GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 461

Emergence and Persistence of Extreme Political Systems

Lukas Buchheim *
Robert Ulbricht **

* University of Munich
** Toulouse School of Economics

February 2014

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

Emergence and Persistence of Extreme Political Systems*

Lukas Buchheim
University of Munich

Robert Ulbricht
Toulouse School of Economics

February 1, 2014

Abstract

We investigate the dynamics of political systems in a framework where transitions are driven by reforms and revolts, and where political systems are *a priori* unconstrained, ranging continuously from single-man dictatorships to full-scale democracies. The dynamics are governed by the likelihood of transitions and their outcome, which are both determined endogenously. We find that reforms and revolts result in extreme political systems—reforms by enfranchising the majority of the population leading to democracies, and revolts by installing autocracies. Reinforcing this polarization, extreme political systems are persistent across time: Democracies are intrinsically stable, leading to long episodes without political change. Autocracies, in contrast, are subject to frequent regime changes. Nevertheless they are persistent, since ensuing revolts lead to autocracies comparable to their predecessors. Taken together, our results suggest that the long-run distribution of political systems is bimodal with mass concentrated on the extremes. The dynamics are consistent with cross-country data.

Keywords: Endogenous dynamics of political systems, invariant distribution, persistence of regime types, polarization, transition paths, unrestricted polity space.

JEL Classification: D74, D78, P16.

*We would like to thank Toke Aidt, Matthias Doepke, Florian Englmaier, Matthias Fahn, Johannes Maier, Torsten Persson, Monika Piazzesi, Klaus M. Schmidt, Martin Schneider, Daniel Sturm, Uwe Sunde, Mirko Wiederholt, and Fabrizio Zilibotti for many helpful comments and discussions. We are also grateful for comments by seminar participants at Munich and Northwestern Universities, the 2012 Silvaplana Workshop in Political Economy, the 2013 Royal Economic Society, and the 2013 EEA meetings. Much of this research was done while the second author was visiting Northwestern. He thanks the Economics Department for their hospitality. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged. *Email Addresses:* lukas.buchheim@econ.lmu.de, robert.ulbricht@tse-fr.eu.

1 Introduction

How do political systems evolve over time? Which political systems are persistent? And what types of political systems should we expect to see in the long-run? Thinking about such questions requires thinking about the inherent dynamics of political systems.

While there is now a growing economics literature exploring causes and circumstances of regime changes, its primary focus has so far been on explaining specific patterns of regime changes, initiated through either reforms or revolts. The unfolding dynamics of political system have, however, so far been largely restricted from the outset, if not abstracted from entirely, typically by restricting either the transition mechanism or the set of originating and emerging political systems. This paper takes a step towards overcoming these limitations, placing the dynamic process that describes the evolution of political systems at the center of analysis.

Building on the previous literature, we construct a model that focuses on reforms and revolts as political transition mechanisms, but generalizes the environment to explore a substantially enriched space of *a priori* attainable dynamics. In particular, our model is based on Acemoglu and Robinson (2000*b*) in that agents with access to political power ("political insiders") may conduct preemptive reforms to alleviate the threat from revolts initiated by "political outsiders". To allow the model dynamics to unfold in an essentially unconstrained way, we augment our framework to meet the following three criteria:

(1) To ensure that the dynamics are potentially driven by both reforms and revolts, we introduce an information asymmetry regarding the regime's vulnerability to a revolt that creates a signaling role for reforms. In equilibrium, this leads insiders to sometimes take "tough stance" rather than to negotiate on moderate reforms in light of revolutionary pressure, guaranteeing the co-existence of reforms and revolts along the equilibrium path. The likelihood of either type of transition (or, equivalently, the stability of a particular political system) is thereby endogenously determined by the equilibrium.

(2) We set up the model so that in principle the whole spectrum of political systems, ranging from single-man dictatorships to full-scale democracies, may emerge in equilibrium. Hereby we follow the literature in that we characterize political systems by the fraction of the population having access to political power.[1] Reforms and

[1]Regimes where political power is concentrated in the hands of small elites are, e.g., Chile (1973–90) and today's North Korea. In contrast, the majority of the population is enfranchised in most

revolts are both implemented such that *a priori* a continuum of political systems may emerge from each of them, with the outcome being endogenously determined by the equilibrium. Specifically, as in the majority of previous works, we model reforms as franchise extensions and assume that, after successful revolts, its supporters form the new regime. Deviating from the literature, we, first, allow for franchise extensions of arbitrary scope and, second, explicitly model revolts as the outcome of a coordination game among heterogeneously adjusted outsiders that equally allows for revolts of arbitrary scope.

(3) Finally, being interested in the dynamics of political systems, we set up the model to allow for repeated transitions and avoid to force it to eventually reach an absorbing state. Owing to the first two criteria, consecutive transitions can be both monotonic and non-monotonic.

Results In equilibrium, the model dynamics are characterized by a Markov process that can be decomposed into two underlying mappings: First, each political system maps to an equilibrium likelihood for either type of transition to occur. Second, conditional on the political system in place, either type of transition maps in turn to a specific distribution over newly emerging political systems. Characterizing the model dynamics is thus equivalent to answering two questions. First, which types of political systems arise from reforms and which arise from revolts? And second, how frequently does either type of transition occur given the political system in place?

The model's answer to the first question is that reforms and revolts lead to a polarization of political systems. While revolts result in "autocracies" in which a minority of the population has access to political power, reforms enfranchise the majority of the population and establish "democratic" political systems. These findings hold independent of the originating regime. In contrast, intermediate types of political regimes do not arise along the equilibrium path, so that emerging political systems tend to be extreme.

An interesting implication of these results is that democracies are only established from within regimes. This gives theoretical support to a long-standing view in political science according to which former autocratic elites are key actors in the establishment of democracies (Rustow, 1970; O'Donnell and Schmitter, 1973; Huntington, 1991). Or, as

Western democracies. Regimes between the two extremes, where parts of the population is deprived from political rights in an otherwise inclusive system, are, e.g., Hungary (1921–31) and Madagascar (1960–72).

Karl (1990, p. 8) puts it: "no stable political democracy [in South America] has resulted from regime transitions in which mass actors have gained control, even momentarily, over traditional ruling classes".

The second question above was concerned with the conditional likelihood of regime changes, or, equivalently, the stability of political systems. Here, our model predicts that democratic regimes are intrinsically stable in the sense that there is a low conditional likelihood of either type of transition. In contrast, autocracies are subject to frequent transition events—both, via revolts or reforms. This is in line with the empirical literature on regime stability, which observes that democratic political systems are significantly more stable than autocratic ones (Przeworski, 2000; Gates et al., 2006; Magaloni and Kricheli, 2010).[2]

With the dynamic process at hand, we simulate our model to explore the "long-run" properties of political systems. The key prediction is that the characteristic Markov process integrates to an invariant distribution that tends to be bimodal with mass concentrated on extreme political systems. There are thus two sets of political systems that (while not necessarily being absorbing) are predicted to be frequently observed in the long-run.

To see the logic behind this result, note that the polarizing effect of reforms and revolts ensures the emergence of (only) extreme political systems along the equilibrium path. For certain types of political systems to have significant mass in the long-run, they, however, need both to emerge with positive probability *and* to be persistent. From the above discussion, it is apparent why democracies are persistent: Facing a low conditional likelihood of either type of transition, democracies are stable and long-lasting.

For autocracies the case is more subtle. Given that autocracies face a high conditional likelihood of regime changes, we have that individual autocracies are relatively short-lived. Nevertheless, our simulations suggest that despite their instability, autocratic systems are persistent across time. Precisely because autocracies are characterized by a high conditional likelihood of revolt, political change is frequently initiated by a small group of insurgents, resulting in autocracies very similar to their predecessors. Hence, while the identity of autocratic leaders may change frequently, autocratic systems tend to persist across regimes.[3]

---

[2]From these results it follows that the mode of transition—peaceful reforms or violent revolts—is important for the characteristics of the resulting regimes. For transitions to democracy, a similar point has been highlighted by Cervellati, Fortunato and Sunde (2012, forthcoming), who show that consensual transitions foster civil liberties and property rights provision in contrast to violent transitions.

[3]As a corollary, the same logic implies that revolts are serially correlated across time, because they

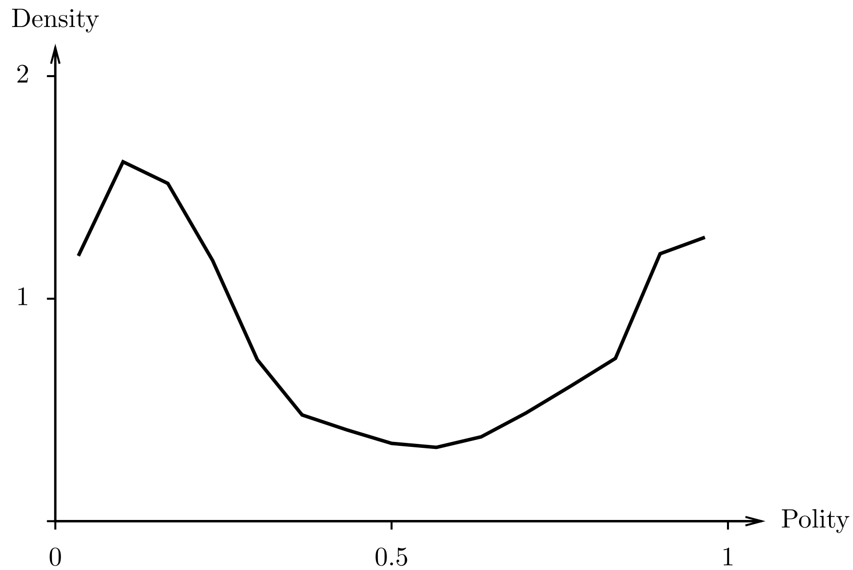

Figure 1. Distribution of political systems since World War I. *Notes:* The figure displays the result of a kernel density regression. Data based on the Polity IV database (Marshall and Jaggers, 2002). Political systems are normalized to range from 0 (extremely autocratic) to 1 (extremely democratic). Units of observation are country-days.

To evaluate how our predictions reconcile with the data, we construct a dataset on political systems and transitions including the majority of countries from 1919 onwards.[4] On an aggregate level, the empirical distribution of political systems since World War I, depicted in Figure 1, matches the predicted bimodal shape. Towards the end of the paper, we use the constructed dataset to take a preliminary look at the empirical counterparts to the two components constituting the Markov process that defines the model dynamics. The exercise suggests that the dynamic process identified by our model is also at work in the data.

Related literature Our paper relates to a growing economics literature on exploring the causes and circumstances of regime changes. In particular, the preemptive logic behind reforms in our paper is based on the seminal rational for why autocratic regimes may want to conduct democratic reforms put forward by Acemoglu and Robinson (2000*b*) (see also, e.g., Conley and Temini, 2001; and Boix, 2003).[5] More closely related

constitute a selection into politically instable regime types.

[4]Specifically, we proxy for political systems using the Polity IV database, and use data from the Archigos Dataset of Political Leaders and the Comparative Constitutions Project to identify transition events linked to changes of political systems. See Section 5 for details.

[5]A related strand of the democratization literature argues that reforms may also be reflective of situations where autocratic decision makers are better off in a democratized political system than

to our paper are a few theoretical studies that allow for preemptive reforms to co-exist with non-democratic transitions along the equilibrium path (e.g., Acemoglu, Ticchi and Vindigni, 2010; Acemoglu and Robinson, 2000*a*; and Ellis and Fender, 2011). Specifically, the latter two of these papers relate to ours in that they choose similar approaches to motivate the co-existence of reforms and revolts via asymmetric information.[6] All of these paper do, however, abstract from repeated transitions, which are at the core of our contribution.

In this respect, our paper relates more closely to Acemoglu and Robinson (2001), who allow for counter-coups in response to newly established democracies, but restrict the space of political systems to two predetermined systems from the outset. Similarly, Justman and Gradstein (1999), Jack and Lagunoff (2006), and Gradstein (2007) allow for multiple (possibly gradual) extensions of the franchise, but do not allow for political change to be initiated from political outsiders via revolts.[7] To the best of our knowledge this is the first paper that allows for both reforms and revolts along the equilibrium path, without restricting their outcomes from the outset. As argued above this is central to our analysis, enabling us to endogenously derive the properties of these transition mechanisms in order to characterize the dynamic process governing the evolution of political systems.

Outline The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 characterizes the equilibrium and outlines the strategic determinants driving policy choices in equilibrium. Section 4 characterizes the law of motion and derives the main predictions regarding the equilibrium dynamics. Section 5 compares our theoretical findings with the data, and Section 6 concludes. All proofs are confined to Appendix A.

---

under the status quo (e.g., Bourguignon and Verdier, 2000; Lizzeri and Persico, 2004; and Llavador and Oxoby, 2005). On the empirical side, Aidt and Jensen (2012) and Przeworski (2009) provide evidence that suggests that preemptive reforms are indeed the driving force behind democratization. In a similar spirit, Besley, Persson and Reynal-Querol (2012) show both theoretically and empirically that a higher risk to lose political power induces leaders to conduct constitutional reforms.

[6]Angeletos, Hellwig and Pavan (2006) and Edmond (2013) also explore how decision makers may strategically manipulate information that affects the outcome of a coordination game in the context of the global games literature. While outsiders in our model share the same posterior beliefs, our model relates to those papers at a methodological level in that it uses heterogeneous opportunity costs as an equilibrium selection device that takes essentially the same role as heterogeneous information does in the global games literature.

[7]Another set of papers that is broadly related includes Acemoglu, Egorov and Sonin (2008, 2012), who characterize the set of stable coalitions or regimes when political status does not act as a commitment, but largely abstract from transitions.

2 The model

We consider an infinite horizon economy with a continuum of two-period lived agents. Each generation has a mass equal to 1. At time t, fraction λ_t of the population has the power to implement political decisions, whereas the remaining agents are excluded from political power. We refer to these two groups as (political) "insiders" and "outsiders".

When born, the distribution of political power among the young is inherited from their parent generation; that is, λ_t agents are born as insiders, while $1 - \lambda_t$ agents are born as outsiders. However, agents who are born as outsiders can attempt to overthrow the current regime and thereby acquire political power. To this end, outsiders choose individually and simultaneously whether or not to participate in a revolt.[8] Because all political change will take effect at the beginning of the next period (see below), only young outsiders have an interest in participating in a revolt. Accordingly, we denote young outsider i's choice by $\phi_{it} \in \{0, 1\}$ and use the aggregated mass of supporters, $s_t = \int \phi_{it}\, \mathrm{d}i$, to refer to the size of the resulting revolt.

Given the mass of supporters s_t, the probability that a revolt is successful is given by

$$p(\theta_t, s_t) = \theta_t h(s_t), \tag{1}$$

where $\theta_t \in \Theta$ is a random state of the world that reflects the vulnerability of the current regime or their ability to put down a revolt, and h is an increasing and twice differentiable function, $h : [0, 1] \to [0, 1]$, with $h(0) = 0$. That is, the threat of a revolt to the current regime is increasing in the mass of its supporters and in the vulnerability of the regime. When a revolt has no supporters ($s_t = 0$) or the regime is not vulnerable ($\theta_t = 0$), it fails with certainty.

The purpose of θ_t in our model is to introduce asymmetric information between insiders and outsiders that, as will become clear below, explains the prevalence of revolts along the equilibrium path. Formally we have that the state θ_t is uniformly distributed on $\Theta = [0, 1]$, is i.i.d. from one period to the next, and is revealed to insiders at the beginning of each period. Outsiders only know the prior distribution of θ_t.

After they learn θ_t, insiders may try to alleviate the threat of a revolt by conducting reforms. We follow Acemoglu and Robinson (2000*b*) by modeling these reforms as an

[8]For notational convenience, we abstract from the possibility of insiders participating in a revolt. In Appendix A.1 we show that this is without loss of generality, since within our framework it is never optimal for insiders to support a revolt against fellow members of the regime.

extension of the franchise to outsiders, which is effective in credibly preventing them from supporting a revolt.[9] However, since our model is aimed at endogenizing the political system λ_t, we generalize this mechanism by allowing insiders to continuously extend the regime by any fraction, $x_t - \lambda_t$, of young outsiders, where $x_t \in [\lambda_t, 1]$ is the reformed political system.[10] Because preferences of insiders will be perfectly aligned, there is no need to specify the decision making process leading to x_t in detail.

Given the (aggregated) policy choices s_t and x_t, and conditional on the outcome of a revolt, the political system evolves as follows:

$$\lambda_{t+1} = \begin{cases} s_t & \text{if the regime is overthrown, and} \\ x_t & \text{otherwise.} \end{cases} \tag{2}$$

When a revolt fails (indicated by $\eta_t = 0$), reforms take effect and the old regime stays in power. The resulting political system in $t+1$ is then given by x_t. In the complementary case, when a revolt succeeds ($\eta_t = 1$), those who have participated will form the new regime. Accordingly, after a successful revolt, the fraction of insiders at $t+1$ is equal to s_t. Note that this specification prevents non-revolting outsiders from reaping the benefits from overthrowing a regime so that there are no gains from free-riding in our model.

To complete the model description, we still have to specify how payoffs are distributed across the two groups of agents at t. As for outsiders, we assume that they receive a constant per period payoff of γ_{it} which is privately assigned to each agent at birth and is drawn from a uniform distribution on $[0, 1]$. We interpret this heterogeneity of outsiders as different degrees of economical or ideological adaptation to a regime, determining their propensity to revolt.

In contrast, insiders enjoy per period payoffs $u(\lambda_t)$, where u is twice differentiable, $u' < 0$, and $u(1)$ is normalized to unity. We think of $u(\cdot)$ as a reduced form function that captures the various benefits of having political power (e.g., from extracting a common resource stock, implementing preferred policies, etc.).[11] One important feature

[9]In Appendix A.1, we show that it is indeed individually rational for enfranchised outsiders to not support a revolt.

[10]By assuming $x_t \in [\lambda_t, 1]$ we are effectively ruling out reforms that withdraw political power once it has been granted. This is in line with the idea that granting someone the status of an insider is a credible and irreversible commitment in the logic of Acemoglu and Robinson (2000*b*).

[11]More specifically, u should be interpreted as a value function where all policy choices that having political power grants access to—except enfranchising political outsiders—are substituted by their optimal policy rules. In particular, this applies to all question of how to organize the economy and

of u is that it is decreasing in the current regime size and, hence, extending the regime is costly for insiders (e.g., because resources have to be shared, or preferences about policies become less aligned). Another thing to note is that $u(\lambda_t) \geq \gamma_{it}$ for all λ_t and γ_{it}; that is, being part of the regime is always desirable. In the case of full democracy ($\lambda_t = 1$) all citizens are insiders and enjoy utility normalized to the one of a perfectly adapted outsider.

To simplify the analysis, we assume that members of an overthrown regime and participants in a failed revolt are worst-adapted to the new regime. Formally, $\gamma_{it} = 0$, resulting in a zero payoff.

For the upcoming analysis it will be convenient to define the expected utility of agents that are born at time t, which is given by:

$$V^I(\theta_t, \lambda_t, s_t, x_t) = u(\lambda_t) + [1 - p(\theta_t, s_t)] \times u(x_t), \tag{3}$$

$$V^O(\theta_t, \gamma_{it}, s_t, \phi_{it}) = \gamma_{it} + \phi_{it}\, p(\theta_t, s_t) \times u(s_t) + (1 - \phi_{it}) \times \gamma_{it}, \tag{4}$$

where superscript I and O denote agents that are born as insiders and outsiders, respectively. In both equations, the first term corresponds to the first period payoff (unaffected by the policy choices of the young agent's generation), while the other terms correspond to second period payoffs. (Since agents do not face an intertemporal tradeoff, we do not need to define a discount rate here).

The timing of events within one period can be summarized as follows:

1. The state of the world θ_t is revealed to insiders.

2. Insiders may extend political power to a fraction $x_t \in [\lambda_t, 1]$ of the population.

3. Observing x_t, outsiders individually and simultaneously decide whether or not to participate in a revolt.

4. Transitions according to (1) and (2) take place, period $t + 1$ starts with the birth of a new generation, and payoffs determined by λ_{t+1} are realized.

---

inasmuch to reallocate resources from outsiders to insiders. Similarly, we also abstract from the question of how exactly to enfranchise political outsiders (i.e., through which political institutions). In abstracting from these issues, we are able to focus on the interplay of changing the inclusiveness of political systems through reforms and through revolts in a tractable way. However, it is important to note that all other policy choices still matter for our analysis in determining the shape of u. Some discussion of how u may vary across different economies and an illustration of how variations in u affect the dynamics of the political system is provided in Section 4.3.

In what follows, we characterize the set of perfect Bayesian equilibria that satisfy the trembling-hand criterion (due to Selten, 1975); that is, perfect Bayesian equilibria that are the limit of some sequence of perturbed games in which strategy profiles are constrained to embody "small" mistakes.[12] To increase the predictive power of our model, we thereby limit attention to equilibria that are consistent with the D1 criterion introduced by Cho and Kreps (1987), a standard refinement for signaling games. The D1 criterion restricts outsiders to believe that whenever they observe a reform x' that is not conducted in equilibrium, the reform has been implemented by a regime with vulnerability θ', for which a deviation to x' would be most attractive.[13]

Anticipating some equilibrium properties, we simplify our notation as follows. First, outsiders' beliefs regarding the regime's vulnerability will be uniquely determined in our setup. We therefore denote the commonly held belief by $\hat{\theta}_t$, dropping the index i. Second, there are no nondegenerate mixed strategy equilibria in our game (see the proofs to Propositions 1 and 2). Accordingly, we restrict the notation in the main text to pure strategies and introduce mixed strategies only to define the perturbations required by trembling-hand perfection.

This leads to the following definition of equilibrium for our economy.

Definition. Given a history $\delta = \{\lambda_0\} \cup \{\{\phi_{i\tau} : i \in [0,1]\}, \theta_\tau, x_\tau, \eta_\tau\}_{\tau=0}^{t-1}$, an equilibrium in this economy consists of policy mappings $x_\delta : (\theta_t, \lambda_t) \mapsto x_t$ and $\{(\phi_{i\delta} : (\hat{\theta}_t, x_t) \mapsto \phi_{it}) : i \in [0,1]\}$, and beliefs $\hat{\theta}_\delta(\lambda_t, x_t) \mapsto \hat{\theta}_t$, such that for all possible histories δ:

a. Reforms x_δ maximize insider's utility (3), given states (θ_t, λ_t), beliefs $\hat{\theta}_\delta$, and perturbed policy mappings $\{\omega_{i\delta}^k : i \in [0,1]\}$ for all values of k;

b. Each outsider's policy choice $\phi_{i\delta}$ maximizes (4), given perturbed policy mappings

[12] Here, the concept of trembling-hand perfection rules out "instable" equilibria, in which $s_t = 0$, but iteratively best-responding to a (perceived) second-order perturbation of s_t would lead to a different equilibrium with a first-order change in s_t. For details see the proof of Proposition 1. Except for these instabilities, the set of trembling-hand perfect equilibria coincides with the set of perfect Bayesian equilibria in our model. An alternative (and outcome-equivalent) approach to rule out these instabilities would be to restrict attention to equilibria which are the limit to a sequence of economies with a finite number of outsiders, where each agent's decision has non-zero weight on s_t.

[13] Formally, let $\bar{V}^I(\theta', \lambda_t)$ be the insiders' payoff in a candidate equilibrium when the regime has a vulnerability θ'. Then the D1 criterion restricts beliefs to the state θ' that maximizes $D_{\theta',x'} = \{\hat{\theta} : V^I(\theta', \lambda_t, s(\hat{\theta}, x'), x') \geq \bar{V}^I(\theta', \lambda_t)\}$, where $s(\hat{\theta}, x')$ is the mass of outsiders supporting a revolt, given the beliefs $\hat{\theta}$ and reform x'. $D_{\theta',x'}$ is said to be maximal here, if there is no θ'', such that $D_{\theta',x'}$ is a proper subset of $D_{\theta'',x'}$. That is, beliefs are attributed to the state in which a deviation to x' is attractive for the largest set of possible inferences about the regime's vulnerability (implying that the regime gains most by deviating).

σ_δ^k, $\{\omega_{j\delta}^k : j \in [0,1] \setminus i\}$, and corresponding beliefs $\hat{\theta}_\delta^k$ for all values of k;

c. Beliefs $\hat{\theta}_\delta = \lim_{k\to\infty} \hat{\theta}_\delta^k(x_t)$, where $\hat{\theta}_\delta^k$ are obtained using Bayes rule given σ_δ^k; and $\hat{\theta}_\delta$ satisfies the D1 criterion;

d. States (λ_t, η_t) are consistent with (1) and (2);

e. The perturbed policy mappings $\{\{\omega_{i\delta}^k : i \in [0,1]\}, \sigma_\delta^k\}_{k=0}^\infty$ are sequences of completely mixed strategy profiles converging to profiles that place all mass on $\{\phi_{i\delta} : i \in [0,1]\}$ and x_δ, respectively.

3 Political equilibrium

In this section, we derive the equilibrium strategies of insiders and outsiders, pinning down the political equilibrium in the model economy. The dynamics of the model economy implied by the equilibrium are investigated in Section 4.

Our analysis is simplified by the overlapping generations structure of our model, which gives rise to a sequence of "generation games" between young insiders and young outsiders. Since the distribution of political power at time t captures all payoff-relevant information of the history up to t, the only link between generations is λ_t. We can therefore characterize the set of equilibria in our model by characterizing the equilibria of the generation games as a function of λ_t. All other elements of the history up to time t may affect the equilibrium at t only by (hypothetically) selecting between multiple equilibria (if the equilibrium in the generation game is not unique).

The generation game consists of two stages that determine the political system at $t+1$. First, outsiders have to choose whether or not to support a revolt. Because the likelihood that a revolt succeeds depends on the total mass of its supporters, outsiders face a coordination problem in their decision to revolt. Second, prior to this coordination problem, insiders decide on the degree to which political power is extended to outsiders. On the one hand this will decrease revolutionary pressure along the extensive margin by contracting the pool of potential insurgents. On the other hand, extending the regime may also contain information about the regime's vulnerability. As a result, reforms may also increase revolutionary pressure along the intensive margin by increasing coordination among outsiders who are not subject to reforms. Insiders' policy choices will therefore be governed by signaling considerations.

We proceed by backward induction in solving for the equilibrium of the generation game, beginning with the outsiders' coordination problem.

3.1 Stage 2: Coordination among outsiders

Consider the outsiders' coordination problem at time t. For any given belief, $(\hat{\theta}_t, \hat{s}_t) \in \Theta \times [0,1]$, individual rationality requires all outsiders to choose a ϕ_{it} that maximizes their expected utility $\mathbb{E}_t\{V^O(\cdot)\}$.[14] At time t, outsider i with adaptation utility γ_{it} will therefore participate in a revolt if and only if

$$\gamma_{it} \leq p(\hat{\theta}_t, \hat{s}_t)\, u(\hat{s}_t) \equiv \bar{\gamma}(\hat{s}_t). \tag{5}$$

Here $\bar{\gamma}(\hat{s}_t)$ is the expected benefit of participating in a revolt that is supported by a mass of $\hat{s}_t$ outsiders. Since $\bar{\gamma}(\hat{s}_t)$ is independent of γ_{it}, it follows that in any equilibrium the set of outsiders who support a revolt at t is given by the agents who are least adapted to the current regime. Suppose for the time being that $\bar{\gamma}(\hat{s}_t) \leq 1$. Then, $\bar{\gamma}(\hat{s}_t)$ defines the fraction of young outsiders that participates in a revolt, and, therefore, the size of a revolt, s_t, that would follow from $\bar{\gamma}(\hat{s}_t)$ is given by

$$f(\hat{s}_t) \equiv (1 - x_t)\, \bar{\gamma}(\hat{s}_t). \tag{6}$$

Further note that in any equilibrium it must hold that $s_t = \hat{s}_t$. Therefore, as long as $\bar{\gamma}(\hat{s}_t) \leq 1$, the share of outsiders that support a revolt at t has to be a fixed point to (6). To guarantee that this is always the case and to further ensure that a well-behaved fixed point exists, we impose the following assumption.

Assumption A1. *For $\psi(s) \equiv h(s) \cdot u(s)$,*

a. $\psi' \geq 0$ *and* $\psi'' \leq 0$*;*

b. $\lim_{s \to 0} \psi'(s) = \infty$.

Intuitively, Assumption A1 states that the participation choices of outsiders are strategic complements; i.e., participating in a revolt becomes more attractive if the total share of supporters grows. This requires that the positive effect of an additional

[14] Note that by our specification of p, V^O is linear in θ_t. For the purpose of computing $\mathbb{E}_t\{V^O\}$, $\hat{\theta}_t \equiv \mathbb{E}_t\{\theta_t\}$ is therefore a sufficient statistic for the full posterior distribution of θ_t. Henceforth we refer to outsider's beliefs accordingly by only keeping track of $\hat{\theta}_t$, disregarding any higher moments.

supporter on the success probability outweighs the negative effect of being in a slightly larger regime after a successful revolt. To ensure existence, we further require that the strategic complementarity is sufficiently strong when a revolt is smallest, and is decreasing as it grows larger.

Using Assumption A1, the above discussion leads to the following proposition.

Proposition 1. *In any equilibrium, the mass of outsiders supporting a revolt at time t is uniquely characterized by a time-invariant function, $s : (\hat{\theta}_t, x_t) \mapsto s_t$, which satisfies $s(0, \cdot) = s(\cdot, 1) = 0$, increases in $\hat{\theta}_t$, and decreases in x_t.*

All formal proofs are in the appendix. Proposition 1 establishes the already discussed tradeoff of conducting reforms: On the one hand, reforms reduce support for a revolt along the extensive margin. In particular, in the limit, as regimes reform to a full-scaled democracy, any threat of revolt is completely dissolved. On the other hand, if reforms signal that the regime is vulnerable, they may backfire by increasing support along the intensive margin.

3.2 Stage 1: Policy choices of insiders

We now turn to the insiders' decision problem. Since more vulnerable regimes have higher incentives to reform than less vulnerable ones, conducting reforms will shift beliefs towards being vulnerable and, therefore, indeed stipulate coordination among outsiders who are unaffected by reforms. This generates the tradeoff established in Proposition 1, which is the main driving force behind the following result.

Proposition 2. *In any equilibrium, policy choices of insiders and beliefs of outsiders are uniquely characterized by time-invariant functions $x : (\theta_t, \lambda_t) \mapsto x_t$, $\xi : \theta_t \mapsto \xi_t$, and $\hat{\theta} : (\lambda_t, x_t) \mapsto \hat{\theta}_t$, such that*

$$x(\theta_t, \lambda_t) = \begin{cases} \lambda_t & \text{if } \theta_t < \bar{\theta}(\lambda_t) \\ \xi(\theta_t) & \text{if } \theta_t \geq \bar{\theta}(\lambda_t), \end{cases}$$

and

$$\hat{\theta}(\lambda_t, x_t) = \begin{cases} \bar{\theta}(\lambda_t)/2 & \text{if } x_t = \lambda_t \\ \bar{\theta}(\lambda_t) & \text{if } \lambda_t < x_t < \xi(\bar{\theta}(\lambda_t)) \\ \xi^{-1}(x_t) & \text{if } \xi(\bar{\theta}(\lambda_t)) \leq x_t \leq \xi(1) \\ 1 & \text{if } x_t > \xi(1), \end{cases}$$

where $\xi' > 0$ *with* $\xi(\theta_t) > \lambda_t + \mu$ *for all* $\theta_t > \bar{\theta}(\lambda_t)$ *and some* $\mu > 0$*, and* $\bar{\theta}(\lambda_t) > 0$ *for all* λ_t*.*

Proposition 2 defines insiders' policy choices for generation t as a function of (θ_t, λ_t). Because the logic behind these choices is the same for all values of λ_t, we can discuss the underlying intuition keeping λ_t fixed. Accordingly, Figure 2 plots reform choices (left panel) and the implied probability to be overthrown (right panel), sliced along a given λ_t plane. It can be seen that whenever a regime is less vulnerable than $\bar{\theta}(\lambda_t)$, insiders prefer to not conduct any reforms (i.e., $x_t = \lambda_t$), leading to a substantial threat for regimes with θ_t close to $\bar{\theta}(\lambda_t)$. Only if $\theta_t \geq \bar{\theta}(\lambda_t)$, reforms will be conducted ($x_t = \xi(\theta_t)$), which in equilibrium effectively mitigate the threat to be overthrown, ruling out marginal reforms where $\xi(\theta_t) \to \lambda_t$.

To see why marginal reforms are not effective in reducing revolutionary pressure consider Figure 3. Here we plot equilibrium beliefs (left panel) and the corresponding mass of insurgents (right panel) as functions of x_t. If the political system is left unchanged by insiders, outsiders only learn the average state $\bar{\theta}(\lambda_t)/2$ of all regimes that pool on $x_t = \lambda_t$ in equilibrium. On the other hand, every extension of the regime—how small it may be—leads to a non-marginal change in outsiders' beliefs from $\hat{\theta}_t = \bar{\theta}(\lambda_t)/2$ to $\hat{\theta}_t \geq \bar{\theta}(\lambda_t)$ and, hence, results in a non-marginal increase in revolutionary pressure along the intensive margin. It follows that there exists some $\tilde{x}(\lambda_t)$, such that for all $x_t < \tilde{x}(\lambda_t)$ the increase of pressure along the intensive margin dominates the decrease along the extensive margin. Thus, reforms smaller than $\tilde{x}(\lambda_t)$ will backfire and *increase* the mass of insurgents (as seen in the right panel of Figure 3), explaining why effective reforms have to be non-marginal.

Furthermore, optimality of reforms requires that the benefit of reducing pressure compensates for insiders' disliking of sharing power. Because $\tilde{x}(\lambda_t) - \lambda_t > 0$, it follows that $u(\tilde{x}(\lambda_t)) - u(\lambda_t) < 0$. Moreover, any reform marginally increasing the regime beyond $\tilde{x}(\lambda_t)$ leads only to a marginal increase in the likelihood to stay in power.

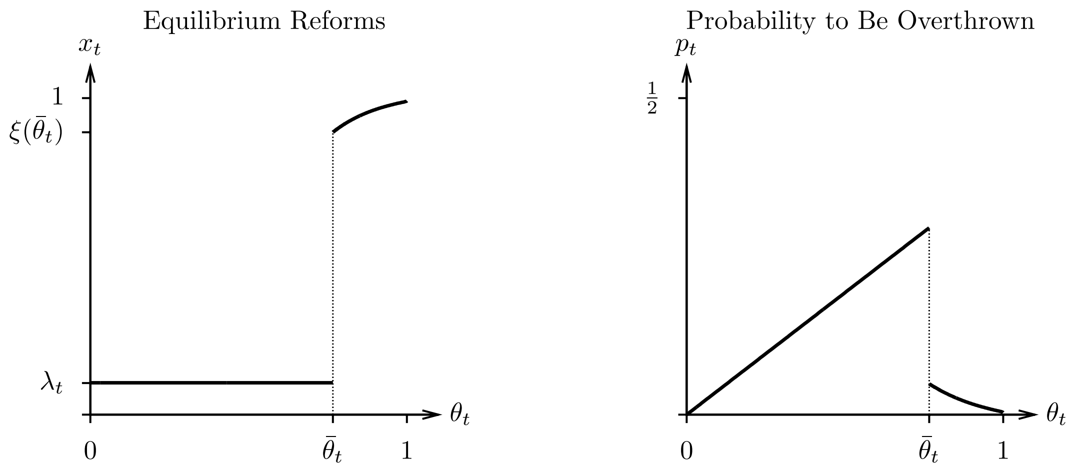

Figure 2. Equilibrium reforms and implied probability to be overthrown.

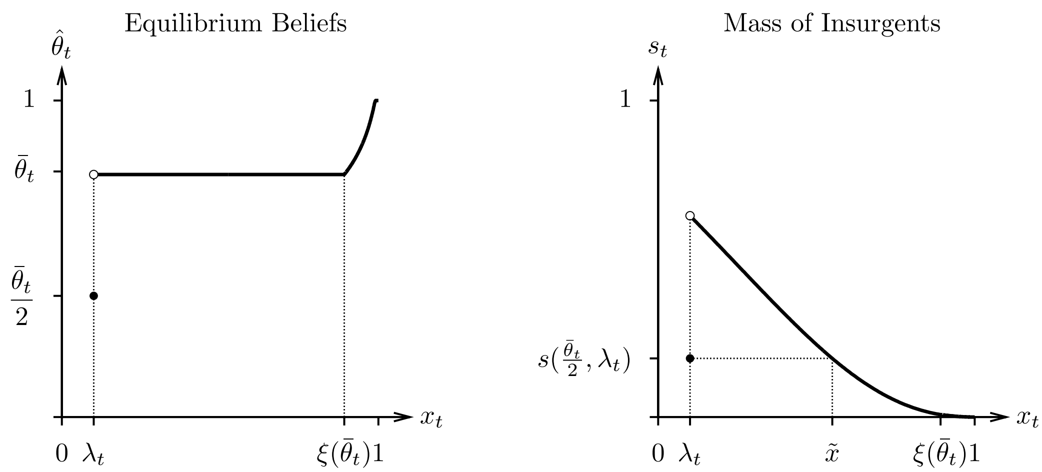

Figure 3. Equilibrium beliefs and implied mass of insurgents.

Hence, there exists a non-empty interval, given by $[\tilde{x}(\lambda_t), \xi(\bar{\theta}(\lambda_t))]$, in which reforms are effective, yet insiders prefer to gamble for their political survival in order to hold on to the benefits of not sharing power in case they survive. This explains the substantial threat for regimes with θ_t close to $\bar{\theta}(\lambda_t)$, as seen in the right panel of Figure 2.[15]

[15] More precisely, gambling for survival increases the likelihood to be overthrown in two ways. First, since at the margin more vulnerable regimes join the pool at $x_t = \lambda_t$, these regimes obviously face a high threat by not conducting reforms. Second, since these regimes also shift the pooling belief towards pooling regimes being more vulnerable, the threat further increases for regimes of all vulnerabilities in the pool.

3.3 Existence and uniqueness of equilibrium

Propositions 1 and 2 uniquely pin down the policy choices in every state, which in return determine the evolution of political systems. We conclude that there is no scope for multiple equilibria in our model economy; if there exists an equilibrium, it must be unique. Verifying the existence then permits us to reach the following conclusion.

Proposition 3. *There exists an equilibrium, in which for all histories δ, policy mappings x_δ and $\{\phi_{i\delta} : i \in [0,1]\}$, as well as beliefs $\hat{\theta}_\delta$ correspond to the time-invariant mappings given by Propositions 1 and 2. Furthermore, for any given initial political system λ_0, the equilibrium is unique.*

4 Transition dynamics

We are now ready to investigate the dynamics of the model economy. By Proposition 3, policy mappings are time-invariant, implying that (λ_t, θ_t) is a sufficient statistic for characterizing the transition dynamics of the political system from time t to $t+1$. Integrating out θ_t, political systems in the unique equilibrium follow a Markov process where the probability that $\lambda_{t+1} \in \Lambda$ can be decomposed into

$$\begin{aligned} Q(\lambda_t, \Lambda) = \rho^S(\lambda_t) \times Q^S(\lambda_t, \Lambda) + \rho^R(\lambda_t) \times Q^R(\lambda_t, \Lambda) \\ + \{1 - \rho^I(\lambda_t) - \rho^R(\lambda_t)\} \times \mathbb{1}_{\lambda_t \in \Lambda}. \quad (7) \end{aligned}$$

Here ρ^S and ρ^R denote the probabilities that in state λ_t a transition occurs via revolts or reforms; Q^S and Q^R are conditional transition functions (specifying the probability that, in state λ_t, $\lambda_{t+1} \in \Lambda$ emerges from a revolt or reform); and $\mathbb{1}$ is an indicator function equal to unity whenever $\lambda_t \in \Lambda$.[16] Accordingly, the first term in (7) defines the probability that state $\lambda_{t+1} \in \Lambda$ emerges through a revolt, the second term defines the

[16]Formally,

$$\begin{aligned} \rho^S(\lambda_t) &= \textstyle\int_0^1 \breve{p}(\theta)\,\mathrm{d}\theta \\ \rho^R(\lambda_t) &= \textstyle\int_{\bar{\theta}(\lambda_t)}^1 \{1 - \breve{p}(\theta)\}\,\mathrm{d}\theta \\ Q^S(\lambda_t, \Lambda) &= \{\rho^S(\lambda_t)\}^{-1} \textstyle\int_{\theta:\breve{s}(\theta)\in\Lambda} \breve{p}(\theta)\,\mathrm{d}\theta \\ Q^R(\lambda_t, \Lambda) &= \{\rho^R(\lambda_t)\}^{-1} \textstyle\int_{\theta:\breve{x}(\theta)\in\Lambda\setminus\lambda_t} \{1 - \breve{p}(\theta)\}\,\mathrm{d}\theta, \end{aligned}$$

where $\breve{x}(\theta) \equiv x(\lambda_t, \theta)$, $\breve{s}(\theta) \equiv s(\hat{\theta}(\lambda_t, \breve{x}(\theta)), \breve{x}(\theta))$, and $\breve{p}(\theta) \equiv p(\theta, \breve{s}(\theta))$.

probability that $\lambda_{t+1} \in \Lambda$ emerges from a reform, and the third term (roughly) refers to the event of no transition.

Decomposing the law of motion into the likelihood maps ρ^S and ρ^R and conditional transition "matrices" Q^S and Q^R is convenient for two reasons. First, it allows us to identify the key forces driving the dynamics of political systems in terms of two intuitively meaningful objects. Second, there exist direct empirical counterparts to the model's likelihood maps and conditional transition matrices, making it in principle possible to investigate whether the forces that drive the dynamics of political systems in the model are also at work in the data.

The next two subsections contain some qualitative characterizations of Q^S, Q^R, ρ^S, and ρ^R. Section 4.3 then simulates the model to explore their interaction and to investigate the long-run dynamics.

4.1 Outcome of transitions

Given the decomposition in (7) the type of political systems that emerge from transitions are defined by the conditional transition "matrices" Q^S and Q^R. Inspecting the equilibrium properties of our model, we get the following polarization result:

Proposition 4. *For all states λ_t,*

$$Q^R(\lambda_t, (\tfrac{1}{2}, 1]) = 1 \qquad \textit{and} \qquad Q^S(\lambda_t, (0, \tfrac{1}{2})) = 1;$$

i.e., reforms lead to majority regimes with $\lambda_{t+1} > 1/2$ and revolts lead to minority regimes with $\lambda_{t+1} < 1/2$.

The first part of Proposition 4 states that any reform leads to a "democratic" system, in which the majority of citizens holds political power. The intuition for this result mirrors the one for Proposition 2. Because conducting reforms will be associated with being intrinsically weak, coordination is increased along the intensive margin. For the benefits along the extensive margin to justify these costs, reforms therefore have to be far-reaching, inducing regimes to enfranchise the majority of the population whenever they conduct reforms.

In contrast, the second part of Proposition 4 establishes that successful revolts always lead to minority regimes, in which a small elite rules over a majority of political outsiders. Underlying this result is that in equilibrium subversive attempts are conducted by only a small group of insurgents. Mass revolutions on the other hand are off-equilibrium. To see what drives this, note that rationality of reforms requires them to be effective; i.e., revolts have to be largest when regimes abstain from reforms and choose to repress the population. However, because abstaining from reforms is optimal for regimes both when they are strong as well as when they hide their weakness through taking tough stance, uncertainty about a regime's weakness is largest from the perspective of outsiders exactly when a regime abstains from reforms. Accordingly, prospects of revolting are at most moderate and only those with large gains from winning political power (i.e., outsiders who are least adapted to the current regime) will find it rational to take the risk of revolting.

An interesting implication of Proposition 4 is that democratic regimes arise if and only if it is optimal for regimes to enfranchise former political outsiders. The commonly made assumption in the previous literature that democracies are established by means of reforms is thus an endogenous outcome of our model. The other channel through which democracies could hypothetically emerge are mass revolutions. But we have just argued that these are events off the equilibrium path. Our model thus supports a long-standing view in political science according to which members of former autocracies are key actors in the establishment of democracies, which is based on, e.g., the observation of Karl (1990, p. 8) that no stable South American democracy has been the result of mass revolutions (see also Rustow, 1970; O'Donnell and Schmitter, 1973; Huntington, 1991).

Finally, note that from Proposition 4 it follows that there is a (possibly quite large) open interval $\bar{\Lambda}$ around 1/2, such that:

Corollary. $Q(\lambda_t, \bar{\Lambda}) = 0$ *for all* λ_t.

That is, there is a range of intermediate regimes that are completely off the equilibrium path. In the simulation below, we will see that $\bar{\Lambda}$ is typically quite large, leading to a long-run distribution with mass only on the extremes.

4.2 Likelihood of transitions

The specific properties of ρ^S and ρ^R depend on the exact specification of u and are investigated below. Here we identify a few limit properties that describe the stability of the most extreme political systems.

Proposition 5. *For* $\lambda_t \to 1$, $\rho^S(\lambda_t) + \rho^R(\lambda_t) \to 0$, *where for all* $\lambda_t > \bar{\lambda}$, $\partial\rho^S/\partial\lambda_t < 0$ *and* $\partial\rho^R/\partial\lambda_t \leq 0$. *For all* $\lambda_t < \underline{\lambda}$, $\rho^S(\lambda_t) + \rho^R(\lambda_t) > \mu$ *for some* $\mu > 0$, $1 >$

$\bar{\lambda} \geq \underline{\lambda} > 0$. *Moreover if* $\bar{\theta}(0) < 1$, *then for all* $\lambda_t < \underline{\lambda}$ *and some* $\{\bar{u}, \underline{u}\} \in \mathbb{R}^2_-$, $\partial \rho^S / \partial \lambda_t < 0$ *and* $\partial \rho^R / \partial \lambda_t > 0$ *if* $\lim_{\lambda \to 0} \partial u / \partial \lambda < \underline{u}$, *and* $\partial \rho^S / \partial \lambda_t > 0$ *and* $\partial \rho^R / \partial \lambda_t < 0$ *if* $\lim_{\lambda \to 0} \partial u / \partial \lambda > \bar{u}$.

Proposition 5 implies that as regimes become more democratic, they eventually become more stable (with $\rho^S(1) = \rho^R(1) = 0$). This is generally true for political systems in which no reforms are conducted; and further holds for sufficiently democratic regimes ($\lambda_t > \bar{\lambda}$). For autocratic systems, in contrast, the likelihood of political change is generally bounded away from zero (but does not necessarily have to be largest for the most autocratic regime).

4.3 Simulation of long-run dynamics

We now simulate our model to explore the long-run dynamics of political systems. For this, let

$$u(\lambda_t) = -\exp(\lambda_t \beta_1) + \beta_0 \quad \text{and} \quad h(s_t) = s_t^{\alpha}.$$

To reduce the number of free parameters, further suppose that $\psi'(1) = 0$; i.e., the strategic effect of an additional outsider supporting a revolt becomes negligible when revolts are supported by the full population. Together with our assumptions on u and h, this pins down α and β_0 in terms of β_1, which is restricted to approximately satisfy $\beta_1 \in (0, 0.56)$.[17]

Interpreting u, one may think of β_0 as a common resource stock or some other type of private benefits that decline at an exponential rate β_1 as power is shared with more insiders. Hence, the larger β_1 the larger the costs of enfranchising political outsiders. In practice, these costs are expected to be high whenever members of the regime have access to a large pool of exogenously given resources, or if there is a large degree of economic and political inequality.[18] If, on the other hand, aggregate income is generated by a production process with strong complementarities between labor inputs or with high returns to capital as in modern Western economies, enfranchising outsiders may come at low costs. This is because enfranchising outsiders constitutes a commitment to

[17]The values implied for α and β_0 are $\alpha = \beta_1 \exp(\beta_1)$ and $\beta_0 = \exp(\beta_1) + 1$, restricting $\beta_1 \in (0, \exp(-\beta_1)) \approx (0, 0.56)$.

[18]In particular, $u(\lambda) = -\exp(\lambda_t \beta_1) + \beta_0 = -\exp(\lambda \beta_1) + \exp(\beta_1) + 1$ is increasing in β_1 for all λ, so that also the inequality between insiders and the average outsider, $\int (u(\lambda) - \gamma)\, \mathrm{d}\gamma$, is increasing in β_1 for all λ.

honor property rights of a larger share of the population and encourages them to acquire human capital, to supply high-skilled labor, or to invest their savings. Accordingly, u (which should be thought of as a value function, cf. Footnote 11) is expected to be relatively flat for modern production economies, so that β_1 is low.

In summary, when β_1 is close to its upper bound, extending the franchise is costly and the incentives to gamble for survival are strong. Consequently, for large β_1, one should expect to observe revolts frequently in equilibrium.[19] On the other hand, if β_1 is low, conducting reforms is cheap and one should expect political insiders to quickly reform to a fully integrated society.

To give an overview of the model dynamics, Figure 4 displays a simulated time series of the model for different values of β_1 and for 500 periods each. For each time path, we plot the political system, λ_t, at time t and indicate the dates where transitions occur via revolts (marked by Δ) and reforms (marked by $\times$). It can be seen that low costs of reforms in Setting 1 ($\beta_1 = 0.35$) result in immediate democratic reforms and the absence of successful subversive attempts. As the costs of reforms are increasing in Setting 2 ($\beta_1 = 0.40$) and Setting 3 ($\beta_1 = 0.45$), successful revolts become more frequent and are followed by periods of frequent regime changes, where autocracies succeed each other. In contrast, democratic reforms give rise to long episodes of political stability.

Across settings, it can be seen that approximately two types of political systems are emerging after transitions *and* are persistent across time. To highlight the model's working, we now briefly refine our previous characterization of Q^S, Q^R, ρ^S, and ρ^R given our parametrization, before taking a closer look at their interplay and the long-run dynamics of political systems.

Outcomes of transitions Figure 5 displays the distribution of political systems that emerge from reforms and revolts for $\beta_1 = 0.4$.[20] From the left panel, it becomes apparent that approximately two types of autocracies emerge after revolts: dictatorships, corresponding to regimes that emerge after revolts against democracies, and autocracies

[19] Given our interpretation of β_1 above, a possible interpretation is a form of "resource curse" leading to civil conflicts, similar to the views expressed in, e.g., Collier and Hoeffler (2004) and Ross (2001). For a critical evaluation of the link between resource richness and civil conflict, see Haber and Menaldo (2011).

[20] The reported distributions weight the conditional distributions $Q^S(\lambda_t, \lambda_{t+1})$ and $Q^R(\lambda_t, \lambda_{t+1})$ with the invariant distribution of λ_t. E.g., letting Ψ denote the invariant distribution, the distribution of political systems after reforms is given by $\int_0^1 Q^R(\lambda_t, \lambda_{t+1})\,\mathrm{d}\Psi(\lambda_t)$. While the long-run distribution itself varies considerably with β_1 (see Figure 8), the conditional distributions displayed here remain largely unaffected by changes in β_1.

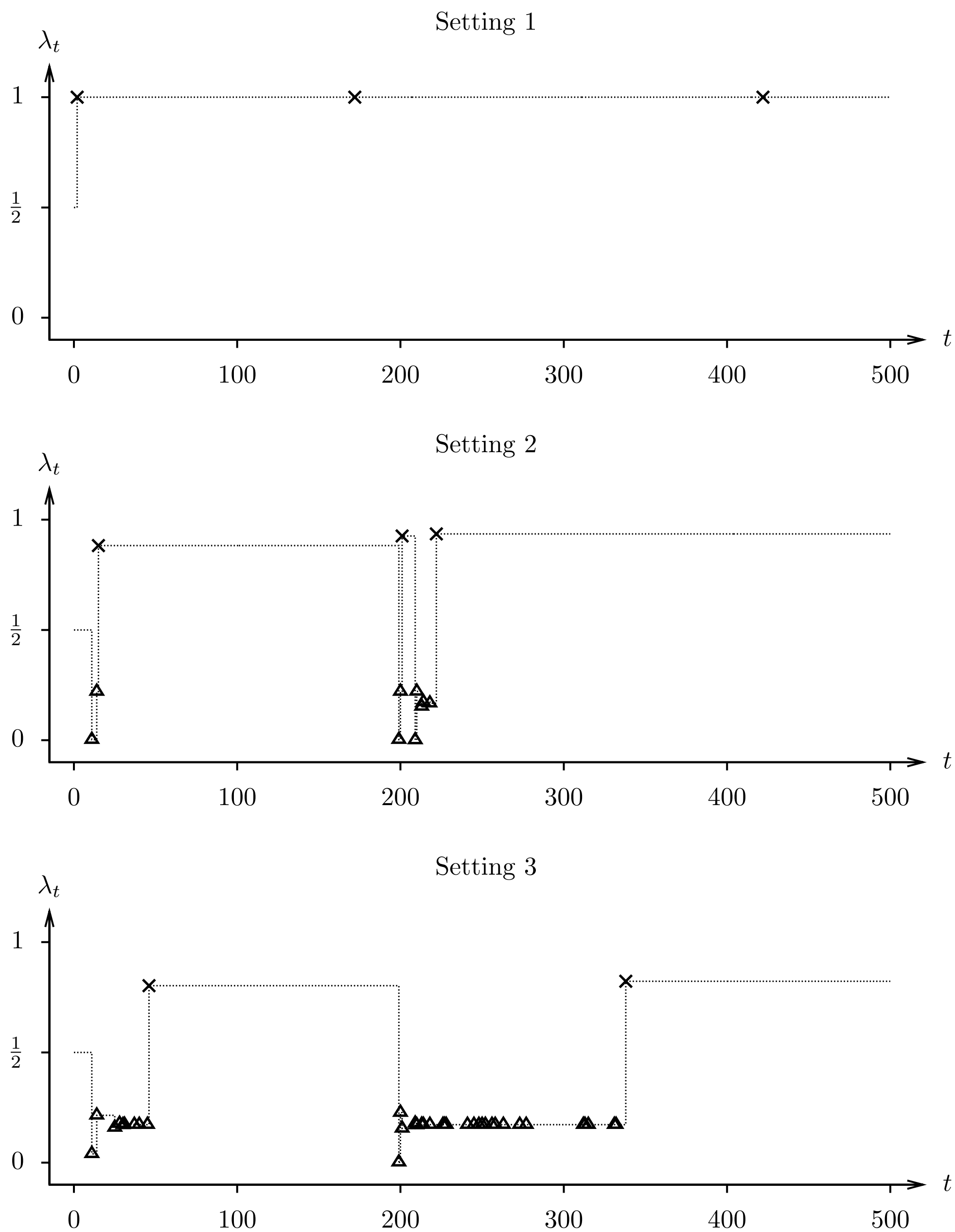

Figure 4. Simulated time series of the model economy. *Notes:* Reforms are marked by "×", successful revolts are marked by "△". Costs of reforms (β_1) are increasing from Setting 1 to 3.

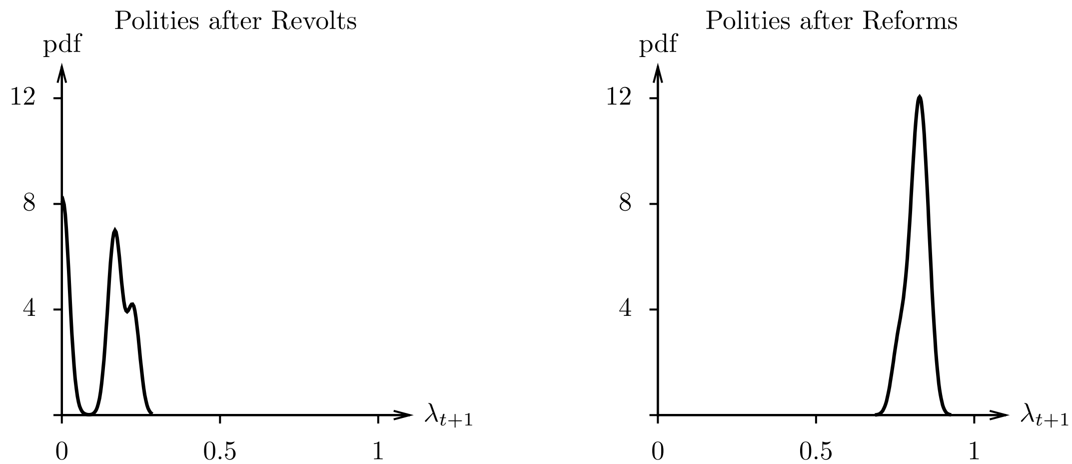

Figure 5. Distribution of political systems after revolts and reforms.

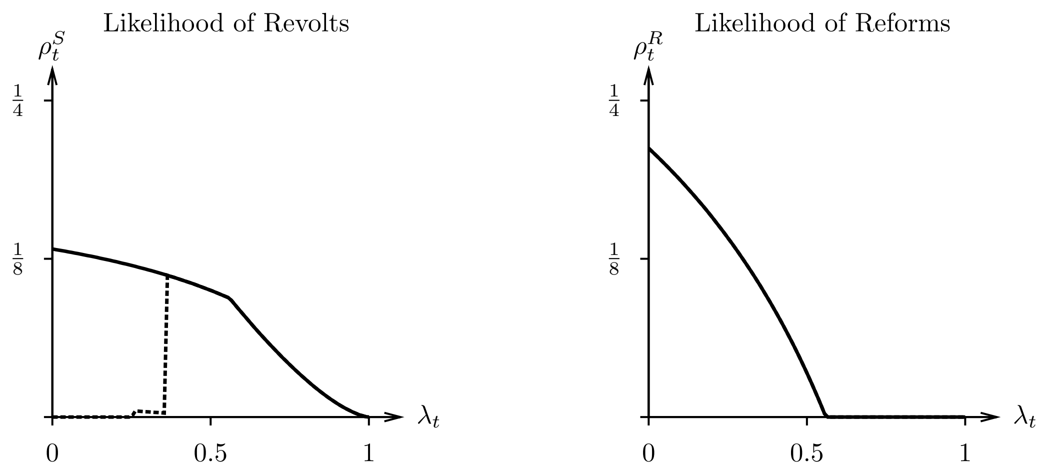

Figure 6. Likelihood of revolts and reforms (solid), and likelihood conditional on an associated change in the political system ≥ 0.25 (dashed).

which emerge after succeeding other non-democratic regimes. The right panel, in turn, displays the distribution of political systems after reforms, which only has positive weight on fairly democratic political systems. In line with Proposition 4, transitions hence lead to a polarization of regimes. In contrast, political systems reaching approximately from 1/4 to 3/4 do neither emerge from reforms, nor from revolts.

Likelihood of transitions The simulations in Figure 4 indicate that autocratic and democratic political systems differ significantly with respect to their stability. In line with Proposition 5, democracies are characterized by long episodes without political change, while autocracies are subject to frequent transitions. The underlying transition probabilities are depicted in Figure 6. Here we plot the likelihood of political transitions via revolts (ρ^S) and reforms (ρ^R) as a function of λ_t (solid lines). It can be seen that both mappings are decreasing in λ_t, such that autocracies are significantly more likely than democracies to experience transitions of either type.

Long-run dynamics While the stability of individual political regimes is sufficiently characterized by $\rho^S + \rho^R$, the "stability" (or persistence) of political *systems* is driven by the interplay of Q^S and Q^R with ρ^S and ρ^R. Specifically, a transition at date t can be seen as a "selection" into certain regime characteristics, which are, in our case, fully defined by the conditional transition functions. Because those characteristics then in turn determine the likely path of future transitions, this selection effect is key to understanding which political systems are frequently observed in the long-run.

From ρ^S and ρ^R depicted in Figure 6 it follows directly that democracies persist over time, because, once established, they merely face low likelihoods of any transition. Autocracies, on the other hand, are relatively short-lived due to their high transition probabilities. Nevertheless Settings 2 and 3 in Figure 4 illustrate a tendency for autocracies to persist *across* regimes. This is because whenever an autocratic regime transforms due to a revolt, the succeeding regime will be very similar to its predecessor, as follows from the conditional distribution of political systems plotted in the left panel of Figure 5. Hence, while the identity of autocratic leaders may change frequently, autocratic systems tend to be persistent. To further illustrate this selection mechanism, the dashed line in the left panel of Figure 6 depicts the likelihood of only those revolts that are associated with a major regime change defined as $|\lambda_{t+1} - \lambda_t| \geq 0.25$. In line with the discussion, revolts against autocracies have a zero probability of establishing a radically different political system.[21]

A corollary to the last result is that revolts are serially correlated across time, since they install regimes that themselves are likely to be overthrown. Conditional on observing a revolt, the future path of the economy is thus likely to be "turbulent" due to a selection into politically instable regime types, as it can be seen in the lower panels of Figure 4. More directly, Figure 7 plots the likelihood of a revolt conditional on observing a successful revolt s periods before (solid line). It can be seen that the conditional likelihood is strictly above the unconditional likelihood of revolts (dashed line) and is highest in periods following a revolt (when the likelihood of not having experienced a reform in the meantime is highest).

[21]Since reforms are generally far-reaching, the probability of a major regime change in the right panel of Figure 6 coincides with the probability of observing a reform (the solid line in the right panel).

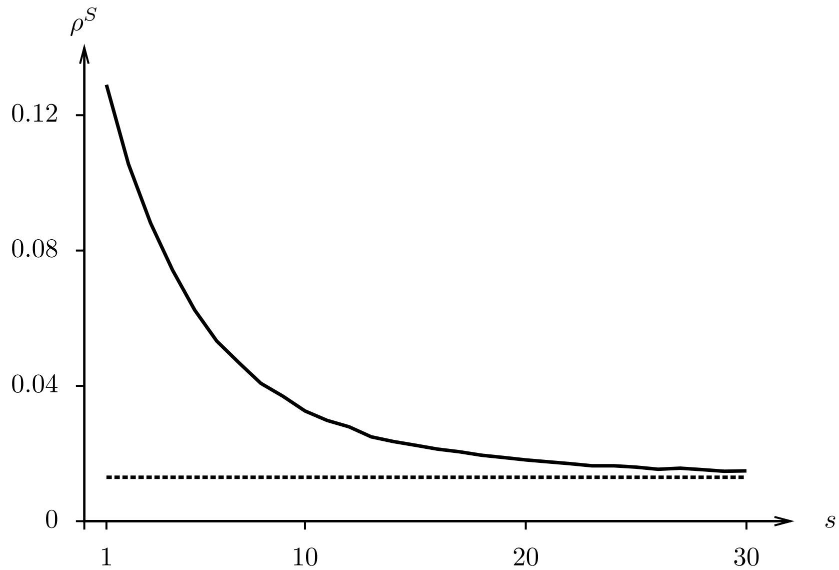

Figure 7. Likelihood of a successful revolt conditional on a revolt s periods before (solid) and unconditional likelihood (dashed).

Finally, the polarization result above on the emergence of extreme political systems lays the ground for an invariant distribution with mass only on the extremes. However, for certain types of political systems to have significant mass in the long-run, they need both to emerge and to be persistent. Figure 8 displays the invariant distribution of political systems for different values of β_1 obtained from running a kernel density regression on simulated time series of 1 Million observations each.[22] Whether political systems are mostly democratic or autocratic thereby depends on the costs of reform as given by β_1. For low values of these costs (Settings 1 and 2), reforms are likely relative to revolts such that mass is mainly concentrated on democratic systems. The converse is true when the costs of conducting reforms are high (Settings 3 and 4).

[22] To retain a constant scale across all settings, we exogenously set a bandwidth = 0.05 in all settings. Somewhat hidden by this is that in Setting 4 all mass is collapsed into a single mass point at $\lambda = 0.12$, which in Setting 4 is absorbing. More generally, there are two scenarios under which a certain political system can be absorbing. First, if $\xi(1) = 1$, then $\lambda = 1$ is reached in equilibrium, which is absorbing, since by Proposition 5 $\rho^S(1) = \rho^R(1) = 0$. However, since $\xi(1) < 1$ in all of the reported settings, we do not observe $\lambda = 1$ along any of the equilibrium paths. Second, if there exists a $\tilde{\lambda}$, such that $\bar{\theta}(\tilde{\lambda}) = 1$ and $s(1/2, \tilde{\lambda}) = \tilde{\lambda}$, then the system $\lambda = \tilde{\lambda}$ is locally attracting and absorbing (despite frequent regime changes), as is the case in Setting 4.

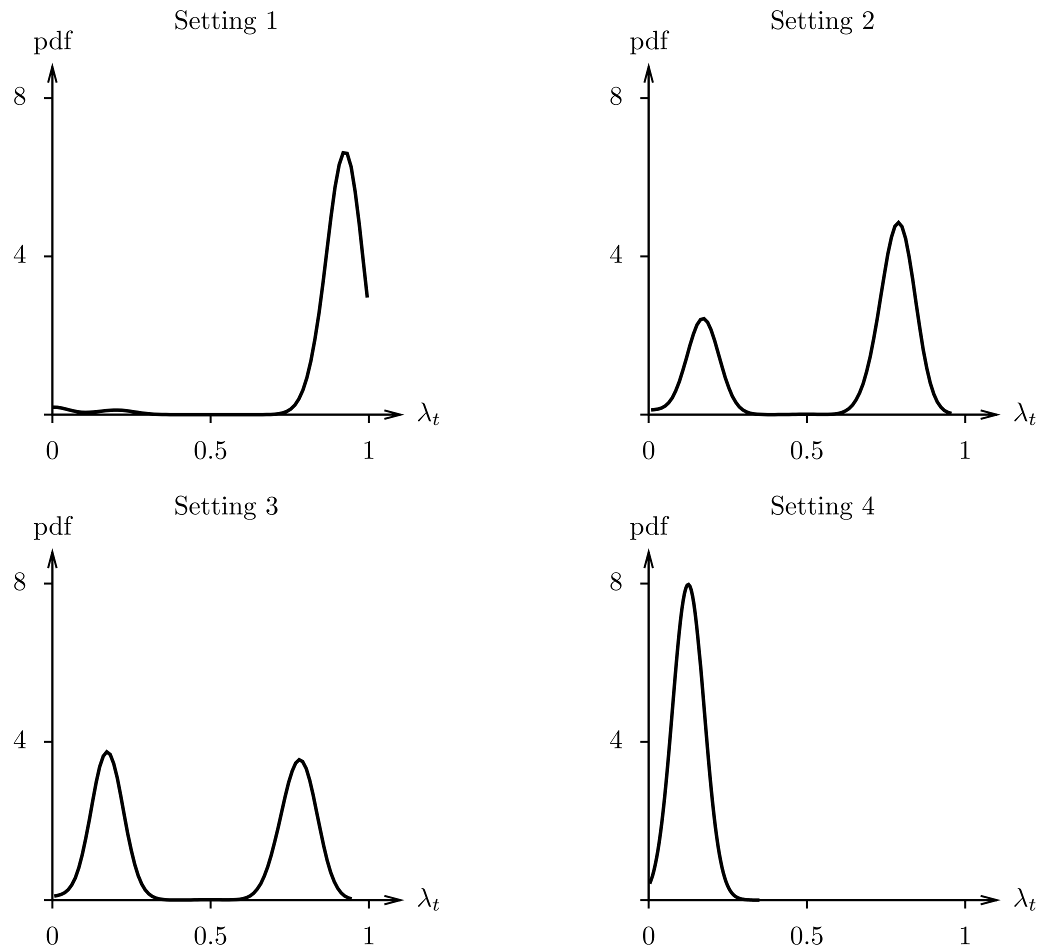

Figure 8. Invariant distribution of political systems. *Note:* Costs of reforms (β_1) are increasing from Setting 1 to 4.

5 A look at the data

The decomposition of the Markov process into ρ^S, ρ^R, Q^S, and Q^R allows us in principle to compare the model's dynamics at an *interim* stage to their empirical counterparts. Comparing the model at an *interim* level ensures that, despite its somewhat abstract nature, the model stands a "fair" chance to speak to the data. At the same time, the level of comparison is sufficiently "disaggregated" to explore whether there is evidence for the forces that drive the long-run dynamics in our model—namely, the interplay of Q^S and Q^R with ρ^S and ρ^R—to also be at work in the data.

In this section, we construct a dataset, containing information on political systems and transitions in the majority of countries from 1919 onwards in order to take an exploratory look at the data along these lines.

5.1 Data construction

As a measure for the inclusiveness of political system, we use the *polity* variable, scaled to $[0, 1]$, from the Polity IV Project (Marshall and Jaggers, 2002), which ranks political regimes on a 21 point scale between autocratic and democratic. In order to examine the model's predictions, we combine this dataset with data on political transitions.

To classify successful revolts, we use the Archigos Dataset of Political Leaders (Goemans, Gleditsch and Chiozza, 2009). The dataset is available for the time period between 1919 and 2004, such that we limit attention to political systems and transition in these years. We record a successful revolt if a leader is irregularly removed from office due to domestic popular protest, rebel groups, or military actors (defined by Archigos' *exitcodes* 2, 4 and 6), and if at the same time the leader's successor takes office in irregular manner (defined by an *entrycode* 1). Furthermore, we take a revolt to be causal for a change in the political system if a change in the political system is recorded in the Polity IV database within a two week window of the revolt.

Finally, we use the dataset on the Chronology of Constitutional Events from the Comparative Constitution Project (Elkins, Ginsburg and Melton, 2010) to classify reforms. We define reforms by a constitutional change (*evnttype* equal to *new*, *reinstated*, or *amendment*) accompanied by a change in the political system (as indicated by the variable *durable* from the Polity IV Project) that is not matched to a revolt or another irregular regime change from the Achigos Dataset. To be consistent with the model's definition of reforms, we restrict attention to positive changes.[23]

The resulting dataset is a daily panel on the country level, which covers 175 countries and records 251 revolts and 97 reforms.

5.2 Empirical properties of political systems and transitions

Table 1 summarizes the dataset. Panel A displays average political systems and annualized empirical likelihoods for a transition of either type. On average, revolts are observed with a frequency of 2.8 percent per year and country, and reforms are observed with a frequency of 1.1 percent. This corresponds on average to a transition every 25 years per country.

[23] In our data we record 49 instances where the political system becomes less democratic after a non-revolutionary change in the constitution, compared to 97 instances where it becomes more democratic. All of the following results are qualitatively robust to including these "negative" constitutional changes in our definition of reforms.

Table 1. Descriptive Statistics

	Mean	Standard Deviation	Observations
	A. Regimes		
Political systems	0.493	0.376	3 289 400
Annual likelihood of a revolt			
Unconditional	0.028		3 289 400
If polity ≤ 0.25	0.030		1 452 533
If polity ≥ 0.75	0.012		1 238 720
Annual likelihood of a reform			
Unconditional	0.011		3 289 400
If polity ≤ 0.25	0.018		1 452 533
If polity ≥ 0.75	0.001		1 238 720
	B. Transitions		
Resulting political systems			
After revolts	0.316	0.235	251
After reforms	0.672	0.242	97

Notes.— Units of observation in Panel A are country-days. Units of observation in Panel B are transitions.

The mean polity is given by 0.49—almost exactly the midpoint of the polity scale. The standard deviation of political systems is, however, quite large. The reason for this becomes clear in light of Figure 1 in the introduction, which displays the distribution of political systems in our dataset: Only a minority of regimes are located in the middle of the polity scale. Instead, in line with our predictions, most mass is concentrated on extreme political systems. More precisely, 44 percent of all regimes are rather autocratic with a polity index of 0.25 and below, while 38 percent of all regimes are rather democratic with an index value of 0.75 and above.

In the remainder of the section, we examine the empirical counterparts of the conditional transition functions, the likelihood maps, and the resulting long-run dynamics discussed in the previous section.

Outcomes of transitions As predicted by Proposition 4 in our model, political systems that emerge from reforms and revolts differ significantly in the data. This can be seen, first, in Panel B of Table 1. Revolts lead to autocratic political systems with a mean polity index of 0.32. Reforms, in contrast, lead to rather democratic political systems with a mean index value of 0.63.

The same picture emerges when looking at the conditional distributions of the polity index depicted in Figure 9 (the mirror image to Figure 5). Clearly, the vast majority of

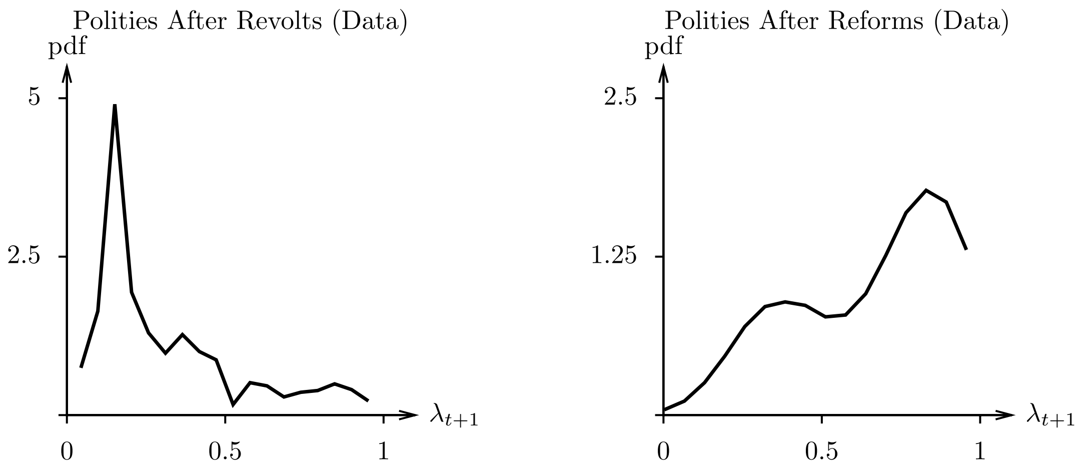

Figure 9. Empirical distribution of political systems after revolts and reforms; estimation by kernel density regressions.

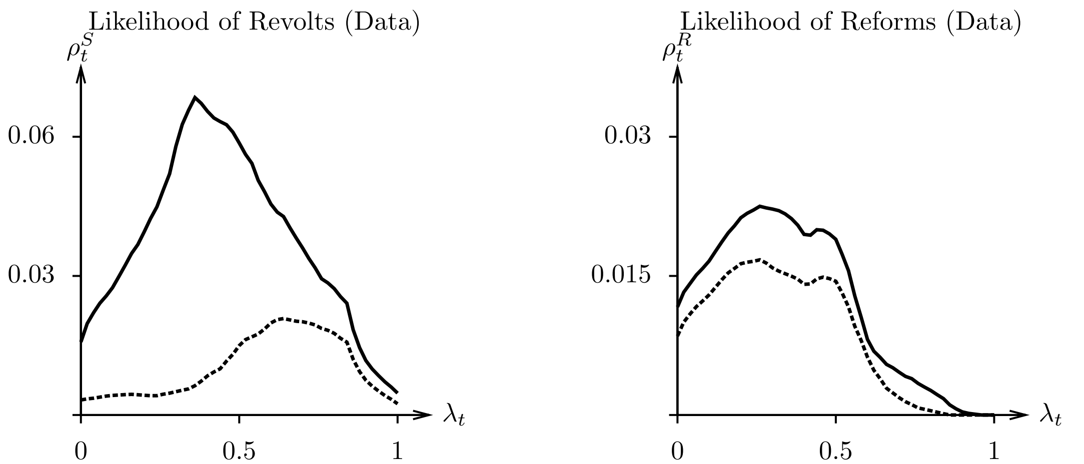

Figure 10. Annual empirical likelihood of revolts and reforms (solid), and likelihood conditional on associated change in the polity index ≥ 0.25 (dashed); estimation by local kernel regressions.

regimes resulting from revolts (left panel) are autocratic. Reforms, on the other hand, by and large lead to democratic political systems, even though a non-negligible number of less democratic systems are emerging after reforms as well.

A possible concern is that the observed correlations may be driven by cross-country heterogeneity or time trends. To address this possibility, we estimate the change in the polity index associated with revolts and reforms while controlling for country and year fixed effects. The estimated coefficients are reported in Column 1 of Table 2. Using only within-country variations, reforms are associated with significant increases in the polity index of 0.44 points on average, hinting at sizable transitions towards democratic systems as predicted by our model. Revolts, in contrast, are associated with statistically

Table 2. Empirical results controlling for country and year fixed effects

Dependent Variable	(1) Δ Polity$_t$	(2) Revolt$_t$	(3) Reform$_t$	(4) Maj. Revolt$_t$	(5) Maj. Reform$_t$
Reform$_t$	0.440				
	(0.025)				
Revolt$_t$	−0.062				
	(0.016)				
Polity$_t$		−0.000	−0.059	0.028	−0.049
		(0.016)	(0.007)	(0.008)	(0.007)
	Control variables				
Year dummies	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes

Notes.— All regressions are OLS. Number of observations are 3 289 400 country-days. Standard errors clustered at the country level are reported in parentheses. With the exception of Column 2, all reported coefficients are significant at the 1 percent level. The dependent variable in Column 1 is the change in the polity index at date t. The dependent variables in Columns 2 to 5 are dummies indicating whether a revolt or reform is observed at date t, whereas the dependents in Columns 4 and 5 are defined with the additional requirement that the associated change in the polity index is not smaller than 0.25. Coefficients and standard errors in Columns 2–5 are multiplied by 365.25 to indicate annual likelihoods.

significant reductions in the polity index by 0.062 points, resulting in (more) autocratic systems.

Note that the difference in magnitude between the estimated coefficients of reforms and revolts should be expected if the mechanisms identified in the model are also at work in the data. Our model predicts that reforms always lead to transitions from autocracies to democracies, implying large changes in the political system. Revolts, in contrast, are predicted to initiate both, transitions from one autocracy to the next with negligible changes in the political system as well as major political changes through transitions from democracies to autocracies. The large effect of reforms compared to the much smaller effect of revolts are thus consistent with the model predictions.

Likelihood of transitions Next, we take a look at the empirical likelihoods of transitions. In line with our model's predictions, the summary statistics in Table 1 suggest that democracies are significantly more stable than autocracies. Autocratic political systems with a polity index below 0.25 are on average more than twice as likely to be overthrown by a revolt than democratic political systems with a polity index value above 0.75. Moreover, autocratic regimes are 18 times more likely to conduct reforms than democratic regimes.

The solid lines in Figure 10 illustrate the relation between the transition likelihoods and the political regime in place (measured by the polity index) in more detail. It can be seen that the likelihoods of reforms and revolts are hump-shaped, so that intermediate political systems have the highest empirical likelihood of a transitions. Comparing the stability of autocracies with democracies, there is a considerably difference in the probability of reforms. The annual probability that a reform is conducted is between one and two percent for autocracies and close to zero for democracies. While for revolts the empirical likelihood also tends to be higher for autocracies than for democracies, the difference is less striking.

Columns 2 and 3 of Table 2 report estimates on these relations where we again control for country and year fixed effects. While there is no significant association between the polity index and the probability of a revolt, full democracies (polity index equals 1) are almost six percentage points less likely to experience a reform than the most autocratic political regimes (index value of 0).

Long-run dynamics With the exception of perhaps the mapping from political systems to the likelihood of revolts, the above results suggest that the likelihood maps and conditional transition functions that drive the dynamics in the model are also at work in the data. Accordingly, we should expect that also more "aggregate" properties of the model's dynamics can be seen in the data.

The first such feature identified by the model is the persistence of extreme political systems. The dashed lines in Figure 10 indicate the empirical probability of observing "major" transitions that are associated with a change in the polity index of at least 0.25 points. The resulting picture qualitatively resembles the corresponding likelihoods of a major political change in the model (see Figure 6). Both in the data and in the model revolts only lead to major changes in the political system when a democratic regime is overthrown. Reforms, in contrast, lead to sizable regime changes in almost all cases, but are conducted only when autocratic political systems are in place, again resembling the conclusion from the model. A linear probability regression, in which we only exploit within-country variation of the polity index and control for year fixed effects, confirms these results. Sizable changes in the polity index are more likely to be initiated via revolts for democratic political systems (Column 4 of Table 2) and via reforms for autocratic political systems (Column 5 of Table 2).[24]

[24]The combination of revolts being likely against fairly autocratic systems but leading to sizable changes in the polity index only when democratic systems are overthrown further supports the previous interpretation of the small impact of revolts on the polity index through the lens of the model.

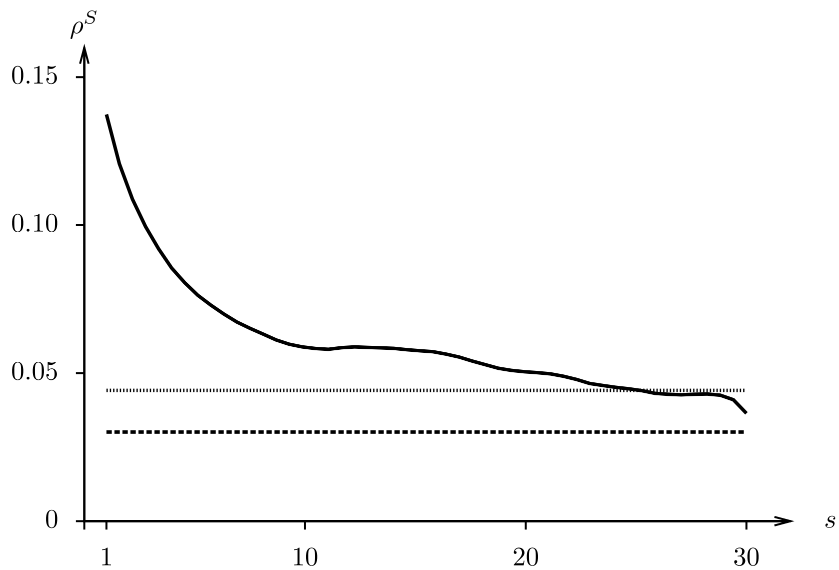

Figure 11. Empirical likelihood of a revolt conditional on a revolt s years before (solid) and unconditional likelihood for all countries (dashed) and countries with at least one transition (dotted); estimation by local kernel regression.

Second, we have seen in the model that the flipside of the selection mechanism that causes the persistence of autocracies is a serial correlation of revolts. Given the above evidence, we therefore expect that a similar serial correlation might be present in the data as well. Figure 11 suggests that this is indeed the case. The solid line in Figure 11 reflects the likelihood of observing a revolt conditional on a successful revolt s years before. This likelihood is considerably larger than the unconditional likelihood of revolts across all countries (dashed line) and also compared to the unconditional likelihood in countries with at least one observed transition (dotted line). Compared to the latter benchmark, the difference is statistically significant at the 5 percent level for $s \leq 15$.

Finally, we already have seen in Figure 1 in the introduction that the empirical distribution of political systems across all countries and times is bimodal in the data, supporting the predictions of the model.

6 Concluding remarks

While there is now a growing economics literature exploring causes and circumstances of regime changes, the unfolding dynamics of political systems across time have so far been largely abstracted from.

This paper makes a contribution towards filling this gap. To achieve this we develop a coherent dynamic framework of political transitions that has three key properties. First, the model is "truly" dynamic, allowing for arbitrary transition paths via repeated regime changes. Second, transitions are driven by both reforms and revolts along the equilibrium path (whereas the likelihood of either transition type at any given point of time is determined endogenously). And, third, political systems are determined endogenously from a continuum of *a priori* possible system. Letting the forces that drive transitions play out freely, we then study the dynamics implied by reforms interacting with revolts.

Our findings suggest that transitions lead to a polarization of political regimes, giving rise to autocracies after revolts and democracies after political reforms. Moreover, while we find democracies to be stable, we find autocracies to be short-lived—characterized by a high likelihood to observe either type of transition. Yet, a selection mechanism gives rise to persistence of autocracies in the long-run as well as to autocorrelation of revolts. As a result, the long-run distribution has mass concentrated on extreme political systems.

To provide a first assessment of inasmuch the forces that drive the evolution of political systems in the model are present in the data, we construct a dataset that combines information on political systems and transitions for the majority of countries since 1919. Looking at the empirical counterparts of the objects that drive the model's dynamics—the likelihood maps and conditional transition functions—we find evidence that they indeed might also define the dynamics in the data. This observation is further substantiated by finding a persistence of autocratic regimes, an autocorrelation of revolts, and a bimodal distribution of political system in the data, which are all predicted by our model. However, currently our evidence is only suggestive and it would be useful to see further empirical investigations of the dynamics outlined in this paper.

Another open question that we have not pursued in this paper is to examine the microfoundations of how the various benefits of having political power determine the shape of insiders' utility function u. In particular, following our discussion in Section 4.3, the costs of sharing power are likely to vary across different economic environments—e.g., as we have argued resource-based endowment economies versus modern production economies—and social institutions. As demonstrated in Figures 4 and 8, such variations would then have important consequences for the types of political systems that are common across different environments and times. Further exploring how cross-country and intertemporal variations in u may explain differences in the dynamics of political systems is a promising direction for future research.

Finally, our analysis implied that revolutionary mass movements are events off the equilibrium path. While from Figure 9 it can be seen that the majority of regimes that emerge after successful revolts are indeed autocratic in the data, there is also a nonzero mass of revolts leading to democratic regimes (around 10 percent). The only way one can account for these events within our framework are strategic mistakes. For example, the elite may erroneously signal weakness by making small concessions, or outsiders may rally because of a commonly held belief that the regime is weak (for example due to information cascades as in Kuran, 1989, or Lohmann, 1994). While it is possible that socially costly mass revolutions are indeed the result of strategic mistakes, extending the analysis by a rational explanation for the emergence of mass revolutions when the regime has the power to counteract them via reforms is another interesting direction for future research.

A Mathematical Appendix

A.1 Insiders never subvert, outsiders always join the regime

Insiders' choice set includes $x_t \in [\lambda_t, 1]$. It thus holds that $(1-p(\cdot, x_t))u(x_t) \geq (1-p(\cdot, 1))u(1) = u(1) \geq \psi(1) \geq \psi(s_t) \geq \hat{\theta}_t \psi(s_t)$, where the first inequality follows from revealed preferences, the second inequality follows from $h(\cdot) \in [0, 1]$, the third inequality follows from ψ increasing, and the last inequality follows from $\theta_t \in [0, 1]$. Hence, it is not attractive for any individual insider to support a revolt against his own regime. As for outsiders we need to differentiate two cases. First, outsiders that are targeted by a reform and would otherwise support a revolt prefer to join the regime using exactly the same argument as above. Second, outsiders that are targeted by a reform and would otherwise not support a revolt prefer to join the regime since again by revealed preferences it holds that $(1 - p(\cdot, x_t))u(x_t) \geq (1 - p(\cdot, 1))u(1) = u(1) \geq \gamma_{it}$ for all i and t. *Q.E.D.*

A.2 Proof of Proposition 1

We first establish that any solution to the outsiders' coordination problem is a fixed point to equation (6). From our discussion in the main body of the paper it is clear that this is the case if and only if $\bar{\gamma}(\hat{s}_t) \leq 1$ for all $\hat{s}_t$. From Assumption A1 it follows that $\bar{\gamma}$ is increasing in $\hat{s}_t$, and therefore $\bar{\gamma}(\hat{s}_t) \leq 1$ holds if $\bar{\gamma}(1) = p(\hat{\theta}_t, 1)\, u(1) \leq 1$. Since $u(1) = 1$ and $p(\cdot) \in [0, 1]$ this is indeed the case.

Hence, consider any fixed point to (6). Since $f(0) = 0$ for all $(\hat{\theta}_t, x_t) \in \Theta \times [0,1]$, there always exists a fixed point at $\hat{s}_t = 0$. Whether or not $\hat{s}_t = 0$ is consistent with the concept of trembling-hand perfection, and whether or not other fixed points exist, depends on the values of $\hat{\theta}_t$ and x_t. We have to distinguish two cases.

First, if $\hat{\theta}_t = 0$ or $x_t = 1$, then $f(\hat{s}_t) = 0$ for all $\hat{s}_t$, and therefore $\hat{s}_t = 0$ is obviously the only fixed point to (6). To establish that $\hat{s}_t = 0$ is also trembling-hand perfect, it suffices to show that for all i, $\phi_{it} = 0$ is a best response to some sequence of totally mixed strategy profiles $\{\omega_{jt}^k : j \in [0,1] \setminus i\}_{k=0}^{\infty}$ that converges to the equilibrium profile where all i play $\phi_{it} = 0$ with probability 1. Since for $\hat{\theta}_t = 0$ and $x_t = 1$ playing $\phi_{it} = 0$ is a (weakly) dominant strategy, this is trivially true.

Second, consider the case where $\hat{\theta}_t \neq 0$ and $x_t \neq 1$. In this case the fixed point at $\hat{s}_t = 0$ is not trembling-hand perfect. To see this let $z^k = \min_i\{\omega_{it}^k(1)\}$ denote the minimum probability with which any agent i plays $\phi_{it} = 0$ in the kth element of sequence ω_{it}^k. The requirement of trembling-hand perfection that $\{\omega_{it}^k\}$ is totally mixed for all i and k implies that $z^k > 0$ for all k. Hence, $s_t^k = (1 - x_t) \int_i \omega_{it}^k(1) \, \mathrm{d}i \geq (1 - x_t) \, z^k > 0$. However, from $h(0) = 0$ in combination with Assumption A1(b) it follows that for any $s_t^k > 0$, $\bar{\gamma}(s_t^k) = \hat{\theta}_t \psi(s_t^k) > 0$ and, hence, a strictly positive fraction of outsiders strictly prefers to choose $\phi_{it} = 1$ in response to $\{\omega_{jt}^k : j \in [0,1]\}$. We conclude that $\hat{s}_t = 0$ can not be supported in any trembling-hand perfect equilibrium if $\hat{\theta}_t \neq 0$ and $x_t \neq 1$.

Having ruled out $\hat{s}_t = 0$ as a solution to the coordination problem for $\hat{\theta}_t \neq 0$ and $x_t \neq 1$, we now show that there is a unique $\hat{s}_t > 0$ solving (6) for $\hat{\theta}_t \neq 0$ and $x_t \neq 1$, which is also consistent with the concept of trembling-hand perfection. From $\bar{\gamma} \in [0,1]$ it follows that f is bounded by its support, $[0, 1 - x_t]$. Moreover, by Assumption A1 we have that $\lim_{\hat{s} \to 0} \psi'(\hat{s}) = \infty$, implying that $\lim_{\hat{s} \to 0} f'(\hat{s}) = \infty$. Hence, there exists a $\tilde{s} > 0$, such that $f(\tilde{s}) > \tilde{s}$. Together with continuity of ψ (and thus of f), it follows that there exists a strictly positive fixed point to (6), which by concavity of ψ (and thus of f) is unique on $(0, 1]$.

Let $s_t^* = f(s_t^*)$ denote this fixed point. It remains to be shown that s_t^* is consistent with the concept of trembling-hand perfection. To show this, consider the following sequences $\omega_{it}^k(1) = 1 - \varepsilon^k$ for all $i \in \{j : \gamma_{jt} \leq \bar{\gamma}(s_t^*)\}$ and $\omega_{it}^k(1) = \frac{\bar{\gamma}(s_t^*)}{1 - \bar{\gamma}(s_t^*)} \varepsilon^k$ for all $i \in \{j : \gamma_{jt} > \bar{\gamma}(s_t^*)\}$, with some $\{\varepsilon^k\}_{k=0}^{\infty}$ such that $\lim_{k \to \infty} \varepsilon^k = 0$. Then, by construction,

$$s_t^k = (1 - x_t) \left((1 - \varepsilon^k) \, \bar{\gamma}(s_t^*) + \frac{\bar{\gamma}(s_t^*)}{1 - \bar{\gamma}(s_t^*)} \, \varepsilon^k (1 - \bar{\gamma}(s_t^*)) \right) = (1 - x_t) \, \bar{\gamma}(s_t^*) = f(s_t^*),$$

and hence $\{\phi_{it} : i \in [0,1]\}$ being mutually best responses implies that $\{\phi_{it} : i \in [0,1]\}$ are best responses to $\{\omega_{it}^k : i \in [0,1]\}$ for all values of k.

The above arguments establish that s_t is uniquely determined by a (time-invariant) function $s : (\hat{\theta}_t, x_t) \to s_t$. It remains to be shown that $\partial s / \partial \hat{\theta}_t \geq 0$ and $\partial s / \partial x_t \leq 0$. Given that s_t is a fixed point to (6), we have that

$$\pi(s_t, x_t) \equiv s_t - (1 - x_t)\,\hat{\theta}_t\,\psi(s_t) = 0.$$

Implicit differentiation implies that

$$\frac{\partial s_t}{\partial x_t} = -\hat{\theta}_t\,\psi(s_t) \times \left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1}$$

and

$$\frac{\partial s_t}{\partial \hat{\theta}_t} = (1 - x_t)\,\psi(s_t) \times \left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1},$$

where

$$\frac{\partial \pi_t}{\partial s_t} = -(1 - x_t)\,\frac{\partial \bar{\gamma}}{\partial s_t} + 1.$$

Since ψ is bounded by $\psi(1) = 1$, (6) implies that $\lim_{\hat{\theta}_t \to 0} s_t^* = \lim_{x_t \to 1} s_t^* = 0$, and therefore the case where $\hat{\theta}_t = 0$ or $x_t = 1$ is a limiting case of $\hat{\theta} \neq 0$ and $x_t \neq 1$. From the implicit function theorem it then follows that s is differentiable on its whole support. Moreover, the previous arguments imply that $f(\tilde{s}) > \tilde{s}$ for all $\tilde{s} < s_t^*$ and $f(\tilde{s}) < \tilde{s}$ for all $\tilde{s} > s_t^*$, implying that $f'(s_t^*) < 1$ or, equivalently, $\partial\bar{\gamma}/\partial s_t < (1 - x_t)^{-1}$ at s_t^*. Thus $\partial \pi_t/\partial s_t > 0$ for all $(\hat{\theta}_t, x_t) \in \Theta \times [0, 1]$, which yields the desired results.

Finally, while we focus on pure strategies above, it is easy to see that the proposition generalizes to mixed strategies. By the law of large numbers, in any mixed strategy equilibrium, beliefs about s are of zero variance and, hence, the arguments above apply, implying that all outsiders, except a zero mass i with $\gamma_i = \bar{\gamma}(s_t^*)$, strictly prefer $\phi_i = 0$ or $\phi_i = 1$. We conclude that there is no scope for (nondegenerate) mixed best responses. *Q.E.D.*

A.3 Proof of Proposition 2

The proof proceeds by a series of lemmas. To simplify notation, in what follows we drop λ_t as an argument of x and $\hat{\theta}$ where no confusion arises. Furthermore, we use $\tilde{V}^I(\theta_t, \hat{\theta}_t, x_t) = (1 - \theta_t h(s_t))\,u(x_t)$ to denote insider's indirect utility (up to a constant $u(\lambda_t)$), as follows from $s_t = s(\hat{\theta}_t, x_t)$ given Proposition 1.

Lemma 1. *x is weakly increasing in θ_t.*

Proof. Suppose to the contrary that $x(\theta'') < x(\theta')$ for $\theta' < \theta''$. Let $x' \equiv x(\theta')$, $x'' \equiv x(\theta'')$, $u' \equiv u(x')$, $u'' \equiv u(x'')$, $h' \equiv h(s(\hat{\theta}(x'), x'))$, and $h'' \equiv h(s(\hat{\theta}(x''), x''))$. Optimality of x' then requires that $\tilde{V}^I(\theta', \hat{\theta}(x''), x'') \leq \tilde{V}^I(\theta', \hat{\theta}(x'), x')$, implying $u'h' - u''h'' \leq (u' - u'')/\theta' < (u' - u'')/\theta''$, where the last inequality follows from $\theta' < \theta''$ and $u' < u''$. Hence, $\tilde{V}^I(\theta', \hat{\theta}(x''), x'') \leq \tilde{V}^I(\theta', \hat{\theta}(x'), x')$ implies that $\tilde{V}^I(\theta'', \hat{\theta}(x''), x'') < \tilde{V}^I(\theta'', \hat{\theta}(x'), x')$, contradicting optimality of x'' for θ''.

Lemma 2. *Suppose x is discontinuous at θ', and define $x^- \equiv \lim_{\varepsilon\uparrow 0} x(\theta' + \varepsilon)$ and $x^+ \equiv \lim_{\varepsilon\downarrow 0} x(\theta' + \varepsilon)$. Then for any $x' \in (x^-, x^+)$, the only beliefs consistent with the D1 criterion are $\hat{\theta}(x') = \theta'$.*

Proof. Let $\theta'' > \theta'$, and let $x'' \equiv x(\theta'')$. Optimality of x'' then requires that $\tilde{V}^I(\theta'', \hat{\theta}(x''), x'') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+)$ and, thus for any $\tilde{\theta}$,

$$\begin{aligned}&\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x''), x'') \quad \text{implies that} \\ &\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+)\,.\end{aligned}$$

Moreover, arguing as in the proof of Lemma 1,

$$\begin{aligned}&\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+) \quad \text{implies that} \\ &\tilde{V}^I(\theta', \tilde{\theta}, x') > \tilde{V}^I(\theta', \hat{\theta}(x^+), x^+)\,.\end{aligned}$$

Hence, if $\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+) = \bar{V}^I(\theta'')$, then $\tilde{V}^I(\theta', \tilde{\theta}, x') > \tilde{V}^I(\theta', \hat{\theta}(x^+), x^+) = \bar{V}^I(\theta')$. Therefore, $D_{\theta'',x'}$ is a proper subset of $D_{\theta',x'}$ if $\theta'' > \theta'$. (For the definition of $D_{\theta,x}$, see Footnote 13.) A similar argument establishes that $D_{\theta'',x'}$ is a proper subset of $D_{\theta',x'}$ if $\theta'' < \theta'$ and, thus, the D1 criterion requires that $\hat{\theta}(x') = \theta'$ for all $x' \in (x^-, x^+)$.

Lemma 3. *There exists $\bar{\theta}(\lambda_t) > 0$, such that $x(\theta_t, \lambda_t) = \lambda_t$ for all $\theta_t < \bar{\theta}(\lambda_t)$. Moreover, $x(\theta'') > x(\theta') > \lambda_t + \mu$ for all $\theta'' > \theta' \geq \bar{\theta}(\lambda_t)$ and some $\mu > 0$.*

Proof. First, consider the existence of a connected pool at $x_t = \lambda_t$. Because for $\theta_t = 0$, $x_t = \lambda_t$ dominates all $x_t > \lambda_t$, we have that $x(0) = \lambda_t$. It follows that there exists a pool at $x_t = \lambda_t$, because otherwise $\hat{\theta}(\lambda_t) = 0$ and, therefore, $p(\cdot, s(\hat{\theta}(\lambda_t), \lambda_t)) = 0$, contradicting optimality of $x(\theta) > \lambda_t$ for all $\theta > 0$. Moreover, by Lemma 1, x is increasing, implying that any pool must be connected. This proves the first part of the claim.

Now consider $x(\theta'') > x(\theta')$ for all $\theta'' > \theta' \geq \bar{\theta}(\lambda_t)$ and suppose to the contrary that $x(\theta'') \leq x(\theta')$ for some $\theta'' > \theta'$. Since x is increasing, it follows that $x(\theta) = x^+$ for all $\theta \in [\theta', \theta'']$ and some $x^+ > \lambda_t$. W.l.o.g. assume that θ' is the lowest state in this pool. Then Bayesian updating implies that $\theta^+ \equiv \hat{\theta}(x^+) \geq (\theta' + \theta'')/2 > \theta'$ and, therefore, $\tilde{V}^I(\theta', \theta^-, x^+) > \tilde{V}^I(\theta', \theta^+, x^+)$ for all $\theta^- \leq \theta'$. Hence, because θ' prefers x^+ over $x(\theta^-)$, it must be that $x(\theta^-) \neq x^+$ for all $\theta^- \leq \theta'$ and, hence, $x(\theta^-) < x^+$ by Lemma 1. Accordingly, let $x^- \equiv \max_{\theta^- \leq \theta'} x(\theta^-)$.

Then from continuity of $\tilde{V}^I$ and $\theta^+ > \theta'$ it follows that there exists an off-equilibrium reform $x' \in (x^-, x^+)$ with $\tilde{V}^I(\theta', \theta', x') > \tilde{V}^I(\theta', \theta^+, x^+)$. Hence, to prevent θ' from choosing x' it must be that $\hat{\theta}(x') > \theta'$. However, from Lemma 2 we have that $\hat{\theta}(x') = \theta'$, a contradiction.

Finally, to see why there must be a jump-discontinuity at $\bar{\theta}(\lambda_t)$ note that $\tilde{V}^I(\bar{\theta}(\lambda_t), \bar{\theta}(\lambda_t)/2, \lambda_t) = \tilde{V}^I(\bar{\theta}(\lambda_t), \bar{\theta}(\lambda_t), x(\bar{\theta}(\lambda_t)))$; otherwise, there necessarily exists a θ in the neighborhood of $\bar{\theta}(\lambda_t)$ with a profitable deviation to either λ_t or $x(\bar{\theta}(\lambda_t))$. From the continuity of $\tilde{V}^I$ and the non-marginal change in beliefs from $\bar{\theta}(\lambda_t)/2$ to $\bar{\theta}(\lambda_t)$ it follows that $x(\bar{\theta}(\lambda_t)) > \lambda_t + \mu$ for all λ_t and some $\mu > 0$.

Lemma 4. *x is continuous and differentiable in θ_t on $[\bar{\theta}(\lambda_t), 1]$.*

Proof. Consider continuity first and suppose to the contrary that x has a discontinuity at $\theta' \in (\bar{\theta}(\lambda_t), 1)$. By Lemma 1, x is monotonically increasing in θ_t. Hence, because x is defined on an interval, it follows that for any discontinuity θ', $x^- \equiv \lim_{\varepsilon \uparrow 0} x(\theta')$ and $x^+ \equiv \lim_{\varepsilon \downarrow 0} x(\theta')$ exist, and that x is differentiable on $(\theta' - \varepsilon, \theta')$ and $(\theta', \theta' + \varepsilon)$ for some $\varepsilon > 0$. Moreover, from Lemmas 2 and 3 it follows that in equilibrium $\hat{\theta}(x') = \theta'$ for all $x' \in [x^-, x^+]$. Hence, $\tilde{V}^I(\theta', \theta', x^-) = \tilde{V}^I(\theta', \theta', x^+)$, since otherwise there necessarily exists a θ in the neighborhood of θ' with a profitable deviation to either x^- or x^+. Accordingly, optimality of $x(\theta')$ requires $\tilde{V}^I(\theta', \theta', x') \leq \tilde{V}^I(\theta', \theta', x^-)$ and, thus, $\tilde{V}^I(\theta', \theta', x^-)$ must be weakly decreasing in x. Therefore, $\partial \tilde{V}^I / \partial \hat{\theta}_t < 0$ and $\lim_{\varepsilon' \downarrow 0} \partial \hat{\theta}(x^- - \varepsilon')/\partial x_t > 0$ (following from Lemma 3) imply that $\lim_{\varepsilon' \downarrow 0} \partial \tilde{V}^I(\theta', \hat{\theta}(x^- - \varepsilon'), x^- - \varepsilon')/\partial x_t < 0$. Hence, a profitable deviation to $x^- - \varepsilon'$ exists for some $\varepsilon' > 0$, contradicting optimality of $x(\theta')$.

We establish differentiability by applying the proof strategy for Proposition 2 in Mailath (1987). Let $g(\theta, \hat{\theta}, x) \equiv \tilde{V}^I(\theta, \hat{\theta}, x) - \tilde{V}^I(\theta, \theta', x(\theta'))$, for a given $\theta' > \bar{\theta}(\lambda_t)$, and let $\theta'' > \theta'$. Then, optimality of $x(\theta')$ implies $g(\theta', \theta'', x(\theta'')) \leq 0$, and optimality of $x(\theta'')$ implies that $g(\theta'', \theta'', x(\theta'')) \geq 0$. Letting $a = (\alpha\theta' + (1-\alpha)\theta'', \theta'', x(\theta''))$, for some $\alpha \in [0, 1]$ this implies

$$0 \geq g(\theta', \theta'', x(\theta'')) \geq -g_\theta(\theta', \theta'', x(\theta''))(\theta'' - \theta') - \tfrac{1}{2} g_{\theta\theta}(a)(\theta'' - \theta')^2,$$

where the second inequality follows from first-order Taylor expanding $g(\theta'', \theta'', x(\theta''))$ around $(\theta', \theta'', x(\theta''))$ and rearranging the expanded terms using the latter optimality condition. Expanding further $g(\theta', \theta'', x(\theta''))$ around $(\theta', \theta', x(\theta'))$, using the mean value theorem on $g_\theta(\theta', \theta'', x(\theta''))$, and noting that $g(\theta', \theta', x(\theta')) = g_\theta(\theta', \theta', x(\theta')) = 0$, these inequalities can be written as

$$\begin{aligned} 0 \geq\ & g_{\hat{\theta}}(\theta', \theta', x(\theta')) + \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'} \times [g_x(\theta', \theta', x(\theta')) \\ & + \tfrac{1}{2} g_{xx}(b(\beta))(x(\theta'') - x(\theta')) + g_{\hat{\theta}x}(b(\beta))(\theta'' - \theta')] + \tfrac{1}{2} g_{\hat{\theta}\hat{\theta}}(b(\beta))(\theta'' - \theta') \\ \geq\ & -[g_{\theta\hat{\theta}}(b(\beta')) + \tfrac{1}{2} g_{\theta\theta}(a)](\theta'' - \theta') - g_{\theta x}(b(\beta'))(x(\theta'') - x(\theta')), \end{aligned}$$

for $b(\beta) = (\theta', \beta\theta' + (1-\beta)\theta'', \beta x(\theta') + (1-\beta)x(\theta''))$ and some $\beta, \beta' \in [0,1]$. Because $\tilde{V}^I$ is twice differentiable, all the derivatives of g are finite. Moreover, continuity of x implies that $x(\theta'') \to x(\theta')$ as $\theta'' \to \theta'$ and, therefore, for $\theta'' \to \theta'$,

$$0 \geq g_{\hat{\theta}}(\theta', \theta', x(\theta')) + \lim_{\theta'' \to \theta'} \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'}\, g_x(\theta', \theta', x(\theta')) \geq 0.$$

By Lemma 3, x and, hence, $\hat{\theta}$ are strictly increasing for all $\theta \geq \bar{\theta}(\lambda_t)$. Arguing similarly as we did to show continuity, optimality of x, therefore, requires that $g_x = \partial\tilde{V}^I/\partial x_t \neq 0$ and, hence, the limit of $(x(\theta'') - x(\theta'))/(\theta'' - \theta')$ is well defined, yielding

$$\frac{\mathrm{d}x}{\mathrm{d}\theta_t} = -\frac{\partial\tilde{V}^I/\partial\hat{\theta}_t}{\partial\tilde{V}^I/\partial x_t}. \tag{8}$$

Lemma 5. *$x(\theta_t, \lambda_t) = \xi(\theta_t)$ for all $\theta_t > \bar{\theta}(\lambda_t)$, where ξ is unique and $\partial\xi/\partial\theta_t > 0$.*

Proof. From Lemma 4 we have that ξ is differentiable, and by Lemma 3, $\partial\xi/\partial\theta_t > 0$. We thus only need to show that ξ is unique. By the proof to Lemma 4, $dx/d\theta_t$ is pinned down by the partial differential equation (8), which must hold for all $x_t \geq x(\bar{\theta}(\lambda_t))$. Moreover, whenever $\bar{\theta}(\lambda_t) < 1$, in equilibrium $\hat{\theta}(x(1)) = 1$ and, therefore, it obviously must hold that $x(1, \lambda_t) = \arg\max_{x_t} \tilde{V}^I(1, 1, x_t)$, providing a boundary condition for (8). Because $\tilde{V}^I$ is independent of λ_t, it follows that $x(\theta_t, \lambda_t)$ is uniquely characterized by a function, i.e., $\xi : \theta_t \mapsto x_t$, for all $\theta_t \geq \bar{\theta}(\lambda_t)$.

Lemma 6. *$\bar{\theta}(\lambda_t)$ is unique.*

Proof. Suppose to the contrary that $\bar{\theta}(\lambda_t)$ is not unique. Then there exist $\bar{\theta}'' > \bar{\theta}'$, defining two distinct equilibria for a given λ_t. By Lemma 5, there is a unique $\xi(\theta)$ characterizing reforms outside the pool for both equilibria. Optimality for type $\theta \in (\bar{\theta}', \bar{\theta}'')$ then requires $\tilde{V}^I(\theta, \theta, \xi(\theta)) \geq \tilde{V}^I(\theta, \bar{\theta}'/2, \lambda_t)$ in the equilibrium defined by $\bar{\theta}'$, and $\tilde{V}^I(\theta, \theta, \xi(\theta)) \leq \tilde{V}^I(\theta, \bar{\theta}''/2, \lambda_t)$ in the equilibrium defined by $\bar{\theta}''$. However, $\tilde{V}^I(\theta, \bar{\theta}'/2, \lambda_t) > \tilde{V}^I(\theta, \bar{\theta}''/2, \lambda_t)$, a contradiction.

This establishes uniqueness of $x(\theta_t, \lambda_t)$, with all properties given by Lemmas 3 and 5, and the corresponding beliefs $\hat{\theta}(\lambda_t, x_t)$ following from Lemma 2 and Bayesian updating. Again, for the purpose of clarity we have established this proposition by focusing on pure strategy equilibria. In the following we outline how the proof generalizes to mixed strategy equilibria; a detailed version of these steps can be attained from the authors on request.

Replicating the proof of Lemma 1, it is trivial to show that if $\tilde{V}^I(\theta', \hat{\theta}(x'), x') = \tilde{V}^I(\theta', \hat{\theta}(x''), x'')$, then $\tilde{V}^I(\theta'', \hat{\theta}(x'), x') < \tilde{V}^I(\theta'', \hat{\theta}(x''), x'')$ for all $\theta' < \theta''$ and $x' < x''$. It follows that (i) supports, $\mathcal{X}(\theta)$, are non-overlapping, and (ii) $\min \mathcal{X}(\theta'') \geq \max \mathcal{X}(\theta')$. Moreover, noting that $\tilde{x}(\theta) \equiv \max \mathcal{X}(\theta)$ has a jump-discontinuity if and only if type θ mixes in a nondegenerate way, (ii) further implies that there can be only finitely many types that mix on the closed interval $[0,1]$. The logic of lemmas 2, 3, and 4 then apply, ruling out any jumps of $\tilde{x}$ on $[\bar{\theta}(\lambda_t), 1]$. This leads to the conclusion that at most a mass zero of types (i.e., $\theta_t = \bar{\theta}(\lambda_t)$) could possibly mix in any equilibrium (with no impact on $\hat{\theta}$) and, thus, there is no need to consider any nondegenerate mixed strategies. *Q.E.D.*

A.4 Proof of Proposition 3

From the discussion in the main body of the paper it is clear that the equilibrium is uniquely pinned down by the time-invariant mappings given by Propositions 1 and 2 if it exists. We are thus left to show existence, which requires us to verify that the equilibrium mappings are consistent with the D1 and trembling-hand criterion. The first is a direct implication from the proof of Proposition 2 where we apply Lemma 2 to restrict off-equilibrium beliefs, such that $\hat{\theta}$ is necessarily consistent with the D1 criterion.

To show consistency with the concept of trembling-hand perfection, we need to show that $\{\phi_i : i \in [0,1]\}$ and x are best responses to a sequence of completely mixed strategy profiles $\{\{\omega_i^k : i \in [0,1]\}, \sigma^k\}_{k=0}^{\infty}$ that converge to a profile that places all mass on $\{\phi_i : i \in [0,1]\}$ and x, respectively.

Accordingly, for $\phi_i(\hat{\theta}^k(\cdot, x_t), x_t)$ to be a best-response to x_t and the perturbed strategy profile $\{\omega_i^k : i \in [0,1]\}$ for the marginal outsider i with $\gamma_i = \bar{\gamma}(s_t)$, we need that $\hat{\theta}^k(\cdot, x_t)\, \psi(s_t^k(x_t)) = \bar{\gamma}(s_t)$, requiring any change in beliefs along the perturbation path to be offset by trembles of outsiders $j \neq i$. Because for $x \in [\xi(1), 1]$, $\hat{\theta}(\cdot, x) = 1$ can never be sustained in a completely mixed equilibrium with a continuum of types, this implies that we need to adjust for $\hat{\theta}^k(\cdot, x) < \hat{\theta}(\cdot, x)$ by introducing asymmetric trembles, leading to $s^k(x) > s(\hat{\theta}(\cdot, x), x)$. Hence, let $s^k(x(1)) = s(x(1)) + \varepsilon_k$ for some $\{\varepsilon^k\}_{k=0}^{\infty}$ such that $\lim_{k\to\infty} \varepsilon^k = 0$ and $\varepsilon^k \in (0, \bar{\varepsilon})$ for all k.

A necessary (and for $\theta \in (\bar{\theta}(\cdot), 1)$ sufficient) condition for $x \in [\xi(\bar{\theta}(\cdot)), \xi(1)]$ to be optimal against s^k is that $s^k(x)$ satisfies the inverse differential equation (8) for $x(\cdot, \theta)$ fixed,

$$\frac{\mathrm{d}s^k}{\mathrm{d}x} = -\frac{\partial V^I/\partial x}{\partial V^I/\partial s}\bigg|_{s=s^k}, \tag{9}$$

which in combination with $s^k(x(1))$ pins down $s^k(x)$ for all $x \in [\xi(\bar{\theta}), \xi(1)]$. Note that $s^k(x(1)) > s(\cdot, x(1))$ implies that $s^k(x) > s(\cdot, x)$ for all $x \in [\xi(\bar{\theta}), \xi(1)]$ since the indifference condition (8) is unique. Moreover, since optimality of x requires that $\bar{\theta}$ is necessarily indifferent between λ_t and $\xi(\bar{\theta})$, $s^k(\xi(\bar{\theta}))$ pins down $s^k(\lambda_t) > s(\cdot, \lambda_t)$.

For off-equilibrium $x \in (\lambda, \xi(\bar{\theta})) \cup (\xi(1), 1]$ we are free to assign any $s^k(x)$ that (1) assures optimality of x, and (2) converges to $s(\cdot, x)$. As to (1), we can for instance set $s^k(x) = s(\bar{\theta}, x) + s^k(\xi(\bar{\theta})) - s(\cdot, \xi(\bar{\theta}))$ for $x \in (\lambda, \xi(\bar{\theta}))$ (which is continuous around $\xi(\bar{\theta})$ and has slope $\mathrm{d}s(\bar{\theta}, x)/\mathrm{d}x \geq \mathrm{d}s^k(\xi(\bar{\theta}))/\mathrm{d}x$, so that by (9) no type has an incentive to deviate), and $s^k(x) = s(\cdot, x) + \varepsilon^k f^k(x)$ for $x \in (\xi(1), 1]$ with some $f^k : [\xi(1), 1] \to \mathbb{R}_+$ such that $\mathrm{d}f^k(\xi(1))/\mathrm{d}x = \{\mathrm{d}s^k(\xi(1))/\mathrm{d}x - \mathrm{d}s(\cdot, \xi(1))/\mathrm{d}x\}/\varepsilon^k$ and f^k sufficiently convex for V^I to be concave on $[\xi(1), 1]$, so that $\xi(1)$ is the global optimum for $\theta = 1$.

Note that these definitions imply that $s^k(x) \downarrow s(\hat{\theta}(\cdot, x), x)$ for all x and, hence, $\hat{\theta}^k(\cdot, x) \uparrow \hat{\theta}(\cdot, x)$ for all x as implied by the indifference condition of the marginal outsider, $\hat{\theta}^k(x) = \bar{\gamma}(s(\cdot, x))/\psi(s^k(x)) \in (0, \hat{\theta}(\cdot, x))$. By construction, these sequences assure optimality of $\{\phi_i : i \in [0, 1]\}$ and x along the perturbation path. To conclude the proof it therefore suffices to show the existence of $\{\{\omega_i^k : i \in [0, 1]\}, \sigma^k\}_{k=0}^{\infty}$ yielding $\{s^k, \hat{\theta}^k\}_{k=0}^{\infty}$.

Consider $\{s^k\}_{k=0}^{\infty}$ first. Define $\tilde{\varepsilon}$ such that $\max_x s^k(x) < 1 - \lambda$ for $\varepsilon^k = \tilde{\varepsilon}$ and suppose that $\bar{\varepsilon} \leq \tilde{\varepsilon}$.[25] Then any s^k can be sustained by setting

$$\omega_i^k(1)(x) = \begin{cases} 1 - \varepsilon^k & \text{for all } i : \gamma_i \leq \bar{\gamma}(s(\hat{\theta}(\cdot, x), x)) \\ c^k(x)\varepsilon^k & \text{for all } i : \gamma_i > \bar{\gamma}(s(\hat{\theta}(\cdot, x), x)), \end{cases}$$

with $c^k(x) = \{s^k(x) - (1 - \varepsilon^k)s(\cdot, x)\}/\{(1 - x)(1 - \bar{\gamma}(x))\varepsilon^k\}$. Note that ω_i^k is completely mixed if $\bar{\varepsilon} < 1$ and $\varepsilon^k c^k(x) \in (0, 1) \iff c^k(x) \in (0, 1/\varepsilon^k) \iff s^k(x) + \varepsilon^k s(\cdot, x) < 1 - x$. From $s^k(x) > s(\cdot, x)$ we have that $c^k(x) > 0$ and because $s^k \to s$, using the same arguments as in Footnote 25, there exists some $\hat{\varepsilon}$ such that $c^k(x) < 1/\varepsilon^k$ holds for all $\bar{\varepsilon} \leq \hat{\varepsilon}$.

Finally, consider $\{\hat{\theta}^k\}_{k=0}^{\infty}$. It is straightforward to verify by Bayes rule that any $\hat{\theta}^k$ with $\hat{\theta}^k(x) > 0$ for all x can be sustained by setting

$$\sigma^k(x)(\theta, \cdot) = \begin{cases} \varepsilon^k & \text{if } \theta > \hat{\theta}^k(x) \text{ and } (x > \lambda_t \text{ or } \theta > \bar{\theta}_t) \\ d^k(x)\varepsilon^k & \text{if } \theta < \hat{\theta}^k(x) \text{ and } x > \lambda_t \\ 1 - R^k(\theta) & \text{if } \theta \geq \bar{\theta}(\lambda_t) \text{ and } x = \xi(\theta) \\ T^k & \text{if } \theta \leq \hat{\theta}^k(\lambda_t) \text{ and } x = \lambda_t \\ Z^k & \text{if } \theta \in (\hat{\theta}^k(\lambda_t), \bar{\theta}(\lambda_t)) \text{ and } x = \lambda_t, \end{cases}$$

with $d^k(x) = (1 - \hat{\theta}^k(x))^2/\hat{\theta}^k(x)^2$, $R^k(\theta) = \int_{\theta > \hat{\theta}(x)} \varepsilon^k \,\mathrm{d}x + \int_{\theta < \hat{\theta}(x)} d^k(x)\varepsilon^k \,\mathrm{d}x$, $T^k = \inf_{\theta < \bar{\theta}(\lambda_t)}(1 - R^k(\theta))$, and $Z^k = \{T^k\hat{\theta}^k(x)^2 + \varepsilon^k[2(1 - \bar{\theta}(\lambda_t))\hat{\theta}^k(\lambda_t) - 1 + \bar{\theta}(\lambda_t)^2]\}/\{\bar{\theta}(\lambda_t) - \hat{\theta}^k(\lambda_t)\}^2$. With a slight abuse of notation, in the definition of σ^k, R^k , T^k and Z^k denote probabilities,

[25] To see that $\tilde{\varepsilon}$ exists, note that $s(\hat{\theta}(\cdot, x), x) < 1 - x \leq 1 - \lambda_t$ since otherwise $\bar{\gamma}_t = 1$, which requires $\bar{\theta}_t = 1$ and $s_t = 1$, contradicting that s is strictly decreasing in x. Convergence of s^k to s then implies that one can always find some $\tilde{\varepsilon}$ that is sufficiently small.

while ε^k are understood to be probability densities. Note that σ^k is completely mixed if $T^k, R^k(\theta) \in (0,1)$ and $Z^k \in (0, R^k(\theta))$ for all θ. This is obviously true for some $\breve{\varepsilon}$, such that $\bar{\varepsilon} < \breve{\varepsilon}$. Finally, note that the above definition is incomplete in the sense that $R^k(\theta) + T^k < 1$ or $R^k(\theta) + Z^k < 1$ for some types $\theta < \bar{\theta}(\lambda_t)$. In these cases the remaining probability mass can be distributed (almost) arbitrary over atoms on $(\lambda_t, 1]$ without impact on the resulting beliefs.[26]

We conclude the proof by setting $\bar{\varepsilon} = \min\{1, \tilde{\varepsilon}, \hat{\varepsilon}, \breve{\varepsilon}\}$. *Q.E.D.*

A.5 Proof of Proposition 4

Consider $Q^R(\lambda_t, (\frac{1}{2}, 1]) = 1$ first. By Proposition 2, for any reform $x_t > \lambda_t$, $x_t = \xi(\theta_t)$, with ξ increasing. To show the claim, it thus suffices to show that $\tilde{x} \equiv \xi(\tilde{\theta}) > 1/2$ for $\tilde{\theta} = \min_\lambda \bar{\theta}(\lambda)$. Also, define $\tilde{\lambda} = \arg\min_\lambda \bar{\theta}(\lambda)$. Then, optimality of $\tilde{x}$ implies $s^* \equiv s(\tilde{\theta}/2, \tilde{\lambda}) > s(\tilde{\theta}, \tilde{x}) \equiv s^{**}$. Using (6),

$$s^* = (\tilde{\theta}/2)(1 - \tilde{\lambda})\, \psi(s^*) \equiv w^* \psi(s^*), \tag{10}$$

$$s^{**} = \tilde{\theta}(1 - \tilde{x})\, \psi(s^{**}) \equiv w^{**} \psi(s^{**}). \tag{11}$$

Note that, in analogue to the proof of Proposition 1, for a general $w_t \equiv \hat{\theta}_t(1 - x_t)$ it holds that

$$\frac{\partial s_t}{\partial w_t} = -\psi(s_t) \left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1} > 0.$$

Hence, $s^* > s^{**}$ implies $w^* > w^{**}$, or $(\tilde{\theta}/2)(1 - \tilde{\lambda}) > \tilde{\theta}(1 - \tilde{x})$. Rearranging, then proves the claim,

$$\tilde{x} > 1 - \frac{1 - \tilde{\lambda}}{2} \geq \frac{1}{2}.$$

Now consider $Q^S(\lambda_t, (0, \frac{1}{2})) = 1$. Again, optimality of x_t implies that $s(\hat{\theta}(\lambda_t, x), x)$ is decreasing in x. Hence, for all λ_t,

$$s(\hat{\theta}(\lambda_t, x_t), x_t) \leq s(\bar{\theta}(\lambda_t)/2, \lambda_t) \leq s(1/2, 0),$$

where the last inequality follows since s is increasing in its first and decreasing in its second argument. Hence, it suffices to show that $s(1/2, 0) < 1/2$.

Let $s^* \equiv s(1, 0) \leq 1$ and let $s^{**} \equiv s(1/2, 0)$. From (6), $s^* = \psi(s^*)$ and $s^{**} = \psi(s^{**})/2$.

[26] For instance, we can dispose of the atomic waste without any hazard by having each type θ place the remaining probability mass on $x = \lambda_t + \theta(1 - \lambda_t)/\bar{\theta}(\lambda_t)$.

Moreover, by Proposition 1, $s^* > s^{**}$. Hence, since ψ is strictly increasing,

$$s^{**} = \frac{\psi(s^{**})}{2} = \frac{\psi(\psi(s^{**})/2)}{2} < \frac{\psi(\psi(s^{*})/2)}{2} = \frac{\psi(s^{*}/2)}{2} < \frac{\psi(s^{*})}{2} = \frac{s^*}{2} \leq \frac{1}{2}.$$

Q.E.D.

A.6 Proof of Proposition 5

From Footnote 16,

$$\rho^S(\lambda_t) = \int_0^{\bar{\theta}(\lambda_t)} \theta h\left(s\left(\bar{\theta}(\lambda_t)/2, \lambda_t\right)\right) \mathrm{d}\theta + \int_{\bar{\theta}(\lambda_t)}^1 \theta h\left(s\left(\theta, x(\theta)\right)\right) \mathrm{d}\theta, \tag{12}$$

and

$$\rho^R(\lambda_t) = \int_{\bar{\theta}(\lambda_t)}^1 \left(1 - \theta h\left(s\left(\theta, x(\theta)\right)\right)\right) \mathrm{d}\theta. \tag{13}$$

Also, note that $\bar{\theta}(\lambda_t) \in (0, 1]$ is implicitly defined as the solution to

$$F(\bar{\theta}, \lambda_t) \equiv \tilde{V}^I(\bar{\theta}, \bar{\theta}/2, \lambda_t) - \tilde{V}^I(\bar{\theta}, \bar{\theta}, \xi(\bar{\theta})) = 0, \tag{14}$$

if an interior solution exists. Otherwise, for λ_t there is a corner solution $\bar{\theta}(\lambda_t) = 1$, which implies $\tilde{V}^I(1, 1/2, \lambda_t) > \tilde{V}^I(1, 1, \xi(1))$.

First, consider $\lambda_t > \bar{\lambda}$. Suppose that there exists $\bar{\lambda}$, such that for all $\lambda_t \in (\bar{\lambda}, 1]$, $\bar{\theta}(\lambda_t)$ is a corner solution. Then clearly for all $\lambda_t > \bar{\lambda}$, $\partial\bar{\theta}(\lambda_t)/\partial\lambda_t = 0$, such that $\partial\rho^S(\lambda_t)/\partial\lambda_t = \partial h(s(1/2, \lambda_t))/\partial\lambda_t < 0$, by Proposition 1. Furthermore, $\partial\rho^R(\lambda_t)/\partial\lambda_t = 0$. Otherwise, if there exists no $\bar{\lambda}$, such that for all $\lambda_t \in (\bar{\lambda}, 1]$, $\bar{\theta}(\lambda_t)$ is a corner solution, then there necessarily exists a λ^*, such that for $\lambda_t \in (\lambda^*, 1]$, $\bar{\theta}(\lambda_t)$ is an interior solution. But then, because it is trivially true that $\rho^S(1) = \rho^R(1) = 0$, continuity of ρ^S and ρ^R implies that $\partial\rho^S(\lambda_t)/\partial\lambda_t < 0$ and $\partial\rho^R(\lambda_t)/\partial\lambda_t < 0$ for all $\lambda_t > \bar{\lambda}$ and some $\bar{\lambda} < 1$, proving the first sentence of the proposition.

Now consider $\lambda_t < \underline{\lambda}$. By Proposition 2, $\bar{\theta}(\lambda_t) > 0$ for all λ_t and, hence, $s(\bar{\theta}(\lambda_t)/2, \lambda_t) > 0$ for all $\lambda_t < 1$ by Proposition 1. Hence the first term in (12) integrates to a strictly positive number for all $\lambda_t < 1$, proving the second sentence of the proposition.

Regarding the third sentence in the proposition, now further let $\bar{\theta}(0) < 1$. Then, F differentiable implies that $\bar{\theta}(\lambda_t)$ has an interior solution and is differentiable for all $\lambda_t \in [0, \lambda^*)$ for some $\lambda^* > 0$. Implicit differentiation of F, substituting for $x'(\bar{\theta})$ from (8), and using

$F(\bar{\theta}, \lambda_t) = 0$ yields

$$\frac{\partial \bar{\theta}(\lambda)}{\partial \lambda} = \frac{-\bar{\theta} h_1^p s_2^p u^p + (1-p^p) u_1^p}{\frac{\bar{\theta}}{2} h_1^p s_1^p u^p + \frac{u^p - u^s}{\bar{\theta}}}, \tag{15}$$

where subscript i denotes the derivative with respect to the ith argument, and superscripts p and s denote that the function is evaluated at the pooling or separating values, respectively (where $\hat{\theta}^p = \frac{\bar{\theta}}{2}$, $x^p = \lambda$ and $\hat{\theta}^s = \bar{\theta}, x^s = x(\bar{\theta})$).

Using this, the signs of $\partial \rho^S / \partial \lambda_t$ and $\partial \rho^I / \partial \lambda_t$ are given by

$$\operatorname{sign}\left\{\frac{\partial \rho^S(\lambda_t)}{\partial \lambda_t}\right\} = \operatorname{sign}\Bigg\{u^P \left(\frac{(p^P - p^S)(1-2p^S)}{1-p^S}\right) + (1-p^P) u_1^P \left((1-\lambda_t) - \frac{2(p^P - p^S)}{\bar{\theta} h_1^P s_2^P}\right)\Bigg\} \tag{16}$$

and

$$\operatorname{sign}\left\{\frac{\partial \rho^R(\lambda_t)}{\partial \lambda_t}\right\} = \operatorname{sign}\left\{-\frac{\partial \bar{\theta}(\lambda_t)}{\partial \lambda_t}(1-p^S)\right\}, \tag{17}$$

where we have used that $(1-p^P)u^P = (1-p^S)u^S$ from (14) and $s_1^P/(-s_2^P) = 2(1-\lambda_t)/\bar{\theta}$ by the proof of Proposition 1.

Evaluated at $\lambda_t = 0$, all terms except u_1 in (16) are strictly positive.[27] Thus, $\partial \rho^S(0)/\partial \lambda_t$ is weakly positive if and only if for $\lambda_t = 0$ it holds that

$$u_1^P \geq -u^P \left(\frac{(p^P - p^S)(1-2p^S)}{1-p^S}\right) \left[(1-p^P)\left((1-\lambda_t) - \frac{2(p^P - p^S)}{\bar{\theta} h_1^P s_2^P}\right)\right]^{-1}. \tag{18}$$

Likewise, note that the sign of $\partial \rho^R / \partial \lambda_t$ is the opposite sign of $\partial \bar{\theta}(\lambda_t)/\partial \lambda_t$. Hence, because all terms except u_1 in (15) are strictly positive, $\partial \rho^R / \partial \lambda_t$ is weakly negative if and only if

$$u_1^P \geq \bar{\theta} h_1^p s_2^p u^p (1-p^P)^{-1}. \tag{19}$$

Let u' and u'' be the values of the right hand sides of (18) and (19) when evaluated at $\lambda_t = 0$. Then, from our discussion above it follows, that $\partial \rho^S(0)/\partial \lambda_t > 0$ and $\partial \rho^R(0)/\partial \lambda_t < 0$ if $u_1(0) > \bar{u} \equiv \max\{u', u''\}$. The converse—that is, $\partial \rho^S(0)/\partial \lambda_t < 0$ and $\partial \rho^R(0)/\partial \lambda_t > 0$—holds

[27] Note that $p^S = \bar{\theta} h(s^S) < 1/2$ for $\lambda_t = 0$ is not obvious. To see that this is indeed the case, assume to the contrary $p^S > 1/2$ implying $p^P = \bar{\theta}/2\, h(s^P) > 1/2$. By Proposition 4, $s^P = \bar{\theta}/2\, h(s^P) u(s^P) = p^P/2\, u(s^P) < 1/2$ and hence $u(s^P) < 1/p^P < 2$ by $p^P > 1/2$. Furthermore, optimality of $\bar{\xi} \equiv \xi(\bar{\theta})$ requires $(1-p^S)u(\bar{\xi}) \geq 1$, since an indirect utility of 1 is always attainable by setting $x = 1$. This implies $u(\bar{\xi}) \geq 2$ by $p^S > 1/2$. Thus, $p^S > 1/2$ implies $u(s^P) < 2 \leq u(\bar{\xi})$ for $\lambda_t = 0$. However, by Proposition 4, $s^P < 1/2 < \bar{\xi}$ such that $u(s^P) > u(\bar{\xi})$, a contradiction.

true, if $u_1(0) < \underline{u} \equiv \min\{u', u''\}$. Differentiability of ρ^S and ρ^R around 0 thus establishes the claim for all $\lambda_t \in [0, \underline{\lambda}]$ for some $\underline{\lambda} > 0$. *Q.E.D.*

References

Acemoglu, Daron, and James A. Robinson. 2000*a*. "Democratization or Repression?" *European Economic Review*, 44: 683–693.

Acemoglu, Daron, and James A. Robinson. 2000*b*. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics*, 115(4): 1167–1199.

Acemoglu, Daron, and James A. Robinson. 2001. "A Theory of Political Transitions." *American Economic Review*, 91(4): 938–963.

Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni. 2010. "A Theory of Military Dictatorships." *American Economic Journal: Macroeconomics*, 2(1): 1–42.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2008. "Coalition Formation in Non-Democracies." *Review of Economic Studies*, 75(4): 987–1009.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2012. "Dynamics and Stability of Constitutions, Coalitions, and Clubs." *American Economic Review*, 102(4): 1446–1476.

Aidt, Toke S., and Peter S. Jensen. 2012. "Workers of the World, Unite! Franchise Extensions and the Threat of Revolution in Europe, 1820-1938." *Cambridge Working Paper in Economics.*

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2006. "Signaling in a Global Game: Coordination and Policy Traps." *Journal of Political Economy*, 114(3): 452–484.

Besley, Timothy, Torsten Persson, and Marta Reynal-Querol. 2012. "Political Instability and Institutional Reform: Theory and Evidence." *Working paper.*

Boix, Carles. 2003. *Democracy and Redistribution.* Cambridge (UK):Cambridge University Press.

Bourguignon, François, and Thierry Verdier. 2000. "Oligarchy, Democracy, Inequality and Growth." *Journal of Development Economics*, 62(2): 285–313.

Cervellati, Matteo, Piergiuseppe Fortunato, and Uwe Sunde. 2012. "Consensual and Conflictual Democratization." *The B.E. Journal of Theoretical Economics (Contributions)*, 12(1).

Cervellati, Matteo, Piergiuseppe Fortunato, and Uwe Sunde. forthcoming. "Violence during Democratization and the Quality of Democratic Institutions." *European Economic Review.*

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics*, 102(2): 179–221.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers*, 56(4): 563–595.

Conley, John P., and Akram Temini. 2001. "Endogenous Enfranchisement when Groups' Preferences Conflict." *Journal of Political Economy*, 109(1): 79–102.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies*, 80(4): 1422–1458.

Elkins, Zachary, Tom Ginsburg, and James Melton. 2010. "Chronology of Constitutional Events, Version 1.1."

Ellis, Christopher J., and John Fender. 2011. "Information Cascades and Revolutionary Regime Transitions." *Economic Journal*, 121(553): 763–792.

Gates, Scott, Håvard Hegre, Mark P. Jones, and Håvard Strand. 2006. "Institutional Inconsistency and Political Instability: Polity Duration, 1800–2000." *American Journal of Political Science*, 50(4): 893–908.

Goemans, Henk E., Kristian S. Gleditsch, and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research*, 46(2): 269–283.

Gradstein, Mark. 2007. "Inequality, Democracy and the Protection of Property Rights." *Economic Journal*, 117(516): 252–269.

Haber, Stephen, and Victor Menaldo. 2011. "Do Natural Resources Fuel Authoritarianism? A Reappraisal of the Resource Curse." *American Political Science Review*, 105(1): 1–26.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century.* Norman (OK):University of Oklahoma Press.

Jack, William, and Roger Lagunoff. 2006. "Dynamic Enfranchisement." *Journal of Public Economics*, 90(4): 551–572.

Justman, Moshe, and Mark Gradstein. 1999. "The Industrial Revolution, Political Transition, and the Subsequent Decline in Inequality in 19th-Century Britain." *Explorations in Economic History*, 36(2): 109–127.

Karl, Terry Lynn. 1990. "Dilemmas of Democratization in Latin America." *Comparative Politics*, 23(1): 1–21.

Kuran, Timur. 1989. "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice*, 61(1): 41–74.

Lizzeri, Alessandro, and Nicola Persico. 2004. "Why did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'." *Quarterly Journal of Economics*, 119(2): 707–765.

Llavador, Humberto, and Robert J. Oxoby. 2005. "Partisan Competition, Growth, and the Franchise." *Quarterly Journal of Economics*, 120(3): 1155–1189.

Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91." *World Politics*, 47(1): 42–101.

Magaloni, Beatriz, and Ruth Kricheli. 2010. "Political Order and One-Party Rule." *Annual Review of Political Science*, 13(1): 123–143.

Mailath, George J. 1987. "Incentive Compatibility in Signaling Games with a Continuum of Types." *Econometrica*, 55(6): 1349–1365.

Marshall, Monty, and Keith Jaggers. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2002."

O'Donnell, Guillermo, and Philipp C. Schmitter. 1973. *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies.* Baltimore (MD):John Hopkins University Press.

Przeworski, Adam. 2000. *Democracy and Development: Political Institutions and Material Well-Being in the World, 1950-1990.* Cambridge (UK):Cambridge University Press.

Przeworski, Adam. 2009. "Conquered or Granted? A History of Suffrage Extensions." *British Journal of Political Science*, 39(2): 291.

Ross, Michael L. 2001. "Does Oil Hinder Democracy?" *World Politics*, 53(3): 325–361.

Rustow, Dankwart A. 1970. "Transitions to Democracy: Toward a Dynamic Model." *Comparative Politics*, 2(3): 337–363.

Selten, Reinhard. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, 4(1): 25–55.