Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Political Camouflage: The Visibility of Dissent under Democratic Backsliding

Valentin Schnellmann
CREMA
Andre Briviba
CREMA
Bruno S. Frey
bruno.frey@bsfrey.ch

CREMA

Research Article

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11701236/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Political Camouflage:
The Visibility of Dissent under Democratic Backsliding

Valentin Schnellmann[a], Andre Briviba[a,c] & Bruno S. Frey[a,b,*]

[a]*CREMA, Center for Research in Economics, Management and the Arts, Südstrasse 11, Zürich, Switzerland*
[b]*Permanent Visiting Professor, University of Basel, Peter-Merian-Weg 6, Basel, Switzerland*
[c]*University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland*

Abstract

We study how institutional deterioration affects not only whether citizens engage in political activity, but also how visible that activity becomes. We use political camouflage to describe opposition expressed through lower-visibility channels when public exposure is costly. We develop a model of channel choice in which citizens sort across inactivity, concealed activity and public dissent, and in which an institutional rupture can raise the cost of public exposure, restrict access to concealed channels, or increase political activation. Each component implies a different combination of the two observable channels. We examine nine institutional ruptures between 2013 and 2023 using activity on the Tor anonymity network, recorded protest events, and case-specific augmented synthetic controls. The nine cases do not produce a common response. Hong Kong provides the clearest evidence of camouflage, where concealed activity rose as public protest contracted, while Cambodia displays the same directional pattern with weaker support. Myanmar instead shows public-led mobilisation, and in Belarus and Russia declining Tor use coincides with deteriorating network accessibility, indicating restricted access rather than reduced demand for concealment. Observed protest alone therefore does not identify the political response: a decline in public dissent need not imply that opposition has receded.

[*]Corresponding author:

Bruno S. Frey; E-mail: bruno.frey@bsfrey.ch; ORCID ID: 0000-0001-5380-6655

Valentin Schnellmann: valentinsascha.schnellmann@student.unisg.ch;
Andre Briviba: andre.briviba@unifr.ch

Acknowledgements:

We thank Marco Broger, Fabian Rullmann, Heinz Buhofer, Reiner Eichenberger, Gerd Folkers, René L. Frey, Jana Gallus, Kai Konrad, Stephan Meier, Christoph Schaltegger, Hans-Werner Sinn, Fritz Schneider, and Margit Osterloh for their encouragement and inspiring comments.

1. Introduction

Democratic backsliding has become a defining feature of the contemporary political landscape. Formal democratic institutions often persist while liberal-democratic constraints are weakened through the concentration of power, the erosion of accountability, and restrictions on political and civil liberties (Bermeo, 2016; Lührmann & Lindberg, 2019). These developments reshape the strategic environment in which citizens respond to political institutions. Hirschman's Exit, Voice, and Loyalty (1970) provides the canonical framework for thinking about such responses, distinguishing departure from the political environment from attempts to influence it through voice. Yet political behaviour may also vary along another dimension: how publicly observable that response is.

An apparent absence of public dissent need not imply satisfaction or loyalty. Where exit is costly and public expression carries substantial personal risk, citizens may withdraw from visible opposition without abandoning political activity. We argue that understanding responses to institutional deterioration requires distinguishing not only between action and inaction, but also between channels of action that differ in public visibility. Frey (2023) develops Camouflage as a response to oppressive political conditions in which individuals hide or disguise their underlying political views, ranging from avoidance of contact and minimal participation to restricted cooperation and full collaboration with the regime. We adapt this idea to the visibility of oppositional activity itself: Political Camouflage here denotes opposition that remains active, but is expressed through lower-visibility channels when public exposure is costly.

Political Camouflage requires a particular combination of incentives. Public expression must be sufficiently costly to make lower-visibility activity attractive, while such channels must remain accessible enough to be used. Institutional backsliding can create this combination: restrictions on public contestation may raise the costs of visible action without eliminating the digital and organisational channels through which less visible activity remains possible. Where public expression carries little cost, concealment offers fewer advantages; where lower-visibility channels become prohibitively costly or inaccessible, the concealed option contracts. Camouflage is therefore most relevant where institutional deterioration raises the costs of public expression without simultaneously foreclosing alternative channels.

This argument is related to, but distinct from, established accounts of preference falsification, the spiral of silence, and hidden resistance (Noelle-Neumann, 1974; Kuran, 1987, 1995; Scott, 1990). These literatures show why private beliefs and public expressions may diverge when truthful expression is costly, and empirical studies document substantial gaps between private and publicly stated political attitudes under authoritarian conditions (Robinson & Tannenberg, 2019; Reuter & Szakonyi, 2021; Zhdanov & Tseitlin, 2024). Political Camouflage concerns a different behavioural margin. Rather than whether citizens misrepresent or withhold their preferences, our interest is in how political activity is expressed when public exposure is costly. In this respect the argument is closer to work on substitution under repression, in which coercion reallocates contention rather than simply reducing it (Lichbach, 1987; Moore, 2000), although our margin is the visibility of citizens' activity rather than its form. The distinction also creates an empirical opening: preferences that are merely withheld remain latent, whereas some forms of lower-visibility activity leave aggregate behavioural traces. Like other forms of concealed behaviour, Political Camouflage is nevertheless difficult to observe directly.

Interpreting movements in a lower-visibility channel poses two related problems. An increase in lower-visibility activity need not reflect a shift away from public dissent, since an

institutional rupture may stimulate political activity more broadly. Conversely, a decline in concealed activity may reflect restrictions on the channel itself rather than a decline in the demand for concealment. We formalise these possibilities in a simple channel-choice model in which citizens choose among inactivity, concealed political activity, and publicly visible activity. Institutional ruptures can alter the costs of public and concealed activity as well as political activation. Within this framework, we classify a response as camouflage when concealed activity increases and gains ground relative to public dissent. It also shows why declines in concealed activity need to be interpreted together with evidence on channel accessibility.

We use these implications to interpret cross-channel responses to institutional ruptures. Direct activity on the Tor anonymity network serves as a proxy for lower-visibility political activity, while protest events recorded by the Armed Conflict Location and Event Data Project capture publicly visible dissent. Tor enables anonymous communication and access to information and can circumvent censorship, but its use is neither exclusively political nor a direct measure of concealed opposition. We therefore do not interpret movements in Tor activity in isolation. Instead, we examine them jointly with the response of public protest and complement both outcomes with Tor bridge activity and reachability data from the Open Observatory of Network Interference, which provide evidence on whether access to the lower-visibility channel itself became more difficult.

We study nine discrete institutional ruptures. We first screen country-years between 2010 and 2023 for large annual declines in V-Dem's Political Civil Liberties Index and retain episodes that can be associated with a clearly datable institutional event. The resulting cases span 2013 to 2023. For each case and outcome, we estimate a separate counterfactual using augmented synthetic control and assess the post-treatment response against in-space placebos. This case-specific approach allows responses to differ across politically heterogeneous ruptures rather than imposing a common treatment effect.

The results show substantial heterogeneity across institutional ruptures. Hong Kong provides the clearest evidence of Political Camouflage. Direct Tor use increases, public protest declines, and Tor activity gains relative to protest. Cambodia exhibits the same directional pattern, although the estimates are less precise. Myanmar shows a different response. Tor use and public protest both increase, but the composition of the response shifts toward visibility. In Belarus and Russia, direct Tor use falls sharply while Tor reachability deteriorates and reliance on bridge relays increases. These cases are consistent with restrictions on access to the concealed channel rather than a simple decline in demand for it. Pakistan shows no clear cross-channel response. Taken together, the results show that publicly observable dissent does not fully characterise the political response to institutional deterioration.

The paper makes three contributions. First, it develops a visibility-based account of political behaviour under institutional deterioration, in which Political Camouflage is one of several configurations. The framework distinguishes the decision to act from the visibility of that action and clarifies how restrictions on public and concealed channels can shape observed dissent. Second, it provides an empirical approach that combines a proxy for lower-visibility activity with public protest and independent evidence on channel accessibility. This combination matters because neither changes in Tor use nor changes in protest can be interpreted in isolation. Third, it documents, case by case, different responses to institutional rupture, including Political Camouflage, public mobilisation, access-constrained responses, and cases with no clear cross-channel pattern.

The remainder of the paper proceeds as follows. Section 2 situates the argument within the literatures on Exit and Voice, preference falsification, political repression, and contentious behaviour. Section 3 develops the channel-choice framework and derives its observable implications. Section 4 describes the data, measurement, case selection, and synthetic-control design. Section 5 presents the empirical evidence across the nine institutional ruptures and examines concealed-channel accessibility. Section 6 discusses the findings and Section 7 concludes.

2. Related Literature

Hirschman's (1970) Exit, Voice, and Loyalty established the canonical framework for analysing responses to decline in firms, organisations, and political systems. Exit denotes withdrawal, voice captures attempts to alter the deteriorating environment, and loyalty affects the relative attractiveness of the two. Early critiques pointed to ambiguities in the original formulation, particularly the role of loyalty and the status of those who neither Exit nor exercise Voice (Barry, 1974; Dowding et al., 2000; Dowding, 2015). Barry (1974) distinguished staying from speaking, making silence analytically distinct from Voice without developing it as a strategic choice in its own right. More recent work treats silence as a meaningful response in itself (Ryu & Chang, 2017; Gray, 2021; Brito Vieira, 2021; Drujon d'Astros et al., 2024), and organisational research distinguishes the motives for withholding Voice (Van Dyne et al., 2003; Knoll et al., 2021; Morrison, 2023). Brito Vieira (2021) is closest to our concern, incorporating silence into the exit-voice framework and showing how it can conceal dissatisfaction under authoritarian rule. Political Camouflage differs from these approaches because concealed activity is not silence. Citizens may remain inactive, act through a lower-visibility channel, or act publicly, so the relevant distinction is not only whether citizens speak but how observable their political activity is.

This argument is closely related to work on preference falsification and hidden resistance. Kuran (1987, 1995) shows how political pressure can generate a gap between private preferences and public expression, while Noelle-Neumann (1974) and Scott (1990) describe related forms of silence and concealed opposition. Empirical work confirms that such gaps can be substantial. Robinson and Tannenberg (2019) find sizeable self-censorship in survey measures of regime support in China, Carter et al. (2024) report still larger differences between direct and indirect measures, and Jiang and Yang (2016) show that a political purge increased expressed support while reducing their measure of actual support; comparable evidence has been reported for Russia (Zhdanov & Tseitlin, 2024). Political Camouflage concerns a different outcome. Preference falsification asks whether public expression accurately reflects private preference, whereas we ask whether political activity is conducted through a public or a lower-visibility channel. The distinction is behavioural rather than attitudinal, and the two accounts are complementary rather than rival. When the concealed channel is unavailable or prohibitively costly, the model in Section 3 collapses toward a choice between inactivity and public expression, bringing the setting closer to the one studied in the preference-falsification literature.

The mechanism also relates to work showing that repression can reallocate contentious behaviour rather than simply suppress it. Lichbach (1987) demonstrates that repression directed at non-violent dissent may reduce that activity while increasing violent opposition, and Moore (2000) models substitution by governments between repression and accommodation. Our margin is different again: citizens allocate activity across channels that differ in public visibility. Recent empirical work provides evidence that contention changes form under autocratisation.

Arslanalp and Erkmen (2025) show that protest persists in authoritarian Turkey but becomes more concentrated in small-scale, spatially contained and less disruptive repertoires, and Ryan and Tran (2024) document the adaptation of online and hybrid pro-democracy activism under intensifying repression in Myanmar. A related literature establishes that anonymity technologies themselves respond to political conditions. Jardine (2018) documents a systematic cross-national relationship between political repression and Tor use, Chen et al. (2024) associate Tor adoption in less-free countries with censorship circumvention, and Strauch and Weidmann (2022) find increased Tor use following periods of intense protest in autocracies. These studies motivate Tor as a politically responsive lower-visibility channel without implying that Tor use is exclusively political, a distinction we return to in Section 4.

A final literature concerns mobilisation following salient political events. Protest depends not only on underlying grievances but also on the costs of participation, information about the behaviour of others, and expectations about the scale of opposition (Kuran, 1991; Lohmann, 1994; Bueno de Mesquita, 2010; Cantoni et al., 2024; Egorov & Sonin, 2024; for the state-side tactics producing these visibility costs, see Hassan et al., 2022). Bueno de Mesquita (2010), for example, formalises how observable political action can change beliefs about the extent of antigovernment sentiment and facilitate mass mobilisation. Repression can likewise have effects beyond simple deterrence: Pan and Siegel (2020) find that imprisoning prominent Saudi dissidents suppressed those individuals' own online activity but did not reduce online dissent overall, as their followers became more critical. These mechanisms matter for our setting because an institutional rupture may simultaneously make public expression more costly and raise political activation. Section 3 therefore allows exposure costs, concealed-channel accessibility, and political activation to move separately. Mobilisation and Camouflage are not ordered responses to increasingly severe shocks; they are different observable configurations generated by different combinations of these underlying margins.

3. Theoretical framework

Political Camouflage concerns the visibility of dissent rather than simply its overall level. To formalise this distinction, we consider how citizens allocate political activity across channels that differ in visibility. The framework distinguishes three components of an institutional rupture and their implications for concealed and publicly visible activity. The central implication is that a decline in publicly observable dissent need not indicate a decline in opposition, and the framework identifies the conditions under which that inference fails. Parameter restrictions and derivations are collected in Appendix A.

3.1 Channel choice

We consider a continuum of citizens of unit mass. Citizen i has a private intensity of grievance against the regime, θ_i, distributed on $[0, 1]$ with cumulative distribution function F and a continuous, strictly positive and bounded density f. Building on Hirschman's (1970) distinction between Exit and Voice and Barry's (1974) extension and critique, we introduce visibility as an additional dimension of political behaviour. Exit is a qualitatively different response that entails leaving the political environment rather than choosing how to act within it, and we set it aside. We focus instead on responses within the existing political environment: citizens may remain inactive, engage through a lower-visibility channel, or act publicly. We do not treat anonymous communication as necessarily constituting Hirschmanian voice, nor the framework as a rigid

behavioural taxonomy. Rather, we take from this literature the broader insight that dissatisfaction can generate alternative behavioural responses and ask how institutional deterioration changes their public visibility. Citizens choose among inactivity s, concealed political activity h, and publicly visible activity v, with payoffs:

$$U(s|\theta_i) = 0, \qquad U(h|\theta_i) = \theta_i - ch, \qquad U(v|\theta_i) = B\theta_i - R, B > 1.$$

Inactivity is the outside option and is normalised to zero. Any political return that accrues to citizens irrespective of their own action, such as a policy concession generated by aggregate protest, is common to the three alternatives and therefore drops out of individual channel choice (Olson, 1965; Tullock, 1971). What remains are the participation-contingent returns and costs. Concealed activity yields a return per unit of grievance, normalised to one, at a fixed cost c_h capturing the practical, informational, and technological barriers to accessing and using lower-visibility channels. Public activity yields the higher return $B > 1$ per unit of grievance, reflecting the expressive, social, and agency benefits associated with public participation, at an exposure cost R associated with surveillance, sanction, or repression.

Because payoffs are affine in grievance and public activity is both the more rewarding and the more costly channel, citizens sort into intervals. Under the conditions stated in Appendix A.1, which place both thresholds inside the unit interval and keep concealed activity on the upper envelope, the indifference points are:

$$\theta_h^* = c_h \text{ (\textit{inactivity versus concealment})}$$

$$\theta_{hv}^* = (R - c_h)/(B - 1) \text{ (\textit{concealment versus public activity})},$$

with $0 < \theta_h^* < \theta_{hv}^* < 1$. Citizens below θ_h^* remain inactive, those between the thresholds conceal, and those above θ_{hv}^* act publicly, giving shares $n_s = F(\theta_h^*), n_h = F(\theta_{hv}^*) - F(\theta_h^*)$, and $n_v = 1 - F(\theta_{hv}^*)$.

The middle region represents concealed political activity. It is neither loyalty nor inactivity, but an active response chosen by citizens whose grievance is sufficient to motivate action while the exposure cost of public activity makes a lower-visibility channel preferable. As discussed in Section 2, this places the mechanism in a tradition in which repression reallocates contention rather than simply reducing it, the margin here being the visibility of citizens' activity. It is also distinct from preference falsification, which concerns the relationship between private and reported preferences rather than the channel through which oppositional behaviour is enacted (Kuran, 1987, 1995). When the concealed channel is unavailable or prohibitively costly, this middle interval vanishes and citizens face only the binary choice between withholding action and expressing it publicly.

3.2 What an institutional rupture does

Treating a coup, an annulled election or a comparable institutional break as a single scalar increase in R would be too restrictive. Raising the exposure cost of public activity is the mechanism of interest, but it is not all that such an event does. Such events may also make lower-visibility channels more costly or difficult to access, raising c_h, and because they are public and salient, they can raise the intensity with which citizens perceive and act on political grievance. We represent this third component as a common increase $g \geq 0$, so that citizen i acts on $\theta_i + g$. Activation is specified as a uniform increase in perceived grievance rather than as a change in the cost or return of either channel: the rupture makes citizens care more, not act more

cheaply. These components need not move together, and an observed rupture may contain more than one of them. Writing φ_h and φ_v for the density of grievance at the lower and upper thresholds, the response of the three shares to a composite shock (dR, dc_h, dg) is:

$$dn_s = \varphi_h(dc_h - \mathrm{dg}), \quad dn_v = \varphi_v(-(dR - dc_h)/(B-1) + dg), \quad dn_h = -dn_s - dn_v$$

The three expressions separate two margins that an institutional rupture can move: how many citizens act, and how they act. The mass of citizens who act at all, $1 - n_s$, responds only to dc_h and dg. It does not depend on dR. In the model, an increase in the exposure cost of public activity alone does not change the mass of citizens who act on their grievance, it changes only how they act. An observer with data only on public dissent would record a contraction and might infer demobilisation, although in the model no citizen becomes inactive. This is the precise sense in which Political Camouflage differs from deterrence.

3.3 Observable implications

Table 1 summarises the predicted responses of the three behavioural channels to the main components of an institutional rupture. To capture changes in the composition of active dissent, we define:

$$\rho = log\left(\frac{n_h}{n_v}\right) = log(n_h) \; - \; log(n_v)$$

where a rise in ρ indicates that concealed activity gains relative to public activity. Section 4 constructs an empirical analogue using Tor and protest activity.

Table 1. *Comparative statics of institutional ruptures*

Shock component	Concealed activity n_h	Public activity n_v	Inactivity n_s	Relative concealment ρ	Empirical configuration
Exposure cost R rises	+	−	0	+	Camouflage: visibility substitution
Concealed-channel cost c_h rises	−	+	+	−	Channel restriction
Political activation g rises	ambiguous	+	−	ambiguous	Political mobilisation
R and c_h rise together	ambiguous	ambiguous	+	ambiguous	Access-constrained response

Note. Signs denote local comparative-static responses around the interior solution in Section 3.1, holding unlisted components fixed. The first three rows vary one component at a time; in the final row, $dR > 0$ and $dc_h > 0$ while $dg = 0$. Under political activation, the sign of ρ is ambiguous and depends on the grievance distribution around the two thresholds and the resulting equilibrium shares (Appendix A.4). In the final row, concealed activity falls whenever dch/dR exceeds φv/(φv + (B − 1)φh).

Three implications organise the empirical analysis. First, a rise in concealed activity is not by itself evidence of camouflage, because political activation can also produce such an increase. Second, a rise in relative concealment is not sufficient either, because it can occur even when concealed activity falls, provided that public dissent falls proportionally more. We therefore define the empirical camouflage pattern by two jointly required conditions: concealed activity must increase and it must gain relative to public dissent, that is, $\Delta n_h > 0$ and $\Delta\rho > 0$. Among the three isolated margins, an increase in the exposure cost of public activity is the one that unambiguously predicts both responses. The criterion does not structurally identify a pure change in R, since an observed rupture may affect several margins simultaneously. Rather, it requires that lower-visibility activity both expands and becomes relatively more prominent.

Because exposure costs and political activation may increase simultaneously, camouflage does not require public dissent to decline. A composite shock can increase both concealed and public activity while concealed activity gains relative ground. Displacement, in which public dissent contracts outright, is therefore one possible manifestation of camouflage rather than a defining feature of it.

Third, a contraction of concealed activity is also uninformative on its own. It can arise when the concealed channel becomes more costly even as the incentive to avoid public exposure increases. When exposure and concealed-channel costs rise jointly, holding political activation fixed, even an increase in the concealed-channel cost that is smaller than the increase in the exposure cost can be sufficient to reverse the observed concealed response. Observed use and underlying demand can therefore move in opposite directions, so a contraction must be interpreted together with evidence on the accessibility of the concealed channel. More generally, when exposure and concealed-channel costs change jointly, the response of public dissent is ambiguous because the two components push it in opposite directions.

The framework also places limits on the patterns it can rationalise. A contraction in both channels without evidence of deterioration in concealed-channel accessibility does not follow from the mechanisms above and would instead point towards deterrence, withdrawal, or another mechanism outside the model. Because accessibility is observed only imperfectly, this is best understood as a limit on what the framework explains rather than as a sharp test.

The model is deliberately minimal. It is static, treats the regime as exogenous, and abstracts from learning, networks, and feedback between concealed and public dissent. Political activity can reveal information and affect subsequent mobilisation (Kuran, 1991; Lohmann, 1994; Edmond, 2013), while regimes may respond to substitution by restricting lower-visibility channels (Roberts, 2018), a possibility examined using the accessibility evidence in Section 5. Appendix B shows that allowing the return to public activity to depend on aggregate participation leaves the sign structure of Table 1 unchanged. The model represents mutually exclusive individual choices for analytical clarity, whereas the empirical proxies measure aggregate activity and need not correspond to distinct individuals: the same citizen may attend a protest and use an anonymity network in the same week. The empirical analysis therefore examines the predicted directional patterns rather than estimating population shares or individual-level switching.

4. Data, measurement, and identification

4.1 Measuring concealed and visible political activity

We combine weekly data on Tor use and protest activity to measure political responses that differ in their public visibility. Our primary proxy for concealed activity is the estimated number of users connecting directly to the Tor network, obtained from Tor Metrics (Tor Project, 2024). Tor provides anonymity and allows users to circumvent surveillance and censorship, but its use is not exclusively political. We therefore interpret changes in Tor activity as changes in the use of an anonymity-preserving channel rather than as counts of hidden dissenters.

This interpretation is supported by previous evidence that Tor use responds systematically to political conditions. Jardine (2018) documents a relationship between political repression and Tor usage across countries. Strauch and Weidmann (2022) show that Tor use increases in autocracies following protest events, while Sandberg (2018) links episodes of political unrest to

increased use of anonymity networks. Chen et al. (2024) further show that Tor adoption in less-free countries is closely associated with censorship circumvention. These studies motivate Tor as a politically responsive measure of low-visibility activity without implying that all Tor use is political.

The limitations of this proxy are important. Tor Metrics reports aggregate, model-based estimates derived from directory requests, so the series cannot be decomposed into political and non-political use or linked to identifiable actors. We therefore measure activity through a low-visibility channel rather than dissent itself.

We aggregate the Tor series to the weekly level and define:

$$T_{\mathrm{ct}} = log(DirectTor_{ct})$$

where $DirectTor_{ct}$ denotes direct Tor use in country c and week t. The logarithmic transformation places changes on an approximately proportional scale and limits the influence of large differences in baseline usage across countries. Internal gaps of up to four consecutive weeks in the Tor series are linearly interpolated; longer gaps are treated as missing rather than as zero use. The bridge series is handled the same way, with weeks reporting no bridge users inside a country's coverage period coded as zero.

Public dissent is measured using protest events from the Armed Conflict Location and Event Data Project (ACLED; Raleigh et al., 2010). We aggregate events to the country-week level and use:

$$V_{ct} = asinh(Protest_{ct})$$

which accommodates weeks without protest while limiting the influence of extreme event counts. Weeks without recorded protest are coded as zero only within a country's ACLED coverage period; observations outside that period are treated as missing. Because the inverse hyperbolic sine transformation is not invariant to the units of the underlying count, effects on this scale are not interpreted as elasticities (Bellemare & Wichman, 2019; Chen & Roth, 2024). We therefore report protest effects in asinh points throughout and do not convert them to percentage changes. ACLED records events rather than participants, just as Tor Metrics measure network activity rather than political preferences. The two outcomes are therefore treated as proxies for activity expressed through channels that differ in public visibility.

The theoretical argument concerns not only activity in either channel separately but also their relative movement. To construct an empirical analogue of the relative-concealment measure ρ, while accommodating weeks without protest events, we define

$$\tilde{\rho}_{ct} = log(DirectTor_{ct}) - log(1 + Protest_{ct})$$

The level of $\tilde{\rho}_{ct}$ has no substantive interpretation because Tor users and protest events are measured in different units. Its treatment effect instead captures whether Tor activity moves upward or downward relative to observable protest activity on logarithmic-type scales. A positive effect indicates an increase in relative concealment. We estimate $\tilde{\rho}_{ct}$ directly as a separate outcome rather than subtracting the separately estimated Tor and protest effects, since those estimates generally rely on different synthetic counterfactuals. Following the criterion derived in Section 3, the empirical camouflage pattern therefore requires both a positive treatment effect on direct Tor use and a positive treatment effect on $\tilde{\rho}_{ct}$.

For estimation, each transformed outcome is centered on the unit's pretreatment mean and re-scaled using a single outcome-specific scale calculated from pretreatment observations across the estimation sample. Using a common scale preserves genuine differences in volatility across countries, while restricting the transformation to pretreatment observations ensures that it cannot be affected by treatment. Treatment effects are converted back to log or asinh units for presentation.

We use bridge connections and Tor reachability measurements from the Open Observatory of Network Interference (OONI; Filastò & Appelbaum, 2012) as auxiliary evidence on the accessibility of the concealed channel. Bridges are non-public entry points intended to preserve access when standard Tor relays are restricted, so migration toward bridges provides evidence of circumvention. OONI records whether Tor can be reached from participating networks. These measures are used to interpret movements in direct Tor activity rather than as additional measures of dissent.

4.2 Treated cases and donor pools

The treated sample is selected independently of the estimated Tor and protest responses. We begin from country-years between 2010 and 2023 in which V-Dem's political civil liberties index (v2x_clpol; Coppedge et al., 2023) declines by at least 0.10 within one year. This identifies 33 candidate country-years across 28 countries. We then require that the institutional deterioration can be associated with a discrete, datable event rather than cumulative erosion, that the outcome data support the comparison window described below, and that ACLED counts are not dominated by active armed conflict. Appendix C reports the screening procedure and the disposition of the candidate cases. The resulting sample contains the nine institutional ruptures listed in Table 2.

Table 2. Treated countries and institutional ruptures

Country	Shock date	V-Dem 1y drop	Pre-shock V-Dem	Triggering event
Egypt	2013-07-03	0.336	0.562	Military removal of Morsi
Thailand	2014-05-22	0.388	0.580	Royal Thai Armed Forces coup
Türkiye	2016-07-15	0.190	0.466	Failed coup attempt and post-coup purge
Cambodia	2017-11-16	0.121	0.361	Dissolution of CNRP opposition party
Hong Kong	2020-06-30	0.140	0.573	National Security Law imposed
Belarus	2020-08-09	0.226	0.411	Disputed presidential election
Myanmar	2021-02-01	0.540	0.658	Military coup
Russia	2022-02-24	0.123	0.303	Invasion of Ukraine, domestic crackdown
Pakistan	2023-05-09	0.106	0.611	Arrest of Imran Khan, mass protests

Notes. Cases are ordered by date. The V-Dem drop is the one-year change in the political civil liberties index (v2x_clpol) in the year of the rupture, and the pre-shock level is the index value in the preceding year. All nine cases enter the analysis of direct Tor use. ACLED coverage of Türkiye begins in 2018, leaving no pretreatment window for the July 2016 rupture, so Türkiye is omitted from the protest and relative-concealment analyses.

The main estimation window contains 96 pretreatment weeks and 52 post-treatment weeks. The week containing the institutional event is excluded because treatment typically begins partway through a calendar week. The treatment-effect window therefore covers the first 52 complete

post-treatment weeks. A common horizon keeps the estimand comparable across cases and limits its exposure to later political developments.

Donor pools are constructed separately for each treated case and outcome. Eligible donors must provide sufficient observations over the comparison window, exhibit non-degenerate pretreatment variation, and have sufficient V-Dem coverage to establish their treatment status. Countries experiencing their own one-year decline of at least 0.10 in the Political Civil Liberties Index during the comparison period are excluded. Donor eligibility therefore depends on institutional histories and data availability rather than on the estimated post-treatment response. The resulting donor pools contain between 17 and 46 countries, depending on the case and outcome. To avoid subjective geographic exclusions, our primary donor pools include all countries with sufficient data.

4.3 Synthetic-control estimation and inference

We estimate a separate synthetic counterfactual for each institutional rupture and outcome, allowing the response to differ across cases rather than imposing a common treatment effect on politically heterogeneous events (Abadie et al., 2010, 2015; Abadie, 2021). Our main estimator is the augmented synthetic control method of Ben-Michael, Feller and Rothstein (2021), which combines synthetic-control weighting with a ridge-regularised outcome model to correct residual pretreatment imbalance when conventional SCM does not achieve exact fit.

Let Y_{ct} denote the outcome for treated country c and $\hat{Y}^0_{ct}$ its estimated untreated counterfactual. Our case-level estimand is the average treated-synthetic gap over the first 52 complete post-treatment weeks,

$$\hat{\tau}_c = \frac{1}{52} \sum_{k=1}^{52} (Y_{c,T_c+k} - \hat{Y}^0_{c,T_c+k})$$

Pretreatment fit remains central to the credibility of the counterfactual. We therefore report pretreatment root mean squared prediction error (RMSPE) and the ratio of post- to pretreatment RMSPE. We do not impose a mechanical fit threshold, estimates with weaker pretreatment fit are reported and interpreted more cautiously.

Inference is based on in-space placebo tests (Abadie et al., 2010, 2015). For each treated case, we reassign the intervention to each eligible donor in turn and re-estimate its synthetic counterfactual using the remaining units. The test statistic is the absolute mean post-treatment gap scaled by pretreatment RMSPE,

$$S_c = \frac{|\hat{\tau}_c|}{RMSPE_{c,pre}}$$

Permutation p-values rank the treated statistic relative to the resulting placebo distribution. Because the number of successfully estimated placebos varies across cases, the smallest attainable p-value is $1/(J_c + 1)$, where J_c denotes the number of placebo units. Inference is case-specific, and we do not combine permutation probabilities across cases because their donor pools overlap substantially.

The identifying assumption is that, absent the institutional rupture, the treated country's outcome would have followed the path represented by its synthetic comparison. Pretreatment fit supports the plausibility of this counterfactual but cannot rule out contemporaneous country-

specific shocks or spillovers onto donor countries. Because the institutional events are not randomly assigned, we interpret the estimates as case-specific reduced-form effects of the dated ruptures rather than as structural effects of any single component of institutional deterioration.

5. Empirical Evidence

We now apply the empirical criterion derived in Section 3 to the nine institutional ruptures. We first estimate responses in direct Tor use, then examine public protest and relative concealment, and finally use Tor accessibility to distinguish contractions in observed use from restrictions on the concealed channel.

5.1 Effects on direct Tor use

Table 3 reports the case-specific synthetic-control estimates for direct Tor use. Pretreatment fit is close in most cases. RMSPE is below 0.09 in seven of the nine countries, although Belarus and Myanmar exhibit substantially greater pretreatment volatility.

Table 3. Effect of institutional ruptures on direct Tor use

Case	ATT	Change (%)	Pre-period RMSPE	Post/pre RMSPE	p	Attainable minimum
Pakistan	−0.08	−8	0.053	2.4	.340	.021
Cambodia	0.33	+40	0.054	8.4	.308	.038
Russia	−0.68	−49	0.075	9.5	.021**	.021
Hong Kong	0.43	+54	0.085	6.8	.065*	.022
Türkiye	−0.12	−11	0.044	13.4	.150	.050
Belarus	−2.20	−89	0.241	10.4	.022**	.022
Egypt	0.62	+86	0.081	9.0	.278	.056
Thailand	0.04	+4	0.075	1.1	.789	.053
Myanmar	0.78	+118	0.239	4.1	.239	.022

Notes. The outcome is the natural logarithm of directly connecting Tor users. The ATT is the mean gap between the treated country and its synthetic comparison over the 52 complete post-treatment weeks, in log points, and the percentage change is $100 \times [\exp(\text{ATT}) - 1]$. The post/pre RMSPE ratio expresses the post-treatment divergence in units of pretreatment prediction error; because it is scaled by each country's own fit, its level is not comparable across cases. p-values are permutation probabilities from in-space placebo tests, and the attainable minimum is $1/(J + 1)$, where J is the number of placebo units (Abadie et al., 2010, 2015); recomputing them after discarding placebos whose pretreatment RMSPE exceeds five times that of the treated unit leaves them unchanged to three decimal places. * $p < .10$. ** $p < .05$.

The estimates reveal substantial heterogeneity. Direct Tor use falls sharply relative to its synthetic counterfactual in Belarus (–89 per cent) and Russia (–49 per cent); both effects lie at the extreme of their placebo distributions. Point estimates are positive in Cambodia (+40 per cent), Hong Kong (+54 per cent), Egypt (+86 per cent), and Myanmar (+118 per cent), although only the Hong Kong effect is distinguishable from placebo variation, and at the ten per cent level ($p = .065$). Pakistan and Türkiye exhibit modest declines, while the estimated change in Thailand is negligible.

5.2 Public dissent and relative concealment

Table 4 reports the corresponding estimates for public protest and the relative-concealment outcome defined in Section 4.1. Pretreatment fit is weaker than for direct Tor use, reflecting the greater week-to-week volatility of protest events: RMSPE ranges from 0.24 to 1.20 asinh points for protest and from 0.27 to 0.95 log points for relative concealment, and is poorest in both for Egypt. Cases with comparatively poor fit are interpreted with greater caution. The discussion below covers the cases with a positive Tor response; Belarus, Russia and Pakistan, whose Tor effects are negative or negligible, are taken up in Section 5.3, where the accessibility evidence bears on their interpretation.

Table 4. Effects on public protest and relative concealment

Case	Public dissent ATT (asinh pts)	p	Relative ATT (log pts)	p	Pre-RMSPE (protest)	Pre-RMSPE (relative)
Pakistan	−0.02	.787	+0.01	.957	0.24	0.27
Cambodia	−1.13	.154	+1.16	.192	0.75	0.58
Russia	−0.45	.234	−0.67	.192	0.51	0.53
Hong Kong	−0.62	.022**	+0.83	.022**	0.48	0.39
Belarus	+2.09	.022**	−4.12	.022**	0.65	0.78
Egypt	+1.01	.316	−0.44	.833	1.20	0.95
Thailand	−1.37	.053*	+0.99	.105	0.93	0.69
Myanmar	+3.30	.022**	−2.61	.022**	0.89	0.75

Notes. Public dissent is the inverse hyperbolic sine of weekly ACLED protest events, reported in asinh points. The relative outcome is log(Tor) − log(1 + protest), estimated as a separate synthetic control; a positive effect indicates that concealed activity grew proportionally more than public dissent. It differs from the difference of the two component ATTs by up to 0.45 log points, although no case classification depends on which is used (Appendix Table C2). Camouflage requires a positive effect both here and in Table 3. Türkiye is omitted; see Table 2. * p < .10. ** p < .05.

The cross-channel estimates reveal two distinct patterns and one case that fits neither. Hong Kong provides the clearest evidence of political camouflage: direct Tor use rises by approximately 54 per cent, public dissent falls, and the relative-concealment effect is positive and distinguishable from the placebo distribution. The case therefore exhibits the displacement form of camouflage identified in Section 3.3, since public dissent contracts outright. Cambodia exhibits the same sign pattern, with higher Tor use, lower protest and a positive relative effect, although none of the three estimates is distinguishable from placebo variation, despite a donor pool large enough for the five per cent level. In Thailand the positive relative effect is driven by the decline in protest, while direct Tor use is essentially unchanged (+4 per cent, p = .79); Thailand therefore shows no evidence for the first condition of the camouflage criterion.

Myanmar and Egypt instead display political mobilisation, in which an increase in political activation raises public activity while leaving the composition of the response unsigned. In both cases the point estimates indicate increases in Tor use and public dissent rather than a shift toward concealment. In Myanmar the relative-concealment effect is significantly negative: the large increase in Tor use occurs alongside, and is outpaced by, visible mobilisation. Egypt displays the same qualitative pattern, but its relative effect is imprecisely estimated and its pretreatment fit weak. An institutional shock may therefore expand activity in both channels while leaving the response relatively more visible.

5.3 Channel accessibility and circumvention

The sharp declines in direct Tor use in Belarus and Russia may reflect restrictions on access rather than lower demand for the network. We examine this possibility using OONI reachability and bridge-relay use, with rising bridge use indicating greater reliance on circumvention when direct access becomes more difficult.

Table 5. Channel accessibility and bridge-relay use

Case	Direct Tor (%)	Total Tor (%)	Bridge share pre (%)	Bridge share post (%)	Change (pp)	OONI reachability (%)
Belarus	−88	−76	8.3	53.4	+45.1	98.0 → 5.4
Russia	−64	−54	5.1	27.3	+22.2	51.9 → 18.9
Cambodia	+76	+103	6.8	19.2	+12.4	n.a.
Türkiye	−12	−0.4	2.3	13.3	+11.0	n.a.
Myanmar	+234	+263	2.4	10.2	+7.8	100 → 99.1
Thailand	−47	−47	0.0	0.1	+0.1	n.a.
Pakistan	−20	−20	1.1	1.1	0.0	98.1 → 97.3
Egypt	+794	+794	0.1	0.0	−0.1	n.a.
Hong Kong	+41	+39	2.9	1.9	−1.0	100 → 99.5

Notes. Percentages compare mean weekly users in the 52 complete weeks after each shock with the 52 weeks before, excluding the event week. A rising bridge share alongside a falling total is consistent with restrictions on direct access and greater reliance on circumvention. OONI reachability is the share of Tor measurements returning a successful connection over the same window, and is unavailable where volunteer coverage is insufficient. Cases are ordered by the change in bridge share.

The evidence for restricted access is strongest in Belarus and Russia. In both, declining direct Tor use coincides with a sharp deterioration in reachability and a pronounced shift toward bridge relays, although circumvention only partially offsets the disruption and total Tor use still falls. The negative direct-Tor effects are therefore more consistent with an accessibility constraint than with a disappearance of demand for the network.

The two cases differ in their public response: protest rises sharply in Belarus and falls, imprecisely, in Russia. Restricting the concealed channel raises the relative return to acting publicly, so public activity rises when only the cost of concealment increases, but is not signed when the cost of public expression rises at the same time. The two cases are consistent with combinations on either side of that condition, although the design does not identify the components separately.

Reachability remains essentially unchanged in the three cases for which OONI data are available – Hong Kong, Myanmar and Pakistan. In Hong Kong the bridge share in fact falls, and reachability stands at 99.5 per cent after the National Security Law, so the rise in direct Tor use is unlikely to reflect a change in how the network was reached. In Pakistan both measures are flat, so the absence of a response there is unlikely to reflect an inability to detect one. For Cambodia, Egypt, Thailand and Türkiye, OONI coverage is insufficient to determine whether accessibility changed materially. Cambodia's bridge share rises by twelve points while total use roughly doubles, which is consistent with some increase in the cost of direct access that did not prevent concealed activity from rising; we therefore record its channel status as not determined rather than unrestricted. In Türkiye the bridge share rises from 2.3 to 13.3 per cent while total Tor use is unchanged, consistent with restricted access to directly connecting relays largely

offset by substitution into bridges, and its incomplete protest series prevents an assessment of relative concealment.

5.4 Patterns across cases

Table 6 brings the three outcome dimensions together with the available accessibility evidence.

Table 6. Observed response patterns across cases

Case	Concealed activity (%)	Public dissent (asinh pts)	Relative concealment	Channel	Classification
Hong Kong	+54*	−0.62**	+0.83**	Open	Camouflage: visibility substitution
Cambodia	+40	−1.13	+1.16	Not determined	Camouflage, directional
Myanmar	+118	+3.30**	−2.61**	Open	Political mobilisation
Egypt	+86	+1.01	−0.44	Not determined	Political mobilisation, directional
Belarus	−89**	+2.09**	−4.12**	Restricted	Channel restriction
Russia	−49**	−0.45	−0.67	Restricted	Access-constrained response
Türkiye	−11	—	—	Restricted	Restriction offset by circumvention
Pakistan	−8	−0.02	+0.01	Open	No detectable response
Thailand	+4	−1.37*	+0.99	Not determined	Unassigned

Notes. Concealed activity is the percentage change from Table 3; public dissent is in asinh points and relative concealment in log points, both from Table 4. Channel status is assessed from the OONI reachability and bridge-relay evidence in Table 5, independently of the estimated effects, and is recorded as not determined where neither source is informative. Classifications use the configuration names of Table 1. Hong Kong's concealed-activity effect is significant at the ten per cent level rather than the five. Türkiye's protest and relative outcomes are not estimated; see Table 2. * $p < .10$. ** $p < .05$.

Six of the nine cases map onto the four configurations of Table 1; the remaining three fall outside them. Hong Kong provides the clearest evidence of Political Camouflage and matches the visibility-substitution row on all three observable dimensions, with Cambodia displaying the same directional pattern but weaker inferential support. Myanmar corresponds to political mobilisation, with Egypt showing a qualitatively similar but imprecisely estimated response. Belarus matches channel restriction on every observable sign, while Russia shows declining concealed activity with a public response that is not distinguishable from zero and therefore corresponds to the joint case; because that configuration restricts neither observable channel, Russia is consistent with it rather than evidence for it. Türkiye shows bridge migration with no change in total Tor use and cannot be assessed on relative concealment. Pakistan shows no detectable response in either channel, with close pretreatment fit and an accessible concealed channel throughout. Thailand shows a large contraction in public dissent with no concealed-channel response, and the available evidence does not identify whether citizens withdrew or whether both channels became costlier.

Read through a single channel, three of these cases would be interpreted differently. Hong Kong, observed only through protest, would appear to show demobilisation. Myanmar, observed only through Tor, would appear to show a shift toward concealment. Belarus, observed only through Tor, would appear to show collapsing demand for anonymity. In each case the second channel, or the accessibility evidence, reverses the reading, so neither channel identifies

the response on its own. Section 6 discusses what these heterogeneous responses imply for the interpretation of observable dissent and the Political Camouflage framework.

5.5 Robustness

Appendix D reports five checks: leave-one-donor-out sensitivity, a 52-week pretreatment window, region-restricted donor pools, the unaugmented synthetic control with weights constrained to the simplex, and a backdated placebo. The sign of the estimate is preserved in 24 of 25 case-outcome combinations under leave-one-donor-out and in all 25 under the unaugmented estimator, the single exception being indistinguishable from zero in the main specification; across every specification the Hong Kong effect on direct Tor use lies between +35 and +59 per cent, and among those in which the relative outcome is re-estimated its relative-concealment effect lies between +0.77 and +0.91 log points. Re-estimating concealed activity on the combined direct-plus-bridge series leaves Hong Kong unchanged at +57 per cent and Türkiye indistinguishable from zero, while the Belarus contraction attenuates from −89 to −59 per cent and its permutation probability rises from .022 to .065 (Table C3). Case-level qualifications, including the attenuation of the Cambodian estimates on regional donor pools and the political events falling inside several pretreatment windows, are reported with the corresponding tables.

6. Discussion

The nine institutional ruptures do not produce a common response. Hong Kong satisfies both conditions of the camouflage criterion, with direct Tor use rising, public protest falling, and relative concealment increasing. Cambodia displays the same directional pattern with weaker inferential support, while Thailand satisfies the relative condition only because protest declines, its concealed-channel response being negligible. Myanmar and Egypt show expansion in both channels with the composition moving toward visibility. Belarus and Russia show contraction in the concealed channel alongside deteriorating reachability and increased reliance on bridge relays. Pakistan shows no detectable response in either channel. These are the configurations set out in Table 1, and the cases distribute across them rather than clustering on one.

The analysis therefore does not support the claim that institutional deterioration generally produces Political Camouflage. It supports a narrower claim, which is the one the framework of Section 3 implies. Because a rupture can raise the exposure cost of public activity, restrict access to lower-visibility channels, and raise political activation, and because these components leave different signatures in the two channels, the movement of either channel on its own is consistent with more than one underlying situation. The cases illustrate this directly. Protest in Hong Kong falls sharply, which taken alone would indicate demobilisation, while concealed activity rises and the composition shifts toward concealment. Concealed activity in Myanmar more than doubles, which taken alone would indicate a shift toward concealment, while visible mobilisation expands more strongly and the relative effect is significantly negative. In the model, an increase in the exposure cost of public activity leaves the mass of citizens who act unchanged and alters only how they act; the empirical counterpart is that a decline in observed protest does not establish a decline in political activity.

The results also show why the second condition of the criterion and the accessibility evidence are both necessary. A positive movement in relative concealment is not sufficient on its own, as Thailand illustrates: the composition shifts toward concealment because protest contracts, not because concealed activity expands. A contraction in concealed activity is likewise

uninformative in isolation. In Belarus and Russia, the two cases with the largest negative estimates for direct Tor use, measured reachability deteriorates sharply and the share of users connecting through bridge relays rises, which is consistent with the concealed-channel cost rising rather than with demand for concealment disappearing. Where only the cost of concealment rises, the model predicts that public dissent increases, and Belarus matches that configuration on every observable sign. Russia shows the same contraction in the concealed channel with a public response that is not distinguishable from zero, which is consistent with the cost of public expression having risen at the same time.

These findings connect to the two literatures set out in Section 2 in different ways. The work following Kuran (1987, 1995) concerns the relationship between privately held and publicly expressed preferences; the margin examined here is behavioural, concerning the channel through which political activity is conducted. The two coincide where the concealed channel is unavailable, since the model then reduces to a choice between inactivity and public expression. The substitution tradition following Lichbach (1987) and Moore (2000) is closer in structure, in that coercion reallocates contention rather than simply reducing it, and the contribution here is to apply that logic to the visibility of citizens' activity rather than to its form or to the state's choice of response. The cases also indicate that any such reallocation is bounded by the operational accessibility of the alternative channel.

The framework does not account for every pattern it might encounter, and one case illustrates the limit rather than the mechanism. Pakistan shows no detectable movement in either channel, which is a null rather than a configuration. More generally, a contraction in both channels without evidence of deteriorating accessibility would not follow from any of the three components and would indicate deterrence or withdrawal on a margin the model does not contain. The results are also consistent with the position taken in Section 3 that the measured magnitude of an institutional rupture does not determine the response: the two largest institutional declines in the sample are associated with mobilisation and with a negligible concealed-channel response respectively, while the clearest camouflage pattern occurs in the lower half of the range. Whether the severity of a rupture systematically shapes the form of the response is not something nine cases can settle, and we do not test it here.

Four limitations qualify these conclusions. The first concerns measurement. Tor Metrics reports aggregate, model-based estimates that cannot be decomposed into political and non-political use or linked to identifiable actors, and we have no independent validation that the movements documented here are political in origin; the interpretation rests on the dating of the ruptures, the untreated counterfactual, and the joint reading of the two channels. The second concerns the strength of the evidence for camouflage specifically. Only Hong Kong satisfies both conditions with inferential support in each component, so the comparative argument rests on the configuration of estimates across cases rather than on any single estimate. The third concerns scope. The design requires a discrete, datable event, so the nine cases are ruptures rather than episodes of gradual erosion, and whether the same configurations arise where institutions deteriorate slowly is outside what this evidence can establish. A fourth is conceptual: the model represents mutually exclusive individual choices while the proxies measure aggregate activity in each channel, so the analysis does not identify individuals moving between channels.

These suggest directions for further work. Tor is one lower-visibility channel among several, and extending the measurement to virtual private network adoption or encrypted-messaging uptake would indicate whether these patterns are specific to one technology. Validating

anonymity-network activity against an independent measure of political behaviour would strengthen the interpretation of the proxy. And measuring the three components of a rupture separately, rather than inferring them from the responses they produce, would allow the framework to be tested rather than applied.

7. Conclusion

This paper examines how citizens respond to institutional deterioration when public expression becomes costly, and argues that responses differ not only in whether citizens act but in how observable their activity is. We use Political Camouflage to describe opposition expressed through lower-visibility channels when public exposure is costly, and develop a channel-choice model in which citizens sort across inactivity, concealed activity, and publicly visible activity. An institutional rupture can raise the exposure cost of public activity, restrict access to concealed channels, and increase political activation. These components imply different configurations of concealed and public activity. The model therefore yields an empirical criterion for camouflage that requires concealed activity both to increase and to gain ground relative to public dissent.

Applying this framework to nine institutional ruptures between 2013 and 2023, using activity on the Tor anonymity network, recorded protest events, and separate synthetic comparisons for each case, we find no uniform response. Hong Kong provides the clearest evidence of Political Camouflage, while Cambodia displays the same directional pattern with weaker inferential support. Myanmar and Egypt instead show expansion in both channels led by visible mobilisation. Belarus and Russia combine declining direct Tor use with substantial deterioration in access to the network, while Pakistan shows no detectable response and two cases cannot be classified with the available evidence. Political Camouflage is therefore not the typical response to institutional deterioration.

The broader implication is that observable dissent is an incomplete measure of political response. Across the nine ruptures studied here, a decline in public protest cannot by itself be read as consent or demobilisation: in Hong Kong protest contracted while lower-visibility activity expanded, and in Belarus and Russia a decline in that lower-visibility channel was more consistent with restricted access to it than with falling demand for concealment. Institutional deterioration can change not only how much political activity occurs, but where it occurs and how visible it becomes.

Literature

Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, *59*(2), 391–425. https://doi.org/10.1257/jel.20191450

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method: COMPARATIVE POLITICS AND THE SYNTHETIC CONTROL METHOD. *American Journal of Political Science*, *59*(2), 495–510. https://doi.org/10.1111/ajps.12116

Arslanalp, M., & Erkmen, T. D. (2025). Resistance under confinement: Resilience of protests and their limits in authoritarian Turkey. *Third World Quarterly*, *46*(2), 170–192. https://doi.org/10.1080/01436597.2024.2355163

Barry, B. (1974). Review Article: 'Exit, Voice, and Loyalty.' *British Journal of Political Science*, *4*(1), 79–107. https://doi.org/10.1017/S0007123400009376

Bellemare, M. F. & Wichman, C. J. (2019). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin Of Economics And Statistics*, *82*(1), 50–61. https://doi.org/10.1111/obes.12325

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The Augmented Synthetic Control Method. *Journal of the American Statistical Association*, *116*(536), 1789–1803. https://doi.org/10.1080/01621459.2021.1929245

Bermeo, N. (2016). On Democratic Backsliding. *Journal of Democracy*, *27*(1), 5–19.

Brito Vieira, M. (2021). The great wall of silence: Voice–silence dynamics in authoritarian regimes. *Critical Review of International Social and Political Philosophy*, *24*(3), 368–391. https://doi.org/10.1080/13698230.2020.1796333

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

Carter, E. B., Carter, B. L., & Schick, S. (2024). Do Chinese Citizens Conceal Opposition to the CCP in Surveys? Evidence from Two Experiments. *The China Quarterly*, *259*, 804–813. https://doi.org/10.1017/S0305741023001819

Chen, Z., Jardine, E., Fan Liu, X., & Zhu, J. J. H. (2024). Seeking anonymity on the Internet: The knowledge accumulation process and global usage of the Tor network. *New Media & Society*, *26*(2), 1074–1095. https://doi.org/10.1177/14614448211072201

Coppedge, M. (2023). *V-Dem's Conceptions of Democracy and Their Consequences* (SSRN Scholarly Paper No. 4410921). Social Science Research Network. https://doi.org/10.2139/ssrn.4410921

De Mesquita, E. B. (2010). Regime Change and Revolutionary Entrepreneurs. *The American Political Science Review*, *104*(3), 446–466.

Dowding, K. (2015). Albert O. Hirschman, Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations, and States. In M. Lodge, E. C. Page, & S. J. Balla (Eds.), *The Oxford Handbook of Classics in Public Policy and Administration* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199646135.013.30

Dowding, K., John, P., Mergoupis, T., & Van Vugt, M. (2000). Exit, voice and loyalty: Analytic and empirical developments. *European Journal of Political Research*, *37*(4), 469–495. https://doi.org/10.1023/A:1007134730724

Drujon d'Astros, C., Morales, J., & Leca, B. (2024). Accounting and silence: The unspeakable, the unsaid, and the inaudible. *Contemporary Accounting Research*, *41*(3), 1449–1476. https://doi.org/10.1111/1911-3846.12944

Dyne, L. V., Ang, S., & Botero, I. C. (2003). Conceptualizing Employee Silence and Employee Voice as Multidimensional Constructs. *Journal of Management Studies*, *40*(6), 1359–1392. https://doi.org/10.1111/1467-6486.00384

Edmond, C. (2013). Information Manipulation, Coordination, and Regime Change. *The Review of Economic Studies*, *80*(4), 1422–1458. https://doi.org/10.1093/restud/rdt020

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

Filasto, A., & Appelbaum, J. (2012). *OONI: Open Observators of Network Inference* [Dataset]. https://ooni.org/

Frey, B. S. (2023). Camouflage: A dominant reaction to worsening conditions. *Rationality and Society*, *35*(3), 366–384. https://doi.org/10.1177/10434631231157588

Gray, S. W. D. (2021). Silence and democratic institutional design. *Critical Review of International Social and Political Philosophy*, *24*(3), 330–345. https://doi.org/10.1080/13698230.2020.1796331

Hassan, M., Mattingly, D., & Nugent, E. R. (2022). Political Control. *Annual Review of Political Science*, *25*(Volume 25, 2022), 155–174. https://doi.org/10.1146/annurev-polisci-051120-013321

Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Fims, Organizations, and State*. Harvard Univ. Press.

Jardine, E. (2018). Tor, what is it good for? Political repression and the use of online anonymity-granting technologies. *New Media & Society*, *20*(2), 435–452. https://doi.org/10.1177/1461444816639976

Jiang, J., & Yang, D. L. (2016). Lying or Believing? Measuring Preference Falsification From a Political Purge in China. *Comparative Political Studies*, *49*(5), 600–634. https://doi.org/10.1177/0010414015626450

Knoll, M., Neves, P., Schyns, B., & Meyer, B. (2021). A Multi-Level Approach to Direct and Indirect Relationships between Organizational Voice Climate, Team Manager Openness, Implicit Voice Theories, and Silence. *Applied Psychology*, *70*(2), 606–642. https://doi.org/10.1111/apps.12242

Kuran, T. (1987). Preference Falsification, Policy Continuity and Collective Conservatism. *The Economic Journal*, *97*(387), 642–665. https://doi.org/10.2307/2232928

Kuran, T. (1991). Now out of Never: The Element of Surprise in the East European Revolution of 1989. *World Politics*, *44*(1), 7–48. https://doi.org/10.2307/2010422

Kuran, T. (1995). The Inevitability of Future Revolutionary Surprises. *American Journal of Sociology*, *100*(6), 1528–1551. https://doi.org/10.1086/230671

Lichbach, M. I. (1987). Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent. *The Journal Of Conflict Resolution*, *31*(2), 266–297. https://www.jstor.org/stable/174013

Lohmann, S. (1994). Information Aggregation Through Costly Political Action. *The American Economic Review*, *84*(3), 518–530.

Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: What is new about it? *Democratization*, *26*(7), 1095–1113. https://doi.org/10.1080/13510347.2019.1582029

Moore, W. H. (2000). The Repression of Dissent. *Journal Of Conflict Resolution*, *44*(1), 107–127. https://doi.org/10.1177/0022002700044001006

Morrison, E. W. (2023). Employee Voice and Silence: Taking Stock a Decade Later. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(Volume 10, 2023), 79–107. https://doi.org/10.1146/annurev-orgpsych-120920-054654

Noelle-Neumann, E. (1974). The Spiral of Silence a Theory of Public Opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups, With a New Preface and Appendix*. Harvard University Press.

Pan, J., & Siegel, A. A. (2020). How Saudi Crackdowns Fail to Silence Online Dissent. *American Political Science Review*, *114*(1), 109–125. https://doi.org/10.1017/S0003055419000650

Raleigh, C., Linke, rew, Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, *47*(5), 651–660. https://doi.org/10.1177/0022343310378914

Reuter, O. J., & Szakonyi, D. (2021). Electoral Manipulation and Regime Support: Survey Evidence from Russia. *World Politics*, *73*(2), 275–314. https://doi.org/10.1017/S0043887120000234

Roberts, M. E. (2018). *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press. https://doi.org/10.2307/j.ctvc77b21

Robinson, D., & Tannenberg, M. (2019). Self-censorship of regime support in authoritarian states: Evidence from list experiments in China. *Research & Politics*, *6*(3), 2053168019856449. https://doi.org/10.1177/2053168019856449

Ryan, M., & Tran, M. V. (2024). Democratic backsliding disrupted: The role of digitalized resistance in Myanmar. *Asian Journal of Comparative Politics*, *9*(1), 133–158. https://doi.org/10.1177/20578911221125511

Ryu, S., & Chang, Y. (2017). Accountability, Political Views, and Bureaucratic Behavior: A Theoretical Approach. *Public Organization Review*, *17*(4), 481–494. https://doi.org/10.1007/s11115-016-0349-z

Sandberg, B. (2018). The Geography of Anonymous Communications: Predicting Escalation of Anonymity Networks During Events of Civil Unrest. *Journal of Mason Graduate Research*, *5*(2), 85–107. https://doi.org/10.13021/G8jmgr.v5i2.1961

Scott, J. C. (1990). *Domination and the Arts of Resistance: Hidden Transcripts*. Yale University Press. https://www.jstor.org/stable/j.ctt1np6zz

Strauch, R., & Weidmann, N. B. (2022). Protest and digital adaptation. *Research & Politics*, *9*(2), 20531680221100440. https://doi.org/10.1177/20531680221100440

The Tor Project. (2024). Tor Metrics. https://metrics.torproject.org

Tullock, G. (1971). The Paradox of Revolution. *Public Choice*, *11*, 89–99. https://www.jstor.org/stable/30022655

Zhdanov, A., & Tseitlin, B. (2024). Revealing the Preference Falsification in Russia: Assessing the Robustness of the Results of Sociological Surveys. *Comparative Sociology*, *23*(6), 747–781. https://doi.org/10.1163/15691330-bja10123

Appendix

Appendix A – Derivations and parameter restrictions

A.1 Interior channel sorting

Citizens choose among $U(s) = 0, U(h) = \theta_i - c_h$ and $U(v) = B\theta_i - R$, with $B > 1$, any return common to all three actions having been normalised away. The distribution F is continuously differentiable on $(0, 1)$ with density f continuous, strictly positive and bounded above.

The inactivity–concealment threshold solves $\theta - c_h = 0$ and the concealment–public threshold solves $\theta - c_h = B\theta - R$, giving

$$\theta_h^* = c_h, \qquad \theta_{hv}^* = \frac{R - c_h}{B - 1}.$$

For the interior ordering $0 < \theta_h^* < \theta_{hv}^* < 1$ it is sufficient that

$$0 < c_h < 1, \qquad R > Bc_h, \qquad R - c_h < B - 1.$$

The middle condition ensures that concealed activity lies on the upper envelope of the three payoff functions: the inactivity–public indifference point is $\theta_v^* = R/B$, and $\theta_h^* < \theta_v^*$ holds if and only if $R > Bc_h$, which in turn implies $\theta_v^* < \theta_{hv}^*$. Since the payoff slopes satisfy $0 < 1 < B$, inactivity is optimal below θ_h^*, concealed activity between the thresholds and public activity above θ_{hv}^*. Applying F gives

$$n_s = F(c_h), \qquad n_h = F(\theta_{hv}^*) - F(c_h), \qquad n_v = 1 - F(\theta_{hv}^*).$$

Write $\varphi_h = f(\theta_h^*)$ and $\varphi_v = f(\theta_{hv}^*)$.

A.2 Exposure cost and concealed-channel cost

Because θ_h^* does not depend on R, $\partial n_s/\partial R = 0$, while $\partial\theta_{hv}^*/\partial R = 1/(B - 1)$ gives

$$\frac{\partial n_h}{\partial R} = \frac{\varphi_v}{B - 1} > 0, \qquad \frac{\partial n_v}{\partial R} = -\frac{\varphi_v}{B - 1} < 0,$$

which establishes the first row of Table 1. For the concealed-channel cost, $\partial\theta_h^*/\partial c_h = 1$ and $\partial\theta_{hv}^*/\partial c_h = -1/(B - 1)$, so

$$\frac{\partial n_s}{\partial c_h} = \varphi_h > 0, \qquad \frac{\partial n_h}{\partial c_h} = -\frac{\varphi_v}{B - 1} - \varphi_h < 0, \qquad \frac{\partial n_v}{\partial c_h} = \frac{\varphi_v}{B - 1} > 0,$$

which establishes the second row. A rise in the cost of concealment therefore compresses the concealed region from both sides: some citizens cease acting and others move to public activity.

When both costs change,

$$dn_s = \varphi_h\, dc_h, \qquad dn_v = -\frac{\varphi_v(dR - dc_h)}{B - 1},$$

$$dn_h = \frac{\varphi_v(dR - dc_h)}{B - 1} - \varphi_h\, dc_h,$$

equivalently $dn_h = -dn_s - dn_v$, as in Section 3.2. Neither dn_h nor dn_v is signed in general. Concealed activity falls whenever $\varphi_v(dR - dc_h)/(B - 1) < \varphi_h\, dc_h$, equivalently whenever

$$\frac{dc_h}{dR} > \frac{\varphi_v}{\varphi_v + (B-1)\varphi_h},$$

a threshold lying strictly between zero and one. An increase in concealed-channel costs smaller than the increase in exposure costs can therefore generate a decline in observed concealed activity, which establishes the fourth row.

A.3 Political activation

Under an activation shock citizens act on $\theta_i + g$, so the relevant thresholds in the original grievance distribution are $c_h - g$ and $\theta^*_{hv} - g$, and the shares become

$$n_s = F(c_h - g), \quad n_h = F(\theta^*_{hv} - g) - F(c_h - g), \quad n_v = 1 - F(\theta^*_{hv} - g).$$

Evaluating locally at $g = 0$,

$$\frac{\partial n_s}{\partial g} = -\varphi_h < 0, \qquad \frac{\partial n_v}{\partial g} = \varphi_v > 0, \qquad \frac{\partial n_h}{\partial g} = \varphi_h - \varphi_v,$$

so that $\partial(n_h + n_v)/\partial g = \varphi_h > 0$ while the effect on concealed activity is ambiguous: citizens enter the concealed channel from inactivity at the lower threshold and leave it for public activity at the upper one. This establishes the third row.

A.4 Relative visibility

With $\rho = log\, n_h - log\, n_v$, $d\rho = dn_h/n_h - dn_v/n_v$. Under a pure increase in *R*,

$$\frac{\partial \rho}{\partial R} = \left[\frac{\varphi_v}{B-1}\right]\left(\frac{1}{n_h} + \frac{1}{n_v}\right) > 0,$$

and under a pure increase in c_h the two terms again have opposite signs, so $\partial\rho/\partial c_h < 0$. Under a pure activation shock,

$$\frac{\partial \rho}{\partial g} = \frac{(\varphi_h - \varphi_v)}{n_h} - \frac{\varphi_v}{n_v},$$

which is negative if and only if $\varphi_h < \varphi_v(1 + n_h/n_v)$. Since this depends on the relative density of grievance at the two thresholds, we make no general sign prediction for relative visibility under activation.

When exposure and concealed-channel costs rise together, dn_h and dn_v are both unsigned by A.2, so $d\rho$ is unsigned as well; this establishes the final entry of Table 1. The model therefore does not claim that the relative-visibility outcome structurally identifies a single underlying parameter; combining the level of concealed activity with its movement relative to public dissent provides a transparent classification of the patterns the framework generates.

Appendix B – Participation externality

This appendix is supplementary; the argument in the main text does not depend on it. The model treats B as a parameter in order to isolate the visibility margin, while existing work shows that political participation also depends on expectations about the participation of others. We verify here that allowing for this changes none of the signs in Table 1.

B.1 Equilibrium

Let the return to public activity depend on the mass n choosing it,

$$b(n) = \beta + \gamma(n),$$

where β is the return that accrues independently of aggregate participation and $\gamma(n)$ the additional benefit of acting alongside others. Public activity now yields $b(n_v^*)\theta_i - R$, while the payoffs to inactivity and concealed activity are unchanged. The baseline of Section 3 is the special case $\gamma \equiv 0$, in which $b(n) = \beta = B$ for every participation rate.

Assumption B.1. Maintaining $0 < c_h < 1$ from Appendix A.1, γ is continuously differentiable and increasing on $[0,1]$ with $\gamma(0) = 0$ and derivative bounded above. In addition, $\beta > 1$, $R - c_h < \beta - 1$, and $c_h b(1) < R$.

The first restriction ensures that public activity has a higher marginal return than concealment even at zero participation, the second that a positive mass of high-grievance citizens acts publicly even when expected turnout is zero, and the third that the three-channel ordering of Appendix A.1 holds for every participation belief.

Definition B.1. A nonatomic Nash equilibrium is a share $n_v^* \in [0,1]$ such that every citizen chooses an optimal action given aggregate participation n_v^*, and the resulting mass choosing public activity is itself n_v^*.

Given a conjectured participation rate n, the sorting result of Appendix A.1 applies with B replaced by $b(n)$, so the implied mass choosing public activity is

$$T(n) = 1 - F\left(\frac{R - c_h}{b(n) - 1}\right),$$

and an equilibrium is a fixed point $n_v^* = T(n_v^*)$. Appendices B.1 and B.2 set $g = 0$; political activation is introduced in B.3.

Proposition B.1 (Existence). At least one equilibrium exists, and every equilibrium has $n_v^* > 0$.

Proof. Under Assumption B.1, $b(n) - 1 \geq \beta - 1 > 0$, so T is continuous and maps $[0,1]$ into itself, and Brouwer's theorem gives a fixed point. Moreover $R - c_h < \beta - 1$ implies $(R - c_h)/(b(0) - 1) < 1$, so $T(0) > 0$ and $n = 0$ cannot be an equilibrium. □

B.2 Uniqueness

The participation externality creates strategic complementarity. Differentiating the best-response map,

$$T'(n) = \frac{f\left(\theta_{hv}^*(n)\right)(R - c_h)\gamma'(n)}{[b(n) - 1]^2} \geq 0,$$

where $\theta_{hv}^*(n) = (R - c_h)/(b(n) - 1)$, with strict inequality wherever $\gamma'(n) > 0$.

Proposition B.2 (Uniqueness under weak complementarity). Let $\bar{f}$ bound f on $[0,1]$ and $\bar{\gamma'}$ bound γ' on $[0,1]$. If $\bar{f}(R - c_h)\bar{\gamma'}/(\beta - 1)^2 < 1$, the equilibrium is unique.

Proof. Since $b(n) - 1 \geq \beta - 1$ for all n, $|T'(n)|$ is bounded by the stated expression, which is smaller than one, so T is a contraction on $[0,1]$ and Banach's theorem gives a unique fixed point. □

The condition is sufficient, not necessary. Where it fails, multiplicity may but need not arise, and we consider local comparative statics at stable equilibria rather than imposing a global selection rule.

B.3 Comparative statics at stable equilibria

To study political activation, extend the best-response map locally around $g = 0$ to

$$T(n;\, g) = 1 - F\left(\frac{R - c_h}{b(n) - 1} - g\right),$$

with the corresponding inactivity threshold $c_h - g$, as in Appendix A.3. All comparative statics below are evaluated locally at $g = 0$.

Write

$$D^* = b(n_v^*) - 1,\ \theta_{hv}^* = \frac{R - c_h}{D^*},\ \varphi_h = f(c_h),\ \varphi_v = f(\theta_{hv}^*),$$

and define the strategic-feedback term

$$\Gamma^* = \frac{\varphi_v(R - c_h)\gamma'(n_v^*)}{(D^*)^2},$$

so that under best-response dynamics an equilibrium is locally stable when $\Gamma^* < 1$.

Proposition B.3. At any stable equilibrium,

$$\frac{dn_v^*}{dR} = -\frac{\varphi_v}{[D^*(1 - \Gamma^*)]} < 0, \qquad \frac{dn_h^*}{dR} = \frac{\varphi_v}{[D^*(1 - \Gamma^*)]} > 0,$$

$$\frac{dn_v^*}{dc_h} = \frac{\varphi_v}{[D^*(1 - \Gamma^*)]} > 0, \qquad \frac{dn_h^*}{dc_h} = -\frac{\varphi_v}{[D^*(1 - \Gamma^*)]} - \varphi_h < 0,$$

$$\frac{dn_v^*}{dg} = \frac{\varphi_v}{(1 - \Gamma^*)} > 0, \qquad \frac{dn_s}{dg} = -\varphi_h < 0,$$

$$\frac{dn_h^*}{dg} = \varphi_h - \frac{\varphi_v}{(1 - \Gamma^*)},$$

the last of which is unsigned. Writing $A^* = \frac{\varphi_v}{[D^*(1 - \Gamma^*)]} > 0$ and $\rho^* = log\, n_h^* - log\, n_v^*$,

$$\frac{d\rho^*}{dR} = A^*\left(\frac{1}{n_h^*} + \frac{1}{n_v^*}\right) > 0,$$

$$\frac{d\rho^*}{dc_h} = -\frac{(A^* + \varphi_h)}{n_h^*} - \frac{A^*}{n_v^*} < 0,$$

$$\frac{d\rho^*}{dg} = \frac{\left(\varphi_h - \frac{\varphi_v}{(1 - \Gamma^*)}\right)}{n_h^*} - \frac{\frac{\varphi_v}{(1 - \Gamma^*)}}{n_v^*},$$

which is unsigned, as is the response of ρ^* to a joint increase in R and c_h.

Proof. Totally differentiating the equilibrium condition $n_v^* = T(n_v^*; x)$ for a generic parameter x gives $dn_v^*/dx = T_x/(1 - T_n)$, with $T_n = \Gamma^* < 1$ at a stable equilibrium. Direct differentiation gives $T_R = -\varphi_v/D^* < 0$, $T_{c_h} = \varphi_v/D^* > 0$ and $T_g = \varphi_v > 0$, which yields the three expressions for n_v^*. Since the inactivity threshold $\theta_h^* = c_h$ is independent of both R and aggregate participation, an increase in R reallocates mass one for one from public to concealed activity; an increase in c_h additionally removes mass through the rising inactivity threshold; and an activation shock lowers that threshold, giving $dn_s/dg = -\varphi_h$. In each case the response of concealed activity follows as the residual, $dn_h^* = -dn_s - dn_v^*$. □

B.4 Scope of the extension

Three points are worth emphasising. First, strategic complementarity strengthens rather than reverses the responses that are sign-determined in the baseline: at a stable equilibrium $0 \leq \Gamma^* < 1$, so the multiplier $1/(1 - \Gamma^*)$ is at least one, and as citizens leave the public channel its expected return falls, inducing further movement away from it. Second, strategic complementarity amplifies the response of public participation without reversing any of the sign-determined comparative statics, and responses that are ambiguous in the baseline remain ambiguous under the extension; for a given parameterisation the multiplier can alter the sign of an ambiguous entry, so what carries over is the classification of responses into determinate and ambiguous rather than their numerical values. Third, the signs are local properties of every stable equilibrium and do not rule out discrete transitions between equilibrium branches following sufficiently large shocks where multiplicity is present. The sign structure of Table 1, including the relative-concealment column, therefore survives the introduction of a participation externality, and the activation component requires no additional restrictions.

Appendix C – Case selection, coverage and measurement

Table C1. Candidate country-years (one-year V-Dem Political Liberties decline ≥ 0.10, 2010–2023)

Country	Year(s)	V-Dem drop	Status	Shock event / Reason for exclusion
Panel A. Headline treated set				
Burma/Myanmar	2021	0.540	Headline	Military coup (1 February 2021)
Thailand	2014	0.388	Headline	Royal Thai Armed Forces coup (22 May 2014)
Egypt	2013	0.336	Headline	Military removal of Morsi (3 July 2013)
Belarus	2020	0.226	Headline	Disputed presidential election (9 August 2020)
Türkiye	2016	0.190	Headline	Failed coup attempt and post-coup purge (15 July 2016)
Hong Kong	2020	0.140	Headline	National Security Law imposed (30 June 2020)
Russia	2022	0.123	Headline	Invasion of Ukraine and domestic crackdown (24 February 2022)
Cambodia	2017	0.121	Headline	Dissolution of CNRP opposition party (16 November 2017)
Pakistan	2023	0.106	Headline	Arrest of Imran Khan and mass protests (9 May 2023)
Panel B. Excluded candidates				
Afghanistan	2021	0.536	Excluded	Tor and internet infrastructure collapse following Taliban takeover; ACLED dominated by armed conflict events.
Nicaragua	2018	0.272	Excluded	ACLED Latin America coverage begins in January 2018 (within the shock window); pre-shock standardisation window unavailable.
Bahrain	2011	0.254	Excluded	ACLED Middle East and North Africa coverage begins in 2015; pre-shock standardisation window unavailable.
Yemen	2015	0.239	Excluded	ACLED Middle East and North Africa coverage begins in 2015 (shock year); active civil war contaminates ACLED counts.
Ukraine	2014, 2022	0.120, 0.222	Excluded	2014: ACLED Eastern Europe coverage begins January 2018; pre-shock window unavailable. 2022: Same global event as Russia 2022; same-event contamination of the donor pool.
Ethiopia	2021	0.144	Excluded	Active Tigray war (from November 2020) dominates ACLED counts with armed conflict rather than political dissent.
Poland	2016	0.143	Excluded	Gradual erosion through PiS judicial reforms across 2015–2018; no single dateable institutional event satisfying the discreteness criterion.
Burkina Faso	2023	0.136	Excluded	Two coups within nine months (January and September 2022) prevent identification of a single discrete shock and contaminate the pre-shock baseline.

Country	Year(s)	V-Dem drop	Status	Shock event / Reason for exclusion
Central African Republic	2013, 2019	0.121, 0.129	Excluded	Active armed conflict throughout the period; ACLED counts dominated by battle and explosion events; very low Tor adoption.
Maldives	2012, 2013	0.113, 0.129	Excluded	ACLED coverage of small states begins January 2020; pre-shock standardisation window unavailable.
South Sudan	2013	0.122	Excluded	Civil war begins in December 2013; ACLED counts dominated by armed conflict events.
Bolivia	2019	0.109	Excluded	ACLED Latin America coverage begins January 2018, leaving an inadequate pre-shock standardisation window; institutional decline rapidly reversed under Áñez/Arce.
India	2015	0.105	Excluded	Gradual democratic backsliding under Modi government from 2014 onward; no single dateable institutional event satisfying the discreteness criterion.
Slovenia	2020	0.104	Excluded	Gradual press-freedom decline under Janša government; no single dateable institutional event satisfying the discreteness criterion.
Serbia	2013	0.102	Excluded	Gradual pre-EU-candidacy changes; no single dateable institutional event satisfying the discreteness criterion.
Mauritania	2013	0.100	Excluded	No single dateable institutional event satisfying the discreteness criterion.

Notes. Candidate country-years are identified by a one-year decline of at least 0.10 on V-Dem's Political Liberties Index (v2x_clpol) between 2010 and 2023, yielding 33 country-years across 28 unique countries (Coppedge et al., 2023), of which the 28 reported here comprise the nine treated cases and nineteen exclusions; five further candidate country-years are excluded on the criteria below and are not listed individually. When a country qualifies in multiple years, all qualifying years are reported on a single line. Panel A reports the nine treated cases of Table 2. Panel B documents the remaining exclusions. The three filtering criteria are a discrete and datable institutional event, outcome coverage over the 96-week pre-shock and 52-week post-shock comparison window, and the absence of active armed conflict dominating ACLED counts.

Table C2. Directly estimated relative concealment against the difference of the component effect

Case	Direct estimate (log pts)	Difference of component ATTs	Absolute difference
Pakistan	+0.01	−0.06	0.07
Cambodia	+1.16	+1.47	0.31
Russia	−0.67	−0.23	0.45
Hong Kong	+0.83	+1.05	0.22
Belarus	−4.12	−4.29	0.17
Egypt	−0.44	−0.39	0.05
Thailand	+0.99	+1.41	0.42
Myanmar	−2.61	−2.52	0.09

Notes. The direct estimate is the synthetic-control effect on log(Tor) − log(1 + protest), estimated as a separate outcome and converted to log points. The difference of component ATTs subtracts the protest effect in asinh points from the Tor effect in log points. The two do not coincide, because the component estimates rest on different synthetic counterfactuals and because the protest series enters the two calculations under different transformations. The sign agrees in seven of the eight cases, differing only for Pakistan, where both quantities are indistinguishable from zero. No case classification in Table 6 depends on which is used. Türkiye is omitted for lack of a protest series.

Table C3. Bridge-augmented concealed activity

Case	Direct only (%)	p	Direct + bridge (%)	p	Pre-RMSPE (direct + bridge)
Pakistan	−8	.340	−7	.340	0.053
Cambodia	+40	.308	+46	.192	0.053
Russia	−49	.021**	−41	.021**	0.068
Hong Kong	+54	.065*	+57	.065*	0.085
Türkiye	−11	.150	+4	.750	0.045
Belarus	−89	.022**	−59	.065*	0.190
Egypt	+86	.278	+87	.278	0.081
Thailand	+4	.789	+4	.790	0.075
Myanmar	+118	.239	+144	.174	0.229

Notes. The combined series is the sum of directly connecting and bridge users, estimated on the same donor pools and pretreatment scale as the direct-only specification in Table 3, with identical in-space placebo inference. A user who switches to a bridge relay remains within the concealed channel, so the combined series is the broader measure of that channel. Weeks with no reported bridge figure are treated as zero bridge users. The Hong Kong estimate is unchanged; the Belarus contraction attenuates by roughly a third and its permutation probability rises from .022 to .065; the Türkiye estimate becomes indistinguishable from zero, indicating that its decline in direct users reflects substitution into bridges rather than a contraction of concealed activity. No bridge-augmented counterpart of the relative-concealment outcome is estimated, so this check covers the first condition of the camouflage criterion only. * $p < .10$. ** $p < .05$.

Table C4. Synthetic-control donor weights, direct Tor use

Case	Pool size	Donors with weight	Five highest-weighted donors
Pakistan	46	43	Albania (0.35), Indonesia (0.26), Morocco (0.25), Vietnam (0.20), Senegal (0.07)
Cambodia	25	18	Lebanon (0.24), Sri Lanka (0.19), Nigeria (0.14), Indonesia (0.11), Vietnam (0.06)
Russia	46	46	Malaysia (0.30), Honduras (0.27), Albania (−0.27), Argentina (0.25), Vietnam (−0.17)
Hong Kong	45	45	Malaysia (0.27), Philippines (−0.22), Kazakhstan (0.19), Malawi (0.18), Dominican Rep. (0.17)
Türkiye	19	13	Morocco (0.37), Algeria (0.19), Tanzania (0.13), Cameroon (0.12), Nepal (0.08)
Belarus	45	45	Kazakhstan (0.78), Rwanda (0.42), Vietnam (−0.35), Malawi (0.35), Argentina (−0.35)
Egypt	17	15	Sri Lanka (0.21), Algeria (0.17), Rwanda (0.16), Cameroon (0.13), Mozambique (0.09)
Thailand	18	18	Vietnam (0.73), Senegal (0.16), Mozambique (0.13), Ghana (−0.09), Cameroon (−0.08)
Myanmar	45	42	Nepal (0.58), Malaysia (0.21), North Macedonia (0.17), Laos (0.09), Sri Lanka (−0.03)

Notes. Pool size is the number of screened donors available to each case; donors with weight is the number receiving an absolute weight above 0.001. Donor pools are screened on coverage over the comparison window, non-degenerate pretreatment variation, and the absence of a concurrent one-year V-Dem decline of 0.10 or more. Weights come from the ridge-augmented estimator, which does not constrain them to the simplex, so some are negative: negative weights permit extrapolation outside the convex hull of the donor outcomes, which the unaugmented synthetic control rules out by construction. This is the bias–variance trade-off discussed by Ben-Michael, Feller and Rothstein (2021), and Table D5 reports the unaugmented specification. Rows are ordered as in Table 3.

Table C5. Tor channel composition pre- and post-shock in levels

Country	Direct pre	Direct post	Bridge pre	Bridge post	Bridge % pre	Bridge % post
Belarus	13,812	1,716	1,249	1,964	8.3	53.4
Russia	292,067	103,739	15,802	38,950	5.1	27.3
Cambodia	352	619	26	147	6.8	19.2
Türkiye	15,068	13,317	358	2,045	2.3	13.3
Myanmar	483	1,616	12	183	2.4	10.2
Thailand	42,794	22,852	8	32	0.0	0.1
Pakistan	18,601	14,801	209	166	1.1	1.1
Egypt	2,003	17,902	1	8	0.1	0.0
Hong Kong	7,235	10,175	218	193	2.9	1.9

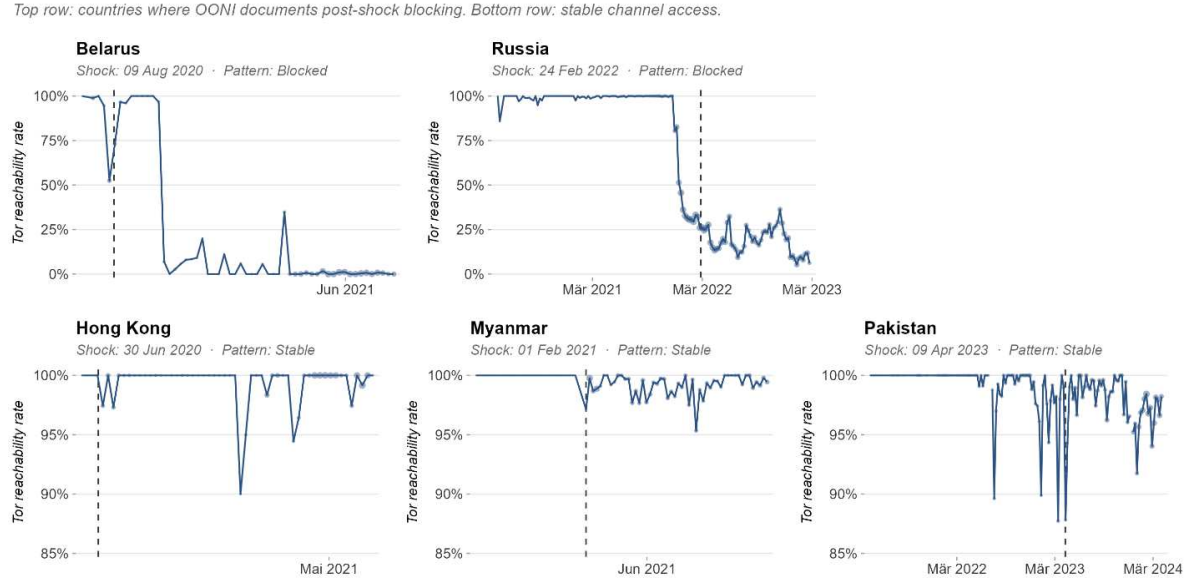
Notes. Mean weekly estimated users in the 52 complete weeks either side of each shock, excluding the event week; bridge share is the share of total Tor users routed through bridges. These are descriptive comparisons, not synthetic-control estimates, and are the levels behind the shares in Table 5. In Belarus, bridge users rise by roughly 57 per cent in absolute terms while total concealed activity falls by three quarters, so circumvention is real but does not compensate. In Türkiye, bridge users rise more than fivefold while the total is unchanged. Rows are ordered by the change in bridge share, as in Table 5.

Table C6. Tor reachability rates pre- and post-shock, by country

Country	Shock date	Pre (%)	Post (%)	Δ (pp)	N tests pre	N tests post	Pattern
Pakistan	2023-05-09	98.1	97.3	−0.8	7,726	15,246	Stable
Hong Kong	2020-06-30	100.0	99.5	−0.5	6	1,698	Stable
Myanmar	2021-02-01	100.0	99.1	−0.9	59	17,731	Stable
Russia	2022-02-24	51.9	18.9	−33.0	124,546	275,968	Blocked
Belarus	2020-08-09	98.0	5.4	−92.6	197	4,946	Blocked
Thailand	*2014-05-22*	—	—	—	—	—	*Insufficient OONI coverage*
Türkiye	*2016-07-15*	—	—	—	—	—	*Insufficient OONI coverage*
Egypt	*2013-07-03*	—	—	—	—	—	*Insufficient OONI coverage*
Cambodia	*2017-11-16*	—	—	—	—	—	*Insufficient OONI coverage*

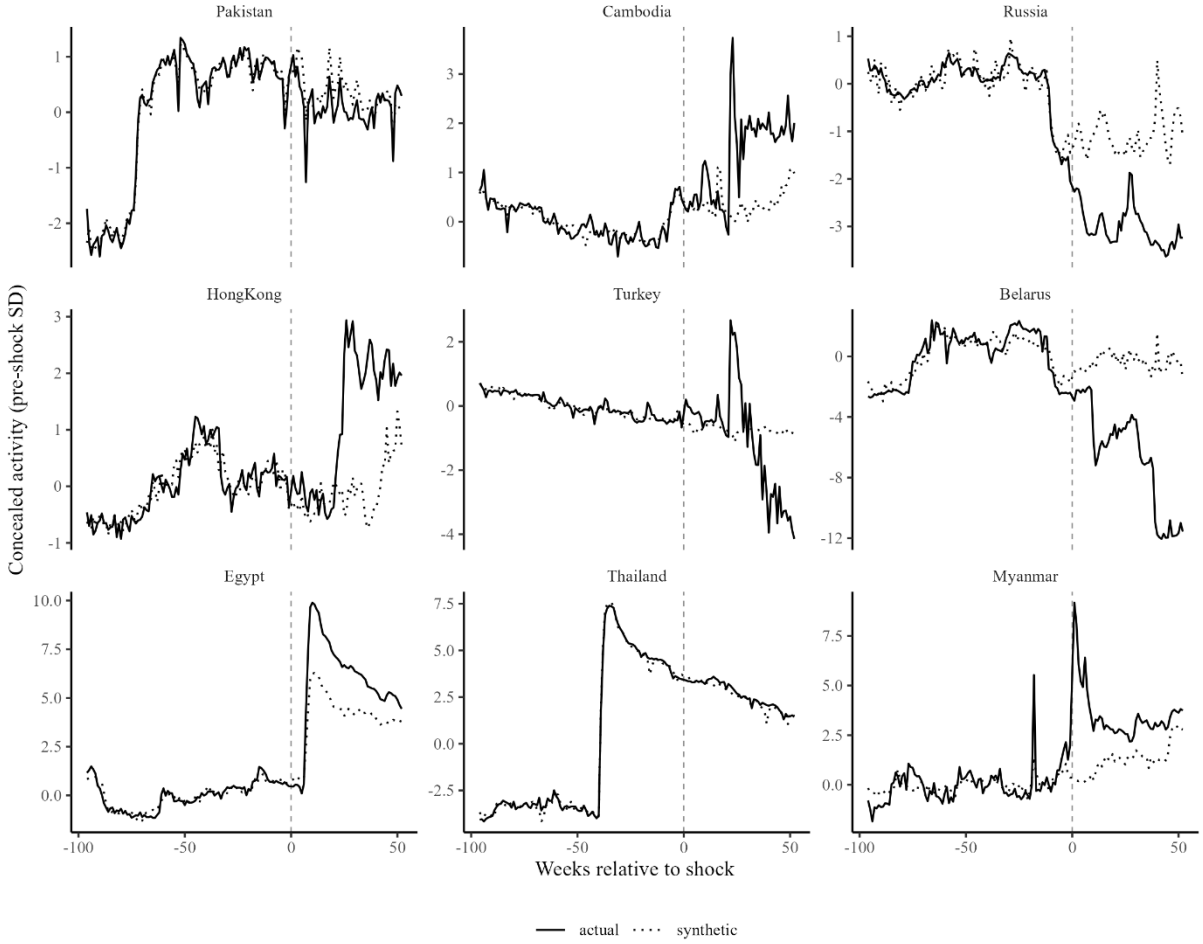
Notes. Rates are computed over the 52 complete weeks either side of each shock, excluding the event week, matching Table 5. A case is read as a major collapse where reachability falls by more than 30 percentage points and as stable where it changes by less than 10. The number of measurements behind these rates varies enormously. On the wider ±104-week window used in an earlier version the pre-shock counts were 6 for Hong Kong, 59 for Myanmar, 197 for Belarus, 7,726 for Pakistan and 124,546 for Russia. The Hong Kong pre-shock figure therefore rests on very few probes, and the stability claim in Section 5.3 should be read with that in mind. Data source: OONI Probe (https://ooni.org/data).

Figure C1. OONI Tor reachability around the shock date, five countries with coverage



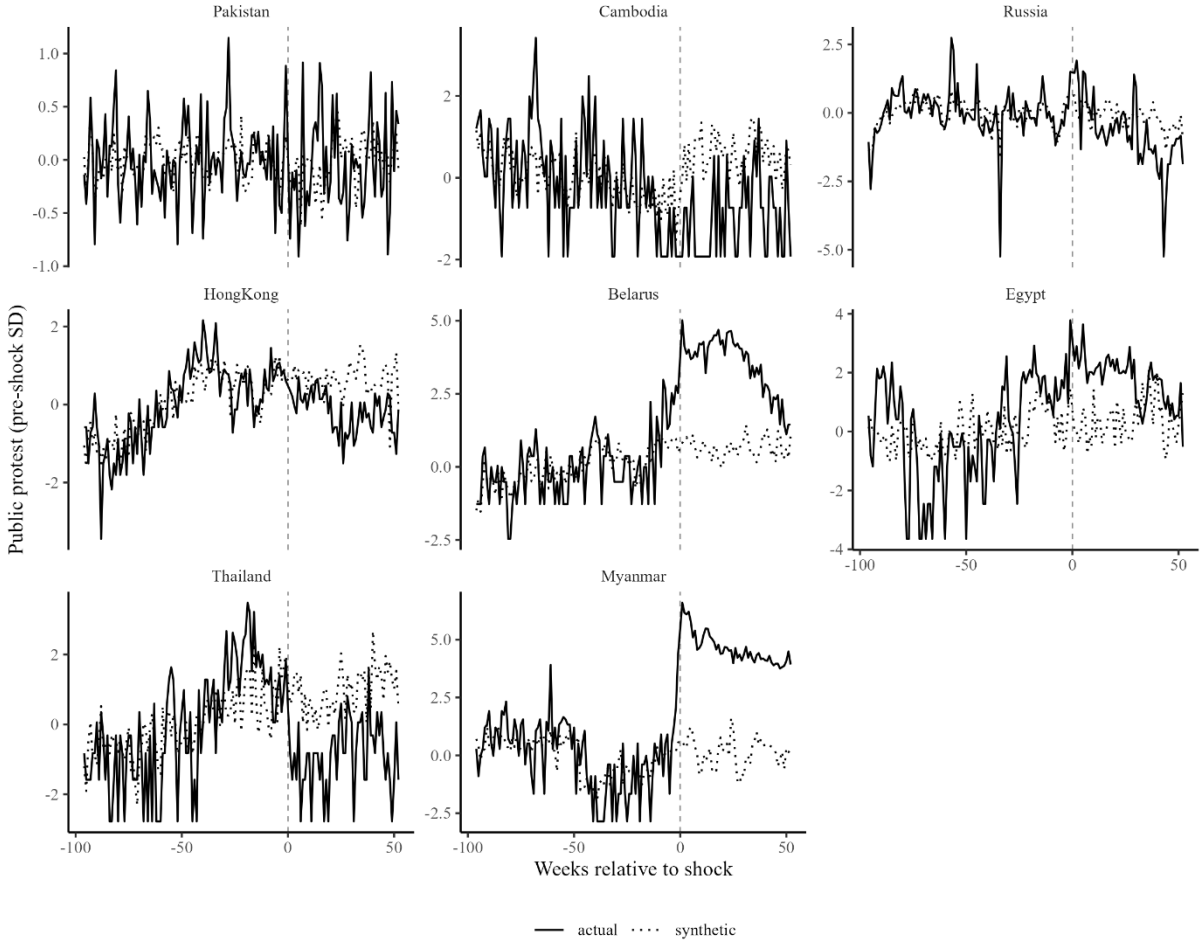
Notes. Weekly Tor reachability rate (share of successful OONI measurements per week), centred on each country's shock date (vertical dashed line). Point size is proportional to the number of OONI measurements in the week. Top row: countries where reachability collapses post-shock, consistent with active blocking of the Tor relay network. Bottom row: countries where reachability remains stable around the shock, supporting the interpretation of the direct Tor user series as a clean substitution measurement. Y-axis scales differ between blocked and stable panels to make the within-group pattern visible. Data source: OONI Probe (https://ooni.org/data).

Figure C2. Direct Tor use: treated and synthetic trajectories



Notes. Weekly direct Tor use for each treated country (solid) and its synthetic comparison (dotted), over the 96 weeks before and 52 weeks after the institutional rupture; the vertical dashed line marks the shock week, which is excluded from estimation. Panels are ordered by the size of the one-year V-Dem decline, as in Table 3, and vertical scales differ across panels. The vertical axis is in the standardised units used for estimation; multiplying by 0.348 converts to the log points reported in Table 3. Pre-treatment fit is visible as the closeness of the two series to the left of the line.

Figure C3. Public protest: treated and synthetic trajectories



Notes. Weekly public protest for each treated country (solid) and its synthetic comparison (dotted), constructed as in Figure C2, for the inverse hyperbolic sine of weekly ACLED protest events. The outcome scale is 0.739, giving asinh points as reported in Table 4. Pretreatment fit is weaker than for direct Tor use, reflecting the greater week-to-week volatility of protest events, and is poorest for Egypt. Türkiye is omitted; see Table 2.

Appendix D – Robustness

Table D1. Leave-one-donor-out sensitivity

Case	Outcome	Baseline	Minimum	Maximum	Sign stable
Pakistan	Tor	−0.08	−0.17	−0.02	Yes
Pakistan	Protest	−0.02	−0.04	−0.01	Yes
Pakistan	Relative	+0.01	−0.03	+0.02	No
Cambodia	Tor	+0.33	+0.26	+0.38	Yes
Cambodia	Protest	−1.13	−1.30	−0.79	Yes
Cambodia	Relative	+1.16	+0.85	+1.37	Yes
Russia	Tor	−0.68	−0.88	−0.51	Yes
Russia	Protest	−0.45	−0.48	−0.33	Yes
Russia	Relative	−0.67	−0.88	−0.58	Yes
Hong Kong	Tor	+0.43	+0.39	+0.46	Yes
Hong Kong	Protest	−0.62	−0.81	−0.49	Yes
Hong Kong	Relative	+0.83	+0.77	+0.91	Yes
Türkiye	Tor	−0.12	−0.14	−0.09	Yes
Belarus	Tor	−2.20	−2.24	−1.90	Yes
Belarus	Protest	+2.09	+2.01	+2.16	Yes
Belarus	Relative	−4.12	−4.34	−4.02	Yes
Egypt	Tor	+0.62	+0.47	+0.68	Yes
Egypt	Protest	+1.01	+0.99	+1.12	Yes
Egypt	Relative	−0.44	−0.76	−0.36	Yes
Thailand	Tor	+0.04	+0.01	+0.17	Yes
Thailand	Protest	−1.37	−1.43	−1.14	Yes
Thailand	Relative	+0.99	+0.94	+1.04	Yes
Myanmar	Tor	+0.78	+0.74	+0.85	Yes
Myanmar	Protest	+3.30	+3.25	+3.35	Yes
Myanmar	Relative	−2.61	−2.70	−2.45	Yes

Notes. Each donor receiving an absolute weight above 0.05 is removed in turn and the case re-estimated, giving 237 refits across the 25 case-outcome combinations. Minimum and maximum are the extremes of the resulting estimates. The sign is stable in 24 of 25 combinations; the exception is the Pakistan relative-concealment estimate, which moves between −0.03 and +0.02 around a baseline of +0.01 and is indistinguishable from zero throughout (p = .957). Among the significant estimates the Hong Kong results are the most stable and the Russia Tor estimate the most sensitive.

Table D2. Pretreatment window sensitivity, direct Tor use

Case	ATT 96 weeks	ATT 52 weeks	Change (%) 96 wk	Change (%) 52 wk	Pre-RMSPE 96 weeks	Pre-RMSPE 52 weeks
Pakistan	−0.08	−0.16	−8	−15	0.053	0.042
Cambodia	+0.33	+0.41	+40	+50	0.054	0.058
Russia	−0.68	−0.68	−49	−49	0.075	0.060
Hong Kong	+0.43	+0.30	+54	+35	0.085	0.091
Türkiye	−0.12	−0.09	−11	−8	0.044	0.044
Belarus	−2.20	−2.59	−89	−93	0.241	0.131

Case	ATT 96 weeks	ATT 52 weeks	Change (%) 96 wk	Change (%) 52 wk	Pre-RMSPE 96 weeks	Pre-RMSPE 52 weeks
Egypt	+0.62	+0.28	+86	+32	0.081	0.039
Thailand	+0.04	+0.06	+4	+6	0.075	0.047
Myanmar	+0.78	+0.80	+118	+123	0.239	0.235

Notes. The 96-week column reproduces the main specification of Table 3. The sign is preserved in all nine cases. Two estimates move materially: Hong Kong attenuates from +54 to +35 per cent and Egypt from +86 to +32 per cent. Pretreatment fit is mechanically closer on the shorter window in seven of nine cases, since fewer weeks have to be matched, so the shorter window buys fit at the cost of a less demanding test of the counterfactual.

Table D3. Region-restricted donor pools, direct Tor use

Case	Region	ATT full pool	ATT regional	Change (%) full	Change (%) regional	Pre-RMSPE full	Pre-RMSPE regional	Donors full	Donors regional
Pakistan	Asia-Pacific	−0.08	+0.04	−8	+4	0.053	0.078	46	13
Cambodia	Asia-Pacific	+0.33	+0.08	+40	+9	0.054	0.060	25	7
Russia	Europe/Eurasia	−0.68	−0.89	−49	−59	0.075	0.320	46	7
Hong Kong	Asia-Pacific	+0.43	+0.35	+54	+43	0.085	0.145	45	12
Türkiye	Europe/Eurasia	−0.12	—	−11	n.e.	0.044	—	19	< 3
Belarus	Europe/Eurasia	−2.20	−2.49	−89	−92	0.241	0.560	45	7
Egypt	MENA	+0.62	—	+86	n.e.	0.081	—	17	< 3
Thailand	Asia-Pacific	+0.04	+0.13	+4	+14	0.075	0.104	18	3
Myanmar	Asia-Pacific	+0.78	+0.90	+118	+146	0.239	0.256	45	12

Notes. Donor pools are restricted to the treated country's own region. Egypt and Türkiye retain fewer than the three donors the estimator requires and are not estimable (n.e.). Among the seven estimable cases the sign is preserved in six; Pakistan changes sign, from −8 to +4 per cent, with both values indistinguishable from zero in the main specification. Cambodia attenuates from +40 to +9 per cent, so the directional pattern there rests on the unrestricted pool. Pretreatment fit deteriorates in every case, by a factor of 4.3 for Russia and 2.3 for Belarus, which also accounts for the larger estimates in those two: the regional synthetic unit no longer tracks the pre-treatment path, so the post-treatment gap inflates.

Table D4. Backdated placebo, treatment reassigned to week −48

Case	Outcome	Real ATT	Placebo ATT	Ratio
Pakistan	Tor	−0.08	−0.01	0.12
Pakistan	Protest	−0.02	+0.21	8.44
Pakistan	Relative	+0.01	−0.05	4.49
Cambodia	Tor	+0.33	+0.04	0.13
Cambodia	Protest	−1.13	−0.36	0.32
Cambodia	Relative	+1.16	+0.39	0.34
Russia	Tor	−0.68	+0.25	0.38
Russia	Protest	−0.45	+0.08	0.18
Russia	Relative	−0.68	−0.69	1.02
Hong Kong	Tor	+0.43	+0.31	0.73
Hong Kong	Protest	−0.62	+0.57	0.91

Case	Outcome	Real ATT	Placebo ATT	Ratio
Hong Kong	Relative	+0.83	−0.40	0.48
Türkiye	Tor	−0.12	−0.01	0.10
Belarus	Tor	−2.20	+0.32	0.15
Belarus	Protest	+2.09	+0.70	0.34
Belarus	Relative	−4.12	−0.17	0.04
Egypt	Tor	+0.62	+0.10	0.17
Egypt	Protest	+1.01	+1.01	1.00
Egypt	Relative	−0.44	−0.84	1.90
Thailand	Tor	+0.04	+1.24	33.2
Thailand	Protest	−1.37	+1.28	0.93
Thailand	Relative	+0.99	+0.17	0.17
Myanmar	Tor	+0.78	−0.19	0.24
Myanmar	Protest	+3.30	−1.18	0.36
Myanmar	Relative	−2.61	+1.21	0.46

Notes. Treatment is reassigned to week −48 and the case re-estimated on the 48 weeks preceding that date, with the true post-treatment period discarded. The ratio is the absolute placebo effect divided by the absolute real effect; small values indicate that the design does not generate comparable effects in the pre-treatment period. The ratio is below 0.5 in sixteen of the twenty-five combinations and below one in nineteen. Three qualifications apply to the exceptions. Where the real effect is indistinguishable from zero the ratio has no substantive interpretation, which accounts for the Pakistan protest and relative entries and for Thailand's Tor entry. The placebo is estimated on 48 rather than 96 pretreatment weeks, so it solves a tighter fitting problem and is noisier by construction. And the pretreatment windows of several cases contain genuine political shocks rather than quiet baselines: week −48 falls in late July 2019 for Hong Kong, during the escalation of that year's protest movement; in August 2012 for Egypt, during the post-revolutionary constitutional crisis; and in March 2021 for Russia, shortly after the Navalny protests. The backdated placebo is therefore less informative in this design than in applications where the pretreatment period is politically quiet, and the Hong Kong and Egypt ratios should be read accordingly rather than as evidence that the estimator manufactures effects.

Table D5. Unaugmented synthetic control with weights on the simplex

Case	Outcome	ATT ridge	ATT unaugmented	Pre-RMSPE ridge	Pre-RMSPE unaugmented
Pakistan	Tor	−0.08	−0.07	0.053	0.077
Pakistan	Protest	−0.02	−0.02	0.242	0.251
Pakistan	Relative	+0.01	+0.02	0.269	0.273
Cambodia	Tor	+0.33	+0.34	0.054	0.055
Cambodia	Protest	−1.13	−1.12	0.754	0.798
Cambodia	Relative	+1.16	+1.18	0.578	0.601
Russia	Tor	−0.68	−1.38	0.075	0.245
Russia	Protest	−0.45	−0.49	0.511	0.548
Russia	Relative	−0.67	−0.73	0.526	0.558
Hong Kong	Tor	+0.43	+0.39	0.085	0.118
Hong Kong	Protest	−0.62	−0.60	0.481	0.494
Hong Kong	Relative	+0.83	+0.80	0.394	0.404
Türkiye	Tor	−0.12	−0.12	0.044	0.044
Belarus	Tor	−2.20	−2.44	0.241	0.466

Case	Outcome	ATT ridge	ATT unaugmented	Pre-RMSPE ridge	Pre-RMSPE unaugmented
Belarus	Protest	+2.09	+2.15	0.651	0.679
Belarus	Relative	−4.12	−4.24	0.783	0.839
Egypt	Tor	+0.62	+0.63	0.081	0.082
Egypt	Protest	+1.01	+1.03	1.197	1.227
Egypt	Relative	−0.44	−0.31	0.951	0.965
Thailand	Tor	+0.04	+0.12	0.075	0.100
Thailand	Protest	−1.37	−1.24	0.931	0.990
Thailand	Relative	+0.99	+0.98	0.692	0.698
Myanmar	Tor	+0.78	+0.79	0.239	0.252
Myanmar	Protest	+3.30	+3.29	0.894	0.946
Myanmar	Relative	−2.61	−2.55	0.755	0.790

Notes. The unaugmented estimator drops the ridge outcome model and retains the synthetic-control weighting, which constrains weights to be non-negative and to sum to one. The smallest weight is exactly zero in all 25 combinations, confirming that no extrapolation outside the convex hull of the donor outcomes occurs in this specification. The sign is preserved in all 25 combinations and the Hong Kong estimates are close to the baseline throughout: +0.39 against +0.43 log points for direct Tor use and +0.80 against +0.83 for relative concealment. Two Tor estimates move materially, Russia from −0.68 to −1.38 log points and Belarus from −2.20 to −2.44, in both cases alongside a pretreatment error that rises by a factor of 3.3 and 1.9 respectively. The augmentation is therefore doing what it is designed to do in those two cases, buying pretreatment fit at the cost of extrapolation, and removing it enlarges rather than removes the estimated contraction