DISCUSSION PAPER SERIES

DP14857

THE OPTIMAL LENGTH OF POLITICAL TERMS

Hans Gersbach, Matthew O. Jackson and Oriol Tejada

PUBLIC ECONOMICS

ISSN 0265-8003

THE OPTIMAL LENGTH OF POLITICAL TERMS

Hans Gersbach, Matthew O. Jackson and Oriol Tejada

Discussion Paper DP14857
Published 09 June 2020
Submitted 03 June 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE OPTIMAL LENGTH OF POLITICAL TERMS

Abstract

We analyze the optimal length of political terms (equivalently, the optimal frequency with which elections should be held) when the candidates of two polarized parties compete for office and the median voter shifts over time. Office-holders determine policy and experience persistent random shocks to their valence. Policy changes are costly for citizens and politicians. Optimal term-length balances the frequency of costly policy changes when parties change office with the incumbent's average valence during tenure. We find that optimal term-length increases with party polarization, with the degree to which the median voter cares about valence, and with the frequency and the size of swings in the electorate. In contrast, optimal term-length decreases when candidates for office undergo less scrutiny or when parties care more about future outcomes. Finally, with small swings in the electorate and large polarization, optimal term-length increases if checks and balances increase.

JEL Classification: C72, C73, D72, D78

Keywords: Elections, term-length, costs of change, Polarization

Hans Gersbach - hgersbach@ethz.ch
ETH Zurich and CEPR

Matthew O. Jackson - jacksonm@stanford.edu
Stanford University and Santa Fe Institute

Oriol Tejada - toriol@ethz.ch
ETH Zurich

Acknowledgements
We would like to thank Alessandra Casella, John Duggan, Daniel Houser, Tobias Kästli, 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, the 2020 ASSA Meeting held in San Diego, and a seminar at George Mason University for comments that helped to improve the paper.

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: June 2020

Abstract

We analyze the optimal length of political terms (equivalently, the optimal frequency with which elections should be held) when the candidates of two polarized parties compete for office and the median voter shifts over time. Office-holders determine policy and experience persistent random shocks to their valence. Policy changes are costly for citizens and politicians. Optimal term-length balances the frequency of costly policy changes when parties change office with the incumbent's average valence during tenure. We find that optimal term-length increases with party polarization, with the degree to which the median voter cares about valence, and with the frequency and the size of swings in the electorate. In contrast, optimal term-length decreases when candidates for office undergo less scrutiny or when parties care more about future outcomes. Finally, with small swings in the electorate and large polarization, optimal term-length increases if checks and balances increase.

Keywords: elections; term-length; costs of change; polarization

JEL Classification: C72, C73, D72, D78

*We would like to thank Alessandra Casella, John Duggan, Daniel Houser, Tobias Kästli, 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, the 2020 ASSA Meeting held in San Diego, and a seminar at George Mason University for comments that helped to improve the paper. All mistakes are our own.

†Matthew Jackson is an external faculty member of the Santa Fe Institute.

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

Periodic elections are a cornerstone of representative democracy. Citizens not only delegate policy-making, but they also have the power to oust elected officials who are deemed to be of low quality. Yet this general principle—*direct rule by the people at certain time intervals*—does not stipulate *how frequently* the people should vote. To determine frequency, many factors have to be taken into account, as acknowledged in the above quote. 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 vary between two years (House of Representatives), four years (President), six years (Senate), and lifetime (Supreme Court). 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, while there has been an increase in 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 addresses the frequency of elections in a systematic and meaningful way? What are the determinants for the optimal design of this key institutional feature?

The model

We address the above questions by analyzing a model in which policy- and office-motivated candidates compete for office over infinitely many periods, grouped in terms. A period is the minimal expanse of time that enables the incumbent to change *one* policy. If they are elected, candidates pursue policies that are in line with one of two polarized parties, which have standard

[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 September 2019.

quadratic utilities over a one-dimensional policy space.[3] As a first main feature, policy changes from one period to another are costly. The *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, and the bid to achieve Catalan independence.

Costs of change create a dynamic link between policy choices that give an electoral advantage to incumbents over challengers. This is not a universal property of costs of change, but is a property of costs that are relatively more convex than the (dis)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. Valence is the way politicians are perceived by the electorate over and above policy concerns, and has to do with their competence in executing governmental duties such as providing public goods efficiently and/or 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 probability in every period that the office-holder will experience a negative and persistent shock to his/her valence, so that expected valence decreases with the number of periods in office. This shock is publicly observed, and it *may* be valued by the average and the median citizens with differing relative intensities.[5]

Finally, as a third main feature, the median voter's preferences shift over time, prompting on occasion a power shift from one party to the other to adjust policy to changes 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

[3] Although our model is most appropriate for an executive office, it can 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. Our model can also be applied to constitutional/supreme courts if we assume that the median voter reflects the executive's/legislative's preferences and elections enable the latter to appoint a pivotal member of the judiciary body.

[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.

[5] In relative terms, a median voter who can influence the level of public goods to be provided by the incumbent, for instance, will suffer more than the average voter from the incumbent's relative inability 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). We provide a micro-foundation for this assumption in Section 7.8.

also any underlying trends in voter preferences.[6] In contrast, by way of normalization, party positions are fixed. This is 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 base—to keep up with the changes in the electorate. While the (exogenous) stochastic process determining valence shocks provides a downward drift that unambiguously harms the incumbent's electoral prospects, the stochastic process determining the median voter's peak generates variance around the downward drift, which can either benefit the incumbent or harm him/her.[7]

The optimal length of political terms

The model generates the following insights:[8] If there were no costs of change, then from the perspective of a voter whose peak is 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. These would be symmetrically located with respect to the equidistant voter's peak, who would then be indifferent between the two.[9] Hence, shortening a term would invariably yield utility gains since office-holders could be ousted as soon as their valence were lower than the challenger's.

The picture changes substantially with costs of change, which introduce trade-offs. 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 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 on different sides of the policy spectrum, the median voter's utility from policies would not be affected by the power shift. Costs of change would then be the decisive factor for determining election outcomes. Absent any valence shocks and changes in the electorate's preferences, it would thus be optimal

[6] The evolution of the median voter's 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, the second dealing with social preferences—enables us to obtain valuable insights regarding policies, elections, and optimal term-length.

[8] Our model further builds on the (technical) assumptions that politicians and voters use Markov strategies, are present-biased, and cannot commit to particular policies before elections.

[9] 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.

from the perspective of a voter with a peak 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 on fundamentals—say, policies—are stable. There is abundant empirical research that bears this out (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 we introduce valence shocks and changes in the median voter's peak, and thus swings in the electorate? 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.[10] This is because it makes for a more frequent change of parties and thus costly policy changes. While such costs affect all voters equally, the median voter (in any particular period) will internalize neither the policy preferences of the entire electorate nor the weight that the incumbent's valence has on welfare. On occasion, this will prompt the median voter to induce a power shift between parties, even if doing so is not socially optimal. In our baseline model, this is possible thanks to exogenous changes to valence (of the incumbent) and preferences (of the median voter). But we also discuss how these changes can be endogenized. In other words, although having the possibility to oust the incumbent in elections would be good for the current median voter, the repeated changes in policy and associated costs can be detrimental to the population's overall welfare over time. In fact, we show that 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 own hands.[11]

Varying term-length and empirical implications

In general, the optimal length of the political term is thus determined by the trade-off between the costs of changing 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

[10]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.

[11]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 term-length.

caused by persistent shocks. This trade-off depends on the primitives of the model, and our analysis reveals comparative statics. In particular, we find that optimal term-length increases with party polarization and with the degree to which the median voter cares about the incumbent's valence in comparison with 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, 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' discount factor. The dependence of optimal term-length on the marginal cost of change is ambiguous. When fluctuations of the median voter's peak are small and polarization is large, optimal term-length will decrease if the marginal cost of change decreases. In contrast, when fluctuations of the median vote's peak are large and polarization is small, optimal term-length decreases as the marginal cost of change increases.

Our comparative statics yield a variety of predictions about when political terms should be longer or shorter, thus extending the scope for empirical study. Specifically, our model rationalizes longer terms when there is high polarization and/or social instability. For instance, in turbulent periods 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 (thus having less uncertainty about their valence). 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 for any length of time within their party organization. In such cases, the higher risk of negative valence shocks or the possibility that such costs are very large despite not being very likely calls for shorter terms. Finally, checks and balances may be one of many reasons for differences in costs of change. We then provide a rationale for a connection between term-length and checks and balances. Increasing (decreasing) levels of checks and balances should accord with longer (shorter) terms, provided that social preferences are relatively stable (unstable) and polarization between parties is large (small).[12]

Further insights

An important feature of the model that ensures mathematical tractability is the planning horizon of politicians and citizens, which we assume to encompass the current period and the next one. Limited farsightedness on the part of politicians and citizens suffices to ensure that the median

[12] Beck et al. (2001) has an index of checks and balances that could be used 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.

voter's behavior will determine policy decisions and vice versa.[13]

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' peaks. We refer to this way of measuring how socially desirable a term length is as *interim welfare*. As it turns out, this second welfare notion coincides with the ex-ante welfare measure when variance in the median voter's fluctuation is low. In cases where the two notions do *not* coincide, the second welfare notion calls for shorter terms, though not necessarily for one-period terms.

Numerous extensions and ramifications of our analysis are imaginable, and we examine some of them in this paper. They include the possibility that incumbents' valence may increase over time—say, through learning by doing—, that elections are entered by third parties or candidates whose preferences coincide with the current median voter's peak, that incumbents have more than one policy dimension at their disposal, or that the incumbent has the option to call early elections. We also investigate the effect on optimal term-length exerted by 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—vary across periods. Finally, we provide a micro-foundation for the assumption that in relative terms the median voter may suffer more than the average voter from an incumbent's low valence, and we discuss how to endogenize the random processes 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 case in which policy changes are costly. In Section 6 we discuss optimal design of political term-length and the empirical relevance of our findings. Some extensions of our baseline setup are discussed in Section 7. Section 8 concludes. Appendices A and B contain the proofs (see supplementary material).

[13]Although our political game typically has more than one equilibrium, these are all similar and yield the same comparative statics. This issue is discussed in Section 5 and analyzed in Appendix B.

2 Relation to the Literature

Our paper is related to various strands of the 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 points out that shorter terms favor the accountability of office-holders and the screening out of bad politicians.[14] By contrast, when information between the office-holder and the electorate is asymmetric, 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 explicitly characterize optimal term length from a welfare perspective, and then to obtain an array of interesting results on comparative statics. As a consequence, our analysis provides potential rationales for the observed variation 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 key to our model. There are a few papers that contain models where policies are difficult to change. 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. Gersbach and Tejada (2018) show that incumbents who are more efficient than the challenger in implementing costly reforms choose extreme policies to obtain an electoral advantage, while Gersbach et al. (2019) show that more political instability moderates policy choices. Finally, Gersbach et al. (2020) 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.[15] From the perspective of this strand of literature, the main

[14]Schultz (2008) considers a model that is very different from ours and analyzes aspects of elections and policies that are not central in our analysis.

[15]It suffices that the degree of convexity of the costs of change is lower than the degree of citizen utility losses 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. We refer to Hwang and Möllerström (2017) for an analysis of unidirectional reforms in the presence of (heterogenous) present-biased voters. There is a vast literature by political economists on time-inconsistent preferences (see e.g. Bisin et al., 2015; Jackson and Yariv, 2015; Piguillem and Riboni, 2015; Lizzeri and Yariv, 2017). From a technical perspective, the nature of the strategies considered by Gersbach et al. (2020) is different from ours. This leads to different off-equilibrium threats and allows Gersbach et al. (2020) to obtain a unique equilibrium under relatively general specifications.

contribution of our paper is to explain how among many other factors that are usually assumed to influence policy and elections optimal term-length is related to costs of change.[16]

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 add some insights on 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 of the incumbency advantage and it endogenizes term-lengths *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 making way for a new politician who might change policy. This bears some resemblance to the effect of shorter terms, which also enable more frequent (costly) policy changes.

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, so 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 necessary for an office-holder to make one policy choice.[17]

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

[16]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; Austen-Smith et al., 2019).

[17]The time span necessary for an office-holder to make one policy choice is fixed throughout. This is made for simplicity and our insights carry over when the time between two different policy decisions is variable.

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, and parties 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 party 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 effect on optimal term-length, we show in Section 6.3 that there are two cases depending on the asymmetry between both parties' peaks relative to zero. If asymmetry is low, our analysis remains intact. If asymmetry is large, then all else being equal, terms should be shorter. We also proceed for simplicity on 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 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 have to engage in political fights and wars of attrition that affect their ability to undertake policy reforms. Finally, a politician may be involved in some scandal that permanently erodes valence. Once it has occurred, the politician cannot 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 is affected by a permanent negative shock to his/her valence.[18] 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.[19] This valence is normalized to zero.

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

[19]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.

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 both the current period and 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. *Costs of change* have been discussed in detail in the Introduction.

Second, a politician k is a citizen, so 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 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, 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.[20] Because we proceed on the assumption that candidates execute party orders, for each candidate k of party K we henceforth write $V_K(\cdot,\cdot,\cdot) = V_k(\cdot,\cdot,\cdot)$, with V_K denoting the party K's utility.[21] In our citizen-candidate setup, 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 shown in Figure 1. Note that i_0, the initial policy before period $t = 1$, is a parameter of the model and is thus exogenously given.

[20] 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.

[21] This means that office rents are the same for the party and the politician. Assuming this is immaterial for our analysis, with the only exception of Section 7.3.

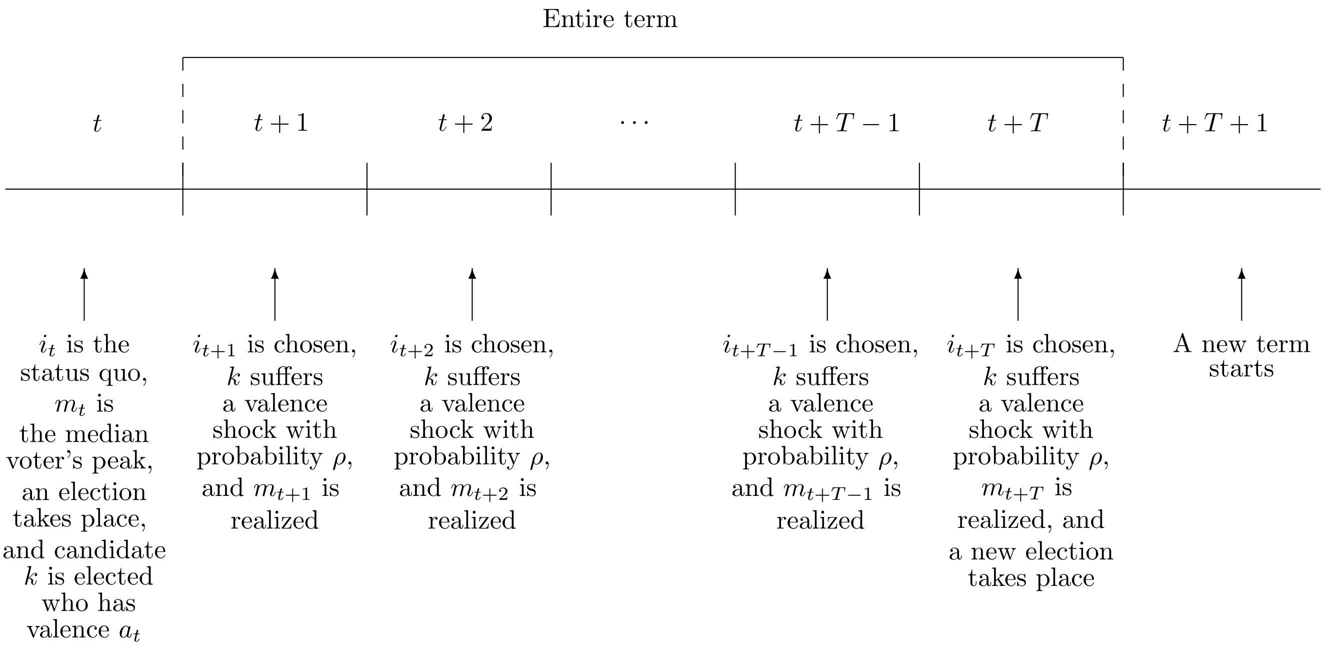

Figure 1: Timeline of the political game, with elections taking place in periods t and $t + T$.

3.4 Equilibrium notion

We look for sequential equilibria with the following refinements: First, we only consider stationary Markov strategies, so voters (the median voter in particular) 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 elapsed.[22] We recall that candidates cannot commit to policies prior to elections, so campaigns are irrelevant. This implies that opposition parties take no payoff-relevant action and are thus not modeled explicitly. Second, citizens (the median voter in particular) vote for the candidate from whom they expect higher 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 the case of indifference. Since office-holders simply comply with the policy objective of their party, we do not explicitly model candidates. Hence, besides the median voter, we consider the two parties, namely L and R, to be the two other players in the political game $\mathcal{G}$. Third, voters and parties care about the current period and one period ahead, which they discount with a common factor θ, with $0 < \theta \leq 1$.[23]

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

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

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 $t \in \mathbb{N}$,

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

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

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

where K denotes the incumbent's party in period t, $-K$ denotes the challenger's party, 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.[24]

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 look one period ahead. One reason for proceeding on this assumption is mathematical tractability.[25] Even so, this assumption suffices to ensure that, regardless of the length of a term, policy choices in 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 are 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.

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

[25] Assuming that agents are present-biased is mainly technical in nature, and it makes for a tractable analysis of the model. Nevertheless, it is not a critical assumption if we focus on Markov dynamics (see Gersbach et al., 2020). As for the electorate, for instance, assuming a two-period horizon guarantees that voters cannot use strategies to reward good behavior of incumbents by punishing bad behavior over some time horizon. Allowing such strategies would not destroy our equilibrium/a but it might add other equilibria. As to the parties, an infinite horizon might also add more equilibria.

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 yet suffered any valence shock.[26] Since we are assuming that m_0 is drawn from some probability distribution, we allow for exogenous changes in citizen preferences. Second,

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

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 of zero. Imposing (2) yields the interesting cases, since large values of c induce no changes in policies and results 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).[27] Fourth and last, we assume that office benefits b are very large, so that regardless of any other consideration incumbents want to be re-elected.

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 setup. 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,

[26] This yields the same results as assuming that the probability of either party being initially in office is arbitrary.

[27] The dynamics with any initial policy and any level of marginal costs of change are examined in Gersbach et al. (2020). Large marginal costs of change yield the property that the status quo is preserved throughout. Very polarized initial policies yield convergence to the policy alternation featured in the present paper.

(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.

Proof: See Appendix A.

□

Hence, in the absence of costs associated with policy changes, the dynamics of our model are simple: policies are polarized since the parties' peaks are chosen. In turn, the incumbent is re-elected if and only if, at the time of elections, the median voter is sufficiently biased towards 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—elections are called seldom because of very long terms—is then not desirable for the median voter with a peak permanently at zero from an ex-ante perspective. That is, in the absence of costs of change, the optimal length is $T = 1$ from the perspective of a citizen with a 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 in our political game, deferring the examination of optimal term-length to the next section. It is 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, there is an equilibrium of $\mathcal{G}$ in which

(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.

Proof: See Appendix A.

□

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. 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 for 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 for the office-holder. This is the manifestation of the incumbency advantage generated by costs of change. It comes about for three reasons. First, although politicians lack the power to commit to arbitrary policies, costs of change enable the incumbent to commit in equilibrium *not* to move the policy further beyond Δ (for party R) or $-\Delta$ (for party L) towards his/her bliss point in the case of re-election.[28] Second, costs of change do *not* enable the challenger to commit to a substantially more moderate policy than the incumbent. In equilibrium, 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, i.e., $-\Delta$ (for party L) or Δ (for party R).[29] Third, to implement this latter policy, the challenger needs to incur a costly policy shift. The net effect of an increase of parameter c on the trade-off between ideology—and hence policy—and valence then depends on the parameters of the model and in particular on the number of valence shocks already experienced by the incumbent.

[28]Gersbach et al. (2020) show that for this first property to hold, it is necessary that costs of change are less convex than the disutility obtained from policies.

[29]The fact that the policies chosen by parties are symmetric with respect to zero follows 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., 2020) or parties' peaks are not symmetrically located with respect to zero (see Section 6.3).

As for party decisions, 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 election takes place. As we have seen, 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 *before* election takes 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 prompts 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 accrue if the status-quo policy—which may have been chosen 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 the quadratic cost of having a policy away from their peak with the linear cost of policy change. If the policy is far away from their peak, the ideological disutility outweighs the adjustment cost, and office-holders change the status quo towards their peak. However, once the policy is no farther away than $c/2$ from his/her peak, the cost of adjustment outweighs the (myopic) gains from moving the policy, 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.

Finally, $\mathcal{G}$ may have other equilibria—see our analysis in Appendix B.[30] However, all these additional equilibria are qualitatively equal to the equilibrium of Theorem 2, since an incumbent chooses the same policy throughout his/her tenure.[31] Moreover, the policies chosen by party R's office-holders must belong to a closed interval of $[\mu - c/2 \cdot (1+\theta), \mu - c/2 \cdot (1-\theta)]$ containing $\mu - c/2 \cdot (1+\theta)$. For their part, the policies chosen by party L's office holders must belong to a closed interval of $[-\mu + c/2 \cdot (1-\theta), -\mu + c/2 \cdot (1+\theta)]$ containing $-\mu + c/2 \cdot (1+\theta)$. These intervals become degenerate as θ converges to zero, in which case the dynamics are described by

[30] In the case of multiple equilibria, there is typically a continuum of them.

[31] If $T = 1$ there may be other equilibria, some of which nevertheless vanish if valence shocks are very large. We refer to Appendix B for details.

Theorem 2 if we take $\theta = 0$ (see above).

There are at least three additional reasons why the equilibrium of Theorem 2 is the best prediction for our model. First, it is the only strategy profile that remains an equilibrium for all parameter constellations. This means, in particular, that this equilibrium is the least demanding in terms of information. For instance, playing the equilibrium of Theorem 2 does not require that either parties (or politicians) know the dynamics of the median voter's peak or the dynamics of the valence shocks. Second, equilibria different from the equilibrium of Theorem 2 also vanish if *(i)* the probability that an incumbent is re-elected is always positive regardless of the number of valence shocks s/he has already suffered, and *(ii)* this probability converges to zero with the number of such shocks. Third and last, the equilibrium of Theorem 2 features the property that both parties choose in all periods the most moderate policy (i.e., the policy closest to zero) among all equilibria. All else being equal, choosing a more moderate policy yields higher utility for parties since a median voter with moderate peak is more likely to elect their candidates.

6 The Optimal Length of a Political Term

To evaluate optimal term-length, we need a measure of social welfare. To this end, we use $\delta \in (0, 1)$ to denote the *social discount factor* used to aggregate utilities of citizens across time for welfare evaluations.[32] One distinct possibility is to equate δ with the quasi-hyperbolic discount factor θ, but our analysis holds for any value of the social discount factor. 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 $\mathcal{S} = (s_1, s_2, \ldots)$, with $s_t \in \{0, 1\}$, to denote an arbitrary realization of the stochastic process describing whether in some period t a shock has occurred to the incumbent's valence ($s_t = 1$) or not ($s_t = 0$).[33] 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 for policies 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, and given $\mathcal{S}$ and $\mathcal{M}$.

[32] The citizens' time horizon consists of the current period and one period ahead and thus make for a lifetime utility in period t equal to the scalar product $(1, \theta, 0, \ldots) \cdot (u(t), u(t+1), u(t+2), \ldots)$. In contrast, the social discount factor applies to all periods, thereby making for a lifetime measure in period t equal to the scalar product $(1, \delta, \delta^2, \ldots) \cdot (W(t), W(t+1), W(t+2), \ldots)$. This means that citizens (and parties) do not internalize the long-term consequences of their actions. This is for mathematical convenience and does not drive our conclusions about optimal term-length. We refer to Section 6.3 for a more in-depth discussion about the role of our extreme form of quasi-hyperbolic discounting.

[33] Valence shocks depend neither on the identity of the incumbent nor on the number of periods in office.

For our analysis in this section, we focus on the equilibrium of Theorem 2.[34] In addition, 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 $\beta \geq 0$.[35] We can interpret β as the variance of shocks to social preferences. 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, i.e. $|a_t|$. 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 $\chi \geq 0$.[36] That is, the median voter suffers *more* from a low-valence incumbent than the average voter, though both voters are affected negatively. This assumption is substantiated 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 of an incumbent being ousted by the median voter despite the fact that this is not socially optimal. This condition is not essential for most of our results in this section but it considerably simplifies the exposition and also enriches the interpretations of our findings; we discuss it further below. 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 broken down into three terms: gains from policies, gains from the incumbent's valence, and losses due to costs

[34] Other equilibria yield qualitatively the same predictions. We do not consider the case that a change of a parameter shifts the equilibrium choice.

[35] 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 optimal term-length.

[36] More generally, one could assume that the median voter's type is $(m_i, \chi_i) \in \mathbb{R} \times \mathbb{R}_+$, which is randomly drawn according to some random process. This would lead to results that are very similar to those that we obtain.

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 of how to weigh policies themselves in welfare, there are at least two reasonable options. These two options enable us to disentangle the different channels through which term length might affect welfare.

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 *ex-ante welfare*. One rationale is that zero is at the same distance from both parties' peaks.[37] 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 society as a whole remains invariant. This leads to the following definition of welfare:

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

where the expected values are taken with respect to the stochastic processes that define $\mathcal{S}$, $\mathcal{M}$, and $\mathcal{I}$, given term-length T.[38] 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 the same policy in all periods t, and oust any incumbent who has experienced a valence shock. However, this outcome cannot be attained through elections, regardless of term-length T. If $T = \infty$, incumbents with a low valence cannot be replaced. If $T < \infty$, there will be costly policy shifts at times, as shown in Theorem 2. In addition, note that 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 favor adaptation of policies to the future preferences of the electorate, and in particular to the median voter's peak, which in our model varies across periods. In the following, we define a welfare measure that internalizes the future evolution of the median voter's peak (but not the fact that the median voter suffers from valence more than the average voter, as captured by parameter χ). Then we investigate the differences in terms of term length between such a definition and the ex-ante welfare measure defined in

[37] As we show in Section 6.3, assuming equal distance from zero for both parties' bliss points does not lead to knife-edged results.

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

Expression (9). Formally, the second definition of welfare—*interim welfare*—is

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

This leads to

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

Again, the expected values are taken with respect to the stochastic processes that define $\mathcal{S}$, $\mathcal{M}$, and $\mathcal{I}$, given term-length T. Equation (11) captures the value in terms of 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 additive term that is independent of T, the following two functions are equal:

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

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

Proof: See Appendix A.

□

To understand the above proposition, let

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

One can verify that $H_0(T)$ is a positive, real-valued function that is decreasing in T.[39] The latter property is not trivial. While shorter terms enable changing policy to adapt to the preferences of the median voter, they also enable policy changes that occur for valence reasons. Yet, in net terms, the above result shows that elections should take place as often as possible if our goal is that policies reflect the median voter's preferences in the best way possible, all else being equal. Also according to Proposition 1, ex-ante welfare and interim welfare are equivalent in certain cases.

[39]See the proof of Proposition 1.

First, they coincide when $\beta \leq c/2$. In this case, the shocks to the median voter's peak are mild enough for incumbents only to be 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 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 in interim welfare. 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 again leads 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 for welfare either.

6.2 Welfare analysis: Costs of change versus valence

Henceforth, we focus on ex-ante welfare as described by Expression (9).

6.2.1 The formula for the optimal length of political terms

The next result characterizes 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, optimal term-length T^* exists, is finite, and maximizes the following expression as a function of T:

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

Proof: See Appendix A.

□

To understand expression $M(T)$ in Proposition 2, it is convenient to split it into a number of parts. 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 ex-ante welfare that can be attributed to 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 positive, real-valued function that captures the ex-ante welfare that can be attributed to 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 positive, real-valued function that captures the ex-ante welfare that can be attributed to 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 one can verify that $H_I(T)$, $(1-(1-\rho^T)) \cdot H_{II}(T)$, and $(1-\rho^T) \cdot H_{III}(T)$ are decreasing functions in T.[40] In particular, $H_I(T)$ is maximal at $T=1$, while $-(1-(1-\rho^T)) \cdot H_{II}(T)$ and $-(1-\rho^T) \cdot H_{III}(T)$ 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 of illustrating the shape of the objective function that determines optimal term-length. It shows that optimal term-length is more than one for some generic parameter values. In such cases, there is a discrepancy between what optimal term-length is from an ex-ante perspective and from the perspective of a voter who wants to have the *opportunity* to oust the incumbent in every period by means of elections if s/he

[40] See the proof of Propositions 3–6.

deems it necessary. Why is it optimal for citizens to tie their own 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 may not be aligned with the interests of the average voter. The latter, whose peak is 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 term-length.[41] 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.

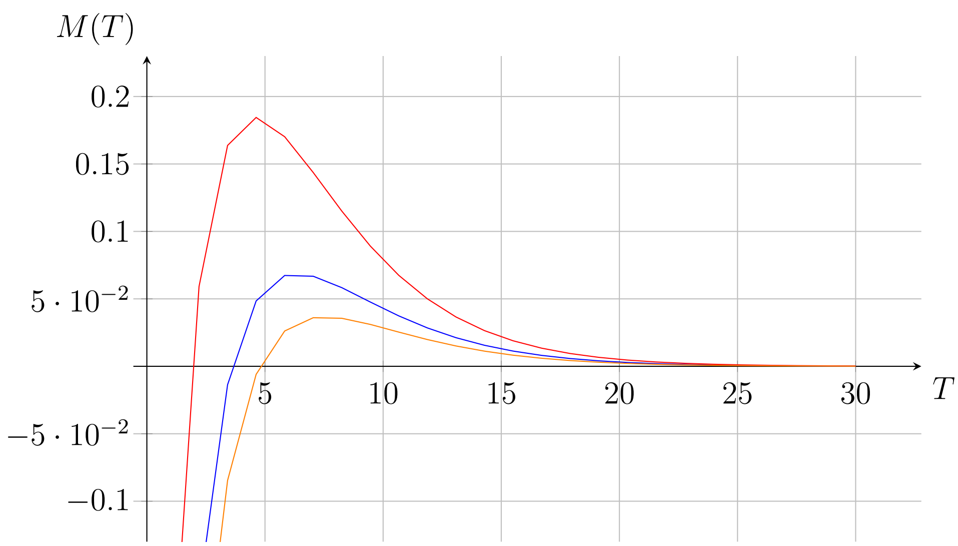

Figure 2: Function $M(T)$ for $\delta = 0.7$, $\mu = 0.4$, $c = 0.35$, $\theta = 0.8$, $\rho = 0.45$, $\beta = 0.2$, $\eta = 0.1$, $\chi = 6$, and $A = 0.016$ (orange line), $A = 0.02$ (blue line) and $A = 0.03$ (red line), where the x-axis encompasses different values of T (say, in years).

6.2.2 Comparative statics and empirical hypotheses

Figure 2 illustrates the effect on optimal term-length of increasing the extent of the valence shocks. In this section, we show that Proposition 2 enables us to obtain a series of insightful comparative statics results with regard to optimal term-length, including the effect of varying parameter A. These results yield hypotheses that could be tested empirically, thereby enabling us to see whether

[41] 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. (2020).

term lengths are set optimally in accordance with our theory. We divide our analysis in four parts, each of which is discussed in the following. The first proposition is concerned with changes in the valence shocks.

Proposition 3

The optimal term length T^*

- *decreases with* A*, and*
- *decreases with* ρ*, provided that* δ *is sufficiently close to* 1 *and* $2\beta < 3c$*.*

Proof: See Appendix A.

□

That is, optimal term-length decreases with the probability of a valence shock occurring and with the extent of such shocks. The second result is not surprising, since the expected valence of incumbents is lower if shocks are greater. A shorter term simply reduces the (expected) disutility generated by low valence incumbents, all else being 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 of a costly policy change occurring—which calls for longer terms. Provided that the future is sufficiently valuable from a social perspective (i.e., δ is sufficiently close to 1) and assuming that the preference shocks are relatively small (i.e., $2\beta < 3c$), we show that the former effect dominates the latter, so an increase of the valence shock probability ρ also calls for shorter periods. This result holds much more generally, but providing an analytical proof is much more cumbersome. 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, probability ρ alone influences how often incumbents are going to be ousted and costly policy changes are going to come about. This means that T^* may have very different values, even if we keep ρA constant.

Proposition 3 has implications for the design of political institutions. In cases where candidates undergo tight scrutiny until they are selected, the optimal length of terms should be larger, all else being equal. This is because the probability and/or the size of the shocks to valence would be smaller (or could even be reduced to 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 more especially 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, the valence of representatives can have major consequences for the importance that such districts gain in terms of policy-making. Terms should be short to prevent constituencies from suffering for a long time from incompetent representatives. This might be the case when parliaments have a large number of members, such as the US House of Representatives or the European Parliament, and it is difficult to assess the representatives' valence. The opposite argument holds for small collective decision bodies, for which it might be easier to observe how competent representatives are.

As a second comparative statics exercise based on Proposition 2, we look at changes in the process determining the median voter's peak.

Proposition 4

The optimal term length T^*

- *increases with β, and*
- *increases with $1 - \eta$.*

Proof: See Appendix A.

□

That is, optimal term-length increases with the variance of the median voter's peak when it changes and with the probability that the median voter's peak changes. First, an increase in β leads to an increase in 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 prefers to oust the incumbent for policy reasons, thereby generating higher costs of change. This property holds not only for our parametrized conditional distribution, but for any distribution determining the new median voter's peak that puts more mass on higher absolute values of the peak. It 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 the case that $2\beta \geq c$. If the latter condition does not hold, an incumbent is never ousted for policy reasons. Second, if citizens' preferences become more stable—in the sense that the median voter's peak changes less frequently—, optimal term-length can never become larger. This is because with more stable social preferences it is less likely that a power shift occurs for policy reasons, so enabling more frequent elections entails fewer risks.

Proposition 4 suggests that term-length should be larger when societies are strongly polarized and/or subject to large swings in the electorate. Otherwise, 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 few decades, our results indicate that those term-lengths that have remained constant may *now* be too short, all else being equal.

The third comparative statics result analyzes changes in party polarization and the individuals' (quasi-hyperbolic) discount factor.

Proposition 5

The optimal term length T^*

- *increases with μ, and*
- *decreases with θ.*[42]

Proof: See Appendix A.

□

That is, optimal term-length increases with party polarization and decreases with the factor with which agents discount the future period. First, an increase in party polarization leads to more polarized policies and thus larger policy shifts, so that ousting the incumbent generates more costs of change. As a result, the optimal term can never become smaller because it will enable more frequent power shifts. Second, when the future becomes more valuable for citizens and candidates, policies become more moderate, so 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 being equal.

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

Proposition 6

The optimal term length T^*

- *increases with c, provided that $2\beta < c < \frac{\mu}{1+\theta}$, and*
- *decreases with c, provided that $\frac{\mu}{1+\theta} < c < 2\beta$.*

[42]The comparative statics with regard to θ may reverse for some equilibria different from the one of Theorem 2.

Proof: See Appendix A.

□

Accordingly, the effect of c on T^* is generally ambiguous. 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 will yield higher optimal term-length. We recall that in such a case, incumbents are only ousted if their valence is (sufficiently) low. A marginal change of c therefore only affects the costs of change that 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 and 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 favors lower terms. Additionally, when $c < 2\beta$, incumbents may also be ousted for policy reasons. In such cases, increasing c reduces the probability that this happens, thereby also making lower terms more appealing, all else being 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 in particular to checks and balances. In most democracies, the larger the policy reform, the more hurdles the proponents of the change have to overcome, and hence the higher the costs of change will be.[43] Examples of such hurdles are judiciary oversight, qualified majority, or double majority. Then, assuming $2\beta < c < \mu/(1+\theta)$ and interpreting that c accounts for the set of institutional hurdles that are necessary to change policy, we find 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 earlier in history confirmation required a super-majority. This double step can be seen as an instance of a high institutional hurdle. Our result of Proposition 6 regarding comparative statics on c can further be used to rationalize the

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

recommendation that term length for Supreme Court members should be longer than, say, for members of the legislative or the executive, provided that fluctuations of the median voter's peak are low. But if the latter are high, higher hurdles call for lower term lengths, provided hurdles were already high before.

Thus far, we have examined how optimal term-length varies with the most crucial parameters of our model of electoral competition. As far as comparative statics exercises are concerned, however, it is also important to understand how changes in the misalignment between the average voter and the median voter with regard to how much they value the inefficient provision of public goods affect optimal term-length, as captured by χ. To this end, we look at an increase of parameter $1-\chi$. We have seen that the exact magnitude of this parameter is immaterial to optimal term-length, provided that Condition (8) holds. This condition guarantees that the (current) median voter will always oust an incumbent who has already suffered a valence shock during tenure. Hence the interesting case occurs when, starting from Condition (8), parameter χ decreases enough so that Condition (8) fails to hold. To illustrate the effect of such a change, assume for simplicity that $\beta = 0$ (i.e., the median voter's peak is zero throughout all periods) and $\rho = 1$ (i.e, the incumbent suffers a valence shock in every period).[44] In addition, let χ_L and χ_H be such that

$$1 + \chi_L < \frac{2c\Delta}{A} < 2(1 + \chi_L) < 1 + \chi_H.$$

The above chain of inequalities guarantees that when $\chi = \chi_L$, an incumbent who has suffered at most one valence shock during tenure is never ousted, while an incumbent who has suffered at least two valence shocks is always ousted. For its part, Condition (8) holds when $\chi = \chi_H$. This is the case we have analyzed throughout this section. If $\chi = \chi_H$, the expected utility from costs of change when the incumbent is in his/her first period in office, denoted by V^0, is

$$V^0 = \delta \cdot V^1,$$

where V^1 denotes the expected utility from costs of change when the incumbent has already been one period in office. In turn,

$$V^1 = \delta \cdot V^0 - 2c\Delta,$$

which yields

$$V^0 = -\frac{\delta}{1-\delta^2} \cdot 2c\Delta. \tag{13}$$

If we let $W_L^1(T)$ ($W_H^1(T)$) denote ex-ante welfare when $\chi = \chi_L$ ($\chi = \chi_H$), it then follows from our previous analysis and Equation (13) that

$$W_L^1(T) - W_H^1(T) = \begin{cases} \frac{\delta^2}{1-\delta^2} \cdot 2c\Delta & \text{if } T = 1, \\ 0 & \text{otherwise.} \end{cases}$$

[44] The analysis of the general case can be provided upon request.

It is clear that the above function is decreasing in T, and it is therefore maximized at $T = 1$. Hence, T^* must be (weakly) smaller if the misalignment between the average voter and the median voter becomes smaller with regard to how much they value the inefficient provision of public goods (i.e., if $1 - \chi$ becomes larger, say by changing from $1 - \chi_H$ to $1 - \chi_L$). The reason is that this makes it less likely that the (current) median voter ousts the incumbent and triggers a policy change.

Finally, for our analysis in this section, we have focused on ex-ante welfare. As for interim welfare, most of our analysis carries over, with some exceptions. For instance, consider that β is very large, in particular $\beta > c/2$, and then it increases further. We recall that parameter β captures the magnitude of the preference shocks. Then one can verify that such an increase of β calls for a shorter optimal term-length T^* (at least, weakly), all else being equal.[45] This is in contradiction with Proposition 4. The reason is that when we take an interim welfare perspective (whereby social preferences vary in line with those of the median voter) instead of an ex-ante perspective (whereby social preferences are fixed), larger preference shocks make it more likely that it is necessary to oust the incumbent and elect a challenger whose preferences are more aligned with those of the current median voter. This can be done by calling elections more often.

6.3 The role of quasi-hyperbolic discounting and party symmetry

In our model, voters are present-biased and we have assumed that parties' peaks are located symmetrically around zero. How do these features influence our analysis of optimal term-length? We first focus on quasi-hyperbolic discounting.[46] Consider a citizen with peak at i, with $i \geq 0$. From Theorem 2, we know that s/he will oust the incumbent, say from party R, in some period t in which elections are called if and only if

$$- a_t > 2c\Delta + 4\Delta i. \tag{14}$$

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.

Assuming that Inequality (14) holds, consider now a fictitious citizen with peak also at i who has a lifetime utility based on standard exponential discounting. Is it possible that s/he does not

[45]This can be done along the lines of the proof of Propositions 2–6. It suffices to show that function $\frac{\delta-\delta^{T+1}}{1-\delta^{T+1}} \cdot \frac{1-\delta^{T+1}}{1-\delta^{T}} \cdot (1-\rho)\frac{1+(\delta\eta)^T(1-(1-\rho)^T)}{1+(\delta\eta)^{T+1}(1-(1-\rho)^{T+1})}$ is decreasing in T.

[46]For parties that are forward-looking, Gersbach et al. (2020) have shown that the main dynamics are similar to the case where parties are present-biased, yet they differ quantitatively. This has no bearing on the main thrust of our analysis on optimal term-length.

want the present-biased voter with peak at i to oust the incumbent in period t? The only way to prevent this from happening is for elections *not* to be called in period t.[47] 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, due to Inequality (14), the fictitious citizen derives less utility in period $t+1$ from the incumbent than from the challenger belonging to party L. In fact, this property holds for every future period until the incumbent is ousted. Hence, the fictitious citizen with peak at i is harmed if s/he does not allow the present-biased voter i to oust the incumbent from party R in period t already. That is, present-biased preferences can never entail too frequent policy changes from the perspective of the fictitious citizen. Such preferences can only lead to policy changes that are too rare. But this cannot be corrected by changing term-length. While *not calling* elections prevents an undesirable policy change, *calling* elections does not ensure a desirable policy change—see Footnote 47.

To sum up, as far as optimal term-length is concerned, a present-biased citizen and a fictitious citizen with the same peak agree. This means that the discrepancy between the fact that citizens are present-biased in our model and having defined ex-ante welfare as the lifetime utility based on exponential discount of a fictitious citizen (with peak at zero) has no bearing for optimal term-length.

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, although still on different sides of the political spectrum. Yet, we assume the counterpart of (2), i.e.,

$$0 \leq c < \frac{\mu_R - \mu_L}{1+\theta}. \tag{15}$$

We proceed by considering that ex-ante welfare is still defined by the lifetime utility of a voter with peak permanently at zero. As for equilibrium behavior, our analysis in Section 5 already yields parties' and voters' choices.[48] That is, party R chooses $\Delta_R := \mu_R - c/2(1+\theta)$ and party L chooses $\Delta_L := \mu_L + c/2(1+\theta)$. As for elections, if we let $2\Delta = \Delta_R - \Delta_L$, then the incumbent $k \in K$ is 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

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

[47] We do not need to consider the case where Inequality (14) does not hold. In this case, by calling elections, the fictitious citizen with peak at i cannot force the present-biased voter with peak at i to oust the incumbent.

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

For instance, if $\mu_R + \mu_L > 0$, it is less (more) likely that an incumbent from party R (party L) is 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 $\beta = 0$.[49] Then we take Equation (9) and investigate the term

$$\mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot i_t^2\right]. \tag{16}$$

When $\mu_R + \mu_L = 0$, Expression (16) is independent of T. Assume without loss of generality that $\mu_R + \mu_L > 0$ and note that a (present-biased) voter with peak at zero re-elects 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 re-elects an incumbent from party L, provided that s/he has not suffered any shock. Therefore we have 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,

$$\mathbb{E}_T\left[\sum_{t\geq 1}\delta^{t-1}\cdot i_t^2\right] = -\frac{\Delta_R^2 + \Delta_L^2}{2}\cdot\frac{1}{1-\delta}. \tag{17}$$

The above term is independent of T. This means that if the parties' peak asymmetry is small relative to costs of change, our analysis about optimal term length remains intact. To obtain Equation (17), define V_R (V_L) as the expected value of Expression (16) when the incumbent belongs to party R (L). Then,

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

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 (17).[50] Second, assume that $\mu_R + \mu_L > c$, which means that party R does not enjoy any net incumbency advantage. Then we have

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

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

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

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

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

One can verify that the above expression is decreasing in T.[51] This implies, *ceteris paribus*, that if one party's peak is much more extreme than the other party's from the perspective of the voter who has peak at zero, terms should be shorter. This prevents the party which has an extreme peak from dictating policy for many periods and ensures that the party with the more moderate peak is only ousted when its candidate has suffered a valence shock. In such a case, s/he is replaced by a candidate from the more extreme party for only one period, after which a fresh candidate from the moderate party is 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 setup is that the politicians' valence can only be affected by negative shocks during tenure. This rules out the possibility that an incumbent may be better able to provide public goods at the end of a given term than at the beginning of that term. There are several reasons why the valence of an incumbent may increase over time. For instance, office-holders may need time to become efficient policy-makers through learning by doing or need 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 political terms should be from a welfare perspective. If incumbents could only experience positive shocks to their valence, in particular, the optimal length would be infinite.

[51]It suffices to realize that $\Delta_R^2 - \Delta_L^2 > c(\Delta_R - \Delta_L) > 0$, where the first inequality holds since $\mu_R + \mu_L > c$ and the second inequality is due to (15) , and then to show that $\frac{\delta^T(1-\rho)^T}{1+\delta^T(1-(1-\rho)^T)}$ is decreasing in T. For the latter, see (64) in Appendix A.

The possibility that incumbents can learn over the course of their tenure adds an interesting aspect to the role of our (extreme) form of quasi-hyperbolic discounting. To illustrate this, we consider the simplest possible scenario, in which it takes the incumbent two periods to *learn*—i.e., to attain the necessary governance skills, 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).[52] 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. To determine the effect of learning on optimal term-length, it suffices to add the term

$$H_{IV}(T) := \frac{\delta^T}{1-\delta^T} \cdot \frac{A}{2c\Delta} \cdot \left[\frac{1}{1-\delta} \cdot \frac{\delta - \delta^T}{\delta^T} + (1-\rho)^T \frac{1-(1-\rho)^T\delta^T - T(1-\rho)^T\rho\delta^T}{1-(1-\rho)^T\delta^T - (T(1-\rho)^T\rho\delta^T)^2} \right] \quad (18)$$

to the expression $M(T)$ defined in Equation (12). To derive the above expression, define $V_{x,y}$ to be the office-holder's expected lifetime valence at the beginning of one term that can be attributed to learning, depending on whether the current incumbent has already learned ($x = 1$) or not ($x = 0$), and depending on the number y of valence shocks s/he has already suffered. For each term-length T, we have[53]

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

Solving the three above equations with standard algebra and noting that $H_{IV}(T) = V_{0,0}$ yields Equation (18). 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 own 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.

[52] A similar, yet more cumbersome argument can be made if it takes incumbents more than two periods to learn.

[53] Note that a politician who has suffered at least two valence shocks is ousted, no matter whether s/he has learned or not.

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 enabled us to build a theory for the optimal determination of the length of a political term, which in its turn provides a number of insightful comparative statics. Of course, many factors that are absent from our analysis in Section 5 are also relevant for actually determining term-length. These include 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 equal to K.[54] If elections take place every T periods, the average cost per period is equal to K/T. To determine the term-length optimally, it then 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 being equal. The relevant point, however, is that our model potentially enables us to capture the *degree* to which compared to other parameters campaign spending influences optimal term-length. In the US federal elections, for example, campaign spending is very large and has risen dramatically in the past decades.[55] This poses 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 term-length should be increased, if at all.

Second, interpreted broadly, our baseline model already incorporates a 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

[54] For the sake of argument, we assume that K is constant over time.

[55] 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.

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 from an ex-ante perspective to implement such a policy. The incumbent has more information about the policy and already realizes in period t that costs Q accrue, so that his/her net valuation of the policy is $q - Q < 0$.[56] 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 implements the policy to increase his/her chances of re-election against possible valence shocks and changes in the preferences of the electorate. Therefore, we can assume that at least for some generic parameter constellations, costs $q - Q \cdot \delta^2$ accrue in any period in which elections are held. To determine the term-length optimally in the presence of our reduced form of pandering, it then 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 of incumbents pandering to the electorate calls for longer terms, all else being equal. Doing so reduces the policy distortion 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 the entry of third parties or third candidates might affect our results. We can assume that parties R and L have the same peaks as (groups of) voters r and l, which together with the median voter (and his/her group) make up for the electorate. 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 numerous and partisan or if there were entry costs in the form of lack of awareness by the electorate, absence of proportionality in the electoral system, or scarce (public) funds.[57] The possibility of party entry at the median voter's peak can also be ruled out in our citizen-candidate setup when there are just two party positions (or peaks), those of R and L, and these are therefore the same as the voters' positions, i.e., those of r and l. In such a scenario, the

[56]For simplicity, we assume that q and Q are constant over time.

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

median voter would change only when the relative size of the two parties flips.

Apart from entry by a third party, another possible scenario is that the opposition party appoints 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 remained in office in the place 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 a party would have little incentive to appoint a less partisan candidate.[58]

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 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.

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 preferences there are two polar cases with linear costs of change. First, in the absence of any capacity constraint, office-holders upon election would change policy towards their bliss point in all dimensions. This would create an incumbency

[58]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 the consequences of such behavior.

advantage in all dimensions where parties have diverging preferences, and our analysis of optimal term-length would then extend from the case of one dimension to multiple dimensions. Second, there may exist capacity constraints precluding more than one policy from being 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, depending on how polarized the parties' positions and how large the marginal costs 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 being equal.

With multiple dimensions and capacity constraints, it might additionally be the case that the incumbent party can set the political agenda before elections, i.e., that it can determine which policy dimension should be at the center of the campaign and thus foremost in voters' minds. 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.[59] The incumbent party would simply choose the dimension where 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 with the US Presidency, for example. In many representative democracies (see e.g. Diermeier and Merlo, 2000, and the references therein), it is possible for incumbents to call early elections. Our model can be easily adapted to include the possibility of calling early elections in every period of a term except the last one, in which elections are automatically called. Naturally, we must assume $T > 1$. For an outline of our analysis, it is convenient to recall that we have proceeded on 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 that follows. As in Section 6, we assume for simplicity that an office-holder who has suffered at least one valence shock is never re-elected. The problem of calling early elections can be described as an optimal stopping problem in a 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 in calling early elections, such as the possibility for the electorate to punish the incumbent for

[59] It is well-known that incumbents use their power as office-holders 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).

opportunistic behavior if s/he calls elections for his/her own benefit.

Given that policy only shifts with office-holder turnover no matter what the term-length is, 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 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). An initial 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 calls 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 extremely biased towards the challenger's peak at election date.

Quite often, however, elections cannot be called immediately but only for a certain date some time in the future. This involves uncertainty, so the decision to call early elections, say at the start of some period t to take place at the end of this period, involves in turn a gambling element. The reason is that both the median voter's peak and the incumbent's valence may change between the moment elections are called and the moment they take place. For the sake of 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, in which s/he holds office when focusing on periods $t+1$ and $t+2$.[60] We concentrate on the case where elections are (automatically) called at the end of period $t+1$ if they are not called in period t. Given the politician's horizon, incumbents never call early elections in periods $t-1, t-2, \ldots$ within the same term. The argument can be nonetheless applied recursively. Without loss of 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.[61] If the incumbent calls early elections 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 is 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,

[60] 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 into his/her maximization problem in period t.

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

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

$$N_1 \geq N_2 \Leftrightarrow \frac{\beta + \min\{\beta, c/2\}}{2\beta} \geq \frac{1 - (1-\rho)\eta(2 - (1-\rho)\eta) \cdot \mathbb{1}_{m_t \geq -c/2}}{(1-\rho)(1-\eta)(2-(1-\rho)(1+\eta))}. \tag{19}$$

Hence, the incumbent calls early elections 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 (19) 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. Because Inequality (19) depends solely on the model primitives, there are two possibilities. First, early elections are never called, which happens if Inequality (19) 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 (19) 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 is only used for low-valence office-holders. In contrast, high-valence office-holders 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.

7.6 Varying costs of change

So 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. Suppose, for example, that the marginal cost of change decreased unexpectedly. Then, using arguments analogous as those used in the proof of Theorem 2, one can see that the policy would move further towards the incumbent's peak. A reduction of c could be the result of a learning process by the office-holder during tenure. The net effect of changes in c in determining the optimal length of a political term has been (partially) described in Proposition 6.

7.7 Endogenizing valence and the median voter's peak

In our baseline setup, 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 is in place for the next period. Moreover, this assumption has yielded rich dynamics. Within our model, it is nevertheless worth asking 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 $1-\eta$ 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 an increase of β and ρ. This would generally result in higher turnover but would not change policy choices—Propositions 3 and 4 describe the marginal effects on optimal term-length of changing such parameters. An alternative possibility is that during tenure, the office-holder can exert some (fixed) effort to either reduce ρ or to increase η but 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 of choosing

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

where ρ^* and η^* denote the probability of a valence shock and the persistence level that is in place for the next term, respectively. The main trade-off for the office-holder is clear from our analysis 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 will change in the direction of the incumbent's peak. There are various ways 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 on 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 Propositions 3 and 4.

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

For our welfare anaylsis in Section 6 we proceeded on the assumption that the median voter suffers (weakly) more from valence shocks than the average voter. This has yielded interesting results regarding the optimal length of a political term, especially 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 this end, 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 $x_t \geq 0$, is 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[62]

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

where $\Psi > 0$ and a_{t-1} is the valence of the incumbent in period t (before any valence shock could take effect). That is, an incumbent with lower valence is less able to provide public goods than an incumbent with higher valence. Crucially, we assume that the median voter with peak $m_t \in [-\beta, \beta]$ determines the level x_t^* to be provided. It is clear that s/he chooses the level $x \geq 0$ that maximizes $v_{m_t}(x)$.[63] 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), i.e.,

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

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

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

[62] 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.

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

That is, assuming that A is sufficiently large, both the median voter and the average voter suffer from an incumbent with lower valence, although 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 of the average voter.

8 Conclusion

An appropriate framework upon which to build a full-scale theory of optimal term-lengths is missing in 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. Our analysis offers an array of hypotheses that can be tested empirically against the assumption that term-length is set optimally for particular political systems in accordance with our theory. While some of the comparative statics are intuitive when taken individually, our analysis provides a quantitative approach that enables us to weigh each factor when we take all of them collectively. Of course, the features of our model can be supplemented with further elements, notably by introducing some asymmetry of information. This we leave to further research.

References

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

Austen-Smith, D., Dziuda, W., Harstad, B., and Loeper, A. (2019). Gridlock and inefficient policy instruments. *Theoretical Economics*, 14(4):1483–1534.

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. (2020). Electoral competition with costly policy changes: A dynamic perspective. *Mimeo.*

Gersbach, H., Muller, P., and Tejada, O. (2019). 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. *American Journal of Political Science, forthcoming*.

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.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, pages 157–180.

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.

Topkis, D. M. (1998). *Supermodularity and complementarity*. Princeton university press.

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 A (for online publication online)

In this appendix we provide the proofs of the results in 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.[64] 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 election takes 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}) respectively in period $t+1$ if they are elected. 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

$$(2m_t - i_{t+1}^{K} - i_{t+1}^{-K}) \cdot \left(i_{t+1}^{K} - i_{t+1}^{-K}\right) \geq A \cdot z_t, \tag{20}$$

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 (20) 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 incumbent k. We distinguish two cases. On the one hand,

[64] 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.

the problem of incumbent k in the beginning of period t where election takes place is

$$\max_{i_t^K \in \mathbb{R}} \Big\{ -(i_t^K - \mu_K)^2 + a_{t-1} - \rho A + p \cdot \theta \left(b - (i_{t+1}^K - \mu_K)^2 + a_{t-1} - \rho A\right) + (1-p) \cdot \theta \left(-(i_{t+1}^{-K} - \mu_K)^2\right) - \theta \rho A \Big\}, \tag{21}$$

where p has been introduced above. Then the problem in (21) 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':

$$\max_{i_{t'}^K \in \mathbb{R}} \Big\{ -(i_{t'}^K - \mu_K)^2 + (1+\theta) \cdot a_{t'-1} + \rho \cdot \left[-A + \theta \cdot (\sigma_K((t'+1)/T, i_{t'}^K, a_{t'-1} - A) - \mu_K)^2\right] . + (1-\rho) \cdot \left[\theta \cdot (\sigma_K((t'+1)/T, i_{t'}^K, a_{t'-1}) - \mu_K)^2\right] - \theta \cdot \rho A \Big\}. \tag{22}$$

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 (22) 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 (20) 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

Here, we tackle the problem of existence of equilibria in game $\mathcal{G}$ by showing that one particular strategy profile is an equilibrium. Whether this equilibrium is unique or there may be other potential equilibria is discussed in Appendix B. For the proof of existence of an equilibrium, we will assume that 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

$$\sigma_m^*(K, i_{t-1}, a_t, m_t) = \begin{cases} K & \text{if } m_t \geq -\frac{c}{2} \cdot \frac{i}{\Delta} - \frac{a_t}{4\Delta}, \\ -K & \text{otherwise.} \end{cases} \tag{23}$$

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

$$\sigma_R^*(t/T, i_{t-1}, a_{t-1}) = \Delta. \tag{24}$$

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

$$\sigma_L^*(t/T, i_{t-1}, a_{t-1}) = -\Delta. \tag{25}$$

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 election takes 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

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

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 (24) and (25), incumbent k will choose Δ in period $t+T+1$ if s/he is re-elected.

In turn, 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

$$p(i, z_+, m_-) \tag{26}$$

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, 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

$$m \geq -\frac{c}{2} \cdot \frac{i}{\Delta} + \frac{A \cdot z}{4\Delta}. \tag{27}$$

From the above expression, it follows that

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

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

$$m \geq -\frac{c}{2} + \frac{Az}{4\Delta}. \tag{28}$$

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

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

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

$$-(m-\Delta)^2 - c \cdot (\Delta - i) - A \cdot z \geq -(m+\Delta)^2 - c \cdot (-\Delta - i), \tag{29}$$

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 incumbent $k \in R$ at the beginning of period $h = t + T$, before s/he can 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:

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

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

$$\begin{aligned} G(i) = & -(i-\mu)^2 - c \cdot |i-j| - \theta \cdot c \cdot |i + \Delta| - \chi' \\ & + \theta \cdot (1-\rho) \cdot p\,(i, z, m_{-}) \cdot \Big[M^z - c \cdot |\Delta - i| + c \cdot |i + \Delta| \Big] \\ & + \theta \cdot \rho \cdot p\,(i, z+1, m_{-}) \cdot \Big[M^{z+1} - c \cdot |\Delta - i| + c \cdot |i + \Delta| \Big], \end{aligned} \tag{30}$$

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:

$$-\Delta \leq j \leq \Delta.$$

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

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

In this case, using (30), 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 because $c > 0$, $\theta \leq 1$, 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 of $G(\cdot)$ 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 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 $p\,(i, z+1, m_{-})$ and $p\,(i, z+1, m_{-})$ 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: $\Delta < i$

In this case, we can write

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

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

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

where last inequality holds 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

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

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

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

where last inequality holds since

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

As in Case I, we can see that

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

To sum up, we have proved that the problem described in (30) 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 the case that $T > 1$. Then the incumbent faces the following problem in a particular period $h = t+1, \ldots, t+T-1$:

$$\max_{i \in \mathbb{R}} H(i) := \max_{i \in \mathbb{R}} -(i-\mu)^2 - c \cdot |j-i| - \theta \cdot c \cdot |\Delta - i| - \chi', \tag{31}$$

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 election takes place, and that we will accordingly proceed by backward induction, from 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

$$j \leq \Delta.$$

Note that $H(i)$ is a continuous function for all $i \in \mathbb{R}$ that is also differentiable for all $i \in \mathbb{R}$, except possibly for a finite number of points. 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 (31) for period h is maximized for $i = \Delta$, thereby completing the proof of existence of equilibrium. For the discussion of the uniqueness of equilibria, we refer to Appendix B.

□

Proof of Proposition 1

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

$$\frac{1}{2} \cdot \left([W^2 - W^1](T) - [W^2 - W^1](T')\right) = \mathbb{E}_T \left[\sum_{t \geq 1} \delta^{t-1} \cdot m_t i_t\right] - \mathbb{E}_{T'} \left[\sum_{t \geq 1} \delta^{t-1} \cdot m_t i_t\right], \tag{32}$$

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 and the initial median voter's peak m_0. We recall that we are assuming that m_0 is drawn according to a uniform distribution on $[-\beta,\beta]$. To compute each of the two terms of the right-hand side of Expression (32), 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

$$B_R(T|m_t) := \mathbb{E}_T\left[\sum_{t'\geq t+1} \delta^{t'-t-1}\cdot m_{t'}i_{t'}\middle| m_t\right]. \tag{33}$$

Analogously, one can define

$$B_L(T|m_t) := \mathbb{E}_T\left[\sum_{t'\geq t+1} \delta^{t'-t-1}\cdot m_{t'}i_{t'}\middle| m_t\right] \tag{34}$$

for the case where L is in office in period $t+1$. Note that up to an additive term that is independent of T (see Equation (11)), we can write

$$[W^2 - W^1](T) = 2\int_{-\beta}^{\beta} B_R(T|m_0)dm_0 = 2\int_{-\beta}^{\beta} B_L(T|m_0)dm_0. \tag{35}$$

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

$$\begin{aligned}
B_R(T|m_t) &= \Delta\cdot\sum_{t'=t+1}^{t+T}\delta^{t'-t-1}\cdot\left[\eta^{t'-t}\cdot m_t + (1-\eta^{t'-t})\cdot\frac{1}{2\beta}\int_{-\beta}^{\beta} idi\right]\\
&+ (1-(1-\rho)^T)\cdot\delta^T\cdot\left(\eta^T\cdot B_L(T|m_t) + (1-\eta^T)\cdot\frac{1}{2\beta}\int_{-\beta}^{\beta} B_L(T|i)di\right)\\
&+ (1-\rho)^T\cdot\delta^T\cdot\Bigg(\eta^T\cdot\left[\mathbb{1}_{m_t\geq -c/2}\cdot B_R(T|m_t) + (1-\mathbb{1}_{m_t\geq -c/2})\cdot B_L(T|m_t)\right]\\
&+ (1-\eta^T)\cdot\left[\frac{1}{2\beta}\int_{-\min\{c/2,\beta\}}^{\beta} B_R(T|i)di + \frac{1}{2\beta}\int_{-\beta}^{-\min\{c/2,\beta\}} B_L(T|i)di\right]\Bigg),
\end{aligned} \tag{36}$$

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 of party R if the latter has suffered a valence shock during the term, which happens with probability $1-(1-\rho)^T$. If the incumbent has not suffered any valence shock, which happens with probability $(1-\rho)^T$, 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 remark will substantially facilitate the analysis. Let $(p_{\mathcal{I},\mathcal{M}})_{\mathcal{I},\mathcal{M}}$ $((p_{-\mathcal{I},-\mathcal{M}})_{-\mathcal{I},-\mathcal{M}})$ denote 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. 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 $-m_t$. This implies that, for all $m_t \in [-\beta, \beta]$,

$$\begin{aligned} B_R(T|m_t) &= \sum_{\substack{\mathcal{I}=(i_{t+1},i_{t+2},\ldots),\\ \mathcal{M}=(m_{t+1},m_{t+2},\ldots)}} p_{\mathcal{I},\mathcal{M}} \cdot \sum_{t'\geq t+1} \delta^{t'-t-1} \cdot m_{t'} \cdot i_{t'} \\ &= \sum_{\substack{-\mathcal{I}=(-i_{t+1},-i_{t+2},\ldots),\\ -\mathcal{M}=(-m_{t+1},-m_{t+2},\ldots)}} p_{-\mathcal{I},-\mathcal{M}} \cdot \sum_{t'\geq t+1} \delta^{t'-t-1} \cdot (-m_{t'}) \cdot (-i_{t'}) = B_L(T| - m_t). \end{aligned} \tag{37}$$

If we use Equation (37), then we can integrate (36) for $m_t \in [-\beta, \beta]$ to obtain

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

which can be rewritten as

$$\begin{aligned} &\int_{-\beta}^{\beta} B_R(T|m_t)dm_t \\ =&\frac{(1-\rho)^T \cdot \delta^T}{1-\delta^T} \cdot \left[\int_{\min\{c/2,\beta\}}^{\beta} B_R(T|m_t)dm_t - \int_{-\beta}^{-\min\{c/2,\beta\}} B_R(T|m_t)dm_t \right]. \end{aligned} \tag{38}$$

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, $\min\{\beta, c/2\} = \beta$. Then, it trivially follows from Equation (38) that

$$\int_{-\beta}^{\beta} B_R(T|m_t)dm_t = 0.$$

We thus have

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

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 Equation (37), it must be the case that

$$\begin{aligned}
\int_{-\beta}^{-c/2} B_L(T|m_t)dm_t &= \int_{c/2}^{\beta} B_R(T|m_t)dm_t := K_{-1},\\
\int_{-c/2}^{c/2} B_L(T|m_t)dm_t &= \int_{-c/2}^{c/2} B_R(T|m_t)dm_t := K_0,\\
\int_{c/2}^{\beta} B_L(T|m_t)dm_t &= \int_{-\beta}^{-c/2} B_R(T|m_t)dm_t := K_1.
\end{aligned}$$

Using the above equations, Equation (38) can be written as

$$\int_{-\beta}^{\beta} B_R(T|m_t)dm_t = K_1 + K_0 + K_{-1} = \frac{(1-\rho)^T \cdot \delta^T}{1-\delta^T} \cdot [K_{-1} - K_1]. \tag{39}$$

On the one hand, by integrating (36) for $m_t \in [c/2, \beta]$ we obtain

$$\begin{aligned}
K_{-1} =& \Delta \cdot \eta \cdot \frac{1-(\delta\eta)^T}{1-\delta\eta} \cdot \frac{1}{2} \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right)\\
&+ (1-(1-\rho)^T) \cdot \delta^T \cdot \left(\eta^T \cdot K_1 + (1-\eta^T) \cdot \frac{\beta - c/2}{2\beta} \cdot (K_1 + K_0 + K_{-1})\right)\\
&+ (1-\rho)^T \cdot \delta^T \cdot \left(\eta^T \cdot K_{-1} + (1-\eta^T) \cdot \frac{\beta - c/2}{2\beta} \cdot (K_0 + 2K_{-1})\right).
\end{aligned} \tag{40}$$

On the other hand, by integrating (36) for $m_t \in [-\beta, -c/2]$ we obtain

$$\begin{aligned}
K_1 =& -\Delta \cdot \eta \cdot \frac{1-(\delta\eta)^T}{1-\delta\eta} \cdot \frac{1}{2} \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right)\\
&+ (1-(1-\rho)^T) \cdot \delta^T \cdot \left(\eta^T \cdot K_{-1} + (1-\eta^T) \cdot \frac{\beta - c/2}{2\beta} \cdot (K_1 + K_0 + K_{-1})\right)\\
&+ (1-\rho)^T \cdot \delta^T \cdot \left(\eta^T \cdot K_{-1} + (1-\eta^T) \cdot \frac{\beta - c/2}{2\beta} \cdot (K_0 + 2K_{-1})\right).
\end{aligned} \tag{41}$$

Equations (39)–(41) yield

$$\int_{-\beta}^{\beta} B_R(T|m_0)dm_0 = \frac{(1-\rho)^T \cdot \delta^T}{1-\delta^T} \cdot \frac{\Delta \cdot \eta}{1-(\delta\eta)} \cdot \frac{1-(\delta\eta)^T}{1+(1-(1-\rho)^T)(\delta\eta)^T} \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right).$$

We thus have that for all $T, T' \in \mathbb{N}$,

$$\begin{aligned}
&[W^2 - W^1](T) - [W^2 - W^1](T')\\
=&\Delta \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right) \cdot \frac{\eta}{1-(\delta\eta)} \cdot \frac{1-(\delta\eta)^T}{1+(\delta\eta)^T(1-(1-\rho)^T)} \cdot \frac{(1-\rho)^T\delta^T}{1-\delta^T}\\
&-\Delta \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right) \cdot \frac{\eta}{1-(\delta\eta)} \cdot \frac{1-(\delta\eta)^{T'}}{1+(\delta\eta)^{T'}(1-(1-\rho)^{T'})} \cdot \frac{(1-\rho)^{T'}\delta^{T'}}{1-\delta^{T'}}.
\end{aligned}$$

Finally, define

$$H_0(T) := \Delta \cdot \left(\beta^2 - \left(\frac{c}{2}\right)^2\right) \cdot \frac{\eta}{1-(\delta\eta)} \cdot \frac{1-(\delta\eta)^T}{1+(\delta\eta)^T(1-(1-\rho)^T)} \cdot \frac{(1-\rho)^T\delta^T}{1-\delta^T}.$$

We claim (and show below) that for all $T \in \{1, 2, \ldots\}$,

$$H_0(T) \geq H_0(T+1).$$

The latter inequality can be arranged as

$$\underbrace{\left[\frac{1-\delta^T}{1-\delta^{T+1}}\right] \Big/ \left[\frac{1-(\delta\eta)^T}{1-(\delta\eta)^{T+1}}\right]}_{\chi_1} \cdot \underbrace{\delta(1-\rho)\frac{1+(1-(1-\rho)^T)(\delta\eta)^T}{1+(1-(1-\rho)^{T+1})(\delta\eta)^{T+1}}}_{\chi_2} \leq 1. \tag{42}$$

Clearly, $\chi_1, \chi_2 \geq 0$. Next we show that $\chi_1 \leq 1$ and $\chi_2 \leq 1$, which implies that Inequality (42) holds and hence our claim was correct. First, we show that

$$\chi_1 \leq 1. \tag{43}$$

For a given $T \in \{1, 2, \ldots\}$, define

$$f(x) := \frac{1-x^T}{1-x^{T+1}}. \tag{44}$$

Then note that

$$\begin{aligned} sgn\big(f'(x)\big) &= sgn\big(-Tx^{T-1}(1-x^{T+1}) - (1-x^T)(-(T+1)x^T)\big) \\ &= sgn\big(-T + (T+1)x - x^{T+1}\big) \leq 0, \end{aligned} \tag{45}$$

where the last inequality follows if we define

$$g(x) := -T + (T+1)x - x^{T+1},$$

and then note that

$$g(1) = -T + (T+1) - 1 = 0$$

and that for all $x \in (0, 1)$,

$$g'(x) = (T+1) - (T+1)x^T = (T+1)(1-x^T) > 0.$$

It then suffices to note that $\delta\eta \leq \delta < 1$ and

$$\chi_1 = \frac{f(\delta)}{f(\delta\eta)},$$

so we have showed Inequality (43).

Second, we show that

$$\chi_2 \leq 1. \tag{46}$$

For a given $T \in \{1, 2, \ldots\}$, define

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

One can easily verify that (46) is implied by

$$F(\rho) \leq 0 \text{ for all } \rho \in [0,1]. \tag{47}$$

Next, note that

$$F'(\rho) = -\delta(1+(\delta\eta)^T) + (1-\rho)^T(T+1)\delta(\delta\eta)^T(1-\eta)$$

and

$$F''(\rho) = -(1-\rho)^{T-1}T(T+1)\delta(\delta\eta)^T(1-\eta) \leq 0.$$

Let ρ^* be such that $F'(\rho^*) = 0$, which can be equivalently written as

$$\delta(1+(\delta\eta)^T) = (1-\rho)^*(T+1)\delta(\delta\eta)^T(1-\eta).$$

Clearly, ρ^* is a maximizer of F, but it might not necessarily belong to $[0,1]$. By means of simple algebra, one can verify that

$$F(\rho^*) = (1-\rho^*)\delta\frac{T}{T+1} - 1 + \underbrace{\delta(\delta\eta)^T\left((1-\rho^*)\frac{T}{T+1} - \eta\right)}_{G(\eta)}. \tag{48}$$

Then,

$$G'(\eta) = \delta^{T+1}\eta^{T-1}\left((1-\rho^*)\frac{T^2}{T+1} - (T+1)\eta\right).$$

Define η^* such that $G'(\eta^*) = 1$. That is,

$$\eta^* = (1-\rho^*)\left(\frac{T}{T+1}\right)^2. \tag{49}$$

Because $G(x) > 0$ if $x > 0$ is sufficiently close to zero and $G(x) < 0$ if x is large enough, then η^* must be the unique maximizer of G. This means that

$$G(\eta) \leq G(\eta^*) = \delta(\delta\eta)^T(1-\rho^*)\frac{T}{(T+1)^2}, \tag{50}$$

where the equality follows from (49). Finally, using (48) and (50), we obtain

$$\begin{aligned} F(\rho^*) &\leq (1-\rho^*)\delta\frac{T}{T+1}\frac{T+1+(\delta\eta^*)^T}{T+1} - 1 \leq (1-\rho^*)\delta\frac{T}{T+1}\frac{T+2}{T+1} - 1 \\ &= (1-\rho^*)\delta\frac{T^2+2T}{T^2+2T+1} - 1 \leq 0. \end{aligned}$$

This completes the proof of the proposition.

$\square$

Proof of Proposition 2

The notion of welfare considered is given by

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

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 and assuming that m_0 is drawn according to a uniform distribution on $[-\beta, \beta]$. 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

$$-\Delta \leq i_0 \leq \Delta$$

and that we have assumed without loss of generality that the first office-holder belongs to party R, so $i_1 = \Delta$ 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 $i \in \{-\Delta, \Delta\}$

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

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

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

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 the case 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's decisions, it must also be the case 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,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^1 = {} & (1-(1-\rho)^T) \cdot (1-\eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ & + (1-(1-\rho)^T) \cdot \eta^T \cdot \left[K^{-1} - 2c \cdot \Delta\right] \\ & + (1-\rho)^T \cdot (1-\eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ & + (1-\rho)^T \cdot \eta^T \cdot K^1. \end{aligned} \tag{51}$$

To understand Equation (51), 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,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^0 = {} & (1-(1-\rho)^T) \cdot (1-\eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ & + (1-(1-\rho)^T) \cdot \eta^T \cdot \left[K^0 - 2c \cdot \Delta\right] \\ & + (1-\rho)^T \cdot (1-\eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ & + (1-\rho)^T \cdot \eta^T \cdot K^0. \end{aligned} \tag{52}$$

Third,

$$\begin{aligned}\frac{1}{\delta^T} \cdot K^{-1} = {} & (1-(1-\rho)^T) \cdot (1-\eta^T) \cdot \left[P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 - 2c \cdot \Delta\right] \\ & + (1-(1-\rho)^T) \cdot \eta^T \cdot \left[K^1 - 2c \cdot \Delta\right] \\ & + (1-\rho)^T \cdot (1-\eta^T) \cdot \left[P \cdot (K^1 - 2c \cdot \Delta) + Q \cdot K^0 + P \cdot K^1\right] \\ & + (1-\rho)^T \cdot \eta^T \cdot [K^1 - 2c \cdot \Delta]. \end{aligned} \tag{53}$$

It is then a matter of algebra to verify that the linear system of equations made up of (51), (52) and (53) has a unique solution, which yields

$$\begin{aligned} EU^c(T) &= P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1 \\ &= -2c\Delta \cdot \frac{\delta^T}{1-\delta^T} \cdot \left[(1-(1-\rho)^T) + (1-\rho)^T \cdot P \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)} \right]. \end{aligned} \tag{54}$$

To find the above expression, define

$$W := P \cdot K^{-1} + Q \cdot K^0 + P \cdot K^1.$$

On the one hand, by adding (51)– (53), one obtains

$$W = \frac{\delta^T}{1-\delta^T} \cdot \left[W - 2c\Delta \cdot \left((1-(1-\rho)^T) + P \cdot (1-\rho)^T \right) + P \cdot (1-\rho)^T \cdot (K^1 - K^{-1}) \right]. \tag{55}$$

On the other hand, by subtracting (53) from (51), one obtains

$$K^1 - K^{-1} = \frac{(\delta\eta)^T(1-\rho)^T}{1+(\delta\eta)^T(1-(1-\rho)^T)} \cdot 2c\Delta. \tag{56}$$

One can easily verify that the combination of (55) and (56) yields (54).

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

$$\begin{aligned} EU^v(T) = \mathbb{E}_T\left[\sum_{t\geq 1} \delta^{t-1} \cdot a_t \right] &= \mathbb{E}_T\left[\sum_{t=1}^{T} \delta^{t-1} \cdot a_t \right] + \delta^T \cdot EU^v(T) \\ &= \sum_{t=1}^{T} \delta^{t-1} \cdot \mathbb{E}_T\left[a_t\right] + \delta^T \cdot EU^v(T), \end{aligned} \tag{57}$$

where for the latter inequality we have used that the expectation operator is linear. Additionally, we derive from the fact that a binomial distribution of parameters t and ρ has expected value equal to $\rho \cdot t$ that

$$\mathbb{E}_T\left[a_t\right] = -\rho A \cdot t.$$

By using the latter equality on Equation (57) we, in turn, obtain

$$\begin{aligned} EU^v(T) &= -A\rho \cdot \sum_{t=1}^{T} t \cdot \delta^{t-1} + \delta^T \cdot EU^v(T) \\ &= -A\rho \cdot \left(\frac{T\delta^T(\delta-1)+1-\delta^T}{(1-\delta)^2} \right) + \delta^T \cdot EU^v(T). \end{aligned} \tag{58}$$

The second inequality in (58) can be easily seen if we define

$$f(\delta) := \sum_{t=0}^{T} \delta^t = \frac{1-\delta^{T+1}}{1-\delta}$$

and note that

$$f'(\delta) = \sum_{t=1}^{T} t \cdot \delta^{t-1} = \frac{T\delta^T(\delta-1)+1-\delta^T}{(1-\delta)^2}.$$

Finally, (58) can be rearranged as

$$EU^v(T) = -\frac{\rho A}{(1-\delta)^2} + \rho A \cdot T \cdot \frac{\delta^T}{1-\delta^T} \cdot \frac{1}{1-\delta}. \tag{59}$$

Note that the first term of the right-hand side of Equation (59) is independent of T.

After the analysis of $EU^c(T)$ and $EU^V(T)$—see Equations (54) and (59)—, maximizing welfare with respect to T is equivalent to finding the value of $T \in \mathbb{N}$, say T^*, that maximizes the following expression:

$$\begin{aligned} M(T) := \frac{\delta^T}{1-\delta^T} \cdot \Bigg[& T \cdot \frac{A}{2c \cdot \left(\mu - \frac{c}{2(1+\theta)}\right)} \cdot \frac{\rho}{1-\delta} - (1-(1-\rho)^T) \\ & - (1-\rho)^T \cdot \frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\} \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)} \Bigg]. \end{aligned} \tag{60}$$

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.

$\square$

Proof of Propositions 3, 4, 5, and 6

Maximizing $W^1(T)$ is equivalent to maximizing

$$M(T) = G_I(T) - G_{II}(T) - G_{III}(T), \tag{61}$$

where

$$G_I(T) = \frac{\delta^T}{1-\delta^T} \cdot T \cdot \underbrace{\frac{A}{2c \cdot \left(\mu - \frac{c}{2} \cdot (1+\theta)\right)} \cdot \frac{\rho}{1-\delta}}_{:=\lambda_I(A,c,\mu,\theta,\rho)},$$

$$G_{II}(T) = (1-(1-\rho)^T) \cdot \frac{\delta^T}{1-\delta^T},$$

$$G_{III}(T) = (1-\rho)^T \cdot \frac{\delta^T}{1-\delta^T} \cdot \underbrace{\frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\}}_{:=\lambda_{III}(c,\beta)} \cdot \frac{1+\delta^T\eta^T(1-2(1-\rho)^T)}{1+\delta^T\eta^T(1-(1-\rho)^T)}.$$

To gain some insight and because the analysis has value in its own right, we start by proving that $G_I(T)$, $G_{II}(T)$, and $G_{III}(T)$ are positive, real-valued decreasing functions. This means, in particular, that function $G_I(T)$ is maximized for $T = 1$, while functions $-G_{II}(T)$ and $-G_{III}(T)$ are maximized for $T = \infty$. First,

$$\frac{G_I(T+1)}{G_I(T)} = \frac{T+1}{T} \cdot \frac{\delta - \delta^{T+1}}{1 - \delta^{T+1}} < 1,$$

where the inequality is explained as follows. For $\delta \in [0, 1)$, define

$$f(\delta) := \frac{\delta - \delta^{T+1}}{1 - \delta^{T+1}} = \frac{\delta + \ldots + \delta^T}{1 + \delta + \ldots + \delta^T}.$$

Clearly, $f(\delta)$ is increasing in δ.[65] Then it suffices to note that

$$\lim_{\delta \to 1^-} f(\delta) = \lim_{\delta \to 1^-} \frac{1 - (T+1) \cdot \delta^{T-1}}{-(T+1) \cdot \delta^T} = \frac{T}{T+1},$$

where the first equality follows from applying L'Hôpital's rule. Second, we know from (44) and (45) (see the proof of Proposition 1) that

$$\frac{1 - (1-\rho)^{T+1}}{1 - (1-\rho)^T}$$

is increasing in $1 - \rho$, with $\rho \in [0, 1]$. Moreover,

$$\lim_{\rho \to 0^+} \frac{1 - (1-\rho)^{T+1}}{1 - (1-\rho)^T} = \lim_{\rho \to 0^+} \frac{-(T+1) \cdot (1-\rho)^{T-1}}{-T \cdot (1-\rho)^T} = \frac{T+1}{T},$$

where the first equality follows from applying L'Hôpital's rule. Hence,

$$\frac{G_{II}(T+1)}{G_{II}(T)} = \frac{\delta - \delta^{T+1}}{1 - \delta^{T+1}} \cdot \frac{1 - (1-\rho)^{T+1}}{1 - (1-\rho)^T} < 1.$$

Third,

$$\frac{G_{III}(T+1)}{G_{III}(T)} = \underbrace{\frac{\delta - \delta^{T+1}}{1 - \delta^{T+1}}}_{:=\phi_1} \cdot \underbrace{\frac{1 + (\delta\eta)^{T+1}(1 - 2(1-\rho)^{T+1})}{1 + (\delta\eta)^T(1 - 2(1-\rho)^T)}}_{:=\phi_2} \cdot \underbrace{(1-\rho) \cdot \frac{1 + (\delta\eta)^T(1 - (1-\rho)^T)}{1 + (\delta\eta)^{T+1}(1 - (1-\rho)^{T+1})}}_{:=\phi_3} < 1,$$

where the inequality holds because, as shown before,

$$\phi_1 \leq \frac{T}{T+1}, \tag{62}$$

while we also have (and show next) that

$$\phi_2 \leq \frac{T+1}{T}, \tag{63}$$

[65] Note that $x/(1+x)$ is increasing in x and $\delta + \ldots \delta^T$ is increasing in δ.

and

$$\phi_3 \leq 1. \tag{64}$$

On the one hand, it is a matter of simple algebra to see that (63) is equivalent to

$$f(\delta\eta, \rho) := (\delta\eta)^T \cdot \Big((\delta\eta)T(1 - 2(1-\rho)^{T+1}) - (T+1)(1 - 2(1-\rho)^T)\Big) \leq 1. \tag{65}$$

One can then verify that

$$\begin{aligned}
\frac{\partial f(\delta\eta, \rho)}{\partial \rho} &= (\delta\eta)^T \cdot \Big((\delta\eta)2T(T+1)(1-\rho)^T - 2(T+1)T(1-\rho)^{T-1}\Big) \\
&= (\delta\eta)^T \cdot 2T(T+1)(1-\rho)^{T-1} \cdot \Big((\delta\eta)(1-\rho) - 1\Big) \leq 0.
\end{aligned}$$

Hence,

$$f(\delta\eta, \rho) \leq f(\delta\eta, 0) = (\delta\eta)^T \cdot (T + 1 - (\delta\eta)T). \tag{66}$$

Finally,

$$\frac{\partial f(\delta\eta, 0)}{\partial \delta\eta} = (\delta\eta)^{T-1}T(T+1) - (\delta\eta)^T(T+1)T = (\delta\eta)^{T-1}T(T+1)(1-\delta\eta) \geq 0,$$

which implies

$$f(\delta\eta, 0) \leq f(1, 0) = 0. \tag{67}$$

Inequalities (66) and (67) imply (65), and hence (63). On the other hand, (64) is equivalent to

$$f(\delta\eta, \rho) := (1-\rho) \cdot \Big(\big(1 + (\delta\eta)^T\big) - (\delta\eta)^T(1-\rho)^T(1-\delta\eta)\Big) - \big(1 + (\delta\eta)^{T+1}\big) \leq 0. \tag{68}$$

Then

$$\frac{\partial f(\delta\eta, \rho)}{\partial \rho} = -\big(1 + (\delta\eta)^T\big) + (T+1)(\delta\eta)^T(1-\rho)^T(1-\delta\eta)$$

and

$$\frac{\partial^2 f(\delta\eta, \rho)}{\partial \rho^2} = -(T+1)T(\delta\eta)^T(1-\rho)^{T-1}(1-\delta\eta) \leq 0.$$

Let ρ^* such that

$$\frac{\partial f(\delta\eta, \rho^*)}{\partial \rho} = 0,$$

i.e.,

$$(1 + (\delta\eta)^T) = (T+1)(\delta\eta)^T(1-\rho^*)^T(1-\delta\eta).$$

We claim (and show next) that

$$g(\delta\eta) := (1 + (\delta\eta)^T) - (T+1)(\delta\eta)^T(1-\delta\eta) > 0. \tag{69}$$

Indeed, by simple algebraic manipulations

$$g'(\delta\eta) = -(\delta\eta)^{T-1}(T^2 - (T+1)^2(\delta\eta)).$$

This means that

$$g(\delta\eta) \geq \min\left\{g(0), g\left(\left(\frac{T}{T+1}\right)^2\right), g(1)\right\} \geq \min\left\{1, 1 - \left(\frac{T}{T+1}\right)^2 \frac{T}{T+1}, 2\right\} > 0.$$

Hence the claim in (69) was correct. Finally, (69) implies that

$$1 - \rho^* > 1,$$

and thus

$$\frac{\partial f(\delta\eta, \rho)}{\partial \rho}$$

is monotone for $\rho \in [0, 1]$. Accordingly,

$$f(\delta\eta, \rho) \leq \max\{f(\delta\eta, 0), f(\delta\eta, 1)\} = \left(1 + (\delta\eta)^T\right) - (\delta\eta)^T(1 - \delta\eta) - \left(1 + (\delta\eta)^{T+1}\right) = 0.$$

This implies (68).

We turn now to proving the different statements from the propositions. We proceed in several steps. Before we do so, we make one important remark that will be applied throughout the remainder of the proof. Let k be one model parameter on which $M(T)$ depends. We make the dependence explicit at times by writing $M(T, k)$ (and similarly for other functions). Then, T^* weakly increases as we increase parameter k if and only if

$$\frac{\partial M(T+1, K)}{\partial k} \geq \frac{\partial M(T, K)}{\partial k}. \tag{70}$$

This is the translation to our setup of the standard supermodularity condition for (T, k) (see e.g. Topkis, 1998; Milgrom and Shannon, 1994). Similarly, T^* weakly decreases as we increase parameter k if and only if

$$\frac{\partial M(T+1, K)}{\partial k} \leq \frac{\partial M(T, K)}{\partial k}. \tag{71}$$

We are now in a position to prove our comparative statics results. First, one can easily verify that

$$\frac{\partial M(T+1, A)}{\partial A} \leq \frac{\partial M(T, A)}{\partial A} \Leftrightarrow \frac{\partial M(T+1, \theta)}{\partial \theta} \leq \frac{\partial M(T, \theta)}{\partial \theta} \Leftrightarrow G_I(T+1) \leq G_I(T).$$

Similarly,

$$\frac{\partial M(T+1, \mu)}{\partial \mu} \geq \frac{\partial M(T, \mu)}{\partial \mu} \Leftrightarrow G_I(T+1) \leq G_I(T).$$

We have proved above that $G_I(T)$ is decreasing in T. Hence, according to (70) and (71), the above two expressions imply that optimal term-length T^* can never increase if A or θ increase or μ decreases.

Second, one can as well verify that

$$\frac{\partial M(T+1,\beta)}{\partial \beta} \geq \frac{\partial M(T,\beta)}{\partial \beta} \Leftrightarrow G_{III}(T+1) \leq G_{III}(T).$$

We have proved above that $G_{III}(T)$ is decreasing in T. This implies that optimal term-length T^* can never decrease if β increases.

Third, we focus on changes in η. Clearly,

$$sgn\left(\frac{\partial M(T+1,\eta)}{\partial \eta} - \frac{\partial M(T,\eta)}{\partial \eta}\right) = -sgn\left(\frac{\partial G_{III}(T+1,\eta)}{\partial \eta} - \frac{\partial G_{III}(T,\eta)}{\partial \eta}\right). \tag{72}$$

Then it is a matter of simple algebra to verify that

$$\begin{aligned} &\frac{\partial G_{III}(T+1,\eta)}{\partial \eta} \geq \frac{\partial G_{III}(T,\eta)}{\partial \eta} \\ \Leftrightarrow &\frac{T+1}{T} \cdot \frac{\delta - \delta^{T+1}}{1-\delta^{T+1}} \cdot (\delta\eta) \cdot \left((1-\rho) \cdot \frac{1+(\delta\eta)^T(1-(1-\rho)^T)}{1+(\delta\eta)^{T+1}(1-(1-\rho)^{T+1})}\right)^2 \leq 1. \end{aligned} \tag{73}$$

Using (62) and (64) together with the fact that $\delta\eta < 1$, we obtain

$$\frac{T+1}{T} \cdot \frac{\delta - \delta^{T+1}}{1-\delta^{T+1}} \cdot (\delta\eta) \cdot \left((1-\rho) \cdot \frac{1+(\delta\eta)^T(1-(1-\rho)^T)}{1+(\delta\eta)^{T+1}(1-(1-\rho)^{T+1})}\right)^2 < 1. \tag{74}$$

Accordingly, (71)–(74) imply that optimal term-length T^* can never increase if η increases.

Fourth, consider a change of c. This affects $G_I(T)$ (through $\lambda_I(A,c,\mu,\theta,\rho)$) and potentially $G_{III}(T)$ (through $\lambda_{III}(c,\beta)$). Specifically, note that

$$sign\left(\frac{\partial \lambda_I(A,c,\mu,\theta,\rho)}{\partial c}\right) = -sign\left(\frac{\partial}{\partial c}\left(c \cdot \left(\mu - \frac{c}{2} \cdot (1+\theta)\right)\right)\right) = \begin{cases} < 0 & \text{if } c < \frac{\mu}{1+\theta}, \\ = 0 & \text{if } c = \frac{\mu}{1+\theta}, \\ > 0 & \text{if } c > \frac{\mu}{1+\theta}, \end{cases} \tag{75}$$

and

$$\frac{\partial \lambda_{III}(c,\beta)}{\partial c} = \frac{\partial}{\partial \beta}\left(\frac{1}{2\beta} \cdot \max\left\{0, \beta - \frac{c}{2}\right\}\right) = \begin{cases} < 0 & \text{if } c < 2\beta, \\ = 0 & \text{if } c \geq 2\beta. \end{cases} \tag{76}$$

We consider two distinct cases. On the one hand, assume that

$$2\beta < c < \frac{\mu}{1+\theta}.$$

That is, polarization is large and preference shocks are small. Using (75), one can easily verify that

$$\frac{\partial M(T+1,c)}{\partial c} \geq \frac{\partial M(T,c)}{\partial c} \Leftrightarrow G_I(T+1) \leq G_I(T).$$

In this case, (70) implies that increasing c yields a (weakly) higher T^*. On the other hand, assume that

$$\frac{\mu}{1+\theta} < c < 2\beta.$$

That is, polarization is low and preference shocks are large. Using (75) and (76), one can verify that

$$\frac{\partial G_I(T+1,c)}{\partial c} \leq \frac{\partial G_I(T,c)}{\partial c} \Leftrightarrow G_I(T) \geq G_I(T+1)$$

and

$$\frac{\partial G_{III}(T+1,c)}{\partial c} \geq \frac{\partial G_{III}(T,c)}{\partial c} \Leftrightarrow G_{III}(T) \geq G_{III}(T+1),$$

respectively. Given that $G_I(T)$ and $G_{III}(T)$ are decreasing functions, we obtain from (61) that

$$\frac{\partial M(T+1,c)}{\partial c} = \frac{\partial G_I(T+1,c)}{\partial c} - \frac{\partial G_{III}(T+1,c)}{\partial c} \leq \frac{\partial G_I(T,c)}{\partial c} - \frac{\partial G_{III}(T,c)}{\partial c} = \frac{\partial M(T,c)}{\partial c}.$$

In this case, (71) implies that increasing c yields a (weakly) lower T^*.

Fifth and last, consider a change in ρ. This has several effects on $G_I(T,\rho)$, $G_{II}(T,\rho)$ and $G_{III}(T,\rho)$. We start by noting that we can write

$$G_{II}(T,\rho) + G_{III}(T,\rho) = \frac{\delta^T}{1-\delta^T} \cdot \left(1 - (1-\rho)^T \cdot \left(1 - P \cdot \left(1 - \frac{(\delta\eta)^T(1-\rho)^T}{1+(\delta\eta)^T(1-(1-\rho)^T)}\right)\right)\right)$$

where we have used

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

Then it is a matter of algebra to verify that

$$\begin{aligned}
&\frac{\partial G_{II}(T,\rho) + G_{III}(T,\rho)}{\partial \rho} \\
=&\frac{\delta^T}{1-\delta^T} \cdot T(1-\rho)^{T-1} \cdot \left(1 - P \cdot \frac{1+(\delta\eta)^T(1-2(1-\rho)^T)}{1+(\delta\eta)^T(1-(1-\rho)^T)} + P \cdot \frac{(\delta\eta)^T(1-\rho)^T(1+(\delta\eta)^T)}{(1+(\delta\eta)^T(1-(1-\rho)^T))^2}\right) \\
=&\frac{\delta^T}{1-\delta^T} \cdot T(1-\rho)^{T-1} \cdot \left(1 - P \cdot f(T)^2 + 2P \cdot (1-f(T))^2\right), \qquad (77)
\end{aligned}$$

where

$$f(T,\rho,\delta) := \frac{1+(\delta\eta)^T(1-2(1-\rho)^T)}{1+(\delta\eta)^T(1-(1-\rho)^T)}.$$

We stress that

$$0 \leq f(T,\rho,\delta) \leq 1.$$

Let now

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

Using Equations (61) and (77), we obtain

$$\frac{\partial M(T,\rho)}{\partial \rho} = \frac{\delta^T}{1-\delta^T} \cdot T \cdot \left(\frac{\lambda}{1-\delta} - (1-\rho)^T \cdot (1 - P \cdot f(T,\rho,\delta)^2 + 2P \cdot (1 - f(T,\rho,\delta))^2 \right),$$

so

$$\frac{\frac{\partial M(T+1,\rho)}{\partial \rho}}{\frac{\partial M(T,\rho)}{\partial \rho}} \tag{78}$$

$$= \underbrace{\frac{\delta - \delta^{T+1}}{1-\delta^{T+1}} \cdot \frac{T+1}{T} \cdot \frac{\frac{\lambda}{1-\delta} - (1-\rho)^{T+1} \cdot (1 - P \cdot f(T+1,\rho,\delta)^2 + 2P \cdot (1 - f(T+1,\rho,\delta))^2}{\frac{\lambda}{1-\delta} - (1-\rho)^T \cdot (1 - P \cdot f(T,\rho,\delta)^2 + 2P \cdot (1 - f(T,\rho,\delta))^2}}_{:=\psi(\delta)}.$$

Moreover, it is easy to verify that

$$g(x) := 1 - P \cdot x^2 + 2P \cdot (1-x)^2$$

is decreasing in x for $x \in [0,1]$, which implies

$$1 - P \cdot f(T+1,\rho,\delta)^2 + 2P \cdot (1 - f(T+1,\rho,\delta))^2 \geq g(1) = 1 - P$$

and

$$1 - P \cdot f(T,\rho,\delta)^2 + 2P \cdot (1 - f(T,\rho,\delta))^2 \leq g(0) = 1 + 2P.$$

Provided that $\delta < 1$ is sufficiently large, the above two inequalities imply that

$$0 \leq \psi(\delta) \leq \frac{\delta - \delta^{T+1}}{1-\delta^{T+1}} \cdot \frac{T+1}{T} \cdot \frac{\lambda - \overbrace{(1-\rho)^{T+1}(1-P)}^{:=\kappa_1}(1-\delta)}{\lambda - \underbrace{(1-\rho)^T(1+2P)}_{:=\kappa_2}(1-\delta)} := h(\delta). \tag{79}$$

Consider now the following two Taylor expansions around $\delta = 1$, namely

$$\frac{\delta - \delta^{T+1}}{1-\delta^{T+1}} = \frac{T}{T+1} + \frac{T}{2} \cdot (\delta - 1) + \sum_{k \geq 2} a_k \cdot (\delta - 1)^k$$

and

$$\frac{\lambda - \kappa_1 \cdot (1-\delta)}{\lambda - \kappa_2 (1-\delta)} = 1 - (\kappa_2 - \kappa_1) \cdot (\delta - 1) + \sum_{k \geq 2} b_k \cdot (\delta - 1)^k.$$

Using the two above Taylor expansions, we obtain

$$h(\delta) = 1 + \left(\frac{T+1}{2} - (\kappa_2 - \kappa_1) \right) (\delta - 1) + \sum_{k \geq 2} c_k \cdot (\delta - 1)^k.$$

Hence,

$$h(1) = 1 \tag{80}$$

and

$$h'(\delta) = \frac{T+1}{2} - (\kappa_2 - \kappa_1) + \sum_{k\geq 1} d_k \cdot (\delta - 1)^k.$$

In particular,

$$\lim_{\delta\to 1^-} h'(\delta) = \frac{T+1}{2} - (\kappa_2 - \kappa_1) > 0, \tag{81}$$

where the inequality holds if

$$P < \frac{1}{3}.$$

Finally, if we use Equations (78)–(81), we obtain that if δ is sufficiently large,

$$\frac{\partial M(T+1,\rho)}{\partial\rho} \leq \frac{\partial M(T,\rho)}{\partial\rho}.$$

That is, (71) implies that increasing ρ yields a (weakly) lower T^*. It is worth noting that this result holds much more generally, even if δ is low or $P > 1/3$. For instance, one can easily verify that $h(\delta) < 1$ if δ is sufficiently low and $\lambda > 1$ regardless of the value of P, or if $\rho > 2/3$.

□

Appendix B (for online publication online)

It remains to investigate whether there can be another equilibrium than the one described in Theorem 2. To this end, we proceed in four steps. 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 already suffered by the incumbent, all else being equal. This is formalized next.

Assumption 1

Let $i_t \in \mathbb{R}$ be the policy chosen by some incumbent k in some period t. Then,

$$i_t > 0 \quad \text{if } k \in R$$

and

$$i_t < 0 \quad \text{if } k \in L.$$

Assumption 2

Let $i_t \in \mathbb{R}$ ($i'_t \in \mathbb{R}$) be the policy chosen by some incumbent k in some period t if s/he has suffered (not suffered) a valence shock in period $t-1$. Then

$$i_t \leq i'_t \quad \text{if } k \in R$$

and

$$i_t \geq i'_t \quad \text{if } k \in L.$$

Also for simplicity, we impose a certain notion of monotonicity for strategies, namely:

Assumption 3

Let $i_t, i_{t+1}, \ldots \in \mathbb{R}$ be all the policies chosen by the same incumbent during his/her tenure. Then either

$$i_{t'} \geq i_{t'-1} \text{ for all } t' \in \{t+1, \ldots\}$$

or

$$i_{t'} \leq i_{t'-1} \text{ for all } t' \in \{t+1, \ldots\}.$$

That is, either incumbents choose along their tenure policies that are (weakly) ever more extreme or (weakly) ever more moderate. Note that this monotonicity property excludes the status-quo policy in place before the incumbent chooses a policy in his/her period in office. Tenure may entail one or several terms. Assumption 3 rules out the possibility that incumbents (but not parties via different incumbents) choose flip-flop policies.

Assumptions 1–2 facilitate the analysis but can be dispensed with.[66] Moreover, since the equilibrium defined in Theorem 2 satisfies the three assumptions, existence of equilibria is guaranteed if we impose them, which can therefore be seen as properties of equilibria rather than restrictions of the action space. We also assume for simplicity that the random process $dF(m|m'_-)$ determining the median voter's peak in each period from that of the previous period has compact support, say $[-\beta, \beta]$, and that for every period t, there is positive probability that each element $m_t \in [-\beta, \beta]$ can be reached from m_{t-1} regardless of the value of the latter peak. Moreover, $dF(m|m_-$ first-order stochastically dominates $dF(m|m'_-)$ if $m_- > m'_-$. For instance, this is the case considered for our analysis of Section 6.

We focus on a term that starts at period $t+1$ and finishes at period $t+T$. As for notation, we use the same shortcuts as in the proof of Theorem 2. In particular, j stands for the status quo policy in place before the office-holder chooses a policy for the current period.

Step 1

We start by considering the median voter's decision in the election that takes place in period $t+T$, in which the incumbent $k \in R$, who has suffered z shocks until the beginning of this period, chooses policy i. We denote by $\Delta^z := \Delta^z(i) > 0$ the policy that the incumbent will choose in period $t+T+1$ if s/he is re-elected and has not suffered a valence shock in period $t+T$. Similarly, we denote by $\Delta^{z+1} := \Delta^{z+1}(i) > 0$ the policy that s/he will choose in period $t+T+1$ if s/he is re-elected and has suffered a valence shock in period $t+T$. By Assumption 2, it must be the case that $\Delta^{z+1} \leq \Delta^z$. In turn, assume that the challenger $k \in L$ will choose some $-\Delta^0 = -\Delta^0(i) < 0$ 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

$$p(i, z_+, m_-) \tag{82}$$

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, before the median voter's peak m is determined according to $F(\cdot|m_-)$. We distinguish four cases.

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

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

[66] Although Assumption 3 is in principle stronger, some of the main mechanics of our proof do not rely on this property. Hence, Assumption 3 could potentially also be dispensed with.

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

$$m \geq \frac{c}{2}\cdot\frac{\Delta^z-\Delta^0-2i}{\Delta^z+\Delta^0} + \frac{A\cdot z}{2(\Delta^z+\Delta^0)} + \frac{\Delta^z-\Delta^0}{2}. \tag{83}$$

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, provided that Δ^z and $-\Delta^0$ do not depend on i as long as the conditions of Case I are satisfied (see Assumption 3 and Remarks 1 and 2 below).

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

$$m \geq \frac{c}{2}\cdot\frac{\Delta^{z+1}-\Delta^0-2i}{\Delta^{z+1}+\Delta^0} + \frac{A\cdot(z+1)}{2(\Delta^{z+1}+\Delta^0)} + \frac{\Delta^{z+1}-\Delta^0}{2}. \tag{84}$$

Using the above expression, one can see that $p(i,z+1,m_-)$ is non-decreasing in i, provided that Δ^{z+1} and $-\Delta^0$ do not depend on i as long as the conditions of Case I are satisfied (see Assumption 3 and Remarks 1 and 2 below).

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

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

$$m \geq -\frac{c}{2} + \frac{Az}{2(\Delta^z+\Delta^0)} + \frac{\Delta^z-\Delta^0}{2}. \tag{85}$$

Using the above expression, one can see that $p(i,z,m_-)$ is constant in i, provided that Δ^z and $-\Delta^0$ do not depend on i as long as the conditions of Case II are satisfied (see Assumption 3 and Remarks 1 and 2 below). This means that in such a case, 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 is 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

$$m \geq -\frac{c}{2} + \frac{A(z+1)}{2(\Delta^{z+1}+\Delta^0)} + \frac{\Delta^{z+1}-\Delta^0}{2}. \tag{86}$$

Using the above expression, one can see that $p(i, z+1, m_-)$ is constant in i, provided that Δ^{z+1} and $-\Delta^0$ do not depend on i as long as the conditions of Case II are satisfied (see Assumption 3 and Remarks 1 and 2 below). This means that in such a case, 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 is re-elected.

Case III: $\Delta^{z+1} < i < \Delta^z$

This case cannot occur on the equilibrium path due to Assumption 3. Yet it is worth noting what would happen off the equilibrium path. On the one hand, if $a_{kT} = -z\cdot A$, the median voter will re-elect k if and only if (85) holds. On the other hand, if $a_{kT} = -(z+1)\cdot A$, the median voter will re-elect k if and only if (84) holds.

Case IV: $i < -\Delta^0$

For Case IV, we distinguish two subcases.

Case IV.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

$$m \geq \frac{c}{2} + \frac{Az}{2(\Delta^z+\Delta^0)} + \frac{\Delta^z-\Delta^0}{2}. \tag{87}$$

Using the above expression, one can see that $p(i, z, m_-)$ is constant in i, provided that Δ^z and $-\Delta^0$ do not depend on i as long as the conditions of Case IV are satisfied (see Assumption 3 and Remarks 1 and 2 below).

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

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

$$m \geq \frac{c}{2} + \frac{A(z+1)}{2(\Delta^{z+1} + \Delta^0)} + \frac{\Delta^{z+1} - \Delta^0}{2}. \tag{88}$$

Using the above expression, one can see that $p(i, z+1, m_-)$ is constant in i, provided that Δ^{z+1} and $-\Delta^0$ do not depend on i as long as the conditions of Case IV are satisfied (see Assumption 3 and Remarks 1 and 2 below).

For simplicity, we shall assume henceforth that $F(\cdot)$ is such that $p(i, z, m_-)$ and $p(i, z+1, m_-)$ are not only continuous but also differentiable for all $i \in \mathbb{R}$. If these probabilities were not differentiable, 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 $i \in \mathbb{R}$.

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, before the median voter's peak will be determined, and before elections will take place, all in the same period $t+T$. In this case, the incumbent faces the following problem:

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

where ξ is independent of i. A first immediate remark is in order.[67]

Remark 1

The maximizer of $G(i)$ does not depend on j if we restrict either to

$$\{i \in \mathbb{R} : i \leq j\}$$

or to

$$\{i \in \mathbb{R} : i \geq j\}.$$

[67] Note that we are using the fact that strategies must be Markov.

Moreover, the maximizer does not change if we keep z constant.

We can rearrange terms to obtain

$$\begin{aligned} G(i) &= -(i-\mu)^2 - c \cdot |i-j| - \theta \cdot c \cdot \left|i + \Delta^0\right| - \xi' \\ &+ \theta \cdot (1-\rho) \cdot p\left(i, z, m_-\right) \cdot \left[M^z - c \cdot |\Delta^z - i| + c \cdot \left|i + \Delta^0\right|\right] \\ &+ \theta \cdot \rho \cdot p\left(i, z+1, m_-\right) \cdot \left[M^{z+1} - c \cdot \left|\Delta^{z+1} - i\right| + c \cdot \left|i + \Delta^0\right|\right], \end{aligned} \tag{89}$$

where ξ' 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, so are M^z and M^{z+1}. We have used the shortcuts $\Delta^z = \Delta^z(i)$, $\Delta^{z+1} = \Delta^{z+1}(i)$, and $\Delta^0 = \Delta^0(i)$. Finally, we assume that $G(i)$ is continuous for all $i \in \mathbb{R}$ and differentiable for all $i \in \mathbb{R}$, except possibly in a finite number of points. We distinguish several cases, but we focus on $i \geq 0$ (see Assumption 1).

Case I: $i < j$

By Assumption 1 and the fact the incumbents are only ousted by the electorate, it must be the case that the status-quo policy j was chosen in period $t + T - 1$ by the same incumbent who is in office in period $t + T$. By Assumptions 1 and 3, it must then be the case that

$$\Delta^z \leq \Delta^{z+1} \leq i < j.$$

Remarks 1 and 2 (see below in Step 3) then imply that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of policy i. Using (89), we have

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

Then

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

where the inequality holds because $p\,(i, z+1, m_-)$ and $p\,(i, z+1, m_-)$ are non-decreasing and M^z and M^{z+1} are very large. To sum up,

$$
i < \mu + \frac{c}{2} \cdot (1-\theta) \Rightarrow G'(i) > 0. \tag{90}
$$

for i satisfying

$$
i < j \tag{91}
$$

to be chosen as part of an equilibrium it cannot be that $G'(i) < 0$. A *sufficient* condition for $G'(i) \geq 0$ is

$$
i \leq \mu + \frac{c}{2} \cdot (1-\theta). \tag{92}
$$

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

By Assumption 3 and Remarks 1 and 2 (see below in Step 3), we can assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. In this case, we can write

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

Hence,

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

Therefore,

$$
i < \mu - \frac{c}{2} \cdot (1+\theta) = \Delta \Rightarrow G'(i) > 0. \tag{93}
$$

Note that for i satisfying

$$
j \leq i \leq \Delta^{z+1} \leq \Delta^{z} \tag{94}
$$

to be chosen as part of an equilibrium it cannot be that $G'(i) < 0$. A *sufficient* condition for $G'(i) \geq 0$ is

$$
i \leq \mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot \left((1-\rho) \cdot p\left(i, z, m_{-}\right) + \rho \cdot p\left(i, z+1, m_{-}\right)\right). \tag{95}
$$

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

By Assumption 3, it must be the case that in equilibrium

$$
i = \Delta^{z+1} \quad \text{ or } i = \Delta^{z}.
$$

Remarks 1 and 2 (see below in Step 3) then enable us to assume that Δ^{z}, Δ^{z+1}, and $-\Delta^{0}$ are independent of i. Therefore, we can write

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

Hence,

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

Therefore,

$$
i < \mu - \frac{c}{2} \cdot (1+\theta) \Rightarrow G'(i) > 0. \tag{96}
$$

Note that for i to be part of an equilibrium it cannot be that $G'(i) > 0$. A *sufficient* condition for $G'(i) \leq 0$ is

$$
i \geq \mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot \Big((1-\rho) \cdot p(i,z,m_-) \Big). \tag{97}
$$

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

By Assumption 3 and Remarks 1 and 2 (see below in Step 3), we can assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. In this case, we can write

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

Hence,

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

In particular, $G'(i)$ is independent of Δ^z and Δ^{z+1}, as well as of m_-. Moreover,

$$
\begin{aligned}
i > \mu - \frac{c}{2} \cdot (1+\theta) &\Leftrightarrow G'(i) < 0, \\
i < \mu - \frac{c}{2} \cdot (1+\theta) &\Leftrightarrow G'(i) > 0.
\end{aligned} \tag{98}
$$

Note that for i satisfying

$$
j \leq i \text{ and } \Delta^{z+1} \leq \Delta^z \leq i \tag{99}
$$

to be chosen as part of an equilibrium it cannot be that $G'(i) > 0$. A *sufficient* condition for $G'(i) \leq 0$ is

$$i \geq \mu - \frac{c}{2} \cdot (1+\theta) = \Delta. \tag{100}$$

Step 3

Next, 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 the case that $T > 1$. Then the incumbent faces the following problem in a particular period $h \in \{t+1, \ldots, t+T-1\}$:

$$\max_{i \in I} H(i) := \max_{i \in I} \left\{ -(i-\mu)^2 - c \cdot |j-i| - \theta c \cdot \left[(1-\rho) \cdot |\Delta^z - i| + \rho \cdot |\Delta^{z+1} - i| \right] - \xi' \right\}, \tag{101}$$

where ξ' is independent of i. Note that $H(i)$ does not depend on m_-. A second immediate remark is the following:[68]

Remark 2

The maximizer of $H(i)$ does not depend on j if we restrict either to

$$\{i \in \mathbb{R} : i \leq j\}$$

or to

$$\{i \in \mathbb{R} : i \geq j\}.$$

Moreover, the maximizer is independent of z.

We assume that in period $h+1$, the incumbent will choose $\Delta^z = \Delta^z(i)$ if s/he has not received a valence shock in period h (which happens with probability $1-\rho$) or $\Delta^{z+1} = \Delta^{z+1}(i)$ 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 Assumption 1).

Case I: $i < j$

By Assumption 1 and the fact the incumbents are only ousted by the electorate, it must be the case that the status-quo policy j was chosen in period $t+T-1$ by the same incumbent who is in office in period $t+T$. By Assumptions 1 and 3, it must be the case that

$$\Delta^z \leq \Delta^{z+1} \leq i < j.$$

[68] Note that we are using the fact that strategies must be Markov.

Remarks 1 and 2 then enable us to assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. Therefore, we have

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

It follows that

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

In particular, $H'(i)$ is independent of Δ^z and Δ^{z+1}. Moreover,

$$i < \mu + \frac{c}{2}\cdot(1-\theta) \Leftrightarrow H'(i) > 0. \tag{102}$$

Note that for i for i satisfying

$$i < j \tag{103}$$

to be chosen as part of an equilibrium it cannot be that $H'(i) < 0$. A *sufficient* condition for $H'(i) \leq 0$ is

$$i \leq \mu + \frac{c}{2}\cdot(1-\theta). \tag{104}$$

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

By Assumption 3 and Remarks 1 and 2, we can assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. Therefore, we have

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

It follows that

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

In particular, $H'(i)$ is independent of Δ^z and Δ^{z+1}. Moreover,

$$\begin{aligned} i < \mu - \frac{c}{2}\cdot(1-\theta) &\Leftrightarrow H'(i) > 0, \\ i > \mu - \frac{c}{2}\cdot(1-\theta) &\Leftrightarrow H'(i) < 0. \end{aligned} \tag{105}$$

Note that for i satisfying

$$j \leq i \leq \Delta^{z+1} \leq \Delta^z \tag{106}$$

to be chosen as part of an equilibrium it cannot be that $H'(i) < 0$. A *sufficient* condition for $H'(i) \geq 0$ is

$$i \leq \mu - \frac{c}{2}\cdot(1-\theta). \tag{107}$$

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

By Assumption 3, it must be the case that in equilibrium

$$i = \Delta^{z+1} \quad \text{or} \quad i = \Delta^z.$$

Remarks 1 and 2 then enable us to assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. Therefore, we have

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

It follows that

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

In particular, $H'(i)$ is independent of Δ^z and Δ^{z+1}. Moreover,

$$\begin{aligned} i &< \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] \Leftrightarrow H'(i) > 0, \\ i &> \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)] \Leftrightarrow H'(i) < 0. \end{aligned} \tag{108}$$

On the one hand, assume that $i = \Delta^z$. Then for i to be chosen as part of an equilibrium it cannot be that $H'(i) < 0$. A *necessary* and *sufficient* condition for $H'(i) \geq 0$ is

$$i \leq \mu - \frac{c}{2} \cdot [1 + \theta \cdot (2\rho - 1)]. \tag{109}$$

On the other hand, assume that $i = \Delta^{z+1}$. Then for i to be chosen as part of an equilibrium it cannot be that $H'(i) > 0$. A *necessary* and *sufficient* condition for $H'(i) \leq 0$ is

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

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

By Assumption 3 and Remarks 1 and 2, we can assume that Δ^z, Δ^{z+1}, and $-\Delta^0$ are independent of i. Therefore, we have

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

It follows that

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

In particular, $H'(i)$ is independent of Δ^z and Δ^{z+1}. Moreover,

$$\begin{aligned} i &> \mu - \frac{c}{2} \cdot (1 + \theta) \Leftrightarrow H'(i) < 0, \\ i &< \mu - \frac{c}{2} \cdot (1 + \theta) \Leftrightarrow H'(i) > 0. \end{aligned} \tag{110}$$

Note that for i satisfying

$$j \leq i \text{ and } \Delta^{z+1} \leq \Delta^{z} \leq i \tag{111}$$

to be chosen as part of an equilibrium it cannot be that $H'(i) > 0$. A *necessary* and *sufficient* condition for $H'(i) \leq 0$ is

$$i \geq \mu - \frac{c}{2} \cdot (1 + \theta). \tag{112}$$

Step 4

Finally, let z, m_{-}, and j be such that when i is chosen according to the (additional) equilibrium by incumbent $k \in R$, we have

$$i \neq \Delta = \mu - \frac{c}{2} \cdot (1 + \theta). \tag{113}$$

We assume that i is the first policy chosen along the equilibrium path by any incumbent from party R to display the property in (113). Together with (3) and Assumption 1, this means that

$$j \leq i. \tag{114}$$

Note that due to Assumption 1, we can further focus without loss of generality on the case where party R deviates from the equilibrium defined in Theorem 2. We distinguish two cases.

Case A: $i < \Delta$

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

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

Based on the analysis of Step 2, we will show that i cannot be the optimal choice by the incumbent k of party R. First, if $i < j$, then Condition (90)—see Case I of Step 2—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 Condition (93)—see Case II of Step 2—shows that increasing i slightly increases the expected utility of incumbent $k \in R$. Third, if $\Delta^{z} \leq i$, then Condition (98)—see Case IV of Step 2—shows that increasing policy i slightly increases the expected utility of incumbent $k \in R$. Fourth and last, note that it cannot be the case that $\Delta^{z+1} < i < \Delta^{z}$, as this would contradict Assumption 3.

Case A.2: $h/T \neq T/T$

Based on the analysis of Step 3, we will show that i cannot be the optimal choice by the incumbent k of party R. First, if $i < j$, then (102)—see Case I of Step 3—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 Condition (105)—see Case II of Step 3—shows that increasing policy i slightly increases the expected utility of incumbent $k \in R$. Third, if $\Delta^{z} \leq i$, then Condition (110)—see Case IV of Step 3—shows that increasing i slightly increases the expected utility of incumbent $k \in R$. Fourth and last, note that it cannot be the case that $\Delta^{z+1} < i < \Delta^z$, as this would contradict Assumption 3. This completes the proof of Case A.

Case B: $i > \Delta$

As in Case A, we distinguish two subcases depending on the period $h \in \{t+1, \ldots, t+T\}$ considered.

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

In this case, election takes place at the end of period h. We stress that we are assuming (114). Then we claim that, provided that $T > 1$, it must be the case that

$$i = \Delta^z. \tag{115}$$

Suppose that this is not the case, i.e.,

$$i \neq \Delta^z. \tag{116}$$

We need to distinguish two cases. First, consider that

$$i < \Delta^z. \tag{117}$$

By Assumptions 2 and 3 (and by (114)), it must be the case that

$$j \leq i \leq \Delta^{z+1} \leq \Delta^z. \tag{118}$$

Assumption 3 further implies that the entire sequence of policies chosen by incumbent k will be increasing (and non-constant), and hence that the relevant cases for analysis are Case II of Step 2 and Case II of Step 3. On the one hand, assume that $T > 1$. This means that both j and Δ^z are chosen in periods in which no election takes place (namely, periods $h-1$ and $h+1$). Then Remark 2 together with the fact that the optimal policy in Case II of Step 3 is independent of the values of Δ^z and Δ^{z+1} (and of m_-) implies that

$$j = \Delta^z,$$

a contradiction with (117) and (118). On the other hand, consider the case $T = 1$. Then, in principle we cannot find any contradiction. The reason is that the optimal policy in Case II of Step 2 may *not* be independent of the values of Δ^z, Δ^{z+1}, and m_-. Second, consider that

$$\Delta^z < i. \tag{119}$$

On the one hand, assume that $T > 1$. Then j was chosen by incumbent k, and Assumptions 2 and 3 imply that

$$\Delta^{z+1} \leq \Delta^z < i \leq j. \tag{120}$$

Assumption 3 further implies that the entire sequence of policies chosen by incumbent k will be decreasing (and non-constant), and hence the relevant case for analysis is Case IV of Step 3. Because $T > 1$, both j and Δ^z are chosen in periods in which no election takes place (namely, periods $h-1$ and $h+1$). Then Remark 2 together with the fact that the optimal policy in Case IV of Step 3 is independent of the values of Δ^z and Δ^{z+1} (and of m_-) implies that

$$j = \Delta^z,$$

a contradiction with (120). On the other hand, assume that $T = 1$ and j was chosen by incumbent k (i.e., k is not in his/her first term). Then it must be the case that (120) holds. Then Remark 1 together with the fact that the optimal policy in Case IV of Step 2 is independent of the values of Δ^z and Δ^{z+1} implies that

$$i = \Delta^z,$$

a contradiction with (119).

To sum up, the claim in (115) was correct. In fact, if $T > 1$, then Remark 2 together with the fact that the optimal policy in Case IV of Step 3 is independent of the values of Δ^z and Δ^{z+1} implies that

$$j = i = \Delta^z = \Delta^{z+1}. \tag{121}$$

As for the case $T = 1$, we know that either the entire sequence of policies chosen by k is (weakly) increasing or there is at most one change to a more moderate policy occurring in incumbent k's second term in office.

Case B.2: $h/T \neq T/T$

Accordingly, no election takes place at the end of period h. Clearly, it must therefore be the case that $T > 1$. We also stress that we are assuming (114). Then we claim that

$$i = \Delta^z. \tag{122}$$

Suppose that this is not the case, i.e.,

$$i \neq \Delta^z. \tag{123}$$

We need to distinguish two cases. First, consider that

$$i < \Delta^z. \tag{124}$$

Then Assumption 3 implies that the entire sequence of policies chosen by incumbent k will be increasing (and non-constant), and together with Assumption 2, that the relevant cases for analysis are Case II of Step 2 and Case II of Step 3. On the one hand, assume that Δ^z and Δ^{z+1} are chosen in a period where election takes place. Then we know from Case B.1—see Condition (121)—that

$$i = \Delta^z,$$

a contradiction with (124). On the other hand, assume that Δ^z and Δ^{z+1} are chosen in a period in which no election takes place. Then Remark 2 together with the fact that the optimal policy in Case II of Step 3 is independent of the values of Δ^z and Δ^{z+1} implies that

$$i = \Delta^z,$$

a contradiction with (124). Second, consider that

$$\Delta^z < i. \tag{125}$$

Assumption 3 implies that the entire sequence of policies chosen by incumbent k will be decreasing (and non-constant), and together with Assumption 2, that the relevant cases for analysis are Case IV of Step 2 and Case IV of Step 3. On the one hand, assume that Δ^z and Δ^{z+1} are chosen in a period where election takes place. Then we know from Case B.1—see Condition (121)—that

$$i = \Delta^z,$$

a contradiction with (125). On the other hand, assume that Δ^z and Δ^{z+1} are chosen in a period in which no election takes place. Then Remark 2 together with the fact that the optimal policy in Case IV of Step 3 is independent of the values of Δ^z and Δ^{z+1} implies that

$$i = \Delta^z,$$

a contradiction with (125).

To sum up, the claim in (122) was correct. In fact, Remark 2 together with Condition (121), and the fact that the optimal policy in Case IV of Step 3 is independent of the values of Δ^z and Δ^{z+1} implies that

$$j = i = \Delta^z = \Delta^{z+1}. \tag{126}$$

This completes the proof of Case B.

Having investigated Cases A and B, suppose now that $T > 1$. Then Conditions (121) and (126) imply two properties for any equilibrium different from that of Theorem 2 (if there is any). First, incumbent $k \in R$ will choose the same policy, say policy Δ', in all periods of his/her first term, as well as in the first period of his/her second term (if s/he is re-elected). One important implication of this observation is that Condition (114) holds at the beginning of the second term, and hence the above arguments can be applied to show that incumbent $k \in R$ will choose Δ' in all periods of his/her second term, as well as in the first period of his/her third term (if s/he is re-elected). Repeating this argument, we obtain that, in fact, incumbent k of party R will choose Δ' in any period in office. The second property is that it must be the case that

$$\Delta < \Delta' \leq \min_{z,m_-} \left\{ \mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot \left((1-\rho) \cdot p\left(\Delta', z, m_-\right) + \rho \cdot p\left(\Delta', z+1, m_-\right) \right) \right\}, \quad (127)$$

where the minimum is taken over all numbers of shocks z and all status-quo median voter's peaks m_- that can occur *in equilibrium* during k's tenure *for a given* T. Note that the last period of any incumbent's tenure is always a period in which elections take place. Thus, Step 2 is the relevant step for determining an upper bound for Δ'—see, in particular, Conditions (95) and (107). From all the above, the analysis of Step 1, and the assumptions made on $dF(m_-)$, it is clear that *(i)* for a given m_-, if z is very large, then k will not be re-elected since his/her low valence will offset any other source of utility for the median voter, *(ii)* all else being equal, the lower m_-, the lower are the chances for k to be re-elected. Because incumbent k knows that the median voter's peak can take in the current period as well as in any future period any value in $[-\beta, \beta]$, there must be an integer $z_{\max} \geq 0$ (depending on the primitives of the model) that is the lowest integer satisfying that

$$z \geq z_{max} \Rightarrow p(\Delta', z_{max}+1, -\beta) = 0.$$

This means that Condition (127) (together with Theorem 2) leads to

$$\Delta \leq \Delta' \leq \mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot (1-\rho) \cdot p\left(\Delta', z_{max}, -\beta\right). \quad (128)$$

Note that

$$\mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot (1-\rho) \cdot p\left(\Delta', z_{max}, -\beta\right) \leq \mu - \frac{c}{2} \cdot (1-\theta).$$

Two remarks are in order. First, Assumption 1 guarantees that any status-quo policy j chosen by party L that is in place before an incumbent of party R starts his/her tenure satisfies Condition (114). Then one can easily verify that any other office-holder of party R will subsequently choose a single policy during his/her tenure satisfying Condition (128). Yet different office-holders may choose different policies. Second, by symmetry, one can also show that an incumbent of party L will choose $-\Delta''$ in any period during his/her tenure, where Δ'' satisfies

$$\Delta \leq \Delta'' \leq \mu - \frac{c}{2} \cdot (1+\theta) + c\theta \cdot (1-\rho) \cdot p\left(\Delta'', z_{max}, -\beta\right). \quad (129)$$

Conditions (128) and (129) are *necessary* for an equilibrium different from that of Theorem 2 to exist. These potential equilibria have been described above: all politicians choose a single policy during tenure, and these policies must satisfy Conditions (128) and (129) for politicians of party R and party L, respectively.

To show that Conditions (128) and (129) are also *sufficient* for an equilibrium to exist, suppose w.l.o.g. that an incumbent k from party R chooses policy Δ' satisfying Condition (128) in every period t in which s/he holds office. We claim that there is no profitable deviation for k in any of these periods, in particular for some period t. To see this, let j be the (status-quo) policy chosen in period $t-1$. Due to Assumption 1, Condition (3), and the fact that incumbents start their tenure only after a politician from the other party is ousted (except for period $t=1$), either j was chosen by a politician from party L or it was chosen by k. In either case, Condition (128) ensures that

$$j \leq \Delta'.$$

Recall that we are assuming $T>1$ and that Δ' must satisfy (128). We distinguish two cases.

First, consider that $t/T = T/T$, i.e., election takes place at the end of period t. We have shown that

$$j = i = \Delta^z = \Delta^{z+1} = \Delta',$$

so both Inequality (94)—see Case II of Step 2—and Inequality (99)—see Case IV of Step 2—hold. On the one hand, suppose that incumbent k increases policy i marginally. Then Condition (100)—see Case IV of Step 2—ensures that doing so will not yield incumbent k a higher utility. On the other hand, suppose that incumbent k decreases policy i marginally. Then Condition (92)—see Case I of Step 2—ensures that doing so will neither yield incumbent k a higher utility.[69]

Second, consider that $t/T \neq T/T$, i.e., no election takes place in period t. We have shown that

$$j \leq i = \Delta^z = \Delta^{z+1} = \Delta',$$

so both Inequality (106)—see Case II of Step 3—and Inequality (111)—see Case IV of Step 3—hold. On the one hand, suppose that incumbent k increases policy i marginally. Then Condition (107)—see Case IV of Step 3—ensures that doing so will not yield incumbent k a higher utility. On the other hand, suppose that incumbent k decreases policy i marginally. If $j=i$, then Condition (104)—see Case I of Step 3—ensures that doing so will not yield incumbent k a higher utility. If $j<i$, then Condition (112)—see Case II of Step 3—ensures that doing so will not yield

[69] If $T=1$, one needs to additionally consider the case where $j<i$. In such a case, Condition (95)—see Case II of Step 2—ensures that decreasing i will not yield incumbent k a higher utility.

incumbent k a higher utility. In other words, there is no profitable deviation for incumbent k, as we claimed.

To sum up, provided that $T > 1$, the strategy profile where any office-holder from party R chooses some Δ' satisfying Condition (128) and any office-holder from party L chooses some Δ'' satisfying Condition (129) is an equilibrium of game $\mathcal{G}$. Moreover, there are no other equilibria for our political game.

Finally, it remains to discuss the case $T = 1$. On the one hand, one can verify that the equilibria found for $T > 1$ remain equilibria also for the case $T = 1$.[70] On the other hand, there might be other equilibria in which policy choices are still bounded like above, but in which an incumbent will either choose a weakly increasing sequence of policies or lower his/her policy only once after having received a valence shock (and having been re-elected). The latter equilibria disappear if A is sufficiently large, in which case incumbents who have experienced one valence shock are always ousted.

[70] See Footnote 69.