The Optimal Length of Political Terms*

Hans Gersbach
CER-ETH – Center of Economic Research at ETH Zurich and CEPR
Zuerichbergstrasse 18
8092 Zurich, Switzerland
hgersbach@ethz.ch

Matthew O. Jackson
Department of Economics
Stanford University
579 Serra Mall
94305-6072 California, USA
jacksonm@stanford.edu

Oriol Tejada
CER-ETH – Center of Economic Research at ETH Zurich
Zuerichbergstrasse 18
8092 Zurich, Switzerland
toriol@ethz.ch

This version: December 2019

Abstract

We analyze the optimal length of political terms (i.e., the inverse of the optimal frequency with which regular elections should be held) in an infinite-horizon model in which candidates of two polarized parties compete for office and the median voter shifts over time. In each period of a term, office-holders determine policy and experience persistent random shocks to their valence, which affect citizens heterogeneously. Policy changes are costly for citizens and politicians. The optimal term length then balances the frequency of costly policy changes when parties change office and with the incumbent's average valence during tenure. We find that the optimal term length increases with party polarization and the degree to which the median voter cares about valence, as well as with the frequency and the size of swings in the electorate. In contrast, the optimal term length decreases when candidates for office undergo less scrutiny or when parties care more about future outcomes. Finally, with small (large) swings in the electorate and small (large) checks and balances, the optimal term length also decreases if the extent of check and balances is further reduced (increased).

Keywords: elections; term length; polarization

JEL Classification: C72, C73, D72, D78

*We would like to thank Alessandra Casella, John Duggan, Daniel Houser, Antoine Loeper, César Martinelli, Andrea Mattozzi, Johanna Mollerstrom, Jan Zápal, as well as participants at the 2018 ETH Zurich Workshop on Political Economy, the 2019 SAET Conference held in Ischia and a seminar at George Mason University for comments that helped to improve the paper. All mistakes are our own.

As it is essential to liberty that the government in general should have a common interest with the people, so it is particularly essential that the branch of it under consideration should have an immediate dependence on, and an intimate sympathy with, the people. Frequent elections are unquestionably the only policy by which this dependence and sympathy can be effectually secured. But what particular degree of frequency may be absolutely necessary for the purpose, does not appear to be susceptible of any precise calculation, and must depend on a variety of circumstances with which it may be connected.

The Federalist Papers, No. 52.

1 Introduction

Motivation

It is widely accepted that periodic elections are a cornerstone of representative democracy. By casting a ballot, citizens not only delegate policy-making, but collectively they have the power to oust elected officials who are deemed to be of low quality or who will choose undesirable policies. Yet, this general principle—*direct rule by the people at certain time intervals*—does not stipulate *how frequently* the people should vote. To determine this frequency, many factors have to be taken into account, as acknowledged in the above quote. As it happens, be it for executive, parliamentary or judiciary office, term lengths—i.e., the inverse of the frequency with which elections are held—vary across, and even within, countries (see e.g. Dal Bó and Rossi, 2011). In the US, for example, term lengths at the federal level vary between two years (House of Representatives), four years (President), six years (Senate) and lifetime (Supreme Court), while in state legislatures we find two-year and four-year terms. In the world, the percentage of constitutions that provide for a four-year term length for the head of state has declined over the years, as opposed to an increase of the percentage of constitutions that provide for a five-year term length.[1] In Ancient Greece, members of a bouleuterion served one-year terms.[2]

Can we build a theory that allows us to address the frequency of elections in a systematic and meaningul way? If so, what are the determinants for setting this key institutional feature optimally?

[1] See `http://comparativeconstitutionsproject.org/files/cm_archives/term_limits.pdf?6c8912`, retrieved 23 August 2019.

[2] See `https://en.wikipedia.org/wiki/Bouleuterion`, retrieved 6 Sep 2019.

The model

This paper addresses the above questions by analyzing a model in which policy- and office-motivated candidates compete for office over infinitely many periods, which are grouped in terms.[3] A period is the minimal spell of time that allows the incumbent to change *one* policy that is in place. If they are elected, candidates pursue policies that are in line with one of two polarized parties, which have standard quadratic utilities over a one-dimensional policy space. As a first main feature, policy changes from one period to another are costly. The so-called *costs of change* increase with the extent of the policy shift and are borne by politicians and citizens. Unlike preserving the status quo, policy changes generate costs *per se*. The origins of such costs are manifold and can include physical investments to dismantle existing facilities, changes in budgets after the modification of the scale and scope of government programs and institutions, human capital to retain individuals or expenditures to smooth the transition. They may also consist of information, communication and lobbying costs. Notable examples of policies with significant costs of change are Obamacare, Brexit, or the attempt to achieve Catalan independence.

Costs of change create a dynamic link between policy choices that can be exploited by incumbents to gain an electoral advantage against challengers. This is not a universal property of costs of change; these costs must be moderately convex compared to the utility derived from policies. In the opposite direction, candidates—and their parties—randomly wear out or erode while holding power in a way that diminishes their *valence*. A politician's valence—and its dynamics—is the second main feature of our model, and it refers to their perceived judgement by the electorate beyond policy concerns, to their competence to execute governmental duties such as providing public goods efficiently, and/or to their character. A candidate of the opposition party always has the highest valence in his/her first period in office.[4] Upon election, however, there is a constant probability in every period that the office-holder will experience a negative and persistent shock to his/her valence, and thus the expected valence decreases with the number of periods in office. This shock is publicly observed and it may be valued by different citizens with differing relative intensities. A median voter who can influence the level of public goods to be provided by the incumbent, for instance, will suffer more in relative terms than the average voter from the

[3]Although our model is most appropriate for an executive office, it can also be applied to a parliament in which the power of agenda-setting is in the hands of one of two polarized groups, with the median member changing over time.

[4]Because we do not consider any information asymmetries, the assumption that a newly-elected candidate has the highest valence is made for simplicity. *Ceteris paribus*, parties have incentives to appoint a candidate with the highest valence.

incumbent's low ability to provide such goods. This is rooted in the conventional wisdom that the relative position between the median voter and the average voter matters for the efficient provision of public goods (see e.g. Lizzeri and Persico, 2001, as well as the references therein).[5]

Finally, as a third main feature, the median voter's preferred policy (or *peak*) shifts over time, prompting on occasion a power shift from one party to the other to adjust policy to the change in preferences (of the median voter). The fluctuation of the median voter captures not only any *noise* in the electoral process that can affect the election outcome and cannot be anticipated by parties, but also any underlying trends in voter preferences.[6] In contrast, by way of normalization, party positions are fixed. This is made for convenience and reflects the situation where parties are simply unable—e.g. because they lack the technology to anticipate electoral trends—or unwilling—e.g. because they are controlled by a radicalized base—to keep up with the changes in the electorate. While the (exogenous) stochastic process determining the valence shocks provides a downward drift that hurts the incumbent's electoral prospects unambiguously, the stochastic process determining the median voter's peak generates variance around the downward drift, which can either benefit the incumbent or hurt him/her.[7]

The optimal length of political terms

The model—which further builds on the (technical) assumptions that politicians and voters use Markov strategies, are present-biased and cannot commit to particular policies before elections—generates the following insights: If there were no costs of change, then from the perspective of a voter whose peak were at the same distance from the two parties' peaks in all periods, it would be optimal for terms to be as short as possible. The reason is that parties would choose policies coinciding with their own peaks, which would be symmetrically located with respect to the equidistant voter's peak, who would then be indifferent between the two.[8] Hence, shortening a term would only yield utility gains since office-holders could be ousted as soon as their valence is lower than the challenger's. This property holds no matter the intensity according to which the equidistant voter values the politicians' valence.

[5] We provide a micro-foundation for this assumption in Section 7.8.

[6] The evolution of the median voter position has been documented empirically. For instance, we refer to Kim and Fording (2003) for an empirical account of the evolution of the median voter's position for several Western democracies in the post-war era, based on party manifesto data.

[7] From a technical perspective, we could simply have one random process that is general enough to add uncertainty to the incumbent's electoral prospects in any possible way. However, splitting the random process into two separate processes—one dealing with valence, a second one dealing with social preferences—enables us to obtain interesting insights regarding policies, elections and the optimal determination of the length of a political term.

[8] Our results are not knife-edged on the assumption that parties' peaks are located symmetrically with respect to the voter who defines welfare. This is discussed in Section 6.3.

The picture changes substantially with costs of change, which introduce the interesting trade-offs in our analysis. For the sake of argument, assume that the median voter's peak is initially located at the same distance from the two parties' peaks. In the absence of valence shocks and changes in the electorate's preferences, the incumbent would always be re-elected. The reason is that changing the party that is in office would result in a costly change of policies, but these costs would not be incurred if the incumbent were re-elected. Since both the policy chosen by the incumbent and the policy chosen by the challenger would be at the same distance from the median voter's peak, albeit located on different sides of the policy spectrum, the median voter's utility derived from policies would not be affected by the power shift. Costs of change would then be the decisive factor for determining the election outcomes. Absent any valence shocks and changes in the electorate's preferences, it would thus be optimal from the perspective of a voter with peak that is permanently equidistant from the parties' peaks for terms to be as long as possible (i.e., infinity). This way of measuring the social desirability of a term length by computing a discounted sum of such a voter's stage utilities will be called *ex-ante welfare*. It captures the idea that aggregate preferences over fundamentals—say, over policies—are stable. This is in line with abundant empirical research (see e.g. Green and Palmquist, 1994; Sears and Funk, 1999; Jennings et al., 2009).

What happens in the presence of costs of change when valence shocks and changes in the median voter's peak, and thus swings in the electorate, are introduced? Clearly, from an overall ex-ante welfare perspective, terms should not be arbitrarily long since the disutility generated by an incumbent who has experienced many valence shocks and has a very low valence will offset any other source of utility for the median voter, including any (moderate) costs of change. At the same time, having excessively short terms in the presence of valence shocks and (enough) variance in the median voter's peak can be ex-ante inefficient.[9] This is because it allows a more frequent change of parties and thus a costly change of policies. While such costs affect all voters equally, the median voter (in any particular period) internalizes neither the *right* weight that the incumbent's valence has on welfare nor the policy preferences of the entire electorate. On occasion, this leads the median voter to prompt a power shift between parties even if doing so is not socially optimal. In our baseline model, this is possible thanks to exogenous changes on valence (of the incumbent) and preferences (of the median voter). Yet, we also discuss how these changes can be endogenized. In other words, although having the possibility to oust the

[9]The minimal degree of variance in the median voter's peak that is necessary to offset the incumbency advantage generated by costs of change is determined by the marginal cost of change itself: higher marginal costs of change generate a larger advantage for incumbents, in which case larger shocks to the median voter are needed for the latter to want to oust the incumbent.

incumbent in elections would be good for the current median voter, over time, the repeated changes in policy and associated costs can be detrimental to the population's overall welfare. In fact, we show that the optimal term length is longer than one period for generic parameter values. In such cases, it would be optimal for the electorate to tie its hands. This property does not originate from our assumption that citizens are present-biased, as we show in Section 6.3. While our extreme case of quasi-hyperbolic discounting *does* generate inefficiencies in policy decisions and elections compared to the case of standard forward-looking discounting, these inefficiencies cannot be corrected by changing the term length.[10] This is because costs of change are incurred in the period where the policy shift occurs and the incumbent's valence has a downward drift.[11]

Varying term length and empirical implications

In general, the *exact* optimal length of the political term is thus determined by the trade-off between the costs of changing the status-quo policy whenever a new candidate is elected—be it due to valence shocks or to preference changes—and the risk that office-holders will have a lower average valence caused by persistent shocks. This trade-off depends on the primitives of the model, and our analysis reveals clear-cut comparative statics results. In particular, we find that in (the unique) equilibrium, the optimal term length increases with party polarization and with the degree to which the median voter cares about the incumbent's valence compared to the average voter. It also increases with the frequency and size of the shocks affecting the median voter's peak, as well as with the degree to which the preferences of the electorate are polarized. In contrast, the optimal term length decreases with the extent of the valence shock to incumbents and with the probability that such shocks occur, as well as with the parties' (quasi-hyperbolic) discount factor. The dependence of the optimal term length on the marginal cost of change is ambiguous. When fluctuations of the median voter's peak and the marginal cost of change are small, the optimal term length decreases if the marginal cost of change further decreases. In contrast, when fluctuations of the median vote's peak and the marginal cost of change are large, the optimal term length decreases if the marginal cost of change further increases.

Our comparative statics exercises yield a variety of suggestions when political terms should be longer or shorter, which extends the scope for empirical. Specifically, our model rationalizes

[10] Gersbach et al. (2019a) show that with fully forward-looking voters, the dynamics are similar to the model we consider from a qualitative perspective, and so is the welfare assessment. We refer to Hwang and Möllerström (2017) for an analysis of one-directional reforms in the presence of (heterogenous) present-biased voters. There is a vast literature in political economy literature on time-inconsistent preferences (see e.g. Bisin et al., 2015; Jackson and Yariv, 2015; Piguillem and Riboni, 2015; Lizzeri and Yariv, 2017).

[11] In Section 7.1 we show how *ceteris paribus*, terms should be longer if incumbents need some time to learn by doing—which increases their valence—and the median voter becomes aware that the incumbent has learnt with some delay.

longer terms when there is high polarization and/or social instability. For instance, in turbulent times with large swings in the electorate, longer terms could avoid excessive costs linked to policy changes. Longer terms are also desirable when politicians undergo in-depth scrutiny before they are appointed, or even nominated to run, for office. Typical examples are courts, and in particular the US Supreme Court. The opposite argument holds when the selection of office-holders includes politicians without an available record in office or outsiders who have not been observed long within their party organization. In such cases, the higher risk of negative valence shocks calls for shorter terms. Finally, we provide a rationale for a connection between the term length and checks and balances (as captured in a reduced form by the extent of the marginal cost of change). Increasing low (high) levels of checks and balances should go along with longer (shorter) terms, provided that social preferences are relatively stable (unstable).[12]

Further insights

The baseline model in which our results are derived is the simplest one in which costs of change play a crucial role in policy decisions and voting, and median voter decisions may not be socially optimal. An important feature of the model that ensures mathematical tractability is the planning horizon of politicians and citizens, which we assume to encompass two periods. With one-period terms, all policy decisions are directly influenced by electoral incentives. For longer terms, the limited farsightedness of politicians and citizens suffices to ensure that the median voter's behavior determines policy decisions and vice versa.

Apart from the ex-ante welfare measure, welfare can also be defined by weighing the utilities derived from policies across periods depending on the evolution of the median voter's peak, which will typically not be equidistant from the two parties' peak. We call *interim welfare* this way of measuring how socially desirable a term length is. As it turns out, this second welfare notion coincides with the ex-ante welfare measure, provided that either the variance in the median voter's fluctuation is low or the initial peak of the median voter is not very biased towards one of the parties. In the cases where the two notions do *not* coincide, the second welfare notion calls for shorter terms, but not necessarily for one-period terms.

Numerous extensions and ramifications of our analysis can be pursued, and we examine some of them. These include the possibility that incumbents' valence increases over time—say, through learning-by-doing—, that third parties' or candidates' entry to the elections whose preferences

[12] See Beck et al. (2001) for an index of checks and balances that could allow to test our theory empirically. A measure of political polarization can be found in `https://cses.org/data-download/download-data-documentation/party-system-polarization-index-for-cses-modules-1-4/`, retrieved 23 August 2019.

coincide with the current median voter's peak, that incumbents have more than one policy at their disposal, or that the incumbent has the option to call early elections. We also investigate the effect on the optimal term length of campaign spending, politician accountability and voter pandering, as well as the effect of having some further parameters of the model—such as the marginal cost of change—varying across periods. Finally, we provide a micro-foundation for the assumption that the median voter can suffer more in relative terms than the average voter from an incumbent's low valence, and we discuss how to endogenize the random process that govern valence and preference shocks. Overall, the main thrust of our results extends to more general settings.

Organization of the paper

The paper is organized as follows. In Section 2 we review the research most closely related to our paper. In Section 3 we present our model of political competition and set up the notation. In Section 4 we study the benchmark case where policy changes are costless. In Section 5 we analyze the general case, in which policy changes are costly. In Section 6 we discuss optimal design of political term length. Some extensions of our baseline set-up are discussed in Section 7. Section 8 concludes. Appendices A and B contain the proofs.

2 Relation to the literature

Our paper is related to various strands of literature.

Optimal term length

The literature investigating term lengths from a normative perspective is sparse and lacks a workhorse model. An exception is Schultz (2008), who pointed out that shorter terms favor accountability of office-holders and screening out of bad politicians. Under asymmetric information between the office-holder and the electorate, in contrast, longer terms reduce the office-holders' incentives to pander and to distort policies.

Our focus on costly policy changes and shocks to the office-holders' valence and the electorate's preferences enables us to characterize optimal term length from a welfare perspective explicitly, and then to obtain an array of interesting comparative statics results. As a consequence, our analysis provides potential rationales for the observed difference in the duration of political terms: differences in features of the political system, such as the level of party polarization or the level of costs associated with policy changes, call for different term lengths.

Costly policy changes

Costs of change are one key ingredient of our model: if policy changes were costless, parties would choose their bliss points in any period in which they hold office, resulting in trivial dynamics. There are a few papers that contain models that are related to ours. Focusing on policies that are enacted for the first time but that can be subsequently amended, Glazer et al. (1998) show that large costs of change generate an incumbency advantage if the challenger is committed to reversing an extreme policy chosen by the incumbent. Also focusing on costly reforms, Gersbach and Tejada (2018) show that incumbents who are more efficient than the challenger in implementing reforms will also choose extreme policies to obtain an electoral advantage, while Gersbach et al. (2019b) show that more political instability will moderate policy choices. Finally, Gersbach et al. (2019a) analyze more general specifications of our model, including convex costs of change and fully forward-looking agents. They show that for the purpose of investigating optimal terms when policy changes are costly, our model specification including linear costs of change does not lead to knife-edge results.[13,14] From the perspective of this strand of literature, the main contribution of our paper is to explain how optimal term lenght is related to costs of change, among many other factors that are usually assumed to influence policy and elections.

Calling early elections and term limits

The optimal design of term lengths is distinct from other aspects of elections that have been analyzed in the literature, such as the possibility to call early elections (see e.g. Lesmono et al., 2003; Kayser, 2005; Keppo et al., 2008) or the effect of term limits (see e.g. Acemoglu et al., 2013; Smart and Sturm, 2013; Duggan, 2017). Yet, our analysis can provide some insights to both types of problems. On the one hand, in some political systems, incumbents have some flexibility as to when to call elections. This flexibility is another source for the incumbency advantage and it endogenizes the term length *de facto*, up to the binding length required by law. Changes in term length set by law can thus affect the endogenous term length decided by incumbents. On the other hand, term limits force the incumbent to step down, thereby allowing a new politician to hold office, who might change policy. This bears some resemblance to the effect of shorter terms, which also enable more frequent (costly) policy changes.

[13]It suffices to assume that the degree of convexity of the costs of change is lower than the degree of utility losses of a citizen accrued when policies differ from his/her peak. This ensures that *ceteris paribus*, the incumbent enjoys an advantage over the challenger when s/he chooses appropriate policies.

[14]Other models of dynamic policies address different issues (see e.g. Forand, 2014; Bowen et al., 2014; Nunnari and Zápal, 2017; Bowen et al., 2017; Chen and Eraslan, 2017).

3 Model

We consider an infinite-horizon model of a two-party electoral competition ($t = 1, 2, \ldots$), in which policy changes are costly for both parties and citizens. In the following, we describe the elements of the model.

3.1 Terms, policies, elections, parties, and voters

Our focus is the optimal length of a single term in office, i.e., the optimal length of time between regular elections. To accommodate varying lengths, we assume that an office-holder's term comprises several periods. We use T to denote the length of a term, with $T \in \{1, 2, \ldots\}$. That is, elections take place every T periods, and thus an incumbent stays in power without facing re-election during $T-1$ consecutive periods. Periods can be months, semesters or years, and they capture the time span that is necessary for an office-holder to make one policy choice.

In each election, citizens cast a vote in favor of one of two candidates, one from each party—see below. Upon election, the winning candidate chooses a policy $i_t \in \mathbb{R}$ at the start of any period t in which s/he holds office. We interpret this policy as the usual left/right ideological dimension. Examples of such policies could be the extent of mandatory insurance coverage within the health care system or the level of taxation and redistribution. We proceed on the assumption that candidates comply with party objectives and cannot commit to any particular policy before election. Parties (and their candidates) and citizens have standard quadratic preferences on $\mathbb{R}$, represented by their *peak* or preferred policy.

Each of the two political parties, denoted by R and L, comprises a large pool of candidates, which are *ex ante* identical with regard to valence and political preferences. We denote candidates—and hence policy-makers—by k, with parties being denoted by K ($k, K \in \{L, R\}$). We assume that any candidate of party R has peak $\mu \in \mathbb{R}$ and any candidate of party L has peak $-\mu \in \mathbb{R}$. Parameter $\mu > 0$ captures the degree of party polarization. The (implicit) assumption that parties' peaks are symmetrically located with respect to zero can be made without loss of generality as far as the description of equilibrium behavior is concerned. As for its role on optimal term length, we show in Section 6.3 that there are two cases depending on the asymmetry between both parties' peak relative to zero. If the asymmetry is low, our analysis remains intact. If the asymmetry is large, then terms should be shorter all else being equal. We also proceed for simplicity with the assumption that in each period t there is a (representative) voter, who is decisive in the elections (if they take place) and whose peak varies over time.

Formally, the median voter's peak in period t is denoted by m_t and is determined according to some cumulative probability distribution function $F(\cdot|m_{t-1})$. The median voter's peak becomes common knowledge immediately after it is realized.

Besides policy choices, voters care about the politicians' *valence*, which includes features such as character, honesty, or ability to deliver public goods. These characteristics may vary over time due to one or more shocks, and may be valued differently across the citizenry. Citizen heterogeneity is immaterial for equilibrium behavior, but it matters for optimal term length—see Section 6. For instance, office-holders may start enjoying power too much and focusing less on providing public goods, or they may endure political fights and wars of attrition that lower their ability to undertake policy reforms. Finally, a politician may be caught in some scandal, which permanently erodes their valence. Once it has occurred, the politician does not recover from the shock. Using a_t to denote incumbent k's valence at the end of period t, we specifically assume that in each period t in office, there is a probability $\rho \in (0, 1)$ that s/he will be affected by a permanent negative shock to his/her valence.[15] This shock results in $a_t = a_{t-1} - A$, with $A > 0$. We further assume that all candidates have the same valence when they assume office for the first time.[16] This valence is normalized to zero. Incumbents stay in office until they are defeated by a challenger.

3.2 Voter and party (stage) preferences

In each period t, a voter with peak μ' derives utility from the policy choice $i_t \in \mathbb{R}$ equal to

$$U_{\mu'}(i_{t-1}, i_t, a_t) := -\left(i_t - \mu'\right)^2 + a_t - c \cdot |i_{t-1} - i_t| \,. \tag{1}$$

The above expression consists of three terms. The term $-(i_t - \mu')^2$ is the cost of having a policy that differs from the voter's peak μ'. The term a_t is the current value of the office-holder's valence. As already mentioned, this value is zero at the beginning of an office-holder's tenure, but decreases every time s/he suffers a valence shock. This happens—if at all—after policy has been chosen, and it affects the current period as well as any subsequent one. The remaining term, $-c \cdot |i_{t-1} - i_t|$, captures the costs of changing policy from one period to the other. The so-called *costs of change* have been discussed in the Introduction in detail.

Second, a politician k is a citizen, and thus his/her utility is similar to the one expressed in (1). The only difference is that a politician also derives private benefits b from each period in which

[15] Since the incumbent is already known at the end of period t, we drop the dependence of a_t on k.

[16] Parties select those candidates for office who have the highest valence. If candidates with lower valence had a chance to enter office, the optimal term length would be shorter than the ones we calculate.

s/he holds office. These benefits are very large and they include wages as well as non-monetary sources of utility such as psychological rewards associated with social status and power and enhanced career opportunities. The benefits may include a share that is enjoyed privately by the politician and another part that is also enjoyed by the party itself. Accordingly, a politician k with peak μ_k derives utility from the policy choice $i_t \in \mathbb{R}$ equal to

$$V_k(i_{t-1}, i_t, a_t) := U_{\mu_k}(i_{t-1}, i_t, a_t) + b \cdot \mathbb{1}_t(k) = -(i_t - \mu_k)^2 + a_t - c \cdot |i_{t-1} - i_t| + b \cdot \mathbb{1}_t(k),$$

where $\mathbb{1}_t(k) = 1$ if k holds office in period t and $\mathbb{1}_t(k) = 0$ otherwise.[17] Because we proceed on the assumption that candidates execute party orders, for each candidate k of party K, we will henceforth write $V_K(\cdot,\cdot,\cdot) = V_k(\cdot,\cdot,\cdot)$, with V_K denoting the party K's utility. In our citizen-candidate set-up, we can think of μ_k as the peak median party member.

3.3 Timeline

The timeline of the political game, which we denote by $\mathcal{G}$, is as follows:

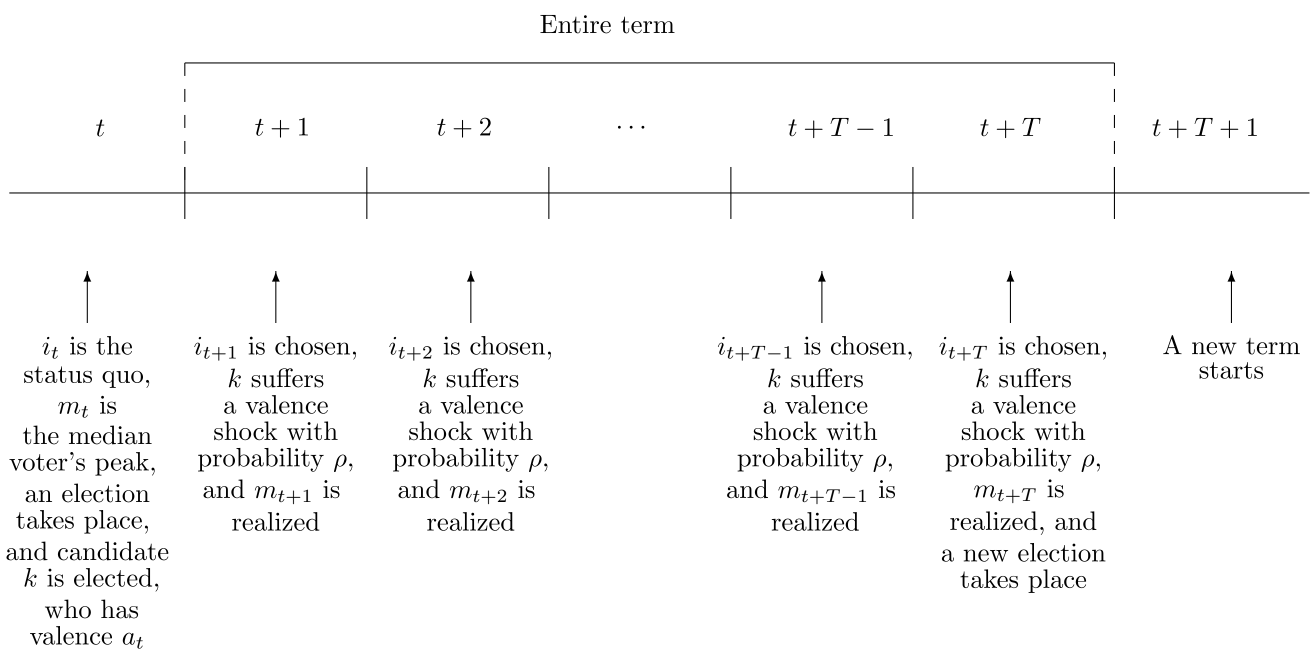

Figure 1: Timeline of the political game, with elections taking place in periods t and $t+T$.

Note that i_0, the initial policy before period $t = 1$, is a parameter of the model and is thus exogenously given.

[17] Assuming that the cost parameter is the same for parties and voters is consistent with a citizen-candidate framework. Moreover, it is not knife-edged since our results carry over to more general cases.

3.4 Equilibrium notion

We look for sequential equilibria with the following refinements. First, we only consider stationary Markov strategies, and hence voters can only condition their strategies on state variables such as the identity and valence of the incumbent, as well as on the status quo policy and the median voter's peak. In turn, incumbents can only condition their strategy on their own valence and the status quo policy, as well as on the current median voter's peak and the number of periods of the current term that have already passed.[18] We recall that candidates cannot commit to policies prior to elections, and hence campaigns are irrelevant. This implies that opposition parties take no payoff-relevant action, and are thus not modelled explicitly. Second, citizens vote for the candidate from whom they expect higher (lifetime) utility. This is a standard refinement in the literature, which rules out implausible voting equilibria. As a (non-essential) tie-breaking rule, citizens prefer the incumbent in case of indifference. Since office-holders simply comply with the policy objective of their party, we do not explicitly model candidates. Hence, besides voters, we consider the two parties, namely L and R, to be the two other players of the political game $\mathcal{G}$. Third and last, we assume that voters and parties care about the current period and one period ahead, which they discount with a common factor θ, with $0 < \theta \leq 1$.[19]

Overall, an *equilibrium* is a triple made up of a voting strategy for the median voter(s) and policy decisions by both parties when in power, denoted by $(\sigma_m, \sigma_L, \sigma_R)$, such that for each $t \in \{1, 2, \ldots\}$ and each status-quo policy $i_{t-1} \in \mathbb{R}$, the following two conditions hold:

(i) for all $t \in \mathbb{N}$,

$$\begin{aligned}\sigma_K(t/T, i_{t-1}, a_{t-1}, m_{t-1}) \in \underset{i_t \in \mathbb{R}}{\operatorname{argmax}}\, \mathbb{E}\,[&V_K(i_{t-1}, i_t, a_t) \\ &+ \theta \cdot V_K(i_t, \sigma_{K'}((t+1)/T, i_t, a_t, m_t), a_{t+1})]\,,\end{aligned}$$

(ii) if $t/T = 0$,

$$\begin{aligned}\sigma_m(K, i_{t-1}, a_t, m_t) = K \Leftrightarrow &\mathbb{E}\left[U_{m_t}\left(i_{t-1}, \sigma_K((t+1)/T, i_{t-1}, a_t, m_t), a_{t+1}\right)\right] \\ &\geq \mathbb{E}\left[U_{m_t}\left(i_{t-1}, \sigma_{-K}((t+1)/T, i_{t-1}, 0, m_t), a_{t+1}\right)\right],\end{aligned}$$

where K denotes the incumbent's party in period t, $-K$ denotes the challenger's party,

[18] Since we will assume that players will only look ahead for one period beyond the current period, the possibilities for the existence of non-Markovian equilibria are quite limited.

[19] That is, we build on the assumption of quasi-hyperbolic discounting, where periods that lie two period ahead of the current one are fully discounted.

and K' denotes the party that is expected to be in power in period $t+1$. We use t/T to denote t modulus T.[20]

The expected values are taken with respect to the current median voter's peak and the identity and current valence of the incumbent, and we assume that they are all well defined. As already mentioned, we allow parties to (potentially) choose different policies in different periods of an office-holder's term and to condition their choices on their own valence as well as on the median voter's peak. More critically, we do not assume that parties and voters are farsighted: they only envision one period ahead. One reason to proceed with this assumption is mathematical tractability.[21] Even so, this assumption suffices to ensure that regardless of the length of a term, policy choices of consecutive periods are linked, and that the median voter's decision on election day influences party decisions and vice versa. Indeed, the median voter's decision is affected by the policies that will be carried out in the first period of the subsequent term by either candidate if they are elected. In turn, these policy decisions depend on the policies that the corresponding office-holder expects to carry out in the second period of the term, and so on. Finally, the last policy decision of an office-holder in a given term takes the median voter's decision into account, thereby establishing dynamic links in the decisions across all periods.

3.5 Technical assumptions on the parameters

We make a number of (technical) assumptions that facilitate exposition but do not affect our results. First, we initiate the model such that in period $t = 1$ a candidate from party R is in office, who has not suffered any valence shock yet.[22] Since we are assuming that m_0 is drawn from some probability distribution, we allow for exogenous changes in the electorate preferences. Second, we assume that

$$0 \leq c < \frac{2\mu}{1+\theta}, \tag{2}$$

and hence costs of change, as expressed by the marginal cost parameter c, are moderate in relation to party polarization. This assumption guarantees that policy choices of right-wing candidates are to the right of zero and policy choices of left-wing candidates are to the left

[20] While the standard notation would be t mod T, here we use the shortcut t/T to simplify notation.

[21] Assuming that agents are present-biased is mainly technical in nature, and it enables a tractable analysis of the model. Nevertheless, this is not a critical assumption if we focus on Markov dynamics (see Gersbach et al., 2019a). As to the electorate, for instance, assuming a two-period horizon guarantees that voters cannot use strategies to reward certain good behavior of incumbents by punishing bad behavior over some time horizon. Allowing such strategies would not destroy our equilibrium, but may add other equilibria. As to the parties, an infinite horizon might add more equilibria, too.

[22] This is equivalent to assuming that the probability that each party is in office is 1/2.

of zero. Imposing (2) yields the interesting cases, since large values of c induce no changes in policies, resulting in very simple dynamics. Third, we assume that

$$-\mu + \frac{c}{2} \cdot (1+\theta) < i_0 < \mu - \frac{c}{2} \cdot (1+\theta). \tag{3}$$

This ensures that initial policy polarization is lower than party polarization, and it is consistent with patterns observed in actual elections (see e.g. Wiesehomeier and Benoît, 2009).[23] Fourth and last, we assume that office benefits b are very large, so that incumbents want to be re-elected regardless of any other consideration.

4 The Case without Costs of Change

Since the purpose of the paper is to discuss the optimal length of a political term in a framework where policy changes are associated with costs, it is convenient to start with the case where policy changes are costless, as a benchmark set-up. Without costs of change, the following result describes (on-path) equilibrium behavior:

Theorem 1

Assume $c = 0$. Then, in the unique equilibrium of $\mathcal{G}$,

(i) office-holders choose their peak in any period,

(ii) the incumbent $k \in K$ is re-elected in period t if and only if

$$m_t \geq \frac{A \cdot z_t}{4\mu}$$

if $K = R$, and if and only if

$$m_t \leq -\frac{A \cdot z_t}{4\mu}$$

if $K = L$, where z_t is the number of valence shocks experienced by the incumbent up to period t.

Hence, in the absence of costs associated with policy changes, the dynamics of our model is simple: policies are polarized since the parties' peaks are chosen. In turn, office-holders are re-elected if and only if at the time of elections, the median voter is sufficiently biased towards

[23] The dynamics with any initial policy and any level of marginal costs of change are examined in Gersbach et al. (2019a). Large marginal costs of change yield the property that the status quo is preserved throughout, while very polarized initial policies yield convergence in the long run to the alternation between policies featured in the present paper, which is a *robust* state.

the incumbent, so that s/he prefers to trade off ideology with lower valence. Not being able to replace a bad politician as soon as possible—because elections are seldom called due to very long terms—is then not desirable for the median voter with peak permanently at zero from an ex-ante perspective. That is, the optimal length is $T = 1$ in the absence of costs of change from the perspective of a citizen with peak permanently at zero.

5 The General Case with Costs of Change

We now address the general case, in which a marginal policy change is costly for both politicians and voters. We start by analyzing equilibrium behavior of our political game, deferring the examination of optimal term length to the next section. It will be convenient to introduce the following notation:

$$\Delta := \mu - \frac{c}{2} \cdot (1 + \theta) . \tag{4}$$

Note that due to our assumptions on the parameters, we have $0 < \Delta < \mu$.

Theorem 2

Assume $c > 0$. Then, in the unique equilibrium of $\mathcal{G}$,

(i) any office-holder $k \in R$ chooses Δ and any office-holder $k \in L$ chooses $-\Delta$ in any period in which they are in power,

(ii) the incumbent $k \in K$ is re-elected in period t if and only if

$$m_t \geq -\frac{c}{2} + \frac{A \cdot z_t}{4\Delta} \tag{5}$$

if $K = R$, and if and only if

$$m_t \leq \frac{c}{2} - \frac{A \cdot z_t}{4\Delta} \tag{6}$$

if $K = L$, where z_t is the number of valence shocks experienced by the incumbent up to period t.

Theorem 2 reveals that the dynamics of our model change substantially when policy changes are costly. As an illustration, assume without loss of generality that $k \in R$ is the incumbent. Let us start with the median voter's decision in the election period. As in the case without costs of change, s/he trades off lower valence of the office-holder with his/her own ideology. With costs of change, this is captured by the term $\frac{A \cdot z_t}{4\Delta}$ in Equation (5). Because Δ is decreasing in c, increasing c requires trading off *more* ideology for the *same* lack of valence of the office-holder.

When $c > 0$, however, there is another term in Equation (5), namely $-\frac{c}{2}$. This second term captures another mechanism affecting the previous trade-off: increasing c requires trading off *less* ideology for the *same* lack of valence of the office-holder. This is the manifestation of the incumbency advantage generated by costs of change. It comes about for three reasons. First, costs of change allow the incumbent to commit *not* to move the policy further beyond Δ (for party R) or $-\Delta$ (for party L) towards his/her bliss point in case s/he is re-elected.[24] Second, costs of change do *not* allow the challenger to commit to a substantially more moderate policy than the incumbent: the challenger can only commit to a policy that is as extreme as the one the incumbent chooses, but on the other side of the political spectrum, namely, $-\Delta$ (for party L) or Δ (for party R).[25] Third, to implement this latter policy, the challenger needs to incur a costly policy shift. Then, the net effect of an increase of parameter c on the trade-off between ideology—and hence policy—and valence depends on the parameters of the model, and in particular on the number of valence shocks already experienced by the incumbent.

As for the parties' decision, say the decision of party R, the effect of costs of change on policy decisions works through different channels across periods. On the one hand, consider the decision in the period where elections take place. As we have see, costs of change *do* generate an electoral advantage, so the incumbent's re-election concerns go hand in hand with indulging in their ideological preferences right *before* elections take place. For his/her calculus, the incumbent also has to take into account the policy that will be chosen, and hence the costs of change that might accrue, *after* elections, be it by him/her or by the challenger. This pushes incumbents to choose a more moderate policy than their own bliss point. All considerations together yield the best response Δ. Remarkably, the optimality of this policy choice is independent of the number of valence shocks experienced by the incumbent and the evolution of the median voter's peak. The dynamics determining the median voter's peak and the incumbent's valence only affect re-election.

On the other hand, consider any other period in which the incumbent must make a policy choice. Immediate re-election concerns are absent, but they matter indirectly through the policy choice in the subsequent period. In this case, the office-holder strikes a balance between ideological objectives and the costs that will accrue if the status-quo policy—which might have been chosen

[24]Gersbach et al. (2019a) show that for this first property to hold, it is necessary that costs of change are less convex than the disutility obtained from policies.

[25]The fact that the policies chosen by parties are symmetric with respect to zero follow from the symmetry of their peaks and the fact that costs of change are linear. Nevertheless, these assumptions do not lead to knife-edged results. Our results still hold even if costs are convex (see Gersbach et al., 2019a) or parties' peaks are not symmetrically located with respect to zero (see Section 6.3).

by the incumbent or by an office-holder of the other party—changes. This can be illustrated in the simplest way for $\theta = 0$, i.e. when politicians are fully myopic. In this case, office-holders trade off a quadratic cost of having a policy away from their peak with a linear cost of policy change. If the policy is far away from their peak, the ideological disutility outweighs the adjustment cost, and office-holders will change the status quo towards their peak. However, once the policy is not farther away than $c/2$ from his/her peak, the cost of adjustment outweighs the (myopic) gains from moving the policy, and so the incumbent leaves the policy as it is. Increasing θ from zero to a positive value increases the distance between the policy chosen by the incumbent and his/her peak, since s/he internalizes the costs of change that will accrue in the future if the other party dictates policy again.

6 The Optimal Length of a Political Term

To evaluate the optimal term length, we need a measure of social welfare. To that end, we denote by $\delta \in (0, 1)$ the *social discount factor* used to aggregate utilities of citizens across time for welfare evaluations. One distinct possibility is to equate δ to the quasi-hyperbolic discount factor θ, but our analysis holds for any value of the social discount factor—we discuss the particular case where $\delta = \theta$ below. Because there are two exogenous stochastic elements in our model, namely the incumbent's valence and the median voter's peak, we have to introduce representations of these stochastic processes to obtain an explicit expression for welfare. In particular, we use $\mathcal{S} = (s_1, s_2, \ldots)$ to denote an arbitrary realization of the stochastic process describing whether or not a shock occurred to the incumbent's valence across all periods.[26] Similarly, we use $\mathcal{M} = (m_0, m_1, \ldots)$ to denote an arbitrary realization of the stochastic process describing the median voter's peak across periods. In addition, we use $\mathcal{I} = (i_0, i_1, i_2, \ldots)$ to denote a path of policies that is generated when the median voter and the two parties decide according to the strategy profile given in Theorem 2, with the initial status-quo policy i_0, given $\mathcal{S}$ and $\mathcal{M}$.

For our analysis in this section, it is convenient to assume that the median voter's peak in period t, namely m_t, is chosen as follows:

$$m_t = \begin{cases} m_{t-1} & \text{with probability } \eta, \\ x & \text{with probability } 1 - \eta, \end{cases} \tag{7}$$

where x—and, in particular, m_0—is drawn according to a uniform distribution on $[-\beta, \beta]$, with

[26] Shocks neither depend on the identity of the incumbent nor on the number of periods in office.

$\beta \geq 0$.[27] We can interpret β as the variance of shocks to social preferences. The probability $\eta \in [0,1]$ measures the degree of persistence of the median voter's preferences across periods. If $\eta = 0$, the median voter's peak is i.i.d. across periods. By contrast, if $\eta = 1$, the median voter's position is fixed across periods and equals m_0.

In the analysis of Sections 4 and 5, we have assumed for simplicity (and without loss of generality) that any voter—the median voter, in particular—derives a disutility equal to the absolute value of the incumbent's valence, namely $|a_t|$. In this section, we assume that while the average voter still derives the same disutility, the median voter has a utility from the incumbent's valence equal to $(1+\chi) \cdot a_t$, with $\chi \geq 0$. That is, the median voter suffers *more* from a low-valence incumbent than the average voter, yet both voters are affected negatively. This assumption is justified in Section 7.8. Furthermore, we also assume throughout this section that

$$(1+\chi) \cdot A > A^* := 2\Delta \cdot (c + 2\beta)\,. \tag{8}$$

This condition guarantees that one valence shock outweighs any other utility source by the median voter, no matter what his/her peak is. Condition (8) is thus sufficient to enable the possibility that an incumbent might be ousted by the median voter despite doing so is not socially optimal. This condition is not essential for most of our results in this section but considerably simplifies the exposition, while enriching the interpretations of our findings. It is straightforward to verify that A^*—the minimal level of the valence shock that guarantees that no incumbent who has received a valence shock is re-elected—increases with μ and β, and decreases with θ. Moreover, A^* increases with c if and only if c is low enough.

Next, recall that the (expected) stage utility of a voter in some period t can be decomposed into three terms: the utility from policies, the utility from the incumbent's valence, and the utility losses due to costs of change. The latter two terms capture the dimensions along which there is no conflict across citizens. Thus, it seems natural that any measure of welfare should consider a δ-discounted sum of such terms. As to the question how to weigh policies themselves in welfare, there are at least two reasonable options, which we now discuss. These two options will allow us to disentangle the different channels through which term length might affect welfare.

[27] The details of the conditional probability distribution determining the median voter's peak are not really important for qualitative equilibrium behavior other than ensuring that the median voter's preferences change over time and that there is some persistence in such preferences. These details are, however, of crucial importance for the precise determination of the optimal term length.

6.1 Two definitions of welfare

The first possibility is to focus on the utility that a voter with peak at zero would obtain in all periods. This is called *ex-ante welfare*. One rationale is that zero is at the same distance from both parties' peaks.[28] Another rationale for measuring welfare this way is to assume that while the median voter changes over time—for reasons outside of our model, such as *noise* in the electoral process—, the distribution of peaks for the entire society remains invariant. This leads to the following definition of welfare:

$$W^1(T) = \mathbb{E}_T\left[-\sum_{t\geq 1}\delta^{t-1}\cdot i_t^2\right] + \mathbb{E}_T\left[-\sum_{t\geq 1}\delta^{t-1}\cdot c\cdot |i_t - i_{t-1}|\right] + \mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot a_t\right], \tag{9}$$

where the expected values are taken with respect to the stochastic processes that define $\mathcal{S}$, $\mathcal{M}$, and $\mathcal{I}$, given the term length T.[29] The latter dependence is denoted explicitly by writing $\mathbb{E}_T[\cdot]$. In terms of ex-ante welfare, a social planner would therefore never call elections, dictate $i_t = 0$ in all periods t, and oust any incumbent who has experienced a valence shock. However, this outcome cannot be attained through elections, no matter the term length T. If $T = \infty$, incumbents with low valence cannot be replaced. If $T < \infty$, there will be costly policy shifts at times, as shown in Theorem 2. Moreover, in terms of utility derived from policies, *any* voter with a fixed peak would gain from having a constant policy situated in the middle, compared to one that switches back and forth around the middle.

Second, it is intuitive that more frequent elections allow to adapt policies to the future preferences of the electorate, and in particular to the median voter's peak, which varies across periods in our model. In the following, we define a welfare measure that internalizes the future evolution of the median voter's peak into account. Then, we investigate the differences in terms of term length between such a definition and the ex-ante welfare measure defined in Expression (9). Formally, the second definition of welfare—which is called *interim welfare*—is

$$\begin{aligned} W^2(T) = {} & \mathbb{E}_T\left[-\sum_{t\geq 1}\delta^{t-1}\cdot (i_t - m_t)^2\right] + \mathbb{E}_T\left[-\sum_{t\geq 1}\delta^{t-1}\cdot c\cdot |i_t - i_{t-1}|\right] \\ & + \mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot a_t\right]. \end{aligned} \tag{10}$$

This leads to

$$[W^2 - W^1](T) = 2\cdot \mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot m_t i_t\right] - \mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot m_t^2\right]. \tag{11}$$

[28] As we show in 6.3, assuming equal distance between parties does not lead to knife-edged results.

[29] Note that the identity of the party in office is immaterial to welfare when it is measured by Expression (9).

Again, the expected values are taken with respect to the stochastic processes that define $\mathcal{S}$, $\mathcal{M}$, and $\mathcal{I}$, given the term length T. Equation (11) captures the value in terms of the policies implemented that calling elections every T periods has for the median voter who shifts over time. Note that the second term in the right-hand side of Equation (11) is independent of T.

The next result characterizes the difference between the two notions of welfare.

Proposition 1

Let $T \in \mathbb{N}$. Then, except for a constant term, the following two functions are equal:

(a) $[W^2(T) - W^1(T)]$,

(b) $\left(\beta^2 - \left(\min\left\{\beta, \frac{c}{2}\right\}\right)^2\right) \cdot 2\Delta \cdot \frac{\eta}{1-\delta\eta} \cdot \frac{\delta^T \eta^T (1-\rho)^T}{1+\delta^T \eta^T (1-(1-\rho)^T)}$

To understand the above proposition, let

$$H_0(T) := \frac{\delta^T \eta^T (1-\rho)^T}{1+\delta^T \eta^T (1-(1-\rho)^T)} = \frac{1+\delta^T \eta^T}{1+\delta^T \eta^T \cdot (1-(1-\rho)^T)} - 1.$$

It is a matter of simple algebra to verify that $H_0(T)$ is a positive, real-valued and decreasing function.[30] This means, not surprisingly, that *ceteris paribus*, terms should be as short as possible if our goal is to adjust policy to the median voter's varying peak. Ex-ante welfare and interim welfare are equivalent in certain cases. First, they coincide when $\beta \leq c/2$. In this case, the shocks to the median voter's peak are mild enough so that incumbents are only ousted if they have suffered a valence shock, and then costs of change generate an incumbency advantage that persists no matter how often the median voter's peak fluctuates. This persistence guarantees that the probabilities of certain policy streams and streams of the median voter's peak are the same for *any* pair of initial peaks of the median voter that are located symmetrically around zero—see the proof of Proposition 1 for details. These pairs of streams have the same weight in welfare, and they cancel each other out with regard to the terms that depend on term length. This leads to equivalence between the two definitions of welfare. Second, $W^2(T) = W^1(T)$ if either $\eta = 0$ or $\rho = 1$. In the former case ($\eta = 0$), there is no persistence in the median voter's peak, which is therefore homogeneously distributed over $[-\beta, \beta]$ in *any* period. This symmetry property leads again to policy streams canceling each other out in interim welfare. In the latter case ($\rho = 1$), turnover occurs after each election since valence shocks themselves occur in every period regardless of the median voter's peak. Because the dynamics of the median voter's peak do not matter for outcomes, they do not matter either for welfare.

[30] Note that $\frac{d}{dx}\left(\frac{1+f(x)}{1+f(x)g(x)}\right) = f'(x)(1-g(x)) - f(x)g'(x)(1-f(x)) < 0$ if $f'(x) < 0$, $g'(x) > 0$ and $0 \leq f(x), g(x) \leq 0$.

A final remark on our approach to welfare is in order. One could additionally define ex-ante welfare and interim welfare for a given initial position m_0 of the median peak, which we can denote by $W^1(T|m_0)$ and $W^2(T|m_0)$, respectively. It can be verified that $[W^2(T|m_0) - W^1(T|m_0)] = 0$ if $m_0 \in [-c/2, c/2]$ and $\beta > c/2$. In this case, incumbents will be ousted when the median voter's peak is sufficiently biased, but this can only occur after the initial peak of median voter has changed for the first time. This guarantees that from the perspective of the first period, the cancellation in the welfare calculus between certain pairs of policy and median voter's peak streams also takes place. The symmetry property no longer holds when m_0 is itself very biased, however. Given that the initial peak of the median voter, m_0, is close enough to the peak of the initial office-holder, the difference $[W^2(T|m_0) - W^1(T|m_0)]$ becomes maximal for $T = \infty$ and minimal for $T = 1$. In contrast, when the initial peak of the median voter, m_0, is close enough to the peak of the party that is not in power initially, the difference $[W^2(T|m_0) - W^1(T|m_0)]$ becomes minimal at $T = \infty$ and maximal at $T = 1$.

6.2 Welfare analysis: Costs of change versus valence

Henceforth, we focus on ex-ante welfare as described by Expression (9). By Proposition 1, this is equivalent to considering interim welfare as described by Expression (10) when $\beta \leq c/2$, i.e., when preference shocks are low or mild.

6.2.1 The formula for the optimal length of political terms

The next result characterizes the term length T that maximizes $W^1(T)$, which we denote by T^* and call the *optimal term length.*

Proposition 2

Suppose Condition (8) holds. Then, the optimal term length T^* exists, is finite, and maximizes the following expression as a function of T:

$$M(T) := \frac{\delta^T}{1-\delta^T} \cdot \Bigg[T \cdot \frac{A}{2c \cdot \left(\mu - \frac{c}{2} \cdot (1+\theta)\right)} \cdot \frac{\rho}{1-\delta} - (1 - (1-\rho)^T) - (1-\rho)^T \cdot \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\} \cdot \frac{1 + \delta^T \eta^T (1 - 2(1-\rho)^T)}{1 + \delta^T \eta^T (1 - (1-\rho)^T)} \Bigg]. \tag{12}$$

To understand expression $M(T)$ in Proposition 2, it is convenient to split it into a number of parts. Specifically, we define the following functions:

$$H_I(T) := \frac{\delta^T}{1-\delta^T} \cdot T \cdot \frac{A}{2c \cdot \left(\mu - \frac{c}{2} \cdot (1+\theta)\right)} \cdot \frac{\rho}{1-\delta},$$
$$H_{II}(T) := -\frac{\delta^T}{1-\delta^T},$$
$$H_{III}(T) := -\frac{\delta^T}{1-\delta^T} \cdot \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\} \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}.$$

Then, we have

$$M(T) = H_I(T) + (1-(1-\rho)^T) \cdot H_{II}(T) + (1-\rho)^T \cdot H_{III}(T).$$

In the latter expression, $H_I(T)$ is a positive, real-valued function that captures the weight in ex-ante welfare of the average valence of the incumbent relative to the costs of change associated with policy shifts. Note that $H_I(T)$ increases with ρ (the probability of a valence shock) and with A (the extent of a valence shock). In turn, $H_{II}(T)$ is a negative, real-valued function that captures the weight in ex-ante welfare of the average number of policy shifts that occur because the incumbent has suffered at least one valence shock during one term, which happens with probability $1-(1-\rho)^T$. Finally, $H_{III}(T)$ is a negative, real-valued function that captures the weight in ex-ante welfare of the average number of policy shifts that occur despite the incumbent having not suffered a valence shock, because the median voter's peak at election date is too far away from the incumbent's peak. The probability that this occurs when the median voter's peak is drawn anew is equal to $1/(2\beta) \cdot \max\{0, \beta - c/2\}$. In fact, the term $H_{III}(T)$ becomes zero if $2\beta \leq c$, and in particular if $\beta \to 0$ and the median voter does not shift over time. Then, it is a matter of algebra to check that not only $H_I(T)$ is decreasing and $H_{II}(T)$ and $H_{III}(T)$ are increasing, but that $(1-(1-\rho^T)) \cdot H_{II}(T) + (1-\rho^T) \cdot H_{III}(T)$ is also decreasing. In particular, $H_I(T)$ is maximal at $T = 1$, while $H_{II}(T)$ and $H_{III}(T)$ are maximal in the limit as T goes to infinity. This is not surprising: having the highest average valence can be attained by calling elections every period ($T = 1$); costs of change—which only materialize if there is government turnover, be it for lack of valence or for changes in the median voter's peak—can be completely avoided if elections are never called ($T = \infty$).

Finally, Figure 2 displays $M(T)$ for some parameter values as a way to illustrate the shape of the objective function that determines the optimal term length.

Figure 2 shows that the optimal term length is more than one for some generic parameter values. In such cases, there is a discrepancy between what the optimal term length is from an ex-ante perspective and from the perspective of a voter who wants to have the *possibility* to oust the

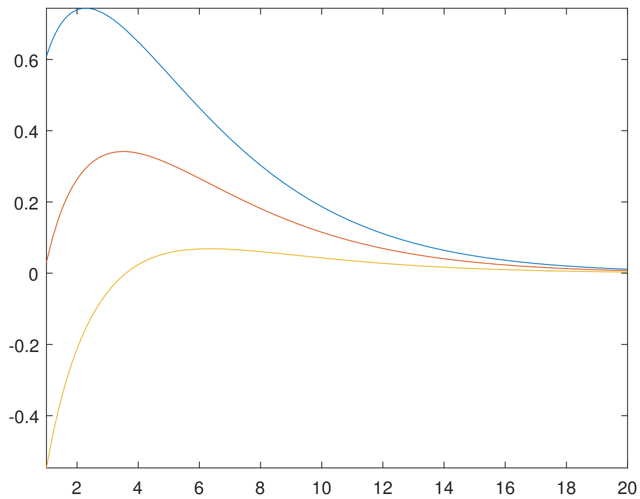

Figure 2: Function $M(T)$ for $\delta = 0.7$, $\mu = 0.4$, $c = 0.35$, $\theta = 0.8$, $\rho = 0.45$, $\beta = 0.2$, $\eta = 0.1$, $\chi = 6$, and $A = 0.01$ (orange line), $A = 0.02$ (red line) and $A = 0.03$ (blue line), where the x-axis encompasses different values of T (say, in years).

incumbent in every period by means of elections if s/he deems it necessary. Why is it optimal for citizens to tie their hands? In our model, ousting an incumbent generates a costly policy shift that lowers the utility of *all* citizens alike. The median voter might want to incur this cost under *one* of two circumstances: *(i)* when the incumbent's valence is too low, and *(ii)* when the preferences of the median voter are much closer to the challenger's than to the incumbent's. It then suffices to note that in *either* of these two cases, the interests of the (current) median voter might not be aligned with the interests of the average voter. The latter, who has peak permanently at zero and suffers less from low valence, defines ex-ante welfare. It is worth noting that while the fact that citizens exhibit an extreme form of quasi-hyperbolic discounting generates inefficiencies in policies and elections, these inefficiencies cannot be corrected by changing the term length.[31] In other words, the need for voters to tie their own hands from an ex-ante welfare perspective does *not* stem from the assumption that voters are present-biased. This is shown in Section 6.3 below.

6.2.2 Comparative statics and empirical hypotheses

Figure 2 illustrates the effect on the optimal term length of increasing the extent of the valence shocks. In this section, we show that Proposition 2 allows us to obtain a series of insightful

[31]If voters were forward-looking and discounted utility in all periods at rate θ, policy choices would differ from our model, but the dynamics would be qualitatively equal—see Gersbach et al. (2019a).

comparative statics results with regard to optimal term length, including the effect of varying A. These results yield hypotheses that could be tested empirically, thereby enabling us to see whether term lengths are set optimally in accordance with our theory. We divide our analysis in three parts, each of which is discussed in the following. The first corollary is concerned with changes in the valence shocks and the discount factor.

Corollary 1

The optimal term length T^*

- *decreases (as a function of A) if and only if A increases,*
- *decreases (as a function of ρ) if and only if ρ increases, and*
- *decreases (as a function of θ) if and only if θ increases.*

That is, the optimal term length decreases with the probability that a valence shock occurs, with the extent of such shocks, and with the factor with which agents discount the future period. The first result is not surprising, since the expected valence of incumbents will be lower if shocks are greater. A shorter term simply reduces the (expected) disutility generated by low valence incumbents, all else equal. For a fixed term length, increasing the probability of valence shocks decreases the expected valence of the incumbent—which calls for shorter terms—but at the same time it increases the probability that a costly policy change occurs—which calls for longer terms. However, we show that the former effect dominates the latter, so an increase of the valence shock probability ρ calls for shorter periods. It is also worth noting that the expected per-period valence variation, ρA, is *not* a sufficient statistic to determine T^*. While the welfare term for valence does depend on ρA, the probability ρ alone influences how often incumbents are going to be ousted and costly policy changes are going to come about. This implies that T^* might have very different values, even if we keep ρA constant. Finally, when the future becomes more valuable for citizens and candidates, policies become more moderate and thus policy shifts carried out when power shifts become smaller. This implies that ousting the incumbent generates fewer costs of change, and as result of this change in θ, the optimal term can never become larger, all else equal. In the particular case where δ is set (and kept) equal to θ, increasing θ also makes the future more valuable from a social viewpoint. This calls further for lower terms.

The first two parts of Corollary 1 have some implications for the design of political institutions. In cases where candidates undergo tight scrutinity until they are selected, the optimal length of terms should be larger, all else equal. This is because the probability and/or the size of the shocks to valence would be smaller (or, even, they could be made zero). Typical examples are courts, and in particular the US Supreme Court, for which the length is maximal, as their members are appointed for life. An argument for longer terms could also be made in democracies where candidates for particular seats are selected through long periods of observation within party organizations, and very particularly when parties have full control over the electoral lists. It has been argued that in the case of the US, party control over Presidential nomination has weakened since the 1970's (see e.g. Levitsky and Ziblatt, 2018). *Ceteris paribus*, this would call for shorter terms. In the case of institutions in which elected officials represent small districts, valence of representatives can have major consequences for the importance that such districts will gain in terms of policy-making. Terms should be short to avoid that constituencies considerably suffer for a long time from incompetent representatives. This could be the case of parliaments with a large number of members, such as the US House of Representatives. The opposite argument can be made for small collective decision bodies.

As a second corollary to Proposition 2, we look at changes in party polarization and the process determining the median voter's peak.

Corollary 2

The optimal term length T^*

- *increases (as a function of μ) if and only if μ increases,*
- *increases (as a function of β) if and only if β increases, and*
- *increases (as a function of $1-\eta$) if and only if $1-\eta$ increases.*
- *increases (as a function of χ) if and only if χ increases.*

That is, the optimal term length increases with party polarization, with the variance of the median voter's peak when it changes, and with the probability that the median voter's peak will actually change.[32] First, an increase of party polarization leads to more polarized policies and thus larger policy shifts, so that ousting the incumbent generates more costs of change. As result, the optimal term can never become smaller because it will enable more frequent power shifts. Second, an increase of β leads to an increase of T^* (at least, weakly). This is because if the extent (not the probability) of the shocks that affect the median voter's peak becomes greater, it also becomes more likely that the (new) median voter will prefer to oust the incumbent due to policy reasons, thereby generating higher costs of change. This property holds not only for

[32] In other words, the optimal term length decreases with the extent of persistence in the median voter's peak.

our parametrized conditional distribution, but for any distribution determining the new median voter's peak that changes as to put more mass on higher absolute values of the peak. This reflects the circumstances under which voters' preferences become more polarized. We stress that for changes in β to have an impact on T^*, it must be that $2\beta \geq c$. If the latter condition does not hold, an incumbent is never ousted for policy reasons. Third, if citizens' preferences become more stable—in the sense that the median voter's peak changes less frequently—, the optimal term length can never become larger. This is because with more stable social preferences, it is less likely that a power shift may occur for policy reasons and then enabling more frequent elections entails fewer risks from this perspective. Fourth and last, T^* must be (weakly) longer if the misalignment between the average voter and the median voter with regard to how much they value the infefficient provision of public goods becomes larger. The reason is that this makes it more likely that the (current) median voter will oust the incumbent and trigger a policy change although the costs associated with the policy shift are too large from from an ex-ante perspective.

Corollary 2 suggests that the term length should be larger when societies are strongly polarized and/or subject to large swings in the electorate. Otherwise, the society may incur excessive costs associated with policy changes. From the evidence on increasing polarization in most democratic societies that has taken place in the last decades, our results indicate that those term lenghts which have remained constant may *now* be too short, all else being equal.

Our last corollary is concerned with a crucial feature of our model of political competition, namely the marginal cost of changing policy.

Corollary 3

The optimal term length T^*

- *increases (as a function of c) if and only if c increases, provided that $2\beta < c < \frac{\mu}{1+\theta}$, and*
- *decreases (as a function of c) if and only if c increases, provided that $\frac{\mu}{1+\theta} < c < 2\beta$.*

Accordingly, the effect of c on T^* is ambiguous in general. If fluctuations of the median voter's peak and the marginal cost of change are small ($2\beta < c < \frac{\mu}{1+\theta}$), a marginal increase of c yields a higher optimal term length. We recall that in such case, incumbents are only ousted if their valence is (sufficiently) low. A marginal change of c therefore only affects the costs of change that will accrue whenever power shifts from one party to the other. These costs are given by the following quadratic function on c:

$$c \cdot \left(\frac{2\mu}{1+\theta} - c \right).$$

The above expression captures the fact that a marginal increase of c has two effects: a direct effect on how costly it is to make marginal changes to policies; an indirect effect on the extent of the policy shift. When $c < \frac{\mu}{1+\theta}$, in particular, the direct effect dominates and an increase of c yields higher costs associated with power shifts. To reduce the disutility generated by such costs, T^* must be (weakly) higher. By contrast, if fluctuations of the median voter's peak and the marginal cost are large ($\frac{\mu}{1+\theta} < c < 2\beta$), the indirect effect dominates and a marginal increase of c reduces the costs associated with turnover. This calls for lower terms. Additionally, when $c < 2\beta$, incumbents may also be ousted due to policy reasons. In such cases, increasing c reduces the probability that this happens, thereby also making lower terms more appealing, all else equal. Overall, T^* becomes (weakly) lower as a result of an increase of c.

The latter results may be relevant for the design of democratic institutions. In a narrow sense, parameter c corresponds to the costs imposed on all citizens per unit of policy change. However, in a broader sense, it can also be related to the institutions that govern the political system, and to checks and balances, in particular. In most democracies, the larger the policy reform, the more hurdles the proponents of the change have to overcome.[33] Such hurdles are judiciary oversight, qualified majority, or double majority, for instance. Then, assuming $2\beta < c < \mu/(1+\theta)$ and interpreting c as the set of institutional hurdles that are necessary to change policy, we obtain that higher hurdles call for higher term lengths. In the US, for instance, members of the Supreme Court are nominated by the President but have to be confirmed by the Senate, and in earlier times confirmation required a super majority. This double step can be seen as an instance of a high institutional hurdle. Our result of Corollary 3 regarding comparative statics on c can further be used to rationalize that term length for the Supreme Court members should be longer than, say, for members of the legislative or executive power, provided that fluctuations of the median voter's peak are low. If the latter are high, by contrast, higher hurdles call for lower term lengths, provided that hurdles were already high.

6.3 The role of quasi-hyperbolic discounting and party symmetry

In our previous analysis, we have assumed that voters are present-biased and that parties' peaks are located symmetrically around zero. What is the role of these assumptions for optimal term length? First, we focus on the role of quasi-hyperbolic discounting. Consider a citizen with peak at i, with $i \geq 0$. From Theorem 2, we know that when s/he is present-biased, s/he will oust the

[33]The potential impact of such hurdles on policy-making has been discussed in the literature: Diermeier and Myerson (1999) studied the consequences of bicameralism, Shepsle and Weingast (1987) discussed the power of committees in legislatures, and Huber (1992) studied further restrictive legislative procedures.

incumbent from party R—see Equation (5)—in some period t if

$$-a_t > 2c\Delta + 4\Delta i. \tag{13}$$

The term $2c\Delta$ captures the *total* costs associated with policy turnover, while the term $4\Delta i$ captures the (relative) utility for voter i of having party R dictate policy instead of party L. Note that Inequality (13) does not depend on the exact value of θ. If Inequality (13) holds—and voter i has present-biased preferences as we have assumed in our model—, s/he will therefore oust the incumbent.

Consider now a central planner with peak at i who has a lifetime utility based on standard exponential discounting, and assume that s/he represents the entire society (i.e., s/he defines what is desirable from a social perspective). Is it possible that the central planner does not want the present-biased voter i to oust the incumbent in period t? The only way to prevent this from happening is that elections are not called in period t, which requires a term length comprising at least two periods. If the incumbent is not ousted, there are two possibilities for period $t+1$ depending on whether the incumbent from party R has suffered another valence shock or not. In either case, the central planner derives less utility in period $t+1$ from the incumbent than from the challenger who belongs to party L. In fact, this property holds for every future period until the incumbent is ousted. Hence, the central planner is hurt if s/he does not allow the present-biased voter i to already oust the incumbent from party R in period t. An analogous argument holds if the incumbent belongs to party L. That is, present-biased preferences can *never* entail too frequent policy changes from a social welfare perspective.[34,35] Such preferences can only lead to too infrequent policy changes, if at all.[36] Because term length T is bounded from below by one period, the optimal term length will always entail one period if there is no misalignment between the median voter and the average voter in terms of preferences for policies and/or valence.

Second, let us assume that parties' peaks μ_R and μ_L do not satisfy the condition $\mu_R + \mu_L = 0$, i.e., they are not at the same distance from zero, albeit being still on different sides of the political spectrum. We also still assume that ex-ante welfare is defined by the lifetime utility of a voter with peak permanently at zero. As for equilibrium behavior, our analysis of Section 5 already

[34]If parties are themselves forward-looking, Gersbach et al. (2019a) have shown that the main dynamics are similar to the case where parties are present-biased, yet they will differ quantitatively. This has no bearing on optimal term length.

[35]Present-biased preferences can lead to too frequent policy changes to happen if costs of change are incurred over more than one period. Mathematically, this is equivalent to requiring the median voter to suffer less from costs of change than the average voter.

[36]This property does not hinge on the assumption that costs of change are linear (see Gersbach et al., 2019a).

yields parties' and voter's choices.[37] That is, party R will choose $\Delta_R := \mu_R - c/2(1+\theta)$ and party L will choose $\Delta_L := \mu_L + c/2(1+\theta)$. As for election, if we let $2\Delta = \Delta_R - \Delta_L$, then the incumbent $k \in K$ will be re-elected in period t if and only if

$$m_t \geq \frac{\mu_R + \mu_L}{2} - \frac{c}{2} + \frac{A \cdot z_t}{4\Delta}$$

if $K = R$, and if and only if

$$m_t \leq \frac{\mu_R + \mu_L}{2} + \frac{c}{2} - \frac{A \cdot z_t}{4\Delta}.$$

For instance, if $\mu_R + \mu_L > 0$, then it is less (more) likely that an incumbent from party R (party L) will be re-elected. The reason is that a voter with peak at zero is closer to party L with peak at μ_L than to party R with peak at μ_R. What is the impact of $\mu_R + \mu_L \neq 0$ on optimal term length? To gain an intuition about the answer to this question, we focus on the particular case where the fluctuations of the median voter are small, by assuming that $\beta = 0$.[38] Then, we take Equation (9) and investigate the term

$$\mathbb{E}_T\left[\sum_{t\geq 1} \delta^{t-1} \cdot i_t^2\right]. \tag{14}$$

When $\mu_R + \mu_L = 0$, Expression (14) is independent of T. Then, assume without loss of generality that $\mu_R + \mu_L > 0$ and note that a (present-biased) voter with peak at zero will re-elect an incumbent from party R who has not suffered any shock if and only if

$$-\Delta_R^2 \geq -\Delta_L^2 - c \cdot (\Delta_R - \Delta_L).$$

The above inequality is equivalent to $\mu_R + \mu_L \leq c$. In turn, a (present-biased) voter with peak at zero will always re-elect an incumbent from party L, provided that s/he has not suffered any shock. Therefore we need to distinguish two cases. First, assume that $\mu_R + \mu_L \leq c$, in which case both parties enjoy a net incumbency advantage (in the absence of valence shocks). One can verify that in this case,

$$\mathbb{E}_T\left[\sum_{t\geq 1} \delta^{t-1} \cdot i_t^2\right] = -\frac{\Delta_R^2 + \Delta_L^2}{2} \cdot \frac{1}{1-\delta}. \tag{15}$$

The above term is independent of T. This means that if the parties' peak asymmetry is small relative to costs of change, our anaylsis about optimal term length remains intact. To obtain Equation (15), define V_R (V_L) as the expected value of Expression (14) when the incumbent belongs to party R (L). Then,

$$V_R = -\Delta_R^2 \cdot \frac{1-\delta^T}{1-\delta} + (1-\rho)^T \delta^T \cdot V_R + (1-(1-\rho)^T)\delta^T \cdot V_L$$

[37] It suffices to make a change of variables so that policy i becomes policy $i - (\mu_R + \mu_L)/2$.

[38] The case $\beta > 0$ yields similar insights from a qualitative perspective.

and

$$V_L = -\Delta_L^2 \cdot \frac{1-\delta^T}{1-\delta} + (1-\rho)^T \delta^T \cdot V_L + (1-(1-\rho)^T)\delta^T \cdot V_R.$$

Adding the two above equations yields Equation (15).[39] Second, assume that $\mu_R + \mu_L > c$, which means that party R does not enjoy any net incumbency advantage. Then, we have

$$V_R = -\Delta_R^2 \cdot \frac{1-\delta^T}{1-\delta} + \delta^T \cdot V_L$$

and

$$V_L = -\Delta_L^2 \cdot \frac{1-\delta^T}{1-\delta} + (1-\rho)^T \delta^T \cdot V_L + (1-(1-\rho)^T)\delta^T \cdot V_R.$$

This leads to

$$\begin{aligned}\mathbb{E}_T\left[\sum_{t\geq 1} \delta^{t-1} \cdot i_t^2\right] =& -\Delta_L^2 \cdot \frac{1}{1-\delta} \\ & - \frac{1}{1-\delta} \cdot \frac{\Delta_R^2 - \Delta_L^2}{1-\delta^T + (1-\rho)^T \delta^T} \cdot \left[(1-\delta^T) + (1-\rho)^T \delta^T \cdot \frac{2(1-(1-\rho)^T)\delta^T}{1+(1-(1-\rho)^T)\delta^T}\right].\end{aligned}$$

One can verify that the above expression is decreasing in T. This implies, *ceteris paribus*, that if a party's peak is much more extreme than the other party's peak from the perspective of the voter who has peak at zero, terms should be shorter. This avoids that the party which has an extreme peak can dictate policy for many periods and ensures that the party with the more moderate peak will only be ousted when its candidate has suffered a valence shock. In such case, s/he will be replaced by a candidate of the more extreme party for only one period, after which a fresh candidate of the moderate party will be elected.

7 Extensions

In this section we discuss some extensions of our baseline model. Combined, they show how the analysis and the results of the previous sections can be applied and extended.

7.1 Learning by doing

One important feature of our baseline set-up is that the politicians' valence can only be affected by negative shocks during tenure. This rules out, for instance, the possibility that an incumbent may be more able to provide public goods at the end of a given term than at the beginning of that same term. There are several reasons why the valence of an incumbent may increase over time.

[39] We assume that it is equally likely for every party to be in power in period $t = 1$.

For instance, office-holders may need time to become efficient in policy-making through learning by doing or need some time to build a team or network that can execute their orders efficiently. The equilibrium analysis of Section 5 can be directly applied to the case where office-holders may experience both negative and positive shocks to their valence. It suffices to consider that in Theorem 2, $-Az_t$, the aggregate value of the shocks experienced by an office-holder up to some period t, may be negative or positive. The likelier it is that an incumbent experiences a positive shock to his/her valence, the longer the political terms should be from a welfare perspective. If incumbents can only experience positive shocks to their valence, in particular, the optimal length would be infinite.

The possibility that incumbents can learn over the course of their tenure adds an interesting aspect of the role of our (extreme) form of quasi-hyperbolic discounting. To illustrate this, we consider the simplest possible scenario, in which it takes for the incumbent two periods to *learn*—i.e, to master his/her abilities to govern, to build a competent team, and/or to create the necessary spillovers and compatibilities between governmental levels—and one additional period for the median voter to realize that the incumbent has learned (although citizens benefit from higher valence as soon as it increases).[40] Learning means that the incumbent's valence increases by an amount (approximately) equal to A. This increase can offset one valence shock and can only be experienced once in a lifetime. For simplicity, we assume that $\beta = 0$, i.e., the median voter's peak does not vary across periods. Then, to determine the effect of this feature on optimal term length, it suffices to add the term

$$H_{IV}(T) := \frac{\delta^T}{1-\delta^T} \cdot \frac{A}{2c\Delta} \cdot \left[\frac{1}{1-\delta} \cdot \frac{\delta - \delta^T}{\delta^T} + (1-\rho)^T \frac{1-(1-\rho)^T\delta^T - T(1-\rho)^T\rho\delta^T}{1-(1-\rho)^T\delta^T - (T(1-\rho)^T\rho\delta^T)^2} \right] \quad (16)$$

to the expression $M(T)$ defined in Equation (12). To derive the above expression, define $r_t = 1$ if the incumbent has already learnt, and $r_t = 0$ if s/he has not. Then, if we use $V_{x,y}$ to denote the expected value of r_t, depending on whether the incumbent has already learned ($x = 1$) or not ($x = 0$), and depending on the number y of valence shocks s/he has already suffered, we have that for each term length T,[41]

$$V_{0,0} = A\frac{\delta - \delta^T}{1-\delta} + (1-\rho)^T\delta^T V_{1,0} + T(1-\rho)^{T-1}\rho\delta^T V_{1,1} + \left(1-(1-\rho)^T - T(1-\rho)^T\rho\right)\delta^T V_{0,0},$$
$$V_{1,0} = A\frac{1-\delta^T}{1-\delta} + (1-\rho)^T\delta^T V_{1,0} + T(1-\rho)^{T-1}\rho\delta^T V_{1,0} + \left(1-(1-\rho)^T - T(1-\rho)^T\rho\right)\delta^T V_{0,0},$$
$$V_{1,1} = A\frac{1-\delta^T}{1-\delta} + (1-\rho)^T\delta^T V_{1,1} + T(1-\rho)^{T-1}\rho\delta^T V_{0,0} + \left(1-(1-\rho)^T - T(1-\rho)^T\rho\right)\delta^T V_{0,0}.$$

[40] A similar, yet more cumbersome, argument can be made if it takes incumbents $\bar{T} > 2$ periods to learn.

[41] Note that a politician who has suffered at least two valence shocks will be ousted no matter s/he has learned or not.

Solving the three above equations through standard algebra and noting that $H_{IV}(T) = V_{0,0}$ yields Equation (16). Then, it is easy to verify that

$$\lim_{T \to \infty} H_{IV}(T) > H_{IV}(1).$$

That is, $H_{IV}(T)$ cannot have a maximum at $T = 1$. This calls for terms that entail more than one period, all else being equal. The reason why it might be socially desirable for the electorate to tie its hands is clear. Doing so gives time to the (present-biased) median voter to realize that the incumbent has increased his/her valence through learning-by-doing.

7.2 Campaign spending, accountability and pandering

The model of electoral competition introduced in Section 3 captures the main effects of costs of change in policy-making and elections. This has allowed us to build a theory for the optimal determination of the length of a political term, which allows in turn a number of insightful comparative statics. Of course, many factors are also relevant for actually determining the term length, which are absent from our analysis in Section 5. These include, among others, campaign spending, politicians' accountability and pandering. Our model can easily be augmented to take these factors into account.

First, assume that carrying out elections generates a per capita cost to each citizen that is equal to K. If elections take place every T periods, the average cost per period is equal to K/T. Then, to determine the term length optimally, it suffices to add the term

$$H_V(T) := -\frac{\delta^T}{1-\delta^T} \cdot \frac{K}{2c\Delta}$$

to the expression $M(T)$ defined in Equation (12). Trivially, $H_V(T)$ is maximized at $T = \infty$. This means, not surprisingly, that (inefficient) campaign spending calls for longer terms all else equal. The relevant point, however, is that our model potentially allows us to capture the *degree* to which campaign spending influences the optimal term length, compared to other parameters. In the US federal elections, to use a real-world example, campaign spending is very large, and has risen dramatically in the past decades.[42] This raises the question whether terms should be longer to avoid such (inefficient) costs, particularly in the House of Representatives. By calibrating our model empirically, our theory could indeed help to provide an assessment of the degree to which such term length should be increased, if at all.

[42]See `https://www.opensecrets.org/overview/cost.php?display=T&infl=Y` and `https://www.huffpost.com/entry/56-years-of-presidential-campaign-spending-how-2016_b_5820bf9ce4b0334571e09fc1`, retrieved 7 September 2019.

Second, if interpreted broadly, our baseline model already incorporates some degree of politician accountability. It suffices to assume that as the politician's valence decreases (via random shocks), it becomes more likely that s/he does not keep a promise or, say, that s/he does not provide enough of the promised public goods. Hence, politicians' with a lower valence are more likely to be ousted for accountability reasons. Taking this reduced-form perspective on accountability, our results thus show that all else being equal, shorter terms *do* favor accountability. This is in line with Schultz (2008).

Third and last, assume that in any period, the incumbent—but not the challenger—has the possibility of implementing some policy that yields voters a benefit q in the period t in which it is implemented, but that generates an average cost Q in period $t+2$, with $q-Q\cdot\delta^2<0$. That is, it is inefficient to implement such a policy from an ex ante perspective. The incumbent has more information about the policy, and already realizes in period t that costs Q will accrue, so that his/her net valuation of the policy is $q-Q<0$. Potential examples are across-the-board tax cuts or increasing pension benefits when it is clear that doing so is not sustainable. However, because the median voter is present-biased—s/he only looks one period ahead—and not retrospective, and the incumbent's benefits from holding office are very large, the incumbent will implement the policy to increase his/her chances of re-election against possible valence shocks and changes of the preferences of the electorate. Therefore, we can assume that at least for some generic parameter constellations, costs $q-Q\cdot\delta^2$ will accrue in any period in which elections are held. Then, to determine the term length optimally in the presence of our reduced form of pandering, it suffices to add the term

$$H_{VI}(T) := \frac{1}{2c\Delta}\cdot\left(q-Q\cdot\delta^2\right)\cdot\frac{\delta^T}{1-\delta^T}$$

to the expression $M(T)$ defined in Equation (12). Trivially, $H_{VI}(T)$ is maximized at $T=\infty$. This means that the possibility that incumbents pander to the electorate calls for longer terms, all else equal. This reduces the distortion in policies introduced by elections and is also in line with Schultz (2008).

7.3 Party entry at the median voter's peak

In our baseline model, the two parties cannot commit to policies before elections and are attached to their (polarized) peaks. This is reasonable when flip-flopping is punished by the electorate. In this section, we explore how entry of third parties or third candidates might affect our results. Recall that parties R and L have the same peaks as (group of) voters r and l. Then, it would not be appealing for a third party to enter the election process against these established parties if l and r were very partisan or if there were entry costs in the form of lack of awareness by the electorate, absence of proportionality of the electoral system or scarce (public) funds.[43] As it happens, the possibility of party entry at the median voter's peak can be ruled out in our citizen-candidate set-up when there are just two party positions (or peaks), and these are therefore the same as the voters' positions. In such a scenario, the median voter would change only when the relative size of the two parties flips.

Absent entry by a third party, another possible scenario is that the opposition party would appoint a candidate who is (credibly) attached to a more moderate position than the party itself, say, to the median voter's peak. Our model identifies the main trade-off in this situation: On the one hand, the chances for this less partisan candidate to be elected would increase compared to a candidate attached to the party's peak, and so would the party's rewards from office, namely b, if these were not private benefits for the office-holder himself/herself. Moreover, the policy implemented by this non-partisan candidate would be closer to the opposition party's peak relative to the incumbent party's peak, and this utility gain from policies would not be offset by the (larger) costs of change incurred by the non-partisan candidate. On the other hand, however, precisely because the policy chosen by the non-partisan candidate would differ from the peak of the party appointing him/her, this would result in higher disutility from policies compared to a candidate attached to the party's peak for as long as the non-partisan candidate would remain in office instead of the former. In particular, if the party's office benefits were zero or very small and the median voter's peak (and its dynamics) and valence shocks were such that the probability of election would not increase (much) if the opposition party appointed a more moderate candidate, such party would have little incentives to appoint less partisan candidate.[44]

As for the optimal term length, even if the median chose the policy in each period—say, by electing a clone of themselves—, the fact that the median voter's peak varies across periods is bad for the rest of the society. This calls for increasing term lengths. But politicians lose valence over time, so one needs to replace them at least occasionally. Overall, all our insights regarding optimal term length would extend to a model in which a candidate located at the median is elected whenever there is an election.

---

[43]With two incumbent parties, Downsian forces do not operate fully when there is a possible entry of a third party (Palfrey, 1984).

[44]Other reasons why candidates are committed to extreme positions include internal party politics. The literature on the causes of party and policy polarization is vast (see e.g. Roberts and Smith, 2003; Theriault, 2006; Heberlig et al., 2006). See Jones (2001); Binder (2003); Fiorina et al. (2005); Testa (2012); Hetherington (2001) for consequences of such behavior.

7.4 Several policy dimensions

In our analysis, we have assumed that there is only one policy dimension voters care about besides valence. On this dimension, policy changes are costly and voters and parties have diverging preferences. With more policy dimensions and separable (Euclidean) preferences, there are two polar cases. First, in the absence of any capacity constraint, office-holders would change policy towards their bliss point in all dimensions upon election. This would create an incumbency advantage in all dimensions where parties have diverging preferences, and our analysis of optimal term length would then easily extend from the case of one dimension to multiple dimensions. Second, there may exist capacity constraints precluding that more than one policy be changed in each period (see e.g. Chen and Eraslan, 2017). In such a scenario, upon election, the incumbent would focus on those dimensions where the other party had implemented the last policy change, and then reverse the change in each period lexicographically for each of these dimensions according to how polarized the parties' position and how large the marginal cost of change were. In such cases, longer terms would enable the incumbent to build a greater electoral advantage along tenure. This would call for longer terms, all else equal.

With multiple dimensions and capacity constraints, it could additionally be that the incumbent party can set the political agenda before elections, namely, that it can determine which policy dimension should be at the center of the campaign, and thus on the voters' mind. There are many ways for the incumbent party to influence this. For instance, it could try to pass some bill dealing with one particular political dimension, e.g. the extent of health coverage or the reform of the education system, or it could influence media coverage in general.[45] The incumbent party would then simply choose the dimension on which it enjoys a larger incumbency advantage.

7.5 Early elections

The existence of a (maximal) term length does not necessarily imply that all terms span over the same number of periods, as is the case for the US Presidency, for example. In many representative democracies (see e.g. Diermeier and Merlo, 2000, and the references therein), for instance, it is possible for incumbents to call early election. Our model can be easily adapted to include the possibility to call early elections in every period of a term, except the last one in which elections

[45]It is well known that incumbents use their power as office-holder to try and influence the political arena in their favor. An extensive literature in political science has addressed the so-called *issue ownership* phenomenon (see Petrocik, 1996; Van der Brug, 2004; Bélanger and Meguid, 2008, among others). For a recent paper on the long-term consequences of initiating a project on political conflict, see Howell et al. (2019).

are automatically called. Naturally, we must assume $T > 1$. For a sketch of our analysis, it is convenient to recall that we have proceeded under the assumption that per-period office benefits are constant and politicians are present-biased—i.e., they only care about the present period and the one the follows after such a period. As in Section 6, we assume for simplicity that an office-holder who has suffered at least one valence shock will never be re-elected. The problem of calling early elections can be described as an optimal stopping problem in finite horizon (see e.g. Kayser, 2005). By focusing on our model with costs of change, we disregard other elements that might play a role for calling early elections, such as the possibility for the electorate to punish the incumbent for opportunistic behavior if s/he calls elections for his/her own benefit.

Given that policy only shifts with office-holder turnover no matter the term length, the only variable relevant for the decision whether to call early elections is the probability of re-election if elections are called in the present period. This probability depends, in turn, on the median voter's peak and the office-holder's valence, and hence also on the probability of a valence shock, the degree of persistence of the median voter's peak and the variance according to which such a peak changes (when it changes). A first possibility is that early elections could be called at the start of each period—except in the first period of a term, since elections have just been called—, knowing the values of the median voter's peak and the office-holder's valence. One can easily see that the incumbent will call early elections only if s/he will be re-elected (with certainty). This happens if and only if the incumbent has not yet suffered a valence shock and the median voter's peak is not very biased towards the challenger's peak at election date.

Quite often, however, elections cannot be called immediately, but only for a certain date in the future at the earliest. This allows uncertainty to kick in, and hence the decision to call early elections, say at the start of some period t to take place at the end of this period, involves a gambling element. This is because the median voter's peak, as well as the incumbent's valence, might change between the moment elections are called and the moment they take place. For the sake of the argument, suppose that when deciding whether to call elections in period t, the incumbent tries to maximize the expected number of periods, denoted by N, that s/he will hold office when focusing on periods $t+1$ and $t+2$.[46] We focus on the case where elections will be (automatically) called at the end of period $t+1$ if they are not called in period t. Given the politician's horizon, incumbents will never call early elections in periods $t-1, t-2, \dots$ within the same term. The argument can be nonetheless applied recursively backwards. Without loss of

[46]If we stick to our baseline model, it is trivial to see that the incumbent will never call early elections if there is a positive probability that s/he will not hold power in period $t+1$. Remember that period $t+2$ does not enter in his/her maximization problem in period t.

generality, assume that the incumbent belongs to party R. Then, let $\mathbb{1}_{m_t \geq -c/2} = 1$ if $m_t \geq -c/2$ and $\mathbb{1}_{m_t \geq -c/2} = 0$ otherwise, and assume that $z_t = 0$, i.e., the incumbent has *not* yet suffered one valence shock at the beginning of period t.[47] On the one hand, if the incumbent calls early election to take place at the end of period t, we have

$$N = 2(1-\rho) \cdot \left[\eta \cdot \mathbb{1}_{m_t \geq -c/2} + (1-\eta) \cdot \frac{\beta + \min\{\beta, c/2\}}{2\beta} \right] := N_1.$$

That is, the incumbent will be appointed for periods $t+1$ and $t+2$, provided that s/he has not suffered a valence shock in period t (which happens with probability $1-\rho$) if the median voter's peak has not changed (which happens with probability η) and was initially not too biased toward party L, or if the the median voter's peak has been drawn anew and it is not too biased toward party L either. The probability that the latter happens is $(\beta + \min\{\beta, c/2\})/(2\beta)$. On the other hand, if the incumbent does not call early election, we have

$$N = 1 + (1-\rho)^2 \cdot \left[\eta^2 \cdot \mathbb{1}_{m_t \geq -c/2} + (1-\eta^2) \cdot \frac{\beta + \min\{\beta, c/2\}}{2\beta} \right] := N_2.$$

To derive the latter equality, we have applied the law of iterated expectation regarding the stochastic process described in (7). It is a matter of simple algebra to verify that

$$N_1 \geq N_2 \Leftrightarrow \frac{\beta + \min\{\beta, c/2\}}{2\beta} \geq \frac{1 - (1-\rho)\eta(2-(1-\rho)\eta) \cdot \mathbb{1}_{m_t \geq -c/2}}{(1-\rho)(1-\eta)(2-(1-\rho)(1+\eta))}. \tag{17}$$

Hence the incumbent will call early election only if c is large enough relative to β (and $z_t = 0$). Given that $(\beta + \min\{\beta, c/2\})/(2\beta) \leq 1$, a necessary condition for Inequality (17) to hold is that $\mathbb{1}_{m_t \geq -c/2} = 1$, as was the case when we assumed that incumbents can call early elections under complete information. Given that Inequality (17) solely depends on the model primitives, there are two possibilities. First, early elections are never called, which happens if Inequality (17) with $\mathbb{1}_{m_t \geq -c/2} = 1$ does *not* hold. Second, early elections are called if and only if the incumbent has not suffered any valence shock, which happens if Inequality (17) with $\mathbb{1}_{m_t \geq -c/2} = 1$ *does* hold. In the former case, the analysis of optimal term length conducted in Section 6 still applies. In the latter case, the full extent of a term will only be used for low-valence office-holders. In contrast, high-valence office-holders will act as if the term length were the shortest possible, regardless of the actual term length T. In this case, longer term lengths are not desirable from a welfare perspective.

[47]If $z_t = 1$, i.e., if the incumbent has already suffered a valence shock, then s/he will never call early elections.

7.6 Varying costs of change

Thus far, we have assumed that the marginal cost of change, c, was fixed across periods. Another possible extension of our baseline model would be to assume that the marginal cost of change may vary from period to period. Changes in c may be expected or unexpected. Following the lines of the proof of Theorem 2, it can be shown that after a politician's first period in office, an unexpected change in the marginal cost of change will only affect the policy choice in any of the subsequent periods in which s/he remains in office if the new marginal cost has become smaller. In this case, the policy will be shifted further towards the incumbent's peak. Note that a reduction of c could actually be the result of a learning process by the office-holder during tenure. An increase of c, by contrast, will only affect policy choices after the incumbent has been ousted, in which case the policy choices by all politicians will become more moderate thereafter. The net effect of changes in c in determining the optimal length of a political term has already been (partially) described in Corollary 3.

7.7 Endogenizing valence and the median voter's peak

In our baseline set-up, office-holders only take one action in every period of their tenure. At the same time, office-holders are also concerned with two (exogenous) random processes that affect their valence and the median voter's peak. Together with policy choices, these random processes determine whether incumbents can retain office or not. Assuming that both processes are exogenous has sufficed to isolate the effect of the most relevant strategic choice of an office-holder on elections, namely the policy that will be in place for the next period. Moreover, this assumption has yielded rich dynamics. Within our model, it is nevertheless worth asking about what happens if the probability ρ that the incumbent experiences a valence shock in a given period and the (conditional) uniform distribution on $[-\beta, \beta]$ that determines the median voter's peak with probability $1 - \eta$ may depend on some decision by the incumbent. One possible way to address this would be to assume that ρ and either η or β depend on $|i_t - i_{t-1}|$, or, more generally, on i_{t-1} and i_t. For instance, changing the status quo may induce changes in the political environment because policies may enter uncharted territory. In particular, such changes may trigger a decrease of η and increase of β and ρ. This would generally result in higher turnover, but would not change policy choices—Corollaries 1 and 2 describe the marginal effects of changing such parameters on the optimal term length. An alternative possibility is that during tenure, the office-holder can exert some (fixed) effort to either reduce ρ or to increase η, but s/he cannot do both. For instance, assume that in some period t before an election, besides choosing $i_t \in \mathbb{R}$, the incumbent has the possibility to choose

$$(\rho^*, \eta^*) \in \{(0,0),(1,1)\},$$

where ρ^* and η^* denote the probability of a valence shock and the persistence level that will be in place for the next term, respectively. The main trade-off for the office-holder is clear from our analysis set in Theorem 2. When the median voter is biased in favor of the incumbent prior to elections, setting either $\eta^* = 1$ or $\rho^* = 0$ would be desirable. The choice would depend on the model parameters—see Expressions (5) and (6). By contrast, if the median voter were biased in favor of the challenger, the incumbent would unambiguously choose $\rho^* = 0$—to avoid suffering a valence shock—and then set $\eta^* = 0$—and expect that the median voter's peak changes in the direction of the incumbent's peak. There are various ways how the incumbent might affect ρ^* and η^*. For instance, s/he could move towards $\rho^* = 0$ by buying influence in the media to suppress negative information about him/her. Alternatively, s/he could set $\eta^* = 1$ by spending resources in gerrymandering, or on manipulating the political system (legally) to leave the peak of the median voter unchanged. The effects on the optimal term length of office-holders determining (ρ^T, η^T) are once more given by Corollaries 1 and 2.

7.8 A micro-foundation for the inefficient provision of public goods

For our welfare anaylsis in Section 6 we have proceeded on the assumption that the median voter suffers (weakly) more from the valence shocks than the average voter. This has yielded interesting results regarding the optimal length of a political term, very particularly when there is not enough variance in the stochastic process that determines the median voter's peak. In this section, we show how our model can be extended to provide a micro-foundation for such an assumption. To that purpose, we focus on any period t of our dynamic game and assume that *within* this period—i.e., between the policy choice of period $t-1$ and the policy choice of period t and elections (if any)—a certain public good level x_t, with $x_t \geq 0$, will be provided by the incumbent. If x_t is provided, a voter with peak $i \in \mathbb{R}$ derives additional utility in period t equal to[48]

$$v_i(x) = i \cdot x - \frac{1 - a_{t-1}}{2} \cdot x^2 + \Psi \cdot a_{t-1},$$

where $\Psi > 0$ and a_{t-1} is the valence of the incumbent in period t (before any valence shock might be realized). That is, an incumbent with lower valence is less able in providing public goods

[48]Of course, there are more general utility specifications that yield the same results from a qualitative perspective. However, they simply make the analysis more cumbersome.

than an incumbent with higher valence. Crucially, we assume that the median voter with peak $m_t \in [-\beta, \beta]$ will choose the level x_t^* to be provided, i.e., s/he will choose the level $x \geq 0$ that maximizes $v_{m_t}(x)$.[49] This can be conceived as reflecting the power of the (current) median voter to determine the provision of public goods in midterm or local elections. It is therefore implicit that the incumbent's valence generates externalities or spillover effects on lower administrative levels, say through the right to influence its approval via the federal budget or by effectively controlling the monetary transfers needed to provide such goods. Then, it is a matter of simple algebra to verify that

$$v_{m_t}(x_t^*|a_{t-1}) = \frac{m_t^2}{2(1-a_{t-1})} + \Psi \cdot a_{t-1}.$$

Assuming that the distribution of peaks for the citizenry is symmetric around zero, the average utility corresponds to the utility of the average voter (who has peak at zero), namely

$$v_0(x_t^*|a_{t-1}) = -\frac{m_t^2}{2(1-a_{t-1})} + \Psi \cdot a_{t-1}.$$

Finally,

$$v_{m_t}(x_t^*|0) - v_{m_t}(x_t^*|-A) = \Psi \cdot A + \frac{m_t^2}{2} \cdot \frac{A}{1+A}$$

and

$$v_0(x_t^*|0) - v_0(x_t^*|-A) = \Psi \cdot A - \frac{m_t^2}{2} \cdot \frac{A}{1+A}.$$

That is, assuming that A is sufficiently large, both the median voter and the average voter suffer from an incumbent with lower valence, albeit the median voter derives a (weakly) higher disutility. Finally, note that for every ε such that $0 < \varepsilon < \beta$, if m_0 is chosen according to a uniform distribution on $[-\beta, -\varepsilon] \cup [\varepsilon, \beta]$, the disutility the median voter derives from an incumbent who has suffered one valence shock is at least $1 + \varepsilon^2 \cdot \frac{A}{1+A}(= 1 + \chi)$ higher (in absolute value) than the disutility the average voter derives.

8 Conclusion

An appropriate framework upon which to build a full-scale theory of optimal term lengths was missing from the literature. This paper has taken a first step towards filling this gap by introducing a model of electoral competition that allows insightful comparative statics about the optimal length of a political term with regard to some parameters that capture essential elements of elections and policy-making. In turn, this offers an array of hypotheses that can be tested empirically against the assumption that the term length is set optimally for particular political

[49] One can easily verify that the maximizer is $\frac{m_t}{1-a_{t-1}}$.

systems in accordance with our theory. While some of the comparative statics are intuitive when taking them individually, our analysis provides a quantitative approach that allows us to weigh each factor when we take all of them collectively. Of course, the features of our model can be complemented with further elements, particularly by introducing some asymmetry of information. This is also left for further research.

References

Acemoglu, D., Egorov, G., and Sonin, K. (2013). A political theory of populism. *Quarterly Journal of Economics*, 128(2):771–805.

Beck, T., Clarke, G., Groff, A., Keefer, P., and Walsh, P. (2001). New tools in comparative political economy: The database of political institutions. *World bank economic review*, 15(1):165–176.

Bélanger, É. and Meguid, B. M. (2008). Issue salience, issue ownership, and issue-based vote choice. *Electoral Studies*, 27(3):477–491.

Binder, S. A. (2003). *Stalemate: Causes and consequences of legislative gridlock*. Brookings Institution, Washington, DC.

Bisin, A., Lizzeri, A., and Yariv, L. (2015). Government policy with time inconsistent voters. *American Economic Review*, 105(6):1711–37.

Bowen, R., Chen, Y., and Eraslan, H. (2014). Mandatory versus discretionary spending. *American Economic Review*, 104(10):2941–2974.

Bowen, T. R., Chen, Y., Eraslan, H., and Zápal, J. (2017). Efficiency of flexible budgetary institutions. *Journal of Economic Theory*, 167:148–176.

Chen, Y. and Eraslan, H. (2017). Dynamic agenda setting. *American Economic Journal: Microeconomics*, 9(2):1–32.

Dal Bó, E. and Rossi, M. A. (2011). Term length and the effort of politicians. *Review of Economic Studies*, 78(4):1237–1263.

Diermeier, D. and Merlo, A. (2000). Government turnover in parliamentary democracies. *Journal of Economic Theory*, 94(1):46–79.

Diermeier, D. and Myerson, R. B. (1999). Bicameralism and its consequences for the internal organization of legislatures. *American Economic Review*, 89(5):1182–1196.

Duggan, J. (2017). Term limits and bounds on policy responsiveness in dynamic elections. *Journal of Economic Theory*, 170:426–463.

Fiorina, M. P., Abrams, S. J., and Pope, J. (2005). *Culture War?* Pearson Longman, New York, NY.

Forand, J. G. (2014). Two-party competition with persistent policies. *Journal of Economic Theory*, 152:64–91.

Gersbach, H., Jackson, M., Muller, P., and Tejada, O. (2019a). Electoral competition with costly policy changes: A dynamic perspective. *Mimeo.*

Gersbach, H., Muller, P., and Tejada, O. (2019b). The effect of costs of change on political polarization. *European Journal of Political Economy, forthcoming.*

Gersbach, H. and Tejada, O. (2018). A reform dilemma in polarized democracies. *Journal of Public Economics*, 160:148–158.

Glazer, A., Gradstein, M., and Konrad, K. A. (1998). The electoral politics of extreme policies. *Economic Journal*, 108(451):1677–1685.

Green, D. P. and Palmquist, B. (1994). How stable is party identification? *Political behavior*, 16(4):437–466.

Heberlig, E., Hetherington, M., and Larson, B. (2006). The price of leadership: Campaign money and the polarization of congressional parties. *Journal of Politics*, 68(4):992–1005.

Hetherington, M. J. (2001). Resurgent mass partisanship: The role of elite polarization. *American Political Science Review*, 95(3):619–632.

Howell, W., Krasa, S., and Polborn, M. (2019). Political conflict over time. *Mimeo.*

Huber, J. D. (1992). Restrictive legislative procedures in France and the United States. *American Political Science Review*, 86(3):675–687.

Hwang, J. and Möllerström, J. (2017). Design of reforms with time-inconsistent voters. *Journal of Public Economic Theory*, 19(3):748–761.

Jackson, M. O. and Yariv, L. (2015). Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics*, 7(4):150–78.

Jennings, M. K., Stoker, L., and Bowers, J. (2009). Politics across generations: Family transmission reexamined. *The Journal of Politics*, 71(3):782–799.

Jones, D. R. (2001). Party polarization and legislative gridlock. *Political Research Quarterly*, 54(1):125–141.

Kayser, M. A. (2005). Who surfs, who manipulates? The determinants of opportunistic election timing and electorally motivated economic intervention. *American Political Science Review*, 99(1):17–27.

Keppo, J., Smith, L., and Davydov, D. (2008). Optimal electoral timing: Exercise wisely and you may live longer. *Review of Economic Studies*, 75(2):597–628.

Kim, H. and Fording, R. C. (2003). Voter ideology in western democracies: An update. *European Journal of Political Research*, 42(1):95–105.

Lesmono, D., Tonkes, E. J., and Burrage, K. (2003). An early political election problem. *ANZIAM Journal*, 45:16–33.

Levitsky, S. and Ziblatt, D. (2018). *How democracies die*. Broadway Books.

Lizzeri, A. and Persico, N. (2001). The provision of public goods under alternative electoral incentives. *American Economic Review*, 91(1):225–239.

Lizzeri, A. and Yariv, L. (2017). Collective self-control. *American Economic Journal: Microeconomics*, 9(3):213–44.

Nunnari, S. and Zápal, J. (2017). Dynamic elections and ideological polarization. *Political Analysis*, 25(4):505–534.

Palfrey, T. R. (1984). Spatial equilibrium with entry. *Review of Economic Studies*, 51(1):139–156.

Petrocik, J. R. (1996). Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, 40(3):825–850.

Piguillem, F. and Riboni, A. (2015). Spending-biased legislators: Discipline through disagreement. *Quarterly Journal of Economics*, 130(2):901–949.

Roberts, J. M. and Smith, S. S. (2003). Procedural contexts, party strategy, and conditional party voting in the U.S. House of Representatives, 1971–2000. *American Journal of Political Science*, 47(2):305–317.

Schultz, C. (2008). Information, polarization and term length in democracy. *Journal of Public Economics*, 92(5-6):1078–1091.

Sears, D. O. and Funk, C. L. (1999). Evidence of the long-term persistence of adults' political predispositions. *The Journal of Politics*, 61(1):1–28.

Shepsle, K. A. and Weingast, B. R. (1987). The institutional foundations of committee power. *American Political Science Review*, 81(1):85–104.

Smart, M. and Sturm, D. M. (2013). Term limits and electoral accountability. *Journal of Public Economics*, 107:93–102.

Testa, C. (2012). Is polarization bad? *European Economic Review*, 56(6):1104–1118.

Theriault, S. M. (2006). Party polarization in the U.S. Congress: Member replacement and member adaptation. *Party Politics*, 12(4):483–503.

Van der Brug, W. (2004). Issue ownership and party choice. *Electoral Studies*, 23(2):209–233.

Wiesehomeier, N. and Benoît, K. (2009). Presidents, parties, and policy competition. *Journal of Politics*, 71(04):1435–1447.

Appendix

In this appendix we first provide the proofs of the results of the paper.

Proof of Theorem 1:

We start with some trivial remarks that follow directly from the equilibrium notion. First, because $c = 0$, the maximization problem any incumbent faces in any period of any term—and, in particular, in the first period of the term—does not depend on previous policy choices through costs of change. Second, when voting, the (present-biased) median voter only cares about the policy choices that either candidate will implement in the subsequent period if they are elected.[50] In particular, the median voter will not condition his/her decision on the policy choices prior to elections in order to reward "good behavior" of politicians, for instance. Third, we recall that valence shocks occur independently of any policy decision by the incumbent.

Now consider the median voter's decision. Let t be the period in which elections take place, and let m_t be the peak of the median voter in such period. Let, in addition,

$$i_{t+1}^{K} := \sigma_K(1/T, i_t, a_t, m_t) \quad \text{and} \quad i_{t+1}^{-K} := \sigma_{-K}(1/T, i_t, 0, m_t)$$

denote the policy choices by the incumbent $k \in K$ (with peak μ_K) and the challenger $-k \in -K$ (with peak μ_{-K}) in period $t+1$ if they are elected, respectively. As explained above, i_{t+1}^{K} and i_{t+1}^{-K} are independent of i_t. Moreover, because the median voter will vote for the candidate from whom s/he expects higher utility in period $t+1$ and valence shocks are additive and independent of any other variable in the model, one can easily verify that incumbent k will be re-elected if and only if

$$(2m_t - i_{t+1}^{K} - i_{t+1}^{-K}) \cdot \left(i_{t+1}^{K} - i_{t+1}^{-K}\right) \geq A \cdot z_t, \tag{18}$$

where z_t is the number of shocks suffered by the (current) incumbent at the end of period t. From the perspective of the incumbent who chooses the policy in t before the median voter's peak is determined and a valence shock might occur, the re-election probability, i.e., the probability that Equation (18) will hold, is then a function of i_{t+1}^{K} and i_{t+1}^{-K} (and z_t) only. Let p denote this probability.

Next, consider the problem faced by the incumbent k. We distinguish two cases. On the one

[50] Since neither politicians nor citizens can commit to policies or voting ahead of elections, promises made during the political campaign have no impact on equilibrium behavior.

hand, the problem of incumbent k in the beginning of period t where elections take place is

$$\max_{i_t^K \in \mathbb{R}} \Big\{ -(i_t^K - \mu_K)^2 + a_{t-1} - \rho A + p \cdot \theta \left(b - (i_{t+1}^K - \mu_K)^2 + a_t\right) \\ + (1-p) \cdot \theta \left(-(i_{t+1}^{-K} - \mu_K)^2\right) - \rho A \Big\}, \tag{19}$$

where p has been introduced above. Then, the problem in (19) is maximized for $i = \mu_K$, regardless of valence, since $-(i_t^K - \mu_K)^2$ is the only term that depends on i_t^K.

On the other hand, suppose that the incumbent has at least one further period in the present term in which s/he can choose a policy before the next election takes place, i.e., we are in some period which is prior to period t but belongs to the same term. Then, the incumbent faces the following problem, say in period t':

$$\max_{i_{t'}^K \in \mathbb{R}} \Big\{ -(i_{t'}^K - \mu_K)^2 + a_{t'-1} + \rho \cdot \left[-A + \theta \cdot (\sigma_K((t'+1)/T, i_{t'}^K, a_{t'-1} - A) - \mu_K)^2\right] . \\ + (1-\rho) \cdot \left[\theta \cdot (\sigma_K((t'+1)/T, i_{t'}^K, a_{t'-1}) - \mu_K)^2\right] - \theta \cdot \rho A \Big\}. \tag{20}$$

Let us assume, in particular, that $t' = t-1$. Then, we have $\sigma_K((t'+1)/T, i_t^K, a_{t'-1} - A) = \sigma_K((t'+1)/T, i_t^K, a_{t'-1}) = \mu_K$, and the maximization of (20) is achieved at $i_{t'}^K = \mu_K$. Iterating the argument backwards to the first period of the term that ends in period t shows that the incumbent will choose his/her peak in all periods of the term.

Finally, given that the incumbent always chooses his/her peak, Equation (18) reduces to

$$m_t \geq \frac{A \cdot z_t}{4\mu}$$

if $K = R$, and to

$$m_t \leq -\frac{A \cdot z_t}{4\mu}$$

if $K = L$.

□

Proof of Theorem 2:

Throughout the proof, we will assume that the incumbent k belongs to party $K = R$. The case where the incumbent belongs to party L follows the same logic. We recall that

$$\Delta = \mu - \frac{c}{2} \cdot (1+\theta).$$

Assuming that $k \in R$ is the incumbent and $t/T = 0$, let

$$\sigma_m^*(K, i_{t-1}, a_t, m_t) = \begin{cases} K & \text{if } m_t \geq -\frac{c}{2} \cdot \frac{i}{\Delta} - \frac{a_t}{4\Delta}, \\ -K & \text{otherwise.} \end{cases} \tag{21}$$

For all $t \geq 1$, if $k \in R$ is the incumbent in period t, let

$$\sigma_R^*(t/T, i_{t-1}, a_{t-1}) = \Delta. \tag{22}$$

Similarly, for all $t \geq 1$, if $k \in L$ is the incumbent in period t, let

$$\sigma_L^*(t/T, i_{t-1}, a_{t-1}) = -\Delta. \tag{23}$$

The remainder of the proof consists in showing that the above strategies are best responses for the parties and the median voter, respectively, given that these same strategies will be played in the future (and have been played in the past). This will establish the result of the theorem. To this end, we will focus on a term that starts in some period $t+1$ and ends in period $t+T$, when elections take place. To facilitate reading, unless there is a possible confusion, we shall henceforth use the following notation for the analysis of period $h \in \{t+1, \ldots, t+T\}$:

$$\begin{aligned} j &:= i_{h-1}, \\ i &:= i_h, \\ m_- &:= m_{h-1}, \\ m &:= m_h, \\ z &:= z_{h-1}, \\ z_+ &:= z_{t+T}. \end{aligned}$$

It will also be convenient to define

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

Then, we proceed in three steps.

Step 1:

We start by considering the median voter's decision in the election that takes place in period $h = t+T$. Given (22) and (23), the incumbent k will choose Δ in period $t+T+1$ if s/he is re-elected. In turn, the challenger $k \in L$ will choose $-\Delta$ in period $t+T+1$ if s/he is elected instead. We stress that at the time of elections in period $t+T$, the median voter knows whether or not the incumbent has suffered a shock. We use

$$p(i, z_+, m_-) \tag{24}$$

to denote the probability that the median voter will elect the incumbent $k \in R$ when the latter has chosen i and has suffered z_+ shocks and before the median voter's peak m is determined according to $F(\cdot|m_-)$. We distinguish three cases.

<u>Case I:</u> $-\Delta \leq i \leq \Delta$

In this case, the median voter will re-elect k if and only if

$$-(m-\Delta)^2 - c \cdot (\Delta - i) - A \cdot z \geq -(m+\Delta)^2 - c \cdot (i + \Delta)\,,$$

which can be rearranged as

$$m \geq -\frac{c}{2} \cdot \frac{i}{\Delta} + \frac{A \cdot z}{4\Delta}. \tag{25}$$

From the above expression, it follows that

$$p(i, z, m_-) = \int_{-\beta}^{\beta} \mathbb{1}_m \left(-\frac{c}{2} \cdot \frac{i}{\Delta} + \frac{A \cdot z}{4\Delta} \right) dF(m|m_-).$$

Hence, $p(i, z, m_-)$ is non-decreasing in i.

<u>Case II:</u> $\Delta \leq i$

In this case, the median voter will re-elect k if and only if

$$-(m-\Delta)^2 - c \cdot (i - \Delta) - A \cdot z \geq -(m+\Delta)^2 - c \cdot (i + \Delta)\,,$$

which can be rearranged as

$$m \geq -\frac{c}{2} + \frac{Az}{4\Delta}. \tag{26}$$

Using the above expression, one can see that $p(i, z, m_-)$ is constant in i.

<u>Case III:</u> $i < -\Delta$

In this case, the median voter will re-elect k if and only if

$$-(m-\Delta)^2 - c \cdot (\Delta - i) - A \cdot z \geq -(m+\Delta)^2 - c \cdot (-\Delta - i)\,, \tag{27}$$

which can be rearranged as

$$m \geq \frac{c}{2} + \frac{Az}{4\Delta}.$$

Using the above expression, one can see that $p(i, z, m_-)$ is constant in i.

Step 2:

We next consider the problem faced by the incumbent $k \in R$ in the beginning of period $h = t+T$, before s/he might experience a valence shock, the median voter's peak will be determined, and elections will take place (all in the same period $t + T$). In this case, the incumbent faces the following problem:

$$\max_{i\in\mathbb{R}} G(i) := \max_{i\in\mathbb{R}} \Bigg\{ -(i-\mu)^2 - c\cdot|i-j| - \chi$$
$$+\theta\cdot(1-\rho)\cdot p(i,z,m_-)\cdot\left[b-(\Delta-\mu)^2 - c\cdot|\Delta-i| - A\cdot z\right]$$
$$+\theta\cdot\rho\cdot p(i,z+1,m_-)\cdot\left[b-(\Delta-\mu)^2 - c\cdot|\Delta-i| - A\cdot(z+1)\right]$$
$$+\theta\cdot(1-\rho)\cdot(1-p(i,z,m_-))\cdot\left[-(\mu+\Delta)^2 - c\cdot|i+\Delta|\right]$$
$$+\theta\cdot\rho\cdot(1-p(i,z+1,m_-))\cdot\left[-(\mu+\Delta)^2 - c\cdot|i+\Delta|\right]\Bigg\}$$

where χ is independent of i. Note that we can rearrange terms to obtain

$$\begin{aligned} G(i) &= -(i-\mu)^2 - c\cdot|i-j| - \theta\cdot c\cdot|i+\Delta| - \chi' \\ &\quad +\theta\cdot(1-\rho)\cdot p(i,z,m_-)\cdot\left[M^z - c\cdot|\Delta-i| + c\cdot|i+\Delta|\right] \\ &\quad +\theta\cdot\rho\cdot p(i,z+1,m_-)\cdot\left[M^{z+1} - c\cdot|\Delta-i| + c\cdot|i+\Delta|\right], \end{aligned} \tag{28}$$

where χ' is also independent of i,

$$M^z := b - A\cdot z + 4\mu\cdot\Delta$$

and

$$M^{z+1} := b - A\cdot(z+1) + 4\mu\cdot\Delta.$$

Given that $b > 0$ is assumed to be very large, so are M^z and M^{z+1}. Moreover, if play occurs in accordance with the proposed equilibrium, the *status-quo* policy j satisfies the following condition:

$$-\Delta \le j \le \Delta.$$

Finally, we note that $G(i)$ is differentiable for all $i \in \mathbb{R}$, except possibly in a finite number of points.

Case I: $-\Delta \leq i \leq j (\leq \Delta)$

In this case, using (28), we have

$$\begin{aligned} G(i) &= -(i-\mu)^2 - c\cdot(j-i) - \theta\cdot c\cdot(i+\Delta) - \chi' \\ &\quad + \theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\Big[M^z + 2c\cdot i\Big] + \theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\Big[M^{z+1} + 2c\cdot i\Big]. \end{aligned}$$

Then, whenever the derivative of $G(i)$ exists, we have (see Case I of Step 1)

$$\begin{aligned} G'(i) &= 2(\mu-i) + c\cdot(1-\theta) + 2c\cdot\theta\cdot\Big[(1-\rho)\cdot p(i,z,m_-) + \rho\cdot p(i,z+1,m_-)\Big] \\ &\quad + \theta\cdot(1-\rho)\cdot\frac{\partial p\left(i,z,m_-\right)}{\partial i}\cdot\Big[M^z + 2c\cdot i\Big] + \theta\cdot\rho\cdot\frac{\partial p\left(i,z+1,m_-\right)}{\partial i}\cdot\Big[M^{z+1} + 2c\cdot i\Big] \\ &\geq 2(\mu-i) + c\cdot(1-\theta) \geq 0, \end{aligned}$$

where the first inequality holds because $p\left(i,z+1,m_-\right)$ and $p\left(i,z+1,m_-\right)$ are non-decreasing probabilities and M^z and M^{z+1} are very large, and the second inequality holds since $c > 0$, $\theta \leq 1$, and

$$i \leq j \leq \Delta = \mu - \frac{c}{2}\cdot(1+\theta) \leq \mu.$$

Finally, given that M^z and M^{z+1} are very large, it follows that for any i_* where the derivative does not exist, we have

$$\lim_{i\to i_*^+} G(i) \geq \lim_{i\to i_*^-} G(i).$$

To sum up, we have shown that

$$G(i) < G(j) \text{ for all } i < j.$$

Case II: $(-\Delta \leq) j \leq i \leq \Delta$

In this case, we can write

$$\begin{aligned} G(i) &= -(i-\mu)^2 - c\cdot(i-j) - \theta\cdot c\cdot(i+\Delta) - \chi' \\ &\quad + \theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\Big[M^z + 2c\cdot i\Big] + \theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\Big[M^{z+1} + 2c\cdot i\Big]. \end{aligned}$$

Hence, we have (see Case II of Stage 1)

$$\begin{aligned} G'(i) &\geq 2(\mu-i) - c\cdot(1+\theta) + 2c\cdot\theta\cdot\Big[(1-\rho)\cdot p(i,z,m_-) + \rho\cdot p(i,z+1,m_-)\Big] \\ &\quad + \theta\cdot(1-\rho)\cdot\frac{\partial p\left(i,z,m_-\right)}{\partial i}\cdot\Big[M^z + 2c\cdot i\Big] + \theta\cdot\rho\cdot\frac{\partial p\left(i,z+1,m_-\right)}{\partial i}\cdot\Big[M^{z+1} + 2c\cdot i\Big] \\ &\geq 2(\mu-i) - c\cdot(1+\theta) \geq 0, \end{aligned}$$

where the first inequality holds because $p\left(i, z+1, m_{-}\right)$ and $p\left(i, z+1, m_{-}\right)$ are non-decreasing probabilities and M^{z} and M^{z+1} are very large, and the second inequality holds since $c>0$, $\theta \leq 1$, and

$$i \leq \Delta=\mu-\frac{c}{2} \cdot(1+\theta).$$

Then,

$$G(i)<G(\Delta) \text{ for all } j \leq i<\Delta.$$

Case III: $\Delta \leq i$

In this case, we can write

$$\begin{aligned} G(i)= & -(i-\mu)^{2}-c \cdot(i-j)-\theta \cdot c \cdot(i+\Delta)-\chi^{\prime} \\ & +\theta \cdot(1-\rho) \cdot p\left(i, z, m_{-}\right) \cdot\left[M^{z}+2 c \cdot \Delta\right]+\theta \cdot \rho \cdot p\left(i, z+1, m_{-}\right) \cdot\left[M^{z+1}+2 c \cdot \Delta\right]. \end{aligned}$$

Hence, we have (see Case III of Stage 1)

$$G^{\prime}(i)=2(\mu-i)-c \cdot(1+\theta) \leq 0,$$

where last inequality holds now since

$$\mu-\frac{c}{2} \cdot(1+\theta)=\Delta \leq i.$$

Then,

$$G(i)<G(\Delta) \text{ for all } \Delta<i.$$

Case IV: $i \leq-\Delta$

In this case, we can write

$$\begin{aligned} G(i)= & -(i-\mu)^{2}-c \cdot(j-i)-\theta \cdot c \cdot(-i-\Delta)-\chi^{\prime} \\ & +\theta \cdot(1-\rho) \cdot p\left(i, z, m_{-}\right) \cdot\left[M^{z}-2 c \cdot \Delta\right]+\theta \cdot \rho \cdot p\left(i, z+1, m_{-}\right) \cdot\left[M^{z+1}-2 c \cdot \Delta\right]. \end{aligned}$$

Hence, whenever the derivative of $G(i)$ exists, we have

$$G^{\prime}(i)=2(\mu-i)+c \cdot(1+\theta) \geq 0,$$

where last inequality holds since

$$i \leq-\mu+\frac{c}{2} \cdot(1+\theta) \leq 0 \leq \mu.$$

As in Case I, one can see that

$$G(i) < G(-\Delta) \text{ for all } i \leq -\Delta.$$

To sum up, we have proved that the problem described in (28) is maximized for $i = \Delta$.

Step 3:

Finally, we consider the case where the incumbent k has at least two periods ahead of him/her in a term that starts in period $t+1$ and ends in period $t+T$. For this to be possible, it must be that $T > 1$. Then, the incumbent faces the following problem in a particular period $h = t+1, \ldots, t+T-1$:

$$\max_{i \in \mathbb{R}} H(i) := \max_{i \in \mathbb{R}} -(i-\mu)^2 - c \cdot |j-i| - \theta \cdot c \cdot |\Delta - i| - \chi', \tag{29}$$

where χ' is independent of i, and we assume that in period $h+1$ the incumbent will choose Δ. Note that this has been proved to be the case for period $t+T$, where elections take place, and that we will accordingly proceed by backward induction, starting in period $T+t-1$ to period $t+1$. As in Step 2, assuming that play has occurred according to the proposed equilibrium, we must have

$$j \leq \Delta.$$

We distinguish three cases.

Case A: $i < j (\leq \Delta)$

In this case, we have

$$H(i) = -(i-\mu)^2 - c \cdot (j-i) - \theta \cdot c \cdot (\Delta - i) - \chi'.$$

Then,

$$H'(i) = 2(\mu - i) + c \cdot (1+\theta) \geq 0,$$

where the last inequality holds because $c > 0$, $\theta \leq 1$, and

$$i \leq j \leq \Delta = \mu - \frac{c}{2} \cdot (1-\theta) \leq \mu.$$

That is,

$$H'(i) > 0 \text{ for all } i \in (-\infty, j).$$

Case B: $j \leq i \leq \Delta$

In this case, we have

$$H(i) = -(i-\mu)^2 - c \cdot (i-j) - \theta \cdot c \cdot (\Delta - i) - \chi'.$$

Then,

$$H'(i) = 2(\mu - i) - c \cdot (1+\theta) \geq 0,$$

where the last inequality holds because $c > 0$, $\theta \leq 1$, and

$$i \leq j \leq \Delta = \mu - \frac{c}{2} \cdot (1+\theta) \leq \mu - \frac{c}{2} \cdot (1-\theta).$$

That is,

$$H'(i) > 0 \text{ for all } i \in (j, \Delta).$$

Case C: $(j \leq)\Delta \leq i$

In this case, we have

$$H(i) = -(i-\mu)^2 - c \cdot (i-j) - \theta \cdot c \cdot (i - \Delta) - \chi'.$$

Then,

$$H'(i) = 2(\mu - i) - c \cdot (1+\theta) \leq 0,$$

where the last inequality holds because $c > 0$, $\theta \leq 1$, and

$$i \geq \Delta = \mu - \frac{c}{2} \cdot (1+\theta).$$

That is,

$$H'(i) < 0 \text{ for all } i \in (\Delta, +\infty).$$

To sum up, we have proved that the problem described in (29) for period h is maximized for $i = \Delta$, thereby completing the proof of existence of equilibrium. For the proof of the uniqueness of equilibrium, we refer to Appendix B.

□

Proof of Proposition 1:

Using Equation (11), we obtain for any $T, T' \in \mathbb{N}$ that

$$\frac{1}{2} \cdot \left([W^2 - W^1](T) - [W^2 - W^1](T')\right) = \mathbb{E}_T \left[\sum_{t \geq 1} \delta^{t-1} \cdot m_t i_t\right] - \mathbb{E}_{T'} \left[\sum_{t \geq 1} \delta^{t-1} \cdot m_t i_t\right]. \quad (30)$$

To compute each of the two terms of the right-hand side of Expression (30), we introduce further definitions. First, assume that party R is in power in the first period of *one* term that consists of T periods, say in some period $t+1$, with the median voter having peak at $m_t \in [-\beta, \beta]$ before the term starts, and then define

$$B_R(T|m_t) = \mathbb{E}_T\left[\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot m_{t'} i_{t'} \middle| m_t\right]. \tag{31}$$

Analogously, one can define

$$B_L(T|m_t) = \mathbb{E}_T\left[\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot m_{t'} i_{t'} \middle| m_t\right], \tag{32}$$

for the case where L is in office in period $t+1$. Note that

$$\mathbb{E}_T\left[\sum_{t \geq 1} \delta^{t-1} \cdot m_t i_t\right] = B_R(T|m_0). \tag{33}$$

Then, for any $m_t \in [-\beta, \beta]$, we have

$$\begin{aligned}
B_R(T|m_t) &= \Delta \cdot \sum_{t'=t+1}^{t+T} \delta^{t'-t-1} \cdot \left[\eta^{t'-t} \cdot m_t + (1-\eta^{t'-t}) \cdot \frac{1}{2\beta} \int_{-\beta}^{\beta} i di\right] \\
&+ (1-\rho)^T \cdot \delta^T \cdot \Bigg(\eta^T \cdot \left[\mathbb{1}_{m_t \geq -c/2} \cdot B_R(T|m_t) + (1 - \mathbb{1}_{m_t \geq -c/2}) \cdot B_L(T|m_t)\right] \\
&+ (1-\eta^T) \cdot \left[\frac{1}{\beta - \min\{c/2, \beta\}} \int_{-\beta}^{-\min\{c/2,\beta\}} B_L(T|i) di + \frac{1}{\beta + \min\{c/2, \beta\}} \int_{-\min\{c/2,\beta\}}^{\beta} B_R(T|i) di\right]\Bigg) \\
&+ (1-(1-\rho)^T) \cdot \delta^T \cdot \left(\eta^T \cdot B_L(T|m_t) + (1-\eta^T) \cdot \frac{1}{2\beta} \int_{-\beta}^{\beta} B_L(T|i) di\right),
\end{aligned} \tag{34}$$

where $\mathbb{1}_{m_t \geq -c/2} = 1$ if $m_t \geq -c/2$ and $\mathbb{1}_{m_t \geq -c/2} = 0$ otherwise. To understand the above expression, note that regardless of his/her peak, the median voter would like to remove the incumbent if the latter has suffered a valence shock during the term. If the incumbent has not suffered any shock, the median voter will still prefer to oust the incumbent if his/her own peak at the time of election, say in period $t+T$, is too far away from the incumbent's peak (and hence from the policy chosen by the incumbent). The latter can only happen in two scenarios, provided that $\beta > c/2$: first, if $m_t < -c/2$ and the peak has not changed during the entire term, which happens with probability η^T; second, if the peak has changed at least once during the term and is to the left of $-c/2$ at the time of election.

The following two remarks will substantially facilitate the analysis. First, assume without loss of generality that some candidate from party R is in his/her first period in office, say in some period

$t+1$, and let $m_t \in [0, \min\{\beta, c/2\}]$. Note that $-m_t \in [-\min\{\beta, c/2\}, 0]$. As already mentioned, this office-holder—and any subsequent one, similarly—will only be ousted from power if he/she has suffered a valence shock or if the median voter's peak is strictly below $-c/2$ (or above $c/2$ for candidates of party L). The latter can only happen if the median voter peak has changed. Given that valence shocks and changes in the median voter peak (when they happen) are stochastically independent of the previous peak, one can write

$$B_R(T|m_t) + B_R(T|-m_t) = \sum_{\mathcal{I}=(i_{t+1},i_{t+2},\ldots)} p_{\mathcal{I}} \cdot \sum_{t' \geq t+1} \delta^{t'-t-1} \cdot (m_{t'} - m_{t'}) \cdot i_{t'} = 0, \tag{35}$$

where $(p_{\mathcal{I}})_{\mathcal{I}}$ denotes the probability distribution of the stream of policies $\mathcal{I}$ from the perspective of period $t+1$, given that the median voter's peak is m_t or, equivalently, $-m_t$. Accordingly,

$$\int_{-\min\{\beta,c/2\}}^{\min\{\beta,c/2\}} B_R(T|m_t) dm_t = \int_{0}^{\min\{\beta,c/2\}} (B_R(T|m_t) + B_R(T|-m_t)) dm_t = 0. \tag{36}$$

Analogously, it holds by symmetry that

$$\int_{-\min\{\beta,c/2\}}^{\min\{\beta,c/2\}} B_L(T|m_t) dm_t = 0. \tag{37}$$

Also by symmetry, the probability of $\mathcal{I} = (i_{t+1}, i_{t+2}, \ldots)$ and $\mathcal{M} = (m_{t+1}, m_{t+2}, \ldots)$ given that party R is in power in period $t+1$ and that the median voter in such period has peak at m_t is equal to the probability of $-\mathcal{I} = (-i_{t+1}, i_{t+2}, -\ldots)$ and $-\mathcal{M} = (-m_{t+1}, -m_{t+2}, \ldots)$ given that party L is in power in period $t+1$ and that the median voter in such period has peak at $-m_t$. Then, for all $m_t \in [-\beta, \beta]$,

$$\begin{aligned} B_R(T|m_t) &= \sum_{\substack{\mathcal{I}=(i_{t+1},i_{t+2},\ldots),\\ \mathcal{M}=(m_{t+1},m_{t+2},\ldots)}} p_{\mathcal{I},\mathcal{M}} \cdot \sum_{t' \geq t+1} \delta^{t'-t-1} \cdot m_{t'} \cdot i_{t'} \\ &= \sum_{\substack{-\mathcal{I}=(-i_{t+1},-i_{t+2},\ldots),\\ -\mathcal{M}=(-m_{t+1},-m_{t+2},\ldots)}} p_{-\mathcal{I},-\mathcal{M}} \cdot \sum_{t' \geq t+1} \delta^{t'-t-1} \cdot (-m_{t'}) \cdot (-i_{t'}) = B_L(T|-m_t). \end{aligned} \tag{38}$$

As before, $(p_{\mathcal{I},\mathcal{M}})_{\mathcal{I},\mathcal{M}}$ $((p_{-\mathcal{I},-\mathcal{M}})_{-\mathcal{I},-\mathcal{M}})$ denotes the probability distribution of the streams of policies $\mathcal{I}$ and $\mathcal{M}$ (of policies $-\mathcal{I}$ and $-\mathcal{M}$) from the perspective of period $t+1$, given that the median voter is initially m_t $(-m_t)$ and party R (party L) is in power. For the remainder of the proof, it will be convenient to distinguish two cases.

Case I: $\beta \leq c/2$

In this case, the variance in the median voter's peak is small relative to half the marginal cost of change, and, in particular, $\mathbb{1}_{m_t \geq -c/2} = 1$ and $\min\{c/2, \beta\} = \beta$. Then, it follows from

Equation (34) that for all $m_t \in [-\beta, \beta]$,

$$
\begin{aligned}
&(1-(1-\rho)^T \cdot (\delta\eta)^T) \cdot B_R(T|m_t) \\
=&\Delta \cdot \eta \cdot \frac{1-(\delta\eta)^T}{1-\delta\eta} \cdot m_t + (1-(1-\rho)^T)(\delta\eta)^T \cdot B_L(T|m_t) + (\delta^T - (\delta\eta)^T) \cdot \frac{1}{2\beta}\int_{-\beta}^{\beta} B_L(T|i)di.
\end{aligned}
$$

Using Equations (35), (37) and (38), we obtain from the previous equation that

$$
B_R(T|m_t) = \frac{\Delta \cdot \eta}{1-\delta\eta} \cdot m_t.
$$

Note that the right-hand side of the above equation is independent of T. That is, if $\beta \leq c/2$,

$$
[W^2 - W^1](T) = [W^2 - W^1](T') \text{ , for all } T, T' \in \mathbb{N}.
$$

As a matter of fact, we have

$$
[W^2 - W^1](T) = 2 \cdot \int_{-\beta}^{\beta} B_R(T|m_t)dm_t = 0.
$$

Case II: $\beta > c/2$

In this case, the variance in the median voter's peak is large relative to half the marginal cost of change, and, in particular, $\min\{\beta, c/2\} = c/2$. By symmetry, it must be that

$$
\int_{-\beta}^{-c/2} B_L(T|i)di = \int_{c/2}^{\beta} B_R(T|i)di := K_{-1}, \tag{39}
$$

$$
\int_{-c/2}^{c/2} B_L(T|i)di = \int_{-c/2}^{c/2} B_R(T|i)di := K_0, \tag{40}
$$

$$
\int_{c/2}^{\beta} B_L(T|i)di = \int_{-c/2}^{-\beta} B_R(T|i)di := K_1. \tag{41}
$$

Note that Equation (36) implies that

$$
K_0 = 0. \tag{42}
$$

We distinguish three subcases.

Case II.A: $-c/2 \leq m_t \leq c/2$

By symmetry, it suffices to focus on party R. Then, it follows from Equation (34) that

$$
\begin{aligned}
&(1-(1-\rho)^T \cdot (\delta\eta)^T) \cdot B_R(T|m_t) \\
=&\Delta \cdot \eta \cdot \frac{1-(\delta\eta)^T}{1-\delta\eta} \cdot m_t + (1-(1-\rho)^T) \cdot (\delta\eta)^T \cdot B_L(T|m_t) \\
&+(1-(1-\rho)^T) \cdot \delta^T \cdot (1-\eta^T) \cdot \frac{1}{2\beta}\int_{-\beta}^{\beta} B_L(T|i)di \\
&+(1-\eta^T) \cdot (1-\rho)^T \cdot \delta^T \cdot \left(\frac{1}{\beta - c/2}\int_{-\beta}^{-c/2} B_L(T|i)di + \frac{1}{\beta + c/2}\int_{-c/2}^{\beta} B_R(T|i)di\right).
\end{aligned} \tag{43}
$$

Using Equation (42), we obtain (assuming $c > 0$) from integrating the above expression for $m_t \in [-c/2, c/2]$ that

$$\begin{aligned} 0 = (1-(\delta\eta)^T)\cdot K_0 =& c\cdot\Bigg[(1-(1-\rho)^T)\cdot\delta^T\cdot(1-\eta^T)\cdot\frac{1}{2\beta}\cdot(K_{-1}+K_0+K_1) \\ &+(1-\eta^T)\cdot(1-\rho)^T\cdot\delta^T\cdot\left(\frac{2\beta}{\beta^2-(c/2)^2}\cdot K_{-1}+\frac{1}{\beta+c/2}\cdot K_0.\right)\Bigg] \end{aligned} \tag{44}$$

Case II.B: $-\beta \le m_t \le -c/2$

On the one hand, it follows from Equation (34) that

$$\begin{aligned} B_R(T|m_t) =& \Delta\cdot\eta\cdot\frac{1-(\delta\eta)^T}{1-\delta\eta}\cdot m_t \\ &+(\delta\eta)^T\cdot B_L(T|m_t)+\Bigg[(1-(1-\rho)^T)\cdot\delta^T\cdot(1-\eta^T)\cdot\frac{1}{2\beta}\cdot(K_{-1}+K_0+K_1) \\ &+(1-\eta^T)\cdot(1-\rho)^T\cdot\delta^T\cdot\left(\frac{2\beta}{\beta^2-(c/2)^2}\cdot K_{-1}+\frac{1}{\beta+c/2}\cdot K_0\right)\Bigg]. \end{aligned}$$

Using Equation (44), the above equation reduces to (assuming $c > 0$)

$$B_R(T|m_t) = \Delta\cdot\eta\cdot\frac{1-(\delta\eta)^T}{1-\delta\eta}\cdot m_t+(\delta\eta)^T\cdot B_L(T|m_t). \tag{45}$$

On the other hand, one can verify that using the counterpart for party L of Equation (34), as well as Equation (44), we obtain (assuming $c > 0$) that[51]

$$\begin{aligned} &(1-(1-\rho)^T\cdot(\delta\eta)^T)\cdot B_L(T|m_t) \\ =&-\Delta\cdot\eta\cdot\frac{1-(\delta\eta)^T}{1-\delta\eta}\cdot m_t+(1-(1-\rho)^T)\cdot(\delta\eta)^T\cdot B_R(T|m_t). \end{aligned} \tag{46}$$

Case II.C: $c/2 \le m_t \le \beta$

This case is analogous to Case II.B. By symmetry, we can dispense with it.

Now, we solve for $B_R(T|m_t)$ and $B_L(T|m_t)$, assuming that $-\beta \le m_t \le c/2$. From Equations (45) and (46), one can find after some algebra that

$$B_L(T|m_t) = -\frac{1-(1-(1-\rho)^T)(\delta\eta)^T}{1+(1-(1-\rho)^T)(\delta\eta)^T}\cdot\frac{\eta}{1-\delta\eta}\cdot\left(\mu-\frac{c}{2(1+\theta)}\right)\cdot m_t$$

and

$$B_R(T|m_t) = \left(1-\frac{2(\delta\eta)^T}{1+(1-(1-\rho)^T)(\delta\eta)^T}\right)\cdot\frac{\eta}{1-\delta\eta}\cdot\left(\mu-\frac{c}{2(1+\theta)}\right)\cdot m_t.$$

[51]Alternatively, one can assume $c/2 \le m_t \le \beta$ and build on Equations (34) and (44), and then use the symmetry conditions between $B_R(\cdot)$ and $B_L(\cdot)$.

Finally, using the above two equations plus Equations (38) and (42), we have

$$B_R(T|m_t) = \begin{cases} \left(1 - \frac{2(\delta\eta)^T}{1+(1-(1-\rho)^T)(\delta\eta)^T}\right) \cdot \frac{\eta}{1-\delta\eta} \cdot \left(\mu - \frac{c}{2(1+\theta)}\right) \cdot m_t & \text{if } -\beta \le m_t < -c/2, \\ 0 & \text{if } -c/2 \le m_t \le c/2, \\ \frac{1-(1-(1-\rho)^T)(\delta\eta)^T}{1+(1-(1-\rho)^T)(\delta\eta)^T} \cdot \frac{\eta}{1-\delta\eta} \cdot \left(\mu - \frac{c}{2(1+\theta)}\right) \cdot m_t & \text{if } c/2 < m_t \le \beta. \end{cases}$$

Then, the result of the proposition follows if we note that

$$\begin{aligned} &[W^2 - W^1](T) = 2 \cdot \int_{-\beta}^{\beta} B_R(T|m_t) dm_t \\ =&2 \cdot \left(\beta^2 - \left(\min\left\{\beta, \frac{c}{2}\right\}\right)^2\right) \cdot \frac{(\delta\eta)^T(1-\rho)^T}{1+\delta^T\eta^T(1-(1-\rho)^T)} \cdot \frac{\eta}{1-\delta\eta} \cdot \left(\mu - \frac{c}{2(1+\theta)}\right). \end{aligned}$$

□

Proof of Proposition 2:

The notion of welfare considered is given by

$$W^1(T) = \underbrace{\mathbb{E}_T\left[-\sum_{t\ge1}\delta^{t-1}\cdot i_t^2\right]}_{EU^0(T)} + \underbrace{\mathbb{E}_T\left[-\sum_{t\ge1}\delta^{t-1}\cdot c\cdot|i_t - i_{t-1}|\right]}_{EU^c(T)} + \underbrace{\mathbb{E}_T\left[\sum_{t\ge1}\delta^{t-1}\cdot a_t\right]}_{EU^v(T)},$$

where the expected values are taken with respect to the stochastic processes that define $\mathcal{S}$, $\mathcal{M}$, $\mathcal{I}$, given the term length T. In the following, we analyze each of the welfare components separately as a function of the term length $T \in \mathbb{N}$. We recall that

$$-\Delta \le i_0 \le \Delta$$

and that we have assumed without loss of generality that the first office-holder belongs to party R, and hence $i_1 = \Delta$ is the first policy choice. Due to Theorem 2, in the subsequent periods policies will alternate between Δ and $-\Delta$, whenever they switch.

Analysis of $EU^0(T)$:

Given the symmetry in the equilibrium policy choices (Δ and $-\Delta$ are chosen alternatively), $EU^0(T)$ is independent of T.

Analysis of EU^c:

From Theorem 2, we know that every policy shift yields a welfare loss equal to $2c \cdot \Delta$. We further note that policy shifts occur if and only if the office-holder is ousted, since a given policy is persistent during the office-holder's tenure. Hence,

$$c \cdot |i_t - i_{t-1}| = \begin{cases} 2c \cdot \Delta & \text{if in period } t \text{ a new candidate was elected,} \\ 0 & \text{otherwise.} \end{cases}$$

We stress that a candidate who has just been elected (whether it is for the first term or not) has experienced no valence shock yet. To compute EU^c, it remains to investigate how often power will shift from one party to the other. This depends on two conditions at the moment of elections: first, on whether the office-holder has suffered a valence shock; second, on whether the median voter's peak is further away than $\mu + c/2$ from the office-holder's. Then, assuming that period $t+1$ is the first period of a new term (not necessarily the first term of the office-holder) and that the office-holder belongs to party R, define for $i \in \{-\Delta, \Delta\}$

$$\begin{aligned} K^{-1}(R,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, m_t < -c/2 \right], \\ K^{0}(R,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, -c/2 \leq m_t \leq c/2 \right], \\ K^{1}(R,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, m_t > c/2 \right]. \end{aligned}$$

Similarly, for the case where the office-holder in period $t+1$ belongs to party L, one can define

$$\begin{aligned} K^{-1}(L,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, m_t > c/2 \right], \\ K^{0}(L,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, -c/2 \leq m_t \leq c/2 \right], \\ K^{1}(L,i) :=& \mathbb{E}_T \left[-\sum_{t' \geq t+1} \delta^{t'-t-1} \cdot c \cdot |i_{t'} - i_{t'+1}| \middle| i_t = i, m_t < -c/2 \right]. \end{aligned}$$

On the one hand, because an incumbent will only change the policy if the incumbent in the period before the elections belonged to the other party and because there will be no differences in the equilibrium path thereafter, it must be that for $j \in \{-1, 0, 1\}$,

$$K^j(R, -\Delta) = K^j(R, \Delta) - 2c \cdot \Delta$$

and

$$K^j(L, \Delta) = K^j(L, -\Delta) - 2c \cdot \Delta.$$

On the other hand, from the symmetry between the two parties regarding their peaks and policy choices, as well as from the symmetry of the electorate decisions, it must be also be that for

$j \in \{-1, 0, 1\}$,

$$K^j := K^j(R, \Delta) = K^j(L, -\Delta),$$

in which case we also let

$$K^{-j} := K^j(R, -\Delta) = K^j(L, \Delta).$$

Then, define

$$P := \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\}$$

and

$$Q := \frac{1}{2\beta} \cdot \left(2\beta - 2\max\left\{0, \beta - \frac{c}{2}\right\}\right).$$

Note that $2P$ is the probability that a newly determined peak for the median voter is extreme (i.e., it is above $c/2$ or, alternatively, it is below $-c/2$), while Q is the probability that a newly determined peak for the median voter is moderate (i.e., it is in between $c/2$ and $-c/2$). Recall that the peak is drawn according to a uniform distribution on $[-\beta, \beta]$. In particular,

$$2P + Q = 1.$$

Then, using all the above equations and Theorem 2, three recursive equations must hold. First,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^1 = &(1 - (1-\rho)^T) \cdot (1 - \eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ &+ (1 - (1-\rho)^T) \cdot \eta^T \cdot \left[K^{-1} - 2c \cdot \Delta\right] \\ &+ (1-\rho)^T \cdot (1 - \eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ &+ (1-\rho)^T \cdot \eta^T \cdot K^1. \end{aligned} \tag{47}$$

To understand Equation (47), assume that a candidate from party R is in the first period of a term—and has not suffered any valence shock so far—and that the median voter's peak at the beginning of the term is above $c/2$. This corresponds to K^1. Then, the incumbent from party R will never be re-elected at the end of the term if s/he suffers a valence shock in at least one period of the term, which happens with probability $1 - (1-\rho)^T$. In this case, a candidate from party L will win office for the next term. If no shock has occurred, the incumbent will still be ousted if at the time of elections, the median voter's peak lies below $-c/2$. Regardless of the valence shocks, the evolution of the median voter is determined as follows: with probability η^T, it does not change during the term, and hence remains above $c/2$; with probability $1 - \eta^T$, it is drawn according to a uniform distribution on $[-\beta, \beta]$. Second,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^0 = &(1-(1-\rho)^T) \cdot (1-\eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ &+ (1-(1-\rho)^T) \cdot \eta^T \cdot \left[K^0 - 2c \cdot \Delta\right] \\ &+ (1-\rho)^T \cdot (1-\eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ &+ (1-\rho)^T \cdot \eta^T \cdot K^0. \end{aligned} \tag{48}$$

Third,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^{-1} = &(1-(1-\rho)^T) \cdot (1-\eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ &+ (1-(1-\rho)^T) \cdot \eta^T \cdot \left[K^1 - 2c \cdot \Delta\right] \\ &+ (1-\rho)^T \cdot (1-\eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ &+ (1-\rho)^T \cdot \eta^T \cdot [K^1 - 2c \cdot \Delta]. \end{aligned} \tag{49}$$

It is then a matter of algebra to verify that the linear system of equations made up of (47), (48) and (49) has a unique solution, which yields

$$\begin{aligned} EU^c(T) &= P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 \\ &= -2c\Delta \cdot \frac{\delta^T}{1-\delta^T} \cdot \left[(1-(1-\rho)^T) + (1-\rho)^T \cdot P \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}\right]. \end{aligned} \tag{50}$$

Analysis of $EU^v(T)$:

Assuming that any incumbent who has received at least one valence shock will be ousted, and given that valence shocks happen independently of any other variable in the model, and in particular independently of the office-holder's identity, we have

$$EU^v(T) = -A \cdot \rho \cdot \sum_{k=1}^{T} \sum_{j=k}^{T} \delta^{j-1} + \delta^T \cdot EU^v(T).$$

The first term in the above formula contains the δ-discounted (dis)utility from valence shocks in *one* term of length T periods. Then, one obtains that

$$EU^v(T) = -\frac{\rho A}{(1-\delta)^2} + \rho A \cdot T \cdot \frac{\delta^T}{1-\delta^T} \cdot \frac{1}{1-\delta}. \tag{51}$$

Note that the first term of the right-hand side of Equation (51) is independent of T.

After the analysis of $EU^c(T)$ and $EU^V(T)$—see Equations (50) and (51)—, maximizing welfare with respect to T is equivalent to finding the value of $T \in \mathbb{N}$, say T^*, that maximizes the following expression:

$$M(T) := \frac{\delta^T}{1-\delta^T} \cdot \Bigg[T \cdot \frac{A}{2c \cdot \left(\mu - \frac{c}{2(1+\theta)}\right)} \cdot \frac{\rho}{1-\delta} - (1-(1-\rho)^T)$$
$$- (1-\rho)^T \cdot \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\} \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}\Bigg]. \tag{52}$$

It is easy to verify that $\lim_{T\to\infty} M(T) = 0$ and that there is $T' > 0$ such that $M(T') > 0$. Hence, the optimal term length T^* exists and is finite. This completes the proof.

□

Proof of Corollaries 1, 2, and 3:

We recall that maximizing $W^1(T)$ is equivalent to maximizing

$$M(T) = H_I(T) + (1-(1-\rho)^T) \cdot H_{II}(T) + (1-\rho)^T \cdot H_{III}(T),$$

where

$$H_I(T) = \frac{\delta^T}{1-\delta^T} \cdot T \cdot \frac{A}{2c \cdot \left(\mu - \frac{c}{2} \cdot (1+\theta)\right)} \cdot \frac{\rho}{1-\delta},$$
$$H_{II}(T) = -\frac{\delta^T}{1-\delta^T},$$
$$H_{III}(T) = -\frac{\delta^T}{1-\delta^T} \cdot \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\} \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}.$$

We start noting that $H_I(T)$ is a positive, real-valued decreasing function, and that $H_{II}(T)$ and $H_{III}(T)$ are negative, real-valued increasing functions. In particular, $H_I(T)$ is maximized for $T = 1$, while $H_{II}(T)$ and $H_{III}(T)$ are maximized for $T = \infty$. We are now in a position to prove the different statements from the corollaries. First, increasing A or θ or decreasing μ gives more importance (i.e., assigns a higher weight in ex ante welfare $W^1(T)$) to $H_I(T)$ compared to $H_{II}(T)$ and $H_{III}(T)$. This implies that the optimal term length T^* can never increase if A or θ increase or μ decreases. Second, increasing β or reducing η gives more importance to $H_{III}(T)$ compared to $H_I(T)$, if at all, and leaves $H_{II}(T)$ invariant.[52] This implies that the optimal term length T^* can never decrease if β increases. In the remainder of the proof, we focus on changes of c and ρ.

On the one hand, consider a change of c. This affects $H_I(T)$ and potentially $H_{III}(T)$. Specifically,

[52]Note that $\frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}$ is decreasing in η.

note that

$$\frac{\partial}{\partial c}\left(c\cdot\left(\mu-\frac{c}{2}\cdot(1+\theta)\right)\right)=\begin{cases}>0 & \text{if } c<\frac{\mu}{1+\theta},\\ =0 & \text{if } c=\frac{\mu}{1+\theta},\\ <0 & \text{if } c>\frac{\mu}{1+\theta},\end{cases}$$

and

$$\frac{\partial}{\partial c}\left(\frac{1}{2\beta}\cdot\max\left\{0,\beta-\frac{c}{2}\right\}\right)=\begin{cases}>0 & \text{if } c<2\beta,\\ =0 & \text{if } c\geq 2\beta.\end{cases}$$

This shows that when $2\beta < c < \frac{\mu}{1+\theta}$, an increase of c yields a higher T^*. By contrast, when $2\beta < c < \frac{\mu}{1+\theta} < c < 2\beta$, an increase of c yields a lower T^*.

On the other hand, consider a change in ρ. This has the following effects: first, it increases the importance of $H_I(T)$ with slope one; second, it increases the importance of $H_{III}(T)$ (which is constant for a given T) by increasing the weight $(1-(1-\rho^T)$, with a slope that is lower than one;[53] third, it decreases the importance of $H_{II}(T)$ by decreasing $(1-\rho^T)$, but increases the term $H_{II}(T)$ itself with a slope lower than one.[54] Overall, an increase of ρ reduces T^*. Finally, we note than if we set and keep $\delta=\theta$, increasing θ increases the importance of $H_I(T)$ compared to $H_{II}(T)$ and $H_{III}(T)$, which calls for lower terms. This completes the proof of Corollaries 1, 2, and 3.

□

[53]Note that $\frac{\partial}{\partial\rho}(1-(1-\rho)^T)=(1-\rho)^T\cdot\ln\frac{1}{1-\rho}<(1-\rho)\cdot\ln\frac{1}{1-\rho}<1$.

[54]Note that $\frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}$ is increasing in ρ (and decreasing in δ).

Appendix B

It remains to show that there cannot be another equilibrium than the one described in Theorem 2. To this end, we proceed in three stages. For simplicity, we assume that candidates of party R choose policies to the right of 0 and candidates of party L choose policies to the left of 0. Because a valence shock decreases the appeal to all voters, we also assume that policy choices are non-increasing in the number of shocks suffered already by the incumbent. These two assumptions facilitate the analysis but can be dispensed with. As for notation, we use the same shortcuts as in the proof of Theorem 2.

Stage 1:

We start by considering the median voter's decision in the election that takes place in period $t+T$. Assume the incumbent $k \in R$ will choose some $\Delta^z > 0$ in period $t+T+1$ if s/he is re-elected and has not suffered a valence shock in period $t+T$, while s/he will choose some $\Delta^{z+1} > 0$ in period $t+T+1$ if s/he is re-elected and has suffered a valence shock in period $t+T$. By assumption, it must be $\Delta^{z+1} \leq \Delta^z$. In turn, assume that the challenger $k \in L$ will choose some $-\Delta^0 < 0$ in period $t+T+1$ if s/he is elected instead. We stress that at the time of elections in period $t+T$, the median voter knows whether or not the incumbent has suffered a shock. We use

$$p(i, z_+, m_-) \tag{53}$$

to denote the probability that the median voter will elect the incumbent $k \in R$ when the latter has chosen i and has suffered z_+ shocks, and before the median voter's peak m is determined according to $F(\cdot|m_-)$. We distinguish three cases.

<u>Case I:</u> $-\Delta^0 \leq i \leq \Delta^{z+1} (\leq \Delta^z)$

For Case I, we distinguish two subcases.

Case I.A: $a_{kT} = -z \cdot A$

In this case, the incumbent has suffered no valence shock in period $t+T$. Then, the median voter will re-elect k if and only if

$$-(m-\Delta^z)^2 - c \cdot (\Delta^z - i) - A \cdot z \geq -(m+\Delta^0)^2 - c \cdot \left(i + \Delta^0\right),$$

which can be rearranged as

$$m \geq \frac{c}{2} \cdot \frac{\Delta^z - \Delta^0 - 2i}{\Delta^z + \Delta^0} + \frac{A \cdot z}{2(\Delta^z + \Delta^0)} + \frac{\Delta^z - \Delta^0}{2}. \tag{54}$$

From the above expression, it follows that

$$p(i,z,m_-) = \int_{-\beta}^{\beta} \mathbb{1}_m \left(\frac{c}{2} \cdot \frac{\Delta^z - \Delta^0 - 2i}{\Delta^z + \Delta^0} + \frac{A \cdot z}{2(\Delta^z + \Delta^0)} + \frac{\Delta^z - \Delta^0}{2} \right) dF(m|m_-).$$

Hence, $p(i,z,m_-)$ is non-decreasing in i.

Case I.B: $a_{kT} = -(z+1) \cdot A$

In this case, the incumbent has suffered one valence shock in period $t+T$. Then, following the logic of Case I.A, one can verify that the median voter will re-elect k if and only if

$$m \geq \frac{c}{2} \cdot \frac{\Delta^{z+1} - \Delta^0 - 2i}{\Delta^{z+1} + \Delta^0} + \frac{A \cdot (z+1)}{2(\Delta^{z+1} + \Delta^0)} + \frac{\Delta^{z+1} - \Delta^0}{2}. \tag{55}$$

Using the above expression, one can see that $p(i, z+1, m_-)$ is non-decreasing in i.

<u>Case II:</u> $(\Delta^{z+1} \leq)\Delta^z \leq i$

For Case II, we distinguish two subcases.

Case II.A: $a_{kT} = -z \cdot A$

In this case, the incumbent has suffered no valence shock in period $t+T$. Then, the median voter will re-elect k if and only if

$$-(m - \Delta^z)^2 - c \cdot (i - \Delta^z) - A \cdot z \geq -(m + \Delta^0)^2 - c \cdot \left(i + \Delta^0\right),$$

which can be rearranged as

$$m \geq -\frac{c}{2} + \frac{Az}{2(\Delta^z + \Delta^0)} + \frac{\Delta^z - \Delta^0}{2}. \tag{56}$$

Using the above expression, one can see that $p(i,z,m_-)$ is constant in i. This means that choosing a policy in period $t+T$ does not increase the probability of re-election (and, hence, the expected benefits that come with office) if this policy is to the right of the policies that will be chosen in period $t+T+1$ by the same incumbent if s/he will be re-elected.

Case II.B: $a_{kT} = -(z+1) \cdot A$

In this case, the incumbent has suffered one valence shock in period $t+T$. Then, following the logic of Case II.A, one can verify that the median voter will re-elect k if and only if

$$m \geq -\frac{c}{2} + \frac{A(z+1)}{2(\Delta^{z+1} + \Delta^0)} + \frac{\Delta^{z+1} - \Delta^0}{2}. \tag{57}$$

Using the above expression, one can see that $p(i, z+1, m_-)$ is constant in i. As before, choosing a policy in period $t+T$ does not increase the probability of re-election (and, hence, the expected benefits that come with office) if this policy is to the right of the policies that will be chosen in period $t+T+1$ by the same incumbent if s/he will be re-elected.

<u>Case III:</u> $\Delta^{z+1} \leq i \leq \Delta^{z}$

On the one hand, if $a_{kT} = -z \cdot A$, the median voter will re-elect k if and only if (56) holds. On the other hand, if $a_{kT} = -(z+1) \cdot A$, the median voter will re-elect k if and only if (55) holds.

<u>Case IV:</u> $i < -\Delta^{0}$

For Case IV, we distinguish two subcases.

Case V.A: $a_{kT} = -z \cdot A$

In this case, the median voter will re-elect k if and only if

$$-(m-\Delta^{z})^2 - c \cdot (\Delta^{z} - i) - A \cdot z \geq -(m+\Delta^{0})^2 - c \cdot \left(-\Delta^{0} - i\right),$$

which can be rearranged as

$$m \geq \frac{c}{2} + \frac{Az}{2(\Delta^{z}+\Delta^{0})} + \frac{\Delta^{z}-\Delta^{0}}{2}. \tag{58}$$

Using the above expression, one can see that $p(i,z,m_{-})$ is constant in i.

Case V.B: $a_{kT} = -(z+1) \cdot A$

In this case, the median voter will re-elect k if and only if

$$m \geq \frac{c}{2} + \frac{A(z+1)}{2(\Delta^{z+1}+\Delta^{0})} + \frac{\Delta^{z+1}-\Delta^{0}}{2}. \tag{59}$$

Using the above expression, one can see that $p(i,z+1,m_{-})$ is constant in i.

For simplicity, we shall assume henceforth that $F(\cdot)$ is such that $p(i,z,m_{-})$ and $p(i,z+1,m_{-})$ are differentiable for all $i \in \mathbb{R}$ (note that they are continuous functions). If they were not, we could apply a limit argument to a sequence $\{F_n(\cdot)\}_{n\geq 1}$ of probability distributions guaranteeing that $p(i,z,m_{-})$ and $p(i,z+1,m_{-})$ are differentiable for all $i \in \mathbb{R}$ to obtain uniqueness of equilibrium (in the limit).

Step 2:

We next consider the problem faced by the incumbent $k \in R$ at the start of period $t+T$, before experiencing any valence shock, the median voter's peak will be determined, and elections will take place, all in the same period $t+T$. In this case, the incumbent faces the following problem:

$$\max_{i\in I} G(i) := \max_{i\in I} \Bigg\{ -(i-\mu)^2 - c\cdot|i-j| - \chi$$
$$+\theta\cdot(1-\rho)\cdot p(i,z,m_-)\cdot\Big[b-(\Delta^z-\mu)^2 - c\cdot|\Delta^z - i| - A\cdot z\Big]$$
$$+\theta\cdot\rho\cdot p(i,z+1,m_-)\cdot\Big[b-\left(\Delta^{z+1}-\mu\right)^2 - c\cdot\left|\Delta^{z+1}-i\right| - A\cdot(z+1)\Big]$$
$$+\theta\cdot(1-\rho)\cdot(1-p(i,z,m_-))\cdot\Big[-\left(\mu+\Delta^0\right)^2 - c\cdot\left|i+\Delta^0\right|\Big]$$
$$+\theta\cdot\rho\cdot(1-p(i,z+1,m_-))\cdot\Big[-\left(\mu+\Delta^0\right)^2 - c\cdot\left|i+\Delta^0\right|\Big]\Bigg\}$$

where χ is independent of i. Note that we can rearrange terms to obtain

$$\begin{aligned} G(i) = &-(i-\mu)^2 - c\cdot|i-j| - \theta\cdot c\cdot\left|i+\Delta^0\right| - \chi' \\ &+\theta\cdot(1-\rho)\cdot p(i,z,m_-)\cdot\Big[M^z - c\cdot|\Delta^z - i| + c\cdot\left|i+\Delta^0\right|\Big] \\ &+\theta\cdot\rho\cdot p(i,z+1,m_-)\cdot\Big[M^{z+1} - c\cdot\left|\Delta^{z+1}-i\right| + c\cdot\left|i+\Delta^0\right|\Big], \end{aligned} \tag{60}$$

where χ' is also independent of i,

$$M^z := b - A\cdot z + (\Delta^0 - \Delta^z + 2\mu)\cdot(\Delta^z + \Delta^0)$$

and

$$M^{z+1} := b - A\cdot(z+1) + (\Delta^0 - \Delta^{z+1} + 2\mu)\cdot(\Delta^{z+1} + \Delta^0).$$

Given that $b > 0$ is assumed to be very large, and so are M^z and M^{z+1}. Finally, we note that $G(i)$ is differentiable for all $i \in I$, except possibly in a finite number of points. We distinguish several cases, but we focus on $i \geq 0$ (see the above assumptions).

Case I: $i \leq j$

In this case, using (60), we have

$$\begin{aligned} G'(i) \geq\ & 2(\mu - i) + c\cdot(1-\theta) \\ &+\theta\cdot(1-\rho)\cdot\frac{\partial p(i,z,m_-)}{\partial i}\cdot\Big[M^z - c\cdot|\Delta^z - i| - c\cdot\left|\Delta^0 - i\right|\Big] \\ &+\theta\cdot\rho\cdot\frac{\partial p(i,z+1,m_-)}{\partial i}\cdot\Big[M^{z+1} - c\cdot\left|\Delta^{z+1}-i\right| - c\cdot\left|\Delta^0 - i\right|\Big] \\ &\geq 2(\mu - i) + c\cdot(1-\theta), \end{aligned}$$

where the first inequality can be shown by distinguishing cases depending on whether i is larger than $-\Delta^0$, Δ^z and Δ^{z+1}, and the second inequality holds because $p\left(i, z+1, m_-\right)$ and $p\left(i, z+1, m_-\right)$ are non-decreasing and M^z and M^{z+1} are very large. To sum up,

$$i < \mu + \frac{c}{2} \cdot (1-\theta) \Rightarrow G'(i) > 0. \tag{61}$$

Case II: $(-\Delta^0 \leq) j \leq i \leq \Delta^{z+1} (\leq \Delta^z)$

In this case, we can write

$$\begin{aligned}
G(i) = & -(i-\mu)^2 - c \cdot (i-j) - \theta \cdot c \cdot \left(i + \Delta^0\right) - \chi' \\
& + \theta \cdot (1-\rho) \cdot p\left(i, z, m_-\right) \cdot \left[M^z - c \cdot \left(\Delta^z - i\right) + c \cdot \left(i + \Delta^0\right)\right] \\
& + \theta \cdot \rho \cdot p\left(i, z+1, m_-\right) \cdot \left[M^{z+1} - c \cdot \left(\Delta^{z+1} - i\right) + c \cdot \left(i + \Delta^0\right)\right] \\
= & -(i-\mu)^2 - c \cdot (i-j) - \theta \cdot c \cdot \left(i + \Delta^0\right) - \chi' \\
& + \theta \cdot (1-\rho) \cdot p\left(i, z, m_-\right) \cdot \left[M^z + c \cdot \left(\Delta^z - \Delta^0\right) + 2c \cdot i\right] \\
& + \theta \cdot \rho \cdot p\left(i, z+1, m_-\right) \cdot \left[M^{z+1} + c \cdot \left(\Delta^{z+1} - \Delta^0\right) + 2c \cdot i\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
G'(i) \geq & \, 2(\mu - i) - c \cdot \left[1 + \theta \cdot \left(1 - 2 \cdot \left((1-\rho) \cdot p\left(i, z, m_-\right) + \rho \cdot p\left(i, z+1, m_-\right)\right)\right)\right] \\
& + \theta \cdot (1-\rho) \cdot \frac{\partial p\left(i, z, m_-\right)}{\partial i} \cdot \left[M^z - c \cdot \left(\Delta^z - i\right) - c \cdot \left(i - \Delta^0\right)\right] \\
& + \theta \cdot \rho \cdot \frac{\partial p\left(i, z+1, m_-\right)}{\partial i} \cdot \left[M^{z+1} - c \cdot \left(\Delta^{z+1} - i\right) - c \cdot \left(\Delta^0 - i\right)\right] \\
\geq & \, 2 \cdot \left[\mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot \left((1-\rho) \cdot p\left(i, z, m_-\right) + \rho \cdot p\left(i, z+1, m_-\right)\right) - i\right].
\end{aligned}$$

Therefore,

$$i < \mu - \frac{c}{2} \cdot (1+\theta) \Rightarrow G'(i) > 0. \tag{62}$$

Case III: $j \leq i$ and $\Delta^{z+1} \leq i \leq \Delta^z$

In this case, we can write

$$\begin{aligned}
G(i) = &-(i-\mu)^2 - c\cdot(i-j) - \theta\cdot c\cdot\left(i+\Delta^0\right) - \chi' \\
&+\theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\left[M^z - c\cdot\left(\Delta^z - i\right) + c\cdot\left(i+\Delta^0\right)\right] \\
&+\theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\left[M^{z+1} - c\cdot\left(i-\Delta^{z+1}\right) + c\cdot\left(i+\Delta^0\right)\right] \\
= &-(i-\mu)^2 - c\cdot(i-j) - \theta\cdot c\cdot\left(i+\Delta^0\right) - \chi' \\
&+\theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\left[M^z + c\cdot\left(\Delta^z - \Delta^0\right) + 2c\cdot i\right] \\
&+\theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\left[M^{z+1} + c\cdot\left(\Delta^{z+1} + \Delta^0\right)\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
G'(i) \geq\ & 2(\mu - i) - c\cdot\left[1+\theta\cdot\left(1-2\cdot\left(\left(1-\rho\right)\cdot p\left(i,z,m_-\right) + \rho\cdot p\left(i,z+1,m_-\right)\right)\right)\right] \\
&+\theta\cdot(1-\rho)\cdot\frac{\partial p\left(i,z,m_-\right)}{\partial i}\cdot\left[M^z - c\cdot\left(\Delta^z - i\right) - c\cdot\left(i-\Delta^0\right)\right] \\
\geq\ & 2\cdot\left[\mu - \frac{c}{2}\cdot(1+\theta) + c\theta\cdot\left(\left(1-\rho\right)\cdot p\left(i,z,m_-\right)\right) - i\right].
\end{aligned}$$

Therefore,

$$i < \mu - \frac{c}{2}\cdot(1+\theta) \Rightarrow G'(i) > 0. \tag{63}$$

Case IV: $(\Delta^{z+1} \leq)\Delta^z \leq i$

In this case, we can write

$$\begin{aligned}
G(i) = &-(i-\mu)^2 - c\cdot(i-j) - \theta\cdot c\cdot\left(i+\Delta^0\right) - \chi' \\
&+\theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\left[M^z - c\cdot\left(i - \Delta^z\right) + c\cdot\left(i+\Delta^0\right)\right] \\
&+\theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\left[M^{z+1} - c\cdot\left(i-\Delta^{z+1}\right) + c\cdot\left(i+\Delta^0\right)\right] \\
= &-(i-\mu)^2 - c\cdot(i-j) - \theta\cdot c\cdot\left(i+\Delta^0\right) - \chi' \\
&+\theta\cdot(1-\rho)\cdot p\left(i,z,m_-\right)\cdot\left[M^z + c\cdot\left(\Delta^z + \Delta^0\right)\right] \\
&+\theta\cdot\rho\cdot p\left(i,z+1,m_-\right)\cdot\left[M^{z+1} + c\cdot\left(\Delta^{z+1} + \Delta^0\right)\right].
\end{aligned}$$

Hence,

$$G'(i) = 2(\mu - i) - c\cdot(1+\theta).$$

Therefore,

$$\begin{aligned} i > \mu - \frac{c}{2} \cdot (1+\theta) \Leftrightarrow G'(i) < 0, \\ i < \mu - \frac{c}{2} \cdot (1+\theta) \Leftrightarrow G'(i) > 0. \end{aligned} \tag{64}$$

Step 3:

Finally, we consider the case where the incumbent k has at least two periods ahead of him/her in a term that starts in period $t+1$ and ends in period $t+T$. For this to be possible, it must be that $T > 1$. Then, the incumbent faces the following problem in a particular period $h = t+1, \dots, t+T-1$:

$$\max_{i \in I} H(i) := \max_{i \in I} \left\{ -(i-\mu)^2 - c \cdot |j-i| - \theta c \cdot \Big[(1-\rho) \cdot |\Delta^z - i| + \rho \cdot |\Delta^{z+1} - i| \Big] - \chi' \right\}, \tag{65}$$

where χ' is independent of i. We assume that in period $h+1$, the incumbent will choose Δ^z if s/he has not received a valence shock in period h (which happens with probability $1-\rho$) or Δ^{z+1} if s/he has received a valence shock in period h (which happens with probability ρ). We note that $H(i)$ is a continuous function. We distinguish several cases, but we focus on $i \geq 0$ (see the above assumptions).

Case A: $i \leq j$

In this case, we have

$$H(i) = -(i-\mu)^2 - c \cdot (j-i) - \theta c \cdot \Big[(1-\rho) \cdot |\Delta^z - i| + \rho \cdot |\Delta^{z+1} - i| \Big] - \chi'.$$

It follows that

$$H'(i) \geq 2(\mu - i) + c \cdot (1-\theta).$$

Hence,

$$i < \mu + \frac{c}{2} \cdot (1-\theta) \Rightarrow H'(i) > 0. \tag{66}$$

Case B: $j \leq i \leq \Delta^{z+1} (\leq \Delta^z)$

In this case, we have

$$H(i) = -(i-\mu)^2 - c \cdot (i-j) - \theta c \cdot \Big[(1-\rho) \cdot (\Delta^z - i) + \rho \cdot (\Delta^{z+1} - i) \Big] - \chi'.$$

It follows that

$$H'(i) = 2(\mu - i) - c \cdot (1-\theta).$$

Hence,

$$
\begin{aligned}
i < \mu - \frac{c}{2} \cdot (1-\theta) \Leftrightarrow H'(i) > 0, \\
i > \mu - \frac{c}{2} \cdot (1-\theta) \Leftrightarrow H'(i) < 0.
\end{aligned}
\tag{67}
$$

Case C: $j \leq i$ *and* $\Delta^{z+1} \leq i \leq \Delta^z$

In this case, we have

$$
H(i) = -(i-\mu)^2 - c \cdot (i-j) - \theta c \cdot \Big[(1-\rho) \cdot (\Delta^z - i) + \rho \cdot (i - \Delta^{z+1})\Big] - \chi'.
$$

It follows that

$$
H'(i) = 2 \cdot \left[\mu - \frac{c}{2} \cdot \big(1 + \theta \cdot (2\rho - 1)\big) - i\right].
$$

Hence,

$$
\begin{aligned}
i < \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] \Rightarrow H'(i) > 0, \\
i > \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] \Rightarrow H'(i) < 0.
\end{aligned}
\tag{68}
$$

Case D: $j \leq i$ *and* $(\Delta^{z+1} \leq)\Delta^z \leq i$

In this case, we have

$$
H(i) = -(i-\mu)^2 - c \cdot (i-j) - \theta c \cdot \Big[(1-\rho) \cdot (i - \Delta^z) + \rho \cdot (i - \Delta^{z+1})\Big] - \chi'.
$$

It follows that

$$
H'(i) = 2(\mu - i) - c \cdot (1+\theta).
$$

Hence,

$$
\begin{aligned}
i > \mu - \frac{c}{2} \cdot (1+\theta) \Leftrightarrow H'(i) < 0, \\
i < \mu - \frac{c}{2} \cdot (1+\theta) \Leftrightarrow H'(i) > 0.
\end{aligned}
\tag{69}
$$

Finally, let z, m_- and j be such that when i is chosen according to the (additional) equilibrium, we have $i \neq \Delta = \mu - \frac{c}{2} \cdot (1+\theta)$. We distinguish two cases.

Case A: $i < \Delta$

In turn, we distinguish two cases, depending on the period $h \in \{t+1, \ldots, t+T\}$ considered.

Case A.1: $h/T = T/T$

In this case, from the analysis of Step 2, it follows that i cannot be the best response. First, if $i \leq j$, then (61)—see Case I—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Second, if $j \leq i \leq \Delta^{z+1}$, then (62)—see Case II—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Third, if $\Delta^{z+1} \leq i \leq \Delta^{z}$, then (63)—see Case III—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Fourth and last, if $\Delta^{z} \leq i$, then (64)—see Case IV—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$.

Case A.2: $h/T < T/T$

In this case, from the analysis of Step 3, it follows that i cannot be the best response. First, if $i \leq j$, then (66)—see Case A—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Second, if $j \leq i \leq \Delta^{z+1}$, then (67)—see Case B—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Third, if $\Delta^{z+1} \leq i \leq \Delta^{z}$, then (68)—see Case C—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$. Fourth and last, if $\Delta^{z} \leq i$, then (69)—see Case D—shows that increasing i slightly increases the expected utility of the incumbent $k \in R$.

Case B: $i > \Delta$

In turn, we distinguish two cases, depending on the period $h \in \{t+1, \ldots, t+T\}$ considered.

Case B.1: $h/T = T/T$

In this case, from the analysis of Steps 2 and 3, it follows that i cannot be the best response. First, if $i \leq j$, then (61) shows that increasing i slightly increases the expected utility of the incumbent $k \in R$, unless (possibly)

$$j \geq i \geq \mu + \frac{c}{2} \cdot (1-\theta) > \Delta.$$

However, from (61)—see Case I of Step 2—and from (66)—see Case A of Step 3—, it follows that j could not have been chosen optimally by the incumbent $k \in R$ (nor of an incumbent of party L), since slightly decreasing j would have increased k's utility. Second, assume that $j \leq i < \Delta^{z}$. Then, taking $i' = \Delta^{z}$ and $j' = i$, we have

$$\Delta \leq j' < i', \tag{70}$$

and hence we are back at Case B.1 (if $T = 1$) or Case B.2 (if $T > 1$). In the latter case ($T > 1$), we will reach a contradiction (see below). Now, if $T = 1$, one can see that for both the problem of choosing i and i', the maximizer of incumbent k's, with $k \in R$, does not depend on the status-quo policy. Hence, it must be that

$$i' = i.$$

This yields a contradiction with (70). Third and last, if $\Delta^z \leq i$, then (64) shows that decreasing i slightly increases the expected utility of the incumbent $k \in R$.

Case B.2: $h/T < T/T$

In this case, from the analysis of Steps 2 and 3, it follows that i cannot be the best response. First, if $i \leq j$, then (66) shows that increasing i slightly increases the expected utility of the incumbent $k \in R$, unless (possibly)

$$j \geq i \geq \mu + \frac{c}{2} \cdot (1 - \theta) > \Delta. \tag{71}$$

However, from (61)—see Case I of Step 2—and from (66)—see Case A of Step 3—, it follows that j could not have been chosen optimally by the incumbent $k \in R$ (nor by an incumbent of party L), since slightly decreasing j would have increased k's utility. Second, if $j \leq i \leq \Delta^{z+1}$, then (67) shows that increasing i slightly decreases expected utility of the incumbent $k \in R$. Third, if $j \leq i$ and $\Delta^{z+1} \leq i \leq \Delta^z$, then (68) shows that decreasing i slightly increases the expected utility of the incumbent $k \in R$, unless (possibly)

$$i = \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] \geq \Delta^{z+1}. \tag{72}$$

Now, taking $i' = \Delta^{z+1}$ and $j' = i$, which satisfy $i' \leq j'$, it follows from (61)—see Case I of Step 2—and from (66)—see Case A of Step 3—that for i' to be optimally chosen by the incumbent $k \in R$ given the status-quo policy j', it must be that

$$i' \geq \mu + \frac{c}{2} \cdot (1 - \theta).$$

However, this leads to the following contradiction:

$$i' \geq \mu + \frac{c}{2} \cdot (1 - \theta) > \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] = j' \geq i'.$$

Fourth and last, if $\Delta^z \leq i$, then (69) shows that decreasing i slightly increases the expected utility of the incumbent $k \in R$.