laws," *FiveThirtyEight (blog)*, October 4, 2018, https://fivethirtyeight.com/features/voters-like-a-political-party-until-it-passes-laws/

[15] Net liberal laws passed is the total number of liberal laws passed by a Congress minus the total number of conservative laws passed by the same Congress.

claim that policy choices have electoral consequences and, again, at a broad level, our claim that as policies move toward the ideal policies of a party, the party's popularity in elections is diminished.

4 Incumbency Disadvantage of Parties in a Two-Period Model

The main result of this paper — incumbency disadvantage of parties — can be illustrated and explained in a two-period setting.

Let $t = 1, 2$. The economy is populated by a continuum of individuals who derive utility from two types of public goods and discount future utility flows at the rate $\beta \in (0, 1)$. The total resources available each period to be spent on the public goods is a constant $\tau > 0$. If $0 \leq g_t \leq \tau$ is spent on the first good in period t, then $0 \leq \tau - g_t \leq 1$ is spent on the second good, i.e., the transformation between the two goods is one-for-one.

People's preferences toward the two public goods vary. In this two-period model, we assume that one-half of the population cares only for the first good and the other half cares only for the second good. We label the first group type D and the second group type R. Also, we assume that the utility flow each type gets from consuming its preferred good is $-(\tau - x_t)^2$. Then, the utility flow of type D from a public spending profile of $(g_t, \tau - g_t)$ is $-(\tau - g_t)^2$ and of type R is $-(\tau - (\tau - g_t))^2 = -g_t^2$.[16]

The different preferences toward the two public goods motivate the model's political structure. Corresponding to the two types, there are two political parties: The D party and the R party. A party's preference over the two goods overlaps with the type it represents but it isn't identical. The period 1 utility of the D party is $-(\tau - g_1)^2 + B \cdot \mathbb{1}_{D \text{ party in power}}$ and the period 1 utility of the R party is $-(g_1)^2 + B \cdot \mathbb{1}_{R \text{ party in power}}$. Here, $\mathbb{1}_{\{\cdot\}}$ is an indicator function and $B > 0$ is the utility flow party representatives obtain if their party is in power.

Period 2 preferences are similar but, in addition, include a cost of changing policies. Specifically, the period 2 preference of the D party is $-(\tau - g_2)^2 + B \cdot \mathbb{1}_{D \text{ party in power}} - \psi(g_1, g_2)$ and that of the R party is $-(g_2)^2 + B \cdot \mathbb{1}_{R \text{ party in power}} - \psi(g_1, g_2)$. Here, $\psi(g_1, g_2)$ is a nonnegative function that is zero only for $g_1 = g_2$. It captures some of the ways policy inertia arises in reality. Most directly, it captures the fact that once policies are put in place, they create their own vested interests that make

[16] In this instance, g_t could also be interpreted as the ideological stance of policies with 0 and τ representing opposite ends of the ideological spectrum.

a reversal difficult.[17] It also captures the inertia that arises from an incumbent party's ability to appoint members of the judiciary (such as Supreme and federal court justices in the U.S.) who then interpret laws in ways that favor the desired policies of that party, making subsequent reversal of policies costly and time-consuming. In this two-period model, ψ has a simple quadratic adjustment cost form, namely, $\eta(g_1 - g_2)^2$, $\eta > 0$.[18]

The final element of the model is elections. People cast their votes in favor of the party whose policies give them the highest expected lifetime utility. Following the probabilistic voting literature, it is assumed that the elected party will make some choices separate from the composition of public goods. An individual's preference toward these other choices is captured by her *net* preference for the D party. This net preference is the sum of an idiosyncratic component e and an aggregate component A. Then, given e_t and A_t, a type D individual's period utility from public goods and the identity of the ruling party is $-(\tau - g_t)^2 + \mathbb{1}_{\{D \text{ party in power}\}}[e_t + A_t]$. Analogously, a type R individual's period utility is $-(g_t)^2 + \mathbb{1}_{\{D \text{ party in power}\}}[e_t + A_t]$. We assume that the net preference shocks e_t and A_t are drawn independently each period from zero-mean distributions with CDF $F(e) : \mathbb{R} \to [0, 1]$ and $H(A) : \mathbb{R} \to [0, 1]$, respectively. Each period, the idiosyncratic component is drawn independently across individuals.

We use this model to show that the party in power in period 1 will choose a policy that implies an electoral disadvantage for itself in period 2 and to explain the key roles of policy inertia and political uncertainty for this result. We proceed by backward induction. To be consistent with the recursive notation used in the next section, we will denote the period 1 policy choice by g' and the period 2 policy choice by g''.

Period 2 post-election decision rule of parties:

If the D party wins the election in period 2, it chooses g'' to maximize $-(\tau - g'')^2 + B - \eta(g' - g'')^2$ given g' and $g'' \in [0, \tau]$. Its optimal decision rule

$$G_D(g') = \frac{1}{1+\eta}\tau + \frac{\eta}{1+\eta}g'$$

[17] For instance, between 2016 and 2018, only some portions of the Affordable Care Act could be reversed.

[18] In the infinite horizon model used in the quantitative section we explore two more.

is a convex combination of the D type's ideal policy, τ, and the inherited policy g'. Similarly, if the R party wins the election in period 2 its optimal decision rule is

$$G_R(g') = \frac{\eta}{1+\eta} g',$$

which is a convex combination of the R type's ideal policy, 0, and the inherited policy g'. The smaller is η the closer is party k's policy to type k's ideal policy. Note that B does not affect the optimal policy.

Period 2 voting decision rule of individuals:

A type k votes for party k if her *net* gain from having party k choose policies is nonnegative. The net gain to a D type from having party D choose policies is

$$-(\tau - G_D(g'))^2 + e' + A' + (\tau - G_R(g'))^2 = -\left[\frac{\eta}{1+\eta}\right]^2 (\tau - g')^2 + e' + A' + \left(\tau - \frac{\eta}{1+\eta} g'\right)^2,$$

and the net gain to an R type from having the R party choose policies is

$$-(G_R(g'))^2 + (G_D(g'))^2 - e' - A' = -\left[\frac{\eta}{1+\eta}\right]^2 g'^2 + \left(\tau - \frac{\eta}{1+\eta}(\tau - g')\right)^2 - e' - A'.$$

Let $e_k(g', A')$, $k \in \{D, R\}$, be the values of e that set the corresponding net gain terms to zero. That is,

$$e_D(g', A') = -\left[-\left[\frac{\eta}{1+\eta}\right]^2 (\tau - g')^2 + \left(\tau - \frac{\eta}{1+\eta} g'\right)^2\right] - A' \tag{1}$$

and

$$e_R(g', A') = \left[-\left[\frac{\eta}{1+\eta}\right]^2 g'^2 + \left(\tau - \frac{\eta}{1+\eta}(\tau - g')\right)^2\right] - A'. \tag{2}$$

Then, a type k person votes for the D party if her $e > e_k(g', A')$, votes for the R party if her $e < e_k(g', A')$ and votes for the k party if her $e = e_k(g', A')$.

The expressions for $e_k(g', A')$ make intuitive sense. First, both thresholds decline with an increase in A': If the D party is more popular with all individuals then more people of both types vote for the D party. Second, the term in square brackets is the net gain from having a person's own party choose policies, ignoring preference shocks. For both types, this net gain is strictly positive and, so, $e_R(g', 0) > 0$ and $e_D(g', 0) < 0$. The larger this net gain is, the more likely it is that the person will vote for her own party.

Period 2 probability of a D party win:

The D party will win the election if more than half the voters vote for it. To determine when this will be the case, consider the value of A' that solves the equation:

$$\frac{1}{2}F(e_D(g', A')) + \frac{1}{2}F(e_R(g', A')) = \frac{1}{2}. \tag{3}$$

The l.h.s. is the fraction of people who vote for the R party when the aggregate preference shock is A'.[19] If e has unbounded support, then, since both thresholds are decreasing in A', the l.h.s. is strictly decreasing in A' and asymptotes to 1 as $A' \to -\infty$ and to 0 as $A' \to \infty$. Thus a unique A' satisfying (3) exists. Denote this value by $A(g')$ and it is the value of A' for which the two parties exactly split the votes, given g'. If A' is higher (lower) than $A(g')$, the D party (R party) wins.

Equation (3) implies $F(e_D(g', A')) = 1 - F(e_R(g', A'))$. Since the distribution of e is symmetric around 0, $e_D(g', A')$ and $e_R(g', A')$ must be on opposite sides of 0 and equidistant from it, i.e., $e_D(g', A') + e_R(g', A') = 0$. Using (1) and (2) then gives

$$A(g') = \frac{2\eta\tau}{(1+\eta)^2}\left(g' - \frac{\tau}{2}\right). \tag{4}$$

When $\eta > 0$, the sign of $A(g')$ depends on the sign of $g - \tau/2$. If g is closer to the ideal choice of the D party (R party), then $A(g')$ is positive (negative). Since A is distributed symmetrically around 0, this means that the probability of the D party (R party) winning the election is *less* than 1/2 when $g' > (<)\ \tau/2$.

To understand this result, we can examine which party wins the election when A is equal to zero. If it is the R party (D party) then A' has to be strictly positive (negative) for the two parties to get equal shares of votes. When $A' = 0$, the fraction voting for the R party is given by

$$\frac{1}{2}\left[F(U_D(G_R(g')) - U_D(G_D(g'))) + F(U_R(G_R(g')) - U_R(G_D(g')))\right].$$

[19] Strictly speaking, the l.h.s. gives the probability that a randomly chosen person votes for the R party. It is well understood that there is no LLN that ensures the identification of probabilities with fractions when the population is a continuum (Judd (1985)). However, for our application it is fine to simply assume that the "law" holds (see Feldman and Gilles (1985) and Uhlig (1996)) and view that l.h.s. as the fraction of people voting for the R party.

Whether this expression is greater or less than 0.5 depends on the magnitude of the net gain for each type from having her own party choose policies (ignoring the idiosyncratic voter preference shocks e), which is

$$|U_k(G_R(g')) - U_k(G_D(g'))| \approx \left| U_k'\left(\frac{G_R(g')+G_D(g')}{2}\right)\right| \cdot |G_R(g') - G_D(g')|. \tag{5}$$

The type for which this magnitude is larger will be on the winning side as fewer of them will be swayed by the idiosyncratic shock and vote for the other party. It will be larger for the type for which $|U_k'\left(\frac{G_R(g')+G_D(g')}{2}\right)|$ is larger. Since the midpoint of the party's desired policies is $\frac{\tau}{2} + \frac{\eta}{1+\eta}\left(g' - \frac{\tau}{2}\right)$, by diminishing marginal utility $|U_D'\left(\frac{G_R(g')+G_D(g')}{2}\right)|$ will be smaller (larger) than $|U_R'\left(\frac{G_R(g')+G_D(g')}{2}\right)|$ for g' greater than (smaller than) $\tau/2$ and so the R party (D party) will win the election when $A = 0$.[20]

Note that $A(g') = 0$ for all g' if $\eta = 0$ or if $\eta = \infty$. If $\eta = 0$, there are no adjustment costs and the winning party implements its ideal policy regardless of g'. Then the midpoint of the desired policies of the two parties is $\tau/2$ and $U_D'(\tau/2)$ is equal to $U_R'(\tau/2)$ and the magnitude of the net gain from voting one's party is the same for both types. If $\eta = \infty$, adjustment costs are infinite so $G_k(g') = g'$ for all k. In this case, $U_D'(g')$ will be higher or lower than $U_R'(g')$ depending on g' but this no longer matters because $|G_D(g') - G_R(g')| = 0$.

Let $\Pi(g')$ denote the probability that D party wins the election in period 2. To get an explicit expression $\Pi(g')$ assume $A' \sim U[-\bar{A}, \bar{A}]$. Then:

$$\Pi(g') \equiv \Pr[A' \geq A(g')] = \begin{cases} 0 & \text{if } A(g') \geq \bar{A} \\ \left[\frac{\bar{A}}{2} - \frac{\eta\tau}{(1+\eta)^2}\left(g' - \frac{\tau}{2}\right)\right] & \text{if } \bar{A} > A(g') > -\bar{A}. \\ 1 & \text{if } -\bar{A} \geq A(g') \end{cases} \tag{6}$$

In what follows, we assume that $\Pi(g') \in (0,1)$ for all $g' \in [0,\tau]$, i.e., the probability of a party winning the period 2 election is strictly positive no matter what the inherited policy. We can verify

[20] Given the key role of cross-voting in explaining the form of the $A(g')$ function, it is surprising that the formula for $A(g')$ does not depend on the specific shape of the F distribution (other than the fact that it is symmetric around 0 and has unbounded support). This is because there are equal measures of the two types. If this were not the case we wouldn't be able to infer that $e_D(g', A(g')) + e_R(g', A(g')) = 0$ and the specific shape of F would matter for $A(g')$.

that this will hold if

$$\bar{A} > \frac{\eta}{(1+\eta)^2}\tau^2, \tag{7}$$

i.e., if there is sufficient uncertainty about the aggregate voter preference shock. For convenience, denote the r.h.s. of (7) as ϕ.

In summary, for $\eta \in (0, \infty)$, the D party (R party) will be at an *electoral* disadvantage (less likely to win than the R party (D party)) if $g' > (<)\ \tau/2$. This means that the D party (R party) will have an *incumbency* disadvantage if it chooses $g' > (<)\ \tau/2$ when it is in power in period 1.

Period 1 choice of policies:

Given post-election choices, the payoffs to the D party from winning and losing the election in period 2 are, respectively,

$$W_D(g') = -\left[\frac{\eta}{1+\eta}\right](\tau - g')^2 + B \text{ and } X_D(g') = -\left(\frac{1}{1+\eta}\right)\tau^2 - \left(\frac{\eta}{1+\eta}\right)(\tau - g')^2,$$

and the payoffs to the R party from winning and losing the election in period 2 are, respectively,

$$W_R(g') = -\left[\frac{\eta}{1+\eta}\right]g'^2 + B \text{ and } X_R(g') = -\left(\frac{1}{1+\eta}\right)\tau^2 - \left(\frac{\eta}{1+\eta}\right)g'^2.$$

Given these payoffs, the D party's period 1 decision problem when it is in power is

$$\max_{g' \in [0,\tau]} \quad -\left(1 + \frac{\beta\eta}{1+\eta}\right)(\tau - g')^2 - (1 - \Pi(g'))\beta\left[\left(\frac{1}{1+\eta}\right)\tau^2 + B\right] + \beta B, \tag{8}$$

and that of the R party is:

$$\max_{g' \in [0,\tau]} \quad -\left(1 + \frac{\beta\eta}{1+\eta}\right)g'^2 - \Pi(g')\beta\left[\left(\frac{1}{1+\eta}\right)\tau^2 + B\right] + \beta B. \tag{9}$$

In both problems the first term records the gain (equivalently, reduction in loss) to party k from moving g' closer to type k's preferred policy. In the second term, the term in square brackets is party's loss if it loses the election in period 2. Symmetry implies that the loss is the same for the two parties. The loss is bigger the larger B is, which is intuitive. The loss is bigger the larger τ^2 is, given $\eta > 0$. In this context, τ is the difference between the ideal policies of the two types of people and, thus, τ^2 is a measure of polarization. The more polarized the economy, the greater the loss from losing an election. The loss is decreasing in η, which is also intuitive: The larger η is, the more costly it is to change policy and, therefore, less is at stake from losing power.

Turning first to the optimal choice of the D party, it follows from (8) and (6) that the net marginal gain to the D party from increasing g' is proportional to

$$2(\tau - g')(1+\eta+\beta\eta) - \beta\left[1 + \frac{(1+\eta)B}{\tau^2}\right]\left(\frac{\phi}{\bar{A}}\right)\tau. \tag{10}$$

This net marginal gain is strictly negative at $g' = \tau$, reflecting the fact that at $g' = \tau$ the marginal loss from reducing g' is second order but the marginal gain due to an increase in the probability of reelection is first order. And it is strictly positive at $g' = 0$, provided the term in square brackets is not too large (which will be true if $B = 0$, for instance).[21] Assuming an interior optimum, the D party's choice of g' must satisfy

$$(\tau - g'_D) = \left[1 + \frac{(1+\eta)B}{\tau^2}\right]\left(\frac{\beta}{1+\eta+\beta\eta}\cdot\frac{\phi}{\bar{A}}\right)\frac{\tau}{2}. \tag{11}$$

Consider first the case where $B = 0$, i.e., party representatives do not obtain any private benefits from holding office. In this case, the term in square brackets is 1 and since the term in parentheses is strictly less than 1, $g'_D > \tau/2$. By (6), the D party's probability of reelection in period 2 will be less than one-half and the D party will be at an electoral disadvantage going into the elections in period 2. By symmetry of the model, the R party's optimal choice of g' in period 1 is less than $\tau/2$ and by (6) again, the R party will be at an electoral disadvantage going into the election in period 2.

We can summarize this in:

Proposition 1. *If $B = 0$, $A' \sim U[-\bar{A}, \bar{A}]$ and $\bar{A} > \phi$, the optimal choice of g' in period 1 by the party in power has the party facing an electoral disadvantage in the period 2 election.*

The assumption that only *parties* bear the costs of adjusting policies is important to the conclusion of Proposition 1. If the costs of changing policies also entered the payoff functions of the two types of people, then, ignoring the voter preference shocks, the net gain to type k from having

[21]When $g' = 0$, the net gain term is proportional to $2(1+\eta+\beta\eta) - \beta\left[1 + \frac{(1+\eta)B}{\tau^2}\right]\left(\frac{\phi}{\bar{A}}\right)$ and the first term is greater than 2 and two out of the three factors in the second term, β and $\phi/\bar{A}$, are less than 1.

the k party choose policies will be given by $W_k(g') - X_k(g')$, $k \in \{D, R\}$. But this difference is $[1/(1+\eta)]\tau^2$, which is independent of k and g', and so $\Pi(g') = 0.5$ for all g'. Therefore, there will be no incumbency disadvantage or advantage going into the elections in period 2. However, this property will not necessarily be true in the other models of inertia we analyze in the quantitative section.

Before closing this section, we point out the sense in which policy inertia substitutes for the lack of pre-election commitment to policies. The context for this discussion is Downs's (1957) insight that if politicians care about holding elected office (i.e., $B > 0$), policies chosen will tend toward centrist ones. But, when $\eta = 0$, the D party always chooses τ and the R party always chooses 0 regardless of the magnitude of B. As first pointed out by Alesina (1988), if neither party can commit to follow a policy that deviates from the one that the party will want to choose once it wins the election then B cannot matter for policy choice (the winning party gets B no matter what policy it chooses). Thus, the inability to commit to policies prior to the election completely eliminates the moderating influence of B on policies.

But this result changes when $\eta > 0$. As can be seen from (11), a marginal increase in B from $B = 0$ causes the D party to choose a policy that is closer to $\tau/2$.[22] Thus, even when there is no pre-election commitment to policies, the future costs of adjusting policies causes B to exert a moderating influence on the choice of policies. The parties understand that because of policy inertia, choosing policies closer to $\tau/2$ means a better chance of winning the election in period 2 and enjoying B. Thus policy inertia acts as a partial substitute for pre-election commitment to policies and reactivates Downs's logic of competitive centrism. In summary, we have:

Proposition 2. *If $\eta = 0$, the party in power chooses its ideal policy regardless of B. If $\eta > 0$, $A' \sim U[-\bar{A}, \bar{A}]$ and $\bar{A} > \phi$, a marginal increase in B from $B = 0$ makes the party in power choose policies closer to $\tau/2$.*

5 Extension to Infinite Horizon

5.1 Environment

In this section we extend the two-period model to an infinite horizon setup. The environment is the same as in the two-period model but utility functions are not restricted to be quadratic and it is not

[22] In this two-period model, it is still the case that the choice of g'' is independent of B.

assumed that types are perfectly polarized. In this form, the model resembles Alesina and Tabellini's (1990) influential model of two parties with different policy preferences circulating in power. The main differences are that election outcomes are determined endogenously via probabilistic voting and policies change inertially.

Let $\alpha \in [0, 1)$. Then, for type D people, the period utility flow from public expenditures on the two goods is

$$U(g_t, m_{1t}) + \alpha U(\tau - g_t, m_{2t}) + [e_t + A_t] \cdot \mathbb{1}_{\{D \text{ party is elected in period } t\}}.$$

Thus, as in the two-period model, type D cares more about the first good than the second. Analogously, the period utility flow of type R is

$$\alpha U(g_t, m_{1t}) + U(\tau - g_t, m_{2t}) + [e_t + A_t] \cdot \mathbb{1}_{\{D \text{ party is elected in period } t\}}.$$

Here m_{1t} and m_{2t} are preference shocks that are drawn independently each period from a continuous probability distribution with support M and $U(\cdot, m_\ell) : \mathbb{R}^+ \times M \to \mathbb{R}$ is continuous in m_ℓ, $\ell = 1, 2$. For the infinite horizon model, these shocks are needed to ensure the existence of a stationary pure strategy Markov equilibrium.[23]

The period utility of the D party is

$$U(g_t, m_{1t}) + \alpha U(\tau - g_t, m_{2t}) - \psi(g_{t-1}, g_t) + B \cdot \mathbb{1}_{\{D \text{ party in power}\}},$$

where ψ is a cost of adjusting policies. Analogously, the period utility accruing to the R party is

$$\alpha U(g_t, m_{1t}) + U(\tau - g_t, m_{2t}) - \psi(g_{t-1}, g_t) + B \cdot \mathbb{1}_{\{R \text{ party in power}\}}.$$

[23] In a finite horizon model one can always compute, via backward induction, the equilibrium decision rules $G_k(g, n)$ — where n is the number of periods to the terminal period — for *any* $n \in \mathbb{N}$ (in the two period model, for instance, $n = 2$). The problem is that $|G_k(g, n) - G_k(g, n + 1)|$ may fail to converge to 0 for all g and k as $n \to \infty$. This problem arises whenever there is time inconsistency, which leads to nonconcavity of continuation value functions. For large n, changes in continuation values can be small as n is incremented, but, even small changes can induce jumps in $G_k(g, n + 1)$ because of the nonconvexities in the decision problem of parties. As a result, neither decision rules nor continuation value functions settle down to *stationary* functions as n is increased. The introduction of the continuously distributed i.i.d. shock, combined with discretization of the state (and action) spaces, ensures the existence of a stationary pure strategy MPE and serves to ameliorate convergence issues. We note in passing that discontinuous decision rules also plague models of industry dynamics (Doraszelski and Satterthwaite (2010), legislative bargaining (Duggan and Kalandrakis (2012)) and sovereign debt and default (Chatterjee and Eyigungor (2012)). In all cases, the existence problem (and convergence issues) is solved by the introducing i.i.d. shocks in the right places. The shock plays a role similar to the additive payoff perturbations in Harsanyi's (1973) purification of mixed strategies.

The timeline of events in a period is as follows. At the start of the period, e shocks for all individuals and the aggregate shock A are realized. Then, the election is held and each individual votes for the party that gives her the highest expected lifetime utility. Following the election, the preference shock of the winning party is realized and the winning party makes its policy choice. Finally, the consumption of the public goods takes place and the period ends.

5.2 Recursive Formulation

Let $g \in [0, \tau]$ denote the inherited policy in the current period and $g' \in [0, \tau]$ denote the policy choice made in the current period. Let $\Pi(g) : [0, \tau] \to [0, 1]$ denote the function that gives the probability of the D party winning the election if the inherited policy is g.

For computational tractability, the preference shock structure will be simplified in one respect: In any period, only the preference shock to the preferred good of the party that wins the election is active. That is, if the D party wins the election, m_2 is set to 0 and, symmetrically, if the R party wins the election, m_1 is set to zero.[24]

For $k \in \{D, R\}$, let V_k denote the value of party k when it is in power and let X_k be its value when it is not in power. Then,

$$V_D(g, m_1) = \max_{g' \in [0,\tau]} U(g', m_1) + \alpha U(\tau - g') - \psi(g, g') + B + \beta \left[\begin{array}{c} \Pi(g')\mathbb{E}_{m_1'} V_D(g', m_1') + \\ \left[1 - \Pi\left(g'\right)\right] \mathbb{E}_{m_2'} X_D(g', m_2') \end{array} \right]. \tag{12}$$

When the D party is in power, it chooses g' taking into account the preference shock m_1 and the costs of changing policies $\psi(g', g)$. The party recognizes that its choice of g' will affect its probability of reelection next period via the function $\Pi(g')$. Let $G_D(g', m_1)$ denote a policy function that attains $V_D(g, m_1)$. When the D party is not in power, it lives with the choices made by the R party. Let $G_R(g, m_2)$ denote the policy function of party R. Then,

$$X_D(g, m_2) = U(g') + \alpha U(\tau - g', m_2) - \psi(g, g') + \beta \left[\begin{array}{c} \Pi(g')\mathbb{E}_{m_1'} V_D(g', m_1') + \\ +[1 - \Pi(g')]\mathbb{E}_{m_2'} X_D(g', m_2') \end{array} \right] \tag{13}$$

$$\text{s.t. } g' = G_R\left(g, m_2\right).$$

[24] The computations require that there be sufficient randomness in the choice of g' conditional on g, and having only one of the two shocks being active is enough for this purpose. Note that if we had only m_1 or only m_2 active every period, the parties would no longer be symmetric.

Symmetrically,

$$V_R(g, m_2) = \max_{g' \in [0,\tau]} \alpha U(g') + U(\tau - g', m_2) - \psi(g', g) + B + \beta \left[\begin{array}{c} \Pi(g') \, \mathbb{E}_{m_1'} X_R(g', m_1') + \\ [1 - \Pi(g')] \, \mathbb{E}_{m_2'} V_R(g', m_2') \end{array} \right], \tag{14}$$

and

$$X_R(g, m_1) = \alpha U(g', m_1) + U(\tau - g') - \psi(g', g) + \beta \left[\begin{array}{c} \Pi(g') \, \mathbb{E}_{m_2'} X_R(g', m_2') + \\ [1 - \Pi(g')] \, \mathbb{E}_{m_1'} V_R(g', m_1') \end{array} \right] \tag{15}$$

$$\text{s.t. } g' = G_D(g, m_1).$$

Next, we turn to the value functions of people. Ignoring her current preference shocks, let W_k and Z_k be the value of type k person when her party is in power and out of power, respectively. Then,

$$W_D(g, m_1) = U(g', m_1) + \alpha U(\tau - g') + \beta \left[\begin{array}{c} \Pi(g') \, \mathbb{E}_{m_1'} W_D(g', m_1') + \mathbb{E}_{A'} A' | (D \text{ party win}) + \\ [1 - \Pi(g')] \, \mathbb{E}_{m_2'} Z_D(g', m_2') \end{array} \right] \tag{16}$$

$$\text{s.t. } g' = G_D(g, m_1),$$

$$Z_D(g, m_2) = U(g') + \alpha U(\tau - g', m_2) + \beta \left[\begin{array}{c} \Pi(g') \, \mathbb{E}_{m_1'} W_D(g', m_1') + \mathbb{E}_{A'} A' | (D \text{ party win}) + \\ [1 - \Pi(g')] \, \mathbb{E}_{m_2'} Z_D(g', m_2') \end{array} \right] \tag{17}$$

$$\text{s.t. } g' = G_R(g, m_2).$$

And, symmetrically,

$$W_R(g,m_2) = \alpha U(g') + U(\tau - g', m_2) + \beta \left[\begin{array}{c} \Pi(g')\, \mathbb{E}_{m_1'} Z_R(g', m_1') + \mathbb{E}_{A'} A' | (D \text{ party win}) + \\ \left[1 - \Pi(g')\right] \mathbb{E}_{m_2'} W_R(g', m_2') \end{array} \right] \tag{18}$$

s.t. $g' = G_R(g, m_2)$,

$$Z_R(g,m_1) = \alpha U(g', m_1) + U(\tau - g') + \beta \left[\begin{array}{c} \Pi(g')\, \mathbb{E}_{m_1'} Z_R(g', m_1') + \mathbb{E}_{A'} A' | (D \text{ party win}) + \\ \left[1 - \Pi(g')\right] \mathbb{E}_{m_2'} W_R(g', m_2') \end{array} \right] \tag{19}$$

s.t. $g' = G_D(g, m_1)$.

While W_k and Z_k are independent of A and e by definition, they are dependent on *future* values of A' through the term $\mathbb{E}_{A'}|(D$ party win). This term recognizes that the election of the D party next period is not independent of the realized value of A'. In particular – as we showed in the two period model – it will be the case that the D party will win only if A' is above some threshold value that depends on g'. Thus, conditional on a D party win, the expectation of A' is generally non-zero.[25]

Since the value of m_1 or m_2 is realized after the election, the individual net gain to a type D person from voting for the D party is $\mathbb{E}_{m_1} W_D(g, m_1) + e + A - \mathbb{E}_{m_2} Z_D(g, m_2)$ and the individual net gain for a type R person from voting for the D party is $\mathbb{E}_{m_1} Z_R(g, m_1) + e + A - \mathbb{E}_{m_2} W_R(g, m_2)$. Given the pair (g, A), these expressions determine thresholds for the idiosyncratic shock, $e_k(g, A)$, $k \in \{D, R\}$, above which a k type will vote for the D party in the election. Specifically,

$$e_D(g, A) = -[\mathbb{E}_{m_1} W_D(g, m_1) - \mathbb{E}_{m_2} Z_D(g, m_2)] - A \tag{20}$$

and

$$e_R(g, A) = [\mathbb{E}_{m_2} W_R(g, m_2) - \mathbb{E}_{m_1} Z_R(g, m_1)] - A. \tag{21}$$

[25] There is no corresponding term for e' because the realization of an individual's e' has no consequence for whether or not the D party wins the election next period. Hence the expectation of e' conditional on D party win is its unconditional expectation, which is zero.

In these threshold expressions, the terms in square brackets represent the expected net gain to a person from their own party coming into power, ignoring the shocks e and A. The larger these terms, the more likely it is that an individual will vote for her own party. In contrast, an increase in A raises the likelihood of both types voting for the D party.

As in the two period case, the threshold value of A above which the D party wins the election satisfies that condition $e_D(g, A) + e_R(g, A) = 0$. Using (20) and (21), this threshold value is determined as:

$$A(g) = \frac{1}{2}\left\{[\mathbb{E}_{m_2} W_R(g, m_2) - \mathbb{E}_{m_1} Z_R(g, m_1)] - [\mathbb{E}_{m_1} W_D(g, m_1) - \mathbb{E}_{m_2} Z_D(g, m_2)]\right\}. \qquad (22)$$

The equilibrium of the model is defined as follows:

Definition 1. *A pure strategy Markov Perfect Equilibrium (MPE) is a collection of party value and policy functions V_k^*, X_k^*, G_k^*, a collection of voter value functions W_k^*, Z_k^*, and a pair of functions $\Pi^*(g)$ and $A^*(g)$ such that:*

- *Given $G_R^*(g, m_2)$ and $\Pi^*(g)$, the functions $V_D^*(g, m_1)$ and $X_D^*(g, m_2)$ solve* (12) - (13) *and $G_D^*(g, m_1)$ attains $V_D^*(g, m_1)$*
- *Given $G_D^*(g, m_1)$ and $\Pi^*(g)$, the functions $V_R^*(g, m_2)$ and $X_R^*(g, m_1)$ solve* (14) - (15) *and $G_R^*(g, m_2)$ attains $V_R^*(g, m_2)$*
- *Given $G_D^*(g, m_1)$, $G_R^*(g, m_2)$, $\Pi^*(g)$ and $A^*(g)$ the functions $W_D^*(g, m_1)$ and $Z_D^*(g, m_2)$ solve* (16) - (17)
- *Given $G_D^*(g, m_1)$, $G{*}_R(g, m_2)$, $\Pi^*(g)$ and $A^*(g)$, the functions $W_R^*(g, m_2)$ and $Z_R^*(g, m_1)$ solve* (18) - (19)
- *Given $W_D^*(g, m_1)$, $Z_D^*(g, m_2)$, $W_R^*(g, m_2)$, and $Z_R^*(g, m_1)$, the function $A^*(g)$ solves* (22) *and the function $\Pi^*(g) = Pr[A \geq A^*(g)]$*

For a model to be suitable for computation, it is important that there be easily verifiable conditions on model primitives for which the existence of an equilibrium is assured. If this is the case, and the conditions hold, one can be certain that failure of an algorithm to find an equilibrium is a failure of the algorithm and not the result of a lack of internal consistency. The following theorem provides such conditions, which cover not only the model described so far but also variants discussed later in the paper.

Theorem 1. *Under Assumptions 1 - 6 stated in Appendix B, a pure strategy MPE exists.*

Proof. See Appendix B.

For the model described thus far, the key requirements are that g and g' belong to a finite set (i.e., the interval $[0, \tau]$ be replaced by a discrete approximation), m_ℓ have compact support, the CDFs for m_ℓ, A and e be continuous, and for any $x \neq \hat{x}$, the difference $U(x, m_\ell) - U(\hat{x}, m_\ell)$, $\ell = \{1, 2\}$, be strictly increasing or strictly decreasing in m_ℓ.

Before moving on to the quantitative analysis, we confirm that without any adjustment costs or other constraints on changing policies, both parties choose their statically ideal policies regardless of the value of B and the likelihood of either party winning an election is exactly one-half.

Theorem 2 (Alesina 1988)**.** *In any MPE equilibrium, if $\psi(g', g) \equiv 0$, parties choose their statically ideal policies regardless of the value of B and the probability of reelection of any party is one-half.*

Proof. See Appendix C

6 Incumbency Disadvantage in a Quantitative Dynamic Model

In this section our goal is to show not only that incumbency disadvantage can arise in the dynamic model but also that the model is capable of delivering the observed *magnitude* of this effect. The demonstration proceeds by choosing a realistic parameterization of the model and examining the relevant properties of the computed MPE. As part of our goal to understand why incumbents act the way they do, we highlight the key roles played by policy inertia and uncertainty in election outcomes in generating incumbency disadvantage.

To proceed with the quantitative analysis, we adopt some parametric assumptions. We assume that $U(x, m_\ell) = (x + m_\ell)^{1-\gamma}/[1 - \gamma]$, $\ell = 1, 2$. For the base model, we assume that $\psi(g, g') = \eta(g - g')^2$, $\eta > 0$. The idiosyncratic shock $e \sim \mathrm{N}(0, \sigma_e^2)$. The aggregate shock $A \sim \mathrm{U}([-\bar{A}, \bar{A}])$ and the distributions of the party preference shocks m_ℓ, $\ell = 1, 2$, are both $\mathrm{U}([-\bar{m}, \bar{m}])$. Note that both A and the m_ℓ distributions are symmetric around 0.

Turning to parameter values, since national elections happen every two years, the value of β is set to 0.92, which corresponds to a biennial discount rate of 8 percent. The value of γ is set to 2. The value of τ is normalized to 1.[26] Since we don't observe large shifts in expenditure patterns when parties controlling the presidency change, α is set conservatively to 0.90 — this implies that voters' static optimum is to have their party spend 51.3 percent of the total budget on their preferred good.

The remaining parameters, namely, η and the dispersions of the distributions of A, e and m_ℓ, have important effects on the magnitude of the incumbency disadvantage in the model. Of these, the dispersion of m_ℓ is special in that having a large enough dispersion helps us compute an equilibrium.

Conditional on a dispersion of m_ℓ and the parameters listed earlier, experimentation showed that the model can deliver the observed magnitude of the incumbency disadvantage for a *range* of values of the remaining three parameters. Among the many constellations of parameter values we could pick, we chose the one in which η is 4.8, the standard deviation of e is 0.02, and the support of (the uniformly distributed) A is ± 0.01.

These parameter choices are summarized in Table 4.

Table 4: Parameter Selections

Parameter	Description	Value
γ	Curvature of utility function	2.00
β	Biennial discount factor	0.92
τ	Total government exp.	1.00
α	Weight given to other party's desired public good	0.90
$\overline{m}$	Support of party preference shock, $\pm\overline{m}$	0.01
η	Adjustment cost parameter	4.80
σ_e	S.D. of idiosyncratic voter preference shock	0.02
$\overline{A}$	Support of aggregate voter preference shock, $\pm\overline{A}$	0.01

To confirm that the model generates incumbency disadvantage, Table 5 reports the results of regressions run on model-generated data that mimic the regressions reported in the top panel of Table 2. The dependent variable is the D party's vote share and the explanatory variables are trichotomous variables that take on the values $+1$, -1, or 0 depending on whether the D party has been in power for 3 or more (or 4 or more) model periods, the R party has been in power for

[26] But we restrict the feasible set of g' to be $(\bar{m}, 1-\bar{m})$. Then $g' + m_1$ and $\tau - g' + m_2$ are strictly positive for any choice of g' and any realization of m_ℓ.

3 or more (or 4 or more) model periods, or neither. For comparison purposes, we also record the outcome of the same regression with the probability of a D victory as the dependent variable.

Table 5: Incumbency Disadvantage

Dep. Var.	$\%D$	$\%D$	Model Prob. of D Win	Prob. of D Win
Constant	50.00	50.00	0.50	0.50
SIX^+	-2.42	-	-0.07	-
$EIGHT^+$	-	-2.44	-	-0.07

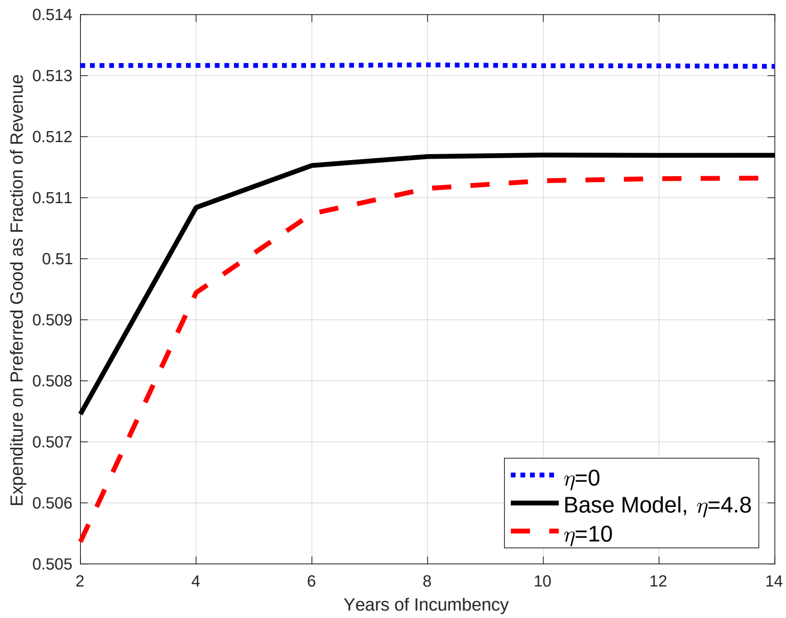
The first two columns in Table 5 report the magnitude of the incumbency disadvantage in the model for both six or more years and eight or more years of incumbency (in the model, these correspond to 3 or more periods or 4 or more periods). For either measure of incumbency, the incumbency disadvantage is 2.4 percentage points. The next two columns report the incumbency disadvantage in terms of the decline in the probability of reelection. For either measure of incumbency, the decline in the reelection probability is 7 percentage points. We also confirmed that adjustment costs remain central to incumbency disadvantage in the full dynamic setting: If we set $\eta = 0$ and run the same regressions on the model output as in Table 5, the coefficients on the incumbency variables are all estimated to be zero.

6.1 Role of Policy Inertia

A positive η introduces inertia in policies and creates a dynamic link between periods. A consequence is that a party's long-run ideal policy, namely, the average composition of spending toward which it tends as its incumbency lengthens, deviates from its no-inertia ideal policy, i.e., its average policy when $\eta = 0$.

Figure 1 charts, for different values of η, the relationship between the average expenditure (over a long simulation) on the preferred good of the incumbent party against the party's years of incumbency. The blue dotted line is the long-run ideal policy, corresponding to the $\eta = 0$ case. As the line shows, the party immediately goes to its long-run ideal and the line is flat at 0.513. The solid black line immediately below corresponds to our base model with $\eta = 4.8$. For the base model, the average expenditure is initially below its long-run level of 0.512 but reaches that level by the sixth year of incumbency and stays flat thereafter. As η is increased, the long-run expenditure level shifts down and the years of incumbency needed to converge to the ideal level lengthen. As seen in the shape of the red dashed line ($\eta = 10$), policies start out closer to 0.50 and continue to move up even beyond the eighth year of incumbency.

Figure 1:
Incumbency and Average Expenditure on the Preferred Good

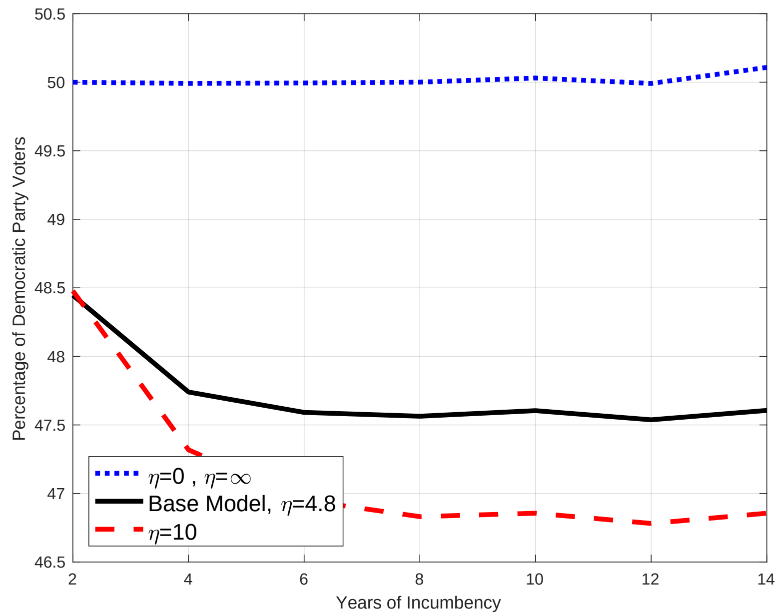


These expenditure dynamics have implications for the time path of incumbency disadvantage. Figure 2 plots the average percentage of voters who cast their ballots in favor of the incumbent. There is no incumbency disadvantage when $\eta = 0$ and the dotted blue line is flat at 50 percent. For the base model, the share falls by about 1.5 percentage points at the end of two years of incumbency and the disadvantage continues to mount until the share stabilizes at around 47.5 percent by the sixth year of incumbency. Incumbency disadvantage increases with η, as shown in the red dashed line corresponding to $\eta = 10$. But if η rises enough, the incumbency disadvantage weakens and eventually disappears when η is very large. The nonmonotonic relationship between the strength of incumbency disadvantage and η follows the same logic as in the two-period model: When the cost of adjustment is high enough, neither party can change policies much and, consequently, incumbency disadvantage weakens.

6.2 Alternative Models of Inertia and Policy Extremism

In this section we show that incumbency disadvantage also occurs for other models of inertia that may seem plausible. We examine two alternative models. In the first the quadratic costs of

Figure 2:
Incumbency and Average Lead in Elections



adjustment are borne only by the party in power; in the second, there are no costs of adjustment but there is an upper bound Δ on how much policies can change in either direction in any period (we call this the constraint-on-change model). In these experiments, all parameters unrelated to η or Δ are the same as in the base model. We pin down the new adjustment cost parameters – the one-sided η and Δ — to match the same SIX^+ coefficient as in the base model. The value of η is now 2.65 and the value of Δ is 0.033.

The first point is that the pattern of incumbency disadvantage documented in Tables 1- 3 can be accounted for by any of these alternative models of inertia. Figure 3 plots the average vote share in elections for the two alternative models along with the base model. The predicted relationships are virtually identical.

Second, the alternative models imply quite different long-run ideal policies of each party. As shown in Figure 4, the long-run ideal policies are *more* extreme than the static ideal policy of their party members. Because of inertia, the incumbent party pushes beyond the static ideal of its party members to assure them policies closer to their static ideal even when the party is out of power.

Interestingly, extremism does not arise in the base model because the costs of changing policies are borne by *all* representatives in government. Since an incumbent party anticipates the swing

Figure 3: Alternative Models of Inertia
Incumbency and Average Lead in Elections

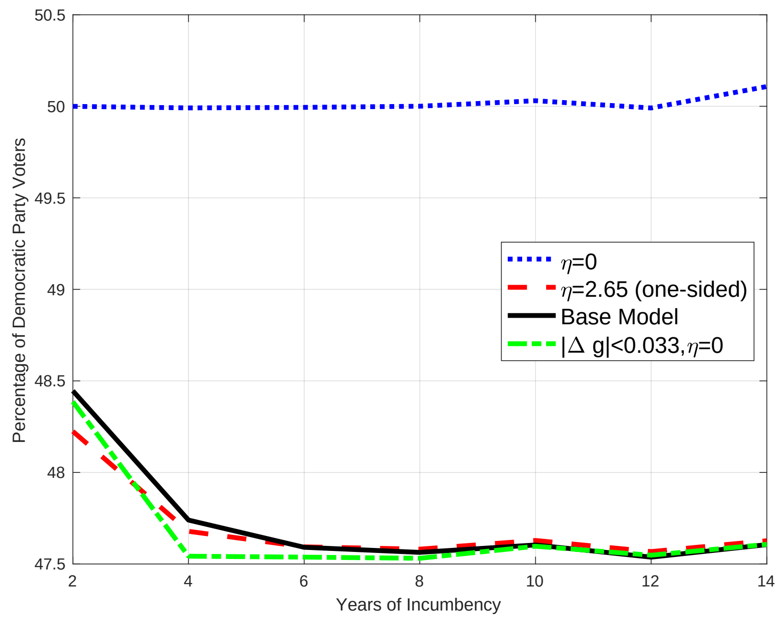
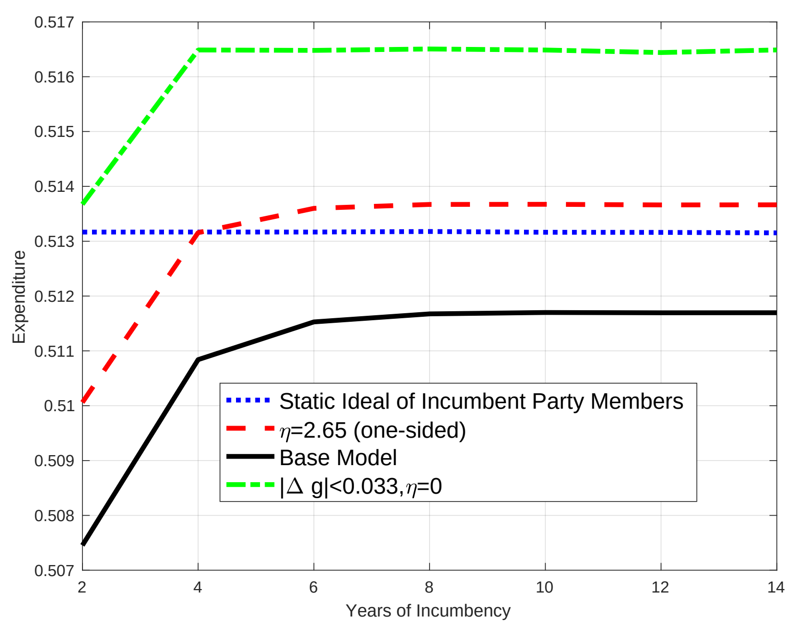


Figure 4: Alternative Models of Inertia
Incumbency and Expenditure on the Preferred Good



back in policy in future periods — and the costs associated with that reversal — it becomes more circumspect about pushing for policies that depart too far from the ideal policies of the other party.

These differences in expenditure patterns across the different models of inertia also mean that the implied welfare of voters differs across the models. From an ex-ante perspective, a more volatile expenditure pattern between the two public goods is costly because of diminishing marginal utility (the existing literature has noted that such cycles can be welfare reducing). Table 6 reports the identical lifetime utility of both types of voters when the inherited policy is $g = 0.50$ and so there is a 50 percent chance of either party being elected. Welfare is highest for the base model where the standard deviation of g is lowest, and welfare is lowest for the constraint-on-change model for which the standard deviation of g is the highest.

Table 6:
Welfare Implications of Policy Inertia

Models	Welfare Loss in % relative to Base Model	Std. Dev of g'
Base model	-	0.0095
$\eta = 2.65$ (one-sided)	0.02	0.0119
$\lvert\Delta g\rvert < 0.033$	0.07	0.0158

6.3 Policy Inertia and Competitive Centrism

If there are no costs of adjusting policies, then, as shown in Theorem 2, the private gain from holding office has no effect on either policies or the likelihood of political turnover. This result changes when there are costs of adjusting policies. As shown in Proposition 2 of the two period model, with inertia, the prospect of losing B motivates the party toward centrism.

We confirm that the same effect is present in the infinite horizon model. Figure 5 plots the average equilibrium value of g' against the number of years of incumbency for the base model ($B = 0$) and the base model with office motivation ($B = 0.04$). While the average g' rises for both model economies, the rise in g' for $B > 0$ model is more muted. Figure 6 plots the average lead of the party in power as incumbency progresses. The average lead falls in both cases, but the decline is less pronounced for the model with $B > 0$.

Interestingly, although office motivation creates a conflict of interest between a party and its adherents, the welfare implications of office motivation are *positive*. This is because office motivation

Figure 5:
Office Motivation, Incumbency and Average Expenditure on the Preferred Good

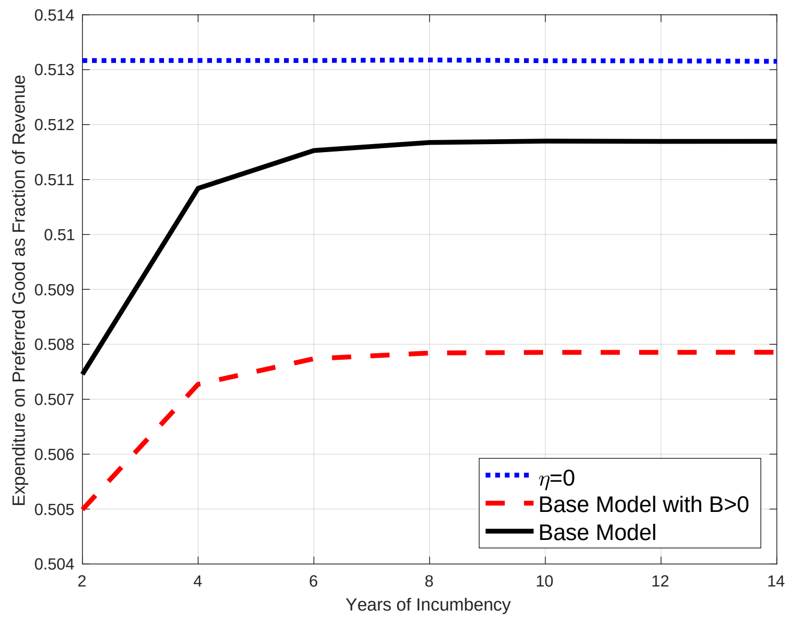
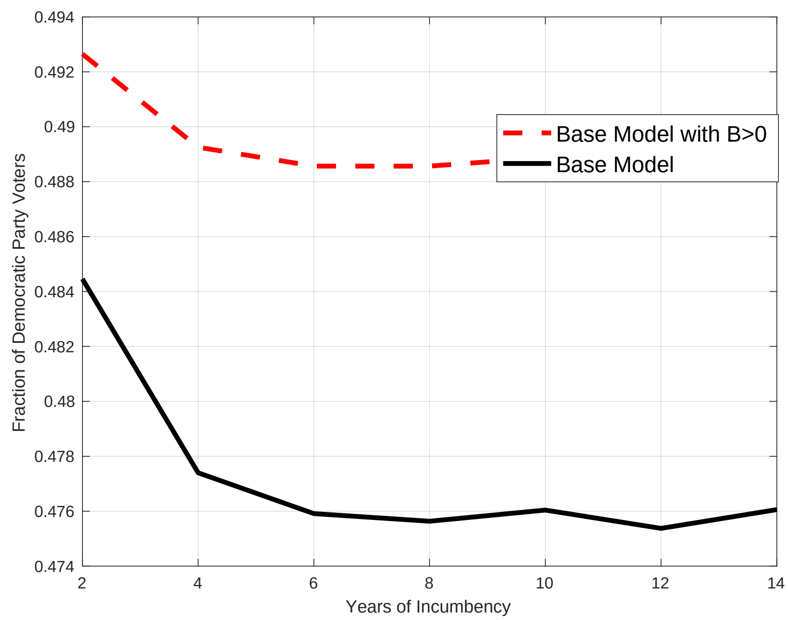


Figure 6:
Office Motivation, Incumbency and Average Lead in Elections



reduces the amplitude of the policy cycles discussed in the previous subsection. Table 7 reports the welfare effects of office motivation for the base model as well as the model with one-sided inertia and the constraint-on-change model. As before, the welfare measure is the (identical) lifetime utility of the two types, conditional on $g = 0.50$. In all three models, welfare is higher and the standard deviation of g is lower with office motivation than without.

Table 7: Welfare and Office Motivation

Models	Relative Welfare Gain Relative in %	Std. Dev of g'
Base Model	–	0.0095
Base Model with $B = 0.04$	0.02	0.0065
One-Sided Inertia Model	–	0.0119
One-Sided Inertia Model with $B = 0.04$	0.03	0.0086
Constraint-on-Change Model	–	0.0158
Constraint-on-Change Model with $B = 0.04$	0.03	0.0134

6.4 Instability, Asymmetry, and Party Behavior in Low-Noise Environments

This section addresses two questions. First, although Theorem 1 guarantees that an equilibrium exists for any $\bar{m} > 0$, our algorithm often fails to converge if $\bar{m}$ is low (i.e., the variance of m_ℓ is positive but low). Why is this? Second, how do parties behave if the election outcome conditional on inherited policy is fully predictable (i.e., if the variance of A' is negligible)?

To answer the first question, we began by looking for symmetric equilibria only. This can be done by restricting the value functions of the two parties to be mirror images of each other. With this restriction, an equilibrium could be found for low values of $\bar{m}$. But these low $\bar{m}$ equilibria are *unstable* in the following sense: If our iterative solution method is started off with equilibrium values functions found by *imposing* symmetry, the iterations diverge away from the symmetric equilibrium. This explains why these equilibria could not be found by our method: Such equilibria exist but they are not a stable solution of our iterative solution method.[27] The instability results from these equilibria being knife-edge cases: Small departures from the equilibrium value functions and the equilibrium $A'(g)$ function alter behavior in ways that accentuate those departures. Since

[27] Our algorithm solves for the value functions of the two parties and the $A(g')$ function via iterations of the following sort: If Ω^n is a stacked vector containing the value functions of the two parties and $A(g')$ function at the end of the nth iteration, then $\Omega^{n+1} = \xi\Omega^n + (1-\xi)T(\Omega^n)$, $\xi \in (0,1)$, where T is the mapping whose fixed points correspond to the value and $A(g')$ functions of an MPE. Rewritten as $\Omega^{n+1} = \Omega^n + [1-\xi] \cdot I \cdot [T(\Omega^n) - \Omega^n]$, where I is a conformable identity matrix, it is an instance of the *parallel chord method* of solving nonlinear equations (Ortega and Rheinboldt (1970, p.181)).

we are uncertain whether knife-edge equilibria are empirically relevant, we focus on the parameter space where equilibria could be found (relatively) easily with our solution method (which does not impose symmetry). This necessitated keeping the $\bar{m}$ values sufficiently large.

To answer the second question, we investigated the equilibrium behavior of parties when $\bar{A} = 0.0001$, i.e., when there is virtually no uncertainty in A'. We sought equilibria for a range of $\bar{m}$ values. Although our method typically fails to converge for low values of $\bar{m}$, with some search we obtained convergence for $\bar{m} = 0.001$ and the convergence was to an *asymmetric* equilibrium.[28] In this equilibrium one out of the two parties stays in power forever. Focusing on the equilibrium in which the D party is in power forever, we found

$$\Pi^*(g') = \begin{cases} 1 & \text{if } g' < 0.5064 \\ 0 & \text{if } g' \geq 0.5064. \end{cases}$$

The form of $\Pi^*(g')$ reflects the asymmetric nature of the equilibrium. In any symmetric equilibrium, $\Pi^*(g' = 0.5)$ is always exactly 1/2, while in this equilibrium it is 1. The reason the D party is more attractive to voters at $g = 0.5$ is that the D party understands that if it chose $g' > 0.5064$, its probability of reelection will fall drastically (this is a feature of the volatility of A being very low). So it sticks with a moderate policy that is at or below 0.5064. In contrast, if the R party were to come into power it faces no such disciplining device: The R party desires a policy that is less than 0.5 (which is less than 0.5064) and, for any such choice, the D party will win the election for sure. Given that it cannot increase its reelection probability by being more disciplined, it follows extreme policies. This behavior is anticipated by voters and, so, in the aggregate, they prefer the D party to the R party.

For values of $\bar{m}$ above 0.001, we got convergence for $\bar{m} = 0.04$. In this case, the convergence was to a symmetric equilibrium in which power changed infrequently: The incumbent party picks a policy that is *closer* to the other party's ideal policy most of the time until a realization of m_ℓ occurs that is low enough (and the marginal utility of its preferred good is high enough) that it pays the party to move policy aggressively toward its ideal policy and then lose power with

[28] If we restrict our algorithm to search only for symmetric equilibrium then a symmetric equilibrium can also be found for this case. So, for this case, we have a clear example of multiple equilibria. We know there exists at least two asymmetric equilibria (with the roles of the two parties reversed) and at least one unstable symmetric equilibrium. We suspect that if the symmetric equilibrium is unstable, then asymmetric equilibria generally exist. However, we also suspect our algorithm is not well-adapted to find asymmetric equilibria.

certainty. Correspondingly, the probability of reelection of the incumbent party is roughly constant and greater than one-half.

The common feature of both these low $\bar{A}$ equilibria is that they do not display the drop in reelection probability with incumbency that is characteristic of the base model.[29] When election outcomes are (virtually) predictable, parties choose policies to *maintain* their probability of reelection. When elections outcomes are substantially unpredictable, the incumbent party understands that it can lose power for random reasons and, consequently, becomes willing to move policies toward its own ideals and accept the electoral disadvantage this entails.

7 Other (Potential) Explanations of Incumbency Disadvantage of Political Parties

7.1 Midterm Balancing Theory

Our model has some of the flavor of the balancing theory of the midterm cycle (Alesina and Rosenthal (1989)). In the midterm balance theory, middle-of-the-road voters act preemptively to ward off extreme policy outcomes and so vote against the president's party in the midterm following a presidential election. In our model, policies favoring the presidential party get implemented and are persistent, which then leads to a backlash not only in the following midterm but also in subsequent elections. It is for this reason that we focus on longer-term effects of incumbency on popularity.

If the midterm cycle is happening only because of a balancing motive among voters (and not because of any persistence in policies implemented in the first two years of the presidency) it might still have the persistent effects we find on congressional vote shares because representatives enjoy *individual* incumbency advantage (once elected, they tend to be reelected). It might be important to know then if the significance of the SIX^{+} variable survives the addition of some measure of the past composition of Congress. Note, however, that the inclusion of any measure of past composition will make the SIX^{+} coefficient harder to interpret from the perpective of our model: For example, once we control for past unpopularity, our SIX^{+} variable might measure how much *more* unpopular the presidential party is relative to it being in power for four years and more.

[29] We note in passing that if we hold the value of $\bar{m}$ at 0.04 but raise the value of $\bar{A}$ to the base value of 0.01, the equilibrium found is symmetric and displays the downward sloping relationship between reelection probability as in the base model.

Keeping this in mind, Table 8 reports the results of including the number of Republicans who won in the *previous* election as a regressor.[30] Notice that all the variables have the expected signs and the SIX^+ variable continues to be significant and is smaller in magnitude.

Table 8:
Presidential Incumbency of a Party and Democratic Share of House Votes

DEP. VAR.	DV
CONSTANT	60.76*** (2.03)
SIX^+	−1.29** (0.51)
$PDPIGR * TWO^+$	0.47*** (0.16)
$MIDTERM * TWO^+$	−1.98*** (0.70)
$REP\_WON(-1)$	−0.05*** (0.01)
SD(DEP VAR)	3.22
NO. OF OBS.	32
R^2	0.65
ADJ. R^2	0.60

7.2 Gradually Mean-Reverting Preference Shocks

Can incumbency disadvantage be explained by popularity shocks that mean revert gradually over time? We show this is not the case for our model: Absent policy inertia, aggregate preference shocks that decay gradually over time imply incumbency advantage: Model regressions yield a *positive* (not negative) coefficient on the SIX^+ variable.

We modify the base model to have the aggregate preference shock be $A + Z$, where A is a continuously distributed uniform variate (as before) and Z a discretized $AR1$ process with innovations distributed $N(0, 0.005^2)$. We set $\eta = 0$ and experiment with two different serial correlation parameters, 0.40 (low) and 0.70 (high). Recall that absent inertia and serial correlation in popularity ($\eta = 0$ and $Z \equiv 0$) there is no incumbency disadvantage and the SIX^+ coefficient in the model regression is 0.

As shown in Table 9 the SIX^+ coefficient in the model regression is strictly positive. If the D party wins the election as a result of a positive Z shock, the Z shock is expected to be positive in the future as well. As such, conditional on being elected today, a party is likely to get more than 50 percent of the votes in the future. It might appear surprising that the coefficient on the

[30] With no indepedendent candidates in Congress, the number of Democrats who won would be 435 minus this number.

Table 9:
Serially Correlated Popularity Shocks and No Policy Inertia

	Model			
Dep. Var.	$\%D$	$\%D$	$\%D$	$\%D$
	$\rho = 0.40$		$\rho = 0.70$	
Constant	50.00	50.00	0.50	50.00
TWO^+	1.30	-	3.69	-
SIX^+	-	1.70	-	5.34

SIX^+ variable is higher than the coefficient on TWO^+ since, with an AR1 coefficient below 1, the popularity shock is, on average, decaying over time. The reason is a selection effect: The incumbencies that last for six years are, on average, the ones that had received a larger preference shock at its inception.

The effect of persistence in popularity shocks sheds light on the role of time trends in the gubernatorial election regressions presented section 3.2 (Table 3). As noted in that section, many Southern states have moved from being aligned with the Democratic party to being aligned with the Republican party (and, correspondingly, some Northern states have moved from being Republican to being Democratic). These persistent shifts in party popularity would, on their own, lead to a positive SIX^+ coefficient and thus would tend to mask any incumbency disadvantage effects arising from shorter-term policy changes. The inclusion of state-level quadratic time trends helps to absorb effects of persistent preference shifts and strengthens the statistical significance of SIX^+ coefficients.[31]

It's worth noting that for the purposes of this paper the effects of persistence cannot be dealt with by having the *change* in vote shares between elections be the dependent variable. This is because the estimate of the SIX^+ coefficient would be negative even if there is no inertia or serial correlation in popularity shock. This is confirmed in Table 10, which reports the results of a regression of the change in vote shares on the TWO^+ or SIX^+ variables, absent inertia and serial correlation in the popularity shock ($\eta = 0$ and $Z \equiv 0$). The coefficient on both TWO^+ and SIX^+ variables are now negative, despite there being no incumbency disadvantage. A party that gets a positive popularity shock and wins the election gets, on average, 50 percent of the votes in the next election. Thus, on average, the *change* in the vote share of the incumbent party is negative.

[31] This is because the national elections data do not show strong time trends and the inclusion of time trend variables in the Congressional elections regressions leaves the coefficient on the SIX^+ variable almost unchanged (see Table 11 in Appendix A).

Table 10: Change-in-Vote-Share Regression with $\eta = 0$ and $Z \equiv 0$

Dep. Var.	$\%D_t - \%D_{t-1}$	$\%D_t - \%D_{t-1}$
Constant	0.00	0.00
TWO^+	-8.56	-
SIX^+	-	-8.56

8 Summary

In this paper, we documented a strong pattern of incumbency punishment in U.S. politics. Postwar evidence of the electoral performance of the two parties in House and state gubernatorial elections show that a long incumbency of a party leads to substantial decline in the popularity of the party in elections. We used a well-known model of partisan politics with elections to explain this finding. We showed that costs of changing policies, or simply constraints on how much policies can change from one period to the next, combined with uncertainty in election outcomes, can generate incumbency disadvantage. We examined the implications of policy inertia for how parties choose policies. We showed that inertia can cause parties to target policies that are more extreme than the policies they would support in the absence of inertia and that such extremism can be welfare reducing. On the other hand, inertia implies that office motivation matters for policy choice, even when there is no preelection commitment to policies, and this can dampen policy cycles and raise welfare.

References

Acemoglu, D., M. Golosov, and A. Tsyvinski (2011): "Power Fluctuations and Political Economy," *Journal of Economic Theory*, 146, 1009–1041.

Aghion, P., and P. Bolton (1990): "Government Debt and the Risk of Default: A Political-Economic Model of the Strategic Role of Debt," in *Public Debt Management: Theory and History*, ed. by R. Dornbusch, and M. Draghi, chap. 11, pp. 315–345. Cambridge University Press.

Aguiar, M., and M. Amador (2011): "Growth in the Shadow of Expropriation," *Quarterly Journal of Economics*, 126, 651–697.

Ales, L., P. Maziero, and P. Yared (2014): "A Theory of Political and Economic Cycles," *Journal of Economic Theory*, 153, 224–251.

Alesina, A. (1988): "Credibility and Policy Convergence in a Two-Party System with Rational Voters," *American Economic Review*, 78(4), 796–805.

Alesina, A., and H. Rosenthal (1989): "Partisan Cycles in Congressional Elections and the Macroeconomy," *The American Political Science Review*, 83(2), 373–398.

Alesina, A., and G. Tabellini (1990): "A Positive Theory of Fiscal Deficits and Government Debt in a Democracy," *Review of Economic Studies*, 57, 403–414.

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

Azzimonti, M. (2011): "Barriers to Investment in Polarized Societies," *American Economic Review*, 101(5), 2182–2204.

Azzimonti, M., M. Battaglini, and S. Coate (2016): "The Costs and Benefits of Balanced Budget Rules: Lessons from a Political Economy Model of Fiscal Policy," *Journal of Public Economics*, 136, 45–61.

Barro, R. J. (1973): "The Control of Politicians: An Economic Model," *Public Choice*, 14(1), 19–42.

Bartels, L., and J. Zaller (2001): "Presidential Vote Models: A Recount," *Political Science and Politics*, 34(1), 9–20.

Battaglini, M., and S. Coate (2008): "A Dynamic Theory of Public Spending, Taxation and Debt," *American Economic Review*, 98, 201–236.

Besley, T., and S. Coate (1998): "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis," *American Economic Review*, 88(1), 139–156.

Bowen, T., Y. Chen, and H. Eraslan (2014): "Mandatory versus Discretionary Spending: The Status Quo Effect," *American Economic Review*, 104(10), 2941–2974.

Brady, D. W., M. P. Fiorina, and A. S. Wilkins (2011): "The 2010 Elections: Why Did Political Science Forecasts Go Awry?," *PS: Political Science and Politics*, 44(2), 247–250.

Calvert, R. L. (1985): "Robustness of the Multidimensional Voting Model: Candidate Motivation, Uncertainty, and Convergence," *American Journal of Political Science*, 29(1), 69–95.

Campbell, J. E. (1997): "The Presidential Pulse and the 1994 Midterm Congressional Elections," *Journal of Politics*, 59(3), 830–857.

Chatterjee, S., and B. Eyigungor (2012): "Maturity, Indebtedness and Default Risk," *American Economic Review*, 102(6), 2674–2699.

——— (2016): "Continuous Markov Equilibria with Quasi-Geometric Discounting," *Journal of Economic Theory*, 163(May), 467–494.

——— (2017): "Endogenous Political Turnover and Fluctuations in Sovereign Default Risk," Working Paper 17-01, Federal Reserve Bank of Philadelphia.

Dixit, A., G. Grossman, and F. Gul (2000): "Dynamics of Political Compromise," *Journal of Political Economy*, 108(3), 531–568.

Doraszelski, U., and M. Satterthwaite (2010): "Computable Markov-Perfect Industry Dynamics," *RAND Journal of Economics*, 41(2), 215–243.

Downs, A. (1957): *An Economic Theory of Democracy.* New York: Harper and Row.

Duggan, J., and T. Kalandrakis (2012): "Dynamic Legislative Policy Making," *Journal of Economic Theory*, 147, 1653–1688.

Dziuda, W., and A. Loeper (2016): "Dynamic Collective Choice with Endogenous Status Quo," *Journal of Political Economy*, 124(4), 1148–1186.

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

Erikson, R. S. (1988): "The Puzzle of Midterm Loss," *Journal of Politics*, 50(4), 1011–1029.

Erikson, R. S., M. B. Mackuen, and J. A. Stimson (2002): *The Macro Polity.* Cambridge University Press.

Fair, R. C. (2009): "Presidential and Congressional Vote-Share Equations," *American Journal of Political Science*, 53(1), 55–72.

Feldman, M., and C. Gilles (1985): "An Expository Note on Individual Risk without Aggregate Uncertainty," *Journal of Economic Theory*, 35, 26–32.

Ferejohn, J. (1986): "Incumbent Performance and Electoral Control," *Public Choice*, 50(1), 5–25.

Gelman, A., and G. King (1990): "Estimating Incumbency Advantage Without Bias," *American Journal of Political Science*, 34(4), 1142–1164.

Harsanyi, J. C. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *International Journal of Game Theory*, 2, 1–23.

Hassler, J., J. V. R. Mora, K. Storesletten, and F. Zilibotti (2003): "Survival of the Welfare State," *American Economic Review*, 93(1), 87–112.

Hutson, V., and J. S. Pym (1980): *Applications of Functional Analysis and Operator Theory*. Academic Press.

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

Judd, K. L. (1985): "Law of Large Numbers with a Continuum of IID Random Variables," *Journal of Economic Theory*, 35, 19–25.

Kroger, G., and M. J. Lebo (2012): "Strategic Party Government and the 2010 Elections," *American Politics Research*, 40(5), 927–945.

Lewis-Beck, M. S., and T. W. Rice (1984): "Forecasting U.S. House Elections," *Legislative Studies Quarterly*, 9(3), 475–486.

Lindbeck, A., and J. Weibull (1987): "Balanced-Budget Redistribution as the Outcome of Political Competition," *Public Choice*, 52(3), 273–297.

Mateos-Planas, X. (2012): "Demographics and the Politics of Capital Taxation in a Life-Cycle Economy," *American Economic Review*, 100(1), 337–363.

Mayhew, D. R. (2008): "Incumbency Advantage in U.S. Presidential Elections: The Historical Record," *Political Science Quarterly*, 123(2), 201–228.

Milesi-Ferretti, G. M., and E. Spolare (1994): "How Cynical Can an Incumbent Be? Strategic Policy in a Model of Government Spending," *Journal of Public Economics*, 55(1), 121–140.

Ortega, J., and W. Rheinboldt (1970): *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, Inc.

Piguillem, F., and A. Riboni (2015): "Spending-Biased Legislators: Discipline Through Disagreement," *Quarterly Journal of Economics*, 130(2), 901–949.

Poole, K. T., and H. Rosenthal (1985): "A Spatial Model for Legislative Roll Call Analysis," *American Journal of Political Science*, 29(2), 357–384.

Rogoff, K. (1990): "Equilibrium Political Budget Cycles," *American Economic Review*, 80(1), 21–36.

Scholl, A. (2017): "The Dynamics of Sovereign Default Risk and Political Turnover," *Journal of International Economics*, 108, 37–53.

Song, Z., K. Storesletten, and F. Zilibotti (2012): "Rotten Parents and Disciplined Children: A Politico-Economic Theory of Public Expenditure and Debt," *Econometrica*, 80(6), 2785–2803.

Stokes, D., and G. Iverson (1962): "On the Existence of Forces Restoring Party Competition," *American Association for Public Opinion Research*, 26(2), 159–171.

Stokey, N. L., and R. E. Lucas Jr. (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.

Uhlig, H. (1996): "A Law of Large Numbers for Large Economies," *Economic Theory*, 8, 41–50.

Wittman, D. (1983): "Candidate Motivation: A Synthesis," *American Political Science Review*, 77(1), 142–157.

APPENDIX

A Empirical Appendix

Table 11 reports the results of two robustness checks for U.S. Congressional elections. In the first, the dependent variable is the Democratic seat share won (DS). The SIX^+ coefficient is again significant and the magnitude is more than double the coefficient in the main regression. It appears that a given decline in vote share leads to a larger proportional loss in seat share. In the second, we check if the inclusion of a time trend changes the results of our main regression. In this case the cubic time trend improved the adjusted R^2 relative to a quadratic time trend. Notice that the SIX^+ coefficient is virtually unchanged relative to our main regression.

Table 11:
Robustness Checks: U.S. Congressional Elections

DEP. VAR.	*DS*	*DV*
CONSTANT	54.33^{***} (1.95)	48.09^{***} (0.72)
SIX^+	-5.76^{***} (1.74)	-2.36^{***} (0.48)
$PDPIGR * TWO^+$	0.88^{**} (0.54)	0.48^{***} (0.14)
$MIDTERM * TWO^+$	-0.28 (1.87)	-1.03^{*} (0.51)
CUBIC TIME TREND	-	YES
SD(DEP VAR)	7.27	3.22
NO. OF OBS.	32	32
R^2	0.39	0.81
ADJ. R^2	0.32	0.77

B Proof of Theorem 1

The goal of this appendix is to provide a secure conceptual and computational foundation for partisan political economy models that feature elections with forward-looking voting behavior. The decision problems that arise in these models can, very naturally, result in nonconcave objective functions. This means that *if* a stationary pure strategy MPE exists, the equilibrium decision rules and continuation value functions need not be continuous in state variables.[32] Because of the potential discontinuity of equilibrium decision rules, ensuring the existence of a stationary pure strategy MPE requires additional structure. The goal of this appendix is to give easily verifiable conditions on primitives for which at least one pure strategy Markov Perfect Equilibrium is assured and can be computed. These conditions apply to a class of models that includes all the models discussed in the main text and, potentially, other models of interest.

[32] See Chatterjee and Eyigungor (2016) for an example of a political economy model with exogenous political turnover with these features. Endogenizing political turnover *via* forward-looking voting behavior does not eliminate nonconcavities, so these issues persist in the class of models being considered in this paper.

B.1 The Assumptions on Primitives

Finite States and Actions:

Let $\mathcal{I} = \{1, 2, \ldots, I\}$, $I \geq 2$ be the set of possible endogenous states the economy can be in at the start of any given period.[33] We use i and j to denote generic elements of $\mathcal{I}$. The action space when the state is i and party k is making decisions is denoted $\Gamma_i^k \subseteq \mathcal{I}$. We say $\{i, j, k\}$ is a *feasible triple* if $j \in \Gamma_i^k$.

Assumption 1. *$\Gamma_i^k \neq \emptyset$ for all $i \in \mathcal{I}$ and all $k \in \{D, R\}$.*

In the models with quadratic adjustment costs, $\Gamma_i^k = \mathcal{I}$ and is independent of i and k. In the "constraint-on-change" model, Γ_i^k is independent of k but not i. In the first case, Assumption 1 is satisfied by virtue of $\mathcal{I}$ being nonempty; in the second case, it is satisfed because $i \in \Gamma_i^k$.

Current period rewards:

Let $u_{i\,j}^k(m_k)$ denote the current period reward to a k type if party k is in power, the state is i, the preference shock is m_k and j is chosen and let $\tilde{u}_{i\,j}^k(m_{\sim k})$ denote the current period reward to a k type if party $\sim k$ is in power, the state is i, the preference shock is $m_{\sim k}$ and j is chosen. Here $m_k \in [\underline{m}, \overline{m}] \equiv M \subset \mathbb{R}$.[34]

When the party in power is D, the utility flows to all individuals are augmented by $e + A$, where $e \in \mathbb{R}$ and $A \in \mathbb{R}$ are the idiosyncratic and aggregate components of an individual's net preference for the D party.

Let $U_{i,\,j}^k(m_k)$ denote the current period reward to party k when party k is in power, the state is i, the preference shock is m_k and j is the chosen and let $\tilde{U}_{i,\,j}^k(m_{\sim k})$ denote the current period reward to party k when party $\sim k$ is in power, the state is i, the preference shock is $m_{\sim k}$ and j is the chosen.

Assumption 2. *For all feasible $\{i, j, k\}$ triples, $u_{i,j}^k(m_k) : M \to \mathbb{R}$ and $U_{i,j}^k(m_k) : M \to \mathbb{R}$ are continuous in m_k and $\tilde{u}_{i,j}^k(m_{\sim k}) : M \to \mathbb{R}$ and $\tilde{U}_{i,j}^k(m_{\sim k}) : M \to \mathbb{R}$ are continuous in $m_{\sim k}$.*

Assumption 3. *Let $\{i, j, k\}$ and $\{i, j', k\}$ be any pairs of feasible triples. Then $U_{i\,j}^k(m_k) - U_{i\,j'}^k(m_k)$ is strictly monotone in $m_k \in M$.*

For the models in the main text, Assumption 3 is satisfied by virtue of the concavity of $U(x)$ and $U(\tau - x)$. To see this, consider $k = D$ and let $\delta(m_k) \equiv U_{i\,j}^D(m_k) - U_{i\,j'}^D(m_k)$. Then (recalling that m_D is denoted m_1 in the main text),

$$\delta(m_1) = U(g_j + m_1) + \alpha U(\tau - g_j) - \psi(g_i, g_j) - U(g_{j'} + m_1) - \alpha U(\tau - g_{j'}) + \psi(g_i, g_{j'}).$$

Observe that $\delta'(m_1) = U'(g_j + m_1) - U'(g_{j'} + m_1)$. Since $g_j \neq g_{j'}$, $g_j + m_1$ is either less than or greater than $g_{j'} + m_1$ for all m_1. By concavity of U, $\delta'(m_1)$ is either strictly positive for all m_1 or strictly negative for all m_1 (an analogous argument establishes the result for $k = R$).

Probability Spaces:

[33] The fact that $\mathcal{I}$ contains only endogenous states is also not restrictive. The proof of existence can be straightforwardly extended to include any number of discrete shocks that affect feasible sets. The same is true for the computation method.

[34] In the main text, m_D is m_1 (the preferred good of type D) and m_R is m_2 (the preferred good of type R).

Let $(M, \mathcal{B}_M, \mu)$ denote a probability space on M, where $\mathcal{B}_M$ denotes the Borel σ-algebra on $\mathbb{R}$ restricted to M.

Assumption 4. *μ is absolutely continuous with respect to the Lebesgue measure on M.*

This assumption means that any subset of M that is of Lebesgue measure 0 has probability zero with respect to the probability measure μ. Any random variable described by a continuous density on M will satisfy this assumption.

Let $(\mathbb{R}, \mathcal{B}, \lambda)$ denote the probability space on $\mathbb{R}$ for the aggregate voter preference shock A.

Assumption 5. *For all $z \in \mathbb{R}$, $\int_{A>z} A\, d\lambda$ exists and is continuous in z and there is a $\overline{A} > 0$ such that $|\int_{A>z} A\, d\lambda| < \overline{A}$.*

This assumption is satisfied by any random variable with a continuous density on a compact support. It is also satisfied by a random variable with unbounded support if it possessed a continuous density over $\mathbb{R}$ and that density converged to 0 exponentially fast as A diverged to $+\infty$ or $-\infty$ (as is the case, for instance, with the normal distribution).

Let $(\mathbb{R}, \mathcal{B}, \varepsilon)$ denote a probability space on $\mathbb{R}$ for the idiosyncratic voter preference shock e and let $F(e) = \varepsilon(-\infty, e])$ be its distribution.

Assumption 6. *$\int_{\mathbb{R}} e d\varepsilon = 0$, $\varepsilon((-e, 0]) = \varepsilon([0, e))$ for all $e \in \mathbb{R}$ and $F(e)$ is continuous and strictly increasing in e.*

Any random variable with a density function that is symmetric around 0 and strictly positive for all $e \in \mathbb{R}$ (such as the normal distribution) will satisfy Assumption 6.

In what follows, we use some standard results from measure theory and functional analysis. In most instances, the proofs of these results can be found in Stokey and Lucas Jr. (1989), and when this is the case, we cite the relevant section of their text.

B.2 Preliminary Lemma

Lemma B.2.1 (Boundedness of Current Period Rewards). *There exists $\overline{U} > 0$ such that $|u^k_{ij}(m_k)|$, $|U^k_{ij}(m_k)|$, $|\tilde{u}^k_{ij}(m_{\sim k})|$, $|\tilde{U}^k_{ij}(m_{\sim k})|$ are all strictly less than $\overline{U}$ for all feasible triples $\{i, j, k\}$ and all m_k, $m_{\sim k} \in M$.*

Proof. Since a real-valued continuous function on compact set must be bounded, each of the functions in the Lemma can be given a bound. And since there is a finite number of such functions, an $\overline{U} > 0$ exceeding all of the individual bounds exists. $\square$

B.3 Decision Problem of the Party in Power

To recall: A period begins in some state i. The voter preference shocks e and A are realized and people vote. If party k wins, the preference shock m_k is realized ($m_{\sim k}$ is automatically zero) and the party chooses next period's state j.

Let $Q^k_i \in \mathbb{R}$, $k \in \{D, R\}$, denote the value of party k of starting a period in state i.

Let $A_i \in \mathbb{R}$ denote the threshold value of A in state i, i.e., if $A > A_i$, D party wins the election in state i.

Let ω_i denote the 3-tuple $\left(Q_i^D, Q_i^R, A_i\right)$. Let $\omega = (\omega_1, \omega_2, \ldots, \omega_I)$ be a vector composed of these 3-tuples. Then ω is an element of $\{\mathbb{R} \times \mathbb{R} \times \mathbb{R}\}^I$.

We use $Q_i^D[\omega]$, $Q_i^R[\omega]$ and $A_i[\omega]$ to denote the specific elements of the ith component of ω.

Since m_k shocks are never active simultaneously, we reduce notational burden by using m to denote realizations of whichever m_k shock is active.

Value Functions:
With these conventions, let $V_{ij}^k(m;\omega)$ be the value to party k when it is in power, the state is i, its preference shock is m, and it chooses j. Then,

$$V_{i,j}^k(m;\omega) = U_{i,j}^k(m) + \beta Q_j^k[\omega], \;\; k \in \{D, R\}. \tag{23}$$

Let $V_i^k(m;\omega)$ be the *optimal* value of party k under the same circumstances. Then,

$$V_i^k(m;\omega) = \max_{j \in \Gamma_i^k} V_{i,j}^k(m;\omega). \tag{24}$$

By Assumption 1, the set of maximizers is nonempty for all i. Let $j_i^k(m;\omega)$ denote the maximizer if it is unique. If the set of maximizers is not unique, we adopt the following tie-breaking rule: $j_i^k(m;\omega)$ is the maximizer with the smallest (index) j.

Proposition B.3.1 (Continuity of V_i^k)**.** *For all i and k, $V_i^k(m;\omega) : M \times \Omega \to \mathbb{R}$ is continuous in m and ω.*

Proof. $V_i^k(m;\omega)$ is the upper envelope of a finite number of functions $V_{i,j}^k(m;\omega) : M \times \Omega \to \mathbb{R}$, $j \in \Gamma_i^k$. By Assumption 2, U_{ij}^k is continuous in m and $Q_j^k[\omega]$ is trivially continuous in ω and so each of these functions is continuous in m and ω. Then the upper envelope of these functions must also be continuous in m and ω. □

Proposition B.3.2 (Integrability of V_i^k w.r.t. m)**.** *Given ω, $V_i^k(\omega) \equiv \mathbb{E}_m V_i^k(m;\omega)$ exists.*

Proof. Since $V_i^k(m;\omega)$ is continuous in m, $V_i^k(m;\omega)$ is measurable with respect to $\mathcal{B}_M$ [SL, Ch. 7, p. 178]. Since M is compact, $\inf_m V_i^k(m;\omega)$ and $\sup_m V_i^k(m;\omega)$ both exist. Then there is some $\overline{V}(\omega) > 0$ for which $\mid V_i^k(m;\omega) \mid < \overline{V}(\omega)$. Therefore $V_i^k(m;\omega)$ is a bounded and measurable function and since $\mu(M)$ is finite (equal to 1), $\int V_i^k(m;\omega) d\mu = \mathbb{E}_m V_i^k(m;\omega)$ exists [SL, Ch. 7, p. 192]. □

To complete the statement of the party's decision problem, let $X_i^k(m;\omega)$ be the value to party k when it is not in power, the state is i, the preference shock of party $\sim k$ is m and party $\sim k$ chooses j optimally. Then,

$$X_i^k(m;\omega) = \tilde{U}^k_{i\, j_i^{(\sim k)}(m;\omega)}(m) + \beta Q^k_{j_i^{(\sim k)}(m;\omega)}[\omega], \;\; k \in \{D, R\}. \tag{25}$$

Observe that the value of party k when it is not in power is not the maximum of an optimization problem. It is, instead, pinned down by the actions chosen by the other party to maximize its own objective function. Thus it is no longer true that $X_i^k(m;\omega)$ is necessarily continuous in m and ω. An inconvenient consequence is that the integrability of $X_i^k(m;\omega)$ w.r.t. m (and of other functions that similarly depend on m via decision rules) cannot be established as easily as for $V_i^k(m;\omega)$ in Proposition B.3.2. More information on the properties of decision rules $j_i^k(m,\omega)$ is needed.

Decision Rules:
The next three lemmas establish key properties of decision rules.

Lemma B.3.1 (Maximizers are almost always unique)**.** *Given k, i and ω, $j_i^k(m;\omega)$ strictly dominates any other feasible choice, except, possibly, at a finite number of m values.*

Proof. Given k, i and ω, the optimal choice at m is unique if $V^k_{i,\,j_i^k(m;\omega)}(m;\omega) > V^k_{i,\,j'}(m;\omega)$ for all $j' \in \Gamma^k_{i,\,j} \setminus j_i^k(m;\omega)$. We show that this inequality holds for all but a finite number (possibly zero) of m values.

Let $j, j' \in \Gamma^k_i$, $j \neq j'$, and let $M^k_i(j, j';\omega) \subseteq M$ be $\{m : V^k_{i,\,j}(m;\omega) = V^k_{i,\,j'}(m;\omega)\}$. Now,

$$V^k_{i,\,j}(m;\omega) - V^k_{i,\,j'}(m;\omega) = U^k_{i,j}(m) - U^k_{i,j'}(m) + \beta Q^k_j[\omega] - \beta Q^k_{j'}[\omega].$$

By Assumption 3, the r.h.s. is strictly monotone in m. Therefore, either $M^k_i(j, j';\omega)$ is empty or it contains exactly one point.

Given ω, let

$$M^k(\omega) = \left\{ \bigcup_{i\in\mathcal{I}} \left\{ \bigcup_{j,\,j'\in\Gamma^k_i,\,j\neq j'} M^k_i(j, j';\omega) \right\} \right\}$$

be the collection of all such (indifference) points for party k. Since $\mathcal{I}$ is a finite set, $M(\omega)$ is a finite set. Now consider $\hat{m} \in M \setminus M^k(\omega)$. Then $V^k_{i,\,j_i^k(\hat{m};\omega)}(m;\omega) > V^k_{i,\,j'}(\hat{m};\omega)$ for any $j' \in \Gamma^k_{i,j} \setminus j_i^k(m;\omega)$. If not, $\hat{m}$ must belong to $M^k_i(j_i^k(\hat{m};\omega), j';\omega)$ for some j' and so must belong to $M^k(\omega)$, which is impossible in view of the choice of $\hat{m}$. Since $M \setminus M(\omega)$ contains all but a finite number of m values, the result follows. □

Lemma B.3.2 (Measurability w.r.t. m)**.** *Given i, k and ω, let $B^k_{i,\,j}(\omega) \subseteq M$ be the set $\{m \in M : j^k_{i\,j}(m;\omega) = j\}$ of m values for which the optimal choice of party k is j. Then, $B^k_{i\,j}(\omega) \in \mathcal{B}_M$ for all $j \in \mathcal{I}$.*

Proof. We prove this by showing that $B^k_{i\,j}(\omega)$ is the union of two Borel sets.

Fix k, i and ω. For each $j \in \mathcal{I}$, let

$$V^k_{i\setminus j}(m;\omega) = \max_{j'\in\mathcal{I}\setminus j} V^k_{i,\,j'}(m;\omega)$$

denote the optimal value of party k excluding policy j. Now consider the difference function $f^k_{i\,j}(m;\omega) : M \to \mathbb{R}$ defined as $V^k_{i\,j}(m;\omega) - V^k_{i\setminus j}(m;\omega)$. Then $\hat{B}^k_{i\,j}(\omega) = \{m \in M : f^k_{i\,j}(m;\omega) > 0\}$ is the set of m points for which j is the unique maximizer. Since $f^k_{i\,j}$ is the difference of two functions continuous in m, $f^k_{i\,j}$ is continuous in m and, hence, $\hat{B}^k_{i\,j}(\omega) \in \mathcal{B}_M$.

Next, given k, i and ω, consider the set of m values for which the maximizer is not unique. By Lemma B.3.1, this set is finite. Of this finite set of m values, let $\Phi^k_{i,\,j}(\omega)$ be the subset of m values for which $j^k_i(m;\omega) = j$, i.e., the subset of m values for which j is the optimal choice because it was smallest j among all optimal j's (the tie-breaking rule).

Then

$$B_{i,j}^k = \hat{B}_{i,j}^k \cup \Phi_{i,j}^k(\omega).$$

Since any finite subset of M is a Borel set and the union of two Borel sets is Borel, it follows that $B_{i\,j}^k(\omega) \in \mathcal{B}_M$. $\square$

Lemma B.3.3. *Given i, k and ω, let $\chi_{B_{i\,j}^k(\omega)}(m)$ denote the indicator function that is 1 if $m \in B_{i\,j}^k(\omega)$ and 0 otherwise. Let $\theta(m) : M \to \mathbb{R}$ be a continuous real-valued function of m. Then the product function $\theta(m)\chi_{B_{i\,j}^k(\omega)}(m)$ is measurable with respect to $\mathcal{B}_M$ and integrable with respect to μ.*

Proof. Since $B_{i\,j}^k(\omega) \in \mathcal{B}_M$ (Lemma B.3.2), $\chi_{B_{i\,j}^k(\omega)}(m)$ is a measurable function. Since $\theta_j(m)$ is continuous, it is also a measurable function. Therefore, $\theta_j(m)\chi_{B_{i\,j}^k(\omega)}(m)$ (being the product of measurable functions) is also a measurable function. Since M is compact, the function $\theta(j)$ is bounded and, therefore, so is $\theta_j(m)\chi_{B_{i\,j}^k(\omega)}(m)$. Since bounded measurable functions are integrable, $\theta_j(m)\chi_{B_{i\,j}^k(\omega)}(m)$ is integrable. $\square$

Lemma B.3.4 (Almost everywhere convergence of decision rules)**.** *Let $\{\omega_n\}$ be a sequence converging to ω. Then, for each i and k, the sequence of functions $\{j_i^k(m; \omega_n) : M \to \mathcal{I}\}$ converges pointwise to the function $j_i^k(m; \omega) : M \to \mathcal{I}$ except, possibly, for a finite number of m values.*

Proof. Pick a point in $\hat{m} \in M$ and suppose that $j_i^k(\hat{m}; \omega)$ is a unique maximizer. Let $V_i^{k-}(\hat{m}; \omega) = \max_{j \in \Gamma_i^k \setminus j_i^k(\hat{m};\omega)} V_{i\,j}^k(\hat{m}; \omega)$. Then, $V_i^k(\hat{m}, \omega) > V_i^{k-}(\hat{m}; \omega)$. Since both $V_i^{k-}(\hat{m}; \omega)$ and $V_i^k(\hat{m}; \omega)$ are continuous in ω, there exists N such that for all $n > N$, $V_i^k(\hat{m}; \omega_n) > V_i^{k-}(\hat{m}; \omega_n)$. Then $j_i^k(\hat{m}; \omega_n) = j_i^k(\hat{m}; \omega)$ for all $n > N$. But this implies that $\lim_n j_i^k(\hat{m}, \omega_n) = j_i^k(\hat{m}, \omega)$. Since the maximizer $j_i^k(m; \omega)$ is unique for all but a finite number of m's (Lemma B.3.1) the result follows. $\square$

Lemma B.3.5 (Continuity of Expected Value)**.** *Let $\theta(m) : M \to \mathbb{R}$ be a continuous real-valued function of m. Then, $\int \theta(m)\chi_{B_{i\,j}^k(\omega)}(m)d\mu$ is continuous in ω.*

Proof. Let $\omega_n \to \omega$

$$\theta(m)\chi_{B_{i,j}^D(\omega_n)}(m) = \begin{cases} \theta(m) & \text{if} j = j_i^D(m; \omega_n) \\ 0 & \text{if} j \neq j_i^D(m; \omega_n) \end{cases}.$$

By Lemma B.3.4, $j_i^D(m, \omega_n)$ converges pointwise to $j_i^D(m, \omega)$ except, possibly, for a finite number of m values. Therefore, $f_n(m) \equiv \theta(m)\chi_{B_{i,j}^D(\omega_n)}(m)$ converges pointwise to $f(m) \equiv \theta(m)\chi_{B_{i,j}^D(\omega)}(m)$ except, possibly, for a finite number of m values. A finite subset of M has Lebesgue measure 0 and so by Assumption 4 $f_n(m)$ converges to $f(m)$, μ - a.e. Furthermore, $|f_n(m)| \leq |\theta(m)|$ is a sequence of bounded functions. By the Lebesgue Dominated Convergence Theorem (SL, Theorem 7.10, p. 192), $\lim_n \int \theta(m)\chi_{B_{i,j}^D}(\omega_n)(m)d\mu = \int \theta(m)\chi_{B_{i,j}^D}(\omega)d\mu$. $\square$

We close this subsection with:

Proposition B.3.3 (Integrability of X_i^k w.r.t. m)**.** *For each k, i and ω, $X_i^k(\omega) \equiv \mathbb{E}_m X_i^k(m; \omega)$ exists.*

Proof. The key step is simply the observation that $X_i^k(m;\omega)$ can be expressed as:

$$X_i^k(m;\omega) = \sum_j \left\{ \tilde{U}_{ij}^k(m) + \beta Q_{ij}^k[\omega] \right\} \chi_{B_{ij}^{\sim k}(\omega)}(m). \tag{26}$$

For each j, $\tilde{U}_{ij}^k(m)+\beta Q_{ij}^k[\omega]$ is a continuous function of m (Assumption 2) and, therefore, $\{\tilde{U}_{ij}^k(m)+\beta Q_{ij}^k[\omega]\}\chi_{B_{ij}^{\sim k}(\omega)}(m)$ is a integrable function of m (Lemma B.3.3). Since a finite sum of integrable functions is also integrable, $X_i^k(m;\omega)$ is integrable. □

B.4 Lifetime Utilities of Voters

The goal of this subsection is to establish that given a pair of decision rules $j_i^k(m;\omega)$, $k \in \{D,R\}$, and values for the thresholds $A_i[\omega]$ of the aggregate voter preference shock is (above which the D party is elected), the voter value functions $\{W_i^k(m;\omega), Z_i^k(m;\omega)\}$ $k \in \{D,R\}$, are uniquely determined.[35] Furthermore, these value functions are continuous in ω.

Let $\Pi_i(\omega) \equiv \int_{z>A_i[\omega]} d\lambda$ denote the probability of a D party win given ω and let $\overline{A}_i(\omega) \equiv \int_{z>A_i[\omega]} A\, d\lambda$ denote the A_i-truncated-expectation of A, which exist by Assumption 5.

Then, the voter value functions, if they exist, must satisfy the following recursions:

$$\begin{aligned} &W_i^D(m;\omega) = &(27)\\ &u^D_{i,j_i^D(m;\omega)}(m) + \beta\left\{\Pi_{j_i^D(m;\omega)}\mathbb{E}_{m'}W^D_{j_i^D(m;\omega)}(m';\omega) + \left[1-\Pi_{j_i^D(m;\omega)}\right]\mathbb{E}_{m'}Z^D_{j_i^D(m;\omega)}(m';\omega) + \overline{A}_{j_i^D(m;\omega)}\right\} \\ &Z_i^D(m;\omega) = &(28)\\ &\tilde{u}^D_{i,j_i^R(m;\omega)}(m) + \beta\left\{\Pi_{j_i^R(m;\omega)}\mathbb{E}_{m'}W^D_{j_i^R(m;\omega)}(m';\omega) + \left[1-\Pi_{j_i^R(m;\omega)}\right]\mathbb{E}_{m'}Z^D_{j_i^R(m;\omega)}(m';\omega) + \overline{A}_{j_i^R(m;\omega)}(\omega)\right\}, \end{aligned}$$

and

$$\begin{aligned} &W_i^R(m;\omega) = &(29)\\ &u^R_{i,j_i^R(m;\omega)}(m) + \beta\left\{\Pi_{j_i^R(m;\omega)}\mathbb{E}_{m'}Z^R_{j_i^R(m;\omega)}(m';\omega) + \left[1-\Pi_{j_i^R(m;\omega)}\right]\mathbb{E}_{m'}W^R_{j_i^R(m;\omega)}(m';\omega) + \overline{A}_{j_i^R(m;\omega)}(\omega)\right\} \\ &Z_i^R(m;\omega) = &(30)\\ &\tilde{u}^R_{i,j_i^D(m;\omega)}(m) + \beta\left\{\Pi_{j_i^D(m;\omega)}\mathbb{E}_{m'}Z^R_{j_i^D(m;\omega)}(m';\omega) + \left[1-\Pi_{j_i^D(m;\omega)}\right]\mathbb{E}_{m'}W^R_{j_i^D(m;\omega)}(m';\omega) + \overline{A}_{j_i^D(m;\omega)}(\omega)\right\}, \end{aligned}$$

Proposition B.4.1. *Let $\mathcal{F}$ denote the set of all $\mathcal{B}_M$-measurable functions $f : M \to \mathbb{R}$ for which $\int f\, d\mu$ exists with respect to the probability space $(M, \mathcal{B}_M, \mu)$. Then, for every ω there exists a set of functions $\left\{W_i^k(m;\omega), Z_i^k(m;\omega)\right\}$, $i \in \mathcal{I}$, all members of $\mathcal{F}$, that satisfy the recursions* (27) - (28) *for $k = D$ and* (29) - (30) *for $k = R$.*

Proof. We will prove the proposition for $k = D$ (the proof for $k = R$ is analogous).

Fix ω. We may view the r.h.s of (27)-(28) as an operator taking as input the set of functions $(W^D(m;\omega), W^R(m;\omega)) \equiv \{W_i^D(m;\omega), Z_i^D(m;\omega)\}_{i\in\mathcal{I}}$. We will establish that if all members of this set belong to $\mathcal{F}$ then the output functions on the l.h.s. of (27)-(28) also belong to $\mathcal{F}$.

[35] Recall that these value functions give the lifetime utility of type k when the state is i and (the active) preference shock is m, ignoring the value of an individual's net preference for the D party when the D party is in power.

First, observe that if the input functions are members of $\mathcal{F}$ then the expectations of these functions with respect to m' exist and the r.h.s. is well-defined.

Next, observe that we can re-express the r.h.s. of (27)-(28) as:

$$W_i^D(m;\omega) = \sum_j \left[u_{i,j}^D(m) + \beta\left\{\Pi_j(\omega)\mathbb{E}_{m'}W_j^D(m';\omega) + [1-\Pi_j(\omega)]\,\mathbb{E}_{m'}Z_j^D(m';\omega) + \overline{A}_j[\omega]\right\}\right]\chi_{B_{i,j}^D(\omega)}(m) \tag{31}$$

$$Z_i^D(m;\omega) = \sum_j \left[\tilde{u}_{i,j}^D(m) + \beta\left\{\Pi_j(\omega)\mathbb{E}_{m'}W_j^D(m';\omega) + [1-\Pi_j(\omega)]\,\mathbb{E}_{m'}Z_j^D(m';\omega) + \overline{A}_j[\omega]\right\}\right]\chi_{B_{i,j}^R(\omega)}(m) \tag{32}$$

By Lemma B.3.3, each term in the summation is an integrable function of m and, therefore, the summation is as well. Hence the output functions in the l.h.s. of (27)-(28) belong to $\mathcal{F}$.

Taking expectations w.r.t. m on both sides (31)-(32) yields a pair of recursions in the expectation (w.r.t. m) of the D types' value functions:

$$\mathbb{E}_m W_i^D(m,\omega) = \\ \sum_j \left[\int u_{i,j}^D(m)\chi_{B_{i,j}^D(\omega)}(m)d\mu + \left[\beta\left\{\Pi_j(\omega)\mathbb{E}_{m'}W_j^D(m';\omega) + [1-\Pi_j(\omega)]\,\mathbb{E}_{m'}Z_j^D(m';\omega) + \overline{A}_j[\omega]\right\}\right]\mu(B_{i,j}^D(\omega))\right] \tag{33}$$

$$\mathbb{E}_m Z_i^D(m;\omega) = \\ \sum_j \left[\int \tilde{u}_{i,j}^D(m)\chi_{B_{i,j}^R(\omega)}(m)d\mu + \left[\beta\left\{\Pi_j(\omega)\mathbb{E}_{m'}W_j^D(m';\omega) + [1-\Pi_j(\omega)]\,\mathbb{E}_{m'}Z_j^D(m';\omega) + \overline{A}_j[\omega]\right\}\right]\mu(B_{i,j}^R(\omega))\right] \tag{34}$$

Or, more compactly,

$$\overline{W}_i^D(\omega) = \sum_j \left[\int u_{i,j}^D(m)\chi_{B_{i,j}^D(\omega)}(m)d\mu + \left[\beta\left\{\Pi_j(\omega)\overline{W}_j^D(\omega) + [1-\Pi_j(\omega)]\,\overline{Z}_j^D(\omega) + \overline{A}_j[\omega]\right\}\right]\mu(B_{i,j}^D(\omega))\right] \tag{35}$$

$$\overline{Z}_i^D(\omega) = \sum_j \left[\int \tilde{u}_{i,j}^D(m)\chi_{B_{i,j}^R(\omega)}(m)d\mu + \left[\beta\left\{\Pi_j(\omega)\overline{W}_j^D(\omega) + [1-\Pi_j(\omega)]\,\overline{Z}_j^D(\omega) + \overline{A}_j[\omega]\right\}\right]\mu(B_{i,j}^R(\omega))\right] \tag{36}$$

We may verify that the operator defined by the r.h.s. of (35)-(36):

$$\left\{W_i^D(\omega)\left(\overline{W}^D(\omega), \overline{Z}^D(\omega)\right),\, Z_i^D(\omega)\left(\overline{W}^D(\omega), \overline{Z}^D(\omega)\right)\right\}_{i\in\mathcal{I}} : \{\mathbb{R}\times\mathbb{R}\}^I \to \{\mathbb{R}\times\mathbb{R}\}^I$$

satisfies Blackwell's sufficiency conditions for a contraction map (with modulus of contraction β) (SL, Theorem 3.3, p. 54). Since $\{\mathbb{R}\times\mathbb{R}\}^I$ is a complete metric space (with, say, the uniform metric), the Contraction Mapping Theorem (SL Theorem 3.2, p. 50) ensures the existence of a unique pair of vectors $(\overline{W}^{*D}, \overline{Z}^{*D})$ satisfying

$$\left(\overline{W}^{*D}, \overline{Z}^{*D}\right) = \left(W_i^D(\overline{W}^{*D}, \overline{Z}^{*D}), Z_i^D(\overline{W}^{*D}, \overline{Z}^{*D})\right)_{i\in\mathcal{I}}.$$

Then the functions, all members of $\mathcal{F}$, whose existence is asserted by the proposition are given by:

$$W_i^D(m;\omega) = u^D_{i,\,j_i^D(m;\omega)}(m) + \beta\left\{\Pi_{j_i^D(m;\omega)}\overline{W}^{*D}_{j_i^D(m',\omega)}(\omega) + \left[1-\Pi_{j_i^D(m;\omega)}\right]\overline{Z}^{*D}_{j_i^D(m',\omega)}(\omega) + \overline{A}_{j_i^D(m;\omega)}\right\}$$
$$Z_i^D(m;\omega) = \tilde{u}^D_{i,\,j_i^R(m;\omega)}(m) + \beta\left\{\Pi_{j_i^R(m;\omega)}\overline{W}^{*D}_{j_i^R(m',\omega)}(\omega) + \left[1-\Pi_{j_i^R(m;\omega)}\right]\overline{Z}^{*D}_{j_i^R(m',\omega)}(\omega) + \overline{A}_{j_i^R(m;\omega)}(\omega)\right\}$$

□

Proposition B.4.2 (Continuity of $\overline{W}^{*k}$ and $\overline{Z}^{*k}$ with respect to ω)**.** *The fixed points* ($\overline{W}^{*k}$, $\overline{Z}^{*k}$), *$k \in \{D, R\}$, vary continuously with ω.*

Proof. We will prove this for ($\overline{W}^{*D}$, $\overline{Z}^{*D}$) (the proof for $k = R$ is entirely analogous). Since the operator $\left(W_i^D(\cdot\,,\cdot),\, Z_i^D(\cdot\,,\cdot)\right)_{i\in\mathcal{I}}$ is a contraction, it is sufficient to show that it is continuous in ω (see, for instance, Theorem 4.3.6 in Hutson and Pym (1980)). That is, *given* the vectors ($\overline{W}^D$, $\overline{Z}^D$), the image $\Big(W_i^D(\overline{W}^D, \overline{Z}^D),\, Z_i^D(\overline{W}^D, \overline{Z}^D)\Big)$ varies continuously with ω for any i. We will show this for $W_i^D(\overline{W}^D, \overline{Z}^D)$ (the proof is analogous for $Z_i^D(\overline{W}^D, \overline{Z}^D)$).

From inspection of the r.h.s. of (33), the image will vary continuously with ω if, for each j, (i) $\overline{A}_j[\omega]$, (ii) $\int u^D_{i,j}(m)\chi_{B^D_{i,j}(\omega)}(m)d\mu$ and (iii) $\mu(B^D_{ij}(\omega))$ are continuous in ω.

Let $\omega_n \to \omega$. (i) Since $\overline{A}_j = \int_{A>A_j} Ad\lambda$ and $A_j[\omega_n]$ is just a sequence A_n converging to some A, continuity of $\overline{A}_j[\omega]$ with respect to ω is part of Assumption 5. (ii) Since $u^D_{i,j}(m)$ is a continuous function of m (Assumption 2), the result follows by setting $\theta(m)$ to $u^D_{i,j}(m)$ in Lemma B.3.5. (iii) Since $\mu(B^D_{i,j}(\omega)) = \int \chi_{B^D_{i,j}(\omega)}(m)$, the result follows by setting $\theta(m) = 1$ in Lemma B.3.5. □

B.5 Existence of a Fixed Point of the MPE Self Map

Let

$$\begin{aligned}
Q_i^D(\omega) &\equiv \left[\Pi_i(\omega)V_i^D(\omega) + [1-\Pi_i(\omega)]X_i^D(\omega)\right] \\
Q_i^R(\omega) &\equiv \left[\Pi_i(\omega)X_i^R(\omega) + [1-\Pi_i(\omega)]V_i^R(\omega)\right] \\
A_i(\omega) &\equiv \frac{[\overline{W}_i^{*R}(\omega) - \overline{Z}_i^{*R}(\omega)] - [\overline{W}_i^{*D}(\omega) - \overline{Z}_i^{*D}(\omega)]}{2}.
\end{aligned}$$

Define $T(\omega) : \{\mathbb{R}\times\mathbb{R}\times\mathbb{R}\}^I \to \{\mathbb{R}\times\mathbb{R}\times\mathbb{R}\}^I$ as:

$$T(\omega) \equiv \left(Q_i^D(\omega), Q_i^R(\omega), A_i(\omega)\right)_{i\in\mathcal{I}}. \tag{37}$$

Then, if an ω^* such that $T(\omega^*) = \omega^*$ exists, functions satisfying the requirements of a Markov Perfect Equilibrium stated in Definition 1 exist as well.

To show the existence of ω^*, we need to show that there is a compact subset $\Omega \subseteq \{\mathbb{R}\times\mathbb{R}\times\mathbb{R}\}^I$ such that $T(\Omega) \subseteq \Omega$ and then establish that $T(\omega) : \Omega \to \Omega$ is a continuous self-map. Brouwer's FPT can then be used to assert the existence of ω^*.

To establish the existence of Ω, suppose that $Q_i^k \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$ for all i, k. Then,

$$\mid V^k_{ij}(m;\omega) \mid \;\le\; \mid U^k_{ij}(m) \mid \; + \beta \mid Q^k_j \mid < \overline{U} + \beta\overline{U}/(1-\beta) \;=\; \overline{U}/(1-\beta).$$

Therefore, $V_i^k(m;\omega) \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$ and so $\mathbb{E}_m V_i^k(m) = V_i^k \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$. The same line of reasoning shows $X_i^k \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$. Thus, if $Q_i^k[\omega] \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$ then $Q_i^k[T(\omega)] \in [-\overline{U}/(1-\beta), \overline{U}/(1-\beta)]$.

To establish a bound for $A_i[T(\omega)]$, we first show that $\overline{W}^{*k}$ and $\overline{Z}^{*k}$ are each contained in $[-(\overline{U}+\overline{A})/(1-\beta), (\overline{U}+\overline{A}/(1-\beta)]^I$. Observe that if $\overline{W}^k$ and $\overline{Z}^k$ belong in $[-(\overline{U}+\overline{A})/(1-\beta), (\overline{U}+\overline{A})/(1-\beta)]^I$ then

$$\mid W_i^k(\overline{W}^k, \overline{Z}^k) \mid < \overline{U} + \beta[(\overline{U}+\overline{A})/(1-\beta) + \overline{A}] < \overline{U} + \overline{A} + \beta[(\overline{U}+\overline{A})/(1-\beta)] = (\overline{U}+\overline{A})/(1-\beta),$$

and, analogously, $\mid Z_i^k(\overline{W}^k, \overline{Z}^k) \mid < (\overline{U}+\overline{A})/(1-\beta)$. Since the map $\Big(W_i^k(\overline{W}^k, \overline{Z}^k), Z_i^k(\overline{W}^k, \overline{Z}^k)\Big)_{i\in\mathcal{I}}$ is a contraction, the fixed points $\overline{W}^{*k}$ and $\overline{Z}^{*k}$ must each lie in $[-(\overline{U}+\overline{A})/(1-\beta), (\overline{u}+\overline{A})/(1-\beta)]^I$. Given these bounds, we may verify that $A_i[T(\omega)] \in [-(\overline{U}+\overline{A})/(1-\beta), (\overline{U}+\overline{A})/(1-\beta)]$ for all i. Since this bound holds for any $(A_i[\omega])_{i\in\mathcal{I}}$, we have, in particular, that if $(A_i[\omega])_{i\in\mathcal{I}} \in [-(\overline{U}+\overline{A})/(1-\beta), (\overline{U}+\overline{A})/(1-\beta)]^I$ then $(A_i[T(\omega)])_{i\in\mathcal{I}} \in [-(\overline{U}+\overline{A})/(1-\beta), (\overline{U}+\overline{A})/(1-\beta)]^I$.

Thus, we may take Ω to be the hypercube $[-\overline{\omega}, \overline{\omega}]^{3I}$, where $\overline{\omega} = [\overline{U}+\overline{A}]/(1-\beta)$.

To establish that $T(\omega)$ is continuous in $\omega \in \Omega$ we need only show that $Q_i^k[T(\omega)]$ is continuous in ω for each i and k, since by Proposition B.4.2 we already know that $\overline{W}^{*k}(\omega)$ and $\overline{Z}^{*k}(\omega)$ are continuous in ω and, hence, $A_i[T(\omega)]$ is continuous in ω.

To proceed, observe that

$$\begin{aligned} Q_i^D[T(\omega)] &\equiv \left[\Pi_i(\omega)V_i^D(\omega) + [1-\Pi_i(\omega)]X_i^D(\omega)\right] \\ Q_i^R[T(\omega)] &\equiv \left[\Pi_i(\omega)X_i^R(\omega) + [1-\Pi_i(\omega)]V_i^R(\omega)\right]. \end{aligned}$$

We need to establish that $V_i^k(\omega) = \int V_i^k(m;\omega)\, d\mu$ is continuous in ω. Let ω_n be a sequence in Ω converging to $\omega \in \Omega$. By Proposition B.3.1, $V_i^k(m;\omega_n)$ converges to $V_i^k(m;\omega)$ pointwise for all $m \in M$. Since $\mid V_i^k(m;\omega_n) \mid < \overline{U}/(1-\beta)$, by the Lebesgue Dominated Convergence Theorem the $\lim_n \int V_i^k(m;\omega_n)\, d\mu = \int V(m,\omega)\, d\mu$. Hence $V_i^k(\omega)$ is continuous in ω. An analogous argument establishes that $X_i^k(\omega)$ is continuous in ω. Finally, the continuity of $\Pi_i(\omega)$ follows from the continuity of $A_i[T(\omega)]$. Hence $Q_i^k[T(\omega)]$ is continuous in ω.

Since $T(\omega) : \Omega \to \Omega$ is continuous and Ω is compact, by Brouwer's FPT there exists ω^* such that $T(\omega^*) = \omega^*$ and the existence of at least one pure strategy MPE is assured.

C Proof of Theorem 2

If $\psi(g', g) \equiv 0$ then g is no longer payoff-relevant and so its value cannot affect equilibrium outcomes. Thus $A^*(g)$ and $G_k^*(g, m_\ell)$, $(j,\ell) \in \{(D,1),(R,2)\}$ are independent of g but, potentially, dependent on B. Assume that the former is $A^*(B)$ and the latter are $G_k^*(m_\ell, B)$. Then, the continuation value of party D when it is in power is

$$\Pr[A \geq A^*(B)]\{\mathbb{E}_{m_1} V_D^*(m_1', B) + \mathbb{E}_A[A|A \geq A^*]\} + [1 - \Pr[A < A^*]]\{\mathbb{E}_{m_2} X_D^*(m_2', B)\}.$$

Since this continuation value is independent of g, once party D is elected, the best it can do is solve its static optimization problem. Thus,

$$G_D^*(m_1, B) = \text{argmax}_{g' \in [0,\tau]} U(g', m_1) + \alpha U(\tau - g') + B,$$

which is evidently independent of B. Symmetrically, the R party will choose its statically ideal policy, independent of the value of B. One may verify that when the two parties act in this way, the net gain terms within square brackets in (22) are equal and so $A^*(B) \equiv 0$ and the probability of reelection is one-half regardless of B.

D Computation of Decision Rules

In this section we describe how, given k, i and ω, we compute the function $j_i^k(m; \omega) : M \to \mathcal{I}$. The following definition of *weakly preferred sets* is useful.

Definition 2 (Weakly Preferred Sets)**.** *Given k, i, m and ω, $P_{ij}^k(m; \omega) \subset \mathcal{I}$ is the weakly preferred set of j at m if and only if $j' \in P_{ij}^k(m; \omega)$ implies $V_{ij'}^k(m; \omega) \geq V_{ij}^k(m; \omega)$.*

The following lemma plays an important role in speeding up the computation.

Lemma D.0.1 (Dominated Choices)**.** *Let $\underline{m} \leq m_1 < m_2 \leq \overline{m}$. Let $j^* = j_i^k(m_1; \omega)$. Then any $j \in \mathcal{I} \setminus P_{ij^*}^k(m_2; \omega)$ is weakly dominated by j^* for all $m \in [m_1, m_2]$.*

Proof. Suppose there is some $m \in (m_1, m_2)$ for which there is an action $j_0 \in \mathcal{I} \setminus P_{ij^*}^k(m_2; \omega)$ such that $V_{ij_0}^k(m; \omega) > V_{ij*}^k(m; \omega)$. First, notice that j^* is always a member of $P_{ij^*}^k(m_2; \omega)$. Since $V_{ij^*}^k(m_1; \omega) \geq V_{ij_0}^k(m_1; \omega)$ (definition of j^*) and $V_{ij^*}^k(m_2; \omega) > V_{ij_0}^k(m_2; \omega)$ (by definition of j_0), it follows that there must be $\hat{m} \in [m_1, m)$ and another $\tilde{m} \in (m, m_2)$ for which $V_{ij_0}^k(\hat{m}; \omega) = V_{ij^*}^k(m; \omega)$ and $V_{ij_0}^k(\tilde{m}; \omega) = V_{ij^*}^k(\tilde{m}; \omega)$. But this contradicts Assumption 3. Hence, $V_{ij^*}^k(m; \omega) \geq V_{ij_0}^k(m; \omega)$ for all $m \in [m_1, m_2]$, with the equality holding, possibly, only at m_1. $\square$

The algorithm proceeds as follows. To begin, separately sort $V_{ij}^k(\underline{m}; \omega)$ and $V_{ij}^k(\overline{m}; \omega)$ with respect to j in descending order. Let $\underline{j}^*$ be the highest-ranked action in the first list and $\overline{j}^*$ be the highest-ranked action in the second list. Set $j_i^k(\underline{m}; \omega) = \underline{j}^*$.

The Initial Step:

Case 1: $\underline{j}^* = \overline{j}^* = j^*$. Then, by Lemma D.0.1, j^* strictly dominates all other actions for all $m \in (\underline{m}, \overline{m}]$. Hence, $j_i^k(m; \omega) = j^*$ for all $m \in (\underline{m}, \overline{m}]$ and we are done.

Case 2: $\underline{j}^* \neq \overline{j}^*$ and $P_{i\underline{j}^*}^k(\overline{m}; \omega)$ (the weakly preferred set of $\underline{j}^*$ at $\overline{m}$) contains only two elements. Then, use bisection to determine the unique $m_1 \in (\underline{m}, \overline{m})$ for which $V_{i\underline{j}^*}^k(m_1; \omega) - V_{i\overline{j}^*}^k(m_1; \omega) = 0$ and set

$$j_i^k(m; \omega) = \begin{cases} \underline{j}^* & \text{for } m \in (\underline{m}, m_1) \\ \min\{\underline{j}^*, \overline{j}^*\} & \text{for } m = m_1 \\ \overline{j}^* & \text{for } m \in (m_1, \overline{m}] \end{cases}$$

and we are done.

Case 3: $\underline{j}^* \neq \overline{j}^*$ and $P^k_{i\underline{j}^*}(\overline{m};\omega)$ contains $n \geq 3$ elements, denoted $\{\overline{j}^*, j_2, \ldots, j_{n-1}, \underline{j}^*\}$. Then, use bisection to determine the indifference points $\{m_{\overline{j}^*}, m_2, m_3, \ldots, m_{n-1}\}$ at which $V^k_{i\underline{j}^*}(m_s;\omega) = V^k_{i j_s}(m_s;\omega)$, $s \in \{\overline{j}^*, 2, 3, \ldots, n-1\}$. Let $\tilde{m}$ be the minimum of this set of indifference points and let $\tilde{j}$ be the corresponding action. Then,

$$j^k_i(m;\omega) = \begin{cases} \underline{j}* & \text{for } m \in (\underline{m}, \tilde{m}) \\ \min\{\underline{j}^*, \tilde{j}\} & \text{for } m = \tilde{m} \\ \in P^k_{i\underline{j}^*}(\overline{m};\omega) \setminus \underline{j}^* & \text{for } m \in (\tilde{m}, \overline{m}] \end{cases} .$$

The top branch follows because $\underline{j}^*$ is the best choice at $\underline{m}$ and $\tilde{m}$ is the *first* m for which some other choice, namely $\tilde{j}$, gives the same utility as $\underline{j}^*$; the middle branch follows from our tie-breaking convention; and the bottom branch follows because by Lemma D.0.1, $\tilde{j}$ dominates $\underline{j}^*$ for all $m > \tilde{m}$.

The Recursive Step:

If the algorithm reaches Case 3, it returns to the Initial Step with $\underline{m} = \tilde{m}$ and $\underline{j}^* = \tilde{j}$. Note that it is legitimate to treat $\tilde{j}$ as a best choice at $\tilde{m}$ because $\tilde{j}$ gives the same utility as $\underline{j}^*$ at $\tilde{m}$ and $\underline{j}^*$ strictly dominates every other choice at $\tilde{m}$ (recall, again, that $\tilde{m}$ is the *first* m for which an action indifferent to $\underline{j}^*$ is encountered). Each return to the Initial Step adds a new m segment of the decision rule. Also, with each return to the Initial Step there is at least one less action to evaluate (for instance, $P^k_{i\tilde{j}}(\overline{m};\omega)$ does not contain $\underline{j}^*$) so the algorithm is guaranteed to deliver the full decision rule in a finite number of steps.

Some remarks about the algorithm. First, for each k, i and ω, the algorithm requires two initial sorts of $V^k_i(m;\omega)$ — one for $m = \underline{m}$ and one for $m = \overline{m}$. For each subsequent return to the Initial Step, no further sorting is necessary because we know that $\tilde{j}$ is a best action at $\tilde{m}$ and, since $V^k_{i\overline{j}^*}(\overline{m};\omega)$ is already sorted, we merely need to locate the position of $\tilde{j}$ in the sorted vector to determine $P^k_{i\tilde{j}}(\overline{m};\omega)$.

Second, if $P^k_{i\underline{j}^*}(\underline{m};\omega)$ has n elements, the maximum number of thresholds calculated is $(n - 1 + n - 2 + \ldots + 1) = (n^2 - n)/2$. This is a maximum because a return to the Initial Step could eliminate more than one choice. For instance, an action that is in $P^k_{i\underline{j}^*} \setminus \{\underline{j}^*, \tilde{j}\}$ may not be in $P^k_{i\tilde{j}}$. In any case, the number of thresholds calculated grows polynomially in n.[36]

Third, it is possible to speed up the algorithm by utilizing a property of the model that, while somewhat special, may hold in other applications as well. The property is that V^k_{ij} can be expressed as a sum of two terms: one that depends monotonically on j and m only and a second term that depends on i and j but is independent of m. Specifically,

$$V^D_{ij}(m;\omega) = u(g_j + m) + B^D_{ij}(\omega) \quad \text{where} \quad B^D_{ij}(\omega) = \alpha u(\tau - g_j) - \eta(g_i - g_j)^2 + \beta Q^D_j[\omega]$$

and

$$V^R_{ij}(m;\omega) = u(\tau - g_j + m) + B^R_{ij}(\omega) \quad \text{where} \quad B^R_{ij}(\omega) = \alpha u(g_j) - \eta(g_i - g_j)^2 + \beta Q^R_j[\omega].$$

[36] The size n depends positively on the number of discrete choices available at each k, i and ω (generally, this depends on the grid size of the state space) and negatively on the width of the support of m (a narrow support means that $\underline{m} \approx \overline{m}$ and so the ranking of js for $V^k_{ij}(\underline{m};\omega)$ will be quite similar to the ranking for $V^k_{ij}(\overline{m};\omega)$).

Since $g_{i+1} > g_i$ (by assumption), the first component is strictly increasing in j for $k = D$ and strictly decreasing in j for $k = R$, regardless of any given value of m. Focusing for the moment on the $k = D$ case, this implies that for an action j to be not dominated by an action $j' > j$, $B^D_{i\,j}(\omega) > B^D_{i\,j'}(\omega)$. If this inequality is violated then j is strictly dominated by j' for all m and so can be dropped from further consideration. Thus, by examining the ordering of $B^k_{i\,j}(\omega)$ over j it is often possible to prune the set of choices the algorithm has to consider.

Finally, there is a property of $j^k_i(m;\omega)$ that holds for our model and which may hold in other applications as well. We do not use this property in the computation but its existence serves as a check on the results. This is the property of monotonicity of $j^k_i(m;\omega)$ with respect to m: For $m' > m$, $j^D_i(m;\omega) \leq j^D_i(m';\omega)$ and $j^R_i(m;\omega) \geq j^R_i(m';\omega)$. The proof follows easily from the concavity of u and we omit it.