Measuring the Tolerance of the State: Theory and Application to Protest

Veli M. Andirin, Brown University*
Yusuf Neggers, University of Michigan
Mehdi Shadmehr, University of North Carolina at Chapel Hill
Jesse M. Shapiro, Brown University and NBER

October 2021

Abstract

We develop a measure of a regime’s tolerance for an action by its citizens. We ground our measure in an economic model and apply it to the setting of political protest. In the model, a regime anticipating a protest can take a costly action to repress it. We define the regime’s tolerance as the ratio of its cost of repression to its cost of protest. Because an intolerant regime will engage in repression whenever protest is sufficiently likely, a regime’s tolerance determines the maximum equilibrium probability of protest. Tolerance can therefore be identified from the distribution of protest probabilities. We construct a novel cross-national database of protest occurrence and protest predictors, and apply machine-learning methods to estimate protest probabilities. We use the estimated protest probabilities to form a measure of tolerance at the country, country-year, and country-month levels. We apply the measure to questions of interest.

JEL Codes: C55, D74
Keywords: lasso, nonparametric identification, structural political economy, text analysis

*We acknowledge funding from the Data Science Institute, the Population Studies and Training Center, the Eastman Professorship, and the JP Morgan Chase Research Assistant Program at Brown University. We are grateful to the United Nations Group of Experts on Geographical Names for sharing their database of country names and official languages with us in a convenient form. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the funding or data sources. We thank Isaiah Andrews, Dan Björkegren, David I. Levine, and audiences at Brown University and Georgetown University for comments and suggestions. We thank our many dedicated research assistants for their contributions to this project. E-mail: veli_andirin@brown.edu, yneggers@umich.edu, mshadmeh@gmail.com, jesse_shapiro_1@brown.edu.
1 Introduction

Citizens often take actions, such as protest or criticism, that the governing regime would rather avoid. Measuring the regime’s tolerance for such actions is complicated by several factors. The *de jure* tolerance of an action, for example as enshrined in a national constitution, may not be a good guide to the *de facto* tolerance of the action.\(^1\) The frequency with which an action occurs may reflect both the extent to which the regime tolerates the action and the extent to which citizens wish to undertake it.\(^2\) The frequency with which an action is repressed may likewise reflect both how often citizens undertake the action and how the regime responds when they do.\(^3\)

In this paper we introduce a new measure of tolerance based not on the frequency of an action but on its predictability. We take political protest as our leading application. We ground our measure in an economic model. We construct our measure using a new daily, cross-national panel of protest occurrence and protest predictors. We illustrate the value of the measure by applying it to questions of interest.

In our model, a regime chooses a level of repression after observing a state of nature and a mobilization decision by an opposition, both of which can influence the probability of protest. Both repression and protest are costly to the regime. We define the regime’s *tolerance* as the ratio of its cost of guaranteeing that no protest occurs, to its cost of one occurring. Under conditions we specify, the regime’s tolerance determines an upper bound on the equilibrium probability of protest—if protest were more likely than this upper bound, the regime would repress it.

We establish further conditions under which the upper bound is attained, or at least approached, in equilibrium. Under these conditions, tolerance is identified from the distribution of equilibrium protest probabilities. The distribution of equilibrium protest probabilities is in turn identified from the joint distribution of protest and the state of nature observed by the regime. In the more realistic situation in which the econometrician observes a coarsening of the regime’s information, a lower bound on tolerance is identified. Our approach to identification is nonparametric in that it does not require knowledge of, or parametric restrictions on, primitive functions such as those governing the level of grievances or the technology of mobilization.

---

\(^1\) Article 67 of the Constitution of the Democratic People’s Republic of Korea (North Korea) states that “Citizens are guaranteed freedom of speech, the press, assembly, demonstration and association” (Constitution Project 2021).

\(^2\) In six of the eight years from 2011 through 2018, the Mass Mobilization Project records the same number (zero) of anti-regime protests in both Finland and Uzbekistan (Clark and Regan 2021).

\(^3\) Carey (2006, Figure 1) finds that semidemocracies have higher rates of both protest and repression than do autocracies. Asal et al. (2018) find that more democratic countries are more dangerous for journalists because they provide more opportunities to be in harm’s way. Munck and Verkuilen (2002, p. 16) discuss related issues.
Building on our approach to identifying tolerance, we lay out an approach to estimating it. The ingredients of the approach are data on protest occurrence and data on predictors of protest. We assemble an original daily, cross-national panel of protest occurrence and protest predictors. The core variables in the panel come from automated text parsing of security alerts obtained from Crisis24, a global risk management firm. We supplement these data with information on search query volume, news media mentions, and social media mentions, all of which can provide advance signals of protest occurrence.

We apply machine-learning methods to predict protest occurrence in these data. We use the estimated protest probabilities to construct our measure of tolerance. We use a sample-splitting approach to avoid overfitting and to facilitate statistical inference. We present Monte Carlo evidence on the performance of our measure.

We use our measure of tolerance in two applications. The first is to the study of political bias in expert ratings of freedom. As part of its annual *Freedom in the World* report (Freedom House 2021a), Freedom House uses expert input to assign numerical freedom ratings to different countries (Freedom House 2021b). Scholars have hypothesized that Freedom House’s ratings are biased toward governments that support US foreign policy positions (e.g., Steiner 2016; Bush 2017). Testing this hypothesis is difficult without a measure of tolerance that is politically unbiased; because our measure is machine-generated, we think it plausibly meets this criterion. We test for political bias in Freedom House ratings by asking whether, for a given Freedom House rating, our estimate of tolerance is lower for countries more closely aligned with US positions according to their votes in the UN. We find no evidence of the hypothesized bias.

The second application is to the role of elections in non-democracies. An existing literature studies the occurrence of political unrest surrounding elections (e.g., Tucker 2007; Harish and Little 2017). Studying the dynamics of tolerance for protest around elections is difficult without a sub-annual measure of tolerance. Using a version of our measure calculated at the country and month level, we find that both the tolerance and the incidence of protest are greater in election months, but the increase in tolerance is greater for non-democracies than for democracies.

A large literature studies methods for comparing human rights or civil liberties across countries and over time. Existing measures of *de facto* freedoms are based on expert ratings (e.g., Cingranelli, Richards, and Clay 2014; Freedom House 2021b), population surveys (e.g, Logan and Mattes 2012; Pickel, Breustedt, and Smolka 2016), or data on the occurrence of repression (e.g.,
Franklin 2008; Fariss 2014; Chilton and Versteeg 2015). Such measures are important for many reasons, including their prominent role in social science research, and in the decision-making processes of governments and international organizations.

We contribute a new approach to measuring tolerance that is grounded in an economic model and is fully automated given data inputs. We are not aware of prior work exhibiting a formal model of strategic behavior in which tolerance is identified even absent observed acts of repression. Grounding our approach in a formal model helps to make our identifying assumptions explicit. Expert ratings have been criticized in the scholarly literature for possible political bias (e.g., Mainwaring, Brinks, and Pérez-Liñan 2001; Steiner 2016; Bush 2017), lack of transparency (e.g., Munck and Verkuilen 2002, p. 21; Bradley 2015, p. 38), and failure to quantify uncertainty (e.g., Høyland, Moene, and Willumsen 2012; see also Armstrong 2011). Because our approach is reproducible given data inputs, and is amenable to statistical inference, it may avoid these drawbacks. Automation also makes it possible to compute our measure at, say, the monthly level, a finer time scale than is available for, say, Freedom House ratings. Our applications highlight some of these advantages.

Our approach also has important limitations. The formal assumptions that we require for identification are substantive, as are the assumptions we make about the input data. We discuss these issues in the paper and show some related sensitivity analyses in an appendix.

A recent literature applies modern statistical methods to predict civil unrest using data from news media, social media, and other sources (e.g., Ramakrishnan et al. 2014; Hoegh et al. 2015; Hoegh, Ferreira, and Leman 2016; Hoegh 2019; Qiao et al. 2017; Bagozzi, Chatterjee, and Makherjee 2019; Ross et al. 2019). The focus of much of this work is on the predictive task itself, whereas our work uses estimated protest probabilities as an input to learning a parameter of

---

4See the typology in Landman (2004). Some scales incorporate information on de jure freedoms including those guaranteed by constitutions (e.g., Merkel et al. 2018). For an analysis of the relationship between such guarantees and de facto freedoms, see, for example, Keith and Poe (2004) and Keith, Tate, and Poe (2009).


7Other recent work studies prediction of related outcomes such as armed conflict (e.g., Mueller and Rauh 2018).
interest defined in an economic model.\textsuperscript{8}

A large theoretical literature, reviewed for example in Gehlbach, Sonin, and Svolik (2016), studies the dynamics of protest, dissent, and repression, especially in autocracies (see also Davenport 2007; Earl 2011; Davenport et al. 2019).\textsuperscript{9} The goal of our model is to support identification of tolerance in the presence of substantial unobserved heterogeneity across environments. As a result, our model is more stylized than in much of the prior literature, with many aspects of the environment subsumed in abstract objects such as the state of nature and the mobilization technology. To our knowledge, the key qualitative implication of our model—that protest is less predictable in less tolerant regimes—is novel.\textsuperscript{10} We are not aware of prior evidence on this prediction.\textsuperscript{11}

The rest of the paper proceeds as follows. Section 2 presents the model, characterizes its equilibrium, and lays out our approach to identification. Section 3 lays out our approach to estimation and inference. Section 4 describes our data, implementation, and some validation exercises. Section 5 presents our results and applications. Section 6 concludes.

\section{Model of Protest and Repression}

\subsection{Setup and Definitions}

There is a set of environments (say, countries) indexed by $i$. Time $t$ is discrete. In each environment $i$, nature determines a state $\omega_{it} \in [0, \bar{\omega}_i]$ in each period $t$ from a time-invariant distribution with $\bar{\omega}_i > 0$. After observing the state $\omega_{it}$, the opposition decides on a mobilization effort $m_{it} \in [0, \bar{m}_i]$. After observing the state $\omega_{it}$ and the mobilization effort $m_{it}$, the regime chooses a level of repression $r_{it} \in [0, \bar{r}_i]$ with $\bar{r}_i > 0$. A protest then occurs with probability $\lambda_i(\omega_{it}, m_{it}, r_{it})$ where $\lambda_i(\cdot)$ is a function increasing in its first two arguments and decreasing in its last.

We impose the following additional structure on the function $\lambda_i(\cdot)$.

\textbf{Assumption 1.} In each environment $i$, the function $\lambda_i(\cdot)$ satisfies the following conditions:

\textsuperscript{8}More broadly, our work relates to recent literature applying innovations in machine learning (Varian 2014; Belloni, Chernozhukow, and Hansen 2014; Athey 2015; Kleinberg et al. 2015; Shapiro 2017; Mullainathan and Spiess 2017) and in the measurement of digital activity (Einav and Levin 2014) to problems in social science.

\textsuperscript{9}Because we model repression as an action by the regime that reduces the \textit{ex ante} likelihood of protest, our work is particularly related to models of preemptive repression (e.g., De Jaegher and Hoyer 2019).

\textsuperscript{10}Langørgen (2016) argues that organized and spontaneous protests are likely to have different causal structures. Kuran (1991) studies the predictability of revolution.

\textsuperscript{11}For past work on the empirical dynamics of protest, dissent, and repression, see, for example, Moore (1998), Carey (2006, 2009), and Ritter and Conrad (2016).
(a) \( \lambda_i(\omega, m, \bar{r}_i) = 0 \) for all \( \omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i] \).

(b) \( \lambda_i(\omega, m, r) \) is concave in \( r \) for all \( \omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i] \).

(c) \( \lambda_i(\omega, m, 0) \) is continuous in \( m \) for all \( \omega \in [0, \bar{\omega}_i] \).

The conditions of Assumption 1 are satisfied, for example, by the function
\[
\lambda_i(\omega, m, r) = \frac{\bar{r}_i - r}{\bar{r}_i} \frac{\omega (m + k)}{(\bar{\omega}_i - \omega) + \omega (m + k)}
\]
where \( k \) is a strictly positive constant.

The regime’s and opposition’s payoffs in period \( t \) are, respectively,
\[
\pi^r_{it} = -L_i z_{it} - r_{it}
\]
\[
\pi^m_{it} = B_i z_{it} - m_{it}
\]
where \( L_i, B_i \geq 0 \) are nonnegative scalars and \( z_{it} \in \{0, 1\} \) is an indicator for whether protest occurs in period \( t \). Payoffs for the regime and opposition are each discounted by some discount factor strictly below 1, possibly differing between the regime and opposition. If the regime is indifferent among two or more levels of repression, it chooses the lowest of these.

If the regime represses fully, choosing \( r = \bar{r}_i \), then under Assumption 1(a) no protest occurs, so \( z_{it} = 0 \), and from (2) the regime’s payoff is \( -\bar{r}_i \). If the regime does not repress at all, choosing \( r = 0 \), and protest does occur, so \( z_{it} = 1 \), then from (2) the regime’s payoff is \(-L_i\). Thus the ratio of the cost of full repression to the cost of protest is \( \bar{r}_i/L_i \). If this ratio exceeds one, then the regime prefers to allow a protest to proceed with certainty (yielding payoff \(-L_i\)) rather than to repress it fully (yielding \(-\bar{r}_i\)). These observations motivate the following definition.

**Definition 1.** The tolerance \( \tau_i \) of the regime in environment \( i \) is given by
\[
\tau_i = \min \left\{ \frac{\bar{r}_i}{L_i}, 1 \right\}.
\]

The goal of our analysis is to establish conditions for the identification of \( \tau_i \).

---

\(^{12}\)If \( L_i = 0 \) we may define this ratio as infinity.
2.2 Solution Concept

The history $H_t$ at time $t$ is the sequence $\{\omega_{it'}, m_{it'}, r_{it'}\}_{i'=1}^{t-1}$. This is a member of the set $\mathcal{H}_i$ of all possible histories at all possible time periods. A pure strategy $\sigma_m : [0, \bar{\omega}_i] \times \mathcal{H}_i \to [0, \bar{m}_i]$ for the opposition prescribes an action for each state and history. A pure strategy $\sigma_r : [0, \bar{\omega}_i] \times [0, \bar{m}_i] \times \mathcal{H}_i \to [0, \bar{r}_i]$ for the regime prescribes an action for each state, action by the opposition, and history. A pair of pure strategies $(\sigma_m, \sigma_r)$ is stationary if

\[ \sigma_m(\omega, H') = \sigma_m(\omega, H'') \]

for all $H', H'' \in \mathcal{H}_i$ and any $\omega \in [0, \bar{\omega}_i]$, and

\[ \sigma_r(\omega, m, H') = \sigma_r(\omega, m, H'') \]

for all $H', H'' \in \mathcal{H}_i$ and any $\omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i]$. For simplicity we will write the prescriptions of stationary pure strategies as $\sigma_m(\omega)$ and $\sigma_r(\omega, m)$.

**Definition 2.** A pair $(\sigma_m, \sigma_r)$ of stationary pure strategies is an equilibrium of the game in environment $i$ if

\[ \sigma_m(\omega) \in \arg \max_m (B_i \lambda_i(\omega, m, \sigma_r(\omega, m)) - m) \] (3)

for all $\omega \in [0, \bar{\omega}_i]$ and

\[ \sigma_r(\omega, m) \in \min \left\{ \arg \max_r (-L_i \lambda_i(\omega, m, r) - r) \right\} \] (4)

for all $\omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i]$.

The use of the minimum in (4) reflects our assumption that ties are broken in favor of lower repression.

2.3 Characterization of Equilibrium

In environment $i$, given equilibrium strategies $(\sigma^*_m, \sigma^*_r)$ the equilibrium probability of protest $\lambda^*_i(\cdot)$ is given by

\[ \lambda^*_i(\omega) = \lambda_i(\omega, \sigma^*_m(\omega), \sigma^*_r(\omega, \sigma^*_m(\omega))) . \]

**Proposition 1.**

(i) Under Assumption [a], in any equilibrium the probability of protest $\lambda^*_i(\cdot)$ satisfies $\lambda^*_i(\omega) \leq \tau_i$ for all $\omega \in [0, \bar{\omega}_i]$.

(ii) Under Assumptions [a]-[c] there exists an equilibrium. In any equilibrium, $\lambda^*_i(\omega) = 0$ whenever $\lambda_i(\omega, 0, 0) > \tau_i$, and $\lambda^*_i(\omega) \in [\lambda_i(\omega, 0, 0), \tau_i]$ otherwise.
An appendix following the main text provides a proof of Proposition 1 and other claims. Here we provide an intuition for Proposition 1.

Begin with part (i) of Proposition 1. If under some strategies \((\sigma_m, \sigma_r)\) the probability of protest exceeds \(\tau_i\) at some state \(\omega\), then from Assumption 1(a), the regime’s payoff in (2), and the definition of \(\tau_i\), it follows that the regime prefers to take \(r = \tau_i\) at state \(\omega\), and hence that \((\sigma_m, \sigma_r)\) are not an equilibrium.

Turn next to part (ii). By Assumption 1(b) and the regime’s payoff in (2), the regime’s expected payoff is convex in \(r\) and therefore the regime chooses either no repression, \(r = 0\), or full repression, \(r = \tau_i\). If \(\lambda_i(\omega, 0, 0) > \tau_i\), then by Assumption 1(a) and the regime’s payoff in (2), the regime will choose full repression regardless of the opposition’s action, and therefore the opposition chooses not to mobilize, \(m = 0\), and \(\lambda^*_i(\omega) = \lambda_i(\omega, 0, \tau_i) = 0\). By contrast, if \(\lambda_i(\omega, 0, 0) \leq \tau_i\), then the regime will choose no repression unless the opposition mobilizes sufficiently to trigger it, and therefore the opposition will choose \(m\) small enough to avoid triggering repression. Therefore \(\lambda^*_i(\omega) = \lambda_i(\omega, \sigma_m^*(\omega), 0) \in [\lambda_i(\omega, 0, 0), \tau_i]\).

Note that these arguments, and the proof of Proposition 1, do not rely on the assumption that the distribution of \(\omega_{it}\) is time-invariant. However, absent that assumption, our focus on stationary strategies seems less natural.

### 2.4 Identification of Tolerance

Motivated by Proposition 1(i), our approach to identification exploits the fact that the equilibrium probability of protest cannot exceed \(\tau_i\).

**Definition 3.** The maximum tolerated protest probability \(\lambda^*_i(\cdot)\) in environment \(i\) with equilibrium protest probability \(\lambda^*_i(\cdot)\) is given by

\[
\lambda^*_i = \inf \{ \lambda \in [0, 1] : \Pr(\lambda^*_i(\omega_{it}) \leq \lambda) = 1 \}.
\]

In words, the maximum tolerated protest probability is the largest probability \(\lambda^*_i(\omega_{it})\) that occurs in the given equilibrium.

It is immediate from Proposition 1(i) and the definition of \(\lambda^*_i\) that \(\lambda^*_i \leq \tau_i\), and therefore that \(\tau_i\) is partially identified from the distribution \(\Pr(\lambda^*_i(\omega_{it}) \leq \lambda)\) of \(\lambda^*_i(\omega_{it})\).

**Corollary 1.** Under Assumption 1(a) in any equilibrium, the tolerance \(\tau_i\) is partially identified from the distribution of \(\lambda^*_i(\omega_{it})\). In particular, \(\tau_i \in [\lambda^*_i, 1]\).
If we further impose Assumptions 1(b) and 1(c), then by Proposition 1(ii),
\[ \lambda^*_i(\omega) \in [\lambda_i(\omega, 0, 0), \tau_i] \]
whenever \( \lambda_i(\omega, 0, 0) \leq \tau_i \). It follows that if \( \lambda_i(\omega_i, 0, 0) \) has sufficiently rich support, then \( \lambda^*_i = \tau_i \), and so \( \tau_i \) is point identified.

**Assumption 2.** The random variable \( \lambda_i(\omega_i, 0, 0) \) has full support on \([0, 1]\).

**Proposition 2.** Under Assumptions 1 and 2, in any equilibrium, the tolerance \( \tau_i \) is identified from the distribution of \( \lambda^*_i(\omega_i) \). In particular, \( \tau_i = \lambda^*_i \).

Under Assumption 1, the conclusion of Proposition 2 holds under weaker or different conditions than Assumption 2. For example, it is sufficient that the support of \( \lambda_i(\omega_i, 0, 0) \) includes a neighborhood \((\tau_i - \varepsilon, \tau_i)\) of \( \tau_i \) with \( \varepsilon \in (0, \tau_i) \) (a weaker condition than Assumption 2), or that \( \Pr(\lambda_i(\omega_i, 0, 0) = \tau_i) > 0 \) (a condition neither weaker nor stronger than Assumption 2).

**Corollary 1 and Proposition 2** rely on knowledge of the marginal distribution of \( \lambda^*_i(\omega_i) \). The marginal distribution is of \( \lambda^*_i(\omega_i) \) is, in turn, identified from the joint distribution of the indicator \( z_{it} \) for whether protest occurs on a given date, and the state of nature \( \omega_i \), because \( \lambda^*_i(\omega_i) = \Pr(z_{it} = 1|\omega_i) \).

Rather than observing the state of nature \( \omega_i \) directly, the econometrician might instead observe a transformation of it. In this case, the maximum protest probability observed by the econometrician is weakly below the maximum tolerated protest probability.

**Claim 1.** For any function \( \chi_i(\cdot) \), we have that in any equilibrium
\[ \inf \{ \lambda \in [0, 1] : \Pr(\Pr(z_{it} = 1|\chi_i(\omega_i)) \leq \lambda) = 1 \} \leq \lambda^*_i, \]
with equality when \( \chi_i(\cdot) \) is one-to-one.

Claim 1 covers, for example, the case where the econometrician observes mobilization effort \( \sigma^*_m(\omega_i) \) rather than the state of nature \( \omega_i \).

Claim 1 implies that if \( \chi_i(\cdot) \) is one-to-one (e.g., strictly increasing), then the conclusions regarding the identification of \( \tau_i \) from the distribution of \( \Pr(z_{it} = 1|\chi_i(\omega_i)) \) are parallel to those above for the identification of \( \tau_i \) from the distribution of \( \lambda^*_i(\omega_i) = \Pr(z_{it} = 1|\omega_i) \), even if \( \chi_i(\cdot) \) is unknown and differs across environments. If instead \( \chi_i(\cdot) \) is many-to-one (i.e., a coarsening), then only a lower bound on \( \tau_i \) can generally be identified from the distribution of \( \Pr(z_{it} = 1|\chi_i(\omega_i)) \), even if Assumptions 1 and 2 hold.
2.5 Discussion

We define the tolerance $\tau_i$ as a function of the ratio of the regime’s cost $r_i$ of full repression to the regime’s cost $L_i$ of allowing a protest. This means that our definition of tolerance makes no distinction between a regime that has little desire to prevent protest (i.e., low $L_i$) and one that has little ability to do so (i.e., high $r_i$).

A related point is that, while Assumption [a] requires that the regime be able, in principle, to prevent protest with certainty, we do not require that the cost $r_i$ of doing so is low enough to make full repression appealing. By allowing for an arbitrarily large $r_i$, our framework can therefore accommodate situations in which full repression is arbitrarily difficult or costly.

The restriction in Assumption [b] seems to us more substantive. In tandem with the assumption in (2) that the regime’s cost of repression is linear in $r_i$, Assumption [b] means the regime will either choose no repression or full repression, and therefore that the regime’s optimal decision depends on $\omega$ and $m$ only through $\lambda_i(\omega, m, 0)$. Without Assumption [b] the regime’s optimal decision can depend directly on $\omega$ and $m$. In that case, following Proposition (1) it remains true that the equilibrium probability of protest cannot exceed $\tau_i$, but the conditions we consider may not suffice to ensure that this bound is actually attained.

Our approach to identification of tolerance $\tau_i$ requires that the econometrician be able to measure the ex ante probability of protest in each period, exactly as the regime would. In the more realistic situation in which the regime has information that the econometrician does not, our approach permits identification of a lower bound on $\tau_i$ via Claim 1.

Importantly, our approach does not require the econometrician to know the functions $\lambda_i(\cdot)$ or $\lambda_i^*(\cdot)$, the parameters $\{L_i, B_i\}$ determining the regime and opposition’s payoffs, or the parameters $\{\bar{r}_i, \bar{m}_i\}$ determining their action spaces. Because, following Claim 1, our approach requires the econometrician to observe the state of nature only up to a one-to-one transformation that may differ across environments, our approach also does not require the econometrician to know the parameter $\bar{\omega}_i$ determining the domain of the state of nature. In this sense, our approach does not require the econometrician to be able to compare the level of grievances, the nature and organization of the opposition, or the technology of mobilization across environments. Our approach also does not require the econometrician to directly observe acts of repression, which in some cases may be clandestine (e.g., arresting opposition figures).

---

13 That is, the regime’s optimal decision is identical for any $\omega, \omega' \in [0, \bar{\omega}_i]$ and $m, m' \in [0, \bar{m}_i]$ such that $\lambda_i(\omega, m, 0) = \lambda_i(\omega', m', 0)$. 
Our approach does require (via Assumption 2) that arbitrarily high protest probabilities would be observed absent mobilization and repression. Substantively, this restriction may be thought of as saying that in any environment there will be some grievances sufficient to motivate predictable protests.\footnote{Many classical accounts of social conflict posit that grievances are pervasive (see, e.g., Jenkins and Perrow 1977, p. 251; Oberschall 1978, p. 298), and that they manifest as protest when the prevailing political structures do not prevent it (e.g., McAdam 1982, Chapter 3).} If this assumption fails, then following Corollary 1, our approach permits identification of a lower bound on $\tau_i$.

3 Estimation and Inference

We now discuss our approach to estimation and inference. Our goal is to learn $\lambda^*_i$ in each environment $i$. In a given environment, $\lambda^*_i$ is identified from the distribution of equilibrium protest probabilities $\lambda^*_i(\omega_t)$, which in turn can be learned from the joint distribution of protest occurrence $z_{it}$ and the state of nature $\omega_t$. In practice, we observe an indicator for protest occurrence $z_{it}$ and a vector $x_{it}$ of predictors in each period $t \in \{1, \ldots, T\}$ and in each environment $i \in \{1, \ldots, N\}$. The sample data are then $\left\{\left\{z_{it}, x_{it}\right\}_{t=1}^T\right\}_{i=1}^N$.

3.1 Procedure

We proceed in the following main steps.

Sample splitting. We partition the periods into cells indexed by $g \in \{1, \ldots, G\}$. For each environment $i$, we further partition the cells into two groups $G^1(i)$ and $G^2(i)$, with corresponding sets of periods $T^1(i)$ and $T^2(i)$. These partitions do not depend on the data.

Predictive model. Using the data $\left\{\left\{z_{it}, x_{it}\right\}_{t \in T^1(i)}\right\}_{i=1}^N$ and $\left\{\left\{z_{it}, x_{it}\right\}_{t \in T^2(i)}\right\}_{i=1}^N$ from the first and second set of periods, respectively, we fit predictive models that yields estimate $\hat{\lambda}^1_i(\cdot)$ and $\hat{\lambda}^2_i(\cdot)$ of the function $\Pr(z_{it} = 1|x_{it} = x)$.

Protest probabilities. Using the predictive models, we compute estimates $\hat{\lambda}^*_i$ of the equilibrium protest probabilities, where $\hat{\lambda}^*_i = \hat{\lambda}^1_i(x_{it})$ for $t \in T^2(i)$ and $\hat{\lambda}^*_i = \hat{\lambda}^2_i(x_{it})$ for $t \in T^1(i)$. That is, we apply the predictive model estimated on the first set of periods to the predictors in the second set of periods, and vice versa.

Estimation. We define, for each environment $i$ and cell $g$, the period $s(g, i) = \arg\max_{t \in g} \hat{\lambda}^*_it$, that has the largest estimated equilibrium protest probability $\hat{\lambda}^*_it$, breaking ties arbitrarily. We then
compute, for each environment $i$, the statistic

$$z_i = \frac{1}{G} \sum_{g=1}^{G} z_{i,s(g,i)},$$

which gives the share of cells $g$ in which a protest occurs in the period $s(g,i)$ with the highest estimated equilibrium probability of protest. Online Appendix Figure 3 presents results from a variant of $z_i$ computed as the share of cells $g$ in which protest occurs within $D > 0$ periods of the period $s(g,i)$. Online Appendix Figure 3 also presents results from excluding the small proportion of cells in which $\arg\max_{t \in g} \lambda^*_{it}$ is not a singleton.

**Inference.** Treating $z_i$ as our point estimate, we perform inference on the random parameter

$$z^*_i = \Pr(z_{i,s(g,i)} = 1),$$

where the probability is implicitly taken with respect to a population of cells $g$. The parameter $z^*_i$ is random because each period $s(g,i)$ is chosen based on the estimated predictive model, which in turn depends on the realization of the data in the group $G_1(i)$ or $G_2(i)$ of which $g$ is not a member. Because the period $s(g,i)$ for each cell $g$ is selected without using the data $\{z_{it}\}_{t \in g}$ on protest for the given cell, we treat the problem of inference on $z^*_i$ from $z_i$ as a standard problem of inference on an unknown proportion. See, for example, Andrews, Kitagawa, and McCloskey (2021) for discussion of related issues.

### 3.2 Discussion

The logic of the procedure in Section 3.1 is as follows. Fix an environment $i$ and an equilibrium, and suppose that $x_{it} = \chi_i(\omega_{it})$ for $\chi_i(\cdot)$ a possibly unknown function. If $s(g,i)$ is an approximate maximizer of $\Pr(z_{it} = 1|x_{it})$ on $t \in g$, then $z^*_i = \Pr(z_{i,s(g,i)} = 1) \approx \mathbb{E}(\max_{t \in g} \Pr(z_{it} = 1|x_{it}))$, where the expectation is taken with respect to a population of cells $g$. If the econometrician has the same information as the regime, i.e., $\chi_i(\cdot)$ is one-to-one, then $\mathbb{E}(\max_{t \in g} \Pr(z_{it} = 1|x_{it})) = \mathbb{E}(\max_{t \in g} \lambda^*_i(\omega_{it}))$ following reasoning analogous to that underlying Claim 1. If the cells $g$ are large enough to contain realizations of $\lambda^*_i(\omega_{it})$ near the maximum, then $\mathbb{E}(\max_{t \in g} \lambda^*_i(\omega_{it})) \approx \overline{\lambda}^*_i$.

Assembling these arguments yields:

$$z^*_i = \Pr(z_{i,s(g,i)} = 1) \approx \mathbb{E}\left(\max_{t \in g} \Pr(z_{it} = 1|x_{it})\right) = \mathbb{E}\left(\max_{t \in g} \lambda^*_i(\omega_{it})\right) \approx \overline{\lambda}^*_i.$$
Thus, under these idealized conditions, the parameter $z_i^*$ approximates $\lambda_i^*$.

In practice, because of estimation error in $\hat{\lambda}_i^*$, in general $s(g, i)$ will not maximize $Pr(z_{it} = 1 | x_{it})$ over $t \in g$. Moreover, because the econometrician is likely to have less information than the regime, say because $\chi_i(\cdot)$ is many-to-one, we should expect that $E(\max_{t \in g} Pr(z_{it} = 1 | x_{it})) < E(\max_{t \in g} \lambda_i^* (\omega_{it}))$ following reasoning analogous to that underlying Claim 1. Lastly, in order to perform inference on $z_i^*$, it is desirable to have a large number of cells $G$, which is in tension with ensuring that each cell $g$ contains a large number of periods. For all of these reasons, we should expect that $z_i^* < \lambda_i^*$.

4 Data and Implementation

Here we describe our main data sources and variable definitions. Additional details, and information on additional data sources used in sensitivity analysis, are in Online Appendix Section A.

4.1 Country Characteristics

We define a universe of 249 countries based on the the ISO 3166 (International Organization for Standardization 2021). From the United Nations Group of Experts on Geographical Names (2019), we obtain the official name and official languages of 195 countries. From the World Bank World Development Indicators (2021), we obtain, for each of the 264 covered countries and economies, population data for 2010.\textsuperscript{15}

From the Freedom in the World (FITW; Freedom House 2021c) report we obtain, for each of the 195 covered countries, a measure of freedom of assembly (category E1), which ranges from 0 to 4, in each year from 2012 (its first year of availability) through 2020. From the Polity Project (Center for Systemic Peace 2021), we obtain, for each of the 167 covered countries, a classification into democracy and non-democracy (defined as autocracy or anocracy) for each available year from 2010 through 2018.

From the United Nations General Assembly Voting Data (Bailey, Strezhnev, and Voeten 2017; Voeten, Strezhnev, and Bailey 2021) we obtain, for each of 192 countries and each year from 2010 through 2020, a measure of the agreement between the country’s votes and those of the US in the session beginning in the given year.

\textsuperscript{15}For Taiwan, we obtain population data from National Statistics, Republic of China (2010).
From ElectionGuide (International Foundation for Electoral Systems 2021), we obtain, for each of 217 countries, the dates of national elections occurring from 1998 through 2021.

4.2 Protest Events

Our main source of information on the occurrence of protest is a database of security alerts obtained from Crisis24 (2020), a global risk management firm. Crisis24 staff produce these alerts to keep clients appraised of current and anticipated events, including those that may impact travel. The alerts occurred in 209 countries between the spring of 2009 and the fall of 2020. Each alert corresponds to an event or set of events that has happened, is happening, or is expected to happen. For each alert, the database includes a short title, longer description, and fields indicating the date and time of the alert.

We parse the text fields using a set of rules described in more detail in Online Appendix Section A.1. From the text fields we identify whether the alert relates to a protest, and, when possible, the dates of the protest. When we are not able to identify the dates of the protest in this way, we attempt to infer from the text fields whether the protest occurred in the past, present, or future relative to the date of the alert, and we impute the date to be either the date preceding the alert (past protests), date of the alert (present protests), or date after the alert (future protests). Online Appendix Figure 3 shows the sensitivity of our main results to excluding protests about whose date we are uncertain.

We define an indicator $z_{lt}$ equal to one on any date where an alert indicates the occurrence of a protest, and zero otherwise. We define an indicator $a_{lt}$ equal to one on any date on which an alert posted at least one day earlier indicates the occurrence of a protest, and zero otherwise.

We supplement these data with data from two other databases of protest activity. From the Mass Mobilization Data Project (MMDP; Clark and Regan 2021) we obtain information on the start and end date of citizen demonstrations against the government in 162 countries from 1990 through 2020, and we create an indicator equal to one on any country-date on which a protest begins. From the Integrated Crisis Early Warning System (ICEWS; Boschee et al. 2015), we obtain information on the number of protest-related stories appearing in the news about a given country for 227 countries from 1995 through 2020, and we create an indicator equal to one on any date on which there is a protest-related story but no protest-related story in the preceding five days.

The MMDP data are hand-curated and sourced from a small number of major international news sources via Lexis-Nexis. The ICEWS data are machine-generated and sourced from hundreds
of news sources. Online Appendix Figure 4 shows event-study plots of the relationship between the occurrence of protest in MMDP (Panel A) and ICEWS (Panel B) and our main indicator $z_{it}$, which is based on Crisis24 alerts. In both cases, the alternative source indicates that a protest is especially likely on a date in which our main indicator indicates a protest.

Our main protest indicator $z_{it}$ may miss some protests that occur. In Online Appendix Figure 3 we show the sensitivity of our findings to including as protest dates any country-dates flagged in MMDP and ICEWS as protest dates. Our main indicator $z_{it}$ may also include some protests that do not occur; this is especially likely for protests for which the alert appears in advance of the event. In Online Appendix Figure 5, we also show the sensitivity of our findings to excluding as protest dates any country-dates for which MMDP does not indicate a beginning or ongoing protest in the same calendar week, and ICEWS does not indicate any protest-related stories in the same calendar week.

4.3 Predictors of Protest

4.3.1 Search Query Volume: Google Trends

From Google Health Trends (Google 2021), we obtain daily data on searches about political demonstrations (entity /m/0gnwz4) in 249 countries from January 2004 through December 2019. We also obtain daily data on searches about each of these countries in both the US and the UK. The search data are reported as a (known) multiple of the probability that a given user session includes a search for the given topic, with a value of zero when the query does not meet reporting standards (Google 2019; see also Zepecki et al. 2020).

4.3.2 News Mentions: The GDELT 1.0 Event Database

From the GDELT 1.0 Event Database (GDELT Project 2020) we obtain daily data on the number of mentions of protest (event code 14), as a share of all mentions of a given country, for each of 274 countries from January 1979 through December 2019. These data are in turn sourced from various international news sources and wire services (Leetaru and Schrodt 2013).

These probabilities are calculated on a random sample of searches. The random sample is redrawn daily. We compute an average of valid returned values across runs executed on at least 15 different days, and therefore corresponding to at least 15 different random samples.
4.3.3 Social Media Mentions: Twitter Data

Using Twitter’s advanced search functionality (Twitter 2020), we obtain daily data on the number of non-withheld tweets containing both the name of the country and either the keyword “protest” or the keyword “demonstration” for each of 195 countries from January 2006 through December 2019. Our list of countries and official languages is based on UNGEGN’s (United Nations Group of Experts on Geographical Names) list of country names (UNGEGN 2019). We perform the search both in English and in the given country’s official languages and obtain separate counts for each of these.17

4.4 Estimation Sample and Variable Standardization

Our sample period is 2010-2019. Our sample countries are the 151 countries with a population of at least 1,000,000 in 2010 and in which there is at least one year in the sample period with at least 10 alerts in the Crisis24 database.

We standardize the predictors described in Section 4.3 by subtracting the mean and dividing by the standard deviation, both calculated over the preceding 90 days, excluding protest days. If the rolling-window standard deviation is zero, or if data are missing for the given variable, country, and date, we set the standardized value to zero. We define a vector $q_{it}$ consisting of indicators for days of the week as well as the seven lags of the standardized values of demonstration search query volume, protest news mentions, and protest Twitter mentions in English. Online Appendix Figure 3 shows the sensitivity of our findings to further including search query volume about the country in the US and UK, protest Twitter mentions in the country’s official languages, a set of financial indices described in Online Appendix A.2, an indicator for whether an election occurred within seven days of the given date in the given country, and an indicator for whether a protest occurred in the given country on the same date in the previous year.

As a prelude to our analysis, Figure 1 shows event-study plots of the standardized value of each of our key predictor variables in the days surrounding alerts about future protests (left column of plots) and the occurrence of protests (right column of plots). The estimated event-time paths of our predictor variables exhibit mild dynamics around an alert about future protest and substantial dynamics around a protest occurrence.

17In some cases we use short versions of country names rather than official ones. For example, we search for “Venezuela” rather than “Bolivarian Republic of Venezuela.”
4.5 Implementation

For our baseline implementation of the procedure in Section 3.1, we let environments $i$ be countries, periods $t$ be calendar dates, and cells $g \in \{1, ..., G\}$ be calendar months. We let the group $G_1(i)$ consist of months in odd-numbered years and the group $G_2(i)$ consist of months in even-numbered years. Online Appendix Figure 3 shows results where we instead let cells $g$ consist of two-month groups, and results where we assign years randomly to the groups $G_1(i)$ and $G_2(i)$.

Our predictive model is a binary logit model, with the protest indicator $z_{it}$ as the dependent variable and the predictors $x_{it} = (1, a_{it}, q_{it})$ as the independent variables. We estimate the model via $L_1$-penalized maximum likelihood with the penalty parameter chosen via 10-fold cross-validation where the folds are equal-sized groups of country-years. Online Appendix Figure 3 shows results based on unpenalized maximum likelihood estimation.

Figure 2 shows simulation evidence on the performance of our estimator. To produce the figure, we randomly generate an indicator of protest occurrence as a sequence of independent Bernoulli draws, with success probabilities given by the estimated equilibrium protest probabilities $\hat{\lambda}_{it}^*$. We then re-implement the entire procedure in Section 3.1 using the simulated protest occurrence indicator in place of the observed indicator. We calculate the percentile of the estimated tolerance in the distribution of equilibrium protest probabilities, and plot statistics of this percentile across simulation replications. The figure shows that, for most countries, the estimated tolerance is centered on a high percentile of the distribution of equilibrium protest probabilities.

We compute variants of the estimator $z_i$ in which an environment $i$ is a country-year or country-month. To compute these variants, we use the same estimated equilibrium protest probabilities $\hat{\lambda}_{it}^*$ as in our baseline implementation, and average the values of $z_{i,s(g,i)}$ over the cells $g$ that are contained in the given country-year or country-month.

5 Results

Panel A of Figure 3 shows the relationship, across country-years, between the estimated tolerance and FITW’s five-category freedom of assembly rating. Countries that are rated as more free tend to have a higher estimated tolerance.

Panel B of Figure 3 shows the relationship between the share of protests that are anticipated by Crisis24 and the freedom of assembly rating. Countries that are rated as more free tend to have a higher share of anticipated protests. Because tolerance is measured based on the predictability of
protest, the pattern in Panel B provides an intuition for the one in Panel A.

Panel C of Figure 3 shows the relationship between the frequency of protest and the freedom of assembly rating. There is some evidence that countries rated as more free have more frequent protest, but the distribution of protest frequency among those rated as most free is similar to that among those rated in two of the three least free categories. This pattern suggests some of the difficulties of measuring tolerance based on frequency of occurrence.

Figure 4 shows a map of the world with countries shaded according to their estimated tolerance. The map shows some qualitative patterns that we find intuitive but also some that we do not. Notably, several countries, such as Japan, New Zealand, and Norway, exhibit estimated tolerance much lower than what we suspect is their true tolerance. Preliminary inspection of the data suggests that one reason for this may be that when protest is rare the conditions for good performance of the estimator are more demanding. Online Appendix Figure 5 shows the estimated tolerance for each country, along with an associated confidence interval.

5.1 Application: Testing for Political Bias in Freedom Ratings

Let \( f_i \) be some observed rating of the freedom of country \( i \) such as one reported in FITW. Let \( y_i \) be some observed and possibly vector-valued characteristic, such as the political alignment of the country with the US, that may influence the rating \( f_i \). Suppose that

\[
f_i = F \left( \tau_i + y_i \beta \right)
\]

where \( F \) is an unknown and strictly increasing function and \( \beta \) is an unknown parameter. We can think of the case where \( \beta = 0 \) as one in which the rating \( f_i \) is not biased in the sense that it is fully determined by the country’s true tolerance.

It is possible to use the estimate \( \bar{z}_i \) to learn about the unknown parameter \( \beta \). Suppose in particular that

\[
\bar{z}_i = \tau_i + \zeta_i
\]

for \( \zeta_i \) an estimation error satisfying \( \mathbb{E}(\zeta_i|\tau_i, y_i) = \mathbb{E}(\zeta_i|f_i, y_i) = \mathbb{E}(\zeta_i) \), where \( \mathbb{E}(\zeta_i) \) need not be zero. Then by applying \( F^{-1} \) to both sides of (5) and substituting into (6) we arrive at

\[
\bar{z}_i = F^{-1}(f_i) - y_i \beta + \zeta_i.
\]
We can thus estimate $\beta$ by a regression of $z_i$ on $y_i$, including a constant and controlling flexibly for $f_i$. Intuitively, if a characteristic $y_i$ is associated with lower estimated tolerance $z_i$ for given freedom rating $f_i$, this is an indication that the freedom rating $f_i$ puts positive weight $\beta > 0$ on that characteristic.

To implement (7), we let $f_i$ be country $i$’s average FITW freedom of assembly score over the sample period. We let $y_i$ be a vector of indicators for quintiles of the average annual fraction of UN General Assembly votes in which country $i$ votes in agreement with the US. And we approximate $F^{-1}(f_i)$ with quintiles in $f_i$.

Figure 5 presents our findings. The plot shows the estimated coefficients $\beta$ for the vector $y_i$ of indicators for quintiles of agreement with the US. The estimates show no evidence of bias in favor of US allies. The estimates are jointly statistically insignificantly different from zero. If anything, the positive point estimate on the coefficient for the top quintile of agreement suggests that the FITW rating penalizes the ratings of the countries that agree most strongly with the US.

5.2 Application: Elections and the Tolerance of Protest

We can think of $z_{i,s(g,t)}$ as a (coarse) measure of tolerance in country $i$ and month $g$. Measuring tolerance at the monthly level makes it possible to study how tolerance varies at monthly frequencies. To illustrate this possibility, Figure 6 shows estimates of the tolerance and incidence of protest in the months before, of, and after an election, separately for countries classified by the Polity Project as non-democracies (Panel A) and for those classified as democracies (Panel B). The figure shows that both the tolerance and incidence of protest are greater in election months, with a larger increase in non-democracies than in democracies. We can reject the hypothesis that the time path of tolerance around elections is identical between democracies and non-democracies at conventional levels of statistical significance ($p = 0.0007$). Online Appendix Figure 6 shows estimates of the time path of tolerance after controlling for incidence.

The fact that tolerance of protest increases in election months, especially in non-democracies, may suggest several interpretations. For example, elections may bring international scrutiny (Hyde 2007) that raises the regime’s cost of repression relative to its cost of protest, or elections may be periods of regime weakness in which repression is difficult to carry out. The fact that incidence of protest increases in election months may reflect greater tolerance of protest, or it may reflect the opposition’s greater incentive to mobilize to influence the election’s outcome (Tucker 2007; Harish and Little 2017). We do not attempt to disentangle these various possibilities here.
6 Conclusions

We develop an economic model in which a regime can take a costly action to prevent protest. We define the regime’s tolerance as the ratio of its cost of repression to its cost of protest. We establish conditions under which the equilibrium probability of protest is bounded above by the regime’s tolerance. We establish further conditions under which the equilibrium probability of protest approaches this bound, such that the regime’s tolerance is identified from the distribution of the equilibrium probability of protest.

We develop an estimator of tolerance that is motivated by this approach to identification. We introduce a novel daily, cross-national panel of protest occurrence and protest predictors, and we use machine-learning methods to implement our estimator for a large cross-section of countries. We demonstrate the utility of the resulting estimates of tolerance by applying them to the question of whether ratings produced by Freedom House are politically biased, and the question of whether regimes become more tolerant of protests in periods close to elections.

Although we focus our analysis on the tolerance of political protest, we believe that our approach could be applied to measure a regime’s tolerance for other kinds of actions by its citizens. For example, one may imagine using text analysis or other tools to predict public criticism of a regime (for example by journalists), and thus to measure the regime’s tolerance of such criticism. As another example, one may imagine using social media and other data to predict acts of violence such as terrorist attacks, and thus to measure the implicit tolerance of such acts.\textsuperscript{18} We believe that such applications are a promising direction for future work.

Appendix: Proofs

Proof of Proposition\textsuperscript{1}

\textit{Part (i)}

If $\tau_i = 1$ then the statement is trivial. So say that $\tau_i < 1$ and pick strategies $(\sigma_m, \sigma_r)$ such that $\lambda_i(\omega, \sigma_m(\omega), \sigma_r(\omega, \sigma_m(\omega))) > \tau_i$ for some state $\omega \in [0, \omega_i]$. Then the expected payoff to the

---

\textsuperscript{18}In the context of Africa, Cox et al. (2018) discuss evidence that social media contain substantial discussion of terrorist attacks.

20
regime at state $\omega$ is

$$-\lambda_i(\omega, \sigma_m(\omega), \sigma_r(\omega, \sigma_m(\omega))) L_i - \sigma_r(\omega, \sigma_m(\omega)) < -\tau_i = -\lambda_i(\omega, \sigma_m(\omega), \tau_i) L_i - \tau_i$$

where the inequality follows from $\lambda_i(\omega, \sigma_m(\omega), \sigma_r(\omega, \sigma_m(\omega))) > \tau_i = \tilde{\tau}_i / L_i$ and $\sigma_r(\omega, \sigma_m(\omega)) \geq 0$, and the equality follows from Assumption $[a]$. Therefore at state $\omega$ the regime prefers to take $r_i = \tilde{\tau}_i$ than to take $r_i = \sigma_r(\omega, \sigma_m(\omega))$, and hence $(\sigma_m, \sigma_r)$ is not an equilibrium.

**Part [ii]**

By Assumption $[b]$ the maximand in (4) is convex in $r$ for all $\omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i]$. Therefore there is a unique mapping $\sigma_r^*(\omega, m)$ that solves (4), with the property that $\sigma_r^*(\omega, m) \in \{0, \tau_i\}$ for all $\omega \in [0, \bar{\omega}_i], m \in [0, \bar{m}_i]$. Further from the structure of payoffs and Assumption $[a]$ we have that

$$\sigma_r^*(\omega, m) = \begin{cases} \tilde{\tau}_i, & \lambda_i(\omega, m, 0) > \tau_i \\ 0, & \lambda_i(\omega, m, 0) \leq \tau_i. \end{cases}$$

In any equilibrium, the opposition chooses

$$\sigma_m(\omega) \in \arg \max_m (B_i \lambda_i(\omega, m, \sigma_r^*(\omega, m)) - m).$$

For any $\omega$, the opposition prefers some $m'' < m'$ to any $m' > 0$ such that $\sigma_r^*(\omega, m') = \tilde{\tau}_i$. So now define

$$\bar{m}(\omega) = \begin{cases} 0, & \sigma_r^*(\omega, 0) = \tilde{\tau}_i \\ \max \{m \in [0, \bar{m}_i] : \sigma_r^*(\omega, m) = 0\}, & \sigma_r^*(\omega, 0) = 0 \end{cases}$$

where the maximum exists by Assumption $[c]$. Now we can say that the opposition chooses

$$\sigma_m(\omega) \in \arg \max_{m \in [0, \bar{m}(\omega)]} (B_i \lambda_i(\omega, m, 0) - m).$$

Because the set $[0, \bar{m}(\omega)]$ is compact and the function $\lambda_i(\omega, m, 0)$ is continuous in $m$, this problem has at least one solution $\sigma_m^*(\omega)$ for each $\omega \in [0, \bar{\omega}_i]$. This establishes the existence of an equilibrium.

Now write

$$\lambda_i^*(\omega) = \lambda_i(\omega, \sigma_m^*(\omega), \sigma_r^*(\omega, \sigma_m^*(\omega))).$$

Following the reasoning above, if $\lambda_i(\omega, 0, 0) > \tau_i$, then $\lambda_i^*(\omega) = 0$, and if $\lambda_i(\omega, 0, 0) \leq \tau_i$, then
\( \sigma_r^*(\omega, \sigma_m^*(\omega)) = 0 \) and
\[
\tau_i \geq \lambda_i^*(\omega) = \lambda_i(\omega, \sigma_m^*(\omega), 0) \geq \lambda_i(\omega, 0, 0).
\]

**Proof of Corollary 1**

Fix an equilibrium. That \( \tau_i \leq 1 \) is immediate from the definition of \( \tau_i \). That \( \tau_i \geq \lambda_i^* \) follows from the definition of \( \lambda_i^* \) because, by Proposition 1(i), \( \Pr(\lambda_i^*(\omega_{it}) \leq \tau_i) = 1 \).

**Proof of Proposition 2**

Fix an equilibrium. By Proposition 1, \( \Pr(\lambda_i^*(\omega_{it}) \leq \tau_i) = 1 \). Notice that for any \( \lambda \in [0, \tau_i) \) we have that
\[
\Pr(\lambda_i^*(\omega_{it}) \leq \lambda) = 1 - \Pr(\lambda_i^*(\omega_{it}) > \lambda)
= 1 - \Pr(\lambda_i^*(\omega_{it}) \in (\lambda, \tau_i])
\leq 1 - \Pr(\lambda_i(\omega_{it}, 0, 0) \in (\lambda, \tau_i])
< 1
\]
where the second step follows from Proposition 1(i), the third from Proposition 1(ii) and the last from Assumption 2. From the definition of \( \lambda_i^* \), it then follows that \( \lambda_i^* \geq \tau_i \) and therefore that \( \lambda_i^* = \tau_i \).

**Proof of Claim 1**

Fix an equilibrium. Observe that
\[
\Pr(z_{it} = 1 | \chi_i(\omega_{it})) = \mathbb{E}(z_{it} | \chi_i(\omega_{it}))
= \mathbb{E}(\mathbb{E}(z_{it} | \chi_i(\omega_{it}), \omega_{it}) | \chi_i(\omega_{it}))
= \mathbb{E}(\mathbb{E}(z_{it} | \omega_{it}) | \chi_i(\omega_{it}))
= \mathbb{E}(\lambda_i^*(\omega_{it}) | \chi_i(\omega_{it}))
\]
where the second step follows from the law of iterated expectations, the third from the fact that \( \chi_i(\omega_{it}) \) is a function of \( \omega_{it} \), and the last from the definition of \( \lambda_i^*(\omega_{it}) \).

Because \( \Pr(z_{it} = 1 | \chi_i(\omega_{it})) = \mathbb{E}(\lambda_i^*(\omega_{it}) | \chi_i(\omega_{it})) \), the support of \( \Pr(z_{it} = 1 | \chi_i(\omega_{it})) \) is in the convex hull of the support of \( \lambda_i^*(\omega_{it}) \), so
\[
\inf\{\lambda \in [0, 1] : \Pr(\Pr(z_{it} = 1 | \chi_i(\omega_{it})) \leq \lambda) = 1\} \leq \lambda_i^*
\]
by Definition 3. If $\chi_i(\cdot)$ is one-to-one, then $E(\lambda^*_i(\omega_t) \mid \chi_i(\omega_t)) = \lambda^*_i(\omega_t)$, so the preceding inequality holds with equality, again by Definition 3.

References


Davenport, Christian, Håvard Mokleiv Nygård, Hanne Fjelde, and David Armstrong. 2019. The


Logan, Carolyn, and Robert Mattes. 2012. Democratising the Measurement of Democratic Qual-


Figure 1: Profile of Predictor Variables Around Key Events

Panel A: Search Query Volume

Panel B: News Mentions

Panel C: Social Media Mentions

Notes: The figure shows daily event-study plots of standardized values of our key predictor variables around alerts of future protests and around the occurrence of protests. Each pair of plots is constructed from a regression in which the unit of analysis is the country-date and the model includes country and date fixed effects. The independent variables of interest are seven leads (relative days -7 through -1) and seven lags (relative days 1 through 7) of both an indicator for the occurrence of an alert indicating a future protest and our main protest indicator $z_{it}$, as well as two variables reflecting the cumulative number of future-protest alerts and protests more than 7 days in the future and two reflecting the cumulative number more than 7 days in the past. The contemporaneous future-protest alert and protest indicators are excluded as a normalization. We recenter the y-axis in each plot by adding a constant equal to the sample mean of the dependent variable on dates with an alert indicating a future protest (left plots) or on dates with a protest (right plots). The inner bars depict 95% pointwise confidence intervals and the outer lines depict 95% uniform sup-t bands, both based on inference clustered by country.
Figure 2: Simulation Evidence on the Performance of the Estimator

Notes: The figure show a box-and-whisker plot summarizing the distribution, across countries, of the median percentile of the estimated tolerance in the distribution of underlying protest probabilities over the 20 replicates of a simulation exercise. For each replicate, we simulate protest occurrence according to the probabilities $\hat{\lambda}_{it}$ estimated in our baseline predictive model, independently across days. We then estimate tolerance following the procedure described in Sections 3.1 and 4.5, replacing the protest occurrence indicator $z_{it}$ with its simulated analogue. For each country and replicate, we compute the percentile of the estimated tolerance $\hat{z}_i$ in the country-specific distribution of the protest probabilities $\hat{\lambda}_{it}$, where a percentile of 1.00 corresponds to estimating the maximum protest probability, $\max \hat{\lambda}_{it}$, in country $i$. For each country, we then compute the median percentile across the replicates.
Figure 3: Tolerance, Frequency, and Anticipation of Protest by Expert-Rated Freedom

Panel A: Tolerance of Protest

Panel B: Anticipation of Protest

Panel C: Frequency of Protest

Notes: Each box-and-whisker plot shows statistics of a given variable of interest across all country-years (y-axis) with a given value of the FITW freedom of assembly rating (x-axis). Outside values are not shown in the box-and-whisker plots. In Panel A, the variable of interest is the estimated tolerance ($z_{it}$) in the given country-year. In Panel B, the variable of interest is the share of protest days (days with $z_{it} = 1$) on which protest is anticipated by an alert at least one day in advance ($a_{it} = 1$). In Panel C, the variable of interest is the number of days with protest ($z_{it} = 1$).
Notes: The map shows the estimated tolerance $\hat{\pi}_i$ for each country in our sample. Darker colors represent greater estimated tolerance. Territories not in our sample are colored white.
Figure 5: A Test for Political Bias in FITW Ratings

Notes: To construct this plot, we regress the estimated tolerance $\tau_i$ of each country $i$ on indicators for quintiles of the average of an annual measure of the agreement between the country’s UN General Assembly votes and those of the US, controlling for indicators for quintiles of the country’s sample average annual FITW freedom of assembly rating. We omit the indicator for the first quintile of agreement from the regression, and recenter the plot by adding a constant equal to the mean of estimated tolerance $\tau_i$ among countries in this first quintile. The depicted intervals are pointwise 95% confidence intervals. The reported $p$-value is for an $F$-test of the hypothesis that the depicted coefficients are jointly equal to zero.
Figure 6: Protest Tolerance and Incidence Around Elections, Democracies vs. Non-Democracies

Panel A: Non-democracies

Tolerance of protest

Incidence of protest

Panel B: Democracies

Tolerance of protest

Incidence of protest

Notes: The figure shows monthly event-study plots of protest tolerance and protest incidence around elections. Each plot is constructed from a separate regression in which the unit of analysis is the country-month and the model includes fixed effects for country and for calendar month. The independent variables of interest are 12 leads and lags of an indicator for the occurrence of an election, and two variables reflecting the cumulative number of elections 13 or more months in the future and the cumulative number of elections 13 or more months in the past. The contemporaneous election indicator is excluded as a normalization. We recenter the y-axis in each plot by adding a constant equal to the sample mean of the dependent variable in election months. The inner bars depict 95% pointwise confidence intervals and the outer lines depict 95% uniform sup-t bands, both based on inference clustered by country. In the plots on the left (“Tolerance of Protest”), the dependent variable is the estimated tolerance $\tilde{z}_{i,g,i}$ of each country $i$ in each month $g$. In the plots on the right (“Incidence of Protest”), the dependent variable is an indicator $\max_{t\in g} z_{i,t}$ for the occurrence of protest in country $i$ in month $g$. In Panel A the sample includes country-years classified as non-democracies (autocracies or anocracies) by the Polity Project. In Panel B the sample includes country-years classified as democracies by the Polity Project.
Online Appendix for
Measuring the Tolerance of the State:
Theory and Application to Protest

Veli M. Andirin, Brown University*
Yusuf Neggers, University of Michigan
Mehdi Shadmehr, University of North Carolina at Chapel Hill
Jesse M. Shapiro, Brown University and NBER

*E-mail: veli_andirin@brown.edu, yneggers@umich.edu, mshadmeh@gmail.com, jesse_shapiro_1@brown.edu.
A Additional Details on Data Sources and Variable Construction

A.1 Parsing of Crisis24 Security Alerts

We identify protest-related alerts using information provided in alert descriptions. We classify an alert as protest-related if its description includes the words “protest,” “demonstration,” or “demonstrator,” or words with those keywords at the root. Because some protests originate or manifest as labor strikes, we also classify as protests alerts whose descriptions contain one of the words “worker,” “union,” “labor,” or “labour” and the root “strike” in the same sentence.\(^1\) We exclude from our search of the description text some generic warnings such as “We advise our clients to stay away from protests.”

We extract, for each alert, any dates mentioned in its title and description.\(^2\) We consider all dates mentioned in the alert’s title to be protest dates. If there is no date in the title, then we consider the latest date mentioned in the description to be a protest date.

We also classify alerts into temporal categories based on syntactic and semantic information in alert titles. We classify an alert as pertaining to a future protest if its title contains a future-related word such as “announce” along with a protest-related keyword, or if its title contains a protest-related verb preceded by “to,” as in “to protest.” We classify an alert as pertaining to a present protest if its title contains present-related words such as “continued,” “ongoing,” or “underway” appearing in the same sentence with a protest-related keyword. We classify a protest alert as pertaining to a past protest if its title contains protest-related verbs in the past tense, such as “protested” or “demonstrated,” or if its title does not contain word structures indicative of being about the present or future.\(^3\)

Online Appendix Figure 1 illustrates the main elements of our approach using example alerts drawn from our data.

A.2 Financial Indices

We obtain daily data on a range of financial indices. Online Appendix Figure 2 gives the dates of coverage of each index for each sample country. We now discuss each index in more detail.

---

\(^1\)An exception is the case where “strike” is part of the word “airstrike.” Online Appendix Figure 3 reports results when we do not include strike-related words in the set of words we use to classify an alert as a protest.

\(^2\)Date information typically includes the month and day of month, e.g., “May 20.” We treat these dates as referring to the year in which the alert is published, unless the alert is published in December and the date is in January, in which case we treat the date as referring to the year following the alert’s publication.

\(^3\)We also search for specific clauses such as “hundreds of people gather” that, according to our reading of alerts, typically indicate that the event was in the past even though they are written in the present tense.
A.2.1 Exchange Rates

We obtain daily data on exchange rates with respect to the US Dollar for a set of currencies from Bloomberg L.P. (2021a). We match currencies to countries using the ISO 4217.4

A.2.2 Sovereign Bond Indices

We obtain from Bloomberg L.P. (2018) daily data on the yield index from J.P. Morgan’s Emerging Market Bond Index Global (EMBIG).

A.2.3 Stock Market Indices

We obtain from Bloomberg L.P (2021b) data on the MSCI Inc. (formerly Morgan Stanley Capital International) indices. For each country, we observe a daily sub-index for MSCI price, measured in the local currency, and a daily sub-index for MSCI trading volume.

---

4Lithuania, Latvia, Malta, Slovenia, Slovakia, and Estonia began using the Euro as their official currency after 2000. For these countries, we use the exchange rate of their official currency prior to the transition to Euro.
Online Appendix Figure 1: Illustrations of Security Alert Parsing

Sample Alert A
Country: Nigeria
Published at: 2017-03-22
Title: Nationwide NULGE Protests Begin
Content: According to local media sources on Wednesday, 22 March, the Nigerian Union of Local Government Employees (NULGE) has commenced a nationwide protest calling for a constitutional amendment granting Local Government Areas autonomy. *Protests began in Nasarawa state and are predicted to continue across the country until President Muhammadu Buhari intervenes.* Travellers are advised to avoid political protests and exercise caution in the vicinity of gathering crowds in order to minimise the risk of exposure to potential crowd disturbances. Monitor local media sources for more information about how the protests will affect specific regions and the effect these may have on overland travel.

Sample Alert B
Country: Peru
Published at: 2018-07-06
Title: Protest to be Held in Arequipa on 10 July over Cost of Public Transport
Content: According to local media reports, a protest march is set to be held in Arequipa city over the high cost of public transportation. The action is being organised by a number of civil society organisations and is set to occur on Tuesday, 10 July. *Protesters are also angry about poor service and regular accidents.* Members in Arequipa are advised to avoid the demonstrations due to the associated risks of exposure to opportunistic crime, unruly crowd behaviour, and police crowd control measures. Monitor local media sources to remain aware of current tensions and for any updates related to planned or ongoing unrest in your area of operation.

Notes: For each sample alert we illustrate (i) terms used to identify protest-related alerts (highlighted in gray), (ii) dates used to identify the date of the protest (italicized), and (iii) generic text that our parser ignores (grayed out).
Online Appendix Figure 2: Date Coverage for Financial Indices, by Country

<table>
<thead>
<tr>
<th>Country name</th>
<th>Bond yields</th>
<th>Stock market prices</th>
<th>Trading volume</th>
<th>Exchange rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albania</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angola</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armenia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azerbaijan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bahrain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belarus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolivia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosnia And Herzegovina</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Botswana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burundi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cameroon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central African Republic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congo, DR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costa Rica</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Côte d'Ivoire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuba</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czechia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country name</th>
<th>Bond yields</th>
<th>Stock market prices</th>
<th>Trading volume</th>
<th>Exchange rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominican Republic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecuador</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El Salvador</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eritrea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eswatini</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guatemala</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinée-Bissau</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haiti</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jamaica</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kazakhstan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuwait</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lao</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notes: For each country (in rows) and each financial indicator (in columns), the filled region of the timeline depicts the dates from October 2009 through December 2019 for which data are available.
B Additional Empirical Results

Online Appendix Figure 3: Estimates of Tolerance Under Alternative Specifications

<table>
<thead>
<tr>
<th></th>
<th>Relationship to baseline</th>
<th>Relationship to freedom rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td><img src="image" alt="Baselines" /></td>
<td><img src="image" alt="Baseline" /></td>
</tr>
<tr>
<td><strong>Modifying the estimator</strong></td>
<td><img src="image" alt="Modifying1" /></td>
<td><img src="image" alt="Modifying2" /></td>
</tr>
<tr>
<td>Protest occurrence within 2 days</td>
<td><img src="image" alt="2days" /></td>
<td><img src="image" alt="2days的路上" /></td>
</tr>
<tr>
<td>Protest occurrence within 3 days</td>
<td><img src="image" alt="3days" /></td>
<td><img src="image" alt="3days的路上" /></td>
</tr>
</tbody>
</table>
Online Appendix Figure 3: Estimates of Tolerance Under Alternative Specifications (Continued)

**Modifying the estimator (continued)**

<table>
<thead>
<tr>
<th>Relationship to baseline</th>
<th>Relationship to freedom rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude cells with ties</td>
<td></td>
</tr>
<tr>
<td>Richer set of covariates</td>
<td></td>
</tr>
<tr>
<td>Cells as two-month groups</td>
<td></td>
</tr>
<tr>
<td>Random grouping of country-years</td>
<td></td>
</tr>
</tbody>
</table>
Online Appendix Figure 3: Estimates of Tolerance Under Alternative Specifications (Continued)

<table>
<thead>
<tr>
<th>Relationship to baseline</th>
<th>Relationship to freedom rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modifying the estimator (continued)</strong></td>
<td></td>
</tr>
<tr>
<td>No penalty</td>
<td><img src="image1.png" alt="Graph" /> <img src="image2.png" alt="Boxplot" /></td>
</tr>
</tbody>
</table>

| **Modifying the input data** |                                  |
| Exclude protests with an uncertain date | ![Graph](image3.png) ![Boxplot](image4.png) |
| Include protests in MMDP/ICEWS | ![Graph](image5.png) ![Boxplot](image6.png) |
Modifying the input data (continued)

Exclude protests not in MMDP/ICEWS

Exclude strikes

Notes: Each row corresponds to a set of estimates of tolerance. The first column (“Relationship to baseline”) shows a scatterplot of the rank of each country’s tolerance under the given set of estimates (y-axis) against the rank of that country’s tolerance under the baseline estimates (x-axis), along with the 45-degree line. The second column (“Relationship to freedom rating”) shows a box-and-whisker plot of the estimated tolerance across all country-years (y-axis) with a given value of the FITW freedom of assembly rating (x-axis), as in Panel A of Figure 3. “Baseline” corresponds to the baseline specification described in Section 4.5. “Protest occurrence within D days” corresponds to modifying our estimator to \( \hat{z}_i = \frac{1}{G} \sum_{g=1}^{G} \max \{ \tilde{z}_{i,g} - \tilde{z}_{i,g+1} + \ldots + \tilde{z}_{i,g+D} \} \). “Exclude cells with ties” corresponds to excluding from the calculation of \( \hat{z}_i \) any country-month \( g \) in which the largest estimated predicted protest probability occurs on two or more distinct dates within the month. “Richer set of covariates” corresponds to including in covariates \( x_t \) the additional variables described in Section 4.4 specifically: search query volume about the country in the US and UK, protest Twitter mentions in the country’s official languages, the set of financial indices described in Online Appendix A.2, an indicator for whether an election occurred in the given country within seven days of the given date, and an indicator for whether a protest occurred in the given country on the same date in the previous year, where all of these except the two indicators are standardized as described in Section 4.4. “Cells as two-month groups” corresponds to letting each \( g \) consist of two-month instead of a single calendar month. “Random grouping of country-years” corresponds to randomly assigning calendar years to the groups \( G_1(i) \) and \( G_2(i) \), separately by country, by placing years in a random order and then assigning every other year to \( G_1(i) \). “No penalty” corresponds to estimating the predictive model via unpenalized maximum likelihood. “Exclude protests with an uncertain date” corresponds to excluding protests whose dates we were not able to infer from text fields. “Include protests in MMDP/ICEWS” corresponds to including in \( z_{it} \) country-dates classified as a protest day according to at least one of MMDP or ICEWS protest indicators described in Section 4.2. “Exclude protests not in MMDP/ICEWS” corresponds to excluding from \( z_{it} \) protests for which MMDP does not indicate a beginning or ongoing protest in the same calendar week, and ICEWS does not indicate any protest-related stories in the same calendar week, provided that the country-week is covered by at least one of these sources. “Exclude strikes” refers to excluding alerts that have only strike-related keywords but lack protest-related ones.
Notes: The figure shows daily event-study plots of protest indicators based on MMDP (Panel A) or ICEWS (Panel B) around the occurrence of protest according to our main indicator $z_{it}$. Each plot is constructed from a regression in which the unit of analysis is the country-date and the model includes country and date fixed effects. The independent variables of interest are seven leads (relative days -7 through -1) and seven lags (relative days 1 through 7) of our main protest indicator $z_{it}$, as well as two variables reflecting the cumulative number of protests more than 7 days in the future and more than 7 days in the past. The contemporaneous protest indicator is excluded as a normalization. We recenter the y-axis in each plot by adding a constant equal to the sample mean of the dependent variable on dates with $z_{it} = 1$. The inner bars depict 95% pointwise confidence intervals and the outer lines depict 95% uniform sup-$t$ bands, both based on inference clustered by country.
Online Appendix Figure 5: Estimated Tolerance by Country

Notes: The figure shows the estimated $\zeta_i$ by continent group, computed following Section 4.5, along with a pointwise 95% Clopper-Pearson confidence interval.
Notes: The figure shows monthly event-study plots of protest tolerance around elections, controlling for protest incidence. Each plot is constructed from a separate regression in which the unit of analysis is the country-month, the dependent variable is the estimated tolerance $z_{i,t(g,i)}$ of each country $i$ in each month $g$, and the model includes a control for an indicator $\max_{t \leq g} z_{i,t}$ for the occurrence of protest in country $i$ in month $g$, as well as fixed effects for country and for calendar month. The independent variables of interest are 12 leads and lags of an indicator for the occurrence of an election, and two variables reflecting the cumulative number of elections 13 or more months in the future and the cumulative number of elections 13 or more months in the past. The contemporaneous election indicator is excluded as a normalization. We recenter the $y$-axis in each plot by adding a constant equal to the sample mean of the dependent variable in election months. The inner bars depict 95% pointwise confidence intervals and the outer lines depict 95% uniform sup-t bands, both based on inference clustered by country. In the plot on the left the sample includes country-years classified as non-democracies (autocracies or anocracies) by the Polity Project. In the plot on the right the sample includes country-years classified as democracies by the Polity Project.
Appendix References Not Appearing in Main Text


