# Online Appendix: A Quantitative Theory of Political Transitions


Lukas Buchheim — LMU Munich

Robert Ulbricht — Boston College


September 10, 2019

## A Data and Additional Empirical Results

### A.1 Data

Our data combines information from three sources. It covers all *regime spells,* as collected by Geddes, Wright and Frantz (2014, GWF), that existed on January 1st of each year between 1946 and 2010 in countries with more than one million inhabitants.[1] GWF define regimes by "the rules that identify the group from which leaders can come and [that] determine who influences leadership choice and policy" (p. 314).[2] Since GWF lists non-autocratic regimes only at a yearly frequency, we impute the begin (end) dates for non-autocratic regimes from (i) the end (begin) dates of the previous (next) regime and (ii) the begin date of the nearest Polity IV case within the same year. If (i) and (ii) yield no match, we encode the begin dates as July 1st and the end dates as June 30th.

We measure the inclusiveness of regimes using the polity score, normalized between 0 and 1, from the Polity IV Project (Marshall, Gurr and Jaggers, 2017), which ranks political regimes on a 21 point scale between autocratic and democratic. Specifically, we merge all polity spells listed in the "Polity IV Polity-Cases" dataset to our sample of GWF regime spells, harmonizing start dates on the basis of GWF spells whenever the start date of a polity case is within half a year (183 days) of a GWF start date. Otherwise, we keep track of changing polity scores by subdividing GWF spells into subspells.[3]

[1] While the published GWF database only includes regimes that existed on January 1st of a given year, the accompanying code book also categorizes spells lasting less than a year when classified as autocratic. We complete the dataset by manually coding spells lasting less than a year as categorized in the GWF codebook; if no information is provided, we code them based on their description in the codebook for the Archigos database of political leaders.

[2] Note that by focusing on the ruling group, the definition allows for leadership succession within regimes (if the identity of the ruling group remains unchanged) as well as regime changes without leadership replacement (if the leader stays in power despite a change in the ruling group, e.g., via reforms). Similarly, the definition allows for transitions that lead to a succession of regimes with similar scores of political inclusiveness.

[3] For some polity spells, Polity IV assigns "standardized authority scores" that do not fall into the autocracy–democracy range. The score of -66 encodes foreign "interruption", which we encode as missing. The polity scores of -77 ("interregnum") and -88 ("transition") identify transitional episodes. We interpret GWF regime transitions that occur during a transitional polity episode as the event defining the polity transition. Accordingly, transitional episodes just before a GWF transition are encoded with the last non-transitional polity score within the old GWF regime, while instances of transitional episodes just after a GWF transition are encoded via the first instance of a non-transitional

Third, we classify the GWF regime transitions primarily based on information provided by GWF (variable "howend"). If the information in GWF is unavailable, we match the GWF transitions to the nearest leader exit, taken from the Archigos database of political leaders by Goemans, Gleditsch and Chiozza, 2009, within half a year, and use the variables on the types of exit and entry to label the regime transitions. All violent regime transitions that are accompanied by popular protest, civil war, or coups are classified as *revolts*. Peaceful transitions where political insiders either actively change rules or newly allow for competitive elections or where there is no irregular leader change, are labeled *democratic reforms* when accompanied by an increase in the polity score and *autocratic consolidations* when accompanied by a decreasing polity score. All transitions influenced by foreign governments are called *foreign imposition*. All remaining transitions are collected in the residual category *other*.

The resulting database covers 494 regime spells in 155 countries covering a total of 8843.87 country-years.

## A.2 Estimation of Transition Hazards and Robustness

**Transition hazards and regime maturity** The hazards, reported in Figure 1, are estimated by differencing and smoothing over Nelson-Aalen estimates for the cumulative hazard rate, correcting for left and right censoring. Here we explore the robustness of the findings controlling for polity and region fixed effects.[4] Specifically, we use a Cox proportional hazard model, with hazard rate

$$p^s(\tau_{i,t}|\lambda_{i,t}, r_i) = h(\tau_{i,t})\exp(f(\lambda_{i,t}) + r_i) \quad \text{for } s \in \{\text{reform, revolt}\}, \tag{A.1}$$

where $h(\tau_{i,t})$ is the baseline hazard, identified non-parametrically as a function of maturity $\tau_{i,t}$, $f$ is a cubic spline in polity $\lambda_{i,t}$, and $r_i$ are the region fixed effects.[5] Figure A.1 plots the baseline hazard rates $h$ for revolts and reforms, respectively. The results are similar to the ones in Figure 1, albeit with slightly larger confidence intervals (the loss in precision is expected given the small number of transition events and the large number of explanatory variables included in the current specification).

**Transition hazards by regime type** Specification (A.1) already controls for the political system, but continues to impose a baseline hazard $h$ that is independent of $\lambda$. To explore inasmuch the documented stabilization patterns equally apply to autocracies and democracies, we also compute

polity score within the new GWF regime. (If the new GWF regime does not include a non-transitional polity score, we use the date of the next GWF transition to assign a date to the polity transition.) Finally, transitional episodes within a given GWF regime spell are encoded using the subsequent polity score.

[4] Region definitions are based on the United Nations geoscheme, which we use to define 10 distinct regions in total (Eastern Europe, Eastern and Central Asia, Middle America, Northern Africa and Arabic Peninsula, South America, South-Eastern Asia, Western and Central Africa, Western Europe, Western Offshoots). Note that disentangling the geographic controls further is likely to cause incidental parameters problems (biasing our estimators given their nonlinearity), as already the region fixed effects turn out to be only weakly identified in our specifications.

[5] Following the recommendations of Harrell (2001), the cubic spline has five knots located at the 5th, 27.5th, 50th, 72.5th and 95th percentile of the distribution of polity (corresponding to the values 0.05, 0.15, 0.45, 0.90, and 1, respectively).

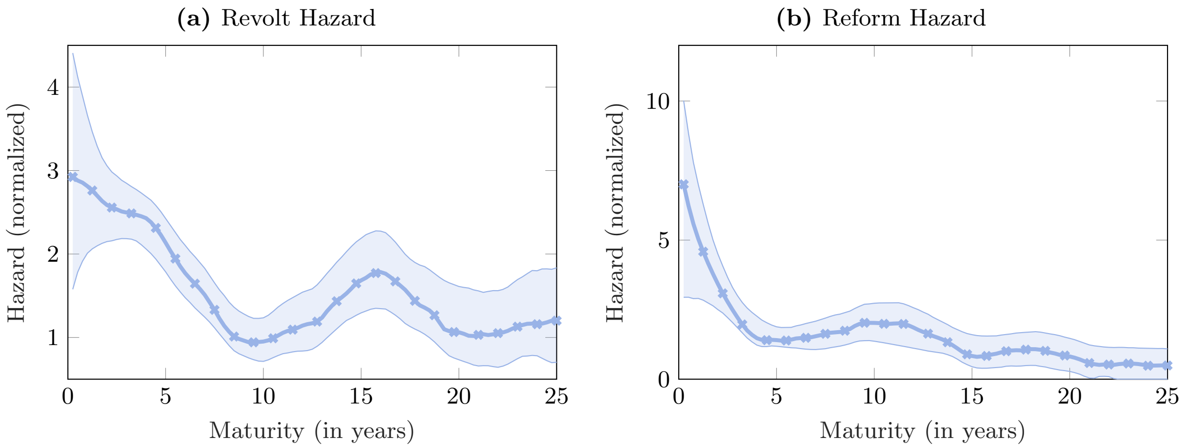


**Figure A.1:** Empirical transition hazards for revolts (left panel) and reforms (right panel). Notes.—Hazards are the baseline hazards estimated via (A.1), normalized relative to the unconditional hazard of revolts and reforms, respectively. Shaded bands correspond to 80 percent bootstrap confidence intervals, clustered at the country level.

the hazard rate separately for regimes with $\lambda \leq .5$ and $\lambda > .5$. The results for the combined revolt and reform hazards are shown in Figure A.2 (all results continue to hold if we separate the hazard by both origin and mode of transition). It is evident that the stabilization equally applies to regimes on the autocratic and democratic side of the spectrum.

**Hazard ratios of political systems** To estimate the relation between political inclusiveness and transition hazards, reported in Figure 2, we re-estimate (A.1) for the combined failures due to reforms and revolts. The estimated relationship is given by the cubic spline $f$.

# B Mathematical Appendix

## B.1 Effectiveness of Reforms

Here we show formally that outsiders have no incentives to ever refuse becoming enfranchised. The argument also implies that agents born as insiders never choose to rebel if given the choice. To show this, we need to show that

$$(1 - p(\cdot, x_t))u(x_t) \geq \max\{\hat{\theta}_t \psi(s_t), \gamma_{it}\}.$$

A lower bound on the utility as an enfranchised insider is $u(1)$, because $x_t = 1$ is in the choice set of insiders; i.e., $(1 - p(\cdot, x_t))u(x_t) \geq (1 - p(\cdot, 1))u(1) = u(1)$. When the best outside option is to not support a revolt, the result trivially follows from $u(1) \geq \gamma_{it}$ for all $i$ and $t$. For the case, where an outsider's best outside option is to revolt, an upper bound on the utility is given by $\psi(1) = h(1)u(1)$, because by Assumption 1 revolts are more rewarding when they have more supporters; i.e., $\hat{\theta}_t \psi(s_t) \leq \psi(s_t) \leq \psi(1)$. Noting that $h(1) \leq 1$ gives the result.

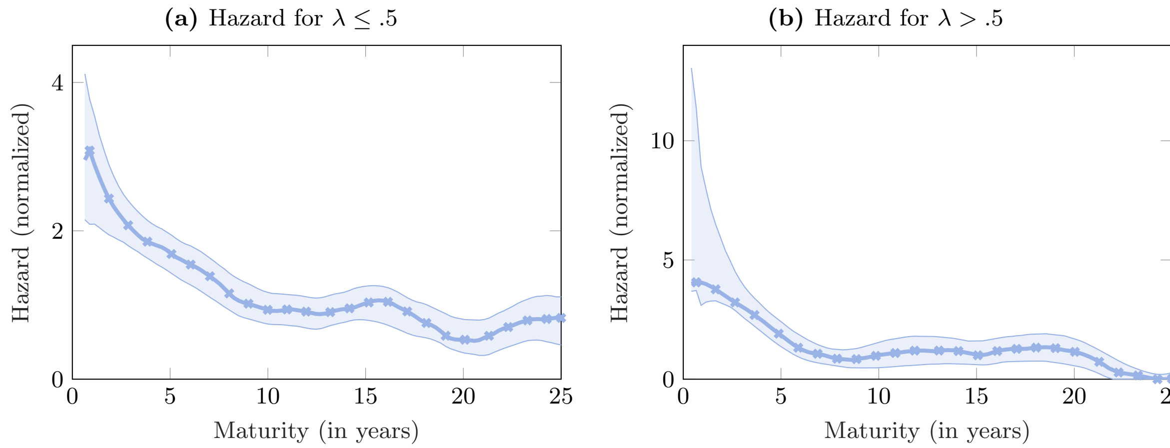


**Figure A.2:** Empirical transition hazards for regimes on autocratic side (left panel) and democratic side (right panel) of the polity spectrum. Notes.—Hazards are normalized relative to the unconditional hazard for autocracies and democracies, respectively. Shaded bands correspond to 80 percent bootstrap confidence intervals, clustered at the country level.

## B.2 Proof of Proposition 1

Using (7), we can rewrite (8) as

$$\pi(s, x, \hat{\theta}) \equiv s - \min\{f(s, (1-x)\hat{\theta}), 1-x\} = 0 \tag{B.1}$$

for

$$f(s, y) \equiv y\psi(s).$$

Recall that $u(1) = 1$ and $h(\cdot) \in [0, 1]$. By Assumption 1, $f$ is therefore increasing and (weakly) concave in both arguments. Accordingly, $f(s, y) \leq y$, allowing us to drop the min-operator from (B.1).

As $f(0, y) = 0$ for all $y \in [0, 1]$, there always exists a solution to (B.1) at $s = 0$. We distinguish two cases. First, let $y = 0$ (i.e,. when $\hat{\theta} = 0$ or $x = 1$). Then $f(s, y) = 0$ for all $s$, so that $s = 0$ is the unique stable solution to (B.1).

Second, let $y > 0$. By Assumption 1, $f_1(0, y) > 1$, so that $s = 0$ is unstable.[6] We now show the existence of a unique stable fixed point $s > 0$. Specifically, $f_1(0, y) > 1$ implies that $f(\tilde{s}, y) > \tilde{s}$ for $\tilde{s} \searrow 0$ and any $y > 0$. On the other hand, as noted above, $f(\tilde{s}, y) \leq y \leq \tilde{s}$ for $\tilde{s} \nearrow 1$. Continuity of $\psi$ (and thus of $f$), therefore imply the existence of a fixed point $s^* > 0$. Monotonicity and concavity of $f$ further imply that $s^*$ is unique on $(0, 1]$. Clearly, it must hold $f_1(s^*, y) < 1$, and so $s^*$ is stable.

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

[6] I.e., iteratively best responding to any perceived $\hat{s} = \varepsilon > 0$ leads to a distinct equilibrium $s^* > 0$ described in the following.

(B.1) implies that

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

and

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

where

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

Since $\psi$ is bounded by $\psi(1) = 1$, (8) 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 $\hat{\theta}_t \neq 0$ and $x_t \neq 1$. From the implicit function theorem it then follows that $s$ is differentiable on its whole support. As noted above, $f_1(s_t, y) < 1$, implying $\bar{\gamma}_2(\hat{\theta}_t, s_t) < (1 - x_t)^{-1}$ at $s_t = s^*$. Thus $\partial \pi_t / \partial s_t > 0$ for all $(\hat{\theta}_t, x_t) \in [0, 1]^2$, which yields the desired results.

Finally, while we developed the proof for 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.

## B.3 Proof of Proposition 2

The proof proceeds by a series of lemmas. To simplify notation, we henceforth drop $(\lambda_t, F_t)$ as arguments of $x$, $\hat{\theta}$ and $\bar{\theta}$ where no confusion arises. Furthermore, we use $\tilde{V}^I(\theta, \hat{\theta}, x) \equiv V^I(\theta h(s(\hat{\theta}, x)), x) = (1 - \theta h(s(\hat{\theta}, x))) \, u(x)$ to denote insider's expected indirect utility, where $s$ is as given by Proposition 1.

**Lemma B.1.** *$x$ is weakly increasing in $\theta_t$.*

*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 $\theta''$. $\square$

**Lemma B.2.** *Suppose $x$ is discontinuous at $\theta'$, 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 $\tilde{\theta}$,

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

Moreover, arguing as in the proof of Lemma B.1,

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

Hence, if $\tilde{V}^I(\theta'',\tilde{\theta},x') \geq \tilde{V}^I(\theta'',\hat{\theta}(x^+),x^+) = V^*(\theta'')$, then $\tilde{V}^I(\theta',\tilde{\theta},x') > \tilde{V}^I(\theta',\hat{\theta}(x^+),x^+) = V^*(\theta')$. Therefore, $D_{\theta'',x'}$ is a proper subset of $D_{\theta',x'}$ if $\theta'' > \theta'$, ruling out $\hat{\theta}(x') > \theta'$ by the D1 criterion.[7] 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 B.3.** *There exists $\bar{\theta}_t > 0$, such that $x_t = \lambda_t$ for all $\theta_t \leq \bar{\theta}_t$. Moreover, $x(\theta'') > x(\theta') > \lambda_t + \mu$ for all $\theta'' > \theta' > \bar{\theta}_t$ 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 B.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 $\theta'' > \theta' > \bar{\theta}_t$ 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 $\theta'$ is the lowest state in this pool. Then Bayesian updating implies that $\theta^+ \equiv \hat{\theta}(x^+) \geq \mathbb{E}_{F_t}\{\theta_t | \theta'' \geq \theta_t \geq \theta'\} > \theta'$ and, therefore, $\tilde{V}^I(\theta', \theta^-, x^+) > \tilde{V}^I(\theta', \theta^+, x^+)$ for all $\theta^- \leq \theta'$. Hence, because $\theta'$ 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 B.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 $\theta'$ from choosing $x'$ it must be that $\hat{\theta}(\lambda_t, x', F_t) > \theta'$. However, from Lemma B.2 we have that $\hat{\theta}(x') = \theta'$, a contradiction.

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

**Lemma B.4.** *$x$ is continuous and differentiable in $\theta_t$ on $(\bar{\theta}_t, 1]$.*

*Proof.* Consider continuity first and suppose to the contrary that $x$ has a discontinuity at $\theta' \in (\bar{\theta}_t, 1)$. By Lemma B.1, $x$ is monotonically increasing in $\theta_t$. Hence, because $x$ is defined on an interval, it follows that for any discontinuity $\theta'$, $x^- \equiv \lim_{\varepsilon \uparrow 0} x(\theta')$ 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 B.2 and B.3 it

[7] The D1 criterion requires that 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. Formally, let $V^*(\theta) \equiv \mathbb{E}\{V^I(\eta, x^*(\theta, \lambda)) | \theta\}$ be the insiders' expected payoff in state $\theta$ under a candidate equilibrium $x^*$. Then the D1 criterion restricts beliefs for off-equilibrium events $x'$ to states $\theta'$ that maximizes $D_{\theta',x'} = \{\hat{\theta} : \mathbb{E}\{V^I(\eta, x') | \theta', s = s(\hat{\theta}, x')\} \geq V^*(\theta')\}$ in the sense that there is no $\theta''$ such that $D_{\theta',x'}$ is a proper subset of $D_{\theta'',x'}$.

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 $\theta$ in the neighborhood of $\theta'$ 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 B.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 $\theta' > \bar{\theta}_t$, 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

$$0 \geq g(\theta', \theta'', x(\theta'')) \geq -g_1(\theta', \theta'', x(\theta''))(\theta'' - \theta') - \tfrac{1}{2} g_{11}(a)(\theta'' - \theta')^2,$$

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 $g_1(\theta', \theta'', x(\theta''))$, and noting that $g(\theta', \theta', x(\theta')) = g_1(\theta', \theta', x(\theta')) = 0$, these inequalities can be written as

$$\begin{aligned} 0 \geq\, & g_2(\theta', \theta', x(\theta')) + \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'} \times [g_3(\theta', \theta', x(\theta')) \\ & + \tfrac{1}{2} g_{33}(b(\beta))(x(\theta'') - x(\theta')) + g_{23}(b(\beta))(\theta'' - \theta')] + \tfrac{1}{2} g_{22}(b(\beta))(\theta'' - \theta') \\ \geq\, & -[g_{12}(b(\beta')) + \tfrac{1}{2} g_{11}(a)](\theta'' - \theta') - g_{13}(b(\beta'))(x(\theta'') - x(\theta')), \end{aligned}$$

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 $\theta'' \to \theta'$,

$$0 \geq g_2(\theta', \theta', x(\theta')) + \lim_{\theta'' \to \theta'} \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'}\, g_3(\theta', \theta', x(\theta')) \geq 0.$$

By Lemma B.3, $x$ and, hence, $\hat{\theta}$ are strictly increasing for all $\theta \geq \bar{\theta}(\lambda_t, F_t)$. Arguing similarly as we did to show continuity, optimality of $x$, therefore, requires that $g_3 = \tilde{V}_3^I(\theta, \theta, x) \neq 0$ and, hence, the limit of $(x(\theta'') - x(\theta'))/(\theta'' - \theta')$ is well defined, yielding

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = -\frac{\tilde{V}_2^I(\theta, \theta, x)}{\tilde{V}_3^I(\theta, \theta, x)}. \tag{B.2}$$

□

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

*Proof.* From Lemma B.4 we have that $\xi$ is differentiable, and by Lemma B.3, $\partial\xi/\partial\theta_t > 0$. We thus only need to show that $\xi$ is unique. By the proof to Lemma B.4, $\mathrm{d}x/\mathrm{d}\theta_t$ is pinned down by the partial differential equation (B.2), which must hold for all $\theta_t > \bar{\theta}_t$. Moreover, whenever $\bar{\theta}_t < 1$, in equilibrium $\hat{\theta}(x(1)) = 1$ and, therefore, it obviously must hold that $x(1) = \arg\max_{x_t} \tilde{V}^I(1, 1, x_t)$, providing a boundary condition for (B.2). Because $\tilde{V}^I$ is independent of $(\lambda_t, F_t)$, it follows that $x(\theta_t)$ is uniquely characterized by a function, i.e., $\xi : \theta_t \mapsto x_t$, for all $\theta_t > \bar{\theta}_t$. □

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

*Proof.* Suppose to the contrary that $\bar{\theta}(\lambda_t, F_t)$ is not unique. Then there exist $\bar{\theta}'' > \bar{\theta}'$, defining two distinct equilibria for a given $\lambda_t$. By Lemma B.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 $\tilde{V}^I(\theta, \theta, \xi(\theta)) \geq \tilde{V}^I(\theta, \mathbb{E}_{F_t}\{\theta_t | \theta_t \leq \bar{\theta}'\}, \lambda_t)$ in the equilibrium defined by $\bar{\theta}'$, and $\tilde{V}^I(\theta, \theta, \xi(\theta)) \leq \tilde{V}^I(\theta, \mathbb{E}_{F_t}\{\theta_t | \theta_t \leq \bar{\theta}''\}, \lambda_t)$ in the equilibrium defined by $\bar{\theta}''$. However, $\tilde{V}^I(\theta, \mathbb{E}_{F_t}\{\theta_t | \theta_t \leq \bar{\theta}'\}, \lambda_t) > \tilde{V}^I(\theta, \mathbb{E}_{F_t}\{\theta_t | \theta_t \leq \bar{\theta}''\}, \lambda_t)$, a contradiction. □

This establishes uniqueness of $x(\theta_t, \lambda_t, F_t)$, with all properties given by Lemmas B.3 and B.5, and the corresponding beliefs $\hat{\theta}(\lambda_t, x_t, F_t)$ following from Lemma B.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.

Replicating the proof of Lemma B.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 $\theta$ mixes in a non-degenerate way, (ii) further implies that there can be only finitely many types that mix on the closed interval $[0, 1]$. The logic of Lemmas B.2, B.3, and B.4 then apply, ruling out any jumps of $\tilde{x}$ on $[\bar{\theta}(\lambda_t, F_t), 1]$. This leads to the conclusion that at most a mass zero of types (i.e., $\theta_t = \bar{\theta}(\lambda_t, F_t)$) could possibly mix in any equilibrium (with no impact on $\hat{\theta}$) and, thus, there is no need to consider any non-degenerate mixed strategies.

## B.4 Proof of Proposition 3

Case (i) follows trivial, as here the state is revealed through insiders' reforms. Cases (ii) and (iii) are a straightforward application of Bayes' law. In particular, for any $\vartheta \leq \bar{\theta}_t$, we get

$$\tilde{F}_t(\vartheta | \eta_t = 1, x_t = \lambda_t) = \frac{\int_0^{\vartheta} p(\theta, s) \mathrm{d}F_t(\theta)}{\int_0^{\bar{\theta}_t} p(\theta, s) \mathrm{d}F_t(\theta)} = \frac{F_t(\vartheta)}{F_t(\bar{\theta}_t)} \frac{M_t^1(\vartheta)}{M_t^1(\bar{\theta}_t)}$$

and

$$\tilde{F}_t(\vartheta | \eta_t = 0, x_t = \lambda_t) = \frac{\int_0^{\vartheta} (1 - p(\theta, s_t)) \mathrm{d}F_t(\theta)}{\int_0^{\bar{\theta}_t} (1 - p(\theta, s_t)) \mathrm{d}F_t(\theta)} = \frac{F_t(\vartheta)}{F_t(\bar{\theta}_t)} \cdot \frac{1 - h(s_t) M_t^1(\vartheta)}{1 - h(s_t) M_t^1(\bar{\theta}_t)}.$$

Note that by letting $\int \mathrm{d}F$ denote the Lebesgue integral, the derivation applies for arbitrary, not necessarily continuous, probability measures $F_t$.

### B.5 Derivation of Equations (11) and (12)

In case (i), $\hat{F}_t|(x_t > \lambda_t)$ is a single mass point on $\theta_t$, so trivially $\tilde{\mu}_t = \theta_t$ and $\tilde{\sigma}_t^2 = 0$.

Consider, case (ii) next. From Proposition 3, we have that for all $\theta \leq \bar{\theta}_t$,

$$\mathrm{d}\tilde{F}_t(\theta) = \frac{1}{F_t(\bar{\theta}_t)M_t^1(\bar{\theta}_t)}\mathrm{d}\big(F_t(\theta)M_t^1(\theta)\big)$$

where, using the definition of $M_t^1$,

$$\mathrm{d}\big(F_t(\theta)M_t^1(\theta)\big) = \mathrm{d}\int_0^\theta \vartheta \mathrm{d}F_t(\vartheta) = \theta \mathrm{d}F_t(\theta).$$

Computing the $i$-the raw moment of $\tilde{F}_t$, we have

$$\mathbb{E}_{\tilde{F}_t}\{\theta^i|\theta \leq \bar{\theta}_t\} = \frac{1}{F_t(\bar{\theta}_t)M_t^1(\bar{\theta}_t)}\int_0^{\bar{\theta}_t} \theta^{i+1}\mathrm{d}F_t(\theta) = \frac{M_t^{i+1}(\bar{\theta})}{M_t^1(\bar{\theta})}$$

and, accordingly,

$$\begin{aligned}
\tilde{\mu}_t|(\eta_t = 1, x_t = \lambda_t) &= \frac{M_t^2(\bar{\theta})}{M_t^1(\bar{\theta})} \\
\tilde{\sigma}_t^2|(\eta_t = 1, x_t = \lambda_t) &= \frac{M_t^3(\bar{\theta})}{M_t^1(\bar{\theta})} - \tilde{\mu}_t^2.
\end{aligned}$$

Case (iii) is analyzed analogously. For all $\theta \leq \bar{\theta}_t$, the probability measure is given by

$$\mathrm{d}\tilde{F}_t(\theta) = \frac{1}{F_t(\bar{\theta}_t)\left(1 - h(s_t)M_t^1(\bar{\theta}_t)\right)}\mathrm{d}\Big(F_t(\theta)\left(1 - h(s_t)M_t^1(\theta)\right)\Big)$$

where

$$\mathrm{d}\Big(F_t(\theta)\left(1 - h(s_t)M_t^1(\theta)\right)\Big) = \mathrm{d}F_t(\theta) - h(s_t)\mathrm{d}\int_0^\theta \vartheta \mathrm{d}F_t(\vartheta) = (1 - \theta h(s_t))\mathrm{d}F_t(\theta).$$

The $i$-th raw moment is thus given by

$$\begin{aligned}
\mathbb{E}_{\tilde{F}_t}\{\theta^i|\theta \leq \bar{\theta}_t\} &= \frac{1}{F_t(\bar{\theta}_t)\left(1 - h(s_t)M_t^1(\bar{\theta}_t)\right)}\int_0^{\bar{\theta}_t} (\theta^i - \theta^{i+1}h(s_t))\mathrm{d}F_t(\theta) \\
&= \frac{M_t^i(\bar{\theta}_t) - h(s_t)M_t^{i+1}(\bar{\theta}_t)}{1 - h(s_t)M_t^1(\bar{\theta}_t)}
\end{aligned}$$

and, hence,

$$\tilde{\mu}_t|(\eta_t = 0, x_t = \lambda_t) = \frac{M_t^1(\bar{\theta}_t) - h(s_t)M_t^2(\bar{\theta}_t)}{1 - h(s_t)M_t^1(\bar{\theta}_t)}$$
$$\tilde{\sigma}_t^2|(\eta_t = 0, x_t = \lambda_t) = \frac{M_t^2(\bar{\theta}_t) - h(s_t)M_t^3(\bar{\theta}_t)}{1 - h(s_t)M_t^1(\bar{\theta}_t)} - \tilde{\mu}_t^2.$$

### B.6 Proof of Proposition 4

The proposition is a straightforward corollary to Propositions 1–3: From Proposition 1 and 2, there exists a unique mapping from $\mathcal{S}_t$ to $\{\{\phi_{it}\}_{i\in[0,1]}, s_t, x_t\}$, which further implies a unique (stochastic) mapping from $\mathcal{S}_t$ to $\eta_t$. Proposition 3, in turn, implies that there exists a unique mapping from $(\mathcal{S}_t, x_t, \eta_t)$ to $\mathcal{S}_{t+1}$. As $\mathcal{S}_t$ is purely-backward looking, we conclude that for any $\mathcal{S}_0$ there exists a unique stochastic equilibrium process.

### B.7 Proof of Proposition 5

Given $\Pr(\theta_t = \theta_{t-1}) \to 1$, we have that $\theta_0 = \theta_1 = \cdots \equiv \theta$ almost surely. We prove two versions of the proposition. Our preferred version amounts to the case where $\theta$ is fixed across time, but is unobserved by the statistician. To estimate transition hazards, the statistician treats $\theta$ as hidden state and refines his estimate for $\theta$ based on the realizations of $(x_t, s_t, \eta_t)$. Accordingly, in our preferred version of the proposition, the statistical probability measure at date $t$ coincides with outsiders' prior $F_t$. However, the proposition also holds conditional on a given realization $\theta_0$ that is known to the statistician. To show this, we first prove the result for a fixed $\theta_0$, and then derive the more general result where $\theta_0$ is treated as hidden state as a corollary.

**Case 1 (fixed $\theta_0$).** Fix some $(\theta_0, \lambda_0, F_0)$, and let $\bar{\theta}_0$ define the pooling threshold as in Proposition 2. We tacitly assume $\theta_0 < \bar{\theta}_0$, so that there is indeed no reform at $t = 0$.

Consider any $t > 0$ and suppose there was no transition until until $t-1$. From $\Pr(\theta_t = \theta_{t-1}) \to 1$, $F_t(\bar{\theta}_{t-1}) = \tilde{F}_{t-1}(\bar{\theta}_{t-1})$. Proposition 3 then implies

$$F_t(\bar{\theta}_{t-1}) = \tilde{F}_{t-1}(\bar{\theta}_{t-1}) = 1, \tag{B.3}$$

so that

$$M_t^1(\bar{\theta}_{t-1}) = \mathbb{E}_{F_t}\{\theta|\theta \le \bar{\theta}_{t-1}\} = \mathbb{E}_{F_t}\{\theta\}. \tag{B.4}$$

Moreover, from (11),

$$\mathbb{E}_{F_t}\{\theta\} = \frac{M_{t-1}^2(\bar{\theta}_{t-1})}{M_{t-1}^1(\bar{\theta}_{t-1})}. \tag{B.5}$$

Combining (B.4) and (B.5) and noting that $\mathrm{Var}_{F_t}[\theta|\theta \le \vartheta] = M_t^2(\vartheta) - [M_t^1(\vartheta)]^2 > 0$ implies $M_t^2(\vartheta) > [M_t^1(\vartheta)]^2$, we have

$$M_t^1(\bar{\theta}_{t-1}) = \frac{M_{t-1}^1(\bar{\theta}_{t-1}) - h(s_{t-1})M_{t-1}^2(\bar{\theta}_{t-1})}{1 - h(s_{t-1})M_{t-1}^1(\bar{\theta}_{t-1})} < M_{t-1}^1(\bar{\theta}_{t-1}). \tag{B.6}$$

Further noting that from (B.3), $M_t^1(\bar{\theta}_{t-1}) = M_t^1(\bar{\theta}_t)$ for all $\bar{\theta}_t \geq \bar{\theta}_{t-1}$, we conclude that

$$\text{haz}_t^{\text{rev}} = \theta_0 h(s(M_t^1(\bar{\theta}_t), \lambda_0)) < \theta_0 h(s(M_{t-1}^1(\bar{\theta}_{t-1}), \lambda_0)) = \text{haz}_{t-1}^{\text{rev}} \tag{B.7}$$

if $\bar{\theta}_t \geq \bar{\theta}_{t-1}$.

To complete the proof of the first step, we need to show that $\bar{\theta}_t \geq \bar{\theta}_{t-1}$, implying that (B.7) indeed holds, and further implying that

$$\text{haz}_t^{\text{ref}} \leq \Pr(\theta_0 \geq \bar{\theta}_t) = 0.$$

To see that this is true, note that from (B.7),

$$\tilde{V}^I(\bar{\theta}_{t-1}, \hat{\theta}(\lambda_0, \lambda_0, F_t), \lambda) > \tilde{V}^I(\bar{\theta}_{t-1}, \hat{\theta}(\lambda_0, \lambda_0, F_{t-1}), \lambda),$$

implying that the right-hand side of condition (9) is increased from $t$ to $t-1$ at $\bar{\theta} = \bar{\theta}_{t-1}$. As the left-hand side of (9) is constant in $t$, it thus must hold that $\bar{\theta}_t > \bar{\theta}_{t-1}$.

**Case 2 ($\theta_0$ is a hidden state)** Now suppose that the statistician does not know the realization of $\theta_0$. Instead they filter through the realized history of the economy, summarized by $(x_\tau, s_\tau, \eta_\tau)_{\tau<t}$, to compute a probability measure for $\theta_0$ and the corresponding transition hazards. As the realized history coincides with outsiders' information set, the statistical probability measure is simply given by $F_t$. Specifically, transition hazards at date $t$ are given by

$$\begin{aligned}
\text{haz}_t^{\text{ref}} &= \int_{\bar{\theta}_t}^1 (1 - \theta h(s(\theta, \xi(\theta))))\, \mathrm{d}F_t(\theta) \\
\text{haz}_t^{\text{rev}} &= \int_0^{\bar{\theta}_t} \theta h(s(M_t^1(\bar{\theta}_t), \lambda_0))\, \mathrm{d}F_t(\theta) + \int_{\bar{\theta}_t}^1 \theta h(s(\theta, \xi(\theta)))\, \mathrm{d}F_t(\theta).
\end{aligned}$$

Step 1 immediately implies that for all $t > 0$, in the absence of any prior transition,

$$\text{haz}_t^{\text{ref}} = 0$$

and

$$\begin{aligned}
\text{haz}_t^{\text{rev}} &= \int_0^{\bar{\theta}_t} \theta h(s(M_t^1(\bar{\theta}_t), \lambda_0))\, \mathrm{d}F_t(\theta) \\
&= h(s(M_t^1(\bar{\theta}_t), \lambda_0)) \cdot M_t^1(\bar{\theta}_t).
\end{aligned}$$

To conclude the proof, note that from above, $M_t^1(\bar{\theta}_t) < M_{t-1}^1(\bar{\theta}_{t-1})$, so that $\text{haz}_t^{\text{rev}}$ is again strictly decreasing in $t$ as both factors are decreasing in $M_t^1(\bar{\theta}_t)$. Intuitively, the first factor captures the decline in the revolt-hazard due to *outsiders* perceiving the regime as more stable (leading to a fall in $s_t$ over time). The second term captures the uncertainty by the statistician, who similarly to outsiders now also infers that the regime is more stable over time, further reducing the revolt-hazard over time compared to the fixed $\theta_0$-case above.

## B.8 Proof of Proposition 6

**Bounding reforms** Insiders' optimality condition implies that $\xi(\bar{\theta}_t)$ is effective in reducing revolutionary pressure; i.e.,

$$s(\bar{\theta}_t, \xi(\bar{\theta}_t)) < s(M_t^1(\bar{\theta}_t), \lambda_t). \tag{B.8}$$

From the proof of Proposition 1, we can write $s(\hat{\theta}, x) = \bar{s}(\omega)$ with $\omega = (1-x)\hat{\theta}$ and $\bar{s}' > 0$. Accordingly, (B.8) implies

$$(1 - \xi(\bar{\theta}_t))\bar{\theta}_t < (1 - \lambda_t) M_t^1(\bar{\theta}_t),$$

and so

$$x_t \geq \xi(\bar{\theta}_t) > 1 - (1 - \lambda_t)\frac{M_t^1(\bar{\theta}_t)}{\bar{\theta}_t} \equiv \bar{\lambda}_t^{\text{ref}}.$$

**Bounding revolts** From (7) and (8), $s_t$ solves the fixed-point equation

$$s_t = \omega_t \psi(s_t) \tag{B.9}$$

with $\omega_t = (1 - x_t)\hat{\theta}_t$. Let $\omega' > \omega$ and, correspondingly, let $s' > s$ as in (B.9). Then

$$s = \omega\psi\left(s\right) < \omega\psi(s') = \frac{\omega}{\omega'} s'. \tag{B.10}$$

Evaluating (B.10) for $\omega = (1 - \lambda_t)M_t^1(\bar{\theta}_t)$ and $\omega' = (1 - \lambda_t)\bar{\theta}_t > \omega$ yields

$$s(M_t^1(\bar{\theta}_t), \lambda_t) < \frac{M_t^1(\bar{\theta}_t)}{\bar{\theta}_t} s(\bar{\theta}_t, \lambda_t). \tag{B.11}$$

Similarly, evaluating (B.10) for $\omega = (1 - \lambda_t)\bar{\theta}_t$ and $\omega' = 1$, we have

$$s(\bar{\theta}_t, \lambda_t) < (1 - \lambda_t)\bar{\theta}_t s(1, 0) \leq (1 - \lambda_t)\bar{\theta}_t. \tag{B.12}$$

Combining (B.11) and (B.12), and recalling that optimization by insiders requires that $s_t$ is weakly below $s(M_t^1(\bar{\theta}_t), \lambda_t)$, yields

$$s_t \leq s(M_t^1(\bar{\theta}_t), \lambda_t) < (1 - \lambda_t)M_t^1(\bar{\theta}_t) \equiv \bar{\lambda}_t^{\text{rev}}.$$

### B.9 Proof of Proposition 7

Differentiating $\tilde{V}^I$ with respect to its third argument, we obtain

$$\lim_{x\to 1}\tilde{V}_3^I(1,1,x) = -\alpha_u - \lim_{x\to 1}\alpha_h s(1,x)^{\alpha_h-1}s_2(1,x),$$

or, substituting for $s_2$ as computed in the proof to Proposition 1 and observing that $x \to 1$ implies $s(1,x) \to 0$,

$$\lim_{x\to 1}\tilde{V}_3^I(1,1,x) = -\alpha_u + \lim_{x\to 1}\alpha_h s(1,x)^{2\alpha_h-1}\frac{1+\alpha_u}{1-(1+\alpha_u)\alpha_h\frac{1-x}{s(1,x)^{1-\alpha_h}}}. \tag{B.13}$$

Using L'Hospital's Rule, we get after some algebra

$$\lim_{x\to 1}\frac{1-x}{s(1,x)^{1-\alpha_h}} = \frac{1}{(1-\alpha_h)(1+\alpha_u)} - \frac{\alpha_h}{1-\alpha_h}\lim_{x\to 1}\frac{1-x}{s(1,x)^{1-\alpha_h}},$$

which has a unique fixed point at

$$\lim_{x\to 1}\frac{1-x}{s(1,x)^{1-\alpha_h}} = \frac{1}{1+\alpha_u}.$$

Substituting back into (B.13), we have that $\xi(1) = 1$ if and only if

$$\lim_{x\to 1}s(1,x)^{2\alpha_h-1} \geq \frac{\alpha_u}{1+\alpha_u}\frac{1-\alpha_h}{\alpha_h}.$$

Note that the right side of the inequality is *strictly* between zero and unity, as $0 < \alpha_u \leq \alpha_h$ given the properties imposed on $u$ and $h$. The left side of the inequality goes to zero for all $\alpha_h > .5$, goes to $\infty$ for all $\alpha_h < .5$, and is constant at unity for $\alpha_h = .5$. We conclude that $\xi(1) = 1$ if and only if $\alpha_h \leq .5$, implying that a regime with $\lambda \to 1$ emerges (almost surely) under the same conditions (as $G$ has full support on $[0,1]$).

## C Numerical Implementation

This section describes the algorithm used to solve and estimate the model.

**Solution to the model** We first describe how to solve the model for a given parametrization $\omega$. The solution is simplified by the block-recursivity of the overlapping generations structure, which let's us break down the algorithm into three successive steps.

*Step 1 (coordination problem).* We solve the functional fixed-point (8) for $s : (\hat{\theta}, x) \mapsto s$ using a spline collocation. Noting that $(1-x)\bar{\gamma}(\hat{\theta}, s) = yh(s)u(s)$ with $y = (1-x)\hat{\theta}$, we can reduce $s$ to a univariate function $\bar{s} : y \mapsto s$. We parametrize $\bar{s}$ using a septic spline with 34 interior break points, with parameters chosen to solve (8) on a fine grid on $[0,1]$. The procedure gives a very accurate approximation to $s$ (evaluating (8) on an equally-spaced grid with 1000 points on $[0, 1]$, yields a maximal error of less than $5 \cdot 10^{-7}$).

*Step 2 (signaling problem).* The solution to the signaling problem characterized by Proposition 2 breaks down in to two substeps. (i) Given $s$, we can solve for $\xi$ using a standard solver for ordinary differential equations. (ii) Given $s$ and $\xi$, $\bar{\theta}$ can be solved using a standard bisection method on $[0, 1]$.

*Step 3 (stationary distribution).* We approximate the stationary distribution on a $(N_\lambda \times N_\mu \times N_\sigma)$-point grid for $(\lambda, \mu, \sigma)$ with $N_\lambda = 21$, $N_\mu = 20$, $N_\sigma = 20$ (see the main body of the paper for more details). The laws of motion are given by (2), (13) and (14), with

$$M_t^i(\vartheta) = B(\vartheta, a_t + i, b_t)/B(\vartheta, a, b)$$

where $B$ is the incomplete Beta function with shape parameters chosen so that the first two moments of the corresponding Beta distribution coincide with $\mu$ and $\sigma^2$; i.e.,

$$a_t = \mu_t \left( \frac{\mu_t(1-\mu_t)}{\sigma_t^2} - 1 \right) \qquad b_t = (1-\mu_t) \left( \frac{\mu_t(1-\mu_t)}{\sigma_t^2} - 1 \right). \tag{C.1}$$

To compute the transition matrix $Q(\lambda_{t+1}, \mu_{t+1}, \sigma_{t+1} | \lambda_t, \mu_t, \sigma_t)$, we first solve the generation game conditional on $(\theta_t, \lambda_t, \mu_t, \sigma_t)$ and integrate out $\theta_t$ using $F_t$ as probability measure (see Footnote 17 for details). For each $(\lambda_{t+1}, \mu_{t+1}, \sigma_{t+1})$, we then discretize the resulting transition probabilities to the eight adjacent grid-points, $\{\lambda_i, \lambda_{i+i}\} \times \{\mu_i, \mu_{i+1}\} \times \{\sigma_i, \sigma_{i+1}\}$, assigning probabilities proportionately to their inverse Euclidean distance to the respective corners of the cube. Once we have $Q$, we first verify that there exist a single recurrent class, consisting of 3322 states at our estimate (the remainder 5078 states are not reached along the equilibrium path). Finally, we iterate on $Q$ until convergence, yielding the unique stationary distribution.

On a Thinkpad X230 with a i5-3230M, the whole process takes about 5 seconds to complete.

**Calibration** We use a combination of global and local minimization tools to solve (15). Specifically, we first use a particle swarm algorithm with 20 chains of 16 particles each to conduct a preliminary global search. The particles are initialized using scrambled Sobol quasi-random numbers, and evolve completely independent across the 20 chains. After running the particle swarm algorithm for up to 200 iterations, we then run 20 local optimizer, initialized at the 20 minima attained across the 16 particles by each of the 20 chains. Our estimator is the minimum across the 20 chains.

The process converged to the exact same estimate for the top 9 out of 20 chains. On two Xeon E5-2630 v4 processors (with 20 physical cores), the whole calibration took about 4.5 hours to complete.

# D Accurateness of Belief Approximation

For the quantitative exploration of the model, we track outsiders' beliefs over time by approximating the one-step ahead projection from the posterior $\tilde{F}_t$ (which we compute exactly as in Proposition 3) to the prior $F_{t+1}$ using a Beta distribution with moments matching (13) and (14). In this section, we explore the accurateness of this approximation. Overall, we find that the approximation of $F_{t+1}$ is remarkably exact, tracing the true prior almost perfectly.

## D.1 Beliefs after reforms

After an attempted or successful reform ($x_t > \lambda_t$, $\eta_t \in \{0, 1\}$), the current state of the regime is fully revealed. Accordingly, the exact prior at $t + 1$ is truncated normal with mean $\rho\theta_t + \mu_\epsilon$ and variance $\sigma_\epsilon^2$. For any interior $\vartheta \in (0, 1)$, the pdf is given by

$$f_{t+1}(\vartheta) = \phi_{\rho\theta_t+\mu_\epsilon,\sigma_\epsilon^2}(\vartheta),$$

where $\phi_{\mu,\sigma^2}$ denotes the density of a $(\mu, \sigma^2)$-normal distribution. At the boundaries, $\vartheta \in \{0, 1\}$, $F_{t+1}$ has mass points corresponding to the tails of $f_{t+1}$.

By contrast, the Beta approximation is given by

$$f_{t+1}^{\text{approx}}(\vartheta) = \beta_{\rho\theta_t+\mu_\epsilon,\sigma_\epsilon^2}(\vartheta),$$

where $\beta_{\mu,\sigma^2}$ denotes the density of a Beta distribution with mean $\mu$ and variance $\sigma^2$ (implemented by shape parameters as in (C.1)).

Panel (a) of Figure D.1 compares $f_{t+1}$ and $f_{t+1}^{\text{approx}}$ for three different values of $\theta_t$. Specifically, the values of $\theta_t$ are set to the 10th, 50th and 90th percentile of the distribution over $\theta_t$ conditional on there being a reform at $t$. In all three cases, the approximation traces the exact shape of $f_{t+1}$ almost perfectly, despite being marginally skewed for the 90th percentile of $\theta_t$. Moreover, because $f_{t+1}^{\text{approx}}$ integrates to unity on $(0, 1)$, the close fit in the interior also implies that $f_{t+1}$ integrates to approximately unity on $(0, 1)$, so that that the residual mass distributed as mass points on $\{0, 1\}$ is negligible.

## D.2 Beliefs after revolts against pooling regimes

Consider next the case where the regime does not attempt any reform and is overthrown by a revolt. From Proposition 3, the posterior density $\tilde{f}_t = \tilde{F}'$ is given by

$$\tilde{f}_t(\vartheta) = \frac{1}{F_t(\bar{\theta}_t)} \frac{\vartheta F_t'(\vartheta)}{M_t^1(\bar{\theta}_t)}.$$

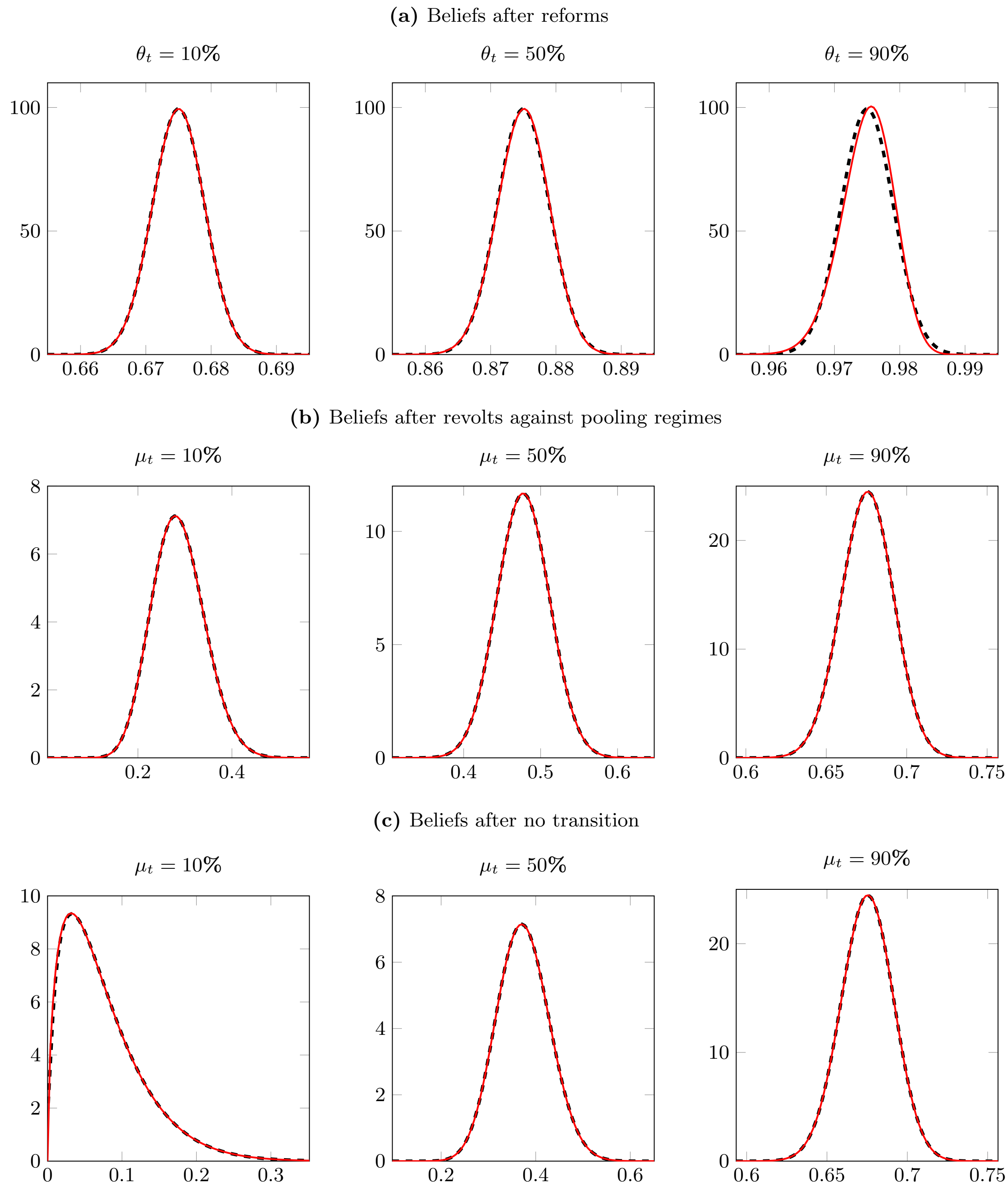


**Figure D.1:** Accurateness of belief approximation. Black dotted lines are exact prior beliefs computed as in (10). Solid red lines approximate the one-step ahead projection using Beta-distributions with their first two moments matching (13) and (14). Top panel compares $f_{t+1}$ with $f_{t+1}^{\text{approx}}$ for $x_t > \lambda_t$, $\eta_t \in \{0, 1\}$ and $\theta_t$ set to the 10th, 50th and 90th percentile of $\mathcal{P}_t(\theta_t | x_t > \lambda_t)$. Middle and bottom panels compare $f_{t+1}$ with $f_{t+1}^{\text{approx}}$ for $x_t = \lambda_t$, $\eta_t = 1$ (middle panel) and $\eta_t = 1$ (bottom panel), $\mu_t$ set to the 10th, 50th and 90th percentile of $\mathcal{P}_t(\mu_t | x_t = \lambda_t, \eta_t)$, and $\lambda_t$ and $\sigma_t$ are set to the conditional (on $\mu_t$) medians.

Substituting $\tilde{f}_t$ into (10), we obtain $f_{t+1}$, which for any interior $\vartheta \in (0,1)$ is given by

$$f_{t+1}(\vartheta) = \frac{1}{F_t(\bar{\theta}_t)} \frac{1}{M_t^1(\bar{\theta}_t)} \int_{-\infty}^{\infty} \phi_{\mu_\epsilon,\sigma_\epsilon^2}(\vartheta - \rho\theta)\theta f_t(\theta)\,\mathrm{d}\theta$$

where $f_t$ is the prior density at $t$. The Beta approximation is given by

$$f_{t+1}^{\text{approx}}(\vartheta) = \beta_{\rho\tilde{\mu}_t+\mu_\epsilon,\rho^2\tilde{\sigma}_t^2+\sigma_\epsilon^2}(\vartheta),$$

with $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ as in (11) and (12).

Panel (b) of Figure D.1 compares $f_{t+1}$ with its approximation $f_{t+1}^{\text{approx}}$ for three different states $\mathcal{S}_t$. Specifically, we set $\mu_t$ to its 10th, 50th and 90th percentile conditional there being no reform and a successful revolt at $t$ ($x_t = \lambda_t$, $\eta_t = 1$). The value of $\sigma_t^2$ is fixed at the associated median (conditional on the corresponding value for $\mu_t$). Again, the approximation closely tracks the exact density $f_{t+1}$ in the interior, and there is no significant mass on $\{0,1\}$.

### D.3 Beliefs after no transition

Finally, consider the case of no transition. From Proposition 3, the posterior density is given by

$$\tilde{f}_t(\vartheta) = \frac{F_t'(\vartheta)}{F_t(\bar{\theta}_t)} \cdot \frac{1-h(s_t)\vartheta}{1-h(s_t)M_t^1(\bar{\theta}_t)},$$

yielding

$$f_{t+1}(\vartheta) = \frac{1}{F_t(\bar{\theta}_t)} \cdot \frac{1}{1-h(s_t)M_t^1(\bar{\theta}_t)} \int_{-\infty}^{\infty} \phi_{\mu_\epsilon,\sigma_\epsilon^2}(\vartheta - \rho\theta) f_t(\theta)(1-h(s_t)\theta)\,\mathrm{d}\theta$$

for any interior $\vartheta \in (0,1)$. The corresponding Beta approximation is given by

$$f_{t+1}^{\text{approx}}(\vartheta) = \beta_{\rho\tilde{\mu}_t+\mu_\epsilon,\rho^2\tilde{\sigma}_t^2+\sigma_\epsilon^2}(\vartheta),$$

with $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ as in (11) and (12).

Panel (c) of Figure D.1 compares $f_{t+1}$ with its approximation $f_{t+1}^{\text{approx}}$ for three different states $\mathcal{S}_t$. Specifically, we set $\mu_t$ to its 10th, 50th and 90th percentile conditional there being no transition at $t$ ($x_t = \lambda_t$, $\eta_t = 0$). The values of $\sigma_t^2$ and $\lambda_t$ (needed to compute $h(s_t)$) is fixed at their associated median (conditional on the corresponding value for $\mu_t$). Again, the approximation closely tracks the exact density $f_{t+1}$ in the interior, and there is no significant mass on $\{0,1\}$.

# E Comparative Statics in the Generation Game

**Comparative statics in $\lambda$** Here we explore how an increase in the regime size $\lambda$ affects the policy mappings depicted in Figures 5 and 6. The primary implication of an increase in $\lambda$ is a reduction in potential supporters of a revolt along the extensive margin. Accordingly, absent reforms, the regime is more stable (seen in the right panel of Figure E.1), which manifests itself in a reduced inclination to implement reforms ($\bar{\theta}$ is higher, see left panel of Figure E.1). A second order implication then is that for increased values for $\bar{\theta}$, the pooling belief $\hat{\theta}^{\text{pool}}$ increases as well (seen in the left panel of Figure E.2), which in turn increases the off-equilibrium support for revolts *conditional* on $x \in (\lambda, \xi(\bar{\theta}))$ as seen in the right panel of Figure E.2.

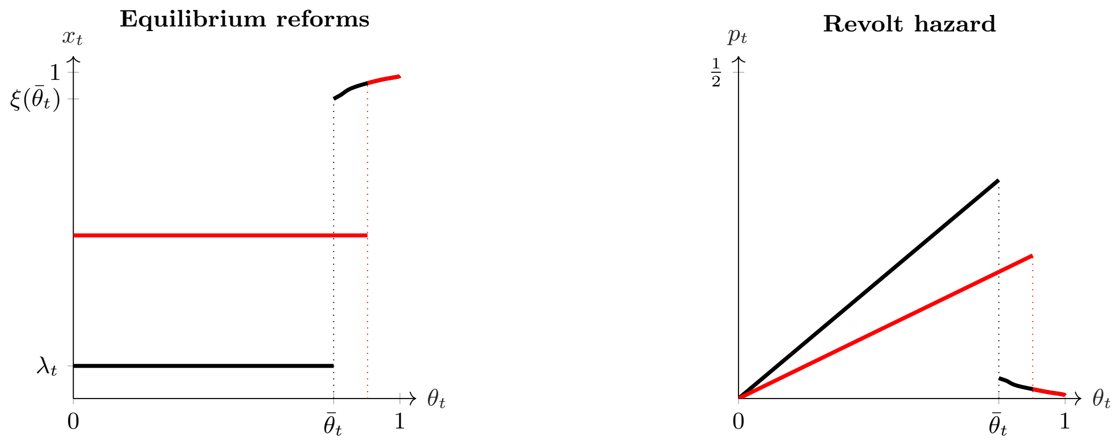


**Figure E.1:** Effect of $\lambda$ on equilibrium reforms and implied probability to be overthrown. Black lines show mappings for $\lambda = .1$, red lines show mappings for $\lambda = .5$.

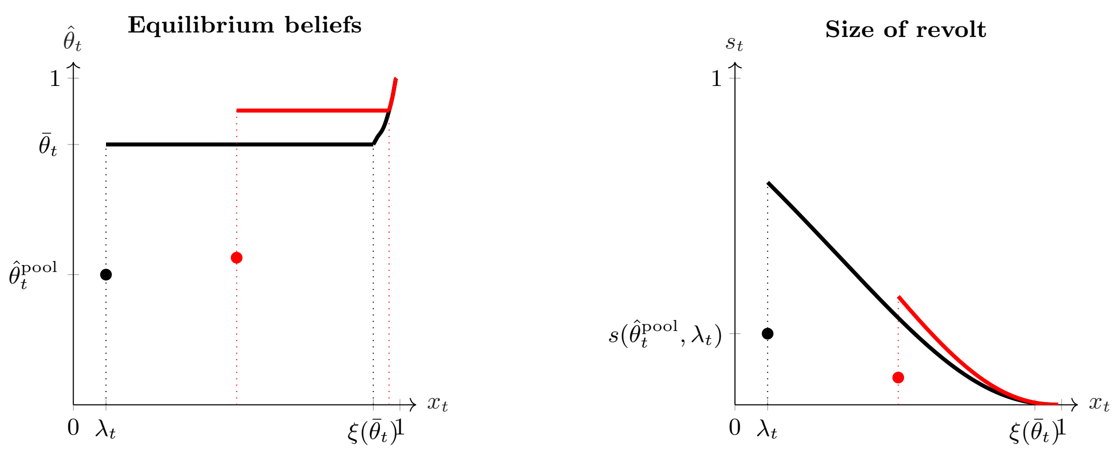


**Figure E.2:** Effect of $\lambda$ on equilibrium beliefs and implied mass of insurgents. Black lines show mappings for $\lambda = .1$, red lines show mappings for $\lambda = .5$.

**Comparative statics in $F$** To demonstrate the effect of outsiders' beliefs on $F$ on the policy mappings, suppose $F$ is parametrized by a Beta distribution with moments $(\mu, \sigma^2)$. Note that the case where $F$ is uniform is a special of the Beta distribution where $\mu = .5$ and $\sigma^2 = 1/12$. We compare this benchmark case, depicted in the main text with the case where $\mu = .35$ and $\sigma^2$ remains fixed at the uniform value of $1/12$. The results are shown in Figures E.3 and E.4. It can be seen that the decline in outsiders' prior expectation (seen in the left panel of Figure E.4) leads again to a drop in revolt hazard (right panels of Figures E.3 and E.4), which makes insiders less inclined to reform ($\bar{\theta}$ is higher, see left panel of Figure E.3).

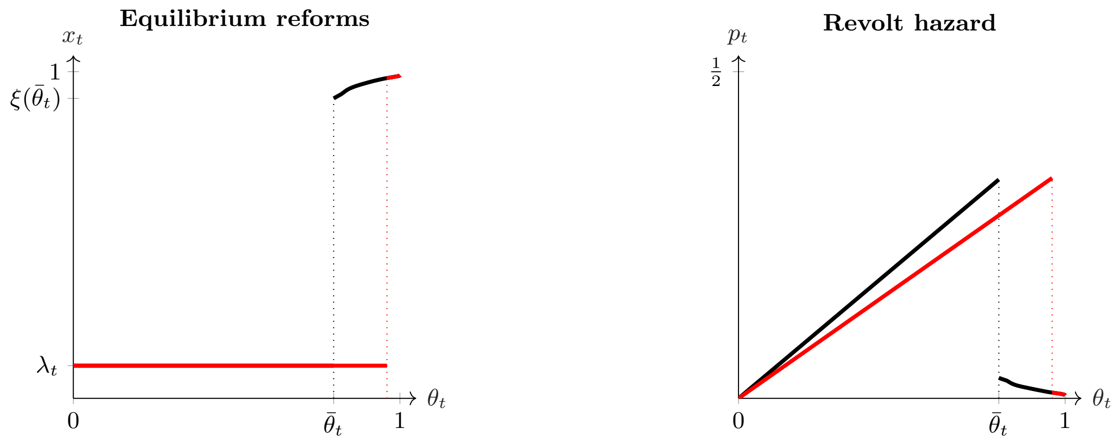


**Figure E.3:** Effect of $\mu$ on equilibrium reforms and implied probability to be overthrown. Black lines show mappings for $\mu = .5$, red lines show mappings for $\mu = .35$.

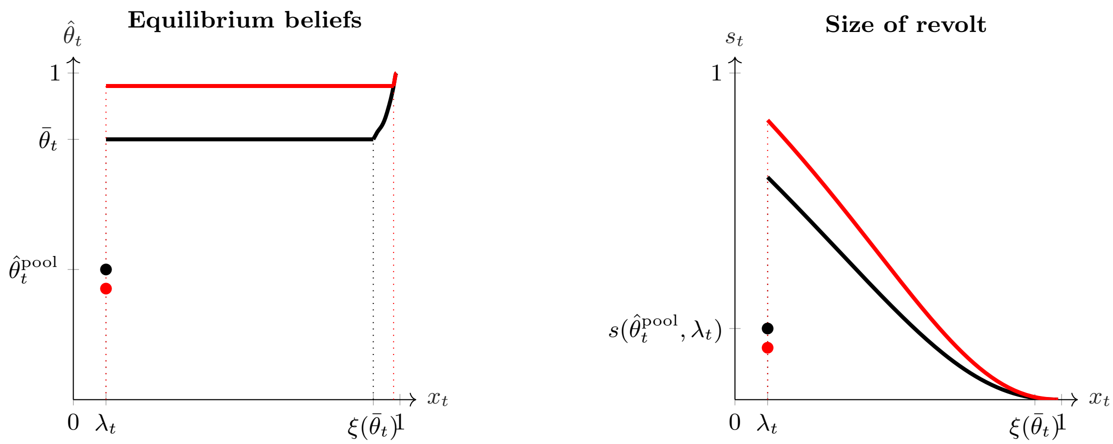


**Figure E.4:** Effect of $\mu$ on equilibrium beliefs and implied mass of insurgents. Black lines show mappings for $\mu = .5$, red lines show mappings for $\mu = .35$.

**Comparison with symmetric information case** Finally, we compare the equilibrium reform mapping with the case where outsiders fully observe $\theta_t$. Full information implies strictly more reforms by insiders compared to the asymmetric information case (see left panel of Figure E.5). This is because asymmetric information essentially imposes an extra cost on reforms associated with revealing that the regime is of a higher type $\theta_t$. On the one hand, this manifests itself in a large pool of regimes not conducting any reform, even though reforms are optimal under full information. On the other hand, since any marginal increase in reforms also implies a marginal change in outsiders' beliefs $\mathrm{d}\hat{\theta}/\mathrm{d}x$, the reform schedule itself (conditionally on conducting reforms) is biased downwards under asymmetric information. As a consequence, revolts tend to be less likely under symmetric information, even though the revolt hazard may point-wise exceed the one under asymmetric information for certain values of $\theta$.[8] Integrating over realizations of $\theta$ (using the uniform prior as probability measure), yields an average revolt hazard of under symmetric information of 5.64 percent as opposed to 13.67 percent under asymmetric information.

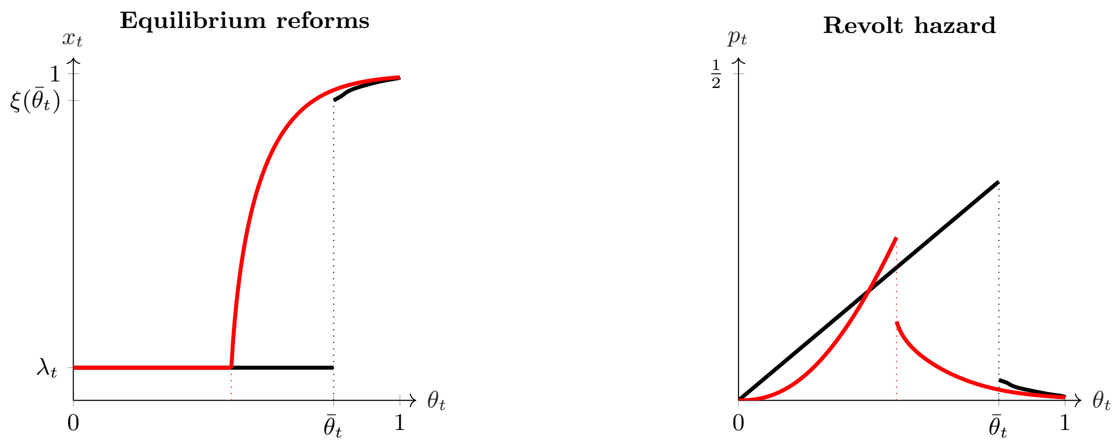


**Figure E.5:** Equilibrium reforms and implied probability to be overthrown under symmetric information. Black lines show equilibrium mappings with asymmetric information, red lines show mappings under full information.

## References


**Geddes, Barbara, Joseph Wright, and Erica Frantz.** 2014. "Autocratic Breakdown and Regime Transitions: A New Data Set." *Perspectives on Politics*, 12(2): 313–331.

**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.

**Harrell, Frank E.** 2001. *Regression Modeling Strategies.* New York, NY:Springer New York.

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

**Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers.** 2017. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2016."


[8]Specifically, the hazard exceeds the one under asymmetric information for $\theta \in (\mathbb{E}_F\{\vartheta|\vartheta \leq \bar{\theta}^{\text{asym}}\}, \bar{\theta}^{\text{sym}}]$.