MEDIA CENSORSHIP UNDER THE THREAT OF A PROTEST

A Master's Thesis

by
KORALP KABAL

Department of
Economics
İhsan Doğramacı Bilkent University
Ankara
July 2025

To the loving memory of my grandparents, Ali and Mihriban Öztürk

MEDIA CENSORSHIP UNDER THE THREAT OF A PROTEST

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

KORALP KABAL

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN ECONOMICS

THE DEPARTMENT OF
ECONOMICS
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

July 2025

MEDIA CENSORSHIP UNDER THE THREAT OF A PROTEST
By Koralp Kabal

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

_____

A. Arda Gitmez
Advisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

_____

Nuh Aygün Dalkıran
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

_____

Emre Ekinci
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

_____

Refet S. Gürkaynak
Director

ABSTRACT

MEDIA CENSORSHIP UNDER THE THREAT OF A PROTEST

Kabal, Koralp

M.A., Department of Economics

Supervisor: Assist. Prof. Dr. A. Arda Gitmez

July 2025

In this thesis, we study, both theoretically and empirically, how polarization affects media freedom when the ruler faces a protest threat. The protest is modeled as a global game, and the ruler censors the media to thwart it, which is costly for the ruler. We show that under this setting, an increase in polarization leads to an increase in media censorship. Using two different measures of polarization and protest data, we provide suggestive empirical evidence that validates the model's main result.

Keywords: Media censorship, protest, polarization, global games, signaling.

ÖZET

PROTESTO TEHDİDİ ALTINDA MEDYA SANSÜRÜ

Kabal, Koralp
Yüksek Lisans, İktisat Bölümü
Tez Danışmanı: Dr. Ögr. Üyesi A. Arda Gitmez

Temmuz 2025

Bu tezde, teorik ve ampirik olarak, protesto tehdidi altında kutuplaşmanın medya özgürlüğünü nasıl etkilediği incelenmektedir. Protesto global oyun olarak modellenmekte ve protestonun engellenmesi için yönetici, kendi açısından maliyeti olan medya sansürüne başvurmaktadır. Bu ortamda, kutuplaşmadaki yükselişin medya sansürünü artırdığını gösteriyoruz. İki farklı kutuplaşma ölçütü ve protesto verisi kullanarak modelin ana sonucunu doğrulayan yön gösterici ampirik kanıtlar sunuyoruz.

Anahtar Kelimeler: Medya sansürü, protesto, kutuplaşma, global oyunlar, sinyalleme.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Arda Gitmez, for his support and guidance in the last two years. Beginning from the time when I was not yet officially a Bilkent student, his line of work has inspired me, and he has left an indelible mark on me as an aspiring political economist. I will always be deeply grateful to him.

I have been fortunate to encounter excellent mentors. My Bilkent journey started with Nuh Aygün Dalkıran, and I have felt his sincerity and warmth ever since I first met him. From my days at Boğaziçi, Murat Güray Kırdar has had a profound impact on my view of what economics is and can be, and he has been consistently supportive. Thanks to Selcen Çakır, I have gained my first research experience, from which I learned invaluable lessons. Alper Yağcı and Başak Taraktaş influenced my approach to political science, and though indirectly, contributed to my agenda of "studying politics using economics". I also owe special thanks to Emre Ekinci for providing valuable feedback on the thesis.

I could not have achieved my goals without the support and love of my family. My father's intellectual curiosity is a true source of motivation. My mother has influenced me to push my limits, ultimately allowing me to pursue economics. Ada is not just a sister, but also a friend I know I can depend on for the rest of my life.

Last but not least, I am truly indebted to my friends. My undergraduate years can be characterized by before and after meeting with Yavuz Selim Kaçmaz. Although we can rarely see each other, I always feel his support miles away. I hope that one day we will be one of the inseparable duos like Gentzkow and Shapiro. Throughout days and nights, Murat Genar Kaplan had been my constant companion in the one and only academic sanctuary of Aptullah Kuran Library. My conversations with Yağmur Barış Berkkan have always enlightened me for more than a decade.

I acknowledge the financial support from the Scientific and Technological Research Council of Turkey (TÜBİTAK), Science Fellowships and Grant Programs Department (BİDEB) 2211 – Domestic Graduate Scholarship Program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Political leaders regularly face resistance from citizens towards their policies. This resistance can be either violent or non-violent, with the latter showing an increasing trend over the past two decades [1]. Examples where citizens engage in protest [2] are abound in weak democracies and autocracies [3](see Appendix H), as well as in democracies such as the US recently. One of the most prominent obstacles to successful protest is solving the coordination problem (Egorov & Sonin, 2024). Accordingly, the attributes of the citizenry are highly consequential in this context, one of which is polarization.

On the other hand, controlling the media is among the most utilized tools of contemporary leaders (Guriev & Treisman, 2019). Media landscape matters, since it can influence political participation (Leeson, 2008), and it also has an impact on political and economic freedoms (Djankov et al., 2003). Then, two questions are in order at this point: 1. Under the threat of a protest, how

[1] See Appendix I.

[2] From this point on, we sometimes use only the word *protest* to refer to non-violent resistance against the regime.

[3] Following Cantoni et al. (2024) and Marshall et al. (2016), weak democracies and autocracies are the countries that have a Polity score below 7.

do leaders influence the media? 2. How much does this influence depend on polarization?

In this study, we provide a game-theoretical model that examines the effect of polarization on media freedom in the face of a protest threat. Combining a global game with costly signaling, our main result suggests that an increase in polarization leads to an increase in media censorship. This is explained using the concept of office insecurity (Rozenas, 2016). A protest threat is one of the factors that increases the ruler's office insecurity. In such a setting, most of the citizenry has a high incentive to protest, and the polarization further amplifies the presence of protest participants. This decreases the ruler's survival probability and leads to more media censorship.

The protest threat is modeled using the machinery of global games (e.g., Carlsson & van Damme, 1993; Morris & Shin, 1998). Each citizen receives a cost signal from which she makes an inference about the cost of protest participation for others. We make a comparative statics analysis using the variation in the correlation among the participation costs of citizens. We conceptualize the decrease in correlation among the participation costs as a measure of polarization.

We model the ruler's media censorship as a costly signaling game. In particular, following the seminal work of Besley and Prat (2006), we interpret media censorship as a cost incurred to hide the bad news (in our case, the fact that the ruler is weak). Costly signaling has also been adopted in a recent paper from the media capture literature (Guriev & Treisman, 2020). After characterizing the equilibrium conditions for the signaling part of the model, we also present our results using monotone comparative statics (Milgrom & Shannon, 1994).

We provide suggestive empirical evidence for our model's main result. Using data from V-Dem (Coppedge et al., 2024) and the Nonviolent and Violent Conflict Outcomes (NAVCO) (Chenoweth & Shay, 2020) [4] we construct a panel data set of up to 179 countries for years 2000-2018. We show that an increase in polarization is associated with a decrease in media freedom during protest years. The measure related to the concept of ideological polarization (e.g., Abramowitz & Saunders, 2008) gives statistically significant results, and this is robust to regime index choice as a control. However, when we use the measure related to affective polarization (Iyengar et al., 2012, 2019), the statistical significance of our results is not robust to regime index choice.

Related Literature. Our paper contributes to three strands of literature. First, our paper belongs to the literature on media capture (e.g., Besley & Prat, 2006; Gehlbach & Sonin, 2014; Guriev & Treisman, 2019, 2020; Petrova, 2008). In particular, our paper is closely related to Edmond (2013), which also introduces signaling in a global game setting. However, in Edmond's model, the ruler's information manipulation involves influencing the signals (related to regime strength) that individuals receive. In our model, citizens receive signals that give the participation cost of protest, and the ruler's information manipulation does not affect these cost signals.

Next, our paper is linked to the protest literature in economics (see Cantoni et al. (2024) for a survey). This literature increasingly focuses on the role of new information technologies in facilitating coordination during protests (e.g., Enikolopov et al., 2020; Fergusson & Molina, 2020; Little, 2016). Our model presents a mechanism in which the ruler can affect the individuals' protest participation incentives without primarily altering their coordination capabilities.

[4] These two are our main data sets. For control variables, we use Fariss et al. (2021), Boix et al. (2022), and Marshall et al. (2020).

Moreover, the model's setting is close to Boleslavsky et al. (2021). Unlike their paper, we have a regime strength parameter (which denotes the type of ruler in our model) that is uncertain. Finally, our paper is also related to Andirin et al. (n.d.). Their paper makes a distinction between two forms of repression of protest events: *prevention* and *suppression*. The model we present belongs to the former type of repression, that is, when faced with a protest threat, the ruler can prevent it by engaging in media censorship.

There is a large literature in economics on the measurement of (e.g., Duclos et al., 2004; Esteban & Ray, 1994) and empirically reporting the trends in (e.g., Boxell et al., 2024; Draca & Schwarz, 2024) polarization. Rather than the measurement of the concept itself, we contribute to the literature on the effects of polarization (e.g., Esteban & Ray, 2011; Svolik, 2020). In particular, our paper is closely related to Gitmez and Molavi (2024-03-09). Rather than using the term polarization, they analyze how *diversity* of attitudes in a society affects media censorship. Our paper presents an alternative setting (protest threat), and uses a different model of strategic communication (signaling instead of Bayesian persuasion), which leads to contrasting results. [5]

The rest of the thesis is organized as follows. In Chapter 2, we present the model. We give the setup, find the equilibrium, and then derive our main result using monotone comparative statics. Then, we present an interpretation of the results. Chapter 3 presents the empirical evidence. Chapter 4 concludes. Most proofs are provided in the Appendices.

[5] In their paper, diversity of attitudes decreases the ruler's media censorship.

CHAPTER 2

MODEL

2.1 Setup

Players: There is a ruler and a continuum one of citizens. The ruler can be of two types. She is either strong $\theta = \overline{\theta}$ with probability μ_0, or weak $\theta = \underline{\theta}$ with probability $1 - \mu_0$ where $0 < \underline{\theta} < \overline{\theta} < 1$.

Actions: The ruler can "bribe" the media by paying $\kappa > 0$. This bribe is interpreted as the cost of hiding the bad news that the ruler is weak. [1] Absent any action by the leader, citizens can observe a "bad" signal [2] that perfectly reveals the leader's type. Then, the citizens simultaneously decide whether to protest ($a_i = 1$) or not ($a_i = 0$).

Payoffs: The payoff of the ruler is determined as follows: If she engages in media censorship, and the regime survives, she gets $\gamma - \kappa$ where $\gamma > 0$ denotes the regime's resources. If the regime collapses despite media censorship, she gets $-\kappa$. If the ruler does not engage in media censorship, and the regime survives, she gets γ. If the regime collapses, and there is no media censorship, she gets a

[1] The suppression of the bad news is what we mean by media censorship in this game.

[2] We use the terms bad news and bad signal interchangeably.

0 payoff. This is summarized in Table 2.1.

	Outcome	
	$A \geq \theta$	$A < \theta$
Media Censorship	$-\kappa$	$\gamma - \kappa$
No Media Censorship	0	γ

Table 2.1: Payoff Matrix of the Ruler

If the measure of protesters $A(\cdot)$ is greater than the type of ruler, the protest succeeds. Those who participate in the protest get a common benefit, denoted by $b > 0$, if the protest succeeds [3]. c_i denotes individual i's cost of protesting. The payoffs of the individuals are as follows:

	Outcome	
	$A \geq \theta$	$A < \theta$
Protest	$b - c_i$	$-c_i$
No Protest	0	0

Table 2.2: Payoff Matrix of the Individuals

Information: The ruler knows her type. If the citizens receive a bad signal, they learn the ruler's type. If the bad signal is suppressed, the citizens do not know the ruler's type. The individual i's private cost of protesting is determined by $c_i = \bar{c} + \rho\epsilon_i$ where $\bar{c} \sim U[0,1]$ and $\epsilon_i \sim U[\alpha - 1, \alpha]$ where $\alpha \in [0,1]$. $\rho > 0$ parametrizes the noise term ϵ_i. This setting is reminiscent of Boleslavsky et al. (2021). However, unlike them, the type of the ruler θ is uncertain in our case.

Timing: The timing of the game is as follows:

1. The ruler learns her type θ.
2. If the ruler is weak, she engages in media censorship with cost κ.
3. Citizens observe news, if any.

[3]The b can be considered as pleasure in agency à la Boleslavsky et al. (2021) and Wood (2003).

4. Each individual i receives her cost signal and decides whether to protest or not. From this, the measure of protesters $A(\cdot)$ is determined.

5. If $A \geq \theta$, the regime collapses.

6. The ruler and the individuals get their payoffs according to Table 1 and Table 2.

Strategies and Equilibrium: The citizens decide to protest according to symmetric monotone strategies such that $a_i = 1$ if and only if $c_i \leq c^*$. In a pure strategy, the weak ruler chooses to engage (or not) in media censorship with probability 1. In a mixed strategy, the weak ruler censors the media with probability σ. We use perfect Bayesian equilibrium as our equilibrium concept.

Assumption 1 *The media can only send a bad signal, i.e., the signal that the ruler is weak. This is why we characterize media censorship as a bribe for the suppression of bad news. If there is good news, i.e., the ruler is strong, the media does not send any signal. Because of this assumption,* ***only the weak ruler engages in media censorship in this model.*** [4]

2.2 Equilibrium

2.2.1 Global Game Given Posterior Belief μ

We begin by characterizing the equilibrium of the global game given citizens' beliefs μ. As discussed above, the equilibrium takes the cutoff form with cutoff c^*.

We first characterize the measure of protesters given c^*, as a function of the realization of the common component of costs.

[4] A similar assumption can be found in Besley and Prat (2006).

Lemma 1 $A(\overline{c}) = \frac{c^* - \overline{c}}{\rho} - \alpha + 1$. *Hence,* $A(\cdot)$ *is decreasing in* $\overline{c}$.

Proof. In the Appendix A. ■

Thanks to this lemma, we have cutoffs $\overline{c_1}^*$ and $\overline{c_2}^*$ that determine the protest success according to different types of rulers. Given this, we can state our main proposition:

Proposition 1 *There exists a symmetric equilibrium where* $a_i = 1$ *if and only if* $c_i \leq c^*(\mu) = [\mu(1 - \overline{\theta}) + (1 - \mu)(1 - \underline{\theta})]b$.

Proof. In the Appendix B. ■

The sketch of the proof is as follows. Given our cutoffs $\overline{c_1}^*$ and $\overline{c_2}^*$, we have two consistency conditions, where the measure of protesters $A(\overline{c})$ is equal to the type of ruler. Combining this with an indifference condition that characterizes the decision to protest and the Laplacian property, which connects the previous two conditions, we get the result.

From now on, to emphasize the dependency of c^* on μ, we will use the notation $c^*(\mu)$.

2.2.2 Survival Probability of the Ruler

Now, we need to characterize an equilibrium regarding the signaling part of the game. To do this, we need to find the ruler's survival probability. Considering Assumption 1, we focus on the probability of the survival of the weak ruler. [5]

[5]Although not necessary for our analysis, the results concerning the probability of the survival of the strong ruler are analogous.

Let us denote the probability of the survival of the ruler by $V(\mu)$. Then we have

$$V(\mu) \equiv \Pr_{\overline{c}}(A(\overline{c}) \leq \underline{\theta}). \tag{2.1}$$

From Lemma 1, we know that $A(\overline{c}) = \frac{c^*(\mu)-\overline{c}}{\rho} - \alpha + 1$. Then, we have

$$\begin{aligned} V(\mu) &= \Pr_{\overline{c}}\left(\frac{c^*(\mu) - \overline{c}}{\rho} - \alpha + 1 < \underline{\theta} \right) \\ &= \Pr_{\overline{c}}(\overline{c} > c^*(\mu) + (1 - \alpha - \underline{\theta})\rho). \end{aligned} \tag{2.2}$$

Let $U[0,1] = H(\cdot)$. Then we have

$$V(\underline{\theta}) = 1 - H(c^*(\mu) + (1 - \alpha - \underline{\theta})\rho). \tag{2.3}$$

Since $U[0,1] = H(\cdot)$, $V(\mu)$ can be written as follows:

$$V(\mu) = \begin{cases} 0, & \text{if } c^*(\mu) > 1 - (1 - \alpha - \underline{\theta})\rho \\ 1 - c^*(\mu) - (1 - \alpha - \underline{\theta})\rho, & \text{if } (\alpha + \underline{\theta} - 1)\rho < c^*(\mu) < 1 - (1 - \alpha - \underline{\theta})\rho \\ 1, & \text{if } c^*(\mu) < (\alpha + \underline{\theta} - 1)\rho \end{cases} \tag{2.4}$$

From this, we have the following lemma, which will be useful in the characterization of the equilibrium:

Lemma 2 *We have* $V(\mu) = 1 - c^*(\mu) - (1 - \alpha - \underline{\theta})\rho$ *if* $(\alpha + \underline{\theta} - 1)\rho < (1 - \overline{\theta})b$ *and* $(1 - \underline{\theta})b < 1 - (1 - \alpha - \underline{\theta})\rho$.

Proof. In the Appendix C. ■

2.2.3 Pooling, Separating, and Hybrid Equilibria

For the rest of the analysis, we are focusing on the case where $V(\mu) = 1 - c^*(\mu) - (1 - \alpha - \underline{\theta})\rho$, the condition for which is specified in Lemma 2.

Proposition 2 *(i) Consider the following beliefs:*

- *If the citizens observe news, they believe that the ruler is weak for certain* $(\mu = 0)$ *(out-of-equilibrium beliefs).*
- *If there is no news, the citizens' posterior belief is equal to their prior* $(\mu = \mu_0)$.

Then, we have a pooling equilibrium in which the weak ruler always engages in media censorship if $V(\mu_0)(\gamma - \kappa) \geq V(0)\gamma$.

(ii) Consider the following beliefs:

- *If the citizens observe news, they believe that the ruler is weak for certain* $(\mu = 0)$.
- *If there is no news, the citizens believe that the ruler is strong for certain* $(\mu = 1)$.

Then, we have a separating equilibrium in which the weak ruler always chooses not to engage in media censorship if $V(0)\gamma \geq V(1)(\gamma - \kappa)$.

(iii) Consider the following beliefs:

- *If the citizens observe news, they believe that the ruler is weak for certain* $(\mu = 0)$.

- *If there is no news, the citizens form a posterior belief μ, which is pinned down by Bayes' rule. That is $\mu = \hat{\mu} \equiv \frac{\mu_0}{\mu_0 + (1-\mu_0)\sigma}$*

Then, we have a hybrid equilibrium in which the weak ruler randomizes between engaging (with probability σ) and not engaging in media censorship if $V(\hat{\mu})(\gamma - \kappa) = V(0)\gamma$.

Proof. For (i) and (ii), we make sure that the ruler does not have an incentive to deviate given the definition of the equilibrium, and for (iii), we use a standard argument in finding an equilibrium involving mixed strategies, that is, the ruler should be indifferent in choosing either action. Beliefs are derived from the definition of the respective equilibrium. ■

For our main result, we will only focus on the hybrid equilibrium. This type of equilibrium has a nice property in that it allows us to do comparative statics analysis between σ and ρ. The following corollary characterizes the case in which we have a unique perfect Bayesian equilibrium, and this equilibrium is of the hybrid type:

Corollary 1 *There exists a unique perfect Bayesian equilibrium where the strong ruler does not engage in media censorship, and the weak ruler randomizes between engaging (with probability σ) and not engaging in media censorship if we have*

$$\frac{(c_2^* - c_1^*)\gamma}{1 - c_1^* - \psi} < \kappa < \frac{(c_2^* - c_3^*)\gamma}{1 - c_3^* - \psi}$$

where $c_1^* \equiv [\mu_0(1-\overline{\theta}) + (1-\mu_0)(1-\underline{\theta})]b$, $c_2^* \equiv (1-\underline{\theta})b$, $c_3^* \equiv (1-\overline{\theta})b$, and $\psi \equiv (1-\alpha-\underline{\theta})\rho$.

Proof. In the Appendix D. ■

Intuitively, the condition in Corollary 1 says that to have a unique perfect Bayesian equilibrium which is of the hybrid type, the cost of media censorship

should not be too high or too low. The former case leads to a separating equilibrium (the ruler cannot afford to engage in media censorship at all), the latter case results in a pooling equilibrium (media censorship is cheap and thus the ruler always does it).

2.2.4 Optimal Level of Media Censorship

Using the results from the previous sections, we can present our main conclusion:

Theorem 1 *In the unique perfect Bayesian equilibrium, the optimal level of media censorship (σ^*) is given by* [6]

$$\sigma^* = \frac{\mu_0[(\overline{\theta} - \underline{\theta})\gamma b + (1 - \overline{\theta})\kappa b + (1 - \alpha - \underline{\theta})\kappa\rho - \kappa]}{(1 - \mu_0)[\kappa - (1 - \underline{\theta})\kappa b - (1 - \alpha - \underline{\theta})\kappa\rho]}.$$

Proof. Setting $V(0)\gamma = V(\hat{\mu})(\gamma - \kappa)$ gives the result. ■

Corollary 2 *An increase in ρ increases the optimal level of media censorship if $\underline{\theta} < \underline{\theta}^* \equiv 1 - \alpha$.*

2.3 Deriving the Main Result Using Monotone Comparative Statics

As can be inferred from Theorem 1, finding the optimal level of media censorship is straightforward, but it involves tedious algebra. Moreover, from Corollary 2, one can notice that ρ and σ^* are complements under some settings (i.e., when $\underline{\theta} < \underline{\theta}^* \equiv 1 - \alpha$). Given these remarks, one can use monotone comparative statics analysis, which can allow us to derive the comparative statics result in Corollary 2 without explicitly finding the optimal level of media censorship. To do this, we

[6] We also need that $\sigma^* \in (0, 1)$. That is, we need $0 < \frac{\mu_0[(\overline{\theta} - \underline{\theta})\gamma b + (1 - \overline{\theta})\kappa b + (1 - \alpha - \underline{\theta})\kappa\rho - \kappa]}{(1 - \mu_0)[\kappa - (1 - \underline{\theta})\kappa b - (1 - \alpha - \underline{\theta})\kappa\rho]} < 1$.

need to focus on the signaling part of the model. [7]

We can characterize the equilibrium in the signaling part as follows:

$$\sigma^* \in \arg\max_{\sigma \in (0,1)} \sigma[V(\mu)(\gamma - \kappa)] + (1 - \sigma)[V(0)\gamma] \tag{2.5}$$

given

$$\mu = \frac{\mu_0}{\mu_0 + (1 - \mu_0)\sigma} \tag{2.6}$$

2.5 and 2.6 combined tell us the following: the ruler maximizes her expected payoff given the posterior belief μ held by citizens.

Let $U(\sigma|\mu, \rho) \equiv \sigma[V(\mu)(\gamma - \kappa)] + (1 - \sigma)[V(0)\gamma]$. Also, consider the function $\sigma^*(\mu, \rho)$, which gives the optimal level of media censorship given μ and ρ. Figure 1 illustrates the conjectured signaling equilibrium. The intersection of the graph of $\sigma^*(\mu, \underline{\rho})$ ($\sigma^*(\mu, \overline{\rho})$) and $\mu(\sigma)$ represents the signaling equilibrium when we have a low (high) value of ρ. If we can formally verify this figure, we can conclude that σ and ρ are complements in the signaling equilibrium.

Before going over the monotone comparative statics analysis regarding our model, let us present the following theorem, which will be useful in the subsequent analysis. We begin with the following definitions (Gitmez et al., 2025; Milgrom & Shannon, 1994). Let $V(\rho, \mu) : \mathbb{R}^2 \to \mathbb{R}$ be a function. We say that

[7] Throughout our analysis, we are making the implicit assumption that the parameters satisfy the conditions that guarantee the existence of the hybrid equilibrium. These conditions are specified in the previous section.

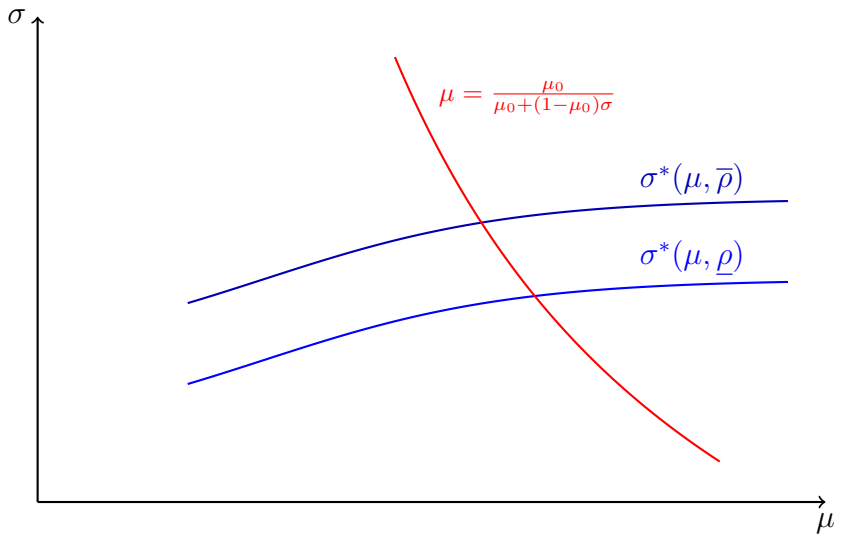

Figure 1: (Generic) Graphical Representation of the Signaling Equilibrium

- $V(\rho, \mu)$ satisfies single-crossing-from-above property in ρ if given μ, we have that $V(\rho, \mu) \geq 0$ implies $V(\rho', \mu) \geq 0$ for all $\rho' < \rho$.
- $V(\rho, \mu)$ satisfies strict single-crossing-from-above property in μ if given ρ, we have that $V(\rho, \mu) \geq 0$ implies $V(\rho, \mu') > 0$ for all $\mu' < \mu$.

Theorem 2 *Suppose that there exists $\mu^*(\rho)$ such that $V(\rho, \mu^*(\rho)) = 0$. If $V(\rho, \mu)$ satisfies the conditions above, we have that $\mu^*(\rho)$ is decreasing in ρ.*

Proof. In the Appendix E. ■

Using Theorem 2, we present the following proposition, which is in line with Corollary 2:

Proposition 3 *Define $h(\rho, \mu) \equiv \sigma(\mu) - \sigma^*(\mu, \rho)$, where $\sigma(\mu) \equiv \mu^{-1}(\sigma)$ [8]. We have that $\mu^*(\rho)$[9] is decreasing in ρ if $\underline{\theta} < \underline{\theta}^* \equiv 1 - \alpha$.*

Proof. In the Appendix F. ■

[8] $\mu(\sigma) = \frac{\mu_0}{\mu_0+(1-\mu_0)\sigma}$.

[9] $\mu^*(\rho)$ as defined in the Theorem 2, that is, $h(\rho, \mu^*(\rho)) = 0$.

2.4 Discussion of the Result

Before we interpret our result in Corollary 2, we need to discuss how we can conceptualize the increase in ρ. A good idea is to inspect the PDF of the cost signal c_i as we increase ρ. To do this, we need the following lemma:

Lemma 3 *Let $\rho \leq 1$ [10] and $Z \equiv c_i$. The PDF of Z is given as follows:*

$$f(z) = \begin{cases} 0, & \textit{if } z < (\alpha-1)\rho \quad \textit{or} \quad z > 1+\alpha\rho \\ \frac{z}{\rho} + (1-\alpha), & \textit{if } (\alpha-1)\rho \leq z \leq \alpha\rho \\ 1, & \textit{if } \alpha\rho \leq z \leq (\alpha-1)\rho + 1 \\ \frac{1-z}{\rho}, & \textit{if } (\alpha-1)\rho + 1 \leq z \leq 1+\alpha\rho \end{cases} \tag{2.7}$$

Proof. In the Appendix. ■

Notice that the PDF of c_i has three main parts: (i) the *increasing region* [11] where $(\alpha-1)\rho \leq z \leq \alpha\rho$, (ii) the *constant region* where $\alpha\rho \leq z \leq (\alpha-1)\rho + 1$ [12], and (iii) the *decreasing* region where $(\alpha-1)\rho + 1 \leq z \leq 1+\alpha\rho$ [13]. From this observation, we give the following definition:

Definition 1 *As ρ increases, the range of the increasing and decreasing regions expands, whereas the range of the constant region shrinks. This is called the* ***polarization*** *of cost signals.*

---

[10] This assumption is made to make an easier discussion of the result without changing the analysis in previous sections.

[11] The region where the graph of PDF has a positive slope.

[12] The region where the PDF is equal to 1.

[13] The region where the graph of PDF has a negative slope.

To understand the intuition behind this definition, we present Figure 2, in which we see the effect of the increase in ρ on the PDF of c_i where $\alpha = \frac{1}{2}$.

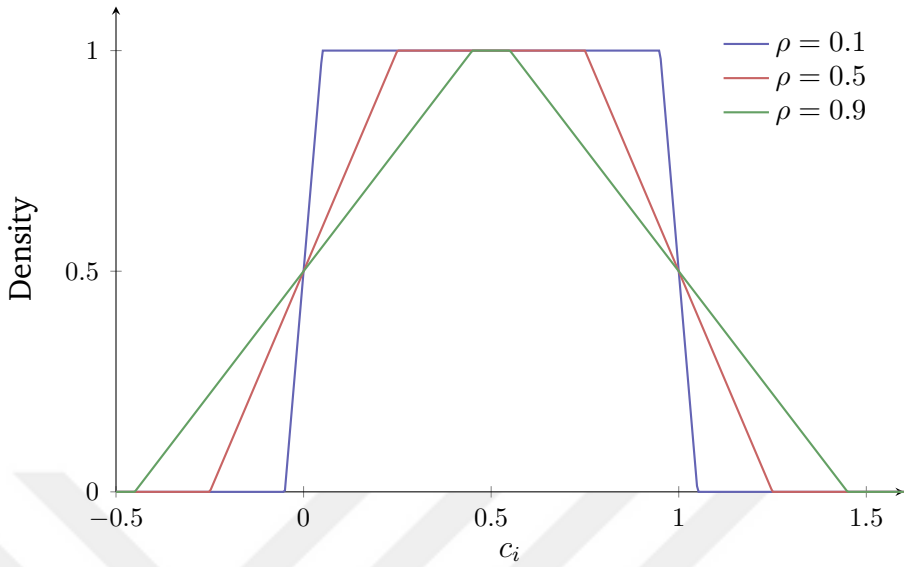

Figure 2: Effect of increasing ρ on the PDF of c_i for $\alpha = \frac{1}{2}$.

As can be observed from the figure, those who are receiving the lowest and highest cost signals are now getting more extreme signals (even lower and higher). In other words, *extremes are becoming more extreme*. On the other hand, the measure of those who are receiving moderately opposing signals (i.e., below and above the median within a certain radius) is decreasing. This is why we conceptualize the effect of an increase in ρ as polarization of cost signals.

Based on this, our concept of polarization is close to the one presented in Johnson and Myatt (2006). Rather than using polarization, they define *spread* as the change in the density away from the center to the lower and upper tails. Notice that in Figure 2, as ρ increases, we observe an increase in the support of the variable c_i and a decrease in the density *around* the center. Therefore, our concept of polarization is not identical to that of spread in Johnson and Myatt (2006).

Next, we need to understand why we have that increase in ρ increases the

optimal level of media censorship when $\underline{\theta} < \underline{\theta}^* \equiv 1 - \alpha$. To get the logic behind this result, note the following: In the hybrid equilibrium, the ruler chooses a higher σ^* when his survival probability $V(\mu)$ decreases (keeping $c^*(\mu)$ fixed). Mechanically, this result follows from the indifference condition $V(0)\gamma = V(\hat{\mu})(\gamma - \kappa)$. This observation also makes sense intuitively in that the ruler needs a more aggressive cover-up when he is confronted with a higher risk of overthrow.[14] The next question is: Why does the increase in ρ affect the ruler's survival probability negatively when $\underline{\theta} < 1 - \alpha$? It turns out that the answer to this question is closely related to the effect of the increase in ρ on $A(\overline{c})$.

In Figure 3, we can observe that the increase in ρ affects the survival probability of the ruler negatively when $\underline{\theta} < \frac{1}{2}$. Denote the x-axis of the intersection of the graph of $A(\overline{c})$ and the horizontal line indicating the ruler's type $\underline{\theta}$ as $\overline{c}^*$. Notice that $\Pr(\overline{c} < \overline{c}^*)$ denotes the probability of the regime collapse. As can be seen from the figure, an increase in ρ increases this probability in the region where $\underline{\theta} < \underline{\theta}^*$. Equivalently, in the region $\underline{\theta} < \underline{\theta}^*$, we see that increase in ρ increases the measure of protesters $A(\overline{c})$. These two observations lead us to the following conclusion: The increase in ρ affects the ruler's survival negatively when $\underline{\theta} < 1 - \alpha$ because in this region increase in ρ increases the measure of protesters. The next crucial thing to get is where the threshold of $\underline{\theta}^*$ comes from.

Given that $A(\overline{c}) = \frac{c^* - \overline{c}}{\rho} - \alpha + 1$, when $\overline{c} > c^*$, an increase in ρ increases the measure of protesters. To see the intuition behind this result, consider the line in Figure 4 [15].

As can be seen, an increase in ρ has two effects: it can amplify the effect of the positive and negative noise terms on the cost signal c_i. If $\overline{c} > c^*$, the former

[14] This is also in line with the empirical result in VonDoepp and Young (2013).

[15] The labels $-\frac{\rho}{2}$ and $\frac{\rho}{2}$ reflect the bounds of the noise term when we have $\alpha = \frac{1}{2}$

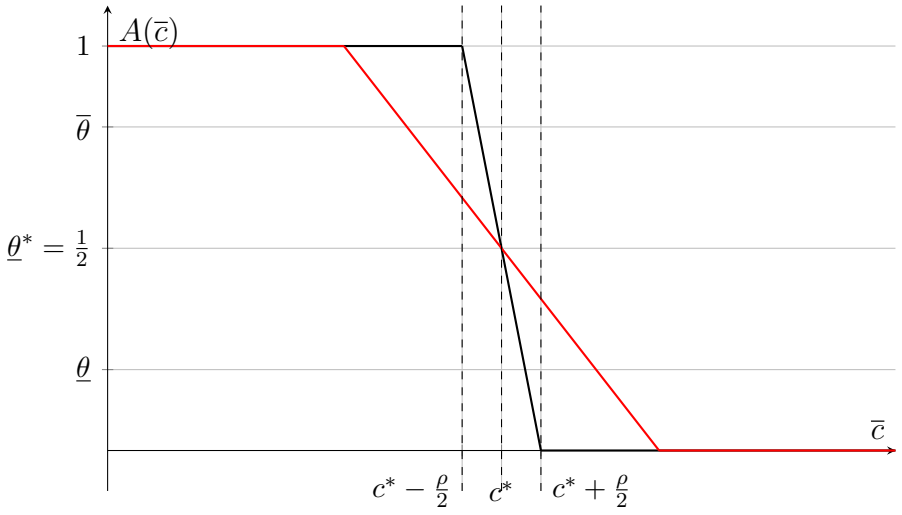

Figure 3: Comparison of $A(\bar{c})$ for low and high ρ values with $\alpha = \frac{1}{2}$

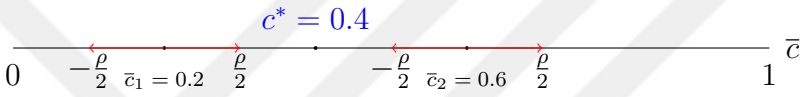

Figure 4: Illustration of $\bar{c}$ with threshold c^* and two realizations $\bar{c}_1$, $\bar{c}_2$

does not affect the measure of protesters. However, the latter can increase the measure of protesters by shifting the cost signals received below the cut-off value c^*. That is, when $\bar{c} > c^*$, an increase in ρ negatively affects the survival probability of the ruler by increasing the measure of protesters. The case of $\bar{c} > c^*$ corresponds to $A(\bar{c}) < 1-\alpha$, which is also the region where $\underline{\theta} < \underline{\theta}^* \equiv 1-\alpha$. This is the reason why when the ruler's type is below the threshold of $1-\alpha$, the increase in ρ decreases the ruler's survival probability.

Finally, the region where we observe that an increase in polarization leads to an increase in the optimal media censorship ($\underline{\theta} < \underline{\theta}^*$) depends on α. A decrease in α increases the region where we observe this relationship. We can interpret α as the ruler's *office security*. The opposite concept, office insecurity, can be defined by "the degree of threat the incumbent faces from the opposition, in the form of a rebellion, a protest, or a coup" (Rozenas, 2016, p. 232). Recall that α enters into our model as the determinant of the support of the noise term ϵ_i. As we move from $\alpha = 1$ to $\alpha = 0$, the decrease in α shifts the support

of the cost signals to the left. That is, for a given $\bar{c}$ and ρ, the decrease in α increases the general public's incentive to protest. Hence, the decrease in α can be interpreted as the increase in the ruler's office insecurity.

Figure 5[16] presents the effect of an increase in α on cost signals in the $\bar{c}$ line. At the top line, we have $\alpha = 0.1$. This represents a high office insecurity case in which most people receive negative noise terms[17] to the shared cost term $\bar{c}$. At the bottom line, we have $\alpha = 0.9$. This represents a low office insecurity case in which most people receive positive noise terms[18] to the shared cost term $\bar{c}$. An increase in ρ magnifies the presence of the negative and positive noise receivers in these extreme cases. In the case of high office insecurity, the polarization amplifies the presence of those who wish to participate, and in the case of low office insecurity, this amplification affects those who are hesitant to protest. Given that conceptually, protest threat and office insecurity are concomitant phenomena, our model's empirically testable implication is as follows: **When the ruler faces a protest threat, an increase in polarization increases media censorship.** This is because in this setting, the increase in polarization has only a unidirectional salient effect. This effect is the increase in those who are a threat to the ruler.

[16]The labels at the end of the arrows reflect the bounds of the noise term as in Figure 4.

[17]This is represented by the long left arrows and short right arrows.

[18]This is represented by the long right arrows and short left arrows.

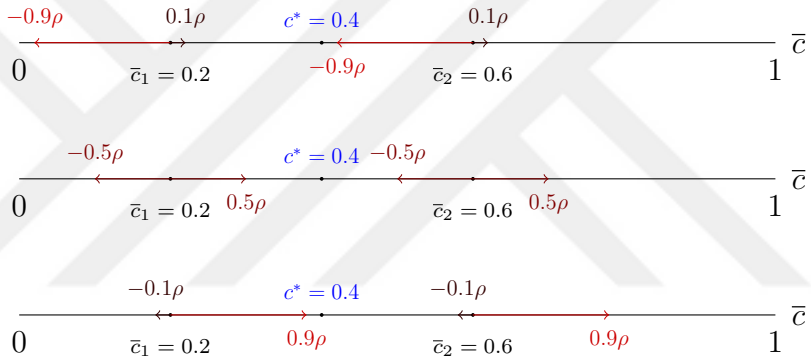

Figure 5: Effect of the increase in polarization on cost signals for varying office insecurity levels.

CHAPTER 3

EMPIRICAL EVIDENCE

3.1 Data

We have five main sources of data. Our dependent variable (media freedom) and independent variable (polarization) are from the Varieties of Democracy (V-Dem) Project (Coppedge et al., 2024). The protest data is from the Nonviolent and Violent Conflict Outcomes (NAVCO) 1.3 Data Set (Chenoweth & Shay, 2020). The GDP per capita and population controls are from Fariss et al. (2021). We will present our results using different regime controls, which are from Boix et al. (2022) and Marshall et al. (2020).

V-Dem gathers expert opinions from scholars worldwide to conceptualize and measure how democratic a regime is (V-Dem Institute, n.d.). The media freedom index we use (v2mecenefm) measures the government censorship effort and is formed by the responses to the following question: "Does the government directly or indirectly attempt to censor the print or broadcast media?" [1] (Coppedge et al., 2024, p. 207). The responses are denoted by an ordinal scale from 0 to 4. 0 denotes the lowest possible media freedom (i.e., direct and routine

[1] Given that in our model, all individuals decide whether to protest or not based on the same cutoff c^*, the ruler's media censorship is perceived the same across citizenry. This is why we focus on traditional media such as print or broadcast, since access to these can be considered more homogeneous across the population than the internet.

censorship), whereas 4 denotes the highest media freedom (i.e., rare attempts at media censorship).

V-Dem data set includes two measures of polarization. The first one is called polarization of society (v2smpolsoc), and it is formed by the responses to the following question: "How would you characterize the differences of opinions on major political issues in this society?" (Coppedge et al., 2024, p. 346). The second one is called political polarization (v2cacamps), and it is formed by the responses to the following question: "Is society polarized into antagonistic, political camps?" (Coppedge et al., 2024, p. 232). Like the media freedom, the answers to these questions are denoted by an ordinal scale from 0 to 4, where higher values indicate more polarization.[2] We use the interval point estimate equivalents of the 0-4 ordinal scales in the V-Dem data set, which is converted by the measurement model (see Pemstein et al., 2022). [3]

NAVCO data set includes 622 maximalist[4] resistance campaigns between 1900-2019 worldwide. For this paper, we focus on non-violent regime change [5] campaigns, which are gathered from news sources, encyclopedic entries, scholarly articles, and the campaigns used are corroborated by experts on the subject (Chenoweth & Shay, 2020). The data set includes information on the beginning

[2]Actually, in the original data set, the answers to the v2smpolsoc are ordered such that lower values indicate higher polarization, and the answers to v2cacamps are vica versa. We multiply the v2smpolsoc by minus one to ease the exposition in the subsequent analysis.

[3]As noted in Coppedge et al. (2021), this model treats ordinal codings of variables as reflections of interval-valued latent characteristics.

[4]This means that the data set focuses on campaigns whose objectives are maximalist (e.g., regime change, secession) rather than limited (e.g., demanding greater civil rights) (Chenoweth & Shay, 2020).

[5]Regime change campaigns are those whose objective is to remove the incumbent national leader.

and the peak year of events of the campaigns. We created a dummy variable called *Protest*, which is equal to 1 if a country-year observation overlaps with the location and it is between the beginning and the peak years of the campaign.

Arguably, one of the important issues in this context is controlling for the regime type. Although our model does not explicitly include a dimension regarding which countries the settings of the model could arise, i.e., in what countries the threat of a regime change protest should occur, one can argue that this type of model is more suitable to analyze non-democracies. [6] To account for such a concern, we check the robustness of our findings using two different regime indicators that are widely paid attention to in the political economy literature (e.g., Acemoglu et al., 2019). The first is a dichotomous coding of democracy by Boix et al. (2022). According to this data, a country is considered democratic if its political leaders are chosen through free and fair elections and there is a certain level of suffrage. [7]. The second regime control is from the Polity 5 data set, which denotes the regime of a country on an autocracy-democracy scale (Marshall et al., 2020). The index ranges from -10 (hereditary monarchy) to +10 (consolidated democracy), which is based on the "key qualities of executive recruitment, constraints on executive authority, and political competition" (Center for Systemic Peace, n.d. para. 5) [8]. Lastly, Fariss et al. (2021)'s estimates of GDP per capita and population are formed by using different data sets (e.g., World Bank (2021) and Penn World Tables (Feenstra et al., 2015), conceptualizing these as latent variables observed with measurement error.

[6] In their survey of the political economics of non-democracy, Egorov and Sonin (2024) devote particular attention to models of information manipulation and protests.

[7] The suffrage condition is not the binding one for our study, considering that we are covering the period 2000-2018.

[8] In particular, we use Polity's revised combined score which converts special values of -66, -77, -88 to conventional polity scores (Coppedge et al., 2024).

Combining these five data sets, we constructed a panel of up to 179 countries for the years 2000-2018. The time frame of our study is limited by the availability of all the variables used (e.g, V-Dem's polarization of society covers the period 2000-2023, the last year that Polity 5 covers is 2018). We argue that although focusing on this recent period is dictated by the limitation of available data, it suits the narrative of our paper [9].

3.2 A Brief Discussion Regarding V-Dem's Polarization Measures

One might ask to what extent V-Dem's polarization of society and political polarization measures can proxy for the concept of polarization as defined in the model. To answer this question, first, we need to understand how these variables are related to the different types of polarization mentioned in the literature. Polarization of society variable is concerned with the difference of opinions in the society and whether this leads to major clashes of views (Coppedge et al., 2024). In contrast, the political polarization variable is related to how much political differences affect social relationships, i.e., whether it results in animosity between the proponents of opposing political camps (Coppedge et al., 2024). [10] Given this, although not explicitly mentioned in the V-Dem codebook, it can be argued that polarization of society is more related to *ideological polarization* (e.g., Abramowitz & Saunders, 2008; Draca & Schwarz, 2024). In Draca and Schwarz's words, this can be characterized by a *disappearing center*, and "shifting away from centrist ideologies into anti-establishment 'anarchist' ideologies" (Draca & Schwarz, 2024, p. 1950). On the other hand, considering that political

[9] See Chapter 1.

[10] It seems that the naming of these concepts is a bit confusing in the V-Dem codebook. This is because political polarization seems more related to societal consequences of polarization, whereas polarization of society is more related to differences in policy opinions.

polarization measures to what extent political differences lead to animosity in the social sphere, it is more related to the *affective polarization* (Iyengar et al., 2012) [11].

Based on this discussion, we can argue that our model's notion of polarization can suit both types of polarization. Consider a case in which we have a high polarization of cost signals. That is, society consists of individuals who have a high incentive to engage in protest (anti-incumbent) and those who virtually do not have an incentive to engage in protest (pro-incumbent). Both ideological and affective polarization can apply to this setting. That is, it is plausible that those who are in the opposing camps have serious differences in opinions and also hold hostile attitudes towards each other. Therefore, it is viable to use both measures of polarization in V-Dem to test the model empirically.

3.3 Results

In this section, we present the regression results using our panel data. In our full specification, we estimate the following model:

$$\begin{aligned} \textit{Media Freedom}_{i,t} = \alpha + \beta_1 \textit{Polarization}_{i,t} + \beta_2 \textit{Protest}_{i,t} + \\ \beta_3 \textit{Polarization}_{i,t} \times \textit{Protest}_{i,t} + \delta' X_{i,t} + \gamma \textit{Trend}_t + \lambda_i + \epsilon_{i,t} \end{aligned} \tag{3.1}$$

where $X_{i,t}$ is the vector of controls including the regime index (Boix et al. (2022) or Polity 5), GDP per capita and population; $\textit{Trend}_t = t - 1999, t$ denoting year t;

[11] The animosity between the opposing political groups is indeed how affective polarization is defined in the literature (e.g., Iyengar et al., 2019).

λ_i denotes country fixed effect, and $\epsilon_{i,t}$ is the error term [12].

For simplicity, we denote polarization of society as *Polarization (1)*, and political polarization as *Polarization (2)*. Our main regressor of interest is the interaction term $Polarization \times Protest$. Recall that our model's main testable implication is that as the ruler faces a protest threat, an increase in polarization leads to an increase in media censorship. Given that in our empirical analysis, we conceptualize our dependent variable as media freedom, we expect the coefficient of this interaction term to be negative. In all of the regressions, we include country fixed effects. Following Petrova (2008), we also include year time trends in some of our specifications. Again, following Petrova, we only focus on the sign and statistical significance of the coefficients, given that the usage of indices does not allow us to make meaningful interpretations about our regression coefficients.

In Table 3.1, we present our first results, using polarization of society as the indicator of polarization and Boix et al. (2022)'s dichotomous measure of democracy as the regime control. In all of the specifications, the coefficient of the interaction term has a negative sign, and the relationship is statistically significant at between 5 and 10% levels.

In Table 3.2, we use political polarization as the indicator of polarization, and again use Boix et al. (2022) as the regime control. In all of the specifications, the coefficient of interest is negative and significant at below the 10% level.

In Table 3.3 and Table 3.4, we go over the same procedure, now using the Polity 5 data set as the regime control. The coefficient of the interaction term is negative and statistically significant across all specifications when we use

[12] In all specifications, standard errors are clustered at the country level.

Table 3.1: Media Freedom and Polarization of Society Using Boix et al. (2022)

	Media Freedom			
	(1)	(2)	(3)	(4)
Polarization (1)	-0.005	-0.005	0.056	0.038
	(0.091)	(0.095)	(0.101)	(0.100)
Protest	-0.153*	-0.153*	-0.131*	-0.133*
	(0.091)	(0.091)	(0.073)	(0.073)
Polarization (1) × Protest	-0.134*	-0.134*	-0.137**	-0.136*
	(0.075)	(0.075)	(0.069)	(0.069)
Year (time trend)		0.000		0.004
		(0.004)		(0.006)
Democracy			0.471***	0.462***
			(0.135)	(0.136)
GDP per capita			-0.010	-0.013
			(0.007)	(0.009)
Population			-0.000***	-0.000***
			(0.000)	(0.000)
R^2	0.913	0.913	0.917	0.917
Adj. R^2	0.908	0.908	0.912	0.912

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

polarization of society as the indicator of polarization. However, when we use political polarization, although the coefficient of interest is still negative, it is not statistically significant in the last two specifications (p-values are around 0.15).

In conclusion, our analysis suggests that our model's main hypothesis has empirical relevance. We showed that the results are robust to regime index choice if we use the polarization of society as the indicator of polarization. The use of political polarization also results in coefficients with the expected sign; however, the statistical significance is not robust to regime index choice.

Table 3.2: Media Freedom and Political Polarization Using Boix et al. (2022)

	Media Freedom			
	(1)	(2)	(3)	(4)
Polarization (2)	-0.159*	-0.172**	-0.112	-0.130
	(0.085)	(0.086)	(0.096)	(0.093)
Protest	-0.084	-0.091	-0.073	-0.079
	(0.073)	(0.072)	(0.068)	(0.065)
Polarization (2) × Protest	-0.144*	-0.143*	-0.143*	-0.146*
	(0.074)	(0.074)	(0.075)	(0.075)
Year (time trend)		0.004		0.008
		(0.004)		(0.006)
Democracy			0.418***	0.398***
			(0.134)	(0.132)
GDP per capita			-0.005	-0.013
			(0.007)	(0.009)
Population			-0.000**	-0.000***
			(0.000)	(0.000)
R^2	0.916	0.916	0.919	0.919
Adj. R^2	0.911	0.911	0.914	0.914

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Media Freedom and Polarization of Society Using Polity 5

	Media Freedom			
	(1)	(2)	(3)	(4)
Polarization (1)	-0.005	-0.005	-0.022	-0.011
	(0.091)	(0.095)	(0.075)	(0.077)
Protest	-0.153*	-0.153*	-0.068	-0.066
	(0.091)	(0.091)	(0.075)	(0.075)
Polarization (1) × Protest	-0.134*	-0.134*	-0.116*	-0.117*
	(0.075)	(0.075)	(0.070)	(0.070)
Year (time trend)		0.000		-0.003
		(0.004)		(0.005)
Polity			0.108***	0.109***
			(0.017)	(0.017)
GDP per capita			-0.010	-0.007
			(0.006)	(0.007)
Population			-0.000***	-0.000***
			(0.000)	(0.000)
R^2	0.913	0.913	0.933	0.933
Adj. R^2	0.908	0.908	0.929	0.929

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Media Freedom and Political Polarization Using Polity 5

	Media Freedom			
	(1)	(2)	(3)	(4)
Polarization (2)	-0.159*	-0.172**	-0.121	-0.120
	(0.085)	(0.086)	(0.082)	(0.081)
Protest	-0.084	-0.091	-0.058	-0.058
	(0.073)	(0.072)	(0.066)	(0.066)
Polarization (2) × Protest	-0.144*	-0.143*	-0.103	-0.103
	(0.074)	(0.074)	(0.071)	(0.071)
Year (time trend)		0.004		-0.000
		(0.004)		(0.005)
Polity			0.103***	0.103***
			(0.019)	(0.018)
GDP per capita			-0.007	-0.007
			(0.006)	(0.008)
Population			-0.000***	-0.000**
			(0.000)	(0.000)
R^2	0.916	0.916	0.935	0.935
Adj. R^2	0.911	0.911	0.931	0.931

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CHAPTER 4

CONCLUSION

There are recent papers that analyze the cross-country trends in polarization (e.g., Boxell et al., 2024; Draca & Schwarz, 2024; Reiljan, 2020) which aim to show that the relevance of the concept is not restricted to the US (for example, the term *affective polarization* is first defined in this context (Iyengar et al., 2012)). This study is another attempt at enlarging the relevance of the concept by showing its potential role in a specific setting. We showed that in the face of a protest threat, polarization has a unidirectional salient effect on attitudes. This is the amplification of the presence of those who are a threat to the regime.

In terms of the interaction between media censorship and collective action, our model presented a mechanism in which the ruler affects the citizens' incentives to protest through an information channel without directly influencing the individuals' coordination capabilities (see also Fergusson and Molina (2020) for information versus coordination debate in the literature). Our model calls for increased attention to media censorship made through traditional media, which can be more associated with information rather than the coordination channel.

Our empirical analysis is only suggestive and serves the purpose of showing that the model's main result is empirically relevant. We believe that finding new measures of polarization that can apply to a wider set of countries is one of the

ways in which we can make our empirical analysis more credible. This is left to future research.

REFERENCES

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, *70*(2), 542–555. https://doi.org/10.1017/S0022381608080493

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, *127*(1), 47–100. https://doi.org/10.1086/700936

Andirin, V. M., Neggers, Y., Shadmehr, M., & Shapiro, J. M. (n.d.). *Surveillance of repression: Theory and implementation* [Unpublished manuscript].

Besley, T., & Prat, A. (2006). Handcuffs for the grabbing hand? media capture and government accountability. *American Economic Review*, *96*(3), 720–736. https://doi.org/10.1257/aer.96.3.720

Boix, C., Miller, M., & Rosato, S. (2022). *Boix-miller-rosato dichotomous coding of democracy, 1800–2020 (v1)*. Harvard Dataverse. https://doi.org/10.7910/DVN/FENWWR

Boleslavsky, R., Shadmehr, M., & Sonin, K. (2021). Media freedom in the shadow of a coup. *Journal of the European Economic Association*, *19*(3), 1782–1815. https://doi.org/10.1093/jeea/jvaa040

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2024). Cross-country trends in affective polarization. *Review of Economics and Statistics*, *106*(2), 557–565. https://doi.org/10.1162/rest_a_01160

Cantoni, D., Kao, A., Yang, D. Y., & Yuchtman, N. (2024). Protests. *Annual Review of Economics*, *16*, 519–543. https://doi.org/10.1146/annurev-economics-082423-032519

Carlsson, H., & van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, *61*(5), 989–1018. https://doi.org/10.2307/2951491

Center for Systemic Peace. (n.d.). Polity project [Retrieved November 26, 2024].

Chenoweth, E., & Shay, C. W. (2020). *List of campaigns in navco 1.3* (Version V1) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/ON9XND

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Angiolillo, F., Bernhard, M., Borella, C., Cornell, A., Fish, M. S., Fox, L., Gastaldi, L., Gjerlow, H., Glynn, A., God, A. G., Grahn, S., Hicken, A., Kinzelbach, K., … Ziblatt, D. (2024). *V-dem [country-year/country-date] dataset v14*. Varieties of Democracy (V-Dem) Project. https://doi.org/10.23696/mcwt-fr58

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Angiolillo, F., Bernhard, M., Borella, C., Cornell, A., Fish, M. S., Fox, L., Gastaldi, L., Gjerløw, H., Glynn, A., God, A. G., Grahn, S., Hicken, A., Kinzelbach, K., … Ziblatt, D. (2024). *V-dem codebook v14*. Varieties of Democracy (V-Dem) Project.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Marquardt, K. L., Medzihorsky, J., Pemstein, D., Alizada, N., Gastaldi, L., Hindle, G., Pernes, J., von Römer, J., Tzelgov, E., Wang, Y.-t., & Wilson, S. (2021). V-dem methodology v11.1.

Djankov, S., McLiesh, C., Nenova, T., & Shleifer, A. (2003). Who owns the media? *The Journal of Law and Economics*, *46*(2), 341–382. https://doi.org/10.1086/377116

Draca, M., & Schwarz, C. (2024). How polarised are citizens? measuring ideology from the ground up. *The Economic Journal*, *134*(661), 1950–1984. https://doi.org/10.1093/ej/ueae010

Duclos, J.-Y., Esteban, J.-M., & Ray, D. (2004). Polarization: Concepts, measurement, estimation. *Econometrica*, *72*(6), 1737–1772. https://doi.org/10.1111/j.1468-0262.2004.00552.x

Edmond, C. (2013). Information manipulation, coordination, and regime change. *The Review of Economic Studies*, *80*(4), 1422–1458. https://doi.org/10.1093/restud/rdt020

Egorov, G., & Sonin, K. (2024). The political economics of non-democracy. *Journal of Economic Literature*, *62*(2), 594–636. https://doi.org/10.1257/jel.20221494

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, *88*(4), 1479–1514. https://doi.org/10.3982/ECTA14281

Esteban, J.-M., & Ray, D. (1994). On the measurement of polarization. *Econometrica*, *62*(4), 819–851. https://doi.org/10.2307/2951734

Esteban, J.-M., & Ray, D. (2011). Linking conflict to inequality and polarization. *American Economic Review*, *101*(4), 1345–1374. https://doi.org/10.1257/aer.101.4.1345

Fariss, C., Anders, T., Markowitz, J., & Barnum, M. (2021). *Replication data for: New estimates of over 500 years of historic gdp and population data* (Version V4) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/DC0ING

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, *105*(10), 3150–3182. https://doi.org/10.1257/aer.20130954

Fergusson, L., & Molina, C. (2020). *Facebook causes protests (hicn working paper no. 323)* (tech. rep.). Households in Conflict Network. https://hicn.org/wp-content/uploads/2020/04/HiCN-WP-323.pdf

Gehlbach, S., & Sonin, K. (2014). Government control of the media. *Journal of Public Economics*, *118*, 163–171. https://doi.org/10.1016/j.jpubeco.2014.06.004

Gitmez, A. A., & Molavi, P. (2024-03-09). *Informational autocrats, diverse societies (cepr discussion paper no. dp18903)* (tech. rep.). Centre for Economic Policy Research. https://cepr.org/publications/dp18903

Gitmez, A. A., Molavi, P., & Sonin, K. (2025). *Propaganda and repression in diverse societies* [Unpublished manuscript]. https://pooyamolavi.com/Propaganda_and_Repression.pdf

Guriev, S., & Treisman, D. (2019). Informational autocrats. *Journal of Economic Perspectives*, *33*(4), 100–127. https://doi.org/10.1257/jep.33.4.100

Guriev, S., & Treisman, D. (2020). A theory of informational autocracy. *Journal of Public Economics*, *186*, 104158. https://doi.org/10.1016/j.jpubeco.2020.104158

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Johnson, J. P., & Myatt, D. P. (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review*, *96*(3), 756–784. https://doi.org/10.1257/aer.96.3.756

Leeson, P. T. (2008). Media freedom, political knowledge, and participation. *Journal of Economic Perspectives*, *22*(2), 155–169. https://doi.org/10.1257/jep.22.2.155

Little, A. T. (2016). Communication technology and protest. *The Journal of Politics*, *78*(1), 152–166. https://doi.org/10.1086/683187

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2016). *Polity iv project: Political regime characteristics and transitions, 1800–2015 (version 13)* (tech. rep.). Center for Systemic Peace. https://www.systemicpeace.org/inscr/p4manualv2016.pdf

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2020). Polity v dataset [Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/inscrdata.html]. http://www.systemicpeace.org/inscrdata.html

Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. *Econometrica*, *62*(1), 157–180. https://doi.org/10.2307/2951479

Morris, S., & Shin, H. S. (1998). Unique equilibrium in a model of self-fulfilling currency attacks. *The American Economic Review*, *88*(3), 587–597. http://www.jstor.org/stable/116850

Pemstein, D., Marquardt, K. L., Tzelgov, E., Wang, Y., Medzihorsky, J., Krusell, J., Miri, F., & von Römer, J. (2022). *The v-dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data (v-dem working paper no. 21, 7th ed.)* (tech. rep.). SSRN. https://doi.org/10.2139/ssrn.3595962

Petrova, M. (2008). Inequality and media capture. *Journal of Public Economics*, *92*(1–2), 183–212. https://doi.org/10.1016/j.jpubeco.2007.04.004

Reiljan, A. (2020). 'fear and loathing across party lines' (also) in europe: Affective polarisation in european party systems. *European Journal of Political Research*, *59*(2), 376–396. https://doi.org/10.1111/1475-6765.12351

Rozenas, A. (2016). Office insecurity and electoral manipulation. *Journal of Politics*, *78*(1), 232–248. https://doi.org/10.1086/683256

Svolik, M. W. (2020). When polarization trumps civic virtue: Partisan conflict and the subversion of democracy by incumbents. *Quarterly Journal of Political Science*, *15*(1), 3–31. https://doi.org/10.1561/100.00018132

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, *26*(2), 305–321. https://doi.org/10.1287/opre.26.2.305

V-Dem Institute. (n.d.). About [Accessed November 26, 2024]. https://v-dem.net/about/

VonDoepp, P., & Young, D. J. (2013). Assaults on the fourth estate: Explaining media harassment in africa. *Journal of Politics*, *75*(1), 36–51. https://doi.org/10.1017/S0022381612000850

Wood, E. J. (2003). *Insurgent collective action and civil war in el salvador*. Cambridge University Press.

World Bank. (2021). World development indicators [Retrieved January 6, 2021]. http://data.worldbank.org/data-catalog/world-development-indicators

APPENDIX A

PROOF OF LEMMA 1

Proof.

$$\begin{aligned} A(\bar{c}) := \Pr(c_i \leq c^* | \bar{c}) &= \Pr(\bar{c} + \rho\epsilon_i < c^*) \\ &= \Pr\left(\epsilon_i < \frac{c^* - \bar{c}}{\rho}\right) \end{aligned} \tag{A.1}$$

Since $\epsilon_i \sim U[\alpha - 1, \alpha]$, we have

$$\Pr\left(\epsilon_i < \frac{c^* - \bar{c}}{\rho}\right) = \frac{c^* - \bar{c}}{\rho} - \alpha + 1. \tag{A.2}$$

That is, $A(\bar{c}) = \frac{c^* - \bar{c}}{\rho} - \alpha + 1$. Hence, $A(\bar{c})$ is decreasing in $\bar{c}$ as needed. ■

APPENDIX B

PROOF OF PROPOSITION 1

Proof. Suppose that there exists a cutoff strategy c^*. Given $\bar{c}$, the measure of protesters is given by $A(\bar{c})$ which is equal to the expression denoted in Lemma 1. Since $A(\cdot)$ is decreasing in $\bar{c}$, there exist $\overline{c_1}^*$ and $\overline{c_2}^*$ such that:

$$A(\bar{c}) \geq \overline{\theta}, \text{if } \bar{c} \leq \overline{c_1}^*$$

$$A(\bar{c}) < \overline{\theta}, \text{if } \bar{c} > \overline{c_1}^*$$

and

$$A(\bar{c}) \geq \underline{\theta}, \text{if } \bar{c} \leq \overline{c_2}^*$$

$$A(\bar{c}) < \underline{\theta}, \text{if } \bar{c} > \overline{c_2}^*$$

When the ruler is strong ($\theta = \overline{\theta}$), the protest succeeds if and only if $\bar{c} \leq \overline{c_1}^*$. If the ruler is weak ($\theta = \underline{\theta}$), the protest succeeds if and only if $\bar{c} \leq \overline{c_2}^*$.

Given the posterior belief μ, three conditions characterize the equilibrium:

(i) Indifference Condition:

$$\mu[\Pr(\bar{c} \leq \overline{c_1}^* | c_i = c^*)(b - c^*) + \Pr(\bar{c} > \overline{c_1}^* | c_i = c^*)(-c^*)] +$$
$$(1 - \mu)[\Pr(\bar{c} \leq \overline{c_2}^* | c_i = c^*)(b - c^*) + \Pr(\bar{c} > \overline{c_2}^* | c_i = c^*)(-c^*)] = 0 \quad \text{(B.1)}$$

(ii) Consistency Condition:

$$\begin{aligned} \Pr(c_i \leq c^* | \overline{c} = \overline{c_1}^*) &= \overline{\theta} \\ \Pr(c_i \leq c^* | \overline{c} = \overline{c_2}^*) &= \underline{\theta} \end{aligned} \tag{B.2}$$

(iii) Laplacian Property:

$$\begin{aligned} \Pr(c_i \leq c^* | \overline{c} = \overline{c_1}^*) &= \Pr(\overline{c} > \overline{c_1}^* | c_i = c^*) \\ \Pr(c_i \leq c^* | \overline{c} = \overline{c_2}^*) &= \Pr(\overline{c} > \overline{c_2}^* | c_i = c^*) \end{aligned} \tag{B.3}$$

Combining B.1, B.2, and B.3; we have

$$\mu[(1-\overline{\theta})(b-c^*) + \overline{\theta}(-c^*)] + (1-\mu)[(1-\underline{\theta})(b-c^*) + \underline{\theta}(-c^*)] = 0. \tag{B.4}$$

Rearranging B.4, we have $c^*(\mu) = [\mu(1-\overline{\theta}) + (1-\mu)(1-\underline{\theta})]b$ as needed. ■

APPENDIX C

PROOF OF LEMMA 2

Proof. Recall that we have $c^*(\mu) = [\mu(1-\overline{\theta}) + (1-\mu)(1-\underline{\theta})]b$. Consider the two extreme cases: $\mu = 0$, and $\mu = 1$. For these cases, c^* takes values $(1-\underline{\theta})b$, and $(1-\overline{\theta})b$ respectively. Since c^* is decreasing in μ, we have

$$c^*(\mu) \in [(1-\overline{\theta})b, (1-\underline{\theta})b]. \tag{C.1}$$

Plugging these critical values in the inequality $(\alpha + \underline{\theta} - 1)\rho < c^*(\mu) < 1 - (1 - \alpha - \underline{\theta})\rho$ yields the result. ■

APPENDIX D

PROOF OF COROLLARY 1

Proof. First, we will find the conditions under which we have a pooling and separating equilibrium. From Proposition 2, we know that there is a pooling equilibrium if

$$[1-(\mu_0(1-\overline{\theta})+(1-\mu_0)(1-\underline{\theta}))b-(1-\alpha-\underline{\theta})\rho](\gamma-\kappa)\geq$$
$$[1-(1-\underline{\theta})b-(1-\alpha-\underline{\theta})\rho]\gamma.$$

Next, we have a separating equilibrium if

$$[1-(1-\underline{\theta})b-(1-\alpha-\underline{\theta})\rho]\gamma\geq[1-(1-\overline{\theta})b-(1-\alpha-\underline{\theta})\rho](\gamma-\kappa).$$

Let $c_1^*\equiv[\mu_0(1-\overline{\theta})+(1-\mu_0)(1-\underline{\theta})]b$, $c_2^*\equiv(1-\underline{\theta})b$, $c_3^*\equiv(1-\overline{\theta})b$, and $\psi\equiv(1-\alpha-\underline{\theta})\rho$. These conditions give us the following: If we have $\frac{(c_2^*-c_1^*)\gamma}{1-c_1^*-\psi}\geq\kappa$, then we have a pooling equilibrium. Moreover, if we have $\kappa\geq\frac{(c_2^*-c_3^*)\gamma}{1-c_3^*-\psi}$, then we have a separating equilibrium. If neither of these inequalities holds, we do not have a pooling or separating equilibrium. This leads to the condition for which we have a unique perfect Bayesian equilibrium in the main text. ■

APPENDIX E

PROOF OF THEOREM 2

Proof. Assume to get a contradiction that there is some ρ, ρ' with $\rho > \rho'$ such that $\mu^*(\rho) > \mu^*(\rho')$. We will use the definition of $\mu^*(\rho)$ and the (strict) single-crossing-from-above property. Note that by definition, we have $V(\rho, \mu^*(\rho)) = 0$ and $V(\rho', \mu^*(\rho')) = 0$. By the single-crossing-from-above property in ρ, we have $V(\rho', \mu^*(\rho)) \geq 0$. Then, by using the strict single-crossing-from-above property in μ, we get $V(\rho', \mu^*(\rho')) > 0$, which is a contradiction. ■

APPENDIX F

PROOF OF PROPOSITION 3

Proof. First, we know that there exists a hybrid equilibrium; hence, there exists $\mu^*(\rho)$ such that $h(\rho, \mu^*(\rho)) = 0$. We need to show that $h(\rho, \mu)$ satisfies the single-crossing-from-above property in ρ and μ. Let us start with the first one. Assume that $h(\rho, \mu) \geq 0$, that is $\sigma(\mu) \geq \sigma^*(\mu, \rho)$. Assume to get a contradiction, $\sigma(\mu) < \sigma^*(\mu, \rho')$ for some $\rho' < \rho$.

Claim 1 *$\sigma^*(\mu, \rho)$ is increasing in ρ.*

Proof. We need to show that $U(\cdot)$ is supermodular with respect to σ and ρ. Then, Topkis's Monotonicity Theorem (Topkis, 1978) will yield the result.

$$\begin{aligned}\frac{\partial^2 U(\cdot)}{\partial\rho\partial\sigma} &= \frac{\partial}{\partial\rho}\left[V(\mu)(\gamma-\kappa) - V(0)\gamma\right] \\ &= -(1-\alpha-\underline{\theta})(\gamma-\kappa) + (1-\alpha-\underline{\theta})\gamma \\ &= (1-\alpha-\underline{\theta})\kappa > 0\end{aligned}$$

The last line holds because we assume that $\underline{\theta} < 1-\alpha$. That is, under this condition, $U(\cdot)$ is indeed supermodular with respect to σ and ρ. ■

Claim 1 suggests that $\sigma^*(\mu, \rho) \geq \sigma^*(\mu, \rho')$ for all $\rho > \rho'$. However, this contradicts that $\sigma^*(\mu, \rho') > \sigma(\mu) \geq \sigma^*(\mu, \rho)$ for some $\rho' < \rho$. This completes the first part. Next, we need to show that $h(\rho, \mu)$ satisfies the strict single-crossing-

from-above-property in μ. Again, assume that $h(\rho, \mu) \geq 0$. Assume to get a contradiction, $\sigma(\mu') \leq \sigma^*(\mu', \rho)$ for some $\mu' < \mu$.

Claim 2 *$\sigma^*(\mu, \rho)$ is increasing in μ.*

Proof. Similar to Claim 1, we need to show that $U(\cdot)$ is supermodular with respect to σ and μ. Again, Topkis's Monotonicity Theorem will yield the result.

$$\begin{aligned}\frac{\partial^2 U(\cdot)}{\partial \mu \partial \sigma} &= \frac{\partial}{\partial \mu}\left[V(\mu)(\gamma - \kappa) - V(0)\gamma\right] \\ &= b(\overline{\theta} - \underline{\theta})(\gamma - \kappa) > 0\end{aligned}$$

$U(\cdot)$ is indeed supermodular with respect to σ and μ. ■

Claim 2 suggests that $\sigma^*(\mu, \rho) \geq \sigma^*(\mu', \rho)$ for all $\mu > \mu'$. Given all the assumptions, we get $\sigma(\mu) \geq \sigma^*(\mu, \rho) \geq \sigma^*(\mu', \rho) \geq \sigma(\mu')$ for some $\mu > \mu'$. However, this contradicts that $\sigma(\mu)$ is strictly decreasing in μ. This concludes the proof. ■

APPENDIX G

PROOF OF LEMMA 3

Proof. We have two distributions to consider: $X \equiv \bar{c} \sim U[0,1]$, and $Y \equiv \epsilon_i \sim U[(\alpha-1)\rho, \alpha\rho]$. Denote the PDF of X and Y as $f_X(\cdot)$ and $f_Y(\cdot)$ respectively. We need to find the PDF of $Z \equiv c_i$ where $Z = X + Y$. From a standard result in probability theory, the PDF of Z, $f(z)$, is given by [1]

$$f(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy \tag{G.1}$$

Using the distributions of X and Y, G.1 can be written as follows:

$$f(z) = \int_{(\alpha-1)\rho}^{\alpha\rho} f_X(z-y) \frac{1}{\rho}\, dy \tag{G.2}$$

Further simplifying the expression by looking at the nonzero values of the integral, we get

$$f(z) = \left[\min\{\alpha\rho, z\} - \max\{(\alpha-1)\rho, z-1\}\right] \frac{1}{\rho}. \tag{G.3}$$

By using the assumption that $\alpha \leq 1$, and recalling that $\text{supp}(Z) = [(\alpha-1)\rho, 1+\alpha\rho]$, we get the result in the main text. ■

[1] The following integral is called the convolution of f_X and f_Y.

APPENDIX H

DISTRIBUTION OF POLITY SCORES AMONG PROTEST CASES

Table H.1: Distribution of Polity Scores Among Protest Cases

Polity	**Percent**
-10	2.85
-9	0.41
-8	0.81
-7	5.69
-6	4.07
-5	0.41
-4	3.66
-3	6.91
-2	4.88
-1	3.25
0	4.88
1	0.41
2	3.66
3	4.47
4	10.16
5	6.50
6	12.60
7	4.88
8	9.76
9	7.32
10	2.44

Note: Author's calculations using the NAVCO and Polity 5 data sets for years 2000-2018. As explained in Chapter 3, a country-year observation has $Protest = 1$ if the country experienced protests and the year is between the beginning and the peak years of protest. The table gives the distribution of Polity scores for the country-year observations that have $Protest = 1$.

APPENDIX I

NUMBER OF PROTESTS PER YEAR

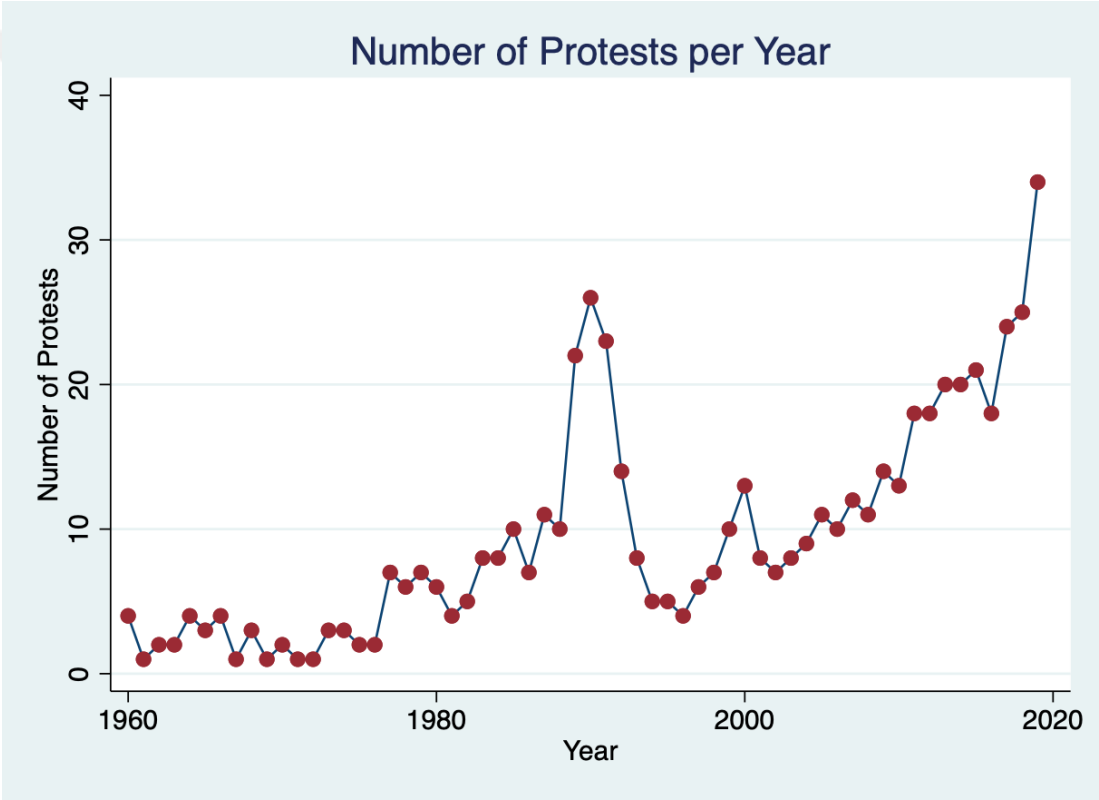

Figure 6: Number of Protests per Year (1960-2019)

Note: Author's calculation based on the NAVCO data set. The graph shows the number of non-violent regime change campaigns (protests) for each year. This is calculated as follows: For example, consider a location (country) that experienced protests that began in 2014 and peaked in 2019. We duplicated this observation six times (for each year). This means that the original observation

contributes to the total number of protests for each year, from 2014 to 2019. Note that the peak around the end of the 1980s and the beginning of the 1990s coincides with the dissolution of the USSR (this observation is also reported in Cantoni et al. (2024)).