Dynamic Electoral Competition with Voter Loss-Aversion and Imperfect Recall

Ben Lockwood
University of Warwick

Minh Le
University of Warwick

James Rockey
University of Birmingham

First version: March 25, 2020
This version: November 9, 2023

Abstract: This paper explores the implications of voter loss-aversion and imperfect recall for the dynamics of electoral competition in a simple Downsian model of repeated elections. The interplay between the median voter's reference point and political parties' choice of platforms generates a dynamic process of (de)polarization, following an initial shift in party ideology. This is consistent with the gradual nature of long-term trends in polarization in the US Congress.

KEYWORDS: electoral competition, repeated elections, loss-aversion, imperfect recall, polarization
JEL CLASSIFICATION: D72, D81

Address for correspondence; Department of Economics, Warwick University, Coventry, CV4 7 AL, United Kingdom. E-mail B.Lockwood@warwick.ac.uk.

1 Introduction

The evidence is clear that, whether computed on the basis of words or deeds, American political elites have become increasingly polarized in recent years. Gentzkow et al. (2019) analyze the corpus of Congressional speeches and show that, measured on the basis of differences in language used, partisanship has been consistently increasing since the 1994 election. Analyses based on deeds, particularly voting patterns, find similar evidence although they tend to identify earlier periods of polarization too.[1] Hare and Poole (2014) provide evidence based on common-space (DW-Nominate) estimates of members of congress' positions that polarization has increased since around 1980. McCarty et al. (2009) document a similar pattern in state legislatures.

What is less well-appreciated is that over the early twentieth century, by the same measures, American political elites became considerably *less* polarized. Figure 1 plots three different measures of the ideological distance between the Democratic and Republican parties during each US Congress from 1880 onwards. These distances are calculated from DW-Nominate ideal point estimated for each member of each congressional delegation provided by Lewis et al. (2022) and discussed in Poole and Rosenthal (2006). They thus capture polarization in the political behavior of US political elites but not necessarily polarization of their preferences or those of voters. It is clear that in the 1880s as now there was substantial political polarization. Moving to the right, however, we can see that on all three of our measures, the distance between them narrows over the subsequent fifty years. We see a particularly rapid decline in both chambers, but especially in the Senate, during the Great Depression and a subsequent period of comparative stability. Yet, from the 1970s onwards, the two parties can be seen to drift apart, first slowly, and then increasingly rapidly, from around 1994 in the House and 2004 in the Senate. Thus, the polarization of elite behavior should be seen as a *gradual* process with long periods of both increasing and decreasing polarization.

In an important recent paper, Callander and Carbajal (2022) proposed a behavioral theory of political competition that could explain a dynamic process of increasing polarization, consistent with what is shown in Figure 1 from the 1950s onwards. Specifically, in their model, a voter updates her ideal point by moving it toward the location of the party she voted for, in order to avoid cognitive dissonance. They show that this updating rule interacts with party behavior in such a way that over time, parties choose more polarized platforms, and the ideal points of voters who do not abstain also become centered around the party platforms, thus also becoming more polarized.[2] However, it cannot explain episodes of *depolarization* as in Figure 1.

This paper introduces alternative behavioral features, namely voter loss-aversion and

[1] Poole and Rosenthal (1984) is an early paper documenting increased polarization in the Senate.

[2] Their model assumes that if a voter is too far from *both* party platforms, she will abstain, and these abstainers move to the political center, so voters as a whole do not necessarily become more polarized.

Figure 1: Polarization in American Politics

Notes: "mean" and "median" are the distance between the mean and median ideal points of each the parties' representatives in House or Senate. These measures do not capture any increase in polarization due to a mean-preserving spread of party representatives. Our third measure, IPQR, is the difference between the 25^{th} percentile of the Democratic party and the 75^{th} of the Republican party.

imperfect recall, into a similar type of model of repeated Downsian competition. In this setting, the interplay between the median voter's reference point and political parties' choice of platforms generates a gradual process of polarization *or* depolarization of preferences, depending on the location of the initial reference point. Specifically, following an initial (de)polarization in platforms, voters become more "tolerant" to additional (de)polarization through a shift in the reference point. As a result, platforms converge monotonically to a stable equilibrium, which depends on the degree of voter loss-aversion and the bias in voter recall. Moreover, starting at this stable equilibrium, an exogenous increase (resp. decrease) in platform polarization due to a shift in elite (party) preferences is *magnified* by the resulting shift in voters' reference points, leading to a dynamic process of additional (de)polarization of platforms, even though party preferences do not change further. This is consistent with the continued gradual changes in (de)polarization over long periods, as shown in Figure 1. However, Figure 1 shows a cycle of de-polarization followed by polarization i.e. a U-shape. In our model, we need two exogenous parameter shifts to explain this cycle; first, an exogenous reduction in the relative weight placed by political elites on policy versus holding office, and then a subsequent increase in that weight.

Our model also captures a somewhat different aspect of polarization from Callander and Carbajal (2022). To see this, it is useful to distinguish, as the literature does, between

1) the polarization of elites; 2) the polarization of policy platforms as depicted in Figure 1; and 3) the polarization of mass preferences. Our paper takes mass preferences as given, although voters' reference points change, and studies the impact of an increase in 1) on 2), whereas Callander and Carbajal (2022) study the process of polarization in the sense of the impact of 2) on 3). So, our two papers give complementary explanations of the emptying of the middle ground in democratic politics, something political scientists have argued is essential for democratic stability (Dahl, 1956).

Our model has some additional attractive features. First, our assumptions are less restrictive than those of Callander and Carbajal (2022). We allow for any distribution of voter ideal points, and we do not require parties to be myopic. Also, over time, the ordinal preferences of voters, i.e. their ideal points, do not change at all, so our model is consistent with the fact that while evidence for elite polarization in the US over the last four decades is very strong, there is much less evidence of polarization at the voter level, at least on policy positions.[3]

Our behavioral characteristics are motivated by two stylized facts about voter behavior. First, there is now considerable evidence that citizens place greater weight on negative news than on positive news when evaluating candidates for office, or the track records of incumbents. In the psychology literature, this is known as negativity bias.[4] For example, several studies find that US presidents are penalized electorally for negative economic performance but reap fewer electoral benefits from positive performance (Bloom and Price, 1975; Lau, 1985; Klein, 1991). Following Lockwood and Rockey (2020), we will interpret this behavior as the outcome of voter loss-aversion.[5]

The second stylized fact is *imperfect recall* by voters of past policy platforms and policies implemented, for which there is considerable evidence. Imperfect recall is a generic feature of human memory, which tends to recall more recent events better than their more distant equivalents (Wixted and Ebbesen, 1991). In the context of voting and elections, voters are known to demonstrate imperfect recall of the platforms and policies of the party they previously voted for (van Elsas et al., 2014).[6]

---

[3] Fiorina et al. (2005, p9) argued that claims of an increasingly polarized US electorate: "...rests on misinterpretation of election returns ..." and that, crucially, that "There is little evidence that Americans' ideological or policy *positions* are more polarized today than they were two or three decades ago, although their *choices* often seem to be." Summarizing the literature more recently, McCarty (2019) concludes: "There is very little survey evidence of overall voter polarization ...". Similarly, Gentzkow (2016) writes: "And, many who have looked closely at the data conclude that the depth of divisions in the current American electorate has been wildly overstated."

[4] See for example, the survey on negativity bias by Baumeister et al. (2001).

[5] Similar asymmetries have also been identified in the UK and other countries. For example, for the UK, Soroka (2006) finds that citizen pessimism about the economy, as measured by a Gallup poll, is much more responsive to increases in unemployment than falls. Kappe (2018) uses similar data to explicitly estimate a threshold or reference point value below which news is "negative", and finds similar results. Nannestad and Paldam (1997) find, using individual-level data for Denmark, that support for the government is about three times more sensitive to a deterioration in the economy than to an improvement.

[6] Imperfect recall is distinct from the phenomenon of *hindsight bias* in which voters tend to believe that they accurately predicted a previous uncertain event, including elections, or a future government's

Related Literature

This paper is related to Alesina and Passarelli (2019) and Lockwood and Rockey (2020). In Lockwood and Rockey (2020), there is an analysis of a similar model to the one studied in this paper, where voters have loss-aversion, but perfect recall. In Lockwood and Rockey (2020), the one-period and two-period versions of the model with perfect recall were studied. However, the focus of Lockwood and Rockey (2020) was mainly on how the platforms adjusted in the second period to shifts in voter preferences.

Alesina and Passarelli (2019) also study a two-period model of electoral competition with loss-aversion. However, their model is rather different from the standard Downsian model, as explained in Lockwood and Rockey (2020). Alesina and Passarelli (2019) prove persistence in policies; if (for example) the right-wing party wins the election, then in the following period, *both* parties' equilibrium platforms will be further to the right. Unlike this paper, they only consider two periods, so they cannot study the long-run behavior of this process.[7] Also, reflecting the fact that their model is very different from ours, the dynamics in our model (which is much closer to a classic Downsian model) are qualitatively completely different. In fact, due to the symmetry of our model, the second period platforms of both parties are *independent* of which party wins the election in the first period.[8] Rather, the intertemporal dependence between equilibrium platforms is in the *polarization* dimension; the amount of polarization in the current period is increasing in the polarization in the previous period.

This paper is also related to the literature studying repeated elections in a Downsian setting, and where there is some kind of linkage between periods in the economic or political environment. These include Battaglini (2014), who studies a model in which two office-motivated parties choose platforms that include the level of public debt. Also related are two papers in which first-period policy decisions by the winning party change induced preferences over taxes or other fiscal policies in the second period. Biais and Perotti (2002) studies the effect of privatization in building support for the right-wing party at the second election, and Prato (2018) has a model in which voters learn about an aggregate shock via home-ownership in the first period, which changes their induced preferences over a tax rate in the second period. Papers in which the political environment provides the linkage include Nunnari and Zápal (2017) and Forand (2014) which assume that if a party wins the election, it is then committed to its winning platform for as long as it remains in power.

Our work also relates to several other recent papers on the causes and effects of elite polarization. Diermeier and Li (2019) study the interaction between partisan affect and elite polarization in a behavioral voting model. They show that parties bias their policies

policy (Blank et al., 2003; Schuett and Wagner, 2011; Pohl and Erdfelder, 2019).

[7] In the introduction, they describe this persistence as a "cycle", but technically, as only two periods are studied, they cannot establish whether the time-path for platforms is cyclic or monotonic.

[8] Both parties' policies in equilibrium are equidistant from the median voter's ideal point.

toward their partisans if voters exhibit in-group responsiveness, i.e., they respond more strongly to their own party's policy deviations than to policy deviations by the other party. This also applies to changes, so greater voter affective polarization leads to greater elite polarization.[9] In contrast, our paper shows how the effects of an exogenous change in elite polarization can become magnified via voter reference points. Levy et al. (2022) has a rather different model in which political office in equilibrium cycles between a populist party (which has a simple, misspecified model of the link between policies and a welfare outcome) and a sophisticated party, which has the correct model. Their model can then explain cycles of polarization in that the populist party chooses more extreme policies when in office.

Finally, several recent papers explore the implications of voter loss-aversion in non-Downsian settings. Passarelli and Tabellini (2017) study a model of political protest in which the protest is partially motivated by policy payoffs relative to an exogenously determined "fair" reference point. Grillo (2016) shows that with loss aversion, honest communication with voters about valence is possible in equilibrium. Grillo and Prato (2020) study a model of democratic backsliding where citizens' retrospective assessment of politicians depend on reference points that are endogenous to incumbent behavior.

The remainder of the paper is organized as follows. Section 2 sets out the model, Section 3 establishes preliminary results for the one-period game, and then provides the main analysis of the finite-period game. Section 4 provides simulation evidence that the model captures key qualitative and quantitative features of the data, and Section 5 concludes.

2 The Model

2.1 The Environment

There are two parties L and R, and a single representative voter (the *median voter*) who interact over finite periods $t = 0, 1, 2, \ldots, T$. The assumption of a single voter is without loss of generality, because, under the assumptions made below, voter preferences are single-peaked, and so with multiple voters, the median voter will, in any case, be decisive.[10] We assume $T < \infty$ to allow us to solve the model by backward induction. This is, without loss of generality, in the sense that the unique equilibrium we identify is also an equilibrium in the infinite-horizon game. It will turn out that the structure of the equilibrium will not depend on the number of time periods T. At $t = 1, \ldots, T$, the

[9] A related paper is Diermeier and Li (2017), which studies electoral control when voters have imperfect recall of previous behavior of the incumbent. In that paper, incumbents do not set policy platforms, but only choose effort.

[10] See Lockwood et al. (2022) for an analysis of the multiple voter case, where the voters have different ideal points. In this case, for analytical tractability, we need the assumptions: (a) that all voters have the same reference platform; (b) the same degree of loss-aversion.

two parties, L and R, choose platforms $x_{L,t}, x_{R,t}$ in the policy space $\Re$. At $t = 0$, the platform x_0 is predetermined. It is assumed that parties are able to commit to implement these platforms. Thus, the basic framework is Downsian competition. Parties are also described by i.i.d time varying shocks to their popularity with the median voter, denoted $v_{R,t}$, $v_{L,t}$. The distinctive feature of our model is that the median voter has loss-aversion over platforms, with the reference point of the voter being her recollection of the winning platform of the previous period, as described in more detail below.

2.2 Order of Events and Information Structure

Within period t, the order of the events is as follows. First, parties L, R simultaneously choose their platforms. Then, the popularity shocks $v_{R,t}, v_{L,t}$ are drawn. The difference $v_t = v_{L,t} - v_{R,t}$ is assumed to be uniformly distributed with support $[-\frac{1}{2\rho}, \frac{1}{2\rho}]$. As we will see, the parameter ρ measures the responsiveness of the median voter to policy changes by the parties. The median voter then votes for one party or the other. We will assume that the voter does not play weakly dominated strategies; with only two alternatives, this implies that she votes sincerely.

This timing implies that the current popularity of both parties is not known at the point when the platforms are chosen. From a modeling point of view, the purpose of this timing assumption is a standard one; it makes the outcome of the election uncertain for the two political parties, thus preventing complete convergence in equilibrium to the median voter's ideal point.

2.3 The Median Voter

We assume that "ordinary" or intrinsic utility over platforms $x \in \Re$ of the median voter is single-peaked and of the absolute value form $u(x) = -|x|$. Following Kőszegi and Rabin (2006, 2007), we specify the gain-loss utility over policy of the voter at period $t = 1, 2, \ldots, T$ as:

$$u(x_t; r_t) = \begin{cases} u(x_t) - u(r_t), & u(x_t) \geq u(r_t) \\ \lambda[u(x_t) - u(r_t)], & u(x_t) < u(r_t), \end{cases} \tag{1}$$

where r_t is the reference platform in period t and the parameter $\lambda > 1$ measures the degree of loss-aversion.

The payoff of the median voter from party $K = L, R$ at date t with platform $x_{K,t}$ and popularity shock $v_{K,t}$ is assumed additive in both terms i.e.

$$u(x_{K,t}; r_t) + v_{K,t}. \tag{2}$$

It remains to specify the reference platform r_t. We will assume that the median voter is "backward looking" in that $r_t = \tilde{x}_{t-1}$, where $\tilde{x}_{t-1}$ is the *recalled* equilibrium policy i.e.

the voter's recollection of the platform of the winning party in the previous period.

We model imperfect recall by assuming that the actual winning platform at $t-1$ is scaled by a random factor ε_t i.e.

$$\tilde{x}_{t-1} = \varepsilon_t x_{t-1}, \tag{3}$$

where ε_t is drawn from a continuously differentiable cumulative distribution function F with support $[0, \infty)$, and a mean of $1 + b$, where $b > -1$ is the degree of bias in the recall. So, ε_t captures the randomness, or imperfectness, of recall, and is realized at the beginning of period t. Specifically, the reference point of the voter at t for evaluation of policy platforms will be $\varepsilon_t x_{t-1}$, not x_{t-1}. Also, $\tilde{\varepsilon}$ is the median value of ε_t; we allow for skewed distributions, such as the exponential i.e. $1 + b \neq \tilde{\varepsilon}$. If $\varepsilon_t \equiv 1$, we have the case of *perfect recall*. The role of imperfect recall is to generate nontrivial dynamics in the evolution of platforms: this point is further discussed in Section 3.5.

Finally, a key assumption is that political parties *do not observe* the stochastic shock ε_t, a reasonable assumption, as it is a mental state of the median voter. This implies that when political parties calculate the expected utility of the median voter (and thus their own election probability), they average over all ε_t and thus over all possible values of the voter's reference point. This is explained in more detail in Section 2.4 below.

2.4 Win Probabilities

Here, we characterize the probability p_t that party R wins the election as perceived by the political parties. We have assumed that the median voter does not use weakly dominated strategies, implying that she votes sincerely. So, from (2), the voter will vote for party R at t, given platforms $x_{L,t}, x_{R,t}$, if and only if

$$u(x_{R,t}; \varepsilon_t x_{t-1}) \geq v_t + u(x_{L,t}; \varepsilon_t x_{t-1}). \tag{4}$$

So, p_t is the probability that the median voter votes for R, conditional on x_{t-1}, as parties do not observe ε_t. From (4), p_t is therefore just the probability that v_t is less than the expectation with respect to ε_t of the utility difference $u(x_{R,t}; \varepsilon_t x_{t-1}) - u(x_{L,t}; \varepsilon_t x_{t-1})$. Generally, we write $\mathbb{E}u(x_t; x_{t-1})$ for the expected utility of the median voter with respect to ε_t, given x_{t-1} fixed. The explicit formula for this is given by equation (A.2) in the Appendix. Then, from the uniform distribution of v_t, as long as $p_t \in (0, 1)$, we have

$$p_t = p(x_{R,t}, x_{L,t}; x_{t-1}) = \frac{1}{2} + \rho \left[\mathbb{E}u(x_{R,t}; x_{t-1}) - \mathbb{E}u(x_{L,t}; x_{t-1}) \right], \tag{5}$$

So, we see that the greater the value of ρ, the more responsive is the election probability to platform changes. Then a sufficient condition for $p_t \in (0, 1)$ in the domain of undominated strategies is that ρ not be too large:

A1. $\frac{1}{2} > \rho\lambda$.

In Lemma 1 in the Appendix, we show that given Assumption **A1**, $p_t \in (0,1)$ for all $x_{R,t}, x_{L,t} \in [-1,1]$ and $x_{t-1} \in \Re$.

2.5 Party Payoffs

As is standard, parties have a payoff for holding office, denoted M. Parties also have policy preferences, with party L having an ideal point of -1 and party R an ideal point of 1. Parties are assumed, like the voter, to have absolute value preferences i.e. $u_R(x) = -|x-1|$, $u_L(x) = -|x+1|$. This assumption will be generalized and discussed in Section 3.4.

In any period t, the expected payoffs for the parties are calculated in the usual way as the probability of winning, times the policy payoff plus M plus the probability of losing, times the resulting policy payoff. So, the parties' payoffs are:

$$\begin{aligned} \pi_{R,t} &= \pi_R(x_{R,t}, x_{R,t}; x_{t-1}) = p_t[u_R(x_{R,t}) + M] + (1-p_t)u_R(x_{L,t}), \\ \pi_{L,t} &= \pi_L(x_{R,t}, x_{R,t}; x_{t-1}) = p_t u_L(x_{R,t}) + (1-p_t)[u_L(x_{L,t}) + M], \end{aligned} \tag{6}$$

where $p_t = p(x_{R,t}, x_{L,t}; x_{t-1})$. Note that party payoffs depend on x_{t-1} via p_t. We assume that both parties are forward-looking, with a discount factor δ. It is convenient to model party (or elite) polarization in this setting by a greater weight on the policy outcome, i.e., a smaller weight on the office payoff, or a smaller M.

Finally, we want to rule out the uninteresting case where the incentive to converge to the median voter's ideal point (as measured by M) is so large that parties do not choose different platforms in equilibrium. At the same time, we want to ensure that the equilibrium platforms in the T-period game are between 0 and 1 in absolute value. So, we will assume:

A2. $\frac{1}{2\lambda\rho} > M > \frac{1}{2\rho} - 2$.

This assumption has the following interpretation. The first inequality says, in the case of party R, that the benefit of a small increase in x_R from zero at the equilibrium election probability of one half, exceeds the expected loss from a lower probability of holding office, which is proportional to M. This ensures that the equilibrium platforms are greater than 0 in absolute value. The second inequality ensures that equilibrium platforms are less than 1 in absolute value.[11]

[11] To ensure that this range of values for M is not empty, we need $\lambda < \frac{1}{1-4\rho}$, which requires $\rho < 0.25$.

3 Multi-Period Electoral Competition

3.1 Equilibrium of the One-Period Game

It is analytically convenient to first analyze the one-period ($T = 1$)version of the game where there is a fixed policy x_0 set in the previous period. In this case, we can, without loss of generality, drop all period subscripts. In this case, we characterize Nash equilibria. Formally, a Nash equilibrium is a pair (x_R^*, x_L^*) where x_R^* maximizes $\pi_R(x_R, x_L^*; x_0)$, and similarly x_L^* maximizes $\pi_L(x_R^*, x_L; x_0)$, where the payoffs π_R, π_L are defined in (6). Also, we say that a Nash equilibrium is *symmetric* if $x_R^* = -x_L^*$. Finally, to lighten the notation, set $s = |x_0|$; then $-s$ is the payoff of the median voter from the previous period's policy platform and so s captures the intertemporal linkage between periods. We can then prove the following very useful intermediate result.

Proposition 1. *Given A1, A2, for any initial policy $x_0 \in \Re$, a Nash equilibrium always exists in the one-period game, and moreover, this equilibrium is unique and symmetric. This symmetric Nash equilibrium $x_R^* = -x_L^* = x^*$ is characterized as follows. For all $s > 0$, $x^* = \phi(s)$ is the unique solution of the implicit equation in x:*

$$\frac{1}{2} - \rho \left[(\lambda - 1)F\left(\frac{x}{s}\right) + 1\right](2x + M) = 0. \tag{7}$$

For $s = 0$, the symmetric Nash equilibrium is $x^* = \frac{1}{4\rho\lambda} - \frac{M}{2} > 0$. For all $s \in (0, \infty)$, $x^* = \phi(s)$ is increasing in s. For any $s \in (0, \infty)$, $\phi(s)$ lies between the values $\lim_{s\to 0} \phi(s) \equiv x^- = \frac{1}{4\rho\lambda} - \frac{M}{2} > 0$ *and* $\lim_{s\to\infty} \phi(s) \equiv x^+ = \frac{1}{4\rho} - \frac{M}{2} < 1$.

This result ensures that a unique and symmetric equilibrium always exists in the one-period version of this game, no matter what the value of x_0. This matters because this will also imply the existence of a unique and symmetric equilibrium in the T-period game. Moreover, for any $s > 0$, the equilibrium platforms are strictly between x^-, x^+. As we shall see, this implies that all equilibrium platforms in the T-period game also lie between these limits.

Here, x^- is the equilibrium platform if the game is played entirely in the loss domain of the median voter, i.e., where the median voter has preferences over policies of $-\lambda |x|$, and similarly x^+ is the equilibrium platform if the game is played entirely in the gain domain of the median voter, i.e., where the median voter has preferences over policies of $-|x|$. Note that $x^- < x^+$ because, in the loss domain, the parties are punished more heavily by the median voter (in terms of a lower election probability) for a given deviation from $x = 0$.

The key result here for the dynamics in the T-period game is that x is strictly increasing in s, i.e., the further the initial policy is away from the median voter's ideal point of zero, the larger in absolute value are the equilibrium platforms $\phi(s), -\phi(s)$. The intuition for this is as follows.

First, note that for a fixed x_0, the realization of ε places a party's chosen platform in either the domain of gains or losses for the median voter. So, at a fixed x_0, the probability that a platform x_R for (say) party R is in the gain domain (i.e. that $x_R \leq \varepsilon s$) is $\mathbb{P}(\varepsilon \geq x_R/s)$ which is clearly increasing in s. In turn, the more likely is x_R to be in the gain domain, the less the electoral penalty (in terms of a lower election probability) from a small increase in x_R. As a consequence, party R will choose a higher x_R in equilibrium, the further is the absolute value of the previous period's policy from zero.

This point can be made more formally. In the proof of Proposition 1, it is shown that the reduction in party R's election probability when the platform x_R is increased, i.e. moved away from the median voter's ideal point, is

$$\frac{\partial p}{\partial x_R} = -\rho \left[(\lambda - 1) F\left(\frac{x_R}{s}\right) + 1 \right] < 0. \tag{8}$$

But then from (8):

$$\frac{\partial^2 p}{\partial x_R \partial s} = \rho(\lambda - 1) f\left(\frac{x_R}{s}\right) \frac{x_R}{s^2} > 0. \tag{9}$$

So, (9) says that *this reduction in the election probability is lower, the higher is s*. This feature creates the dynamic linkage between periods. Finally, note from (7) that when $\lambda = 1$, $x^* = x^+$ and is thus independent of x_0. In other words, without loss-aversion, there is no dynamic linkage between periods.

3.2 Equilibrium of the T-Period Game

Our equilibrium concept will be subgame-perfect equilibrium. Then, as we will see, the subgame-perfect equilibrium will be unique and symmetric in the sense that $x_{R,t} = -x_{L,t}$, for all t. A possible complication in this case is the existence of dynamic incentives. Generally, when the outcome is x_t at t, this helps determining r_{t+1}. Now, if the expected equilibrium payoff in $t+1$ depends directly on r_{t+1}, then forward-looking parties will take into account the effect of their choice of x_t on their expected equilibrium payoff in $t+1$ when choosing their actions at t. While interesting, dynamic incentives make characterizing the path of equilibrium platforms very complex.

However, it turns out that in our setting, parties behave *as if* they are completely myopic, even though their payoffs are forward-looking. The argument is by backward induction. In the last period, by Proposition 1, the equilibrium platforms will be $x_T^*, -x_T^*$, with $s = |x_{T-1}|$. So, the expected payoffs from symmetric equilibrium in the final period are then

$$\frac{M}{2} + \frac{1}{2} \left[u_K(x_T^*) + u_K(-x_T^*) \right], \; K = R, L \tag{10}$$

as each of $x_T^*, -x_T^*$ occurs with equal probability. Also, by Proposition 1, $0 < x_T^* < 1$.

But then $u_R(x) = x - 1, u_L(x) = -(x + 1)$, and so:

$$u_R(x) + u_R(-x) = u_L(x) + u_L(-x) = -2. \tag{11}$$

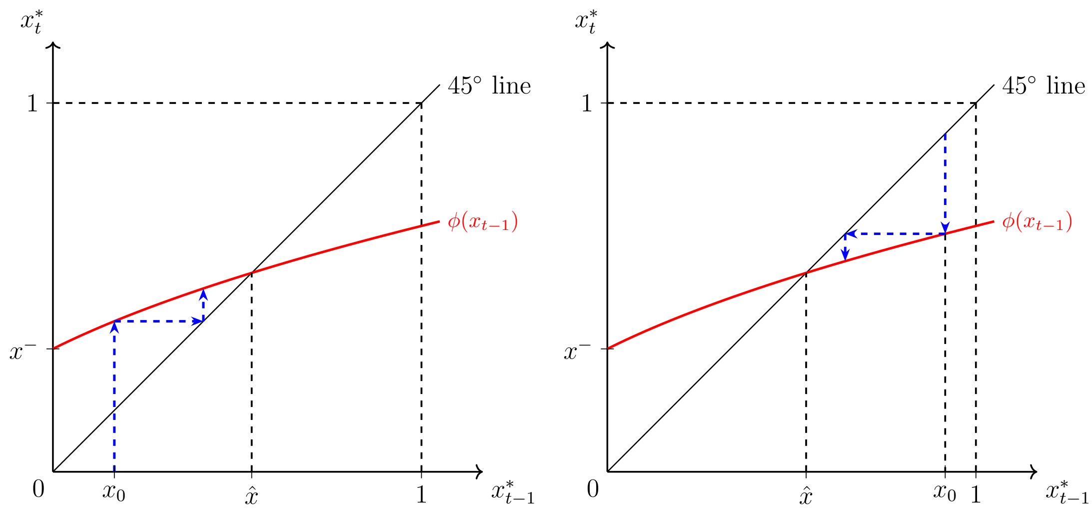
Then, we can combine (10), (11) to conclude that the expected continuation payoffs from symmetric equilibrium in the final period are simply $\frac{M}{2} - 1$ for each party, and thus independent of x_{T-1}.

This, of course, implies that the equilibrium at $T - 1$ will also be the equilibrium in the static game as in Proposition 1, with $s = |x_{T-2}|$, and so on. So, we can solve for the political equilibrium as a sequence of static problems where only the median voter's reference point is varying over time. This implies that the current period's equilibrium platforms will depend on the expectation of the median voter's reference point, which is in turn the previous period's equilibrium platform, as recalled by the median voter. This creates a dynamic linkage between equilibrium platforms in successive periods. Formally, we can characterize this linkage as follows:

Proposition 2. *There is a unique and symmetric equilibrium $x_{R,t} = -x_{L,t} = x_t^*$, $t = 1, \ldots, T$. The equilibrium platform x_t^* at t is given by*

$$x_t^* = \phi(x_{t-1}^*), \tag{12}$$

where $\phi(.)$ is defined in Proposition 1. Because $\phi(.)$ is strictly increasing, there is monotonic convergence to the unique long-run platform $\hat{x}$ that solves $\hat{x} = \phi(\hat{x})$.



(a) Polarization with moderate x_0 (b) Depolarization with extreme x_0

Figure 2: Equilibrium in the T-Period Game

The evolution of the equilibrium path over time is shown in Figure 2 above. Figure 2(a) shows the case where, starting from a relatively moderate historically determined platform $x_0 < \hat{x}$, both parties have an incentive to choose more polarized platforms $x_{R,1} = -x_{L,1} = x_1^* > x_0$. This, in turn, leads to an outward shift in the expected value of the median voter's reference point, which creates an incentive for further polarization in the parties' platforms, and so on. A reverse process of depolarization occurs if the initial platform is extreme, i.e. $x_0 > \hat{x}$, as shown in Figure 2(b).

Note that whatever the starting point, the unique long-run platform $\hat{x}$ solves $\hat{x} = \phi(\hat{x})$. Solving (7) for $\hat{x}$, noting that in this case $F(x/s) = F(1)$, we then get:

$$\hat{x} = \frac{1}{4\rho[(\lambda - 1)F(1) + 1]} - \frac{M}{2}. \tag{13}$$

Note that $x^- < \hat{x} < x^+$ so from A2, it follows that $\hat{x} \in (0, 1)$.[12] To interpret this long-run equilibrium, note that we can write

$$F(1) - 0.5 = \underbrace{F(1) - F(1+b)}_{\text{bias, } -\beta} + \underbrace{F(1+b) - 0.5.}_{\text{skewness, } \kappa} \tag{14}$$

So, $\beta > 0$ if the median voter has a positive bias in recall, that is, $b > 0$, and vice versa. The skewness parameter is positive if ε_t is skewed to the right, i.e. $\mathbb{E}\varepsilon_t = 1 + b > \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is the median value of ε_t.

Then, combining (13), and (14), we get:

$$\hat{x} = \frac{1}{4\rho[(\lambda - 1)(0.5 + \kappa - \beta) + 1]} - \frac{M}{2}. \tag{15}$$

So, by inspection of (15) we can summarize as follows:

Proposition 3. *With imperfect recall, the equilibrium platforms converge monotonically over time to the long-run equilibrium $\hat{x}$. The long-run equilibrium platforms are less polarized, the larger is loss-aversion λ, or the skewness of ε_t. The long-run equilibrium platforms are more (less) polarized if there is positive (negative) bias b in recall.*

The intuition for these results is fairly straightforward. The higher the degree of loss-aversion, the more the median voter dislikes polarization of platforms, and so the less polarization there will be in long-run equilibrium. If there is a positive bias in recall, the recalled reference platform from last period will, other things equal, be larger, so electoral competition is more likely to occur in the gain domain for the voter, leading to more polarization in the long run. The reverse applies if there is a negative bias in recall. Finally, the skewness of ε_t matters for the long-run equilibrium; for example, if ε_t is skewed to the right, more than half the realizations of ε_t will be below the mean,

[12] As $0 < F(1) < 1$, $\frac{1}{4\rho} - \frac{M}{2} > \hat{x} > \frac{1}{4\rho\lambda} - \frac{M}{2}$.

which for a given mean, makes it relatively likely that the median voter will evaluate the platforms using a smaller reference point. This, in turn, means that electoral competition is more likely to take place in the loss domain, leading to a smaller $\hat{x}$.

3.3 Elite Polarization and the Magnification Effect

The discussion in the Introduction suggests that, in the US at least, there is robust evidence for elite polarization. Here, we analyze how an increase in elite (de)polarization interacts with voter loss-aversion to create a dynamic process, which magnifies the initial effect of elite (de)polarization on party platforms. As already noted, a parsimonious way to model elite polarization is as a decrease in the weight that parties place on the value of office relative to policy, i.e., a fall in M.

The effects of this are shown in Figure 3 below. There is initially a long-run political equilibrium at $(x_0, -x_0)$ for some $x_0 > 0$. From (9), it is clear that a decrease in M shifts the equilibrium mapping $\phi(\cdot)$ upwards. In the short run, i.e., in the following period, with the median voter's reference point fixed at $\tilde{x}_1 = \varepsilon_1 x_0$, the parties' equilibrium platforms move to $(x_1^*, -x_1^*)$. But, of course, this then shifts the median voter's reference point upwards, making the voter more tolerant of polarization. This, in turn, leads to more polarized platforms and so on, until a new long-run equilibrium is reached. In other words, the shift of the voter's reference point magnifies the initial effect of elite polarization.

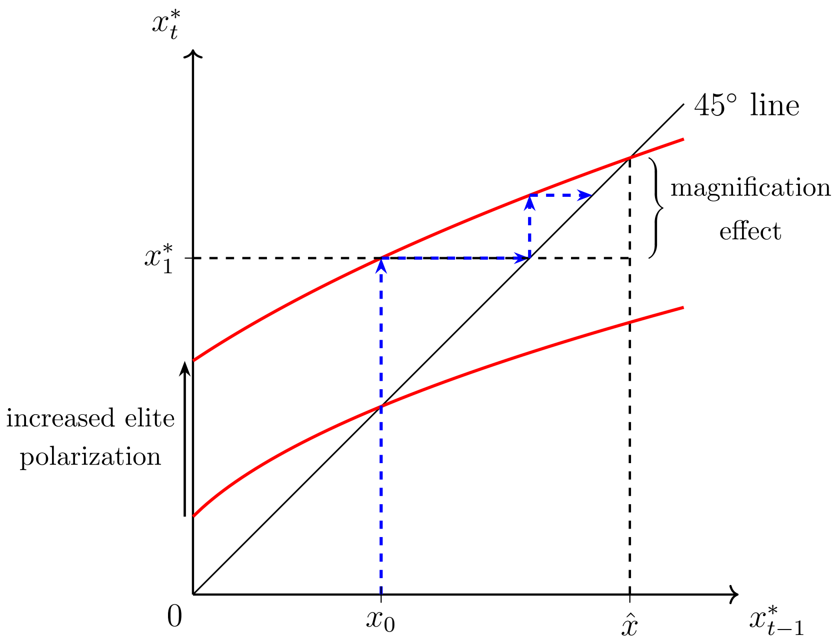


Figure 3: Equilibrium Platform Dynamics with Elite Polarization

This magnification effect is consistent with continued monotonic changes in (de)polarization over long periods, as shown in Figure 1.

3.4 More General Preferences

So far, we have assumed that parties have absolute value preferences. These have the important implication that parties are risk-neutral over symmetric equilibrium outcomes in the one-shot game. As explained in Section 3.2 above, this in turn implies that in the T-period game, parties behave *as if* they are myopic. What happens if the parties have more general preferences? In this case, in the one-shot game, Proposition 1 is generalized straightforwardly, except for the fact that we cannot rule out asymmetric equilibria.

Suppose that payoffs of the L and R party members are $u_L(x) \equiv -l(|x+1|)$, $u_R(x) \equiv -l(|x-1|)$ respectively, where l is twice differentiable, strictly increasing, symmetric and convex in $|x - x_i|$, and $l(0) = l'(0) = 0$. This specification allows for parties to be risk-neutral ($l'' = 0$) or strictly risk-averse ($l'' > 0$) over policy outcomes. In this more general setting, it can be shown that the mapping from $s = |x_0|$ to the equilibrium platforms is $x_R^* = -x_L^* = x^*$ where $x^* = \phi(s)$ is the unique solution of the implicit equation in x:

$$\frac{1}{2}u_R'(x) - \rho\left[(\lambda-1)F\left(\frac{x}{s}\right)+1\right][u_R(x) - u_R(-x) + M] = 0. \tag{16}$$

Comparing (16) and (7), we see that $u_R(x)$ replaces $-|1-x|$, $u_R'(x)$ replaces 1 and so on. Subject to this change, Proposition 1 continues to apply, given the appropriate changes in the definitions of x^-, x^+, except for the fact that we cannot now rule out asymmetric equilibria because the one-shot game between the parties is no longer zero-sum.

Now, in the T-period game, parties are generally risk-averse over the next period's equilibrium outcomes. Specifically, parties at date t would prefer the lottery $(x_{t+1}, -x_{t+1})$ with probability $(0.5, 0.5)$ to be less risky i.e. a smaller value of x_{t+1}, so this creates a *dynamic incentive* to reduce x_{t+1} by moderating the current platform $x_{R,t}$ or $x_{L,t}$. A general T-period analysis of this game is analytically intractable, although some results are available in the 2-period case in an earlier version of this paper (Lockwood et al., 2022). However, we conjecture that the basic "positive feedback loop" between the reference point and the equilibrium platforms would continue to operate.

An alternative is to assume, following Callander and Carbajal (2022), that parties are myopic, in which case this dynamic incentive disappears. In that case, Proposition 3 continues to apply, modulo the new definition of $x^* = \phi(s)$ in (16). So, with risk-averse but myopic parties, we can also have episodes of gradual (de)polarization.

3.5 The Role of Imperfect Recall

To understand why imperfect recall is needed to generate interesting dynamics, suppose that recall were perfect, i.e. $\varepsilon_t \equiv 1$. In this case, it can be shown that the mapping from the previous period's equilibrium to the current period's one- the red line in Figure 2 - collapses to the 45 degree line between the lower and upper bounds, as shown in Figure 4.[13]

[13] A formal proof is given in Proposition 1 from Lockwood and Rockey (2020).

The lower and upper bounds x^-, x^+ are defined as in Proposition 1, and, as already noted, are the symmetric equilibria in the one-shot games where the median voter's payoff is entirely in the loss and gain domain, respectively.

The intuition for this is as follows. If $\varepsilon_t \equiv 1$, then political parties now know the exact value of the median voter's reference point, which is x_{t-1}. Moreover, from (1), the median voter's utility, as a function of x_t, is “kinked" i.e. not differentiable at $x_t = x_{t-1}$. Then, from (5), this also generates a kink in the parties' perceived election probabilities as a function of x_t. This kink in turn implies that if $x^- < x_{t-1} < x^+$, the best response of party R to $x_{L,t} = -x_{t-1}$ is to neither increase nor decrease $x_{R,t}$ from x_{t-1} and vice versa for party L i.e. there is a “zone of inaction" for each party. This generates the mapping along the 45 degree line.

Figure 4 shows the evolution of platforms over time for three different initial values x_0, x_0', x_0''. In each case, there is a “one-step” convergence to a long-run equilibrium. So, whatever x_0, x_1^* will move to some value in the interval $[x^-, x^+]$ and then stay there in all subsequent periods. So, the long-term equilibrium is *completely predetermined* by the initial condition, and thus cannot be affected by other parameters of the model. This is a drawback in the sense that we cannot explain the long-run outcome as depending on the underlying parameters of the model λ, b, M, ρ.

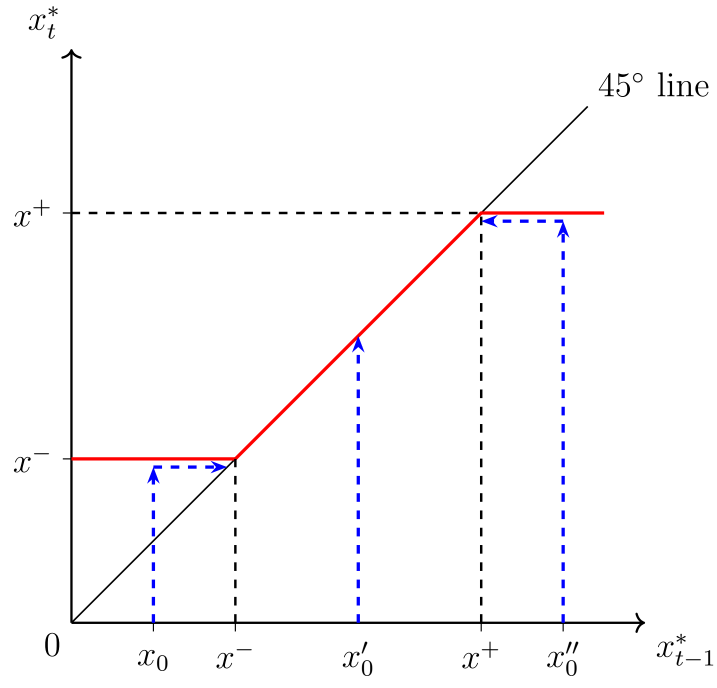


Figure 4: Equilibrium Platform Dynamics with Perfect Recall

4 Simulations

We have shown that a model of dynamic political competition that incorporates loss-aversion and imperfect recall can capture two key features of Figure 2 – periods of depolarization and, from a long-term perspective, periods of consistent increases in polarization due to the magnification effect.

We now study the extent to which our model can capture the third feature of the data - periods of relatively gradual change and occasional periods of rapid change. Looking at data for the US house, on which we focus for simplicity, in Figure 2 we see that there was a sharp fall in polarization in the 1920s, and a sharp increase around 1990, although in both cases there were pre-existing trends in the same direction. The increase in polarization in the early 1990s, is sharp and sudden, and coincides with the 1994 mid-term election which saw the Republican party take control of both the House and the Senate with gains of 54 and 8 seats respectively, under the leadership of Newt Gingrich and his *Contract with America*.

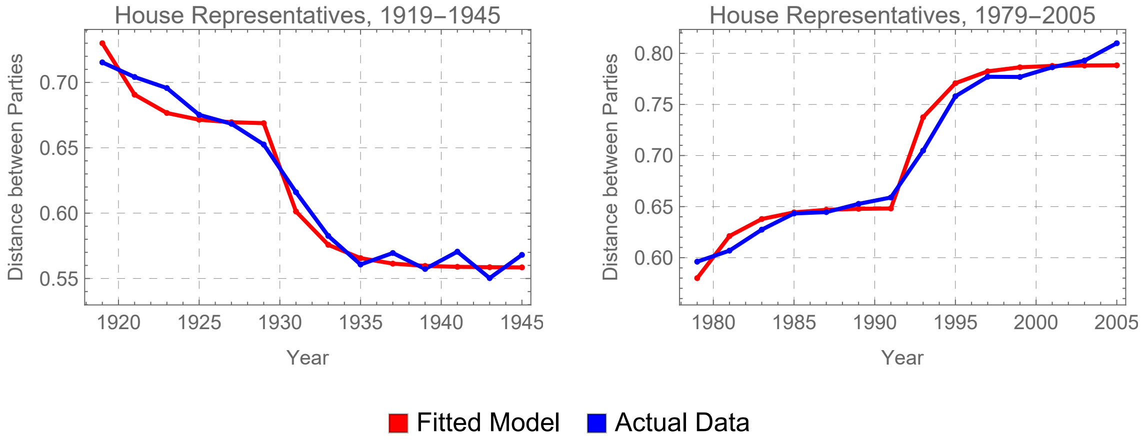
The decline in polarization in the 1920s is a little different. There is a downward trend from 1920-1927, with an acceleration around 1927, and a period of stability from 1935 onwards. This decline in polarization is less-well understood than the rise in the 1990s. Chatfield et al. (2021) provide evidence that it is driven first by the rise of the Farm Bloc, a precursor to the Conservative Coalition, and later by the progressive coalition. As with the early 1990s, changes in representatives drove this change. In particular, they point to the role of changing mass-politics and particularly the consequent election of more moderate Democrats (relative to other Democrats) elected to replace Republicans and vice versa.

It is important to emphasize that this is not a conventional quantitative exercise. Our model is comparatively abstract, designed to elucidate key mechanisms through which loss aversion affects political competition, and to be analytically tractable. As such, it involves a minimal number of parameters, which in a number of cases do not correspond to observable quantities, in contrast to most quantitative models in political economy or otherwise. Because of this, we do not attempt to calibrate our model to the data. Instead, we show that, as written, it captures the key quantitative and qualitative features of the two periods 1919-1945 and 1979-2005.

We focus, as before, on the shock to M, the relative value of office rents. This is equivalent to a (symmetric) shock to the political preferences of the party elites, which here are normalized to 1. Figure 5 overlays the simulated equilibrium paths of the model (in red) on the empirical data for mean polarization for the US House for 1919-1945 and 1979-2005 (in blue). In both cases, we set $\lambda = 2.1$, $\rho = 0.2$, and the recall shocks follow an exponential distribution with mean 1.5.

In broad terms, the fact that for both periods the blue and red series are close to each other suggests that the model captures well both the periods of rapid decrease (increase) in polarization and also the periods of comparative stability before and after.

Figure 5: Episodes of Rapid Polarization and Depolarization in the US House



Notes: The blue lines depict mean polarization for the US house calculated using the Voteview data (Lewis et al., 2022). The red lines are the the simulated equilibrium paths of the model assuming $\lambda = 2.1$, $\rho = 0.2$ and $b = 0.5$. For the left-hand figure, M changes from 0.96 to 1.07 in 1927. For the right-hand figure, M changes from 0.98 to 0.84 in 1993.

Looking more closely at the left-hand panel, we see a large fall in observed polarization, shown by the blue line, between 1925 and 1935. The simulated equilibrium path (in red) models this as an assumed increase in M, the rents from office, 0.96 to 1.07 in 1927. Likewise, we can see the rapid increase in polarization between 1991-2001 in the right-hand panel of the empirical data. This is modeled as an assumed negative shock to M in 1993, which reduces office rents from 0.98 to 0.84. Note that the key qualitative features of the data are captured: a steady period; a rapid increase or decrease in polarization; and then a new period of comparative stability. The fit is also good in quantitative terms—the mean absolute deviation is 0.0096 and 0.0105 for the two periods, respectively, which amounts to 1.3% or 1.7% of mean polarization. Our results are not sensitive to the particular set of parameters chosen, as can be seen in Figures B.1a to B.1c in the Appendix.

It is the case, however, that the simulated (de)polarization trajectory converges to the new equilibrium more rapidly than we observe in the data. This could reflect many factors that are not captured by our simple model, but one possibility is that the shock may operate more slowly than in our model. This would be the case if, as argued by Poole (2007); Chatfield et al. (2021), changes in parties' positions are largely driven by the turnover of representatives rather than a shift in the views of existing representatives. In this case, given, for example, that some US House representatives have substantial incumbency advantages, we would expect a more gradual change. In Figure B.2 in the Appendix, we show results modeling this by replacing the sudden jump in the office rent, M, with a gradual change. We can see that the model now is much better able to fit the data with a smaller immediate change and a more gradual convergence to the new long-run equilibrium, both improving the fit of the model to the data.[14]

5 Conclusions

This paper has explored the implications of voter loss-aversion for the dynamics of electoral competition in a simple Downsian model of repeated elections. When the representative voter is backward-looking, i.e. the reference point is the last period's recalled policy, interesting dynamics emerge when the voter has *imperfect recall* about that policy. Then, the interplay between the median voter's reference point and political parties' choice of platforms generates a dynamic process of (de)polarization, where platforms monotonically converge over time to a new long-run equilibrium. Exogenous shifts in elite (party) (de)polarization lead to a dynamic process of further (de)polarization, consistent with US evidence.

Whilst the model is comparatively abstract and is not designed for simulation it nevertheless is able to explain over 98% of the fall in polarization in the 1920s and '30s and the rise from the mid-1990s onwards. It would be valuable for future research to develop methods to directly measure the primitives of the model. This could include surveys or lab-in-the-field techniques such that voters' preferences, reference points, and their perceptions of representatives' positions could be captured in a consistent policy space such that they were comparable over time and across districts. This might allow for analysis of the causal relationships between elite and mass polarization.

[14]It would, of course, be possible to choose a series of shocks such that our model reproduced the entire time-series but this exercise is not informative in the absence of obvious constrains with which to reduce the number of degrees of freedom.

References

Alesina, Alberto and Francesco Passarelli, "Loss Aversion in Politics," *American Journal of Political Science*, 2019.

Battaglini, Marco, "A dynamic theory of electoral competition," *Theoretical Economics*, 2014, *9* (2), 515–554.

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs, "Bad is stronger than good.," *Review of General Psychology*, 2001, *5* (4), 323–370.

Biais, Bruno and Enrico Perotti, "Machiavellian privatization," *American Economic Review*, 2002, *92* (1), 240–258.

Blank, Hartmut, Volkhard Fischer, and Edgar Erdfelder, "Hindsight Bias in Political Elections," *Memory*, January 2003, *11* (4-5), 491–504.

Bloom, Howard S and H Douglas Price, "Voter response to short-run economic conditions: The asymmetric effect of prosperity and recession," *American Political Science Review*, 1975, *69* (04), 1240–1254.

Callander, Steven and Juan Carlos Carbajal, "Cause and Effect in Political Polarization: A Dynamic Analysis," *Journal of Political Economy*, April 2022, *130* (4), 825–880.

Chatfield, Sara, Jeffery A. Jenkins, and Charles Stewart, "Polarization Lost: Exploring the Decline of Ideological Voting in Congress after the Gilded Age," *Journal of Historical Political Economy*, 2021, *1* (2), 183–214.

Dahl, Robert A, *A preface to democratic theory*, Vol. 10, University of Chicago Press, 1956.

Diermeier, Daniel and Christopher Li, "Electoral control with behavioral voters," *The Journal of Politics*, 2017, *79* (3), 890–902.

_ **and** _ , "Partisan affect and elite polarization," *American Political Science Review*, 2019, *113* (1), 277–281.

Fiorina, Morris P, Samuel J Abrams, and Jeremy C Pope, *Culture war? The myth of a polarized America*, New York: Pearson Longman, 2005.

Forand, Jean Guillaume, "Two-party competition with persistent policies," *Journal of Economic Theory*, 2014, *152* (1), 64–91.

Fudenberg, Drew and Jean Tirole, *Game theory*, MIT press, 1991.

Gentzkow, M, "Polarization in 2016," *Toulouse Network of Information Technology whitepaper*, 2016, pp. 1–23.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech ," *Econometrica*, 2019, *87* (4), 1307–1340.

Grillo, Edoardo, "The hidden cost of raising voters' expectations: Reference dependence and politicians' credibility," *Journal of Economic Behavior & Organization*, 2016, *130*, 126–143.

_ **and Carlo Prato**, "Reference points and democratic backsliding," *American Journal of Political Science*, 2020.

Hare, Christopher and Keith T. Poole, "The polarization of contemporary American politics," *Polity*, 2014, *46* (3), 411–429.

Kappe, Roland, "Asymmetric Evaluations: Government Popularity and Economic Performance in the United Kingdom," *Electoral Studies*, 2018, *53*, 133–138.

Klein, Jill G., "Negativity Effects in Impression Formation: A Test in the Political Arena," *Personality and Social Psychology Bulletin*, 1991, *17* (4), 412–418.

Kőszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, November 2006, *121* (4), 1133–1165.

_ **and** _ , "Reference-Dependent Risk Attitudes," *American Economic Review*, September 2007, *97* (4), 1047–1073.

Lau, Richard R., "Two Explanations for Negativity Effects in Political Behavior," *American Journal of Political Science*, February 1985, *29* (1), 119.

Levy, Gilat, Ronny Razin, and Alwyn Young, "Misspecified politics and the recurrence of populism," *American Economic Review*, 2022, *112* (3), 928–962.

Lewis, Jeffrey B, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet, "Voteview: Congressional roll-call votes database," *See https://voteview. com/(accessed 12 October 2022)*, 2022.

Lockwood, Ben and James Rockey, "Negative voters? electoral competition with loss-aversion," *The Economic Journal*, 2020, *130* (632), 2619–2648.

_ **, Minh Le, and James Rockey**, "Dynamic Electoral Competition with Voter Loss-Aversion and Imperfect Recall," *Available at SSRN 4052626*, 2022.

McCarty, Nolan, *Polarization: What everyone needs to know®*, Oxford University Press, 2019.

— , **Keith T. Poole, and Howard Rosenthal**, "Does gerrymandering cause polarization?," *American Journal of Political Science*, 2009, *53* (3), 666–680.

Nannestad, Peter and Martin Paldam, "The grievance asymmetry revisited: A micro study of economic voting in Denmark, 1986–1992," 1997.

Nunnari, Salvatore and Jan Zápal, "Dynamic elections and ideological polarization," *Political Analysis*, 2017, *25* (4), 505–534.

Passarelli, Francesco and Guido Tabellini, "Emotions and political unrest," *Journal of Political Economy*, 2017, *125* (3), 903–946.

Pohl, Rüdiger F and Edgar Erdfelder, "Hindsight bias in political decision making," in "Oxford Research Encyclopedia of Politics" 2019.

Poole, Keith T, "Changing minds? Not in congress!," *Public Choice*, 2007, *131* (3-4), 435–451.

Poole, Keith T. and Howard Rosenthal, "The Polarization of American Politics," *The Journal of Politics*, 1984, *46* (4), 1061–1079.

— **and** — , *Ideology and Congress*, 1st ed., Edison, NJ: Transaction Publishers, 2006.

Prato, Carlo, "Electoral competition and policy feedback effects," *The Journal of Politics*, 2018, *80* (1), 195–210.

Schuett, Florian and Alexander K. Wagner, "Hindsight-Biased Evaluation of Political Decision Makers," *Journal of Public Economics*, December 2011, *95* (11), 1621–1634.

Soroka, Stuart N., "Good news and bad news: Asymmetric responses to economic information," *Journal of Politics*, 2006, *68* (2), 372–385.

van Elsas, Erika J., Rozemarijn Lubbe, Tom W.G. van der Meer, and Wouter van der Brug, "Vote Recall: A Panel Study on the Mechanisms That Explain Vote Recall Inconsistency," *International Journal of Public Opinion Research*, March 2014, *26* (1), 18–40.

Wixted, John T. and Ebbe B. Ebbesen, "On the Form of Forgetting," *Psychological Science*, November 1991, *2* (6), 409–415.

A Appendix A

Lemma A1. *In the one-period game, given Assumption A1, $p \in (0,1)$ for all $x_R, x_L \in [-1,1]$ and $x_0 \in \Re$.*

Proof. (i) We begin by developing a formula for the expected utility of the median voter in the one-period case. Note that $u(x) = -|x|$, and because ε is non-negative, the median voter's utility from the "recalled" platform $u(\tilde{x}_0) = -\varepsilon |x_0| \equiv -\varepsilon s$. Combining these with (1), we see that the utility of the median voter from platform x given the status quo x_0 can be written as

$$u(x; \varepsilon x_0) = \begin{cases} \lambda(\varepsilon s - |x|), & \varepsilon s < |x|, \\ \varepsilon s - |x|, & \varepsilon s \geq |x|. \end{cases} \tag{A.1}$$

So, taking expectations in (A.1) over ε, for all $x \in \Re$, we have[15]

$$\mathbb{E}u(x; x_0) = \lambda \int_0^{|x|/s} (\varepsilon s - |x|) f(\varepsilon) d\varepsilon + \int_{|x|/s}^{\infty} (\varepsilon s - |x|) f(\varepsilon) d\varepsilon. \tag{A.2}$$

(ii) Note from (A.2) that $\mathbb{E}u(x; x_0) = \mathbb{E}(-x; x_0)$, all $x, x_0 \in \Re$. Using this fact, and as the parties are symmetric, if there exist $x_R, x_L \in [-1,1]$ and $x_0 \in \Re$ such that $p(x_R, x_L; x_0) = 0$, then $p(-x_L, -x_R; x_0) = 1 - p(x_R, x_L; x_0) = 1$. Hence, it is sufficient to check that under Assumption A1, $p < 1$ for all $x_R, x_L \in [-1,1]$ and $x_0 \in \Re$. Recall from (5) that $p \in (0,1)$ if and only if

$$p = \frac{1}{2} + \rho \left[\mathbb{E}u(x_R; x_0) - \mathbb{E}u(x_L; x_0) \right] \in (0,1). \tag{A.3}$$

Notice that for $x_R, x_L \in [-1,1]$, $\mathbb{E}u(x_R; x_0)$ is bounded above by $\mathbb{E}u(0; x_0)$ and $\mathbb{E}u(x_L; x_0)$ is bounded below by $\mathbb{E}u(-1; x_0)$. From (A.2), it is easy to calculate that

$$\mathbb{E}u(0; x_0) = \mathbb{E}\varepsilon |x_0| = (1+b)s, \tag{A.4}$$

$$\begin{aligned} \mathbb{E}u(-1; x_0) &= \lambda \int_0^{1/s} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon + \int_{1/s}^{\infty} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon \\ &= (\lambda - 1) \int_0^{1/s} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon + \int_0^{\infty} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon \\ &= (\lambda - 1) \int_0^{1/s} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon - 1 + (1+b)s. \end{aligned} \tag{A.5}$$

Then we have

$$\begin{aligned} \mathbb{E}u(x_R; x_0) - \mathbb{E}u(x_L; x_0) &\leq \mathbb{E}u(0; x_0) - \mathbb{E}u(-1; x_0) \\ &= 1 - (\lambda - 1) \int_0^{1/s} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon \\ &\leq \lambda. \end{aligned} \tag{A.6}$$

In the last line, we have used the fact that $\int_0^{1/s} (-1 + \varepsilon s) f(\varepsilon) d\varepsilon$ is strictly negative and strictly

[15] As the utility of the median voter is continuous in s, we have $\mathbb{E}u(x; 0) = \lim_{x_0 \to 0} \mathbb{E}u(x; x_0) = -\lambda |x|$.

increasing in s and it approaches -1 as $s \to 0$. So, from (A.3), (A.6), we have

$$p \leq \frac{1}{2} + \rho\lambda < 1, \tag{A.7}$$

where the second inequality is from Assumption A1. □

Lemma A2. *In the one-period game, given Assumption A1, for party R (resp. L), any platform outside* $[0, 1]$ *(resp.* $[-1, 0]$*) is a strictly dominated strategy.*
Proof. Fix any $x_0 \in \Re$. We first show that for party R, any x such that $|x| > 1$ is a strictly dominated strategy. By rewriting (A.2), for any $s > 0$, we have[16]

$$\begin{aligned}
\mathbb{E}u(x; x_0) &= \lambda \int_0^{1/s} (\varepsilon s - |x|)dF(\varepsilon) + \lambda \int_{1/s}^{|x|/s} (\varepsilon s - |x|)dF(\varepsilon) + \int_{|x|/s}^{\infty} (\varepsilon s - |x|)dF(\varepsilon) \\
&< \lambda \int_0^{1/s} (\varepsilon s - 1)dF(\varepsilon) + \underbrace{\lambda \int_{1/s}^{|x|/s} (\varepsilon s - |x|)dF(\varepsilon)}_{(-)} + \int_{|x|/s}^{\infty} (\varepsilon s - 1)dF(\varepsilon) \\
&< \lambda \int_0^{1/s} (\varepsilon s - 1)dF(\varepsilon) + \underbrace{\int_{1/s}^{|x|/s} (\varepsilon s - 1)dF(\varepsilon)}_{(+)} + \int_{|x|/s}^{\infty} (\varepsilon s - 1)dF(\varepsilon) \\
&= \mathbb{E}u(1; x_0).
\end{aligned} \tag{A.8}$$

Combining with (A.3), it must be that $p(x, x_L) \leq p(1, x_L)$ where, for convenience, we suppress the dependence of p on x_0. Also, since $u_R(x) < u_R(1) = 0$, we have

$$\pi_R(1, x_L; x_0) - \pi_R(x, x_L; x_0) = \underbrace{[M - u_R(x_L)]}_{(+)}[p(1, x_L) - p(x, x_L)] - \underbrace{u_R(x)}_{(-)} p(x, x_L). \tag{A.9}$$

Notice that for $|x_L| > 1$, we have $p(1, x_L) > 0.5$, and for $|x_L| \leq 1$, by Lemma A1, we have $p(1, x_L) > 0$. Hence, given Assumption A1, $p(1, x_L) > 0$ for all $x_L \in \Re$. If $p(1, x_L) = p(x, x_L)$, then it must be that $p(x, x_L) > 0$; and if $p(x, x_L) = 0$, then it must be that $p(1, x_L) > p(x, x_L)$. In either case, we have $\pi_R(1, x_L; x_0) - \pi_R(x, x_L; x_0) > 0$, which implies, for party R, proposing its own ideal point, 1, strictly dominates any policy outside $[-1, 1]$ for any $x_L, x_0 \in \Re$. By symmetry, any policy outside [-1,1] is also a strictly dominated strategy of party L.

Fix any $1 \geq x' > 0$. By iterated elimination of the dominated strategy, we now need to show that, for party R, $-x'$ is a strictly dominated strategy for any $x_L \in [-1, 1]$ and $x_0 \in \Re$. From (A.2) we have $\mathbb{E}u(x'; x_0) = \mathbb{E}u(-x'; x_0)$. But, from (5), this implies

$$p(x', x_L) = p(-x', x_L). \tag{A.10}$$

Also, as $u_R(x) = -|x - 1|$, it follows that $u_R(x') > u_R(-x')$. Then we have

$$\pi_R(x', x_L; x_0) - \pi_R(-x', x_L; x_0) = \underbrace{[u_R(x') - u_R(-x')]}_{(+)} p(x', x_L) > 0. \tag{A.11}$$

where the last inequality follows the results that $p(x_R, x_L) > 0$ for all $x_R, x_L \in [-1, 1]$. Thus, for party R, x' strictly dominates $-x'$ for any $x_L \in [-1, 1]$ and $x_0 \in \Re$. By symmetry, any policy outside $[-1, 0]$ is also a strictly dominated strategy for party L, as required. □

Proof of Proposition 1. (i) First, we solve the one-period game for the case $x_0 = 0$. We show

[16] As $|x| > 1$, it is also true that $\mathbb{E}u(x; 0) = -\lambda|x| < -\lambda = \mathbb{E}u(1; 0)$

that when $x_0 = 0$, a Nash equilibrium exists and, moreover, this equilibrium is symmetric and uniquely defined. We take the limit as $s \to 0$ in (A.2) to get $\mathbb{E}u(x;0) = -\lambda |x|$. TThen,for any $x_L \in [-1,0]$, party R's probability of winning can be simply written as

$$p(x_R, x_L; 0) = \frac{1}{2} - \rho\lambda(|x_R| - |x_L|),$$

which is a linear function of x_R for all $x_R \in [0,1]$. Given $u_R(x) = -|1-x|$, it is then straightforward to check that

$$\pi_R(x_R, x_L; 0) = p(x_R, x_L; 0)[u_R(x_R) + M] + [1 - p(x_R, x_L; 0)]u_R(x_L),$$

is quadratic and also strictly concave over $x_R \in [0,1]$ for all $x_L \in [-1,0]$. Then the equilibrium policy of party R can be characterized by the first-order condition:

$$\frac{1}{2} - \rho\lambda(2x^* + M) = 0. \tag{A.12}$$

Applying the same arguments for party L, it is clear that for the case $x_0 = 0$, a Nash equilibrium exists and, moreover, this equilibrium is symmetric and uniquely defined, $x_R = -x_L = \lim_{s\to 0}\phi(s) = x^-$.

(ii) Next, we show that the same results apply for the case $|x_0| > 0$. To prove the existence, we apply the Debreu-Glicksberg-Fan Theorem (see Fudenberg and Tirole, 1991, Theorem 1.2). First, by Lemma A2, the sets of undominated strategies $(x_R, x_L) \in [0,1] \times [-1,0]$ are compact. Next, from (5), p is continuous in x_R, x_L for any $x_0 \in \Re$ because from (1), the utility of the median voter is continuous in x_L, x_R for any fixed reference point. Therefore, the payoffs $\pi_R(x_R, x_L; x_0), \pi_L(x_R, x_L; x_0)$ are also continuous in x_R, x_L. So, it remains to prove that $\pi_R(x_R, x_L; x_0), \pi_L(x_R, x_L; x_0)$ are quasi-concave in x_R, x_L, respectively. To do this, given the symmetry of the model, we only need to do this for party R. Note that $u_R(x) = x - 1$ is linear in x on $[0,1]$, and that p is continuously differentiable in x_R on $[0,1]$ from the twice differentiability of F. So, it is clear that $\pi_R(x_R, x_L; x_0)$ is twice continuously differentiable in $x_R \in [0,1]$ for all $x_L \in [-1,0]$ and $x_0 \in [-1,1]$. From (6), for any $x_R \in (0,1)$, $x_L \in [-1,0]$ and $x_0 \in [-1,1]$, we have

$$\frac{\partial \pi_R}{\partial x_R}(x_R, x_L; x_0) = p u_R'(x_R) + \frac{\partial p}{\partial x_R}\left[u_R(x_R) + M - u_R(x_L)\right]. \tag{A.13}$$

Using the fact that $u_R(x) = -|x-1| = x - 1$ when $x \le 1$, we get

$$\frac{\partial \pi_R}{\partial x_R}(x_R, x_L; x_0) = p + \frac{\partial p}{\partial x_R}\left(x_R + M - x_L\right). \tag{A.14}$$

Moreover, from (4), (A.2), it is easy to compute that

$$\frac{\partial p}{\partial x_R} = \rho\frac{\partial \mathbb{E}u(x_R; x_0)}{\partial x_R} = -\rho\left[(\lambda - 1)F\left(\frac{x_R}{s}\right) + 1\right] < 0. \tag{A.15}$$

It is also straightforward to calculate that

$$\frac{\partial^2 p}{\partial x_R^2} = -\frac{\rho(\lambda-1)}{s} f(\frac{x_R}{s}) < 0. \tag{A.16}$$

Recall that for Assumption A2 to hold, we need $\rho < 0.25$. Also, by Assumption A2, we have $M > \frac{1}{2\rho} - 2$. Hence, given A2, it is necessary that $M > 0$. So, from (A.14), (A.15) and (A.16):

$$\frac{\partial^2 \pi_R}{\partial x_R^2} = 2\underbrace{\frac{\partial p}{\partial x_R}}_{(-)} - \frac{\rho(\lambda-1)}{s} f(\frac{x_R}{s})\underbrace{(x_R + M - x_L)}_{(+)} < 0. \tag{A.17}$$

So, π_R is strictly concave and thus quasi-concave in $x_R \in [0,1]$ as required.

(iii) We need to show that the symmetric equilibrium is the unique pure strategy Nash equilibrium of the one-period game. Fix $x_0 \in [-1,1]$ and $x_0 \neq 0$. Suppose, by contradiction, that the one-period game admits an asymmetric equilibrium (x'_R, x'_L) such that $(x'_R, x'_L) \in [0,1] \times [-1,0]$ and $x'_R \neq -x'_L$. If $(-x'_L, x'_L)$ is an equilibrium, then it must be that $\pi_R(x'_R, x'_L; x_0) = \pi_R(-x'_L, x'_L; x_0)$. As proved above, under Assumption A1, $\pi_R(x_R, x_L; x_0)$ is strictly concave in $x_R \in [0,1]$ for all $x_L \in [-1,0]$. Thus, for all $\alpha \in (0,1)$, we have

$$\pi_R(\alpha x'_R + (1-\alpha)(-x'_L), x'_L; x_0) > \pi_R(x'_R, x'_L; x_0),$$

contradicting that (x'_R, x'_L) is an equilibrium. Therefore, $(-x'_L, x'_L)$ is not an equilibrium and $\pi_R(x'_R, x'_L; x_0) > \pi_R(-x'_L, x'_L; x_0)$. By the same token, $\pi_L(x'_R, x'_L; x_0) > \pi_L(x'_R, -x'_R; x_0)$. Adding up the two inequalities, it follows that

$$\pi_R(x'_R, x'_L; x_0) + \pi_L(x'_R, x'_L; x_0) > \pi_R(-x'_L, x'_L; x_0) + \pi_L(x'_R, -x'_R; x_0). \tag{A.18}$$

Let $p(x'_R, x'_L) = p'$. Using the fact that $u_R(x) = -(1-x)$ and $u_L(x) = -(1+x)$ for all $x \in [-1,1]$, we have

$$\begin{aligned}
\pi_R(x'_R, x'_L; x_0) + \pi_L(x'_R, x'_L; x_0) &= p'[u_R(x'_R) + u_L(x'_R)] + (1-p')[u_R(x'_L) + u_L(x'_L)] + M \\
&= p'[-(1-x'_R) - (1+x'_R)] + [1-p'][-(1-x'_L) - (1+x'_L)] + M \\
&= M - 2.
\end{aligned}$$

Note that at any $x_R = -x_L = x$, we have

$$p(x, -x; x_0) = \frac{1}{2} + \rho[\mathbb{E}u(x; x_0) - \mathbb{E}u(-x; x_0)] = \frac{1}{2},$$

where the last equality comes from the fact that $u(x) = u(-x) = -x$. Hence, $p(-x'_L, x'_L) = p(x'_R, -x'_R) = 0.5$. Then it is straightforward that

$$\begin{aligned}
\pi_R(-x'_L, x'_L; x_0) + \pi_L(x'_R, -x'_R; x_0) &= \frac{1}{2}[u_R(-x'_L) + u_R(x'_L)] + \frac{1}{2}[u_L(x'_R) + u_L(-x'_R)] + M \\
&= \frac{1}{2}[-(1+x'_L) - (1-x'_L)] + \frac{1}{2}[-(1+x'_R) - (1-x'_R)] + M \\
&= M - 2,
\end{aligned}$$

which stands in contradiction with (A.18). Hence, there does not exist any asymmetric equilibrium and the symmetric equilibrium is the unique equilibrium of the one-period game.

(iv) We characterize the unique and symmetric equilibrium for any $s > 0$. As noted above, at any symmetric equilibrium $x_R = -x_L$, we have $p(x_R, -x_R; x_0) = 0.5$. We then combine (A.14) and (A.15), and set $\frac{\partial \pi_R}{\partial x_R} = 0$ to obtain the first-order condition characterizing the symmetric equilibrium $x_R = -x_L = x^*$:

$$\frac{1}{2} - \rho \left[(\lambda - 1) F\left(\frac{x^*}{s}\right) + 1 \right] (2x^* + M) = 0. \tag{A.19}$$

which is (7) above. We need to prove that (A.19) has a unique solution in the domain $(0,1)$ for any fixed $s > 0$. First, write (7) more compactly as $g(x; s) = 0$. Then we can write

$$g(0; s) = \frac{1}{2} - \rho\left[(\lambda - 1)F(0) + 1\right] M = \frac{1}{2} - \rho M > \frac{1}{2} - \lambda \rho M > 0, \tag{A.20}$$

where the last inequality follows from Assumption A2. Similarly, by inspection,

$$g(1;s) = \frac{1}{2} - \rho\left[(\lambda-1)F\left(\frac{1}{s}\right)+1\right](2+M) < \frac{1}{2} - \rho\,(2+M) < 0, \tag{A.21}$$

where the last inequality also follows from Assumption A2. Finally, for all $x, s \in (0,1)$, note that

$$\frac{\partial g}{\partial x}(x;s) = -\rho\frac{\lambda-1}{s}f\left(\frac{x}{s}\right)(2x+M) - 2\rho\left[(\lambda-1)F\left(\frac{x}{s}\right)+1\right] < 0. \tag{A.22}$$

So, $g(x;s)$ is strictly decreasing in x. Combining (A.20), (A.21), (A.22), we see that $g(x;s)=0$ has a unique solution $x^* = \phi(s)$ strictly between 0 and 1.

(v) We now show that x^* is strictly increasing in s for $s > 0$. Total differentiation of $g(x^*;s) = 0$ gives:

$$\frac{\partial x^*}{\partial s} = -\frac{g_s(x^*;s)}{g_x(x^*;s)} \tag{A.23}$$

where g_x, g_s denote the partial derivatives of g with respect to x, s, respectively. Now, from (A.22), we have, for all $x, s \in (0,1)$, $g_x(x;s) < 0$. So, as $x^* \in (0,1)$, we also have $g_x(x^*;s) < 0$. Then we only need to show that $g_s(x^*;s) > 0$. At symmetric equilibrium, from (A.19):

$$g_s(x^*;s) = \frac{\partial^2 \pi_R}{\partial x_R \partial s}(x^*, x^*; s) = \rho(\lambda-1)\frac{x^*}{s^2}f\left(\frac{x^*}{s}\right)(2x^*+M) > 0,$$

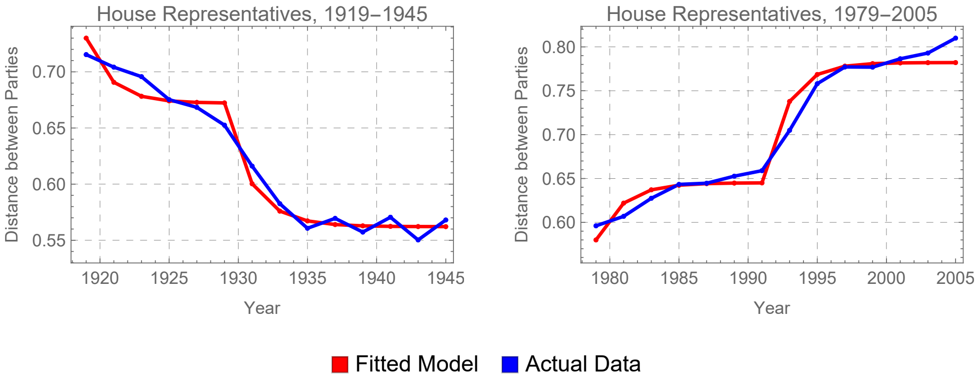
as required. As x^* increases in s, it is clear that $x^* \in [x^-, x^+)$. To find x^- and x^+, we take the limits as $s \to 0$ and $s \to \infty$ in (A.19) to get, respectively:

$$\frac{1}{2} - \rho\lambda\,(2x+M) = 0, \quad \frac{1}{2} - \rho\,(2x+M) = 0. \tag{A.24}$$

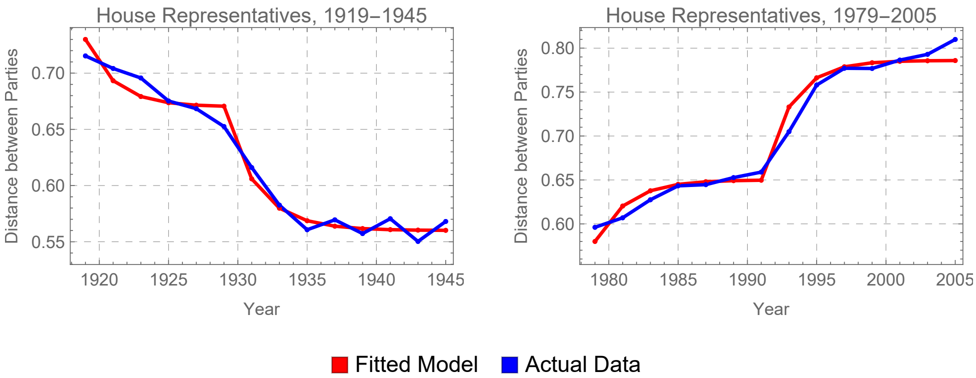
These solve to give x^-, x^+ respectively in Proposition 1. Finally, from A2, it is clear that $x^- > 0, x^+ < 1$. $\square$

B Appendix B

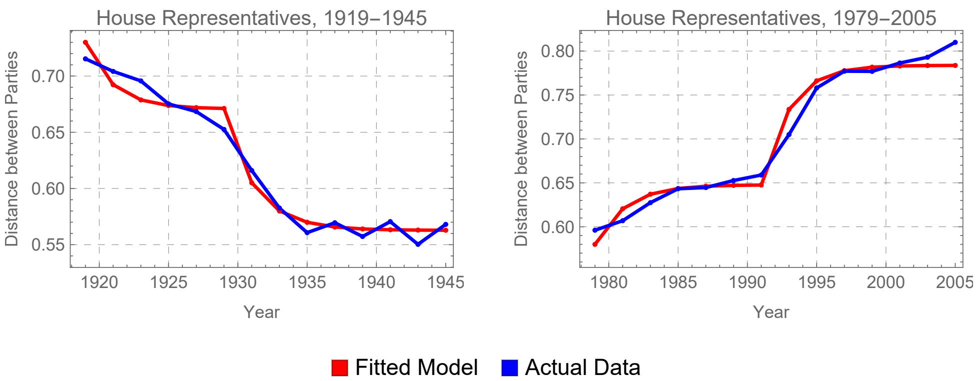
Figure B.1: Robustness Checks for the Equilibrium Path of the Fitted Model



(a) $\lambda = 1.9, \rho = 0.22$ and the recall shock follows an exponential distribution with $b = 0.35$. For the 1919-1945 figure, M changes from 0.873 to 0.983. For the 1979-2005 figure, M changes from 0.9 to 0.763. The MADs for the two figures are 0.0097 and 0.0111, respectively.



(b) $\lambda = 2.5, \rho = 0.2$ and the recall shock follows an exponential distribution with $b = 0.45$. For the 1919-1945 figure, M changes from 0.802 to 0.915. For the 1979-2005 figure, M changes from 0.825 to 0.687. The MADs for the two figures are 0.0088 and 0.0096, respectively.



(c) $\lambda = 2.3, \rho = 0.21$ and the recall shock follows an exponential distribution with mean $b = 0.7$. For the 1919-1945 figure, M changes from 0.838 to 0.946. For the 1979-2005 figure, M changes from 0.861 to 0.725. The MADs for the two figures are 0.0089 and 0.0099, respectively.

B.1 Extension to Gradual Changes in M

Here we study how the fit of the model to the data improves if we extend the model to allow for changes in M to occur gradually rather than instantaneously.

More specifically, instead of having a sudden shock of size Δ in the value of M, the dynamics of the office rent term at time t are modelled as:

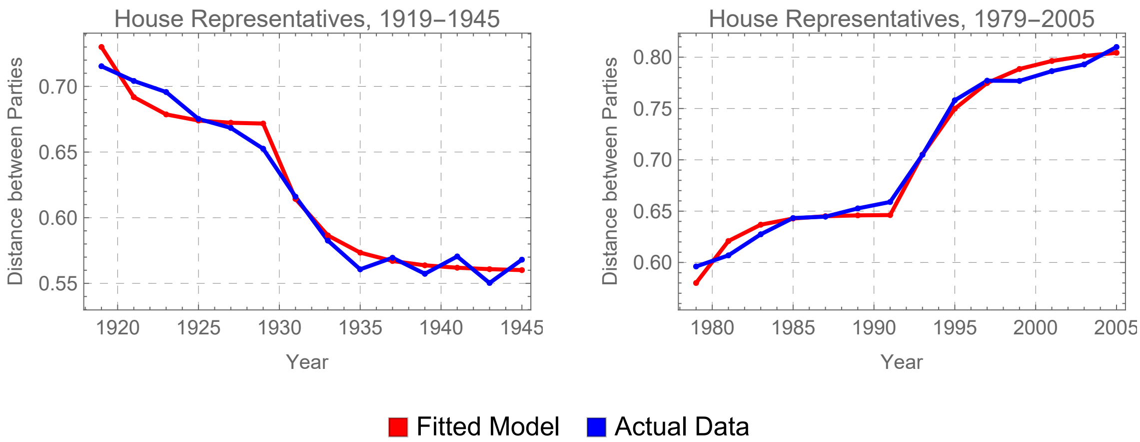
$$M = M_0 + \Delta e^{-r/[(t+1)-t_0]}, \tag{B.1}$$

where M_0 is the pre-shock value of M, t_0 is when the shock begins, and $r > 0$ captures how gradual the shock is. The larger the value of r, the slower the rate of change. The setup ensures that as time passes, $t \to \infty$, without any further changes in the values of M, the office rents will eventually converge to their new value, $M_0 + \Delta$. Then, for $t = t_0, t_0 + 1, \ldots$, the equilibrium platform x_t^* is the unique solution of the following equation:

$$\frac{1}{2} - \rho \left[(\lambda - 1) F \left(\frac{x}{|x_{t-1}^*|} \right) + 1 \right] \left(2x + M_0 + \Delta e^{-r/(t-t_0+1)} \right) = 0.$$

As discussed in Section 4, the fit of the model is now improved, particularly for the later period. In quantitative terms, the mean absolute deviations are now 0.009 and 0.008, or 1.2% and 1.3%, respectively.

Figure B.2: Parties' House Representatives with Gradual Shocks



Notes: In these figures, the office rent term evolves as specified in equation (B.1). For both figures, we have $\lambda = 2.1, \rho = 0.2$ and the recall shocks follow the exponential distribution with $b = 0.5$. In the first figure, the parameter values are $M_0 = 0.87$, $\Delta = 0.125$ and $r = 0.23$. In the second figure, the parameter values are $M_0 = 0.895$, $\Delta = 0.175$ and $r = 0.29$.