UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

Working Papers

Col·lecció d'Economia E26/495

DEMOCRATIC BACKSLIDING IN TIMES OF CRISIS

Artyom Jelnov

Maxim Senkov

ISSN 1136-8365

UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

Working Papers

UB Economics *Working Paper No. 495*

Title: Democratic backsliding in times of crisis

Abstract: In a political-agency model, an incumbent can initiate a restrictive policy in response to a crisis state of the world. Both the opposition and the citizen value the incumbent's policy matching the state; however, they are uncertain about the incumbent's true motives. If the incumbent is of the dictatorial type, a restrictive policy that is not protested by both the opposition and the citizen leads to the start of authoritarian rule. We show that when the incumbent is relatively unlikely to be dictatorial, the presence of radical opposition, protesting the restrictive policy regardless circumstances, can reduce voter welfare: it eliminates the efficient state-matching equilibrium, since the opposition never fully reveals dictatorial incumbents. Conversely, when the incumbent is relatively likely to be dictatorial, a high probability of radical opposition can increase voter welfare by deterring the dictatorial type from implementing the restrictive policy.

JEL Codes: D72, D82, D83.

Keywords: democratic backsliding, autocratization, emergency powers, populist radical parties.

Authors:

Artyom Jelnov	Maxim Senkov
Ariel University	Department of Actuarial, Financial and Economic Mathematics. Universitat de Barcelona and BEAT.
Email: artyomj@ariel.ac.il	**Email**: msenkov@ub.edu

UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

Working Papers

Date: March 2026

Acknowledgements: Thanks to Bruce Bueno de Mesquita, Jordi Massó, Oriol Tejada, Matthew Ellman, Nikolas Tsakas, Dmitriy Vorobyev, Ole Jann, Yiman Sun, Peter Bayer, Antoine Zerbini, Konuray Mutluer, and Vladimir Shchukin, seminar participants at Universitat Autonoma de Barcelona and Universitat de Barcelona, as well as the participants of 52nd Annual Meeting of the European Public Choice Society and REES XXII Workshop for helpful comments.

.

1 Introduction

Democratic backsliding, or the erosion of democracy, is a process where a democratic or semi-authoritarian country moves toward authoritarian rule through a deterioration in the quality of democratic institutions. In recent years, there has been a new wave of democratic backsliding globally (Lührmann and Lindberg, 2019). Democratic backsliding involves imposing restrictions on civilian, political, or economic liberties. These limitations may sometimes be justified, especially in a time of emergency.[1] However, there is a concern that a ruler will misuse a state of emergency to expand his or her power (*power as a pretext*, see Stenberg et al., 2022. See also Hertzler, 1940).

A recent example of a crisis is the COVID-19 pandemic, during which severe restrictions were imposed on the population for public health purposes. Most of them were lifted after the emergency period ended; however, in many cases, pandemic-related restrictions on freedom of assembly and movement were widely misused to enhance the incumbent's power, suppress the opposition, and were maintained longer than expected (Lewkowicz et al., 2022). Pandemic-related emergency legislation restricting public gatherings in Morocco and Egypt has helped local regimes curb political dissent (Badran and Turnbull, 2023). In Hungary, during the COVID-19-related state of emergency, Viktor Orbán granted himself the power to rule by decree, preventing parliamentary oversight and imposing jail penalties on journalists for "disseminating misleading information."[2]

Not surprisingly, these broad and often misused pandemic-related restrictions frequently caused public outrage and centralized protest, sometimes resulting in the overturning of the restrictions.[3] While the political opposition often accepts restrictions when it trusts the incumbent and agrees that the threat is serious, it is not uncommon for populist radicals to protest against the restrictions regardless of the circumstances. A recent example comes from Austria, where the Freedom Party (FPÖ) organized anti-lockdown demonstrations at the end of 2021, with protest signs comparing the measures to a "fascist dictatorship" and party leader Herbert Kickl using terms like "totalitarism" and "health communism" in parliamentary speeches opposing the mandate.[4][5]

[1] See Bjørnskov and Voigt (2022) for an empirical analysis of the relationship between constitution and state of emergency.

[2] China exploited the crisis to expand its control over Hong Kong (Stenberg et al., 2022). In Malaysia, the primer minister Muhyiddin Yassin, whose majority in the parliament was not safe, delayed meetings of the parliament, using pandemic as an excuse (Tew, 2020, Talib, 2023).

[3] A health crisis is only one example of an emergency, along with military and financial crises or natural disasters. Following a terrorist attack in Beslan, Russia, in 2004, Russian President Vladimir Putin consolidated his power, in particular by replacing the direct election of district heads with a system in which candidates are proposed by the president.

[4] Schocher, S. 2021. "Violent protests in Vienna against new vaccination requirement". *Die Welt*, November 20. https://www.welt.de/politik/ausland/article235182294/Corona-in-Oesterreich-Heftige-Proteste-in-Wien-gegen-neue-Impfpflicht.html

[5] Radical oppositional protests are not limited to right-wing populists: in the Czech Republic, both left- and right-wing populists staged joint demonstrations throughout 2023 against financial and military support for Ukraine despite the military conflict unfolding nearby.

In this paper, we ask: How does the interplay between the incumbent's and the opposition's uncertain motives affect citizens' welfare during crises? We show that the answer is not straightforward. When the incumbent is relatively unlikely to have dictatorial motives, the possibility of opposition being radical reduces citizens' welfare by undermining trust between citizens and the opposition. As a result, citizens refrain from joining opposition-led protests against restrictive measures, which opens the door for the dictatorial incumbent to capture the power. By contrast, when the incumbent is more likely motivated by power capture, a highly probable radical opposition benefits citizens: in this case, citizens are willing to join protests initiated by radicals, thereby preventing the dictatorial incumbent from imposing restrictive policies.

We consider a model, where the state of the world may be "risky" (emergency situations such as a pandemic, natural disaster, international or internal conflict, or economic crisis) or "safe" (a regular state). An incumbent politician (he) observes the state of the world and chooses between obtaining executive empowerment and limiting the civil freedoms ("restrictive policy"), or maintaining the status quo ("liberal policy"). The incumbent ruler may be either of "benevolent" or "dictatorial" type. If he is dictatorial, adopting the restrictive policy triggers a transition to an autocratic regime. The incumbent's type is his private knowledge.

Another player is the "opposition" (she), who is supposed to preserve the social interest from possible violations by the ruler. In different countries, various institutions may play this role, e.g., a legislature, elites, or a constitutional court. Thus, the opposition can oppose the restrictive policy, with some chance of success.[6] Importantly, opposing is a costly action for the opposition. However, the opposition can be of "radical" type, which opposes the restrictive policy under any circumstances. The opposition's type is her private knowledge.

The third actor is a representative citizen. If the opposition calls for a protest, he can join it. The opposition succeeds in overruling the incumbent's policy if and only if the citizen joins the protest.

Beyond the opposition having a private type, there is further information asymmetry between the opposition and the citizen as the opposition knows the state of the world, while the citizen does not. In reality, even if the existence of a crisis is commonly known, information on whether regular policy tools are sufficient and whether a restrictive policy is required may not be available to all. If the opposition succeeds in overruling the incumbent's restrictive policy, the incumbent pays a cost for it.[7]

The players' preferences are structured as follows. Both the citizen and the benevolent opposition face a *tradeoff between flexibly adjusting policy to the state of the world and pre-*

[6] Anti-lockdown protests during COVID-19 were often positioned as preemptive measures to prevent sliding into autocracy. For instance, protesters in US states like Michigan in 2020 compared the imposition of lockdowns to dictatorship-like measures (Censky, A. 2020. "Heavily Armed Protesters Gather Again At Michigan Capitol To Decry Stay-At-Home Order". *National Public Radio (NPR)*, May 14. https://www.npr.org/2020/05/14/855918852/heavily-armed-protesters-gather-again-at-michigans-capitol-denouncing-home-order).

[7] For example, in 2024 President Yoon Suk Yeol of South Korea was impeached after an unsuccessful attempt to impose martial law.

venting potential autocracy. Similarly, the benevolent incumbent gets utility from matching the state. The dictatorial incumbent derives utility primarily from capturing power by implementing the restrictive policy.

We find that there exist only three Perfect Bayesian equilibria in pure strategies: *pooling liberal* (Proposition 1), in which only the liberal policy is chosen in every state of the world; *partially separating* (Proposition 2), in which the dictatorial type of the incumbent chooses the restrictive policy in every state of the world, while the benevolent type chooses the policy that matches the state, and *pooling efficient* (Proposition 3), in which both types of the incumbent choose the same policy, matching the state of the world.[8] Proposition 4 fully characterizes the welfare of the representative citizen under these equilibria.

The central result of the paper, Theorem 1, establishes the relationship between radical opposition and citizens' welfare. When the prior probability that the incumbent is a dictator is relatively low, both the citizen and the benevolent opposition, who bears a protest cost, care primarily about matching the state and are therefore unwilling to protest. In this case, and when the opposition is non-radical with certainty, a welfare-dominant efficient pooling equilibrium exists for all such priors. The logic is straightforward: if the benevolent opposition initiates a protest, this action fully reveals that the incumbent is a dictator, which motivates the citizen to join. On the equilibrium path, then, the dictatorial incumbent is disciplined, and his policy choices match the state of the world. By contrast, if there is a positive probability of radical opposition, any protest observed on the equilibrium path is initiated by the radical opposition. Consequently, the citizen refuses to join. Knowing this, the dictatorial incumbent is effectively unconstrained and no longer matches the state in the safe state of the world. To sum up, as we show in the first part of Theorem 1, the absence of radical opposition supports the welfare-dominant efficient pooling equilibrium, providing the citizen a strictly higher payoff than any equilibrium that exists when radical opposition is present (with positive probability).

Conversely, when the prior probability that the incumbent is a dictator is relatively high, citizen cares more about preventing autocracy and therefore joins any protest, even those initiated by the radical opposition. The benevolent opposition, by contrast, does not initiate the protest on-path as protest is costly to her. As protests led by radical opposition are sufficient to block restrictive policy, when the opposition is sufficiently likely to be radical, the dictatorial incumbent chooses not to initiate the restrictive policy in the first place. Hence, the only equilibrium that remains is the pooling liberal equilibrium, which is welfare-dominant in this case. Intuitively, in this equilibrium the dictator never has the opportunity to establish autocracy, thereby maximizing citizens' welfare. To summarize, as shown in the second part of Theorem 1, sufficiently high polarization between the incumbent and the opposition (as captured by the probability of radical opposition) eliminates the welfare-detrimental equilibrium and thus weakly improves citizens' welfare.

Corollary 1 presents another surprising result. When the probability of a dictatorial incumbent is sufficiently low and the probability of radical opposition is positive, citizens are

[8] We restrict attention to pure-strategy equilibria to keep the analysis simple and tractable.

better off under the separating equilibrium – in which the dictator adopts restrictive policies even in the safe state – than under the efficient pooling equilibrium, where radical opposition protests restrictive policies in the risky state. Effectively, when the dictatorship is sufficiently unlikely, leaving the door for dictatorship open rather than allowing for false alarms from the radical opposition leads to more efficient adjustment of policy to the state of the world, which benefits the citizen.

Finally, Section 4 uses our model to explain the end of major national, politically independent television in Russia in 2001 and Ukraine's experience during the Euromaidan of 2013–2014.

1.1 Literature

There is a burgeoning economic-theoretical literature on democratic backsliding. See Grillo et al. (2023) for a survey.

Gratton and Lee (2024) consider a model in which voters have incomplete information about the security shock and might choose an illiberal incumbent who can provide them with better insurance against the shock at the cost of deteriorating accountability, leading to illiberal democracy. Similar to their paper, our paper centers on imperfect information and the liberty–security tradeoff, but looks at a different accountability channel. While Gratton and Lee (2024) emphasize how illiberal rule can endogenously weaken electoral accountability through information control, we emphasize the crisis-related signaling and coordination problem: citizens mobilize (or not) based on what opposition behavior implies about the incumbent's motives. This yields a sharp comparative static in how the polarization between the citizens and the opposition affects the citizens' welfare. Thus, our main contribution is to incorporate the opposition as an information-driven source of accountability.

Other closely related papers include Miller (2021) and Chiopris et al. (2025). In Miller (2021), the actors are the incumbent, the opposition, and the citizens, like in our model. The incumbent can violate a law to grab more power. The opposition knows whether there has been a violation, but the citizens do not. Then, the opposition can call on citizens to protest. In our setting incumbent's action is observed by the citizen. Moreover, we consider asymmetric information about the incumbent and the opposition preferences.

Similar to our model, in Chiopris et al. (2025), the incumbent also has a private type. In their setting, the "closet autocrat" type can choose an autocratic policy, which is disliked by voters. The incumbent first chooses an institutional reform, which may allow autocracy. This reform is approved or not by citizens, and only the incumbent then chooses a policy. In our model policy choice may lead to autocratic rule in the future.

Our paper relates to recent developments in the theory of selectorate (Bueno de Mesquita and Smith, 2023, 2024). Similar to our paper, Bueno de Mesquita and Smith (2024) study how anticipation affects the incumbent's policy choice, which drives politics in either a liberal or autocratic direction. However, there are a number of differences in the focus and scope of the research. First, we focus on the study of policy choice in the specific context of crises or emergency situations, and anticipation concerns either the presence or absence of

a crisis situation tomorrow. Moreover, Bueno de Mesquita and Smith (2024) explore the adjustment of political and economic institutions in anticipation of a potential revolt or a coup. Second, information asymmetry between politicians and the public serves as one of the key components of our story. Bueno de Mesquita and Smith (2024) assume that the preferences of the politicians are known to the public and that politicians care only about political survival.

Svolik (2018) suggests a model in which a politician's preferences are known. The politician can manipulate voting (although the manipulation process itself is not modeled). In his setting, informed voters react to policy, whereas uninformed voters react to the fact of manipulation.

In Hollyer et al. (2023), the incumbent chooses a two-dimensional policy, where one of the dimensions is a 'level of autocracy.' In our model, autocracy is a 'side effect' of the policy chosen. Luo and Przeworski (2021) consider a dynamic model where in each period a new challenger is born, which may be more or less appealing than an incumbent. The incumbent decides on his effort in undermining democracy.

Fox and Stephenson (2011) is also related. In their model a leader, which can be competent or not, decides about policy, which is preferred by a voter only in a special state of the world. The policy can be vetoed by a court. After that, voter decides to support the leader or not, depending on belief about leader's competence. In our model, the ruler is competent, but may became a dictator.

In a dynamic model by Howell et al. (2023), the executive branch of government gradually accumulates more power when the judicial system constrains it. The judicial system, in its turn, is limited by precedents made in previous periods. In Fox and Polborn (2021) legislative and executive branches check each other, when two competing compete on control on them.

We analyze a decline in democracy, which stands in contrast to the opposing process of democratization. The literature concerning democratization is very rich: see Bueno de Mesquita et al. (2005), Acemoglu and Robinson (2006) and Rosendorff (2001). For the political economy of non-democratic countries, see Egorov and Sonin (2024).

2 Model

We consider a game between an incumbent ruler (I, he), an opposition (O, she) and a representative citizen (C, he). Both incumbent and opposition are of uncertain types.

The state of the world θ is either risky (r) or safe (s), with $\rho \equiv \mathrm{P}(r)$. The state of the world is known to I and to O, but not to C. The incumbent's type τ_I is either dictatorial d or benevolent $\bar{d}$, with $\mu \equiv \mathrm{P}(d)$. This type is the private knowledge of I. The incumbent chooses a policy π, that can be either liberal L or restrictive R. If the restrictive policy is chosen, and the incumbent is dictatorial, then the rule becomes autocratic in future.

The opposition is one of two types τ_O: benevolent ($\tau_O = \bar{e}$) or radical ($\tau_O = e$). The type is the private knowledge of O, while $\lambda \equiv \mathrm{P}(e)$. Following the liberal policy, the opposition does nothing. Following the restrictive policy, the benevolent opposition chooses an action b,

where $b = 1$ means to oppose the policy and to call to protest and $b = 0$ means not to oppose. The radical opposition always opposes the restrictive policy and chooses $b = 1$. Opposing is a costly action for O, with $c > 0$ as the cost.

If $b = 1$ is chosen, the citizen chooses to join the protest ($j = 1$) or not to join ($j = 0$). Protest is costless to C. If he joins, opposing the restrictive policy is successful, and in this case, the liberal policy is imposed instead of the restrictive one. If C does not join, the opposition fails and the restrictive policy remains.

Payoffs. For the benevolent type of the incumbent, the utility is

$$u_I^{\overline{d}} = w_I(\theta, \pi) - b \cdot j \cdot f.$$

First, it includes the state-matching component $w_I(\theta, \pi)$, which depends on the state of the world and the implemented policy. We assume $w_I(s, L) > w_I(s, R)$ and $w_I(r, R) > w_I(r, L)$, i.e. in the safe state of the world the liberal policy is preferable and in the risky state the restrictive one is superior. Second, it includes the loss from having the restrictive policy overruled: if the policy chosen by I (of each type) is successfully opposed, a positive f is subtracted from I's payoff. This penalty f reflects image or other political damage due to policy change.

The dictatorial type of the incumbent differs from the benevolent type as he enjoys autocratic rule:

$$u_I^d = w_I(\theta, \pi) - b \cdot j \cdot f + \underbrace{1(\pi = R) \cdot [1 - b \cdot j] \cdot h}_{\text{benefit from power capture}}.$$

That is, if he chooses the policy R, and this was not opposed by O through $b = 1$ and by C through $j = 1$, then he gets $w_I(\theta, R) + h$, where the increment h captures the gain from autocratic rule. Moreover, we assume $w_I(s, L) < w_I(s, R) + h$, so the dictatorial incumbent prefers R in any state of the world.

Benevolent opposition's utility is

$$u_O^{\overline{e}} = w_O(\theta, \pi) - b \cdot c \underbrace{-1(\pi = R) \cdot 1(\tau_I = d)\,[1 - b \cdot j] \cdot g}_{\text{loss from autocracy}}.$$

First, there is the state-matching component, where we assume $w_O(s, L) > w_O(s, R)$ and $w_O(r, R) > w_O(r, L)$. Second, $b \cdot c$ captures the cost of opposing the restrictive policy. Third, g is subtracted, if the rule becomes autocratic. We assume $w_O(r, R) - g < w_O(r, L) - c$. Thus, O is willing to bear costs of protest and suffer the state mismatch if she knows with certainty that the incumbent is the dictatorial type.

The opposition of type e prefers the liberal policy L. To simplify exposition, we assume this is the case in any state of the world. Thus, the radical opposition automatically opposes the restrictive policy.

The citizen's payoff is given by

$$u_C = w_C(\theta, \pi) \underbrace{-1(\pi = R) \cdot 1(\tau_I = d)\,[1 - b \cdot j] \cdot g}_{\text{loss from autocracy}}.$$

We assume $w_C(s, L) > w_C(s, R)$ and $w_C(r, R) > w_C(r, L)$. Further, if the rule becomes autocratic in the future, it yields a disutility $-g$ to the citizen, so the total utility of C in this case is $w_C(r, R) - g$. Let $w_C(r, R) - g < w_C(r, L)$, so the citizen is willing to prevent autocracy even at the cost of having the liberal policy when the state of the world is risky.

Timing. State θ, incumbent type τ_I and opposition type τ_O are drawn. The incumbent and the opposition observe θ. I chooses policy π; if $\pi = R$, then the opposition chooses b. If $b = 1$, then C chooses his action j. If $j = 1$, then the opposition succeeds and policy becomes L; otherwise, the policy remains R. If $b = 0$, the policy remains R. If policy R is not overruled after O's and C's actions, and the politician is d, then the incumbent becomes an autocratic ruler. The start of autocratic rule is observable to all players. Payoffs are then realized. The timing of the game is summarized in Figure 1.

The solution concept we use is the pure-strategy Perfect Bayesian Equilibrium ("equilibrium" in the rest of the text).

- State $\theta \in \{s, r\}$ is realized. I and O observe it
- I's private type $\tau_I \in \{d, \bar{d}\}$ and O's private type $\tau_O \in \{e, \bar{e}\}$ are drawn
- I chooses between liberal and restrictive policy, $\pi \in \{L, R\}$
- O observes I's choice and then chooses either to protest ($b = 1$) or not ($b = 0$)
- C observes I's and O's choices and then chooses either to join the protest ($j = 1$) or not ($j = 0$)
- Policy $\pi \in \{L, R\}$ is implemented and the political regime is determined
- Payoffs are realized

Figure 1: Timing of the game

3 Analysis

Observe that the benevolent incumbent always chooses L if $\theta = s$. Moreover, if $\bar{d}$ chooses L in both states of the world, then d does so as well. Otherwise, if I chooses R, then both O and C know that he is a dictator and successfully oppose him. Further, it cannot be the case that d chooses R only in s. To see this, note that if he chooses a restrictive policy in the safe state, he does not anticipate effective opposition. Observing R in s, O knows for sure that I is dictatorial. The only possible case in which O chooses not to oppose is when C does not join. However, since C does not distinguish between states of the world, he will also choose not to join the protest in the risky state as well. Thus, it is optimal for d to also choose the restrictive policy in the risky state too.

Therefore, only three pure-strategy equilibria are possible: *pooling liberal* (both types of I choose L regardless of the state of the world, "liberal equilibrium" hereafter); *partially*

separating (d chooses R regardless of the state of the world, $\bar{d}$ chooses R in r and L in s, "separating equilibrium" hereafter) and *pooling efficient* (both types of I choose R in r and L in s, "efficient equilibrium" hereafter).

Proposition 1 establishes the existence of the liberal equilibrium.

Proposition 1 (Liberal equilibrium)**.** *There is a Perfect Bayesian Equilibrium in which both types of I choose L regardless of the state of the world. On the off-equilibrium path, following R both types of O choose $b = 1$, and C chooses $j = 1$.*

To illustrate the case of the *liberal equilibrium*, consider instances where anti-epidemic measures were not taken owing to concerns about human rights violations. Witt (2020) provides an account of such cases from American history.[9] For instance, in the 1890s, health authorities in Milwaukee were barred from isolating persons infected with smallpox without their consent.

Another example of a liberal outcome is Russia in 1917, which faced political turmoil, a financial crisis, and military failure. However, the Provisional Government, which led the country, did not support an attempt at a military coup by General Kornilov, whose intention was to stabilize the situation. Among the reasons for the lack of support was that supporting the military coup would undermine the democratic and progressive ideals the Provisional Government was attempting to uphold, and thus jeopardize its power. Shortly afterward, however, the Bolsheviks took power.

Next, we show that the separating equilibrium exists if the prior that the incumbent is a dictator is low.

Proposition 2 (Separating equilibrium)**.** *There is $\hat{\mu}$ such that for $\mu < \hat{\mu}$ and $\lambda > 0$ there is a Perfect Bayesian Equilibrium, where d chooses R regardless of the state of the world, $\bar{d}$ chooses R in r and L in s. The benevolent type of O chooses $b = 0$ after any action by I. Following $b = 1$, chosen by e, C chooses $j = 0$.*

As an illustration of a *separating equilibrium*, consider the case of Franklin D. Roosevelt's presidency in the United States. He was elected four times, breaking with tradition, during the crises of the Great Depression and the Second World War. It cannot be ruled out that, had he lived, he might have sought a fifth or even sixth term, effectively consolidating power. His death before the end of the war leaves us with asymmetric information: was he a benevolent leader who believed that extraordinary circumstances required extraordinary measures, or was he ultimately striving to expand his personal authority?

In Proposition 3 we find conditions for the existence of the efficient equilibrium.

Proposition 3 (Efficient equilibrium)**.** *If $\lambda = 0$ and $\mu < \max\{\tilde{\mu} \equiv \frac{w_C(r,R)-w_C(r,L)}{g}, \tilde{\tilde{\mu}} \equiv \frac{w_O(r,R)-w_O(r,L)+c}{g}\}$ or $0 < \lambda < \tilde{\lambda} \equiv \frac{w_I(r,R)-w_I(r,L)}{w_I(r,R)-w_I(r,L)+f}$ and $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$, then there is a Perfect Bayesian Equilibrium, where both types of I choose R in r and L in s. The benevolent type of O chooses $b = 1$ only on the off-equilibrium path, where R is chosen in s. Otherwise, she*

[9] On the other hand, Witt (2020) also shows how, on different occasions, quarantine measures were discriminatory against immigrants, African Americans, or other minorities.

chooses $b = 0$. *Following* $b = 1$ *chosen by* e, C *chooses* $j = 1$. *For other values of* λ *and* μ *equilibrium of this kind does not exist.*

The threshold $\tilde{\mu} = \frac{w_C(r,R) - w_C(r,L)}{g}$ is a ratio between value the citizen assigns to the correct action in the risky state and magnitude of disutility from the autocratic rule. $\tilde{\tilde{\mu}} = \frac{w_O(r,R) - w_O(r,L) + c}{g}$ is the similar threshold for the moderate opposition.

To understand Proposition 3, note that when either $\mu > \tilde{\tilde{\mu}}$ or $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$ and $\lambda > \tilde{\lambda}$, following R, the incumbent is so likely to face a successful protest, that he avoids the restrictive policy. Further, for $\mu < \tilde{\mu}$ and $\lambda > 0$, based on the prior regarding the incumbent's type, the citizen never joins the protest, and as the radical opposition initiates the protest even in risky state, there is no way to fully reveal the dictatorial incumbent to the citizen. Thus, the dictatorial incumbent is not disciplined, and deviates to R. However, if $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$ and $0 < \lambda < \tilde{\lambda}$, both the former and the latter is absent: it is sufficiently unlikely that R in risky state will lead to a successful protest, and R in safe state reveals the dictator and leads to a successful protest. Thus, the efficient equilibrium exists. It is worth noting that the crucial factor at play here is that for $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$, there is a misalignment between the moderate opposition and the citizen: while the citizen wants to have the restrictive policy overruled, the opposition does not want to protest as the protest is too costly for the opposition. We summarize this logic in Figure 2.

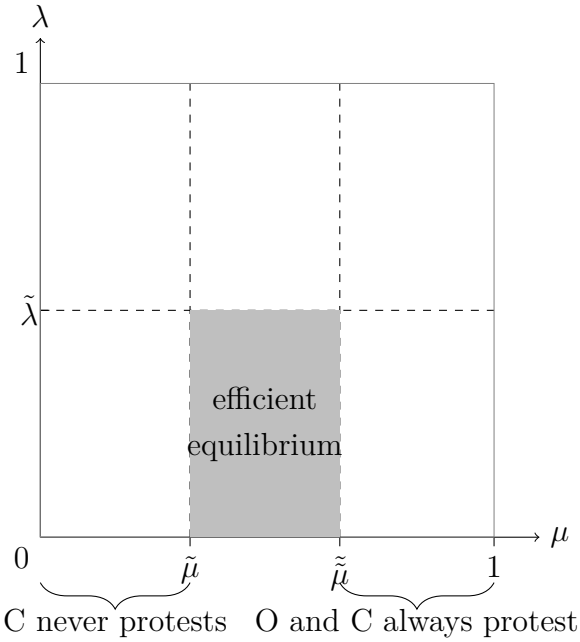


Figure 2: Existence of the efficient equilibrium when $\lambda > 0$ and $\tilde{\mu} < \tilde{\tilde{\mu}}$.

Interestingly, when the opposition is non-radical with certainty, $\lambda = 0$, the efficient equilibrium exists for any $\mu < \max\{\tilde{\mu}, \tilde{\tilde{\mu}}\}$. This is because if $\mu < \max\{\tilde{\mu}, \tilde{\tilde{\mu}}\}$, the moderate opposition only protests off-path when R is chosen in the safe state, which fully reveals the dictator to the citizen. However, if $\mu > \max\{\tilde{\mu}, \tilde{\tilde{\mu}}\}$, then the moderate opposition initiates the protest against the restrictive policy even in the risky state, and she is joined by the citizen. Thus the incumbent chooses to avoid the restrictive policy.

To illustrate the *efficient equilibrium*, consider the misuse of anti-COVID restrictions in several semi-democratic countries. In Hungary, for instance, Viktor Orbán used the COVID-19 state of emergency to grant himself the power to rule by decree, expanding his political opportunities and further accelerating the erosion of democracy.

We proceed with a normative exercise. We consider the citizen's expected payoff at the strategy profiles corresponding to the liberal, efficient, and separating equilibrium, and study how it changes as we change the probability of dictatorial incumbent μ and the probability of radical opposition λ. This way, we discover what would be socially desirable if each equilibrium could be achieved. We refer here to citizen's expected payoff as the citizen's welfare.[10]

Proposition 4 (Citizen's welfare)**.** *There exist such $\bar{\mu}$ and $\tilde{\mu}$, $\bar{\mu} > 0$ if $\lambda > 0$, and $\bar{\mu} = 0$, if $\lambda = 0$, that the highest citizen's welfare is achieved by*

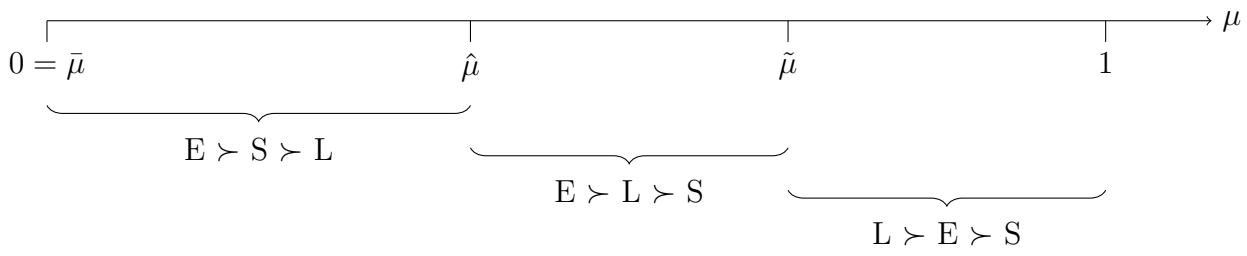
1. *the separating equilibrium strategy profile, if $\mu < \bar{\mu}$,*
2. *the efficient equilibrium strategy profile, if $\bar{\mu} < \mu < \tilde{\mu}$,*
3. *the liberal equilibrium strategy profile, if $\mu > \tilde{\mu}$.*

Proposition 4 is visualized in Figure 3. To better understand Proposition 4, recall that the citizen faces a tradeoff between matching the state and preventing autocracy. Thus, for very high μ, the incumbent is very likely to be dictatorial, so the liberal equilibrium in which R is never implemented yields the citizen the highest welfare. Consider now very low μ and assume $\lambda > 0$. The separating equilibrium strategy profile guarantees that R is never opposed, and that L is chosen in state l if the incumbent is benevolent. For very low μ, the incumbent is very likely to be benevolent, and so the outcome under the separating equilibrium strategy profile is very close to the state-matching outcome. Hence, for very low μ, the separating equilibrium strategy profile provides the citizen the highest welfare. Finally, consider the intermediate values of μ. Under the efficient equilibrium strategy profile, the outcome is not as close to state-matching as in the case described above, because with probability λ, R in state r is overruled through successful protest. Hence, when radicality of opposition is possible, $\lambda > 0$, the outcomes under the efficient equilibrium could provide a "balance" between preventing autocracy under the liberal equilibrium and state-matching under the partial separation: under the efficient equilibrium strategy profile, the dictatorial type is disciplined not to play R in s, which prevents autocracy in those contingencies; yet, R is still often played in state r, which allows more policy adjustment to the state of the world.

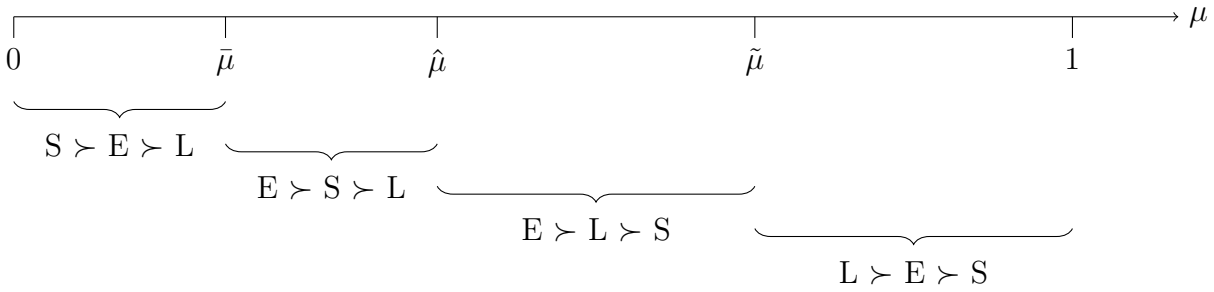
We continue with a Corollary of Proposition 4.

Corollary 1. *Assume $\mu < \bar{\mu}$ and $\lambda > 0$. Equilibrium in which the dictatorial type of incumbent initiates restrictive policy in the safe state (separating equilibrium) welfare-dominates the state-matching equilibrium (efficient equilibrium).*

[10]Note that Proposition 4 can be right away extended to show the actual highest equilibrium welfare in the game using Figure 4 by removing the non-existent equilibria for the particular values of μ and λ.



(a) $\lambda = 0$



(b) $\lambda > 0$

Figure 3: Welfare ranking of equilibria. L is liberal, E is efficient equilibrium, S is separating equilibrium

In this counterintuitive case, the equilibrium in which the dictatorial type is not disciplined (and always chooses the restrictive policy) is better for the citizen than the one in which the dictatorial type mimics the benevolent type of the incumbent. This is because in the case of efficient equilibrium, restrictive policy is overruled wherever the incumbent faces the radical opposition, and not overruled otherwise. Thus, the policy mismatches the state with probability λ. Meanwhile, in the case of the separating equilibrium, the restrictive policy is never overruled on-path, and the mismatch occurs whenever the incumbent is dictatorial, i.e. with probability μ. When $\mu < \bar{\mu}(\lambda)$, the citizen cares relatively more regarding the state-matching rather than preventing the autocracy. Further, from the state-matching perspective, the probability of restrictive policy in safe state is sufficiently low. As a result, when dictatorship is very unlikely, it is better from the citizen's perspective to leave the "door to dictatorship" wide open than to permit false anti-incumbent alarms from the radical opposition that undermine restrictive policies.

We now proceed to visualize the existence of equilibria. Given the existence conditions established in Propositions 1, 2, and 3 we have a complete picture of when each of the three equilibria exists, depending on the prior beliefs λ and μ. The existence conditions are summarized in Figure 4.

An important remark regarding Figure 4 is that it depicts existence of equilibria for the parametric case $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$. In the case $\tilde{\tilde{\mu}} < \tilde{\mu}$, efficient equilibrium exists only when $\lambda = 0$ and $\mu < \tilde{\mu}$.

We now turn to the central results of the paper.

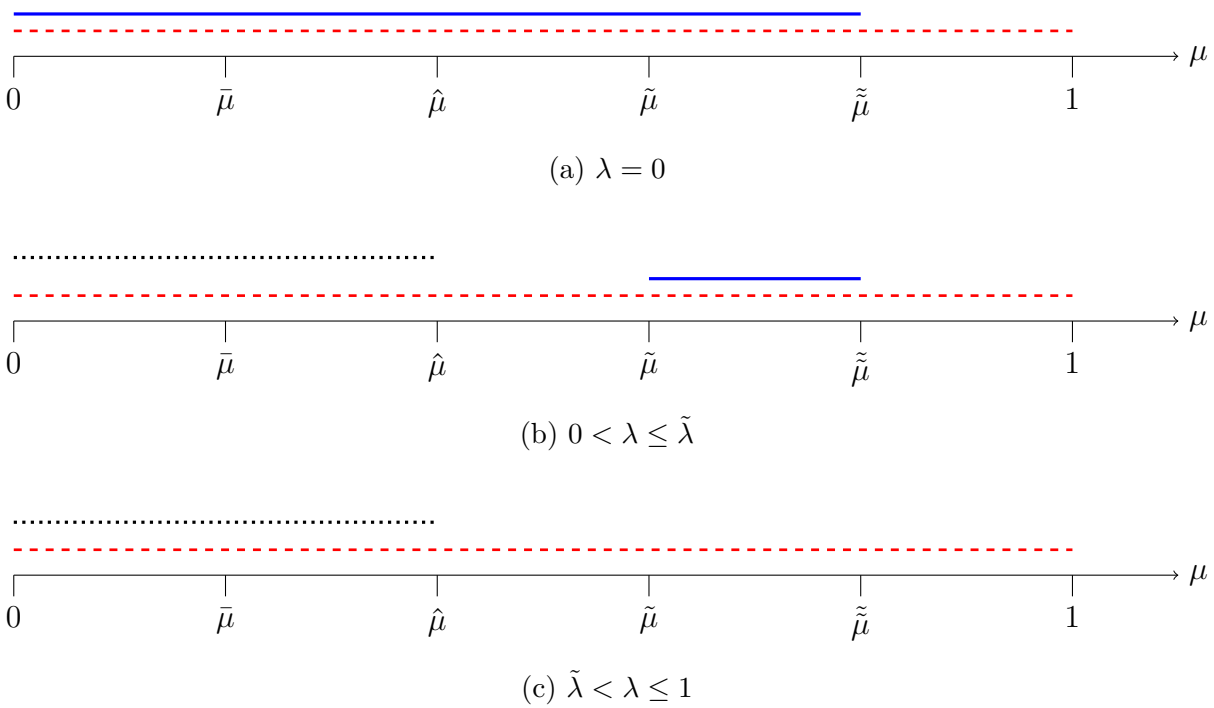


Figure 4: Existence of the pure-strategy Perfect Bayesian equilibria for different values of opposition type prior λ and incumbent type prior μ

Theorem 1 (Radical opposition and citizen's welfare)**.** *1. Assume the incumbent is unlikely to be of the dictatorial type, $0 < \mu < \tilde{\mu}$. Then, in the absence of radical opposition ($\lambda = 0$), there exists an equilibrium in which the citizen's payoff is strictly higher than under any equilibrium with radical opposition ($\lambda > 0$).*

2. Assume the incumbent is likely to be of the dictatorial type, $\mu > \tilde{\mu}$. Then, making the radical opposition sufficiently likely ($\lambda > \tilde{\lambda}$) weakly increases the citizen's payoff as this makes the welfare dominant equilibrium the unique equilibrium.

We start with the intuition behind the first result of the Theorem. Given $\mu < \tilde{\mu}$, the citizen's payoff under the efficient equilibrium is decreasing in λ, and it does not depend on λ under the separating and liberal equilibrium. Finally, as we know from Propositions 3 and 4, when λ gets as low as 0, the welfare-dominant efficient equilibrium starts to exist. Thus, presence of the radical opposition $\lambda > 0$ can impede existence of the welfare-superior efficient equilibrium. If a positive λ is interpreted as an indicator of a political polarization, we can conclude that political polarization may lead to an inefficiency when the incumbent is unlikely to be dictatorial.

The first result of the Theorem does not claim that decreasing λ to $\lambda = 0$ will induce improvement in the citizen's welfare. This is because by construction the model has multiple equilibria, and for the parameters we consider we can not exclude the welfare-dominated

liberal equilibrium, which also exists under $\lambda = 0$. This can be seen from Figure 4.

We proceed with the intuition behind the second result of the Theorem. Given $\mu > \tilde{\mu}$, the separating equilibrium does not exist, and the citizen's payoff under the liberal equilibrium is higher than under the efficient for any λ. Two cases are possible. If $\tilde{\tilde{\mu}} < \tilde{\mu}$, then efficient equilibrium does not exist and the unique equilibrium is the welfare-dominant liberal equilibrium. If $\tilde{\mu} < \tilde{\tilde{\mu}}$, then efficient equilibrium exists only for $0 < \lambda \leq \tilde{\lambda}$. Thus, increasing λ above $\tilde{\lambda}$ removes the efficient equilibrium, leaving the welfare-dominant liberal equilibrium as the unique one.

As there are multiple equilibria, the citizen's welfare increase due to higher λ needs not be strict. Yet, it is a strict increase if $\tilde{\mu} < \tilde{\tilde{\mu}}$ and the switch to liberal equilibrium occurs from the dominated efficient equilibrium.

4 Case studies

In this section, we illustrate the model with two examples in which the policy choice motives of both the incumbent and the opposition were uncertain, and we study the implications of the key factors for the policy choice and the subsequent protests.

4.1 NTV channel state takeover in Russia, 2001

After the collapse of the Soviet Union, NTV was the primary oppositional TV channel offering explicit critique of the policies of Boris Yeltsin's, and later Vladimir Putin's, government. In 2000, it emerged that NTV had outstanding debt to the state-owned Gazprom corporation. A subsequent court decision in 2001 granted Gazprom an additional block of shares and effectively control over the board and top management. This was possible thanks to active state enforcement actions, which were likely guided by the Kremlin (Becker, 2004).

It was the first major political conflict of Vladimir Putin's presidency. At that point, Putin did not yet have a history of consolidating power or a reputation for power-mongering. For the young post-Soviet Russian society, the dispute regarding the NTV channel raised a tradeoff between sticking to the rule of law and ensuring that an important watchdog media outlet survives. Further, there was uncertainty regarding the state of the world, i.e., which of the two was of higher social value. Putin reacted to the Gazprom–NTV conflict by denying any political relation to the court decision and focusing on the necessity of establishing the rule of law, stating that it was an economic dispute and that the state must not and would not interfere.[11] In terms of our model, all this can be summarized by a relatively low prior, μ, that Putin has dictatorial intentions.

The Gazprom-facilitated change of the board and top management in early April 2001 was followed by a series of "Save NTV" protests led by the liberal opposition. The protests were headed by liberal democratic parties such as "Yabloko" and "SPS", and were further supported by journalists and public figures such as Yevgeny Kiselyov from NTV and Alexei

[11] Newsline, 2001. *Radio Free Europe/Radio Liberty*, April 5. https://www.rferl.org/a/1142380.html

Venediktov from "Echo of Moscow". Notably, the opinion of broader popular masses toward the liberal-democratic opposition was quite negative by that time: they blamed liberal democracy for the economic slump of the 1990s and associated the liberal opposition with Russian oligarchs (Miller, 2024). Thus, they perceived the opposition as extremely biased in favor of the old NTV owners, represented by the media tycoon Vladimir Gusinsky. In the language of our model, this points to a positive prior regarding the opposition being "radical", i.e., $\lambda > 0$.

The outcome of the standoff between Gazprom and the old NTV owners and management favored the former. The protests initiated by the liberal opposition managed to attract only 5–10 thousand people, who were predominantly from the educated middle class. Further, the protests were largely confined to large cities such as Moscow and Saint Petersburg, with no substantial protest activity in the provinces. This was followed by the forceful physical takeover of NTV's newsroom by the new security staff on 13–14 April 2001, and the exodus of a substantial part of the old staff. As is clear today, the change in NTV's team effectively ended the modern history of politically independent major national television in Russia and facilitated further erosion of democracy, which was clearly welfare-detrimental for Russian civil society (Becker, 2004).

According to the results of our model, this was to be expected. Theorem 1, part 1 establishes that the combination of a low prior regarding the incumbent being dictatorial ($\mu < \tilde{\mu}$) and a positive probability of the opposition being radical ($\lambda > 0$) destroys the welfare-dominant efficient equilibrium. Relating this to the case of the NTV takeover, the dictatorial incumbent was able to implement the takeover of the main oppositional TV channel ("restrictive policy") when saving the oppositional TV channel was socially preferable ("safe state"). This happened because broader popular masses saw Putin as a non-dictatorial type and did not take the opposition's calls to protest at face value, believing that the opposition's interests were likely misaligned with theirs.

4.2 Euromaidan in Ukraine, 2013–14

In 2013, the Ukrainian government headed by Viktor Yanukovych was preparing to sign the EU Association Agreement, which would facilitate tighter economic and political connections. Ukrainian society faced a tradeoff between moving closer to Russia or to the EU, and there was uncertainty regarding the better course. In November 2013, the Yanukovych government suddenly suspended preparations to sign the EU deal, motivating this decision by threats to national security.

Viktor Yanukovych's Party of Regions had a clear reputation as a pro-Russia party. Further, by that time Russia had established a reputation as a major force meddling in fair elections and suppressing democracy in neighboring countries. Yanukovych's government also received criticism over concerns about judicial independence and the rule of law, including the Tymoshenko case (Trochev, 2018). Thus, it is plausible to state that the prior μ regarding Yanukovych's type being dictatorial was substantially high.

The suspension of the EU Agreement in November 2013 immediately caused grassroots

protests, with thousands joining. After the incident of police violence on 30 November, opposition parties (Batkivshchyna, UDAR, Svoboda) set up coordination structures and called people out. The opposition parties had a fierce stake against Yanukovych's government. Batkivshchyna had ending the "Yanukovych anti-people regime" as one of its main goals. UDAR's Klitschko was an opposition challenger competing directly for power with Yanukovych. Svoboda was a nationalist party with a consistent and explicit anti-Yanukovych and anti-Russian stance. Hence, in the language of our model, the prior λ regarding the opposition having a radical type was clearly high.

The opposition parties successfully used the impulse created by police violence to scale up the protests. On 1 December 2013, the crowd in Kyiv already included hundreds of thousands. Eventually, the protests further spread and hardened in January–February, leading to Yanukovych leaving Kyiv on 22 February 2014. From today's perspective, electoral competition and turnovers improved after 2014 (Fedorenko et al., 2016); thus, it is plausible that the change of government moved Ukraine toward more competitive politics and a stronger civil society, and benefited citizens' welfare.

Again, the Yanukovych's loss and eventual exile could be expected given our model. The high prior regarding the incumbent being dictatorial ($\mu > \tilde{\mu}$) ensured that, given any protest initiated by the opposition, citizens were willing to join. As a result, the protest facilitated by the opposition, which was very likely to be radical, was still joined by the popular masses and successfully reached its goals. Notably, Yanukovych was likely not playing an equilibrium strategy, as it would have been optimal for him to reverse the suspension of the EU deal earlier, which could have potentially saved him politically.

Why did Yanukovych choose an off-equilibrium strategy? One possible explanation is that he was constrained by an external power—namely, Russia (Svoboda, 2019). For Russian President Vladimir Putin, the agreement between Ukraine and the EU was unacceptable, and Russia possessed leverage that enabled it to pressure Yanukovych to withdraw from the agreement.

5 Discussion

We assume that if under pressure from the opposition the incumbent withdraws the restrictive policy, he pays a cost. In reality, such a flip-flop could shorten the ruler's term, at least in the long run, and he will be replaced then by the opposition. Thus, the "radical" opposition, which automatically opposes a restrictive policy, may be driven not only by ideology, but also by the office seeking motivation and a desire to replace the incumbent in office.

In our model the citizens consider matching of the policy to the state of the world and possibility of authoritarian rule only. However, additional dimensions of decision may exist. In highly polarized societies there are citizens who always oppose a ruler, or always support him. In terms of our model, if the citizen does not oppose the ruler, then the partial separating equilibrium exists. If he always opposes, the outcome depends on λ. If λ is high (again, this fits polarized society), only the liberal equilibrium is possible.

Appendix

Proof of Proposition 1. Assume that, off the equilibrium path, after observing R the benevolent opposition and the citizen assign a sufficiently high probability to the incumbent being of the dictatorial type. Then $b = 1$ and $j = 1$ are their best replies. □

Proof of Proposition 2. After observing $b = 1$, the citizen concludes that the opposition is radical. C believes that I is dictator with probability $\frac{\mu}{\mu+(1-\mu)\rho}$. He also believes that the state of the world is risky with probability $\frac{\rho}{\rho+(1-\rho)\mu}$. Then he still chooses not to join the protest if

$$\frac{\rho w_C(r,L)}{\rho+(1-\rho)\mu}+\frac{(1-\rho)\mu w_C(s,L)}{\rho+(1-\rho)\mu}<\frac{\rho w_C(r,R)}{\rho+(1-\rho)\mu}+\frac{(1-\rho)\mu w_C(s,R)}{\rho+(1-\rho)\mu}-\frac{\mu g}{\mu+(1-\mu)\rho} \quad (1)$$

By $w_C(r,L) < w_C(r,R)$, there is $\hat{\mu}$ such that for $\mu < \hat{\mu}$ (1) holds. We can explicitly solve for this $\hat{\mu}$

$$\hat{\mu}=\frac{\left(w_C(r,R)-w_C(r,L)\right)\rho}{\left(w_C(s,L)-w_C(s,R)\right)(1-\rho)+g}. \quad (2)$$

Finally, observe that given $\lambda = 0$, the separating equilibrium does not exist under reasonable assumptions on off-path beliefs. If $\lambda = 0$, then $b = 1$ is never played on the path. Hence, after observing R in s off-path, O has a profitable deviation to playing $b = 1$, which should induce a reasonable off-path belief $\mu = 1$ for C, and thus induce $j = 1$. □

Proof of Proposition 3. **Case 1.** $\lambda > 0$.

Consider "efficient" candidate equilibrium strategies: I, both types: R in r, L in s; O type $\bar{e}$: $b = 0$ after (R, r) (and $b = 1$ off-path (R, s)); O type e: (automatically) $b = 1$ after any (π, θ).

Let's derive citizen's best-replies given candidate equilibrium strategies of I and O. Note the I's on-path strategy fully reveals the state of the world θ.

Hereafter, for any prior belief ω on some parameter, we denote by $\omega'(\pi, b)$ a posterior belief of C about this parameter, after observing actions π and b.

Recall that by definition C joins protest only after $b = 1$. Let $b = 1, \pi = R$. In this case there is only one on-path possibility: the opposition is radical ($\lambda'(R, 1) = 1$). Since radical opposition acts automatically, her action is uninformative, and the citizen's inference is $\rho'(R, 1) = 1$, $\mu'(R, 1) = \mu$.

After inference:

$$\begin{aligned} U(j=0) &= w_C(r,R)-\mu g,\\ U(j=1) &= w_C(r,L), \end{aligned}$$

where $U(j)$ is C's utility after action j.

Then cutoff for choice of $j = 0$ is:

$$\begin{aligned} & w_C(r,R)-\mu g > w_C(r,L)\\ \iff & \mu<\tilde{\mu}\equiv\frac{w_C(r,R)-w_C(r,L)}{g}. \end{aligned} \quad (3)$$

Opposition of type $\bar{e}$ does not deviate to $b = 1$ in the state r, when expects $j = 1$ by C, if

$$w_O(r, L) - c < w_O(r, R) - \mu g \Leftrightarrow \mu < \frac{w_O(r, R) - w_O(r, L) + c}{g} \equiv \tilde{\tilde{\mu}}. \tag{4}$$

Case 1.1 $\tilde{\tilde{\mu}} < \tilde{\mu}$.

For $\mu < \tilde{\mu}$, C does not join any protest, so I of type d deviates to R in state s. For $\mu > \tilde{\mu}$, O of type $\bar{e}$ best-replies to R in state r with $b = 1$, which is followed by $j = 1$ from C. Given this, I of any type never chooses R. Thus, efficient equilibrium does not exist.

Case 1.2 $\tilde{\mu} < \tilde{\tilde{\mu}}$.

Assume that $\mu < \tilde{\mu}$, then $j = 0$ is the best reply of the citizen to $b = 1$. Note that the dictator has a profitable deviation from the efficient equilibrium strategy: he deviates to R in state s. Thus, the considered strategy profile is not an equilibrium.

Assume

$$\mu > \tilde{\tilde{\mu}},$$

O of both types choose $b = 1$ after R even in r. Given this, for any λ, any type of I deviates to L in any state, so efficient equilibrium does not exist.

Assume now $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$. Then after $b = 1$ (which is chosen in equilibrium only by $\bar{e}$) the citizen always chooses $j = 1$. Incumbent's decision to deviate from R to L in state r depends on the value of λ.

R is chosen in r by the benevolent incumbent iff

$$\begin{aligned} &\lambda\left(w_I(r, L) - f\right) + (1 - \lambda) w_I(r, R) > w_I(r, L) \\ &\iff \lambda < \tilde{\lambda}_{\bar{d}} \equiv \frac{w_I(r, R) - w_I(r, L)}{w_I(r, R) - w_I(r, L) + f}, \end{aligned}$$

and for dictatorial type the condition is

$$\begin{aligned} &\lambda\left(w_I(r, L) - f\right) + (1 - \lambda)(w_I(r, R) + h) > w_I(r, L) \\ &\iff \lambda < \tilde{\lambda}_{d} \equiv \frac{w_I(r, R) - w_I(r, L) + h}{w_I(r, R) - w_I(r, L) + f + h}, \end{aligned}$$

Since $h > 0$, $\tilde{\lambda}_{\bar{d}} < \tilde{\lambda}_{d}$. Thus, efficient equilibrium exists for $\frac{w_C(r,R) - w_C(r,L)}{g} < \mu < \frac{w_O(r,R) - w_O(r,L) + c}{g}$ iff

$$\lambda < \tilde{\lambda}_{\bar{d}}.$$

The dictatorial type does not deviate to R in $\theta = s$, such otherwise both types of O successfully oppose.

Case 2. $\lambda = 0$. *Case 2.1* $\tilde{\tilde{\mu}} < \tilde{\mu}$. For $\mu > \tilde{\mu}$, R in state r is optimally followed by $b = 1$ and $j = 1$. Thus, I has a profitable deviation to L.

Note for any $\mu < \tilde{\tilde{\mu}}$, O of type $\bar{e}$ trivially chooses $b = 0$ after R in state r. Further, she chooses $b = 0$ after R in state r for any $\tilde{\tilde{\mu}} < \mu < \tilde{\mu}$ because C chooses $j = 0$. Thus, I does not have any profitable deviations in state r. In state s, deviating to R fully reveals that O has type d, and is thus followed by $b = 1$ and $j = 1$. Hence, I does not have any profitable deviations in state s either.

To conclude, efficient equilibrium exists for $\mu < \tilde{\mu}$.

Case 2.2 $\tilde{\mu} < \tilde{\tilde{\mu}}$.

For $\mu > \tilde{\tilde{\mu}}$, R in state r is optimally followed by $b = 1$ and $j = 1$. Thus, I has a profitable deviation to L.

For $\mu < \tilde{\tilde{\mu}}$, O of type $\bar{e}$ trivially chooses $b = 0$ after R in state r. In state s, deviating to R fully reveals that O has type d, and is thus followed by $b = 1$ and $j = 1$. To conclude, efficient equilibrium exists for $\mu < \tilde{\tilde{\mu}}$. □

Proof of Proposition 4. The expected payoff of C in the liberal equilibrium is

$$\rho w_C(r, L) + (1 - \rho) w_C(s, L). \tag{5}$$

In the efficient equilibrium, his expected payoff is

$$\rho(1 - \lambda)[w_C(r, R) - \mu g] + \rho\lambda w_C(r, L) + (1 - \rho) w_C(s, L). \tag{6}$$

Finally, in the separating equilibrium, the expected payoff of C is

$$\rho[w_C(r, R) - \mu g] + (1 - \rho)[\mu(w_C(s, R) - g) + (1 - \mu) w_C(s, L)] \tag{7}$$

By (5) and (6), the expected payoff in the liberal equilibrium is higher than in the efficient one, if $\frac{w_C(r,R) - w_C(r,L)}{g} < \mu$.

By (5) and (7), the expected payoff in the liberal equilibrium is lower than in the separating one, if

$$\mu < \hat{\mu} \equiv \frac{\rho[w_C(r, R) - w_C(r, L)]}{(1 - \rho)[w_C(s, L) - w_C(s, R)] + g}. \tag{8}$$

Finally, comparing (6) and (7), we get the condition on partial separation to dominate the efficient equilibrium

$$\mu < \bar{\mu} \equiv \frac{\lambda\rho[w_C(r, R) - w_C(r, L)]}{(1 - \rho)[w_C(s, L) - w_C(s, R)] + g\,(1 - (1 - \lambda)\rho)}.$$

Comparing the cutoffs, one observes that

$$sign\left(\frac{\rho[w_C(r, R) - w_C(r, L)]}{(1 - \rho)[w_C(s, L) - w_C(s, R)] + g} - \frac{\lambda\rho[w_C(r, R) - w_C(r, L)]}{(1 - \rho)[w_C(s, L) - w_C(s, R)] + g\,(1 - (1 - \lambda)\rho)}\right)$$
$$= sign\,((1 - \lambda)(1 - \rho)(g + w_C(s, L) - w_C(s, R))) > 0,$$

so

$$\hat{\mu} > \bar{\mu}.$$

Thus, we can completely characterize the welfare ranking for any μ, see Figure 3. □

Proof of Corollary 1. Proposition 4 established that if $\mu < \bar{\mu}$, then out of all equilibria, separating equilibrium yields the highest welfare. □

Proof of Theorem 1. Proof of the first statement. The citizen's payoffs under the liberal equilibrium (5) and the separating equilibrium (7) do not depend on λ, while the payoff under the efficient equilibrium decreases in λ if and only if $\mu < \tilde{\mu}$.

Given this, consider two parametric cases. First, assume $\bar{\mu} < \mu < \tilde{\mu}$. In this case, Proposition 4 implies that the efficient equilibrium dominates both the separating and the liberal equilibria for any λ. Furthermore, Proposition 3 shows that for such μ, the efficient equilibrium exists only when $\lambda = 0$.

Second, consider $0 < \mu < \bar{\mu}$. In this case, Propositions 3 and 4 imply that the efficient equilibrium dominates both the separating and the liberal equilibria if and only if it exists, i.e., when $\lambda = 0$. To see this, note that $\bar{\mu}(\lambda = 0) = 0$, so given $\lambda = 0$ and $\mu > 0$, it follows that $\mu > \bar{\mu}$.

Thus, when $\lambda = 0$, the efficient equilibrium exists and gives the citizen a strictly higher payoff than any equilibrium under $\lambda > 0$.

Proof of the second statement. The citizen's payoff under the liberal equilibrium (5) does not depend on λ. Further, for $\mu > \tilde{\mu}$ the citizen's payoff under the efficient equilibrium (6) increases in λ and achieves maximum for $\lambda = 1$, when it is equal to the payoff under the liberal equilibrium. Hence, for $\mu > \tilde{\mu}$ and $\lambda < 1$, the citizen's payoff under the efficient equilibrium is strictly lower than under the liberal equilibrium.

From Proposition 3, two cases are possible. If $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$ then the efficient equilibrium ceases to exist for $\lambda \geq \tilde{\lambda}$. If $\tilde{\tilde{\mu}} < \tilde{\mu} < \mu$ then the efficient equilibrium does not exist for any λ. Further, from Proposition 2, and by $\hat{\mu} < \tilde{\mu}$, separating equilibrium does not exist for $\mu > \tilde{\mu}$. Thus, the only pure strategy equilibrium for $\lambda \geq \tilde{\lambda}$ is the liberal equilibrium. Hence, increasing λ from some $\lambda_1 < \tilde{\lambda}$ to some $\lambda_2 \geq \tilde{\lambda}$ weakly increases the citizen's payoff, with a strict increase if $\tilde{\mu} < \mu < \tilde{\tilde{\mu}}$ and equilibrium under λ_1 was the efficient one.

□

References

Acemoglu, D. and J. A. Robinson (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Badran, S. and B. Turnbull (2023). The Covid-19 pandemic and authoritarian consolidation in North Africa. In *Rights at Stake and the COVID-19 Pandemic*, pp. 147–166. Routledge.

Becker, J. (2004). Lessons from russia: A neo-authoritarian media system. *European journal of communication 19*(2), 139–163.

Bjørnskov, C. and S. Voigt (2022). Emergencies: on the misuse of government powers. *Public Choice 190*(1), 1–32.

Bueno de Mesquita, B. and A. Smith (2023). Institutional change as a response to unrealized threats: An empirical analysis. *Journal of Conflict Resolution 67*(5), 1032–1062.

Bueno de Mesquita, B. and A. Smith (2024). *Anticipating Instability*. Unpublished manuscript.

Bueno de Mesquita, B., A. Smith, R. M. Siverson, and J. D. Morrow (2005). *The logic of political survival*. MIT press.

Chiopris, C., M. Nalepa, and G. Vanberg (2025). A wolf in sheep's clothing: Citizen uncertainty and democratic backsliding. *The Journal of Politics 87*(4), 000–000.

Egorov, G. and K. Sonin (2024). The political economics of non-democracy. *Journal of Economic Literature 62*(2), 594–636.

Fedorenko, K., O. Rybiy, and A. Umland (2016). The ukrainian party system before and after the 2013–2014 euromaidan. *Europe-Asia Studies 68*(4), 609–630.

Fox, J. and M. Polborn (2021). On the separation of executive and legislative powers: Executive independence, liberty, and social welfare. *Journal of Theoretical Politics 33*(4), 430–454.

Fox, J. and M. C. Stephenson (2011). Judicial review as a response to political posturing. *American Political Science Review 105*(2), 397–414.

Gratton, G. and B. E. Lee (2024). Liberty, security, and accountability: The rise and fall of illiberal democracies. *Review of Economic Studies 91*(1), 340–371.

Grillo, E., Z. Luo, M. Nalepa, and C. Prato (2023). Theories of democratic backsliding. *Annual Review of Political Science 27*.

Hertzler, J. O. (1940). Crises and dictatorships. *American Sociological Review 5*(2), 157–169.

Hollyer, J. R., M. Klašnja, and R. Titiunik (2023). Charismatic leaders and democratic backsliding.

Howell, W. G., K. A. Shepsle, and S. Wolton (2023). Executive absolutism: the dynamics of authority acquisition in a system of separated powers. *Quarterly Journal of Political Science 18*(2), 243–275.

Lewkowicz, J., M. Woźniak, and M. Wrzesiński (2022). Covid-19 and erosion of democracy. *Economic Modelling 106*, 105682.

Lührmann, A. and S. I. Lindberg (2019). A third wave of autocratization is here: what is new about it? *Democratization 26*(7), 1095–1113.

Luo, Z. and A. Przeworski (2021). Democracy and its vulnerabilities: Dynamics of democratic backsliding. *Available at SSRN 3469373*.

Miller, C. (2024, 05). Why russia's democracy broke. In *When Democracy Breaks: Studies in Democratic Erosion and Collapse, from Ancient Athens to the Present Day*. Oxford University Press.

Miller, M. K. (2021). A republic, if you can keep it: Breakdown and erosion in modern democracies. *The Journal of Politics 83*(1), 198–213.

Rosendorff, B. P. (2001). Choosing democracy. *Economics & Politics 13*(1), 1–29.

Stenberg, M., P. Rocco, and S. A. Farole (2022). Calling in "sick": Covid-19, opportunism, pretext, and subnational autocratization. *Global Studies Quarterly 2*(3), 1–11.

Svoboda, K. (2019). On the road to maidan: Russia's economic statecraft towards ukraine in 2013. *Europe-Asia Studies 71*(10), 1685–1704.

Svolik, M. (2018). When polarization trumps civic virtue: Partisan conflict and the subversion of democracy by incumbents. *Available at SSRN 3243470*.

Talib, N. S. (2023). Democratic backsliding in Malaysia: Executive aggrandizement under Muhyiddin Yassin's government. In *Democratic Recession, Autocratization, and Democratic Backlash in Southeast Asia*, pp. 117–138. Springer.

Tew, Y. (2020). Constitutionalism in the time of corona. *International Journal of Constitutional Law Blog*.

Trochev, A. (2018). Patronal politics, judicial networks and collective judicial autonomy in post-soviet ukraine. *International Political Science Review 39*(5), 662–678.

Witt, J. F. (2020). *American contagions: Epidemics and the law from smallpox to COVID-19*. Yale University Press.