# Between Order and Anarchy:

# A Hobbesian Logic of State Coercion

Hansol Kwak

**Abstract**

How do governments calibrate the level of coercion against protests? Drawing on the perspective of a Hobbesian balance between order and anarchy, this study argues that repression operates as a continual adjustment around a socially intelligible normal line rather than a fixed response to dissent. Using ACLED, repression is measured in two complementary ways, as state-involved fatalities and as the use of excessive force, and a country-specific reference point summarizes coercive history. The key explanatory variable captures the gap between prior repression and this reference baseline. Results consistently reveal a corrective dynamic: larger deviations are followed by countervailing adjustments, with stronger effects where coercion is more publicly observable and under heightened background insecurity. Autoregressive benchmarks, a buffered reference point, and a separated specification indicate that this pattern is not fully accounted for by the mechanical coupling of the design or the serial dependence of repression alone. A sign-split specification shows that correction falls more sharply on over-repression than on under-repression, a pattern beyond what mean reversion alone produces. Governments thus appear to counteract departures from the level of coercion that observers have come to expect, stabilizing the perceived boundary between order and disorder.

# Introduction

How do governments decide how much coercive repression to use when protests erupt? A dominant answer in the literature is straightforward: states respond to threats. When dissent becomes larger, more violent, more organizationally cohesive, or more likely to diffuse, governments increase repression; when the perceived threat subsides, coercion declines. Across regime types and contexts, repression is commonly modeled as a function of opposition capacity, mobilizational strength, and the risks such challenges pose to regime survival or public order. In this framework, protest size, violence, coalition breadth, and structural vulnerabilities serve as signals to which rulers react strategically.

This threat-centered approach has generated a rich body of findings, showing that governments tend to escalate coercion when protests intensify, when challengers appear organizationally robust, or when rulers anticipate future instability. Repression, in short, is primarily understood as a reaction to an immediate threat, whether to ongoing protest, anticipated risk, or accumulated strategic experience. Yet this framework leaves an important puzzle unresolved. In practice, repression has not always followed a proportional relationship with protest threat.[1] Historical research on protest control in Britain and continental Europe documents recurring oscillations between more coercive and more restrained strategies, rather than steady escalation or decline (Della Porta and Reiter, 1998). British responses to industrial unrest shifted markedly. Late nineteenth-century episodes involved lethal force against striking workers, whereas by the early and mid-twentieth century, strikes were managed with comparatively limited bloodshed. Germany and Italy also exhibited recurring swings between more tolerant and more violent modes of repression from the 1950s to the 1990s. These patterns suggest that repression often moves within a bounded range, alternating be-

[1] For example, during the 1999 Seattle WTO protests, police repeatedly used pepper spray against nonviolent demonstrators (Chasan and Walker, 2000). At the 2001 G8 summit in Genoa, numerous reports of excessive force emerged despite the largely peaceful character of the protests (AMNESTY, 2006). Similar concerns were raised during the policing of Occupy Wall Street (Knuckey, Glenn and MacLean, 2012).

tween mild coercion and crackdown. The central puzzle is thus why repression frequently presents oscillation rather than proportionality in response to the level of protest threats.

This article approaches repression from a different angle, drawing on a Hobbesian perspective on political authority. For Hobbes, the sovereign's defining task is to prevent the collapse of social order into a war of all against all. From this viewpoint, coercion is not only an instrument deployed against dissent; it is part of the broader practice of governance through which the state sustains a recognizable boundary between order and disorder. The political problem facing governments is therefore not simply how to neutralize dissent, but how to do so in a manner that preserves the intelligibility of authority itself: force must appear neither arbitrarily excessive nor dangerously insufficient.

In this sense, repression carries symbolic meaning beyond its immediate tactical effects. When the state uses far more force than citizens have come to expect, coercion can begin to resemble the unregulated violence it claims to prevent. When it uses substantially less force than expected, it risks appearing incapable of maintaining minimal order. In both directions, large departures from established patterns blur the distinction between regulated authority and anarchy. What matters, then, is not only the level of repression in response to a given protest, but how that level compares to the patterns of state behavior observers have come to expect.

Building on this insight, the article advances a reference-dependent argument. Over time, repeated exposure to state coercion generates a socially intelligible normal line of repression, a baseline against which new episodes of force are interpreted. Citizens, protesters, and elites evaluate current repression relative to this accumulated history rather than in isolation. Governments, in turn, attend to these historically formed expectations, even if they observe them only imperfectly through media reactions, elite discourse, and public responses. The central claim is that visible departures from this baseline, in either direction, generate pressure for countervailing adjustment. When prior repression exceeds what recent history makes expectable, subsequent repression is more likely to decline; when it falls short, subse-

quent repression is more likely to increase. Repression thus unfolds as a process of continual recalibration rather than as a sequence of independent reactions to protest threat.

This corrective dynamic should also vary in strength across contexts. It should be more pronounced where the coercive act is more publicly observable, since a deviation that reaches a larger audience makes the gap between current and expected conduct more legible and correction more urgent. It should likewise intensify under conditions of heightened background insecurity. When generalized violence is high and disorder feels proximate, the political costs of miscalibration rise. Excessive force more readily signals sovereign instability; insufficient force more readily signals incapacity. In such contexts, deviation itself becomes destabilizing, and governments face stronger incentives to limit departures from the socially intelligible baseline.

The article evaluates these expectations using the Armed Conflict Location and Event Data (ACLED), measuring repression in two complementary ways across consecutive protest events: the fatalities resulting from state coercion and a binary indicator of whether state forces used excessive force. A country-specific reference point is constructed from temporally weighted histories of repression, intended as an empirical proxy for this accumulated collective memory. To guard against the concern that the corrective association merely restates the mechanical coupling and serial dependence built into the design, the empirical strategy benchmarks the gap against increasingly deep autoregressive controls, buffers the most recent events out of the reference point, and separates the lagged level from the reference point as distinct regressors. A sign-split specification further distinguishes the corrective pattern from simple regression to the mean. Across both measures and these alternative constructions, the results reveal a consistent corrective pattern: larger deviations from the reference point are followed by countervailing adjustments, and this pattern strengthens where coercion is more publicly observable and where background insecurity is higher.

# Literature

A large body of scholarship conceptualizes state repression as a response to perceived threats from domestic challengers. Foundational studies argue that governments employ coercion when dissent poses risks to regime survival, public order, or the broader stability of the political system (Davenport, 2007; Gartner and Regan, 1996; Poe and Tate, 1994; Regan and Henderson, 2002; Ritter, 2014). Across regime types and institutional contexts, rulers assess the magnitude and character of opposition activity and calibrate force accordingly (Earl, Maher and Pan, 2022; Hendrix and Salehyan, 2017; Hultquist, Asal and Rethemeyer, 2025).

Contributions emphasized that threat is multidimensional rather than reducible to a single indicator. Governments do not respond solely to the presence of protest, but to combinations of factors that signal the severity and scope of opposition challenges (Davenport, 1995, 1996). Subsequent research reinforces this view across diverse empirical domains. In protest and civil conflict settings alike, more violent, more sustained, or more organizationally robust forms of dissent are associated with higher levels of repression (Carey, 2006, 2010; Young, 2013). The cumulative implication of this literature is that repression tracks observable threat environments: when dissent intensifies or becomes more dangerous, coercion tends to increase.

Further work highlights that the political meaning of dissent depends on coalition structure and social breadth. Broad, cross-cleavage, or multiethnic alliances can signal organizational strength and resilience, thereby altering both perceived threat and the costs of repression (Dahlum, 2023; Manekin, Mitts and Zeira, 2024; Stephan and Chenoweth, 2008). In such contexts, the decision to repress involves weighing the risks of emboldening opposition against the reputational and political consequences of visible coercion.

At the same time, a related strand of scholarship expands the threat framework by distinguishing between realized and latent risks. Governments may respond not only to current mobilization but also to structural conditions that increase the likelihood of future insta-

bility. Research on demographic pressures, minority configurations, and external security environments demonstrates that rulers sometimes act preemptively in anticipation of potential contention (Artabe et al., 2023; Cunningham, Ward and Owens, 2019; Heffington, 2021; Nordås and Davenport, 2013). In protest contexts, authorities often interpret demonstrations as signals of broader mobilizational capacity or diffusion potential rather than isolated events (Aytaç, Schiumerini and Stokes, 2017; Earl, Soule and McCarthy, 2003). States thus sometimes repress in response to anticipated contagion effects rather than contemporaneous violence alone (Danneman and Ritter, 2014; Karreth, Tir and Gibler, 2022). These studies suggest that repression is shaped not only by what dissidents are currently doing, but also by what they might plausibly do under changing political conditions.

While much of the preceding literature emphasizes governments' responses to contemporaneous or anticipated future threats, another trend of research extends the temporal horizon backward by examining how past repression shapes present coercive behavior. A notable work is Davenport's analysis of lagged determinants, which explicitly demonstrates that repression exhibits substantial temporal depth (Davenport, 1996). By incorporating distributed lags, this work shows that contemporary coercion reflects accumulated historical experience rather than immediate conditions alone. This work advances a synthetic framework in which short-term dynamics are consistent with strategic responsiveness, while longer-term effects are attributed to habituation and the routinization of decision-making. This move broadens the temporal scope of repression research and highlights the weight of the past.

Despite these contributions, several limitations remain. First, much of the literature retains an almost automatic connection between threat and state response, in which repression is modeled primarily as a function of variations in protest activity, political conflict, or structural vulnerability. Even when institutional or contextual factors are incorporated, the underlying explanatory structure typically assumes that changes in external threat environments should produce corresponding shifts in repression levels. This orientation leaves

limited space for the patterned oscillations in repression. As Della Porta and Reiter (1998) identify, repression often involves recurring movements between restraint and crackdown that cannot be reduced to proportionality to protest threats. This perspective thus struggles to account for why repression often adjusts within a bounded range, repeatedly shifting direction even in the absence of commensurate changes in observable threat conditions.[2]

Second, although temporal depth has been incorporated through lagged specifications, the decision logic that connects past experience to present coercion remains incompletely specified. Short-term repression tends to be framed as a strategic response to immediate threats, while longer-term dynamics are attributed to habituation, routinized practice, or institutional carryover. Yet the conditions under which governments shift between forward-looking calculation and reliance on established routines are not clearly delineated. Temporal dependence is thus demonstrated empirically, but the mechanisms through which past experiences structure present decision-making are not fully theorized.

Third, existing frameworks devote comparatively limited attention to the broader political meaning of coercion. Repression is typically modeled as a reaction to threatening environments, with primary emphasis on the material conditions that provoke state force. Less systematic attention is given to how coercive acts are situated within ongoing political contexts or how their significance may extend beyond immediate threat mitigation. As a result, while the determinants of repression have been extensively analyzed, how coercion functions within longer political trajectories remains less fully elaborated.

In summary, these limitations suggest the need to reconsider repression not only as a reaction to threat, nor solely as a product of accumulated historical influence, but as a political practice embedded within broader processes of governance. Existing studies pay comparatively less attention to how repression participates in the ongoing construction and

[2] Della Porta and Reiter (1998) explain oscillations in repression as contingent on specific circumstances, including public demands and broader institutional frameworks. However, their account does not fully clarify why these underlying conditions themselves shift over time.

maintenance of state authority. Coercion does not operate only as an instrument for managing dissent; it also carries symbolic and political significance within evolving contexts of rule. Understanding repression thus requires moving beyond models that treat it exclusively as proportional responsiveness or temporal persistence, and toward an account that situates coercion within the larger political project of sustaining a recognizable order.

# Theory

Echoing Hobbes's insight that political authority is grounded on the sovereign's ability to prevent a war of all against all (see Abizadeh, 2011), state coercion is not simply a reaction to protest or unrest, but is part of a broader effort to maintain a recognizable boundary between order and anarchy (Davenport, 2007). Thus, governments do not rely on force by itself; what matters equally, if not more, is whether the use of force appears consistent and intelligible to citizens, as a routine preservation of order rather than arbitrary violence or a sign of losing control. In everyday political life, this balance may be fragile. If the state uses far more force than people expect, it can begin to look excessive or even tyrannical, blurring the moral distinction between lawful authority and the very chaos it claims to prevent. But if the state uses too little force, it risks appearing weak or incapable, inviting challenges from more radical actors, eroding confidence in its ability to preserve stability. In both cases, what is lost is not only effectiveness but clarity: citizens struggle to tell the difference between regulated power and unregulated violence. Put differently, there exists a socially recognized normal line of coercion, and large deviations in either direction shrink the perceived gap between order and disorder.

This socially recognized normal line develops through repeated exposure to patterns of state behavior, as individuals and groups evaluate current uses of force in relation to what they have historically observed. Citizens, protesters, media actors, and even state agents themselves gradually form expectations about what level of coercion appears typical, tol-

erable, or predictable in their political trajectory. In this respect, the interpretation of repression follows the reference-dependent logic emphasized in prospect theory (Kahneman and Tversky, 1979): political reactions are shaped less by the absolute magnitude of force than by how current actions compare to an accumulated baseline of prior experience (Kwak, 2025). These accumulated impressions serve as informal reference points that people rely on when judging state behavior in real time: whether a given response seems routine, unexpectedly harsh, or surprisingly restrained. Such expectations may not be purely individual or static; they are shaped through social interaction, media narratives, and observation of how others respond, gradually producing shared understandings of what counts as ordinary versus exceptional coercion.

From the government's perspective, this expectation structure generates not merely a tactical consideration but a Hobbesian constraint tied to the preservation of sovereign authority itself. The state is not only choosing how much force to apply in response to dissent, but also managing how far its actions depart from the socially accumulated reference point that shapes how coercion is interpreted. In this sense, the problem governments confront is not simply the presence of unrest, but the risk that their response will be perceived as a sharp deviation from established patterns of force. Excessive unpredictability in coercion threatens the sovereign's role by blurring the very boundary between regulated authority and unregulated violence that legitimizes its existence.

Specifically, a sudden spike in repression may restore short-term control yet simultaneously signal weakness (Kricheli, Livne and Magaloni, 2010), factional fragmentation among elites (Yuen and Cheng, 2017), or the erosion of procedural restraint (Davenport, 2007), suggesting that the sovereign is reacting to disorder rather than containing it. Conversely, an unexpectedly limited response may reduce immediate confrontation but risk signaling that the state is no longer capable of enforcing the minimal conditions of order. In both cases, what destabilizes authority is not merely the level of force itself but its misalignment with the expectations formed through prior experience. Coercion thus becomes not simply

an instrument of deterrence, but a mechanism whose political meaning depends on how far it departs from an accumulated baseline of past conduct. Hence, governments are pushed toward adjustment rather than maximization: because deviations in either direction blur the boundary between regulated authority and unregulated violence, they generate pressure for correction.

Hypothesis 1a: Greater deviations between a government's recent level of coercion and a socially recognized reference point will be associated with larger subsequent corrective changes in state repression.

The pressure toward correction, however, need not operate symmetrically across the two directions of deviation. Over-repression and under-repression both threaten the intelligibility of authority, but they do so through different mechanisms. First, over-repression creates a normative problem: when force exceeds the reference point, the state risks appearing not as the guardian of order but as a source of arbitrary violence. Under-repression, by contrast, creates a capacity problem: when force falls below the reference point, the state may appear unable to maintain order. Second, the two deviations differ in reversibility. Excessive force is difficult to present as legitimate authority once enacted. By contrast, restraint leaves more interpretive flexibility: it can be understood as tactical patience or temporary accommodation. For these reasons, deviations above the reference point should generate stronger incentives for correction than deviations below it. The corrective mechanism is therefore expected to be asymmetric: governments should counteract over-repression more sharply than equivalent under-repression.

Hypothesis 1b: Corrective adjustments will be larger for deviations above the reference point (over-repression) than for deviations of comparable magnitude below it (under-repression).

Protests are not uniform in how widely they are observed. Some demonstrations command broad public attention, drawing sustained coverage and circulating across social networks, while others remain largely confined to a local audience and pass with little notice. This variation implies a conditional structure to the correction mechanism. The pressure to counteract departures from socially formed expectations should not be constant across all instances of state coercion, but should intensify as the event's observability increases. Where a deviation is broadly visible, the government confronts a larger and more attentive audience for whom the gap between current and expected conduct is legible, and correction becomes more urgent. Where a deviation unfolds with little public notice, the reference point exerts a weaker pull, because the social comparison through which it operates is itself attenuated. The corrective dynamic, then, is expected to be strongest precisely where the coercive act is most publicly observable.

Hypothesis 2: Corrective adjustments will be larger when deviations from the reference point occur in more publicly observable events than in less observable ones.

Another condition concerns the broader environment in which deviations from the reference point are interpreted. The corrective dynamic should become especially pronounced in contexts of deep insecurity and a widespread sense that social breakdown is a real possibility. In such environments, the Hobbesian dilemma is immediately felt in everyday political life, and people increasingly seek reassurance that the state remains capable of preventing disorder from spreading. This heightened desire for reassurance raises the value of predictability in state behavior; any volatility in the government's use of force itself becomes a source of fear. Excessive force can appear as uncontrolled sovereign violence, suggesting that the institution meant to contain chaos is itself becoming erratic. Too little force, by contrast, raises doubts about whether the state can still perform its basic protective function. Because both extremes more easily blur the line between authority and disorder, the political

cost of misjudging the appropriate level of coercion rises sharply. Governments operating under high social insecurity are therefore pushed not simply toward harsher repression, but toward avoiding sharp departures from established patterns, which sustains the image of the state as a stable buffer against anarchy. In this sense, social insecurity compresses the room for deviation and strengthens the pressure to limit departures from socially formed expectations.[3]

Two clarifications about the reference point are warranted, since they bear directly on what this argument does and does not claim. First, the reference point is not a normatively optimal level of coercion, nor a level to which governments ought to return. A government with a long history of severe repression generates a high reference point, and one with a history of restraint generates a low one; the argument does not treat either baseline as appropriate or just. The reference point is a predictive anchor (the level of force observers have come to expect), and the mechanism operates not through the reference level itself but through the political cost of departing from it. Sharp deviation in either direction, regardless of where the baseline sits, is what renders authority less legible. Second, and relatedly, the claim is not that governments mechanically revert to their own historical average. It is that visible departures from socially formed expectations carry political costs, and that governments, observing those costs through imperfect feedback, face pressure to counteract

[3] Prospect theory provides a micro foundational mechanism for why insecurity intensifies this recalibration dynamic. Prospect theory predicts that actors evaluate outcomes relative to a psychological status quo and tend to perceive outcomes below that status quo as losses, even when they do not constitute objective losses in strict cost–benefit terms (Mercer, 2005). Once in a domain of loss, actors become more willing to engage in risk-taking behavior in an effort to recover what they perceive as deterioration; conversely, when outcomes exceed the status quo, actors enter a domain of gain and are more likely to behave in a risk-averse manner (Mercer, 2005; Vis, 2011). In contexts of heightened social insecurity, however, individuals are more likely to already experience the status quo as deteriorating. Widespread violence, instability, or perceived breakdown in public order can thus increase the probability that citizens situate themselves in a psychological domain of loss. Under such conditions, even moderate shortfalls in state protection or unexpected volatility in coercive responses may reinforce perceptions of further loss, amplifying the likelihood of risk-acceptant behaviors such as dissent, backlash mobilization, or violent protest. Governments, whether through explicit strategic calculation or accumulated experience, may recognize that misalignment in coercion is more likely to trigger such risk-seeking reactions when insecurity is high. As a result, the political costs of deviation increase, strengthening incentives for corrective recalibration and narrowing the corridor within which state coercion appears stable.

them. The reference point is thus better understood as a moving expectation generated by observed conduct than as an equilibrium toward which coercion is drawn.

Hypothesis 3: Corrective adjustments will be larger when deviations from the reference point occur in contexts of heightened social insecurity than in contexts of lower social insecurity.

## Alternative Arguments and Scope Conditions

An alternative mechanism suggests that governments may at times deliberately cultivate disorder rather than counteract departures from an established baseline of coercion. Instead of correcting deviations to preserve predictability, authorities may strategically withdraw security provision in order to generate visible chaos and insecurity. By allowing looting, crime, or localized breakdowns in public order, the state can impose the tangible costs of disorder on the broader population while signaling that protesters are unable to govern or maintain stability. Under this logic, volatility is not a constraint but an instrument: the government reframes the political choice as one between anarchy and a repressive yet order-providing state. Historical accounts of the 1988 Rangoon uprising describe the Burmese military withdrawing forces, releasing prisoners, and permitting widespread looting, thereby undermining popular support for the rebellion (Ferrara, 2003). Similarly, analyses of Egypt in 2013 highlight how perceived security vacuums following the retreat of police forces weakened public sympathy for protest movements (Elzoughby, 2011). In such cases, disorder itself becomes a political tool for reasserting the indispensability of sovereignty.

This mechanism is most plausible when the legitimacy of the existing order itself has substantially eroded. The corrective adjustment dynamic emphasized in the main theory presumes that the prevailing order remains normatively meaningful, that citizens still recognize a socially intelligible boundary between order and anarchy. However, when mobilization becomes mass-based, cross-cleavage, and explicitly oriented toward regime change, and when

a broad segment of society no longer regards the incumbent order as legitimate, adjusting repression may cease to function as a stabilizing signal. In such contexts, measured recalibration cannot restore authority because the baseline of order itself is no longer socially accepted. Governments may therefore abandon efforts to counteract departures within the existing baseline and instead attempt to reconstruct the boundary altogether. By permitting or inducing temporary breakdowns in security, authorities intensify the perceived costs of disorder and compel citizens to reassess the tradeoff between instability and coercive rule. When the legitimacy of order collapses, the strategic production of disorder can emerge as an alternative means of reasserting sovereignty.

Yet this strategy presupposes a near collapse of the normative legitimacy of the existing order and may thus be unlikely to characterize routine protest governance. The corrective adjustment dynamic advanced in this study assumes a more common condition: that the sovereign order, while contested, remains socially intelligible and normatively meaningful. In such contexts, governments retain incentives to preserve predictability in coercion, since visible volatility risks further undermining authority. Deliberately inducing disorder is inherently risky and can erode control rather than restore it.

Beyond the strategic production of disorder, three further mechanisms could in principle generate the negative gap coefficient without the Hobbesian logic of legibility. The first is adaptive policymaking: governments may simply learn by trial and error, scaling back coercion after a costly overreach and escalating it after an ineffective response. The second is a principal-agent dynamic: because frontline police and military units operate with partial autonomy, central authorities may rein in agents after excessive force and press them after visible restraint. The third is ordinary deterrence: harsh repression that suppresses dissent reduces the threat that elicits further force, while lenient repression that emboldens dissent invites escalation. All three would produce coercion that oscillates around its recent level, and none can be dismissed on the basis of the negative gap coefficient alone.

However, the corrective pattern documented here is bidirectional: higher-than-expected

repression is followed by decline, and lower-than-expected repression by increase. A pure deterrence account privileges the escalatory direction, in which rising threat draws rising force, and less naturally explains why unusually severe repression is systematically followed by retrenchment even as protest threat is held constant. The reference-dependent account treats both directions symmetrically, since departures of either sign blur the legibility of authority. The moderating role of background insecurity, examined below, further distinguishes the accounts: neither trial-and-error adjustment nor the disciplining of agents implies that correction should intensify specifically as ambient violence rises, whereas the reference-dependent logic predicts exactly this.

# Research Design

## Data

The empirical analysis draws on the Armed Conflict Location and Event Data (ACLED) Raleigh et al. (2010), an event-based dataset that records the date, location, actors, and reported fatalities of political violence and protest events worldwide. ACLED is well-suited to the present analysis because it records events at fine temporal resolution and identifies the actors involved, allowing repression to be measured for events involving state forces. Raw event records are aggregated by country and event date. When multiple events occur within the same country on the same date, both repression measures are aggregated as their mean: the fatality count as the mean number of fatalities, and the binary excessive-force indicator as the share of that country-date's state-involved events in which excessive force was used. The ACLED sample covers events worldwide between January 1997 and June 2026. The analysis focuses on protest and riot events involving state forces, since these are the instances in which state coercion in response to contention can be directly observed.

This study measures state repression in two complementary ways: the fatalities resulting from state coercion, and a binary indicator of whether state forces used excessive force

against demonstrators. Consistent findings across them guard against the possibility that the results depend on any single operationalization. The first captures the lethal magnitude of coercion, while the second reflects coercive intensity that need not result in death.

Neither measure claims to exhaust the concept of repression. State coercion encompasses a broad repertoire, including arrests, surveillance, crowd-control tactics, and various forms of nonlethal force (Earl, 2011). The focus on lethal violence and excessive force is motivated by their being among the most publicly visible and politically consequential components of that repertoire. Fatalities, in particular, represent among the most observable manifestations of state violence and are more likely to be consistently recorded across countries and over time, thereby providing a relatively comparable cross-national signal of the intensity of force deployed by the state. Criminological research, for example, treats fatalities as among the most reliably documented forms of violent behavior; intentional homicide rates are therefore often used as proxies for broader violent crime, and in some cases even overall crime (Malby, 2010). This logic is especially relevant for cross-national repression data, where nonlethal abuses such as arrests, harassment, detention, or ill-treatment are more vulnerable to underreporting, definitional inconsistency, and shifts in documentation practices. As Fariss (2014) shows, measures of human rights abuse can reflect changes in monitoring intensity and standards of accountability rather than behavioral change alone. Fatalities are not immune to these problems, but their severity makes them comparatively less sensitive to changing classification thresholds, and thus a more stable indicator of coercive force across time and countries. Moreover, lethal outcomes are disproportionately salient in media coverage, collective memory, and public discourse, and exert especially strong effects on political attitudes, including trust in law enforcement (Yameogo, 2025). Even when lethal outcomes arise from spontaneous escalation rather than deliberate targeting (see Nassauer, 2019), what informs future expectations is the observable magnitude of state violence itself. Together, the two measures provide a substantively meaningful and comparatively stable account of the most consequential forms of coercive state action, without claiming to capture its full repertoire.

## Dependent Variable: Change in Repression

The outcome of interest is the change in repression between consecutive events,

$$\Delta R_{it} = R_{it} - R_{i,t-1}.$$

Substantively, $\Delta R_{it}$ captures whether the government increases, decreases, or stabilizes its use of coercion relative to its immediately preceding action. Focusing on changes rather than levels reflects the theoretical claim that governments are engaged in a process of continual recalibration rather than static maximization of force. Constructed within each country across consecutive events, $\Delta R_{it}$ reflects movement relative to the government's own preceding action.

Here, $R_{it}$ denotes the level of state coercion at event time $t$ in country $i$, operationalized in two ways that correspond to the two measures of repression described above. The first is the mean number of reported fatalities in state-involved events on that day. Because fatality counts are highly skewed and changes can occur in either direction, a signed logarithmic transformation is applied to this measure in estimation. This transformation preserves the direction of change while reducing the influence of extreme outliers that could otherwise dominate the analysis. The second is a binary indicator of whether state forces used excessive force against demonstrators, which is bounded between zero and one and therefore requires no transformation.

## Reference Point and Misalignment

The central theoretical construct is the existence of a socially intelligible normal line of coercion, an implicit expectation about the level of force that appears routine rather than arbitrary. This expectation is a socially formed construct and is not directly observable. It is therefore operationalized with an empirical proxy: a reference point that summarizes recent repression history using temporally declining weights. The use of observed coercive

history as a proxy rests on the assumption that accumulated patterns of observed force are the principal input through which such expectations form.[4] This behavioral history is not claimed to be the social reference point itself. Citizens, protesters, and observers form expectations based on recent conduct, but also through media framing, selective memory, and salient episodes that a behavioral summary captures only imperfectly.

In practice, the reference point is constructed as a weighted average of past repression levels in which more recent events exert greater influence than distant ones. Recency is modeled through exponential decay calibrated by a one-year half-life, assuming that memory declines smoothly and continuously over time. The exponential form is preferred because its relative weighting does not depend on the total time span of the data, yielding a psychologically interpretable memory horizon rather than one that varies mechanically with the length of the observed series. The formal construction of this reference metric follows the procedures introduced in Kwak (2025).[5]

---

[4] Recent empirical research on public perceptions of police use-of-force provides important support for the plausibility of socially formed and socially shifting reference points in evaluations of state coercion. Mourtgos and Adams (2020) show that substantial portions of the American public disapprove of police actions that are legally reasonable under prevailing constitutional standards, and that such disapproval varies systematically across groups and over time, indicating that citizens evaluate coercion against socially constructed expectations of what appears acceptable rather than against fixed legal thresholds alone. Complementing this evidence, research on the pathologic adaptation model of violence exposure demonstrates that repeated exposure to violence can produce emotional desensitization and increase later violent behavior, suggesting that chronic exposure normalizes coercion and alters affective responses to it (Gaylord-Harden et al., 2017). Considered together, these findings support the claim that evaluations of state violence are reference-dependent: they are anchored to collectively formed standards, and those standards themselves can shift through cumulative exposure and normalization processes, making it plausible that protest repression is interpreted relative to historically conditioned expectations rather than to stable, absolute criteria.

[5] To construct individuals' reference point of expected repression at time $t$, this study employs a time-weighted average of prior repression levels. The exponential decay reference point is defined as

$$RP_t^{\exp} = \frac{\sum_{s=1}^{t-1} w_{ts}^{\exp} R_s}{\sum_{s=1}^{t-1} w_{ts}^{\exp}}, \quad \text{where} \quad w_{ts}^{\exp} = e^{-\lambda d_{ts}}, \quad d_{ts} = t - s. \tag{1}$$

Here $R_s$ denotes the level of repression at past event $s$, which is the mean number of fatalities for the fatalities-based measure and the binary excessive-force indicator for the intensity-based measure; a separate reference point is constructed for each measure from its own past values using the identical weighting scheme. Past repression levels are discounted at a constant proportional rate governed by the decay parameter $\lambda$, which is determined by a user-defined half-life. The exponential structure ensures that relative weighting does not depend on the total time span of the dataset, thereby providing a

The key independent variable is the degree of misalignment between the government's most recent action and this reference point. Formally, the gap is defined as

$$G_{it} = R_{i,t-1} - RP_{it},$$

where $RP_{it}$ denotes the reference baseline implied by repression history. A positive value of $G_{it}$ indicates that the prior level of repression exceeded what would be expected given recent patterns, whereas a negative value indicates that it fell short. This construction directly reflects the theoretical mechanism: what matters is not the absolute level of force, but whether current behavior appears excessive or insufficient relative to a socially recognized baseline. As with the fatalities-based outcome, a signed logarithmic transformation is applied to the gap constructed from fatalities, while the gap based on the excessive-force indicator enters without transformation.

In the empirical models, the coefficient on $G_{it}$ captures the extent to which governments adjust subsequent repression in response to this misalignment. A negative coefficient indicates corrective adjustment, meaning that higher-than-expected prior repression is followed by decreases in coercion, while lower-than-expected repression is followed by increases. The theoretical expectation is therefore a negative association, consistent with a stabilizing dynamic in which governments counteract departures from the socially recognized baseline rather than allowing deviations to persist or escalate.

## Identification Diagnostics: Mechanical Coupling and Serial Dependence

A potential concern in this design is that both the outcome and the key explanatory variable contain the same lagged repression term $R_{i,t-1}$, which may mechanically induce a negative

psychologically interpretable memory horizon. The one-year half-life implies that an event's contribution to the reference point falls to half its initial weight after roughly 365 days.

relationship even in the absence of a substantive adjustment process. This issue becomes clear when the two quantities are written explicitly:

$$\text{Dependent Variable: } \Delta R_{it} = R_{it} - R_{i,t-1}$$

$$\text{Independent Variable: } G_{it} = R_{i,t-1} - RP_{it}.$$

In both expressions, the term $R_{i,t-1}$ appears directly, entering the outcome with a negative sign and the explanatory variable with a positive sign. This means that the same prior repression level is mathematically embedded on both sides of the empirical relationship. Intuitively, when the previous level of repression $R_{i,t-1}$ is large, the quantity $\Delta R_{it} = R_{it} - R_{i,t-1}$ is mechanically more likely to be negative even if the current repression $R_{it}$ is unrelated to past expectations. At the same time, a large $R_{i,t-1}$ mechanically increases the value of the gap $G_{it} = R_{i,t-1} - RP_{it}$. Conversely, when $R_{i,t-1}$ is small, $\Delta R_{it}$ is mechanically more likely to be positive while $G_{it}$ becomes smaller or negative.

As a result, even if $R_{it}$ were statistically independent of the reference point $RP_{it}$, a negative association between $\Delta R_{it}$ and $G_{it}$ could arise purely from this shared arithmetic structure rather than from a genuine behavioral response. This is the regression–to–the–mean concern: large prior values tend to be followed by decreases and small prior values tend to be followed by increases, not necessarily because actors are adjusting strategically, but because of the way differences are constructed mathematically.

Beyond this direct coupling, because the reference point $RP_{it}$ is itself a weighted average of past repression, it also inherits the serial dependence of the repression series as a whole. A negative association between $\Delta R_{it}$ and the gap could therefore arise not only through the directly shared lagged term, but through the broader autoregressive structure of repression: in a persistent series, a change in repression is partly predictable from earlier values, and any function of those values, including the gap, will tend to correlate with the change for purely time-series reasons.

To benchmark the corrective association against both of these channels, the gap is entered alongside an explicit autoregressive control of order $p$,

$$\Delta R_{it} = \beta\, G_{it} + \sum_{k=1}^{p} \phi_k\, R_{i,t-k} + \mathbf{X}'_{it}\boldsymbol{\gamma} + \alpha_i + \delta_t + \varepsilon_{it},$$

where $R_{i,t-1}, \ldots, R_{i,t-p}$ are successive lagged levels of repression, $\mathbf{X}_{it}$ collects the controls, and $\alpha_i$ and $\delta_t$ are country and year fixed effects. By including the full set of lagged levels, this specification absorbs both the directly shared $R_{i,t-1}$ term and the broader serial dependence carried by the reference point. The model is estimated for $p = 1$ through $p = 10$. As deeper lags are included, the autoregressive terms absorb an increasing share of the predictable time-series variation in $\Delta R_{it}$. If the gap coefficient $\beta$ retained its sign and significance only in the absence of such controls, this would indicate that the corrective association is largely a restatement of the mechanical and autoregressive dependence between the gap and the change in repression. If instead $\beta$ remains negative and significant as $p$ increases, the corrective association reflects variation in the gap that neither the directly shared lagged term nor the broader autoregressive structure can account for.

Two further specifications address the same mechanical and autoregressive dependence. The first is a buffered reference point, which excludes the $k$ most recent events, including $R_{i,t-1}$, from the baseline before recomputing the gap, so that the directly shared lagged term cannot drive the association; it is estimated for buffer widths of $k = 3$, 5, and 10. The second is a separated specification, which enters the lagged level $R_{i,t-1}$ and the reference point as distinct regressors rather than as a single gap, isolating whether the reference point predicts current repression net of the autoregressive term.

## Asymmetry of Correction

Hypothesis 1b predicts that correction is not symmetric across the two directions of deviation: over-repression, where prior force exceeded the reference point, should draw a stronger

corrective response than under-repression of comparable magnitude. Testing this requires separating the gap by its sign rather than entering it as a single term. The gap is therefore decomposed into a positive part, which records deviations above the reference point (over-repression), and a negative part, which records deviations below it (under-repression); each event contributes to only one of the two. As with the pooled gap, each part is signed-log transformed for the fatalities-based measure and left untransformed for the excessive-force measure. Both parts enter the model jointly, replacing the single gap term in the specification above, so that their coefficients capture the strength of correction following over- and under-repression, respectively. A negative value of each indicates corrective adjustment in that direction, and Hypothesis 1b implies that correction of over-repression is the stronger of the two.

The sign-split specification also speaks to the regression-to-the-mean concern raised above. Pure regression to the mean is symmetric: in a mean-reverting series, deviations above and below the local average are pulled back at the same rate, so it implies $\beta^+ = \beta^-$. A corrective process driven by the political costs of misalignment carries no such restriction, since the costs of over- and under-repression differ. A statistically distinguishable asymmetry between the two coefficients therefore points beyond symmetric mean reversion, and its presence suggests that the corrective association does not reduce to the mean reversion in the repression series alone. To probe the robustness of this logic, a block-bootstrap simulation reshuffles the repression series within each country, preserving its distribution while removing its directional ordering, to recover the asymmetry the distributional structure alone would produce.

## Observability

The design evaluates whether the corrective adjustment becomes more pronounced when the coercion is more publicly observable. Where a given act of repression is more observable, both excessive and insufficient force are more likely to be noticed and to carry reputational

consequences, so the same degree of misalignment $G_{it}$ should be associated with larger subsequent adjustment. Observability is operationalized using the size of the local population exposed to the event, measured as the number of residents living within one kilometer of the event location. A larger resident population constitutes a larger pool of potential witnesses to both the contention and the state's response, and thus a more legible setting in which deviations from the expected baseline are likely to be noticed. The theoretical expectation is therefore that the corrective association strengthens as the local exposed population increases.

## Background Insecurity

Finally, the design evaluates whether corrective adjustment becomes more pronounced under heightened background insecurity. Background insecurity is operationalized using two country-level indicators. The first is the country-year homicide rate (World Bank, 2023), which captures the prevalence of criminal violence. The second is the recent density of contentious activity, measured as the number of events recorded in the country over the preceding year, which captures political disorder. Together, they indicate the general climate of insecurity within which state decisions are interpreted.

The theoretical expectation is that as generalized violence rises and the perceived proximity of disorder increases, the political costs of deviating from the expected level of coercion intensify. In more insecure environments, both excessive and insufficient force are more likely to be interpreted as signs of instability. Consequently, the same degree of misalignment $G_{it}$ is expected to be associated with larger subsequent changes in repression, reflecting narrower margins for acceptable deviation and stronger incentives to counteract such departures.

## Control Variables

Across all specifications, the models include the following control variables. The Electoral Democracy Index Teorell et al. (2019), the Local Government Index from the Varieties of

Democracy dataset Coppedge et al. (2022), GDP per capita, and total population (Arel-Bundock, 2025) are included to account for differences in political and socioeconomic contexts that may influence how a socially intelligible baseline of coercion is formed. Democratic institutions affect the normative expectations surrounding state force; more democratic environments may sustain lower tolerance for visible coercion. The Local Government Index captures the extent to which subnational authorities are elected and institutionally autonomous, which may shape how coercion is implemented and attributed across levels of government. Economic development and demographic scale may influence state capacity and administrative reach, thereby affecting the background conditions under which coercion is perceived as routine or excessive. An indicator for whether the event was a nonviolent protest and a measure of civilian targeting are also included as alternative proxies for the immediate threat environment.[6]

# Results

Table 1 reports two-way fixed-effects estimates of the prediction in Hypothesis 1a, with standard errors clustered at the country level. The gap coefficient is negative and statistically significant under both measures of repression. The models thus provide consistent support for Hypothesis 1a: across both the lethal and the excessive-force measures, governments counteract prior departures from a historically grounded baseline rather than allowing those departures to persist or escalate. The buffered reference point and the separated specification, whose construction and full results are reported in the Appendix, address the mechanical coupling and serial dependence discussed above, and both continue to support this main

[6] Indicators of participant size are omitted from the main specifications. Although protest size may independently shape both threat perceptions and coercive responses, including participation measures substantially reduces the number of usable observations due to extensive missingness. Because the core empirical leverage of the design comes from within-country temporal comparisons, preserving sample size is important for avoiding compositional changes in the set of events used for estimation.

Table 1: Main Models of Change in Repression

| | Fatalities | Intensity |
|---|---|---|
| Gap (fatalities RP) | −0.768*** | |
| | (0.005) | |
| Gap (intensity RP) | | −0.649*** |
| | | (0.040) |
| Protest (t-1) | 0.001 | −0.002 |
| | (0.004) | (0.002) |
| Civilian Targeting (t-1) | 0.015 | −0.161*** |
| | (0.009) | (0.029) |
| Polyarchy (t-1) | 0.065* | −0.041*** |
| | (0.035) | (0.015) |
| Local Government (t-1) | 0.010 | 0.029*** |
| | (0.020) | (0.009) |
| GDP pc (t-1) | −0.094*** | −0.031*** |
| | (0.021) | (0.009) |
| Population (t-1) | 0.064 | 0.096*** |
| | (0.061) | (0.026) |
| Observations | 40 567 | 40 567 |
| Within R-sq | 0.384 | 0.438 |
| Adj. Within R-sq | 0.265 | 0.329 |

* p < 0.1, ** p < 0.05, *** p < 0.01.
Two-way country and year fixed effects. Standard errors clustered by country. Fatalities use the signed-log gap; intensity uses the excessive-force gap.

finding.

Figure 1 plots the gap coefficient as successively deeper autoregressive controls are introduced, from AR(1) through AR(10), for the fatalities and the intensity measures. Across all orders, the coefficient remains negative and significant at the 5% level. As deeper lags absorb a larger share of the predictable time-series variation in the change in repression, the magnitude of the coefficient attenuates modestly. This pattern indicates that the negative gap coefficient does not simply restate the mechanical coupling between the outcome and the explanatory variable or the broader serial dependence of the repression series. The full estimates are reported in the Appendix. The buffered reference point and the separated specification, also reported in the Appendix, are consistent with this pattern: the corrective association persists when the most recent events are excluded from the reference point and

when the lagged level and the reference point enter as distinct regressors.

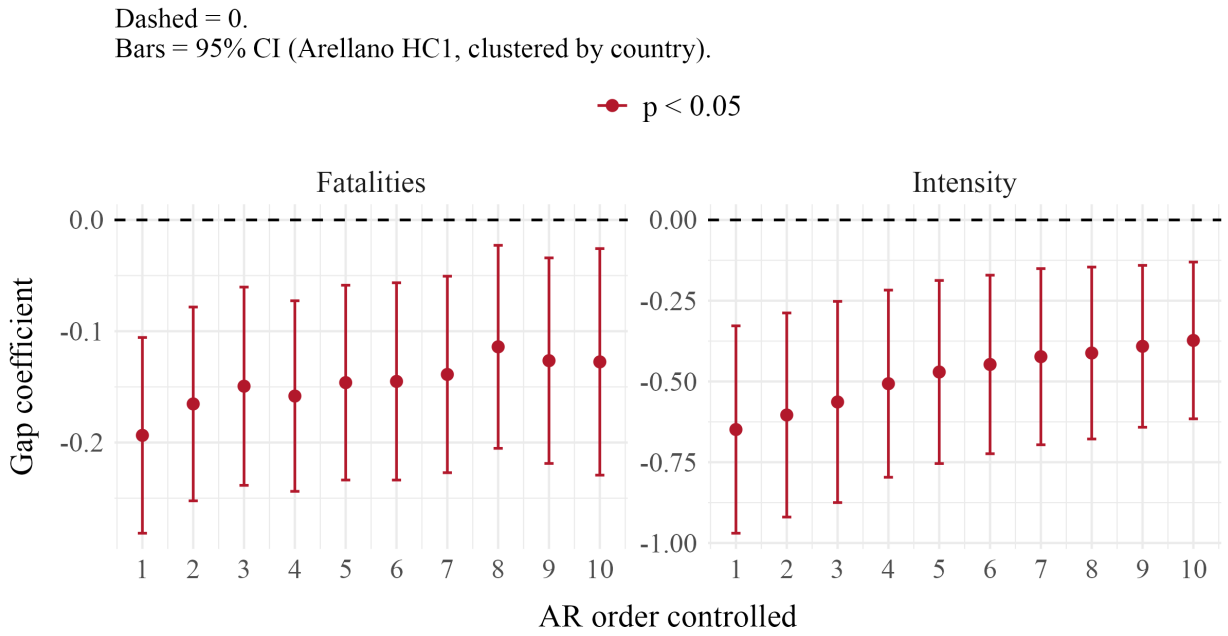



Figure 1: Gap Coefficient Under Increasing Autoregressive Controls

Figure 2 reports the sign-split estimates that decompose the gap into its over- and under-repression components, for the fatalities and the intensity measures. Both components are negative and significant, indicating corrective adjustment in either direction. The coefficient on over-repression is substantially larger in magnitude than the coefficient on under-repression under both measures, and a Wald test rejects the equality of the two at conventional levels in each case. This pattern supports Hypothesis 1b: governments counteract deviations above the reference point more sharply than deviations of comparable magnitude below it. A block-bootstrap simulation that removes the directional ordering of the repression series also confirms that the observed asymmetry exceeds what the distributional structure alone would produce under both measures, consistent with this interpretation; the full estimates and details are reported in the Appendix.

Figure 3 reports the interaction between the gap and the size of the local exposed population across the autoregressive benchmarks, from AR(1) through AR(10), for the fatalities and the intensity measures. For the intensity measure, the interaction is negative and sig-

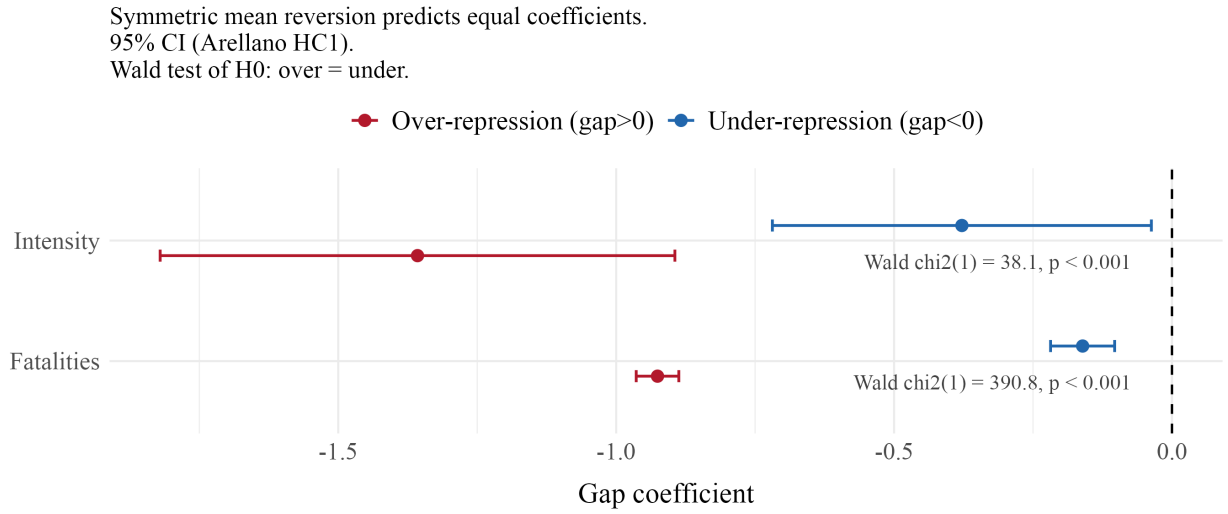



Figure 2: Sign-Split Gap Coefficients (Asymmetry Test)

nificant at the 5% level across all orders, and its magnitude is stable as deeper lags are introduced. This pattern supports Hypothesis 2: the corrective adjustment strengthens as the act of repression becomes more publicly observable. For the fatalities measure, by contrast, the interaction is not distinguishable from zero at any order, so the moderating role of observability receives support under the intensity measure but not the lethal one. The full estimates are reported in the Appendix.

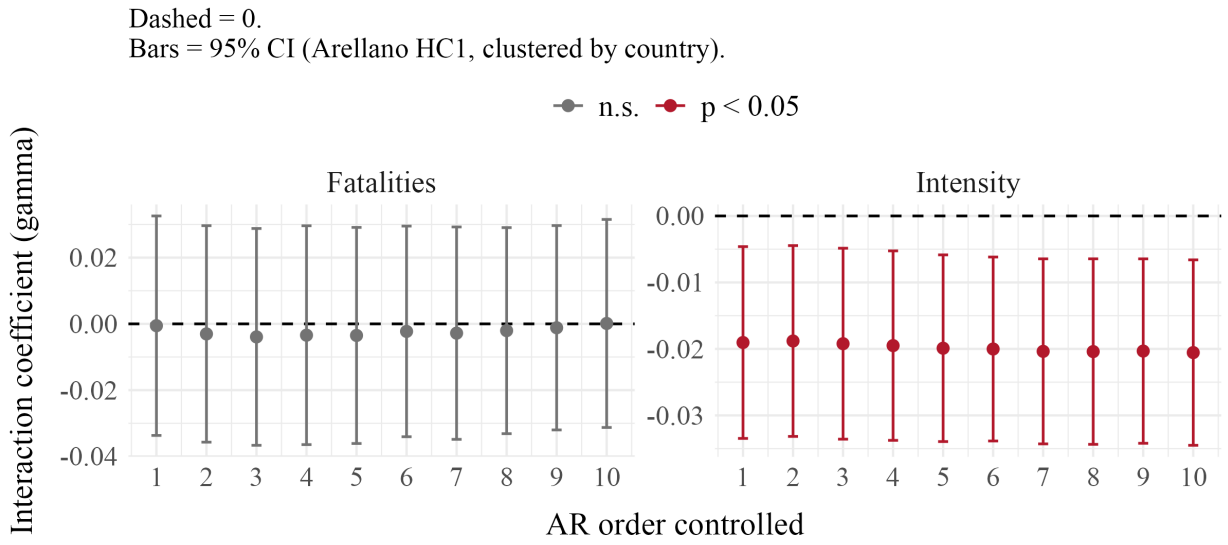



Figure 3: Observability (Local Population) Moderation Across Autoregressive Controls

Background insecurity is examined next, using two indicators. Figure 4 plots the inter-

action between the gap and the country-year homicide rate across the same autoregressive benchmarks. Here, the moderating effect holds under both measures: the interaction is negative throughout, remains significant at the 5% level at every order, and changes little as deeper lags enter. Figure 5 repeats the exercise with the recent density of contentious activity as the moderator. The interaction is again negative across both measures and at every order, significant throughout for the intensity measure and at all but two of the early orders for the fatalities measure. Hypothesis 3 is therefore supported across both proxies, with corrective adjustment intensifying where criminal violence is more prevalent and where contention has recently been denser. This evidence further distinguishes the reference-dependent account from the adaptive-policymaking and principal-agent alternatives, neither of which predicts that correction should sharpen specifically with rising ambient insecurity. The Appendix reports the full estimates.

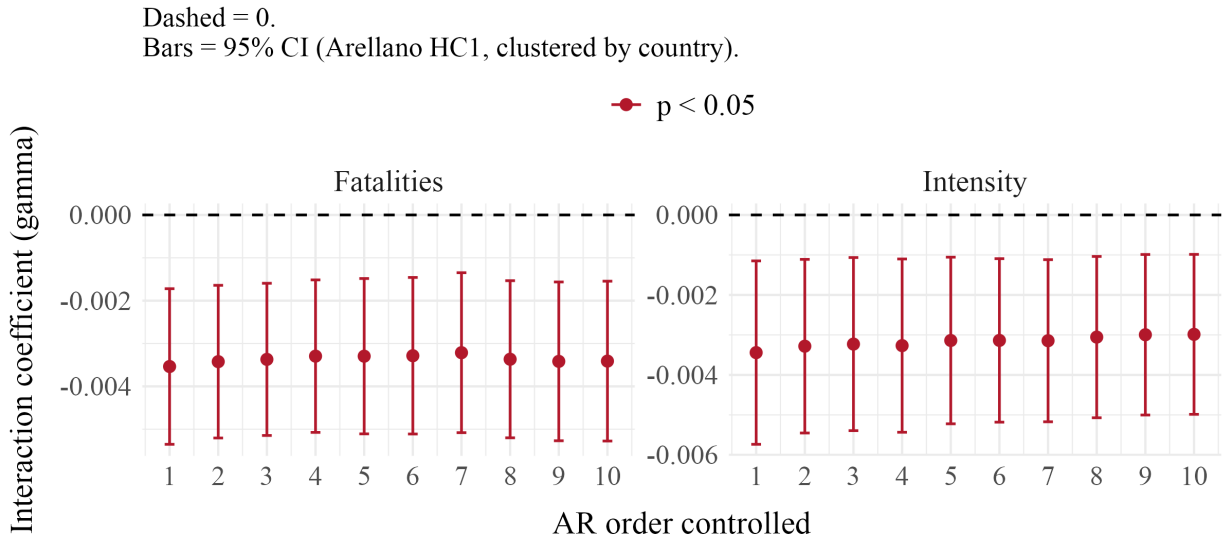



Figure 4: Insecurity (Homicide) Moderation Across Autoregressive Controls

In summary, the empirical evidence reveals a coherent adjustment pattern. Under both the lethal and the excessive-force measures, governments respond to prior deviations in repression with countervailing shifts, indicating a stabilizing dynamic oriented toward an internally accumulated baseline of coercion. This corrective association is not an artifact of the design: it persists as deeper autoregressive controls absorb the serial dependence of

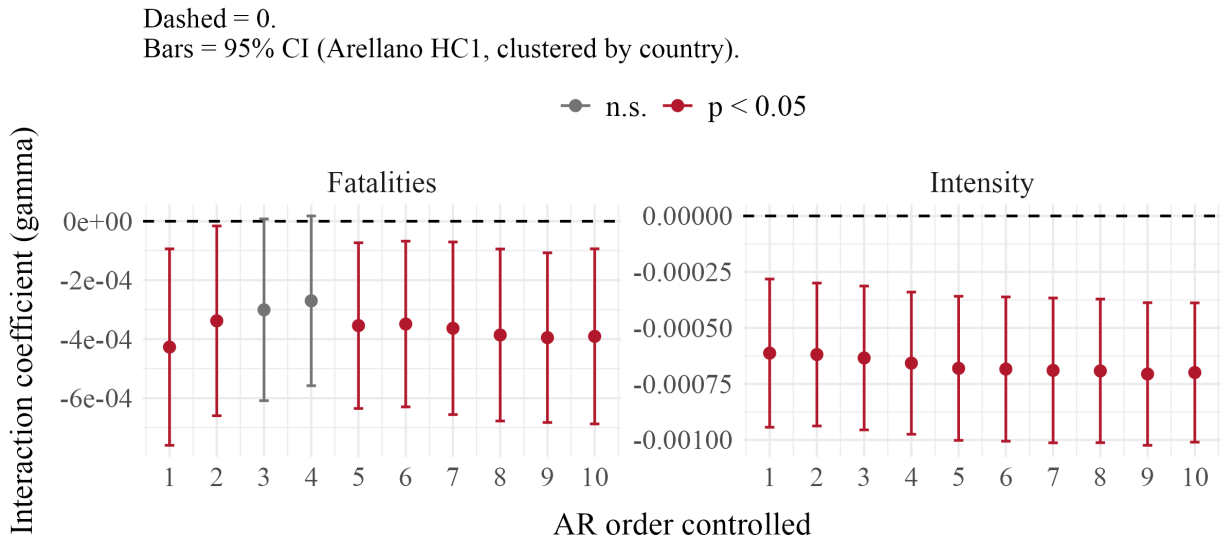



Figure 5: Insecurity (Recent Events) Moderation Across Autoregressive Controls

the repression series, when the most recent events are excluded from the reference point, and when the lagged level and the reference point enter as distinct regressors, indicating that the pattern does not reduce to the mechanical coupling between the outcome and the explanatory variable. The sign-split estimates show that correction is asymmetric, falling more sharply on over-repression than on under-repression, a pattern beyond what regression to the mean alone produces. The interaction analyses indicate that the corrective tendency intensifies where repression is more publicly observable, though only under the excessive-force measure, and where background insecurity is higher, across both the homicide rate and the recent density of contention. Considered together, these findings support the interpretation that repression during protest is best understood not as episodic escalation or regime-type reflex, but as a continual process of recalibration through which governments attempt to preserve a socially intelligible boundary between order and anarchy.

# Conclusion

This article contends that repression during protests should be viewed not merely as a reaction to threats posed by protestors but as a deliberate governance strategy aimed at

maintaining the clarity of state authority. Instead of seeing coercion as an isolated decision made independently in each instance of contention, the Hobbesian framework presented here emphasizes a persistent political challenge: the sovereign must ensure that the boundary between order and disorder remains socially understandable. Coercive actions that appear erratic, either unexpectedly severe or unusually lenient, risk blurring this boundary by making legitimate authority seem like arbitrary violence or, conversely, signaling weakness. Empirical data from ACLED support this reasoning across two complementary measures of repression, state-involved fatalities and the use of excessive force. When recent repression diverges from a country-specific baseline shaped by historical patterns of coercion, subsequent repression tends to move in the opposite direction, reflecting a systematic process of adjustment rather than a simple upward trend. This corrective mechanism intensifies where coercion is more publicly observable, at least under the excessive-force measure, and in contexts of heightened insecurity, consistent with the idea that greater visibility and widespread violence narrow the acceptable range of variation and increase the political costs associated with misjudging repression levels.

Theoretically, these findings suggest two broader implications for how repression should be conceptualized in the study of contentious politics. First, the results underscore that repression has an interpretive dimension that cannot be reduced to the immediate deterrence of challengers. A threat-centered framework remains essential for explaining why coercion occurs and why it sometimes intensifies, but it is incomplete to the extent that it assumes that repression should primarily shift with observable protest attributes and structural contingencies. The adjustment dynamic documented here provides a complementary logic: governments also manage the trajectory of coercion, because the meaning of force depends on its fit with historically accumulated expectations. This perspective helps explain why coercion sometimes changes even when protest characteristics do not: what is being stabilized is not only short-term control, but the recognizability of control. Second, the Hobbesian argument clarifies that the politics of repression are not only about punishing dissent but about sus-

taining the state as the actor that prevents society from sliding toward disorder. In this sense, repression functions as a signaling practice within governance, where volatility itself can be interpreted as dangerous. This emphasis shifts attention from whether governments repress to how they manage the credibility of sovereign authority over time.

The findings also offer policy implications, particularly for governments and international actors concerned with protest management and the prevention of escalation. The results of the study suggest that predictability and consistency may be at least as consequential as the absolute level. When coercion is experienced as volatile, swinging sharply between restraint and crackdowns, citizens may interpret the state as unstable, either because it appears unrestrained or because it appears incapable. Policies that standardize rules of engagement, strengthen accountability mechanisms, and improve transparency in crowd-control practices may reduce not only the intensity of violence but also the variance of coercive responses, thereby mitigating the political risks associated with unpredictable repression. The moderating role of background insecurity further implies that interventions aimed at reducing generalized violence, such as strengthening community safety, reducing homicide rates, and building trust in law enforcement, may indirectly lower the incentives for sharp corrective swings in protest repression by widening the range within which state responses remain socially intelligible.

The findings carry implications for activists and protest organizers as well. Activists may consider not only the absolute level of state violence but its position relative to recent patterns. Sharp deviations from the reference point can shift how future repression unfolds by triggering corrective adjustment. In some contexts, strategic restraint in moments of anticipated escalation may reduce the likelihood of sustained crackdowns by lowering the perceived need for correction. In others, publicizing unusually harsh repression may increase the political costs of further escalation, reinforcing pressures toward recalibration. While activists do not control state responses, understanding repression as a trajectory rather than a static level highlights that protest–repression dynamics are iterative, and that expectations

can shape not only mobilization but the evolution of coercion itself.

Several limitations remain, indicating directions for future research. First, the empirical analysis relies on two measures, state-involved fatalities and the use of excessive force. While both provide relatively comparable and substantively meaningful indicators of severe coercion, they capture only the most visible and lethal dimensions of repression and may understate forms of nonlethal violence, harassment, arrests, and coercive intimidation. Future work should extend the reference-point framework to broader repertoires of repression, including nonlethal force and administrative coercion, and evaluate whether similar adjustment dynamics characterize these tactics. Second, the reference point is an empirical proxy: it is constructed from observed coercive histories, whereas the theoretical construct is a socially formed expectation. Behavioral histories may diverge from perceived expectations through selective memory, the disproportionate salience of widely publicized episodes, partisan framing, and international attention, none of which a behavioral summary fully captures. Incorporating measures of the information environment, such as media content, social media salience, or elite discourse, would help clarify how reference points form and shift, and represents an important task for future research. Third, the argument treats governments as unitary actors, even though repression decisions often involve executives, police, military, and local authorities with potentially divergent preferences. Adjustment dynamics may thus vary depending on how centralized coercive control is. In highly centralized systems, corrective recalibration may be implemented consistently, whereas in fragmented systems, volatility may reflect coordination failures or competing elite strategies rather than deliberate adjustment.

Finally, the scope conditions outlined in the alternative argument invite further empirical scrutiny. The theory presumes that the existing boundary between order and anarchy remains normatively intelligible, so that limiting departures from expected coercion carries political value. In episodes of regime crisis or revolutionary contention, governments may instead seek to reconstruct the boundary by cultivating insecurity. Identifying when and

how states switch from counteracting departures to the strategic production of disorder is an important next step.

# Appendix

## Descriptive Statistics

Table 2: Descriptive Statistics of Model Variables

| Variable | N | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| *Dependent variables* | | | | | | |
| Δ Repression (fatalities) | 52,830 | -0.000 | 0.376 | -5.303 | 0.000 | 5.303 |
| Δ Repression (intensity) | 52,830 | -0.000 | 0.164 | -0.693 | 0.000 | 0.693 |
| *Key regressors* | | | | | | |
| Gap (fatalities RP) | 52,830 | -0.007 | 0.320 | -3.420 | -0.010 | 5.300 |
| Gap (intensity RP) | 52,830 | 0.018 | 0.127 | -0.378 | -0.002 | 0.693 |
| *Controls* | | | | | | |
| Protest (t-1) | 52,830 | 0.489 | 0.472 | 0.000 | 0.500 | 1.000 |
| Civilian targeting (t-1) | 52,830 | 0.041 | 0.186 | 0.000 | 0.000 | 1.000 |
| Polyarchy (t-1) | 41,400 | 0.497 | 0.236 | 0.014 | 0.484 | 0.915 |
| Local government (t-1) | 40,957 | 0.680 | 0.347 | 0.000 | 0.869 | 0.997 |
| GDP per capita (t-1) | 41,000 | 8.276 | 1.285 | 5.514 | 8.138 | 11.616 |
| Population (t-1) | 41,218 | 17.590 | 1.574 | 12.310 | 17.538 | 21.087 |
| *Moderators* | | | | | | |
| Homicide rate | 27,187 | 8.165 | 11.101 | 0.067 | 2.967 | 54.359 |
| Recent events (t-1) | 52,830 | 95.415 | 87.891 | 1.000 | 74.000 | 362.000 |
| Local population (log) | 46,038 | 9.415 | 1.459 | 0.000 | 9.743 | 12.753 |

Statistics computed on the estimation sample (df_mod).
Gap variables are on the signed-log scale at which they enter the models.

## Results from Main Models

Table 3: Main Gap Coefficient Across Autoregressive Benchmarks

| AR order | Fatalities | Intensity |
|---|---|---|
| AR(1) | -0.194** | -0.649** |
| | (0.045) | (0.164) |
| AR(2) | -0.165** | -0.604** |
| | (0.044) | (0.161) |
| AR(3) | -0.149** | -0.563** |
| | (0.045) | (0.159) |
| AR(4) | -0.158** | -0.507** |
| | (0.044) | (0.148) |
| AR(5) | -0.146** | -0.471** |
| | (0.045) | (0.145) |
| AR(6) | -0.145** | -0.447** |
| | (0.045) | (0.141) |
| AR(7) | -0.139** | -0.423** |
| | (0.045) | (0.139) |
| AR(8) | -0.114** | -0.412** |
| | (0.047) | (0.136) |
| AR(9) | -0.126** | -0.391** |
| | (0.047) | (0.128) |
| AR(10) | -0.128** | -0.373** |
| | (0.052) | (0.124) |
| Observations (AR1 / AR10) | 40,567 / 39,282 | |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Two-way country and year fixed effects. Standard errors clustered by country.

Each cell reports the coefficient with its standard error below.

Gap coefficient with R(t-1) through R(t-p) controlled, p = 1 to 10.

Significance marked at the 5% level from the 95% confidence interval.

## Results from Buffered Reference Point

The buffered reference point removes the direct coupling between the outcome and the gap, both of which contain $R_{i,t-1}$. The $k$ most recent events, including $R_{i,t-1}$, are excluded from the exponentially weighted baseline before the gap is recomputed. Exclusion is counted in events, so the removal of $R_{i,t-1}, \ldots, R_{i,t-k}$ is exact. The buffered gap enters signed-log transformed for the fatalities measure and untransformed for the excessive-force measure, and the diagnostic is estimated for $k = 3$, 5, and 10 under the same controls and fixed effects as the main models. A gap coefficient that survives as $k$ widens cannot restate the shared lagged term.

Table 4: Buffered Reference Point: Gap Coefficient Only

| | fat_buf3 | int_buf3 | fat_buf5 | int_buf5 | fat_buf10 | int_buf10 |
|---|---|---|---|---|---|---|
| Gap (buffer k=3, fat) | −0.757*** | | | | | |
| | (0.005) | | | | | |
| Gap (buffer k=3, int) | | −0.557*** | | | | |
| | | (0.038) | | | | |
| Gap (buffer k=5, fat) | | | −0.753*** | | | |
| | | | (0.005) | | | |
| Gap (buffer k=5, int) | | | | −0.574*** | | |
| | | | | (0.038) | | |
| Gap (buffer k=10, fat) | | | | | −0.745*** | |
| | | | | | (0.005) | |
| Gap (buffer k=10, int) | | | | | | −0.551*** |
| | | | | | | (0.038) |
| Observations | 40 538 | 40 538 | 40 509 | 40 509 | 40 405 | 40 405 |
| Within R-sq | 0.379 | 0.437 | 0.377 | 0.437 | 0.372 | 0.438 |
| Adj. Within R-sq | 0.259 | 0.328 | 0.257 | 0.329 | 0.252 | 0.330 |

* p < 0.1, ** p < 0.05, *** p < 0.01.
Two-way country and year fixed effects. Standard errors clustered by country.
Key coefficient(s) only; controls and AR lags estimated but not shown.
Buffer k excludes the k most recent events (including R(t-1)) from the reference point.

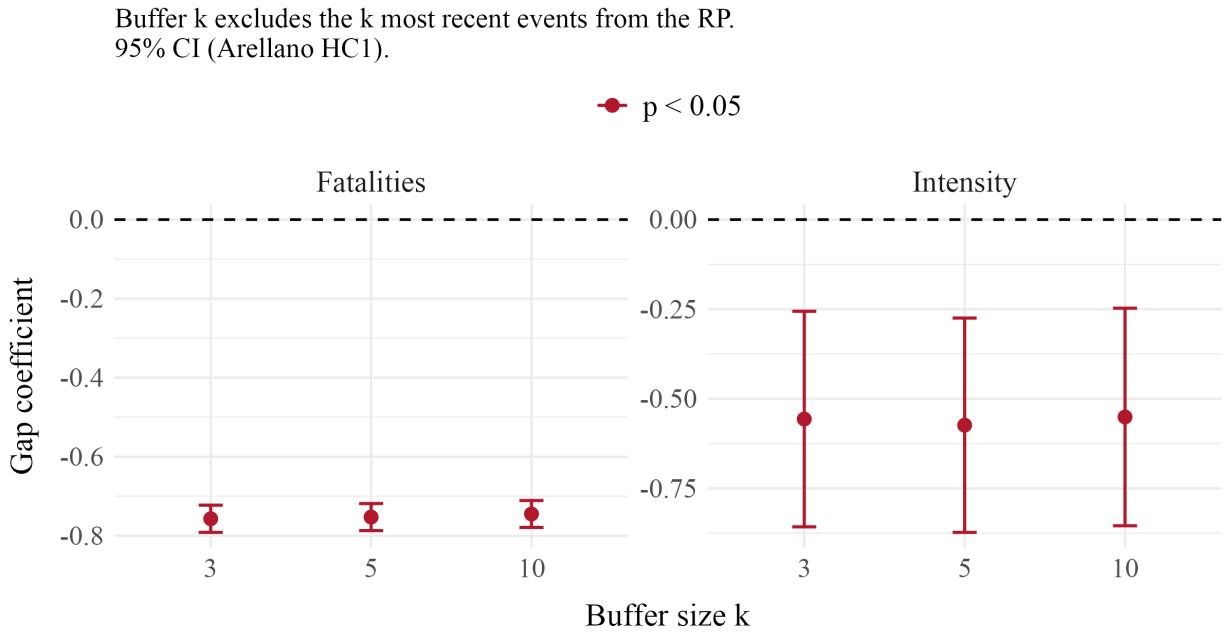



Figure 6: Buffered Reference Point

## Results from Separated Specification

The separated specification enters the lagged level and the reference point as distinct regressors rather than as a single gap,

$$R_{it} = \gamma\, R_{i,t-1} + \delta\, RP_{it}^{\text{buffered}} + \mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \delta_t + \varepsilon_{it},$$

with the buffered reference point so that it shares no events with $R_{i,t-1}$. Here $\gamma$ captures serial dependence and $\delta$ the distinct pull of the baseline net of it; a corrective process implies $\delta \neq 0$. Estimated for $k = 3$, 5, and 10, this is the most demanding test: $R_{i,t-1}$ and the reference point remain collinear after buffering, which inflates the standard errors.

Table 5: Separated Specification: Reference-Point Effect Net of R(t-1)

| | fat_buf3 | fat_buf5 | fat_buf10 | int_buf3 | int_buf5 | int_buf10 |
|---|---|---|---|---|---|---|
| $\delta$ (RP), fat k=3 | 0.132*** | | | | | |
| | (0.015) | | | | | |
| $\delta$ (RP), fat k=5 | | 0.132*** | | | | |
| | | (0.015) | | | | |
| $\delta$ (RP), fat k=10 | | | 0.111*** | | | |
| | | | (0.015) | | | |
| $\delta$ (RP), int k=3 | | | | 0.938*** | | |
| | | | | (0.047) | | |
| $\delta$ (RP), int k=5 | | | | | 0.999*** | |
| | | | | | (0.048) | |
| $\delta$ (RP), int k=10 | | | | | | 0.945*** |
| | | | | | | (0.048) |
| $\gamma$ R(t-1), fat | 0.148*** | 0.150*** | 0.152*** | | | |
| | (0.006) | (0.006) | (0.006) | | | |
| $\gamma$ R(t-1), int | | | | 0.107*** | 0.107*** | 0.109*** |
| | | | | (0.006) | (0.006) | (0.006) |
| Observations | 40 538 | 40 509 | 40 405 | 40 538 | 40 509 | 40 405 |
| Within R-sq | 0.027 | 0.027 | 0.027 | 0.028 | 0.030 | 0.028 |
| Adj. Within R-sq | −0.161 | −0.161 | −0.160 | −0.159 | −0.158 | −0.158 |

* p < 0.1, ** p < 0.05, *** p < 0.01.
Two-way country and year fixed effects. Standard errors clustered by country.
$\delta$ is the coefficient on the reference point; $\gamma$ is the coefficient on R(t-1).
Key coefficient(s) only; controls and AR lags estimated but not shown.
Most conservative test; high R(t-1)/RP collinearity limits precision.

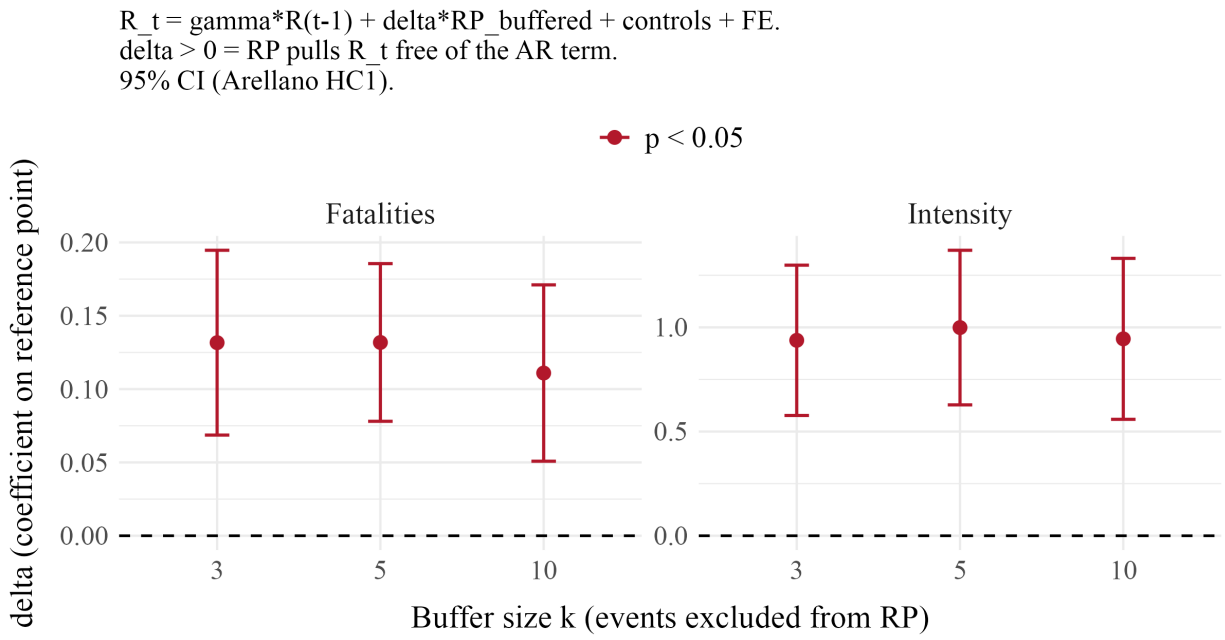



Figure 7: Separated Specification

## Asymmetry Test

Table 6: Asymmetry of Correction: Sign-Split Gap Coefficients

| | Fatalities | Intensity |
|---|---|---|
| Gap+ (over-repression, fatalities) | −0.926*** | |
| | (0.007) | |
| Gap- (under-repression, fatalities) | −0.161*** | |
| | (0.017) | |
| Gap+ (over-repression, intensity) | | −1.357*** |
| | | (0.061) |
| Gap- (under-repression, intensity) | | −0.378*** |
| | | (0.044) |
| Observations | 40 567 | 40 567 |
| Within R-sq | 0.410 | 0.442 |
| Adj. Within R-sq | 0.296 | 0.334 |

* p < 0.1, ** p < 0.05, *** p < 0.01.
Two-way country and year fixed effects. Standard errors clustered by country.
Key coefficient(s) only; controls and AR lags estimated but not shown.
Gap+ = max(gap, 0); Gap- = min(gap, 0). Wald test of equality reported in text.

The sign-split result in Table 6 could in principle arise from the distributional structure of the repression series rather than from a behavioral corrective process. To assess whether the observed asymmetry exceeds this mechanical baseline, a block-bootstrap null reorders the repression series within each country in contiguous blocks on a fixed event grid, preserving its distribution and short-run dependence while removing the directional ordering through which a corrective process would operate, and recomputes the sign-split coefficients on each reshuffled series across one thousand replications. As Figure 8 shows, the null distributions center below zero, confirming that part of the asymmetry is mechanical; the observed asymmetry, however, lies in the left tail of each ($p < 0.001$ for fatalities, $p = 0.002$ for intensity), indicating that it is larger than the bounded, skewed structure of the series can generate on its own.

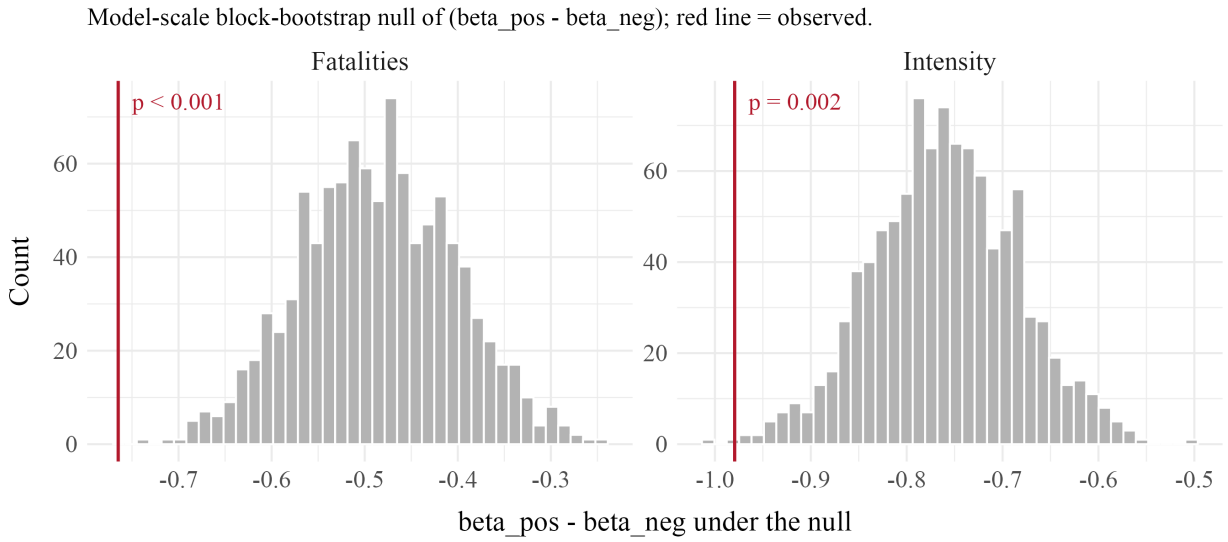



Figure 8: Block-Bootstrap Placebo for the Sign-Split Asymmetry

## Moderation Models

Table 7: Moderation Interactions Across Autoregressive Benchmarks

| | Local population | | Homicide rate | | Recent events | |
|---|---|---|---|---|---|---|
| AR order | Fatalities | Intensity | Fatalities | Intensity | Fatalities | Intensity |
| AR(1) | -0.001 | -0.019*** | -0.004*** | -0.003*** | -0.000** | -0.001*** |
| | (0.017) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(2) | -0.003 | -0.019** | -0.003*** | -0.003*** | -0.000** | -0.001*** |
| | (0.017) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(3) | -0.004 | -0.019*** | -0.003*** | -0.003*** | -0.000* | -0.001*** |
| | (0.017) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(4) | -0.003 | -0.020*** | -0.003*** | -0.003*** | -0.000* | -0.001*** |
| | (0.017) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(5) | -0.003 | -0.020*** | -0.003*** | -0.003*** | -0.000** | -0.001*** |
| | (0.017) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(6) | -0.002 | -0.020*** | -0.003*** | -0.003*** | -0.000** | -0.001*** |
| | (0.016) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(7) | -0.003 | -0.020*** | -0.003*** | -0.003*** | -0.000** | -0.001*** |
| | (0.016) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(8) | -0.002 | -0.020*** | -0.003*** | -0.003*** | -0.000*** | -0.001*** |
| | (0.016) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(9) | -0.001 | -0.020*** | -0.003*** | -0.003*** | -0.000*** | -0.001*** |
| | (0.016) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| AR(10) | 0.000 | -0.021*** | -0.003*** | -0.003*** | -0.000*** | -0.001*** |
| | (0.016) | (0.007) | (0.001) | (0.001) | (0.000) | (0.000) |
| Obs. (AR1) | 34,034 | | 27,030 | | 40,567 | |
| Obs. (AR10) | 33,171 | | 26,170 | | 39,282 | |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Two-way country and year fixed effects. Standard errors clustered by country.
Each cell reports the interaction coefficient with its standard error below.
Interaction of the gap with each moderator (local population, homicide rate, recent events), with R(t-1) through R(t-p) controlled, p = 1 to 10.
Negative interaction = correction strengthens as the moderator rises.

# References


Abizadeh, Arash. 2011. "Hobbes on the Causes of War: A Disagreement Theory." *American Political Science Review* 105(2):298–315.

AMNESTY. 2006. "Five Years after the G8 Genoa Policing Operations: Italian Authorities Must Take Concrete Action to Prevent and Prosecute Police Brutality in all Circumstances.".

Arel-Bundock, Vincent. 2025. "WDI: World Development Indicators and Other World Bank Data.".
**URL:** *https://CRAN.R-project.org/package=WDI*

Artabe, Eugenia, Samantha Chapa, Leah Sparkman and Patrick E. Shea. 2023. "External Threats, Capacity, and Repression: How the Threat of War Affects Political Development and Physical Integrity Rights." *British Journal of Political Science* 53(4):1311–1327.

Aytaç, S. Erdem, Luis Schiumerini and Susan Stokes. 2017. "Protests and Repression in New Democracies." *Perspectives on Politics* 15(1):62–82.

Carey, Sabine C. 2006. "The Dynamic Relationship Between Protest and Repression." *Political Research Quarterly* 59(1):1–11.

Carey, Sabine C. 2010. "The Use of Repression as a Response to Domestic Dissent." *Political Studies* 58(1):167–186.

Chasan, Daniel Jack and Christianne Walker. 2000. "Out of Control: Seattle's Flawed Response to Protests Against the World Trade Organization: a Special Report.".

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig and Daniel Ziblatt. 2022. "V-dem codebook v12.".
**URL:** *https://www.v-dem.net/data/the-v-dem-dataset/*

Cunningham, David, Geoff Ward and Peter B. Owens. 2019. "Configuring Political Repression: Anti-Civil Rights Enforcement in Mississippi*." *Mobilization: An International Quarterly* 24(3):319–343.

Dahlum, Sirianne. 2023. "Joining Forces: Social Coalitions and Democratic Revolutions." *Journal of Peace Research* 60(1):42–57.

Danneman, Nathan and Emily Hencken Ritter. 2014. "Contagious Rebellion and Preemptive Repression." *Journal of Conflict Resolution* 58(2):254–279.

Davenport, Christian. 1995. "Multi-Dimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions." *American Journal of Political Science* 39(3):683.

Davenport, Christian. 1996. "The Weight of the Past: Exploring Lagged Determinants of Political Repression." *Political Research Quarterly* 49(2):377–403.

Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10(1):1–23.

Della Porta, Donatella and Herbert Reiter, eds. 1998. *Policing Protest: The Control of Mass Demonstrations in Western Democracies*. Number v. 6 *in* "Social movements, protest, and contention" Minneapolis: University of Minnesota Press.

Earl, Jennifer. 2011. "Political Repression: Iron Fists, Velvet Gloves, and Diffuse Control." *Annual Review of Sociology* 37(1):261–284.

Earl, Jennifer, Sarah A. Soule and John D. McCarthy. 2003. "Protest Under Fire? Explaining the Policing of Protest." *American Sociological Review* 68(4):581–606.

Earl, Jennifer, Thomas V. Maher and Jennifer Pan. 2022. "The Digital Repression of Social Movements, Protest, and Activism: A Synthetic Review." *Science Advances* 8(10):eabl8198.

Elzoughby, Moaaz. 2011. "The Dynamics of Egypt's Protest: An Inside View." *Arab Reform Brief* 3.

Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297–318.

Ferrara, Federico. 2003. "Why Regimes Create Disorder: Hobbes's Dilemma During a Rangoon Summer." *Journal of Conflict Resolution* 47(3):302–325.

Gartner, Scott Sigmund and Patrick M. Regan. 1996. "Threat and Repression: The Non-Linear Relationship Between Government and Opposition Violence." *Journal of Peace Research* 33(3):273–287.

Gaylord-Harden, Noni K., Suzanna So, Grace J. Bai, David B. Henry and Patrick H. Tolan. 2017. "Examining the Pathologic Adaptation Model of Community Violence Exposure in Male Adolescents of Color." *Journal of Clinical Child & Adolescent Psychology* 46(1):125–135.

Heffington, Colton. 2021. "External Threat and Human Rights: How International Conflict Leads to Domestic Repression." *Journal of Human Rights* 20(1):2–19.

Hendrix, Cullen S. and Idean Salehyan. 2017. "A House Divided: Threat Perception, Military Factionalism, and Repression in Africa." *Journal of Conflict Resolution* 61(8):1653–1681.

Hultquist, Agatha Skierkowski, Victor Asal and R. Karl Rethemeyer. 2025. "Why Do States Violently Repress Some Political Organizations, but Not Others? Evidence from Ethnopolitical Actors in the Middle East and North Africa." *Studies in Conflict & Terrorism* 48(5):557–580.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 42(2):263–292.

Karreth, Johannes, Jaroslav Tir and Douglas M Gibler. 2022. "Latent Territorial Threat and Democratic Regime Reversals." *Journal of Peace Research* 59(2):197–212.

Knuckey, Sarah, Katherine Glenn and Emi MacLean. 2012. "Supressing Protest: Human Rights Violations in the US Response to Occupy Wall Street." *NYU School of Law* .

Kricheli, Ruth, Yair Livne and Beatriz Magaloni. 2010. "Taking to the Streets: Theory and Evidence on Protest under Authoritarianism." *APSA 2010 Annual Meeting Paper* .

Kwak, Hansol. 2025. "When Repression Surprises: Contextual Deviations and Protest Mobilization." *Research & Politics* 12(3):20531680251381521.

Malby, Steven. 2010. Homicide. In *International Statistics on Crime and Criminal Justice*, ed. Stefan Harrendorf, Heiskanen Markku and Steven Malby. Viena: United Stations Office on Drugs and Crime pp. 7–20.

Manekin, Devorah, Tamar Mitts and Yael Zeira. 2024. "The Politics of Allyship: Multiethnic Coalitions and Mass Attitudes Toward Protest." *Proceedings of the National Academy of Sciences* 121(19):e2314653121.

Mercer, Jonathan. 2005. "Prospect Theory and Political Science." *Annual Review of Political Science* 8(1):1–21.

Mourtgos, Scott M. and Ian T. Adams. 2020. "Assessing Public Perceptions of Police Use-of-Force: Legal Reasonableness and Community Standards." *Justice Quarterly* 37(5):869–899.

Nassauer, Anne. 2019. *Situational Breakdowns: Understanding Protest Violence and Other Surprising Outcomes*. Oxford studies in culture and politics New York, NY: Oxford university press.

Nordås, Ragnhild and Christian Davenport. 2013. "Fight the Youth: Youth Bulges and State Repression." *American Journal of Political Science* 57(4):926–940.

Poe, Steven C. and C. Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4):853–872.

Raleigh, Clionadh, Rew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47(5):651–660.

Regan, Patrick M and Errol A Henderson. 2002. "Democracy, Threats and Political Repression in Developing Countries: Are Democracies Internally Less Violent?" *Third World Quarterly* 23(1):119–136.

Ritter, Emily Hencken. 2014. "Policy Disputes, Political Survival, and the Onset and Severity of State Repression." *Journal of Conflict Resolution* 58(1):143–168.

Stephan, Maria J. and Erica Chenoweth. 2008. "Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict." *International Security* 33(1):7–44.

Teorell, Jan, Michael Coppedge, Staffan Lindberg and Svend-Erik Skaaning. 2019. "Measuring Polyarchy Across the Globe, 1900–2017." *Studies in Comparative International Development* 54(1):71–95.

Vis, Barbara. 2011. "Prospect Theory and Political Decision Making." *Political Studies Review* 9(3):334–343.

World Bank. 2023. "Intentional Homicides (per 100,000 people)." *World Bank* .
**URL:** *https://data.worldbank.org/indicator/VC.IHR.PSRC.P5*

Yameogo, Souleymane. 2025. "Proximity to Riots: Spatial Exposure and Attitude Toward the Police in Africa." *Social Forces* p. soaf203.

Young, Joseph K. 2013. "Repression, Dissent, and the Onset of Civil War." *Political Research Quarterly* 66(3):516–532.

Yuen, Samson and Edmund W Cheng. 2017. "Neither Repression Nor Concession? A Regime's Attrition against Mass Protests." *Political Studies* 65(3):611–630.