Incumbency Disadvantage in U.S. National Politics: The Role of Policy Inertia and Prospective Voting

Satyajit Chatterjee and Burcu Eyigungor
Federal Reserve Bank of Philadelphia

December 2017

Abstract

We document that postwar U.S. national elections show a strong pattern of "incumbency disadvantage": If the presidency has been held by a party for some time, that party tends to lose seats in Congress. To generate this, we employ Alesina and Tabellini's (1990) model of partisan politics extended to have elections with prospective voting. We show that inertia in policies (combined with sufficient uncertainty in election outcomes) implies incumbency disadvantage. We find 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.

Keywords: rational partisan model, incumbency disadvantage, policy inertia, prospective voting, median voter

JEL Codes: D72 H50

Satyajit.Chatterjee@phil.frb.org. The authors thank Hülya Eraslan and seminar participants at Rice University for very helpful suggestions on an early draft of this paper. Comments from conference participants at the 2017 Society for Economic Dynamics Meetings, 2017 Stony Brook Summer Workshop in Political Economy, 2017 Midwest Macro Meetings at Pittsburgh and from seminar participants at Ohio State University are gratefully acknowledged.

1 Introduction

This paper is motivated by the following observation. 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. Based on this history, the probability that the U.S. presidency switches parties after being with any party for eight or more years is 88 percent (7/8). On the other hand, the fraction of years the Democratic party has held the presidency since 1954 is 0.45, implying that if the outcome of each election was an independent draw, we would expect the presidency to switch parties with a probability close to one-half in every election. These two facts suggest that there is an *incumbency disadvantage* in U.S. national politics: When any party has held the presidency for two or more terms, the popularity of the party with voters is strongly diminished.[1]

We make several contributions. First, we verify that the suggestion of incumbency disadvantage noted above is in fact present in the electoral performance of the two parties in the postwar era. Specifically, we show that the Democratic lead in the House is strongly affected by how long the two parties have held the presidency going into each election.[2] For instance, if the Democratic (Republican) party has held the presidency for six or more years going into an election, the Democratic lead in the House declines (increases) by 38 seats, on average. This finding persists if we measure electoral performance with the Democratic House vote share lead instead of the Democratic House lead: The House Democratic vote share lead declines (rises) by 6 percentage points going into an election in which the Democrats (Republicans) have held the presidency for six or more years. These effects persist in roughly the same magnitude when detrended values of the electoral performance measures are used.

Second, we show that incumbency disadvantage is implied by the time-consistent equilibrium of Alesina and Tabellini's (1990) model of partisan politics, if their model 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,

[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] Since every seat in the House is generally contested in every national election (held every two years), the scope of the electorate to express approval or disapproval of current policies is the greatest in House elections.

and, second, there is policy inertia stemming from costs of (or constraints on) changing policies quickly. These two features imply that after a long Democratic incumbency (which can happen randomly), inherited policies will be relatively far from a Republican voter's preferred policies. Given diminishing marginal utility, the utility loss to a Democratic voter from a relatively small change in policy (the change is small due to inertia) toward those preferred by Republicans will be smaller in magnitude than the utility gain for a Republican from the same change in policy. As a result, preference shocks will be less determinative for Republicans, and they are more likely to vote along partisan lines and win the election. The same logic applies in reverse after a long Republican incumbency.

Third, we show that the policy inertia needed to account for incumbency disadvantage has important positive and normative implications. On the positive side, each party's long-run ideal policy may become more extreme than the incumbent party member's static ideal policy; i.e., the policy that maximizes the party member's period utility. This is significant because when such an extreme policy is followed, a less extreme policy would improve *all* voters' period utility. Since inertia leads to incumbency disadvantage, parties target more extreme policies so that they can enjoy policies closer to their ideal even when the opposition is calling the shots. On the normative side, this extremism may entail a welfare loss for voters.

Fourth, we show that policy inertia and prospective voting enlarge the domain of application of the median voter and probabilistic voting paradigms wherein policies are selected via Downsian competition between two parties that care about winning the election.[3] As Alesina (1988) first pointed out, if the two parties have their own agendas and neither party can commit to policies before the election, then in a Markov equilibrium, the winning party will implement its preferred, not median, policy *regardless* of its desire to hold office (i.e., regardless of the intensity of office motivation).[4] In our model as well, without inertia and pre-election commitment, the Markov equilibrium will have parties choosing their preferred policies regardless of their desire to hold office; i.e., there will be no movement toward centrist policies stemming from a desire to win and retain elected office. But policy inertia and prospective voting changes this result and office motivation begins to matter for policy choice: Choosing policies closer to the voters' ideal policies

[3]For recent examples of median voter and probabilistic voting in quantitative macroeconomics, see Mateos-Planas (2012) and Song, Storesletten, and Zilibotti (2012)

[4]If parties can commit to a platform before the election, the Downsian outcome becomes possible even with candidates who compete because they care about policies (Wittman (1983), Calvert (1985)). The Downsian outcome can also be resurrected if purely ideologically motivated parties play a repeated game with trigger strategies, as shown in Alesina (1988).

means a higher chance of losing the next election and, thus, a higher chance of losing the utility from being in office. The additional private cost of extremism pulls equilibrium policies closer to the center. Thus, the median voter logic remains active even when parties are ideologically motivated and there is no pre-election commitment.

Finally, our paper makes a methodological contribution by augmenting the toolkit of researchers using quantitative theory to study partisan politics with endogenous reelection probabilities. Such games can be hard to compute because of the possible lack of continuity of Markovian decision rules (Chatterjee and Eyigungor (2016)) and, consequently, the possible nonexistence of a pure strategy equilibrium. By incorporating a continuously distributed i.i.d. preference shock, we ensure the existence of a Markov perfect equilibrium in pure strategies despite equilibrium decision rules being potentially discontinuous. The existence result puts our model — and other models that share key features of our environment — on a secure computational footing. To this end, we also offer an algorithm to compute the (potentially discontinuous) Markovian decision rules.[5]

Our paper is related to several literatures. On the empirical side, our finding of incumbency disadvantage echoes 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, and the same remains true for the 14 presidential elections since 1960. They interpret their findings as evidence of "restoring forces" that work to elevate the popularity of the party that has been less popular in the past. More recently, Bartels and Zaller (2001) study a large set of empirical presidential vote models and identify an "incumbent fatigue" effect 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 are similar to theirs, except that we focus directly on the electoral performance of the two parties in all House elections rather than the two-party presidential vote share. Relatedly, Alesina and Rosenthal (1995)

[5] The potential lack of continuity of decision rules and, therefore, of the mapping whose fixed points are pure strategy Markov equilibria, also plagues models of sovereign debt and default (Chatterjee and Eyigungor (2012)) and models of legislative bargaining (Duggan and Kalandrakis (2012)). In both instances, the existence problem is solved by the addition of continuously distributed random shocks in the appropriate places. However, the existence proofs given in these papers cannot be used to claim existence for the present model since a key element of the present model — endogenous reelection probability — is missing in these earlier papers. In Duggan and Kalandrakis (2012), proposer election probabilities are an *exogenous* function of the state and in Chatterjee and Eyigungor (2012) there is no political turnover of the sovereign at all.

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.

In the area of macro political economy, we contribute to the growing literature on quantitative-theoretic partisan political economy models featuring endogenous reelection probabilities. In terms of model structure and quantitative focus, the closest is Azzimonti (2011). We borrow from her our model of elections wherein people vote prospectively subject to preference shocks. However, while the reelection probability is endogenous in her model, her assumption regarding preferences (and the symmetry of the model) implied that the probability of reelection is *state independent*. Thus, by construction, her model is silent on how reelection probabilities evolve over time. Prospective voting also features importantly in Alesina and Rosenthal's (1995) explanation of the mid-term cycle. Their explanation centered around non-polarized voters attempting to get closer to median policies by counterbalancing a partisan president's power. In our approach, 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, we predict an electoral disadvantage that grows with incumbency. Prospective voting and endogenous reelection probability also features in Rogoff's (1990) model of the political budget cycle. The friction in his model is asymmetric information about the competency of the incumbent leader and the cycle arises from the competent type choosing policies that separate her from the incompetent type. There are cycles in taxes and expenditure policies but not necessarily incumbency disadvantage.

Prospective voting and endogenous reelection probabilities have recently featured in quantitative models of sovereign debt and default.[6] Scholl (2017) uses a version of the Persson and Svensson's (1989) partisan politics model to quantitatively explore the implications of endogenous reelection probabilities on sovereign borrowing and default behavior, and Chatterjee and Eyigungor (2017) study the role of prospective voting in accounting for observed fluctuations in the risk of sovereign default in emerging economies. But incumbency disadvantage is not a necessary feature of these models.

Endogenous political turnover is a feature of Ales, Maziero, and Yared's (2014) model of political cycles. These authors approach political turnover as the outcome of a principal/agent contracting problem (citizens as principal and the government as agent). Because the agent (government) has

[6] Aghion and Bolton (1990) is an early precursor that analyzed the role of default risk in a stylized partisan political economy model.

private information about the state of the budget, the optimal contract has the feature that after a sequence of bad outcomes the contract is terminated and a new contract is entered into with a new agent. Replacement of the government is, thus, endogenous and the logic is closely tied to those in models of political control and retrospective voting (Barro (1973) and Ferejohn (1986)). But there is no incumbency disadvantage.

Cycles in expenditure tied to exogenous shifts in parties occur in Dixit, Grossman, and Gul (2000), Acemoglu, Golosov, and Tsyvinski (2011) and Aguiar and Amador (2011). In contrast to us, these authors study self-enforcing equilibria supported by trigger strategies. Cuadra and Sapriza (2008) and Hatchondo, Martinez, and Sapriza (2009), examine the implications of exogenous political turnover for sovereign borrowing and default. In comparison to these sets of studies, the probability of turnover is endogenous in our model.

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. 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 because it endogenizes reelection probabilities. In contrast, in Bowen, Chen, and Eraslan's (2014) environment (and in environments of legislative bargaining models in general) the probability of a legislator being chosen to propose a policy is exogenously given and does not depend on the policies the legislator is expected to propose.

The paper is organized as follows. In Section 2, we present our empirical findings regarding incumbency disadvantage. In Section 3, we develop the model outlined previously. In Section 4, we analyze a static version of the model to provide intuition for why policy inertia leads to incumbency disadvantage. In Section 5, we explore the full dynamic model computationally — the goal of this section is to show that analogs of the empirical findings reported in Section 2 can arise in the full equilibrium of the model and to understand the key factors that lead to this result. In Section 6 and 7 we examine the positive and normative implications of policy inertia and incumbency disadvantage, and Section 8 concludes. The Appendix gives the proof of existence of pure strategy Markov perfect equilibrium for this class of models and gives an algorithm to compute the (potentially discontinuous) Markovian decision rule of the dynamic game.

2 Incumbency Disadvantage in U.S. National Elections: 1954-2016

In this section we show that the incumbency disadvantage mentioned in the opening paragraph manifests itself in the outcome of House elections. Examining House elections allows us to include all national elections, as opposed to only presidential elections. We examine the electoral performance of the two parties when the parties have held the presidency for specified lengths of time – six or more years and eight or more years. Our aim is to establish that a long presidential incumbency of a party leads to worse electoral outcomes for the party.[7]

We measure the electoral performance of parties in House elections by the Democratic lead in the House (which can range from 0 to 435), denoted DL, and by the Democratic vote share lead in House elections (which, in percentage points, can range from -100 to 100), denoted DVL. Figure 1 shows the time variation in these measures for all the election years since 1954.

Our main empirical specification is:

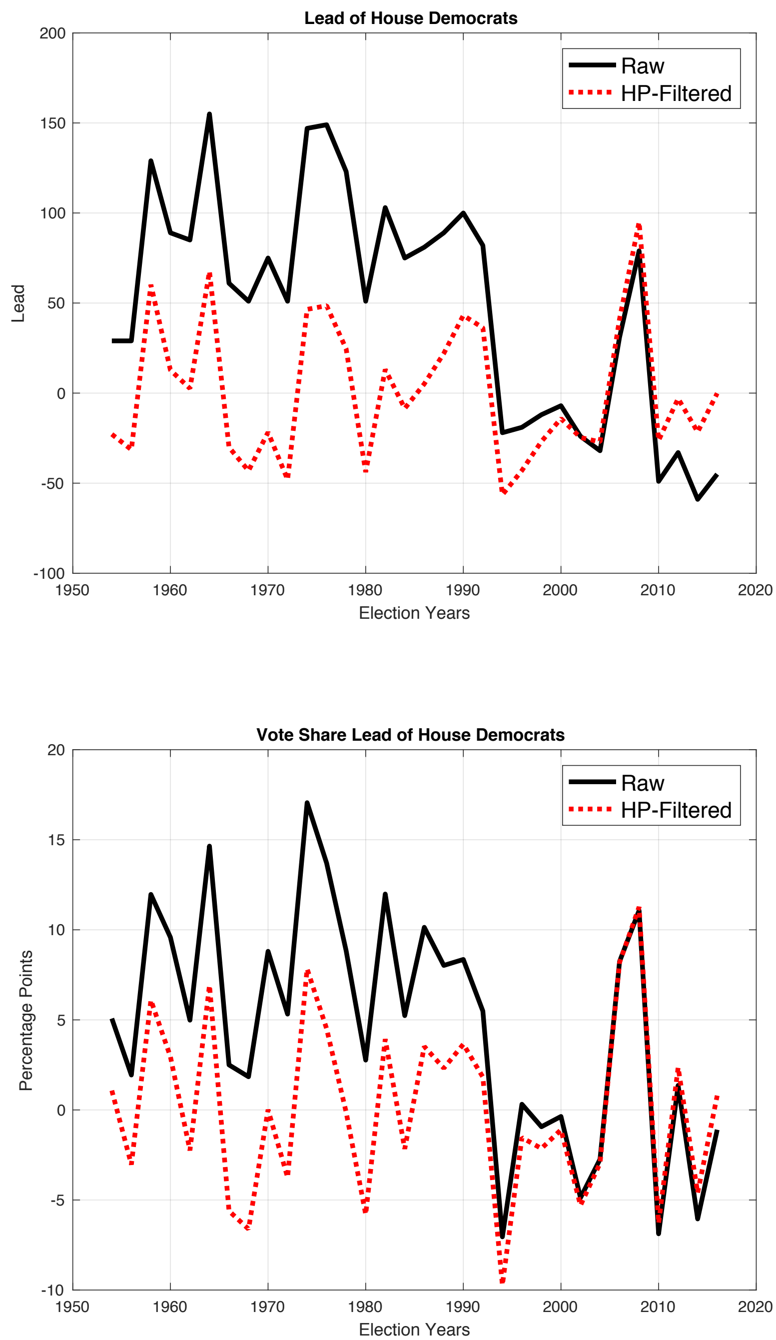
$$DL_t \text{ or } DVL_t = \beta_0 + \beta_1 SIX_t^+ + \beta_2 (TWO_t^+ \cdot GR_t).$$

Here SIX_t^+ is a trichotomous variable that takes a value of 1 if, at the time of 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, we expect β_1 to be negative. A negative coefficient implies that, after 6 years of a Democratic presidency, the Democratic lead in the House falls, and, after 6 years of a Republican presidency, the Democratic lead in the House increases. We also experiment with $EIGHT^+$ which takes the value 1 if the presidency has been held by Democrats for eight or more years and -1 if it has been held by Republicans for eight or more years and 0 otherwise.

$TWO_t^+ \cdot GR_t$ is a control 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 it takes the value -1 if the presidency was held by a Republican in the preceding two or more years.

[7] In a presidential system such as in the U.S., the president's party gets to set the policy agenda and, so, during elections, we expect voters to vote against (or for) the members of the president's party if they disapprove (or approve) of the party's current policies.

Figure 1:
Relative Electoral Performance of the Democratic Party, 1954-2016



And, GR_t is the average HP detrended (with smoothing parameter equal to 100) growth rate of real GDP from the third quarter of $t-2$ to the third quarter of t (because elections are held in the fourth quarter). This interaction term takes into account that above-trend economic performance in the inter-election period may be attributed to the success of policies of the presidential party and, so, the presidential party gains more seats. If so, we expect β_2 to be positive.

Tables 1 and 2 present our estimation results. The results pretty much speak for themselves. Regardless of the measure of electoral performance or the specific length of duration of the presidential incumbency of a party, a long incumbency is costly for the party. On average, if the party has held the presidency for 6 or more years, it loses about 54 seats on a nondetrended basis and about 37 seats on detrended basis. For the vote share lead, a long incumbency leads to a 5 to 6 percentage point decline in the vote share lead on a nondetrended basis and a 4 to 5 percentage point decline on a detrended basis. Although this is not the focus of our paper, our results confirm the common finding that good economic performance has a salutatory effect on the election success of the incumbent party. The statistical significance of all estimated coefficients is high.

Table 1:
Presidential Incumbency of a Party and House Dem. Lead

Dep. Var.		DL	
Constant	$37.6^{**}(17.9)$	$36.2^{**}(16.6)$	$38.7^{**}(17.7)$
SIX^+		$-54.1^{***}(12.2)$	
$EIGHT^+$			$-52.8^{***}(13.0)$
$TWO^+ \cdot GR$	$16.4^{**}(7.5)$	$14.6^{**}(7.0)$	$12.9^{*}(7.0)$
Adj. R^2	0.28	0.39	0.22
S.D. Dep Var	63.2	63.2	63.2
Dep. Var. (detrended)	*DL*	*DL*	*DL*
Constant	$-7.4(6.2)$	$-8.8^{*}(5.1)$	$-7.3(5.5)$
SIX^+		$-37.1^{***}(6.3)$	
$EIGHT^+$			$-37.3^{***}(8.2)$
$TWO^+ \cdot GR$	$11.5^{***}(2.5)$	$10.6^{***}(2.2)$	$9.5^{***}(3.2)$
Adj. R^2	0.31	0.54	0.33
S.D. Dep Var	38.4	38.4	38.4

Table 2:
Presidential Incumbency of a Party and House Dem. Vote Share Lead

Dep. Var.	DVL	DVL	DVL
Constant	$3.5^{**}(1.4)$	$3.3^{**}(1.3)$	$3.7^{**}(1.5)$
SIX^+		$-6.1^{***}(0.8)$	
$EIGHT^+$			$-5.0^{***}(1.0)$
$TWO^+ * GR$	$1.7^{**}(0.6)$	$1.6^{**}(0.6)$	$1.4^{**}(0.6)$
Adj. R^2	0.31	0.48	0.19
S.D. Dep Var	6.4	6.4	6.4

Dep. Var. (detrended)	DVL	DVL	DVL
Constant	$-0.9(0.7)$	$-1.1^{*}(0.6)$	$-0.8(0.7)$
SIX^+		$-4.8^{***}(1.0)$	
$EIGHT^+$			$-3.9^{***}(1.1)$
$TWO^+ \cdot GY$	$1.4^{***}(0.3)$	$1.3^{***}(0.2)$	$1.1^{***}(0.4)$
Adj. R^2	0.27	0.52	0.21
S.D. Dep Var	5.0	5.0	5.0

3 Model

We build on 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 and policies change inertially. We denote one party by D and the other by R. We assume that tax revenues are constant and given by $\tau > 0$. We use $g \in [0, \tau]$ to denote the type of expenditures preferred more by the D party and $(\tau - g) \in [0, \tau]$ as the type of expenditure that is more preferred by the R party. Alternatively, g may be interpreted as the ideological stance of policies in which case τ represents the D end of the ideological spectrum and 0 represents the R end. In what follows, g denotes the expenditures/policies inherited from the previous period and g' denotes the expenditures/policy chosen in the current period.

Each period starts with an election. Given the inherited policy g and an i.i.d. aggregate voter preference shock $A \in \mathbb{R}$ (the shock is explained in more detail below), let $\mathbb{1}(g, A)$ be the indicator function that takes the value 1 if the election is won by the D party and 0 if it is won by the R party. Following an election, the party in power experiences an i.i.d. shock $m \in [\underline{m}, \overline{m}]$ that negatively affects the marginal utility from its preferred good (the higher the shock, the less urgency there is to spend on the party's preferred good). The party in power then chooses g' and all period payoffs are realized.

For $j \in \{D, R\}$, let V_j denote the value of party j when it is in power and by X_j its value when it is not in power. Then,

$$V_D(g,m) =$$

$$\max_{g'\in[0,\tau] \text{ and } g'+m\geq 0} U(g'+m) + \alpha U(\tau - g') - \eta\left(g'-g\right)^2 + \beta \mathbb{E}_{(A',m')}\begin{bmatrix} \mathbb{1}\left(g',A'\right)V_D(g',m') + \\ \left[1-\mathbb{1}\left(g',A'\right)\right]X_D(g',m') \end{bmatrix} \tag{1}$$

Here, $U(c)$ is a continuous, strictly increasing and strictly concave function defined over $c \geq 0$, $0 \leq \alpha < 1$ and $\eta\left(g'-g\right)^2$, $\eta > 0$, is the cost of adjusting policies. Letting $G_R(g,m)$ denote the policy of party R when it is in power, party D's value when not in power is

$$X_D(g,m) = U(g') + \alpha U(\tau - g') - \eta\left(g'-g\right)^2 + \beta \mathbb{E}_{(A',m')}\begin{bmatrix} \mathbb{1}\left(g',A'\right)V_D(g',m') + \\ +[1-\mathbb{1}(g',A')]X_D(g',m') \end{bmatrix} \tag{2}$$

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

When the D party is not in power, it does not make any choices but must live with the choices made by the R party.[8]

Symmetrically,

$$V_R(g,m) =$$

$$\max_{g'\in[0,\tau] \text{ and } \tau-g'+m\geq 0} \alpha U(g') + U(\tau - g' + m) - \eta\left(g'-g\right)^2 + \beta \mathbb{E}_{(A',m')}\begin{bmatrix} \mathbb{1}\left(g',A'\right)X_R(g',m') + \\ \left[1-\mathbb{1}\left(g',A'\right)\right]V_R(g',m') \end{bmatrix}, \tag{3}$$

[8]The preference shock m does not enter directly into the expressions for X_D but it enters indirectly through how the shock affects the choices made by the R party.

and

$$X_R(g,m) = \alpha U(g') + U(\tau - g') - \eta \left(g' - g\right)^2 + \beta \mathbb{E}_{(A',m')} \left[\begin{array}{c} \mathbb{1}\left(g', A'\right) X_R(g',m') + \\ \left[1 - \mathbb{1}(g', A')\right] V_R(g',m') \end{array} \right] \quad (4)$$

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

We assume that $\tau + \underline{m} > 0$ so that the choice sets for both parties are nonempty for all $m \in [\underline{m}, \overline{m}]$. Also, for our base case we assume that the adjustment costs of changing inherited policies are borne equally by both parties. Later, we consider the case where adjustment costs are borne only by the party in power and the case where there are no adjustment costs but there is a constraint on how large policy changes can be in any period.

Next, we turn to the decision problem of voters that provides the underpinning of the election outcome function $\mathbb{1}(g, A)$. There is a continuum of voters. In terms of utility from g' and $\tau - g'$, half of them have the same preferences as the D party and the other half have the same preferences as the R party. Voters differ from their affiliated party in two ways. First, they do not experience the costs of changing policies borne by their representatives in government. Second, as mentioned earlier, voters experience an aggregate preference shock $A \in \mathbb{R}$ that is i.i.d. over time. In addition, voters also experience an idiosyncratic preference shock that is independent across voters and also i.i.d. over time, which we denote by $e \in \mathbb{R}$. It is assumed that $\mathbb{E}A = 0$ and $\mathbb{E}e = 0$ and that both shocks are symmetrically distributed around their means and that both affect, additively, the utility that a voter gets from party D being in power.[9]

For $j \in \{D, R\}$, let W_j be the value function of voters belonging to party j when their party is in power, and let Z_j be their value function when their party is out of power. Then,

$$W_D(g,m,e,A) = U(g') + \alpha U(\tau - g') + e + A + \beta \mathbb{E}_{(e',A',m')} \left[\begin{array}{c} \mathbb{1}\left(g', A'\right) W_D(g',m',e',A') + \\ \left[1 - \mathbb{1}\left(g', A'\right)\right] Z_D(g',m') \end{array} \right] \quad (5)$$

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

[9] The fact that these shocks affect preferences for the D party is without any loss of generality. As will become clear below, voter preference for the R party will occur for negative values of e and A.

and

$$Z_D(g,m) = U(g') + \alpha U(\tau - g') + \beta \mathbb{E}_{(e',A',m')} \left[\begin{array}{c} \mathbb{1}(g',A')\, W_D(g',m',e',A') \,+ \\ \left[1 - \mathbb{1}(g',A')\right] Z_D(g',m') \end{array} \right] \tag{6}$$

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

The above recursions can be written more compactly. Observe that (5) implies $W(g,m,e,A) = W(g,m,0,0) + e + A$. Denoting $W(g,m,0,0)$ by $W(g,m)$ and then substituting $W(g,m) + e + A$ for $W(g,m,e,A)$ in (5) - (6) and using $\mathbb{E}_{(e',A',m')} e' = \mathbb{E}_{e'} e' = 0$ (the first equality follows from independence of the shocks and the second from $\mathbb{E}e = 0$) yields:

$$W_D(g,m) = U(g') + \alpha U(\tau - g') + \beta \mathbb{E}_{(A',m')} \left[\begin{array}{c} \mathbb{1}(g',A')\left[W_D(g',m') + A'\right] \,+ \\ \left[1 - \mathbb{1}(g',A')\right] Z_D(g',m') \end{array} \right] \tag{7}$$

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

and

$$Z_D(g,m) = U(g') + \alpha U(\tau - g') + \beta \mathbb{E}_{(A',m')} \left[\begin{array}{c} \mathbb{1}(g',A')\left[W_D(g',m') + A'\right] \,+ \\ \left[1 - \mathbb{1}(g',A')\right] Z_D(g',m') \end{array} \right] \tag{8}$$

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

Proceeding symmetrically,

$$W_R(g,m) = \alpha U(g') + U(\tau - g') + \beta \mathbb{E}_{(A',m')} \left[\begin{array}{c} \mathbb{1}(g',A')\left[Z_R(g',m') + A'\right] \,+ \\ \left[1 - \mathbb{1}(g',A')\right] W_R(g',m') \end{array} \right] \tag{9}$$

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

and

$$Z_R(g,m) = \alpha U(g') + U(\tau - g') + \beta \mathbb{E}_{(A',m')} \left[\begin{array}{c} \mathbb{1}(g',A')\left[Z_R(g',m') + A'\right] \,+ \\ \left[1 - \mathbb{1}(g',A')\right] W_R(g',m') \end{array} \right] \tag{10}$$

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

At the time of an election, voters compute their individual net gain from voting for each party and vote for the party for which this net gain is nonnegative. Since the value of m is realized after the election, the individual net gain to a D party member from voting for his own party is $\mathbb{E}_m W(g,m) + e + A - \mathbb{E}_m Z_D(g,m)$ and the individual net gain for an R party member from voting for the D party is $\mathbb{E}_m Z_R(g,m) + e + A - \mathbb{E}_m W_R(g,m)$. Given (g, A), these expressions determine thresholds for the idiosyncratic shock, $e_j(g,A)$, $j \in \{D,R\}$, above which a j party member will vote for the D party in the election. Specifically,

$$e_D(g,A) = -[\mathbb{E}_m W_D(g,m) - \mathbb{E}_m Z_D(g,m)] - A \tag{11}$$

and

$$e_R(g,A) = [\mathbb{E}_m W_R(g,m) - \mathbb{E}_m Z_R(g,m)] - A. \tag{12}$$

In these threshold expressions, the terms in square brackets represent the expected net gain to members of a party from their own party coming into power, ignoring all preference shocks. Holding fixed A, the bigger the expected net gain term in (11), the lower the threshold e_D and the bigger the expected net gain in (12), the higher the threshold e_R. Hence, the larger these expected net gain terms, the more partisan voting will be. In contrast, an increase in A lowers *both* thresholds and increases the fraction of voters in favor of the D party.

Theorem 1. *Given W_j and Z_j, $j \in \{D,R\}$, the election outcome function $\mathbb{1}(g,A)$ has the following step-function form:*

$$\mathbb{1}(g,A) = \begin{cases} 1 & \text{if } A \geq A(g) \\ 0 & \text{otherwise}. \end{cases} \tag{13}$$

where

$$A(g) = \frac{1}{2}\left\{[\mathbb{E}_m W_R(g,m) - \mathbb{E}_m Z_R(g,m)] - [\mathbb{E}_m W_D(g,m) - \mathbb{E}_m Z_D(g,m)]\right\}. \tag{14}$$

Proof. By definition of $e_j(g,A)$, $j \in \{D,R\}$, the probability that a randomly selected voter will cast his vote in favor of the R party is $(1/2)F(e_D(g,A) + (1/2)F(e_R(g,A)$. In what follows, we will assume that this expression also gives the *fraction* of voters who vote for the R party, given

(g, A).[10] Since this expression is decreasing in A, it follows that there is a unique value of A such that exactly half of the voters vote for the D party. This cut-off value of the aggregate shock, denoted $A(g)$, solves:,

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

or,

$$F(e_D(g, A(g)) = 1 - F(e_R(g, A(g)). \tag{15}$$

Since e is symmetrically distributed around 0, we may infer that at $A(g)$, e_D must equal e_R in magnitude, i.e., $e_D(g, A(g)) + e_R(g, A(g)) = 0$. Using (11) and (12) then gives:

$$A(g) = \frac{1}{2}\left\{[\mathbb{E}_m W_R(g,m) - \mathbb{E}_m Z_R(g,m)] - [\mathbb{E}_m W_D(g,m) - \mathbb{E}_m Z_D(g,m)]\right\}.$$

Thus, we have

$$\mathbb{1}(g, A) = \begin{cases} 1 & \text{if } A \geq A(g) \\ 0 & \text{otherwise}. \end{cases}$$

□

Theorem 1 implies that $\mathbb{E}_A \mathbb{1}(g, A) = \text{Prob}[A \geq A(g)]$. Thus, if the two expected net gain terms in square brackets in (14) are equal in value then $A(g) = 0$ and, from the symmetry of the A distribution, the probability of any party winning the election is one-half. If the net gain term for D- party members is smaller than the net gain term for R-party members, then $A(g) > 0$ and (from the symmetry of the A distribution) the probability of a D-party win is less than one-half.

It is worth commenting on the role played by the idiosyncratic shock e in the derivation of the expression for $A(g)$. To arrive at the recursions (7)-(10) that determine $\{W_j, Z_j\}$, the assumptions needed on e were that e is independent of all other shocks and that $\mathbb{E}e = 0$. And the assumption needed on e to go from (15) to the expression for $A(g)$ is that e is distributed symmetrically around

[10] It is well understood that there is no "law of large numbers" that ensures this identification of probabilities as fractions, when we are dealing with a continuum of voters (Judd (1985)). However, for our application it is fine to simply assume that the "law" holds (see Feldman and Gilles (1985) and Uhlig (1996)).

its mean and has unbounded support. Aside from these properties, the precise nature of the e distribution does not matter for the determination of $A(g)$.[11]

The equilibrium of the model is defined as follows:

Definition 1. *A Markov Perfect Equilibrium (MPE) is a collection of party value and policy functions $V_j^*(g,m)$, $X_j^*(g,m)$, $G_j^*(g,m)$, a collection of voter value functions $W_j^*(g,m)$, $Z_j^*(g,m)$, and an election outcome function $\mathbb{1}^*(g,A)$ such that (i) given $G_{\sim j}^*(g,m)$ and $\mathbb{1}^*(g,A)$, the functions $V_j^*(g,m)$, $X_j^*(g,m)$ and $G_j^*(g,m)$ solve* (1) - (2) *for $j = D$ and* (3) - (4) *for $j = R$; (ii) given $G_j^*(g,m)$ and $\mathbb{1}^*(g,A)$, the functions $W_j^*(g,m)$ and $Z_j^*(g,m)$ solve* (7) - (8) *for $j = D$ and* (9) - (10) *for $j = R$; and (iii) given $W_j^*(g,m)$ and $Z_j^*(g,m)$, $j \in \{D,R\}$, the function $\mathbb{1}^*(g,A)$ solves* (13) - (14)*.*

In Section A, we establish the existence of MPE for a class of models which includes the base model (and other models discussed later in the paper). In light of this, we have:

Theorem 2. *If g (and, therefore, g') is restricted to lie in a discrete subset of $[0,\tau]$ then under mild regularity conditions, an MPE exists.*

4 Policy Inertia and Incumbency Disadvantage in a Static Model

The basic idea underlying the key result of this paper — incumbency disadvantage — can be best illustrated in a static model. Imagine that the economy has arrived into a "last" period with some g. Our goal is to understand how the probability of a party winning the election varies with g.

For this illustration, we ignore the m shock and assume that $\alpha = 0$ and that $U(g') = -(\tau - g')^2$. Hence, $U(\tau - g') = -(\tau - (\tau - g'))^2 = -(g')^2$. Thus, the D party's ideal g' is τ, and the R party's ideal g' is 0. Here g is best interpreted as an ideological stance, with τ being the liberal end and 0 being the conservative end.

If the D party wins the election, it will choose g' to maximize $-(\tau - g')^2 - \eta(g - g')^2$ subject to $g' \in [0,\tau]$. This maximization implies

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

[11] This results from the feature that all payoffs depend on the identity of the party in power only and not on the party's margin of victory in the election.

The optimal decision is, thus, a convex combination of the D party's ideal policy, τ, and the inherited policy g. We can verify that if the R party were to win, then

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

This optimal decision is also a convex combination of the R party's ideal policy, 0, and the inherited policy g. Thus, the lower η is, i.e., the smaller is the adjustment cost, the closer g' is to the elected party's ideal policy.

The net gain to D party members from electing their own party over the R party is

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

and the gain to R party members from electing their own party over the D party is

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

Then

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

and

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

Thus the threshold level of A above which the D party wins (given by the A value for which $e_D(g, A) + e_R(g, A) = 0$) is:

$$A(g) = \frac{2\eta\tau}{(1+\eta)^2} \left(g - \tau/2\right) . \tag{16}$$

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, then $A(g)$ is positive, which means that the probability of the D party winning the election is *less* than 1/2. Furthermore, the probability of a D-party victory falls with g. These results stem from the fact that the symmetry of adjustment costs makes it equally costly to move away from the inherited policy in either direction (toward τ or toward 0). However, if g is closer to τ than 0, then, by diminishing marginal utility, the expected gain to members of the D party from electing their own party will be smaller than the expected gain to members of the R party from electing their own party. This same logic applies in reverse if g is closer to 0 than to τ: the likelihood of the R party winning the election would then be less than 1/2.

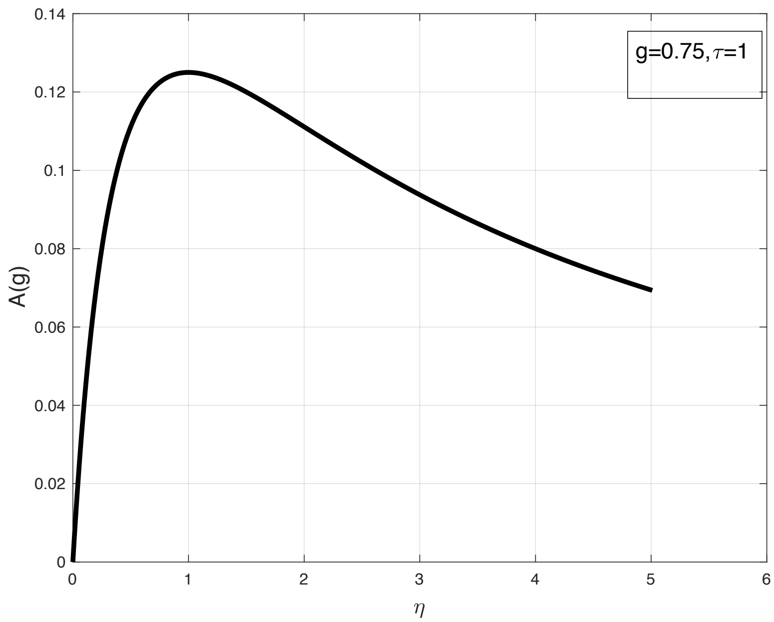
The result that the likelihood of a D party victory falls with g is echoed in how the fraction of people voting for the D party varies with g. As can be verified by direct differentiation, the rate of change of both $e_D(g, A)$ and $e_R(g, A)$ with respect to g is given by $2[\eta/(1+\eta)^2]\tau$, which is strictly positive for $\eta > 0$. Since the fraction voting for the D party is $1-(1/2)[F(e_D(g, A))+F(e_R(g, A))]$, the fraction declines with g.

These results are consistent with an incumbency disadvantage if we expect that g will be closer to τ (to 0) when the D party (R party) has been in power for some time.

The expression in (16) shows that $A(g)$ will be zero and, so, there will be no incumbency disadvantage (with respect to likelihood of a win) under two sets of circumstances. First, $A(g) = 0$ if $g = \tau/2$. This is the case where the inherited g is halfway between the ideal choices of the two parties, and so, given the symmetry in preferences and adjustment costs, the expected net gain from electing one's own party is the same for D and R party members.

Second, $A(g) = 0$ 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. Again, given the symmetry in preferences, the net gain from electing one's own party is the same for D and R party members. If $\eta = \infty$, adjustment costs are infinite so $g' = g$ regardless of which party is elected. The expected net gain from electing one's own party over the other party is thus 0, and therefore the same, for members of both parties. These properties suggest that when $g > \tau/2$, the functional relationship between $A(g)$ and η has an inverted-U shape as shown in Figure 2 (the figure assumes $\tau = 1$ and $g = 0.75$). As η increases from zero, even if the R party comes into power, the resulting policies remain relatively close to what D party members prefer and, so, D party members are more likely to defect and vote for the R party and, hence, there is an increase in $A(g)$. As η increases further, policies respond less to a change in leadership and cross-voting by members of both parties increases. However, the effect is greater for R party members (since initially there were not cross voting very much), which works to mute the increase in $A(g)$. For this model $A(g)$ peaks at $\eta = 1$ and then begins to decline.

Figure 2:
Cost of Adjustment and the A Cut-off for a D Party Win



5 Incumbency Disadvantage in the Dynamic Model

The static model discussed in the previous section gave the basic intuition for why incumbency disadvantage arises in our model. The result depends on voters recognizing that the winning party will move policy toward its own ideal and that voters whose ideal policy is closer to the inherited g don't care as much about this movement (and, therefore, are more likely to be swayed by other idiosyncratic factors in their voting). But since the game ended after g' was chosen, neither party took into account that by moving policy toward its ideal policy it might adversely affect its chances of reelection in the *future*. When elections happen every period, it is not obvious that an incumbent party will act in a way that generates a disadvantage for itself in future elections.

In this section our goal is to show not only that incumbency disadvantage can arise in the dynamic model but to show that the model is capable of delivering the observed *magnitude* of this effect. Given the quantitative nature of our goal, the demonstration proceeds by choosing a realistic parametrization 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 $U(g) = g^{1-\gamma}/[1-\gamma]$, the idiosyncratic shock e has a normal distribution with 0 mean, and the aggregate shock A as well as the party preference shock m have zero mean uniform (hence, symmetric around 0) distributions. Turning to parameter values, since national elections happen every two years, the value of β is set to 0.92, which corresponds to an biennial discount rate of 8 percent. The value of γ is set to 2, a standard value in the macroeconomics literature. The value of τ is normalized to 1. 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, have important effects on the magnitude of the incumbency disadvantage in the model. Of these, the dispersion of m is special in that its size also governs the speed with which we can converge to an equilibrium in our computations. Conditional on the settings of the other parameters, the speed drops as the standard deviation of m is lowered. Experimentation showed that a standard deviation of around 3 percent is large enough to ensure quick convergence to an equilibrium for a range of parameter values. Thus the support of m was set to ± 0.05.

Conditional on the setting of the dispersion of m and the parameters listed earlier, experimentation also 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 to be consistent with the observed magnitude of incumbency disadvantage, we chose the one in which η is 5, 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 3.

To confirm that the model generates incumbency disadvantage, Table 4 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 lead (i.e., the percentage of the total vote cast in favor of the D party minus and the percentage of votes cast for the R party) 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, or the R party has been power for 3 or more (or 4 or more) model periods, or neither. For comparison purposes,

Table 3: Parameter Selections

Parameter	Description	Value
γ	Curvature of utility function	2.00
β	Biennual 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, $\pm\overline{m}$	0.05
η	Adjustment cost parameter	5
σ_e	S.D. of idiosyncratic voter preference shock	0.02
$\overline{A}$	Support of aggregate voter preference shock, $\pm\overline{A}$	0.01

we also record the outcome of the same regression with probability of a D victory as the dependent variable.

Table 4: Incumbency Disadvantage

	Model			
Dep. Var.	$\%D$-$\%R$	$\%D$-$\%R$	Prob. of D Win	Prob. of D Win
Constant	0.00	0.00	0.50	0.50
SIX^+	-5.00	-	-0.07	-
$EIGHT^+$	-	-5.00	-	-0.07

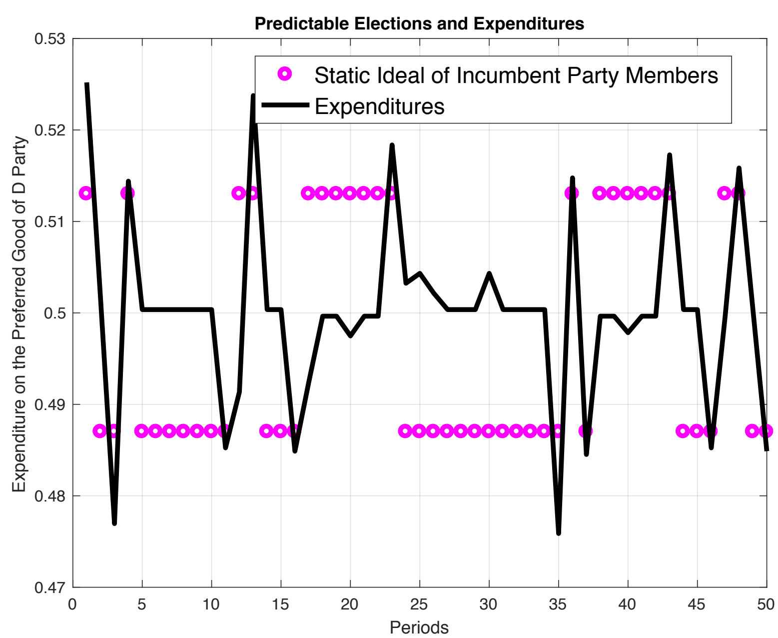
The first two columns in Table 4 report the magnitude of the incumbency disadvantage in the model for both six or more years and eight or more years of incumbency (in the model, these correspond to 3 or more periods or 4 or more periods). For either measure of incumbency, the incumbency disadvantage is 5 percentage points. The next two columns report the incumbency disadvantage in terms of 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 4, the coefficients on the incumbency variables are all estimated to be zero.

5.1 Role of Election Uncertainty

As we have seen (and argued via the static example) inertia in policies is a necessary condition for incumbency disadvantage. However, it is not sufficient: There must be sufficient uncertainty in election outcomes as well. The reason is that there are two offsetting forces at work: Moving policy toward its own ideals increases the incumbent's current payoff but it also increases the likelihood that the party will be voted out of power in the next election. The latter effect is weakened if the aggregate voter preference shock is sufficiently volatile, for then there is a good chance the incumbent will retain office despite moving policy toward its own ideals. This attenuation allows the first force to dominate and leads to incumbency disadvantage.

We can confirm this logic by studying what happens if elections become perfectly predictable, i.e., there are no aggregate voter preference shocks and $A = 0$ always. Figure 3 shows one 30-period segment of a long simulation of the baseline model with variance of A set to zero. In the Figure, the magenta dots record which party is in power in any given period: When the D party (R party) is in power, the dots are above (below) 0.5. For convenience, dots are placed at 0.513 and 0.487, the values of g that D party and R party members, respectively, would most prefer if there were no costs to changing policies.[12] The solid line shows the path of g (since $\tau = 1$, g can be thought of the portion of spending that is allocated to the preferred good of the D party). It shows that the incumbent party either keeps policy stable at 0.5 or indulges the preferences of the other party and once in a while moves policy strongly toward its constituents' ideal policy and then goes out of power.

Figure 3:
Predictable Elections and Expenditures



[12]Since voters do not experience the m shock, the statically ideal expenditure on the voter's preferred good is always 0.513. In terms of g, the ideal expenditure of D and R party members is 0.513 and $1 - 0.513$, respectively.

A closer look at the path of g reveals that when the D party (R party) is in power then it shades its policies toward the ideal policy of the R party (D party) by choosing g' below 0.5 (above 0.5) until the period in which it moves policy strongly toward τ (toward 0). Thus, the optimal strategy of the incumbent is to compromise during "normal" times by picking an expenditure pattern that is more preferred by constituents of the other party until the time when the gain from moving policy toward its preferred good overwhelms the loss from electoral defeat in the next election. This occurs when m is sufficiently low (i.e., the marginal utility from it preferred good is sufficiently high).

The key implication of this behavior is that once a party comes into power it stays on until a large enough shock occurs to make it want to respond and relinquish power. Since large shocks occur infrequently, reelection is the norm rather than the exception. Indeed, if we run the same regression on model output as we did for the baseline model, we find that the probability of reelection conditional on having been in power for six or more (or eight or more) years is 0.64 (in contrast to 0.43 in the baseline model) as shown in Table 5. Thus, when election outcomes are perfectly predictable, incumbency does *not* predict disadvantage. However, note that incumbency disadvantage, as measured by vote share lead, is still present. This effect is entirely the result of policies swinging strongly toward the incumbent's ideal right before the incumbent goes out of power.

Table 5: Election Uncertainty

	Base model		$var(A) = 0$	
Dep. Var.	$\%D$-$\%R$	Prob. of D Win	$\%D$-$\%R$	Prob. of D Win
Constant	0.00	0.50	0.00	0.50
SIX^+	-5.00	-0.07	-2.70	0.14

To summarize: In the dynamic model an incumbent has the opportunity to move policy toward its ideal policy. If the connection between today's policy choice and tomorrow's election outcome is rendered weak by a volatile A shock, the incumbent will utilize the opportunity to move policy toward its ideal and risk losing the election next period. If elections become very predictable, the incumbent will delay utilizing the opportunity until such time it is very valuable to do so. The equilibrium will display incumbency disadvantage in the former case only.

5.2 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 4 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 $\eta = 0$ 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 $\eta = 5$. For the base model, the average expenditure is initially below its long-run level of 0.512 but reaches that level (essentially) 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 lengthens. As seen in the shape of the red dashed line ($\eta = 50$) and the green dashed line ($\eta = 200$), policies start out closer to 0.50 and continue to move up even beyond the eighth year of incumbency.

Figure 4:
Incumbency and Average Expenditure on the Preferred Good

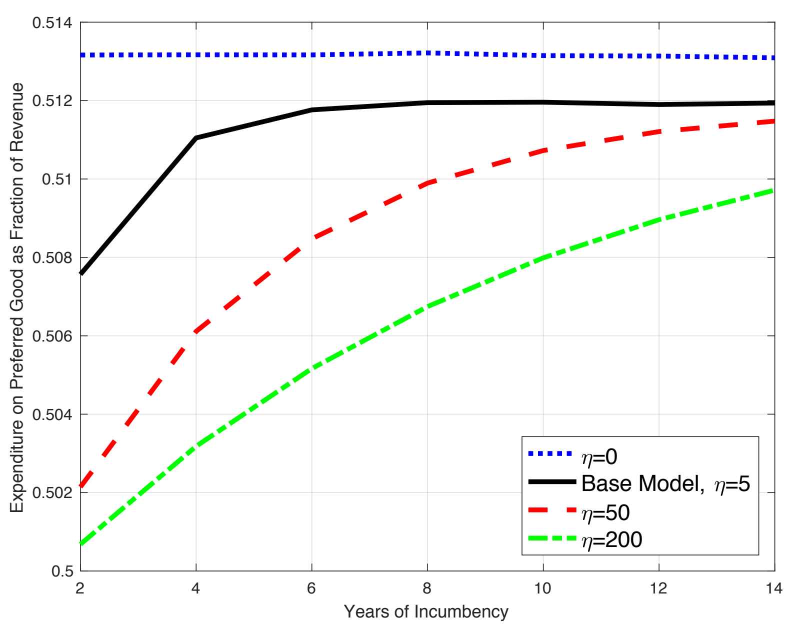
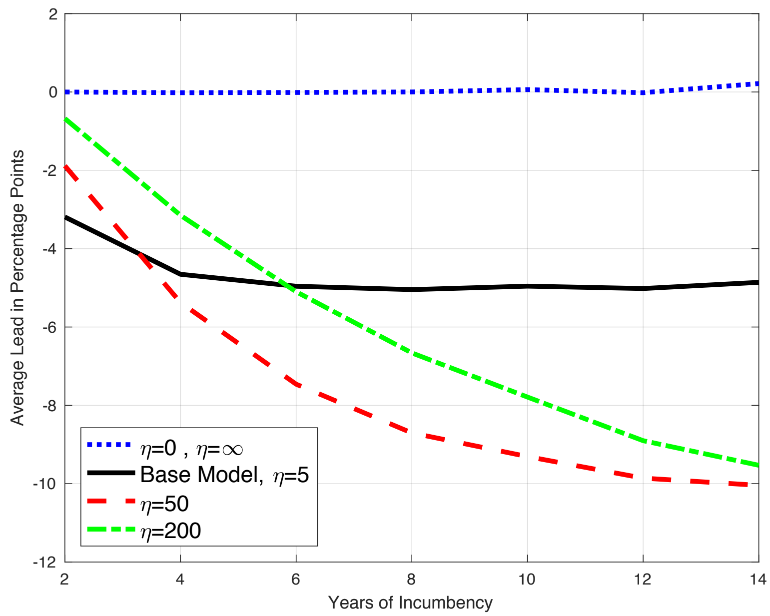


Figure 5:
Incumbency and Average Lead in Elections



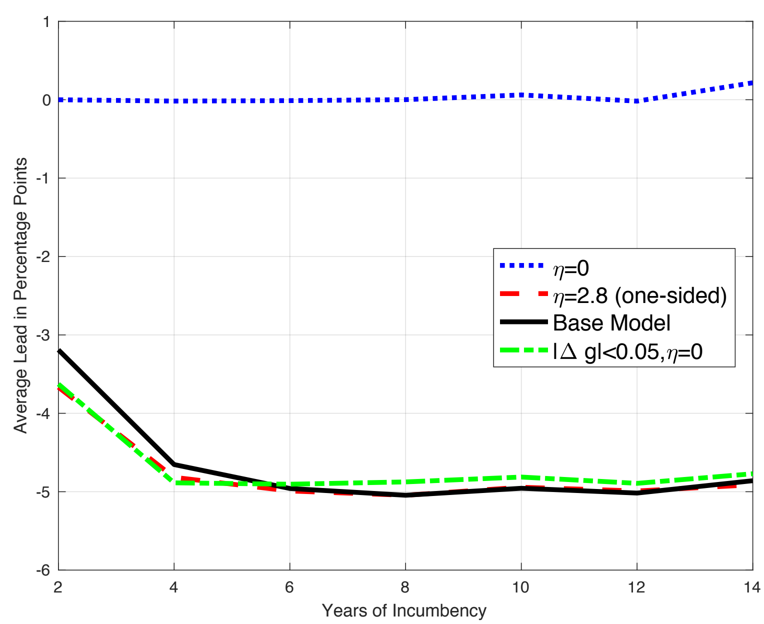
These expenditure dynamics have implications for the time path of incumbency disadvantage. Figure 5 plots the average percentage of voters who cast their ballots in favor of the incumbent minus the percentage who don't. There is no incumbency disadvantage when $\eta = 0$ and the dotted blue line is flat at zero. For the base model, the difference is about -3 percentage points at the end of two years of incumbency and the difference continues to drop until it (essentially) stabilizes at -5 percentage points by the sixth year of incumbency. Interestingly, the incumbency disadvantage is *stronger* for $\eta = 50$ than for $\eta = 200$: The red dashed line is everywhere below the green dashed line. This ordering echoes the inverted-U shape of $A(g)$ function shown for the static model (Figure 2). In the full dynamic model as well, $A(g)$ is initially increasing in η but then it falls. Notice also that the red dashed line crosses the solid black line and the incumbency disadvantage for $\eta = 50$ and $\eta = 5$ is roughly similar for incumbency of 4 years. However, unlike the $\eta = 5$ case, the $\eta = 50$ case has incumbency disadvantage declining strongly all the way to 12 years of incumbency. Since this is *not* the pattern we see in the data, the model "identifies" a relatively small value of η as the correct parameter for matching the pattern of incumbency disadvantage documented in Table in 2.

5.3 Alternative Models of Inertia

In this section we show that incumbency disadvantage also occurs for other models of inertia that may seem plausible. For this we examine two alternative models. In the first alternative there is a quadratic cost of adjustment just as in the base model but the costs of adjustment are borne only by the party in power. In the second alternative, there is no cost of adjustment but there is an upper bound on how much policies can change (in either direction) in any period (we call this the constraint-on-change model)

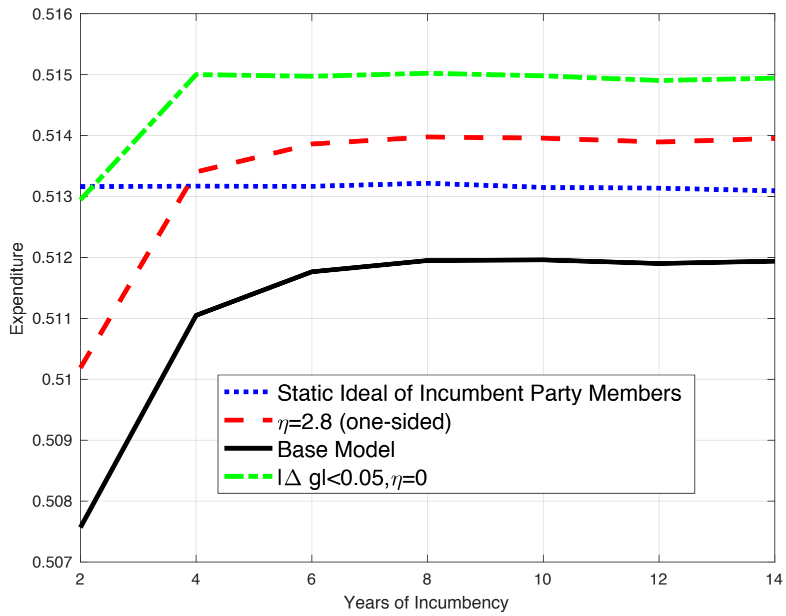
The first point is that the pattern of incumbency disadvantage documented in Table 2 can be accounted for by any of these alternative models of inertia. Figure 6 plots the average lead in elections for the two alternative models along with the base model. The predicted relationships are virtually identical. Indeed, regressions on the simulated output from any of these models implies incumbency disadvantage of about 5 percentage points after six (eight) years of incumbency. The one-sided adjustment cost model can account for the pattern for $\eta = 2.8$ and the constraint-on-change model can account for it by setting the constraint at 0.05 units of g.

Figure 6: Alternative Models of Inertia
Incumbency and Average Lead in Elections



However, the alternative models imply something quite different about the long-run ideal policy of each party. In contrast to the base model, the long-run ideal policies are *more* extreme than the

Figure 7: Alternative Models of Inertia
Incumbency and Expenditure on the Preferred Good

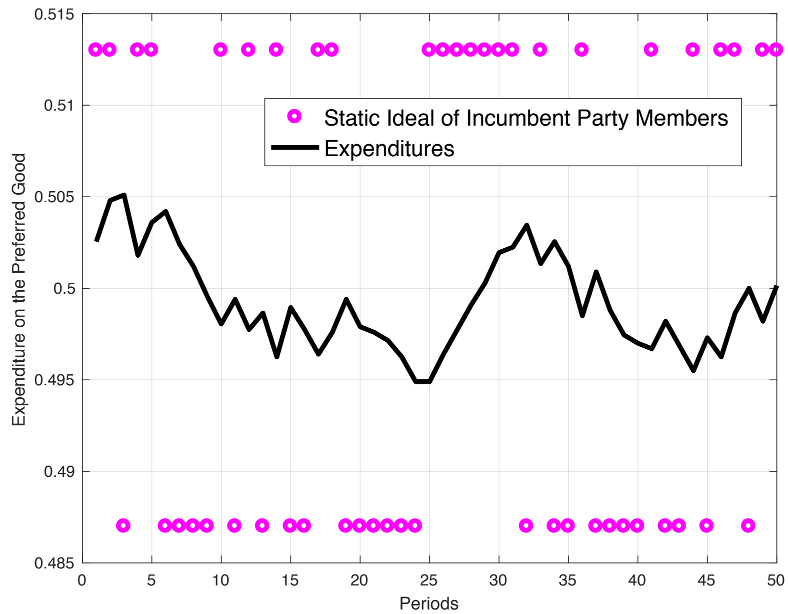


static ideal policy of their party members, as shown in Figure 7. Because of inertia, the incumbent pushes beyond the static ideal policy of its party members to assure them policies "close" to their static ideal even when it is out of power and the other party is choosing policies. Thus, inertia can explain why parties throw their weight behind extreme policies. Interestingly, this extremism does not arise in the base model because the costs of changing policies are borne by *all* representatives in government. Since an incumbent anticipates the swing 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.

Next we turn to the welfare implications of inertia. Policy inertia and the implied incumbency disadvantage mean that the composition of government spending exhibits cycles. When the D party comes into power, it typically inherits a g that is generally below its ideal level because it tends to win elections precisely when the composition of spending has strayed significantly from its ideal composition. Upon getting into power, it works to reverse the shift in composition that occurred during the previous R party reign. Figure 8 shows one 30-period segment (from a long simulation of the baseline model) that captures this pattern particularly well.

These spending composition cycles impose a welfare cost on voters in that there is a path of g and a path for the governing party that delivers a higher utility to each voter, ex ante. This is the

Figure 8:
Political Expenditure Cycles



path in which g' is constant at 0.50 and the party favored by the aggregate shock (i.e., the D party if $A_t > 0$ and the R party, otherwise) is elected. For this policy, the mean lifetime utility, averaged over both types of voters and over all possible values of the shocks A and e, solves the following recursion:

$$\mathcal{W} = [1+\alpha]U(0.5) + \mathbb{E}_A \mathbb{1}_{\{A \geq 0\}} A + \beta \mathcal{W},$$

where we have used the fact that $\mathbb{E}_{(A,m)} \mathbb{1}_{\{A\geq 0\}}[A+e] = \mathbb{E}_A \mathbb{1}_{\{A\geq 0\}} A$, and that the fraction of voters of each type is one-half. In contrast, starting with an initial g of 0.5, the analogous voter welfare expression for equilibrium policies is:

$$\mathcal{W}_{\text{EQ}} = \frac{1}{2}\left[\frac{\mathbb{E}_m W_D(0.5,m) + \mathbb{E}_m Z_R(0.5,m)}{2}\right] + \frac{1}{2}\left[\frac{\mathbb{E}_m Z_D(0.5,m) + \mathbb{E}_m W_R(0.5,m)}{2}\right] + \mathbb{E}_A \mathbb{1}_{\{A\geq 0\}} A,$$

where we have used the fact that $A(0.5) = 0$ and, so, $\mathbb{E}_A \mathbb{1}(0.5, A) = 0.5$.

Table 6 reports the value of $[(\mathcal{W} - \mathcal{W}_{\text{EQ}})/\mathcal{W}] \times 100$, in ascending order, for different levels and models of inertia. Observe that equilibrium welfare never exceeds $\mathcal{W}$. Predictably, the welfare loss is zero for large η since g stays constant at 0.5. For $\eta = 50$, the welfare loss is two-tenths of a percentage point and the standard deviation of g is 0.08. For the base model, the welfare loss is seven-tenths of a percentage point and the standard deviation of g is 0.12.

Table 6:
Welfare Implications of Policy Inertia

Models	Welfare Loss in %	Std. Dev of g'
$\eta = 50000$	0.00	0.00
$\eta = 50$	0.02	0.08
Base Model	0.07	0.12
$\eta = 2.8$ for party in power	0.12	0.13
$\eta = 0$	0.15	0.14
$\eta = 0$ but $\lvert\Delta g\rvert < 0.05$	0.18	0.14

As is to be expected, the welfare loss in the base model is *less* than the loss for $\eta = 0$ model. Recall that in the base model, the long-run ideal policy of the parties is less dispersed relative to the no-inertia model. Thus, inertia works to reduce fluctuations in g by slowing down the movement of policies toward the long-run targets *and* by bringing the targets closer to 0.5. Interestingly, the welfare loss for the model where the costs of adjustment are borne only by the party in power is also less than the no-inertia model even though the long-run ideal policies are further apart than the no-inertia model counterparts. The welfare loss is attenuated because inertia and the resulting incumbency disadvantage serve to keep realized policies close to 0.5. For the constraint-on-change model, the extremism of the long-run policies win out and welfare loss is greater than in the no-inertia model.

6 Policy Inertia and Median Voter Logic without Pre-Election Commitment

Downs (1957) famously observed that if politicians care only about winning elections, policies enacted will converge to the one most preferred by the median voter. Alesina (1988) pointed out that if parties also care about the policies they enact and cannot commit to policies before an election, then equilibrium policy will be the one that the winning party most prefers, *regardless* of how motivated the party is to win the election. In other words, the inability to commit can completely short-circuit the moderating influence of office motivation.

Here we show that inertia and prospective voting can reintroduce median voter logic even wherein there is pre-election commitment to policies. For this we extend the base model to include a utility benefit enjoyed by the political party when it is in power. Specifically, the period utility from being in power for the D party is $U(g') + \alpha U(\tau - g') - \eta(g' - g)^2 + B$, where $B > 0$. Symmetrically, the period utility from being in power for the R party is $\alpha U(g') + U(\tau - g') - \eta(g' - g)^2 + B$. The constant B is a stand-in for the office motivation of the representatives seeking election.

We confirm Alesina's result for the no-inertia version this model.

Theorem 3 (Alesina 1988)**.** *If there are no adjustment costs or constraints on changing policies, parties choose their statically ideal policies regardless of the value of B.*

Proof. If $\eta = 0$, g is no longer payoff-relevant and, so, its value cannot affect equilibrium outcomes. Thus $A^*(g)$ and $G_j^*(g, m)$, $j \in \{D, R\}$, are independent of g but, potentially, dependent on B. Assume that the former is $A^*(B)$ and the latter are $G_j^*(m, B)$. Then, the continuation value of party D when it is power is

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

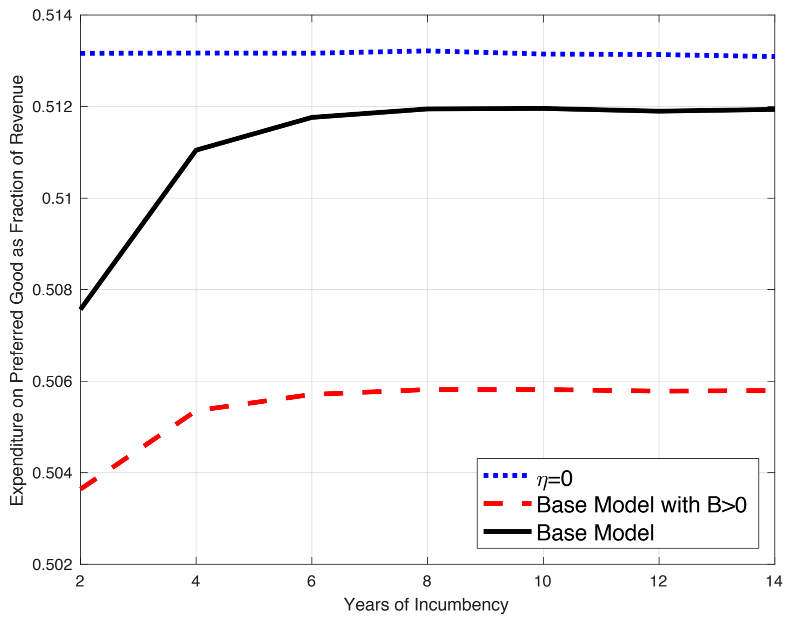
Since this continuation value is independent of g, once party D is elected the best it can do is solve its static optimization problem. Thus,

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

which is evidently independent of B. Symmetrically, the R party will choose its statically ideal policy, independent of the value of B. As a corollary, one may verify that when the two parties act in this way, the net gain terms within square brackets in (14) are equal and, so, $A^*(B) \equiv 0$ and the probability of reelection is 1/2 regardless of B. $\square$

This result changes when $\eta > 0$. With inertia, the presence of B creates a conflict of interest between representatives and their constituencies: Upon electoral defeat, the party loses B and the prospect of this loss restrains the party from pushing as hard for its constituents as it otherwise would. We confirm this in Figure 9. The top panel plots the average equilibrium value of g' against the number of years of incumbency for the baseline model and the model with office motivation. For this graph, we assume that the initial $g = 0.5$. The graph shows that as the D party comes into power and continues in power, the average g rises for both models. However, the rise in g' for the base model is more pronounced than in the model with office motivation. Figure 10 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 office motivation.

Figure 9:
Office Motivation, Incumbency and Average Expenditure on the Preferred Good

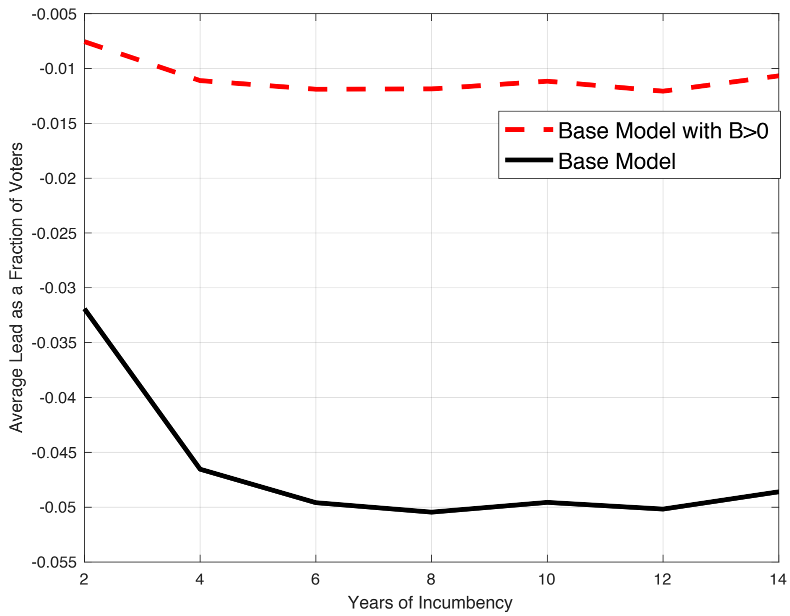


Interestingly, although office motivation creates a conflict of interest between the party and its constituents, the welfare implications of office motivation for the constituents is *positive*. This is because office motivation reduces the amplitude of the policy cycles discussed in the previous subsection. Table 7 reports the welfare loss from the stable g benchmark (i.e., relative to $\mathcal{W}$) for the base model and the base model augmented with office motivation. Observe that the standard deviation of g' is lower and the welfare loss is smaller with office motivation than without.

Table 7: Welfare and Office Motivation

Models	Welfare Loss in %	Std. Dev of g'
Base Model	0.07	0.12
Base Model with $B = 0.9$	0.05	0.10

Figure 10:
Office Motivation, Incumbency and Average Lead in Elections



7 Summary

In this paper, we documented a strong pattern of incumbency punishment in U.S. national politics. Postwar evidence of the electoral performance of the two parties in House elections shows that a long presidential incumbency of a party leads to substantial decline in the popularity of the party in national elections. We used Alesina and Tabellini's (1990) model of partisan politics, extended to have elections with prospective voting, 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 pre-election commitment to policies, and this can dampen policy cycles and raise welfare.

APPENDIX

The goal of this appendix is to provide a secure computational foundation for models with endogenous reelection probability. We establish that there is at least one pure strategy Markov Perfect Equilibrium for a class of models that includes all the models discussed in the main text and, potentially, other models of interest. The two main assumptions needed to obtain this result are that there is a finite number of possible states and that there is a continuously distributed shock to primitives that affects the optimal choice in any given state.

Because of nonconvexities that may arise from strategic interactions, MPE of the type of game analyzed in this paper cannot, in general, to be shown to possess continuous decision rules. This lack of continuity, in turn, implies nondifferentiability of continuation value functions. Thus, computational methods that rely on interpolation of functions of continuous variables are ill-suited for this class of models. Given this, these types of models are best solved on a grid. But if the state space is discrete, a focus on pure strategy equilibria is problematic because standard theory guarantees the existence of mixed strategy equilibria only. The role of the continuously distributed shock to primitives (the m shock in the main model) is to allow mixing when it is called for in equilibrium: When a player is very close to being indifferent between two actions, the realization of the shock determines which action is chosen.[13]

However, the presence of this *continuously* distributed shock poses a challenge for computation because the equilibrium decision rule may fail to be continuous with respect to this shock as well. Since this shock cannot be put on a grid without compromising existence, we need to be able to compute the optimal decision for *all* values of the shock (given a value of the other discrete state). In this regard, if it is the case that for any two feasible actions there is at most *one* value of the shock for which the two actions give the same payoff, then an algorithm exists — described in detail in the Appendix B — to compute the equilibrium decision rule to any arbitrary precision.

A Existence a Pure Strategy Markov Perfect Equilibrium

A.1 A General Environment

Probability Spaces:

[13] In this sense, the continuously distributed shock to primitives plays the same role as the additive payoff perturbations introduced in Harsanyi (1973) to purify mixed strategies.

If $X \subseteq \mathbb{R}$, $\mathcal{B}_X$ is the Borel σ-algebra on $\mathbb{R}$ restricted to X. Let $M = [\underline{m}, \overline{m}]$ be a closed interval in $\mathbb{R}$ and let $(M, \mathcal{B}_M, \mu)$ denote the probability space for the preference shock m.

Assumption 1. *μ is absolutely continuous with respect to the Lebesgue measure on M.*

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

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

Assumption 2. *For all $a \in \mathbb{R}$, $\int_{A>a} A\, d\lambda$ exists and is continuous in a and there is a $\Delta > 0$ such that $|\int_{A>a} A\, d\lambda| < \Delta$.*

For the proof of existence, it is not necessary to assume that A has compact support, only that all truncated integrals exist, vary continuously with the truncation point and be bounded. These assumptions would be satisfied if A possessed a continuous density over a compact support but would also be satisfied if it possessed a continuous density over $\mathbb{R}$ as long as the density converged to 0 exponentially fast as a diverged to $\pm\infty$ (as is the case, for instance, with the Normal distribution).

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

Assumption 3. *$\int_{\mathbb{R}} e d\varepsilon = 0$, $\varepsilon((-e, 0]) = \varepsilon([0, e))$ for all $e \in \mathbb{R}$ and $F(e)$ is continuous and strictly increasing in e.*

Any density function that is symmetric around 0 and strictly positive for all $e \in \mathbb{R}$ will satisfy this assumption (such as the Normal distribution). The assumption implies the following result:

Lemma 1. *If $F(e') = 1 - F(e)$ then $e' = -e$.*

Proof. First, we show that e' and e cannot both be strictly positive or strictly negative. Suppose, to get a contradiction, that $e', e > 0$. By symmetry around zero, $F(e') = 1/2 + \varepsilon((0, e'])$ and $F(e) = 1/2 + \varepsilon((0, e])$. Since $F(e)$ is strictly increasing in e, both $\varepsilon((0, e'])$ and $\varepsilon((0, e'])$ are strictly positive. But, $F(e') = 1 - F(e)$ implies $\varepsilon((0, e']) = -\varepsilon((0, e])$, a contradiction. The proof that e' and e cannot both be negative is analogous.

Assume, then, that $e' \leq 0$ and $e \geq 0$. By symmetry around 0, $F(e') = 1 - F(e)$ implies $F(e') = \varepsilon((-\infty, 0]) + \varepsilon(-\infty, 0]) - F(e)$. Hence, $F(e') - \varepsilon((-\infty, 0]) = \varepsilon((-\infty, 0]) - F(e)$. Since $e' \leq 0$, the l.h.s. of this equality is $-\varepsilon((e', 0])$ and since $e \geq 0$ the r.h.s. is $-\varepsilon([0, e))$. Hence we have that $\varepsilon((e', 0]) = \varepsilon([0, e))$. Now, fix e. By symmetry we know that $\varepsilon((e', 0]) = \varepsilon([0, e))$ holds for $e' = -e$. Since $F(e)$ is strictly increasing in e, it follows that it can hold *only* for $e' = -e$. $\square$

States and Actions:

Let $\mathcal{I} = \{1, 2, \ldots, I\}$ be the set of possible endogenous states the economy can be in at the start of any given period and let $\Gamma_i^k \subseteq \mathcal{I}$ be the set of states that can be selected as a state for the next period given that the current state is i and the current party in power is k. We will use i and j to denote generic elements of $\mathcal{I}$ and Γ_i^k, respectively. We call $\{i, j, k\}$ a *feasible triple* if $j \in \Gamma_i^k$.

Assumption 4. *$\Gamma_i^k \neq \emptyset$ for all $i \in \mathcal{I}$ and all $k \in \{D, R\}$.*

That there is a feasible selection for every possible endogenous state is a mild requirement. The fact that $\mathcal{I}$ contains only endogenous states is 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. A more important restriction embodied in the definition of Γ_i^k is that these sets are unaffected by the realization of m. In our application, $M = [-0.05, 0.05]$ and restriction can be met by requiring that $g' \in (0.05, 0.95)$. Then $g' + m$ and $\tau - g' + m$ will be strictly positive for all g', no matter what value of m is realized.

Current period rewards:

$U_{i,j}^k(m) : \mathcal{I} \times \Gamma_i^k \times M \to \mathbb{R}$, $k \in \{D, R\}$, denotes the current period reward to party k when party k is in power, the state is i, the preference shock is m and $j \in \Gamma_i^k$ is the chosen state for the next period. And $U_{i,j}^k : \mathcal{I} \times \Gamma_i^{\sim k} \to \mathbb{R}$, denotes the current period reward to party k when party $\sim k$ is in power, the state is i and $j \in \Gamma_i^{\sim k}$ is the state chosen for the next period.

Assumption 5. *$U_{ij}^k(m)$ is continuous in m.*

The assumption is unremarkable but helps to establish the first part of the following result.

Lemma 2. *There exists $\overline{U} > 0$ such that $|U_{ij}^k(m)| < \overline{U}$ for all feasible triples $\{i, j, k\}$ and all m and $|U_{ij}^k| < \overline{U}$ for all feasible triples $\{i, j, \sim k\}$.*

Proof. Since $M \subset R$ is compact, Assumption 5 ensures that for any feasible triple $\{i, j, k\}$ there exist $\underline{B}^k_{ij} \leq U^k_{ij}(m) \leq \overline{B}^k_{ij}$. Then $|U^k_{ij}(m)| \leq B^k_{ij} \equiv \max\{|\underline{B}^k_{ij}|, |\overline{B}^k_{ij}|\}$. Let B_1 be the maximum over all these individual $B^k_{ij}s$. Let B_2 be the maximum of $|U^k_{ij}|$ over all feasible triples $\{i, j, \sim k\}$. The Lemma follows by setting $\overline{U} > \max\{B_1, B_2\}$. □

Assumption 6. *Let $\{i, j, k\}$ and $\{i, j', k\}$ be a pair of feasible triples and let $\widetilde{M} = \{m \in M : U^k_{ij}(m) - U^k_{ij'}(m) = 0\}$. Then $\widetilde{M}$ either empty or contains a finite number of points.*

Assumption 6 ensures that party k can be indifferent between any two feasible choices for next period's state for only a finite number of m. This assumption, together with Assumption 1, plays a key role in ensuring the continuity of the self-map whose fixed points are MPE. In our application this assumption is satisfied by virtue of the fact that $U^k_{ij}(m)$ is differentiable and concave in m (see the discussion following Assumption 8 below).

Let u^k_{ij} denote the current period reward to a k-party member if j is chosen when the state is i. Our notation assumes that the identity of the party choosing j does not matter for the utility derived from j (of course, the utility derived could be different for D party and R party members). However, if party D is in power the utility flow u^k_{ij} is augmented by $e + A$ for both D and R party members.

Assumption 7. $|u^k_{ij}|$ *is less than $\overline{U}$ for all k, i and $j \in \Gamma^k_i$.*

This is an assumption of convenience only. The fact that the bound is the same as in Lemma 2 does not restrict u^k_{ij} since $\overline{U}$ can be set large enough to also serve as a bound for individual utility flows.

In what follows, we use some standard definition and results from measure theory and functional analysis. In most instances, the proofs or discussions 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.

A.2 Decision Problem of Parties

The timing of events is the same as in the main text. The period begins in some state i chosen in the previous period. The voter preference shocks e and A are realized and people vote. Following the election, the preference shock m of the party in power is realized and the party chooses the next period's state j. All period rewards are realized and the period closes.

Let $Q_i^k \in \mathbb{R}$, $k \in \{D, R\}$, denote the continuation value of party k of starting a period when the state is i.

Let $A_i \in \mathbb{R}$ be the value of A above which the D party wins the election when the state is i. Let $\mathbb{1}_i(A)$ be the step function that takes the value 1 if $A \geq A_i$ and 0 otherwise.

Let $\omega \in \mathbb{R}^{3I}$ denote $\left(A_i, Q_i^D, Q_i^R\right), i \in \mathcal{I}$. We will use the notation $A_i[\omega]$, $Q_i^k[\omega]$ etc, to denote specific components of ω.

Let $V_{ij}^k(m; \omega) : \mathcal{I} \times M \times \mathbb{R}^{3I} \to \mathbb{R}$ be the value to party k when it is in power and chooses j given state i and preference shock m. Then,

$$V_{i,j}^k(m; \omega) = U_{i,j}^k(m) + \beta Q_j^k[\omega], \;\; k \in \{D, R\}. \tag{17}$$

Let $V_i^k(m; \omega) : \mathcal{I} \times M \times \mathbb{R}^{3I} \to \mathbb{R}$ be the *optimal* value of party k under the same circumstances. Then,

$$V_i^k(m; \omega) = \max_{j \in \Gamma_i^k} V_{i,j}^k(m; \omega) \tag{18}$$

Since $\mathcal{I}$ is a finite set, by Assumption 4 the set of maximizers is nonempty. Let $j_i^k(m; \omega) : \mathcal{I} \times M \times \mathbb{R}^{3I} \to \mathcal{I}$, $k \in \{D, R\}$, denote the maximizer if it is unique, or, if the set of maximizers contains more than one element, the *smallest* member of $\mathcal{I}$ that attains $V_i^k(m; \omega)$ (this is just a tie-breaking rule).

Proposition 1 (Monotonicity and Continuity of V_i^k)**.** *$V_i^k(m; \omega)$ is strictly increasing in ω and it is continuous in both m and ω.*

Proof. That $V_i^k(m; \omega)$ is strictly increasing in ω follows from the fact that for each i, the objective function is trivially strictly increasing in ω (holding m constant). To prove continuity with respect to m and ω, observe that given i, $V_i^k(m; \omega)$ is the upper envelope of the a finite number of functions $V_{i,j}^k(m; \omega) : M \times \Omega \to \mathbb{R}$, $j \in \Gamma_i^k$. By Assumption 5, U_{ij}^k is continuous in m and $Q_j^k[\omega]$ is trivially continuous in ω and, so, each of these $\#(\Gamma_i^k)$ functions is continuous in m and ω. It follows that $V_i^k(m; \omega)$ is continuous in m and ω. □

Proposition 2 (Integrability of V_i^k)**.** *For each 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 closed interval in $\mathbb{R}$ (and therefore 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]. $\square$

Lemma 3 (Maximizers are almost always unique)**.** *Given ω, $V^k_{i,j_i^k(m;\omega)} > V^k_{i,j'}(m;\omega)$ for $j' \neq j_i^k(m;\omega)$, except, possibly, for a finite number of m values.*

Proof. Let $j, j' \in \Gamma_i^k$ and let $M_i^k(j,j';\omega) \subseteq M$ be the set of m values for which $V^k_{i,j}(m;\omega) = V^k_{i,j'}(m;\omega)$. By Assumption 6, $M_i^k(j,j';\omega)$ is either empty or contains a finite number of points. Let

$$M(\omega) = \bigcup_{k\in\{D,R\}} \left\{ \bigcup_{i\in\mathcal{I}} \left\{ \bigcup_{j,j'\in\Gamma_i^k} M_i^k(j,j';\omega) \right\} \right\}$$

be the collection of all such (indifference) points. Since $\mathcal{I}$ is a finite set, $M(\omega)$ is a finite set. Now consider $\hat{m} \in M \setminus M(\omega)$. Then $V^k_{i,j_i^k(\hat{m};\omega)} > V^k_{i,j'}(\hat{m};\omega)$ for any $j' \neq j_i^k(\hat{m};\omega)$. If not, $\hat{m}$ must belong to $M_i^k(j_i^k(\hat{m};\omega), j';\omega)$ for some j' and, so, must belong to $M(\omega)$, 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 lemma follows. $\square$

Proposition 3 (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 $\{j_i^k(m;\omega_n)\}$ converges pointwise to $j_i^k(m;\omega)$ 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 $V_i^{k-}(\hat{m};\omega) = \max_{j\in\Gamma_i^k\setminus j_i^k(\hat{m};\omega)} V^k_{ij}(\hat{m};\omega)$. 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 each 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, the lemma follows. $\square$

Definition 2. *Let $\{B_j \in \mathcal{B}_X\}$, $j = 1,2,\ldots N$ be disjoint, measurable sets. Then $f : X \to \mathbb{R}$ is a simple function if*

$$f(x) = \sum_{j=1}^{N} \theta_j \chi_{B_j}(x).$$

where $\theta_j \in \mathbb{R}$ and $\chi_{B_j}(x)$ is an indicator function that takes the value 1 if $x \in B_j$ and 0 otherwise. If $\{B_j\}$ is also a partition of X, the r.h.s. is called the standard representation of f [SL, Ch. 7, p. 179] .

Definition 3. *Let $(X, \mathcal{B}_X, \nu)$ be a probability space and let $f(x) : X \to \mathbb{R}$ be as in Definition 1 and suppose that $\{B_j \in \mathcal{B}_X\}$, $j = 1, 2, \ldots N$, is a partition of X. By the definition of the Lebesgue integral,*

$$\mathbb{E}_x f(x) = \int f(x)\, d\nu = \sum_{j=1}^{N} \theta_j \int \chi_{B_j}(x) d\nu = \sum_{j=1}^{N} \theta_j \nu(B_j).$$

Proposition 4 (Decision rules are simple functions of m)**.** *There exists $\{B^k_j(\omega)\}$, $j \in \mathcal{I}$, such that $B^k_j(\omega) \in \mathcal{B}_M$, $B^k_j(\omega) \cap B^k_{j'}(\omega) = \emptyset$, $\cup_{j \in \mathcal{I}} B^k_j(\omega) = M$ and*

$$j^k_i(m; \omega) = \sum_{j \in \mathcal{I}} j \cdot \chi_{B^k_j(\omega)}(m). \tag{19}$$

Proof. Fix k. For each $j \in \mathcal{I}$, let

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

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 $V^k_{i\,j}(m; \omega) - V^k_{i \backslash j}(m; \omega)$. Then j is the unique maximizer for m (given k, i, and ω) if and only if $f^k_{i\,j}(m; \omega) > 0$. Let $B^k_{i\,j}(\omega) = \{m \in M : f^k_{i\,j}(m; \omega) > 0\}$ be 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, $B^k_{i\,j}(\omega) \in \mathcal{B}_M$. Next, consider the finite set of m values with nonunique maximizers. For each m in this set, determine the value of the smallest maximizer. If this value is j, assign m to $B^k_{i\,j}(\omega)$. Then:

- $B^k_{i\,j}(\omega) \in \mathcal{B}_M$ (since any finite subset of M is a Borel set and the union of two Borel sets is Borel)
- $B^k_{i\,j}(\omega) \cap B^k_{i,\,j'}(\omega) = \emptyset$ (since an m with a unique maximizer cannot appear in two different $B^k_j(\omega)$'s and nonunique maximizers are assigned to exactly one $B^k_j(\omega)$) and
- $\cup_{j \in \mathcal{I}} B^k_{i\,j}(\omega) = M$ (by Lemma 3, all but a finite number of m's have unique maximizers and so must appear in some $B^k_{i\,j}(\omega)$; if m has nonunique maximizers then it has a smallest maximizer – call it j – and m must belong to $B^k_{i\,j}(\omega)$)

Finally, to show that (19) holds, suppose $j_i^k(m;\omega) = j$. If j is a unique maximizer then $m \in B_{i\,j}^k(\omega)$; if it is a nonunique maximizer it must be the smallest such maximizer by definition of $j_i^k(m;\omega)$ and, therefore, $m \in B_{i\,j}^k(\omega)$ as well. Hence the r.h.s. of (19) is equal to j. □

Corollary 1 (Continuity of decision rules almost everywhere)**.** *$j_i^k(m;\omega)$ is continuous in m except, possibly, at a finite number of points.*

To complete the statement of the party's decision problem we note the value to party k when party $\sim k$ is choosing policy. Let $X_i^k(m;\omega) : \mathcal{I} \times M \times \mathbb{R}^{3I} \to \mathbb{R}$ be the value to party k when the inherited policy is g_i, and continuation values are given by ω and party $\sim k$ is choosing its optimal policy. Then,

$$X_i^k(m;\omega) = U^k_{i\,j_i^{(\sim k)}(m;\omega)} + \beta Q^k_{j_i^{(\sim k)}(m;\omega)}[\omega], \;\; k \in \{D, R\}. \tag{20}$$

We may easily verify that $X_i^k(m;\omega)$ is a simple function of m with the following standard representation:

$$X_i^k(m;\omega) = \sum_j \left\{ U_{i\,j}^k + \beta Q_{i\,j}^k[\omega] \right\} \chi_{B_{i\,j}^{\sim k}(\omega)}(m) \;\; k \in \{D, R\}. \tag{21}$$

Proposition 5 (Integrability of X_i^k)**.** *For each ω, $X_i^k(\omega) \equiv \mathbb{E}_m X_i^k(m;\omega)$ exists.*

Proof. Follows directly from Definition 2.

□

A.3 Lifetime Utilities of Voters

The value functions of voters satisfy the following recursions:

$$W_i^D(m;\omega) = \tag{22}$$
$$u^D_{j_i^D(m;\omega)} + \beta \mathbb{E}_{A'} \left[\mathbb{1}_{j_i^D(m;\omega)} \left(A';\omega \right) \left[\mathbb{E}_{m'} W^D_{j_i^D(m;\omega)}(m';\omega) + A' \right] + \left[1 - \mathbb{1}_{j_i^D(m;\omega)} \left(A';\omega \right) \right] \mathbb{E}_{m'} Z^D_{j_i^D(m;\omega)}(m';\omega) \right],$$
$$Z_i^D(m;\omega) = \tag{23}$$
$$u^D_{j_i^R(m;\omega)} + \beta \mathbb{E}_{A'} \left[\mathbb{1}_{j_i^R(m;\omega)} \left(A';\omega \right) \left[\mathbb{E}_{m'} W^D_{j_i^R(m;\omega)}(m';\omega) + A' \right] + \left[1 - \mathbb{1}_{j_i^R(m;\omega)} \left(A';\omega \right) \right] \mathbb{E}_{m'} Z^D_{j_i^R(m;\omega)}(m';\omega) \right],$$

and

$$W_i^R(m;\omega) = \tag{24}$$
$$u^R_{j_i^R(m;\omega)} + \beta \mathbb{E}_{A'}\left[\mathbb{1}_{j_i^R(m;\omega)}\left(A';\omega\right)\left[\mathbb{E}_{m'} Z^R_{j_i^R(m;\omega)}(m';\omega) + A'\right] + \left[1 - \mathbb{1}_{j_i^R(m;\omega)}\left(A';\omega\right)\right]\mathbb{E}_{m'} W^R_{j_i^R(m;\omega)}(m';\omega)\right],$$
$$Z_i^R(m;\omega) = \tag{25}$$
$$u^R_{j_i^D(m;\omega)} + \beta \mathbb{E}_{A'}\left[\mathbb{1}_{j_i^D(m;\omega)}\left(A';\omega\right)\left[\mathbb{E}_{m'} Z^R_{j_i^D(m;\omega)}(m';\omega) + A'\right] + \left[1 - \mathbb{1}_{j_i^D(m;\omega)}\left(A';\omega\right)\right]\mathbb{E}_{m'} W^R_{j_i^D(m;\omega)}(m';\omega)\right].$$

Proposition 6. *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 each $\omega \in \mathbb{R}^{3I}$ 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* (22) - (23) *for $k = D$ and* (24) - (25) *for $k = R$.*

Proof. We will prove the proposition for $k = D$ (the proof for $k = R$ is analogous). Observe that for each i and ω, the r.h.s. of (22) can be viewed as an operator that takes as inputs a *set* of functions $(W^D(m), Z^D(m)) \equiv \left(\{W_j^D(m), Z_j^D(m)\}_{j\in\mathcal{I}}\right)$ and returns a *single* function $W_i^D\left(W^D, Z^D\right)(m)$. If the input functions are members of $\mathcal{F}$ then their expectation w.r.t. to m exists. Denote these expectations by $\{\overline{W}_j^D, \overline{Z}_j^D\}_{j\in\mathcal{I}}$. Then,

$$W_i^D\left(W^D, Z^D\right)(m) = u^D_{j_i^D(m;\omega)} +$$
$$\beta\left[\Pr\left[A' > A_{j_i^D(m;\omega)}[\omega]\right]\overline{W}^D_{j_i^D(m;\omega)} + \int_{A' > A_{j_i^D(m;\omega)}[\omega]} A'\, d\lambda + \left[1 - \Pr[A' > A_{j_i^D(m;\omega)}[\omega]]\right]\overline{Z}^D_{j_i^D(m;\omega)}\right].$$

where the truncated expectation terms involving A' exist by virtue of Assumption 2. We may verify that the r.h.s. of the above equation can be expressed as a simple function (in standard representation form):

$$W_i^D\left(W^D, Z^D\right)(m) =$$
$$\sum_{j\in\mathcal{I}}\left[u_j^D + \beta\left[\Pr\left[A' > A_j[\omega]\right]\overline{W}_j^D + \int_{A' > A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right)\overline{Z}_j^D\right]\right]\chi_{B^D_{ij}(\omega)}(m)$$

Hence $W_i^D\left(W^D, Z^D\right)(m) \in \mathcal{F}$. Taking expectations with respect to m then yields

$$\overline{W}_i^D\left(W^D, Z^D\right) = \sum_{j\in\mathcal{I}}\left[u_j^D + \beta\left[\Pr\left[A' > A_j[\omega]\right]\overline{W}_j^D + \int_{A'>A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right)\overline{Z}_j^D\right]\right]\mu\left(B_{ij}^D(\omega)\right)$$

Observe that $\overline{W}_i^D(W^D, Z^D)$ depends on the input functions $W^D(m)$ and $Z^D(m)$ only through the expected values (w.r.t. m) of these functions. Therefore, the recursion can be reduced to a recursion in expected values only. Namely,

$$W_i^D(\overline{W}^D, \overline{Z}^D) = \sum_{j\in\mathcal{I}}\left[u_j^D + \beta\left[\Pr\left[A' > A_j[\omega]\right]\overline{W}_j^D + \int_{A'>A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right)\overline{Z}_j^D\right]\right]\mu\left(B_{ij}^D(\omega)\right),$$

where (with a slight abuse of notation) we continue to use $W_i^D(\cdot\,,\cdot)$ to denote the new operator that takes in a vector of expected values and returns an expected value.

Analogously,

$$Z_i^D(\overline{W}^D, \overline{Z}^D) = \sum_{j\in\mathcal{I}}\left[u_j^D + \beta\left[\Pr\left[A' > A_j[\omega]\right]\overline{W}_j^D + \int_{A'>A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right)\overline{Z}_j^D\right]\right]\mu\left(B_{ij}^R(\omega)\right).$$

We may verify easily that the operator:

$$\left(W_i^D(\overline{W}^D, \overline{Z}^D),\ Z_i^D(\overline{W}^D, \overline{Z}^D)\right)_{i\in\mathcal{I}} : \mathbb{R}^{2I} \to \mathbb{R}^{2I}$$

satisfies Blackwell's sufficiency conditions for a contraction map (with modulus of contraction β). Since $\mathbb{R}^{2I}$ is a complete metric space (with, say, the uniform metric), there exists 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:

$$W_i^D(m;\omega) = \sum_{j\in\mathcal{I}} \left[u_j^D + \beta \left[\Pr\left[A' > A_j[\omega]\right] \overline{W}_j^{*D} + \int_{A'>A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right) \overline{Z}_j^{*D}\right]\right] \chi_{B_{ij}^D(\omega)}(m)$$

$$Z_i^D(m;\omega) = \sum_{j\in\mathcal{I}} \left[u_j^D + \beta \left[\Pr\left[A' > A_j[\omega]\right] \overline{W}_j^{*D} + \int_{A'>A_j[\omega]} A'\, d\lambda + \left(1 - \Pr\left[A' > A_j[\omega]\right]\right) \overline{Z}_j^{*D}\right]\right] \chi_{B_{ij}^R(\omega)}(m).$$

□

Proposition 7 (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 $\left(W_i^D(\overline{W}^D, \overline{Z}^D),\, Z_i^D(\overline{W}^D, \overline{Z}^D)\right)$ 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, the image will vary continuously with ω if, for each j, (i) $\Pr\left[A' > A_j[\omega]\right]$ and (ii) $\mu\left(B_{ij}^D(\omega)\right)$ vary continuously with ω. (i) follows from Assumption 2. For (ii), observe that $\mu\left(B_{ij}(\omega) = \int \chi_{[j_i^D(m;\omega)=j]} d\mu\right.$. Let ω_n be a sequence converging to ω. By Proposition 3, $\chi_{[j_i^D(m;\omega_n)=j]}$ converges pointwise to $\chi_{[j_i^D(m;\omega)=j]}$ except, possibly, for a finite number of m values. Since a set composed of a finite number of m has Lebesgue measure 0, by Assumption 1 the probability measure of a finite number of points is also 0. Hence, $\lim_n \chi_{[j_i^D(m;\omega_n)=j]} = \chi_{[j_i^D(m;\omega)=j]}$ $\mu - a.e.$ By the Lebesgue Dominated Convergence Theorem, $\lim_n \mu\left(B_{ij}^D(\omega_n)\right) = \mu\left(B_{ij}^D(\omega)\right)$. □

A.4 Equilibrium Mapping

Define $T(\omega) : \mathbb{R}^{3I} \to \mathbb{R}^{3I}$ to be the mapping:

$$\begin{aligned}
Q_i^D[T(\omega)] &= \mathbb{E}_{(A,m)}\left[\mathbb{1}_i(A;\omega)V_i^D(m;\omega) + [1 - \mathbb{1}_i(A;\omega)]X_i^D(m;\omega)\right] \\
Q_i^R[T(\omega)] &= \mathbb{E}_{(A,m)}\left[\mathbb{1}_i(A;\omega)X_i^R(m;\omega) + [1 - \mathbb{1}_i(A;\omega)]V_i^R(m;\omega)\right] \\
A_i[T(\omega)] &= \frac{[\mathbb{E}_m W_i^R(m;\omega) - \mathbb{E}_m Z_i^R(m;\omega)] - [\mathbb{E}_m W_i^D(m;\omega) - \mathbb{E}_m Z_i^D(m;\omega)]}{2}
\end{aligned}$$

Then, a fixed point of T constitutes a Markov perfect equilibrium. We will show that there is a compact subset $\Omega \subseteq \mathbb{R}^{3I}$ such that $T(\Omega) \subseteq \Omega$ and $T(\omega) : \Omega \to \Omega$ is continuous.

To proceed, use the independence of A and m to re-express the mapping as:

$$\begin{aligned}
Q_i^D[T(\omega)] &= \Pr\left[A > A_i[\omega]\right] \cdot V_i^D(\omega) + \Pr\left[A \leq A_i[\omega]\right] \cdot X_i^D(\omega) \\
Q_i^R[T(\omega)] &= \Pr[A > A_i[\omega]] \cdot X_i^R(\omega) + \Pr[A \leq A_i[\omega]] \cdot V_i^R(\omega) \\
A_i[T(\omega)] &= \frac{1}{2}\left[[\overline{W}_i^{*R}(\omega) - \overline{Z}_i^{*R}(\omega)] - [\overline{W}_i^{*D}(\omega) - \overline{Z}_i^{*D}(\omega)]\right]
\end{aligned}$$

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,

$$\mid V_{ij}^k(m;\omega) \mid \;\leq\; \mid U_{ij}^k(m) \mid + \beta \mid Q_j^k \mid < \overline{U} + \beta\overline{U}/(1-\beta) \;=\; \overline{U}/(1-\beta).$$

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 $[-(\overline{U}+\Delta)/(1-\beta), (\overline{U}+\Delta)/(1-\beta)]^I$. Observe that if $\overline{W}^k$ and $\overline{Z}^k$ belong in $[-(\overline{U}+\Delta)/(1-\beta), (\overline{U}+\Delta)/(1-\beta)]^I$ then

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

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

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

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 $k \in \{D, R\}$ (since by Proposition 7 we 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 establish this, 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

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.

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

B Computation of Decision Rules

In this section we describe how, given k, i and ω, we compute the function $j_i^k(m;\omega) : M \to \mathcal{I}$. We strengthen Assumption 6 to:

Assumption 8 (At most one indifference point)**.** *For any pair $\{j, j'\} \in \mathcal{I}$, there is at most one value of m for which $V_{i,j}^k(m;\omega) = V_{i,j'}^k(m;\omega)$.*

The assumption is satisfied in our application. To see this, let $k = D$ and define $\Delta(m) = V_{i,j}^D(m;\omega) - V_{i,j'}^D(m;\omega)$. Then, $d\Delta(m)/dm = U'(g_j + m) - U'(g_{j'} + m)$. Since u is strictly concave, $d\Delta(m)/dm$ is either strictly positive for all m or strictly negative for all m. Hence $\Delta(m)$ can cross 0 only once. An analogous argument establishes the result for $k = R$.

The following definition of *weakly preferred sets* is useful.

Definition 4 (Weakly Preferred Sets)**.** *Given k, i, m and ω, $P_{i\,j}^k(m;\omega) \subset \mathcal{I}$ is the weakly preferred set of j at m if and only if $j' \in P_{i\,j}^k(m;\omega)$ implies $V_{i\,j'}^k(m;\omega) \geq V_{i\,j}^k(m;\omega)$.*

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

Lemma 4 (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 $j_0 \in \mathcal{I} \setminus P_{i\,j^*}^k(m_2;\omega)$ such that $V_{i\,j_0}^k(m;\omega) > V_{i\,j*}^k(m;\omega)$. First, notice that j^* is always a member of $P_{ij^*}^k(m_2;\omega)$. Since $V_{i\,j^*}^k(m_1;\omega) \geq V_{i\,j_0}^k(m_1;\omega)$ (definition of j^*) and $V_{i\,j^*}^k(m_2;\omega) > V_{i\,j_0}^k(m_2;\omega)$ (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 $V_{i\,j_0}^k(\hat{m};\omega) = V_{i\,j^*}^k(m;\omega)$ and $V_{i\,j_0}^k(\tilde{m};\omega) = V_{i\,j^*}^k(\tilde{m};\omega)$. But this contradicts Assumption 8. Hence, $V_{i\,j^*}^k(m;\omega) \geq V_{i\,j_0}^k(m;\omega)$ 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 4, 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 $P_{i\underline{j}^*}^k(\overline{m};\omega)$ 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 $V_{i\underline{j}^*}^k(m_s;\omega) = V_{ij_s}^k(m_s;\omega)$, $s \in \{\overline{j}^*, 2, 3, \ldots, n-1\}$. Let $\tilde{m}$ be the minimum of this set of indifference points and let $\tilde{j}$ be the corresponding action. Then,

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

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 4, $\tilde{j}$ 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.[14]

Third, it is possible to speed up the algorithm by utilizing a property of the model which, 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

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

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 $j'>j$, $B^D_{ij}(\omega)>B^D_{ij'}(\omega)$. If this inequality is violated then j is strictly dominated by j' for all m and, so,

[14]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 $V^k_{ij}(\overline{m};\omega)$.

can be dropped from further consideration. Thus, by examining the ordering of $B^k_{ij}(\omega)$ over j it is often possible to prune the set of choices the algorithm has to consider.

Finally, there is a property of $j^k_i(m;\omega)$ which 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.

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," *The American Economic Review*, 78(4), 796–805.

Alesina, A., and H. Rosenthal (1995): *Partisan Politics, Divided Government and the Economy*. Cambridge University Press, Cambridge, U.K.

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.

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

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.

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

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

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

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

——— (2017): "Endogenous Political Turnover and Fluctuations in Sovereign Default Risk," Working Paper No 2017-01, Federal Reserve Bank of Philadelphia.

Cuadra, G., and H. Sapriza (2008): "Sovereign Defaults, Interest Rates and Political Uncertainty in Emerging Markets," *Journal of International Economics*, 76, 77–88.

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

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.

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.

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.

Hatchondo, J. C., L. Martinez, and H. Sapriza (2009): "Heterogeneous Borrowers in Quantitative Models of Sovereign Default," *International Economic Review*, 50(4), 1129–1151.

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

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

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

Persson, T., and L. Svensson (1989): "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences," *Quarterly Journal of Economics*, 104, 325–345.

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

Rogoff, K. S. (1990): "Equilibrium Political Budget Cycles," *American Economic Review*, 80, 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," *The American Political Science Review*, 77(1), 142–157.