Can Civil Society Organizations Improve Public Service Delivery? Evidence from Construction Monitoring in Brazil*

Manoel Galdino[†], Bianca Vaz Mondo[‡], Juliana Mari Sakai[§], Natália Paiva[¶]

2026

Abstract

This paper examines the effect of bottom-up accountability on public service delivery. We differentiate between information-driven interventions and the mobilization and monitoring efforts of organized Civil Society Organizations (CSOs), and argue that the latter type of interventions can drive significant policy change. The study evaluates the effectiveness of the Obra Transparente project by Brazilian NGO Transparência Brasil, engaging 21 local CSOs in South and Southeast Brazil. Using a difference-in-differences design, we estimate that the intervention increased construction completion rates by approximately 8 percentage points on average (static specification), with suggestive evidence that the dynamic effect may reach up to 18 percentage points approximately five years after the start of the intervention, though this estimate should be interpreted with caution due to a gap in data collection spanning the COVID-19 pandemic. Our findings emphasize the importance of sustained, coordinated efforts by socially embedded CSOs. These efforts, involving direct engagement with municipal officials and ensuring that their complaints cannot be ignored, lead to more substantial outcomes when compared with information-driven interventions targeting individual citizens.

*We thank all participants in the Obra Transparente project; external evaluator Vanda Medeiros; and the teams at Transparência Brasil and OSB. The project was funded by a UNDEF grant (UDF-15-BRA-681). We also thank participants at a USP seminar and at ANPOCS for valuable suggestions. Any remaining errors are our sole responsibility.

[†]University of São Paulo. Corresponding author: mgaldino@usp.br

[‡]Government Transparency Institute, Germany

[§]Transparência Brasil, Brazil

[¶]Alandar Consulting, Brazil

Word count: 5929

1 Introduction

Accountability is a foundational element of democratic governance, crucial for ensuring the effective delivery of public services (Besley and Ghatak 2003; W. Cameron 2004; O'donnell 1998). Traditionally defined as the obligation of public officials to explain and justify their actions (Mulgan 2000; O'Loughlin 1990), the concept has since expanded to include both institutional oversight (O'donnell 2003; Kenney 2003) and popular control (Bertelli and Van Ryzin 2020; Relly 2012; Paul 1992).

Scholars have distinguished between horizontal accountability—oversight among state institutions (O'donnell 2003; Kenney 2003)—and vertical accountability, which refers to how citizens, civil society, and media hold politicians accountable (Bertelli and Van Ryzin 2020; Relly 2012; Paul 1992). A growing body of work has explored "bottom-up" or social accountability mechanisms, in which citizens use information to monitor public policy implementation (Fox 2015). Influential studies, such as *Power to the People* (Björkman and Svensson 2009), inspired optimism about the potential of community-based oversight to improve public services. However, subsequent evaluations have yielded mixed results, particularly for interventions that rely on disorganized individuals and technology-based tools like SMS or apps (Freire, Galdino, and Mignozzetti 2020; Raffler, Posner, and Parkerson 2019; Ankamah 2019).

While these evaluations have advanced the field, they often overlook a key distinction: the difference between citizen-focused interventions and those that strengthen established Civil Society Organizations (CSOs). Embedded in both political and social networks, CSOs may possess greater capacity to generate policy change than isolated individuals (Gong and Xiao 2017; Joshi and Houtzager 2012).

This paper argues that interventions which empower CSOs are more effective than those targeting disorganized citizens. Projects that enhance the ability of CSOs to act differ substantially from those that simply provide information to the public (Mattoni and Odilla 2021; Odilla 2023). Our study contributes to this debate by providing empirical evidence on the effectiveness of organized civil

society action in improving public service delivery.

We evaluate Obra Transparente, a project by the Brazilian NGO Transparência Brasil, which supported 21 local CSOs across municipalities in the South and Southeast regions of Brazil between 2017 and 2019. The initiative aimed to improve educational infrastructure by addressing inefficiencies and delays in the construction of public schools and nurseries. Using a difference-in-differences approach, we estimate that the intervention increased the delivery rate of construction projects by approximately 8 percentage points on average (static specification). The dynamic event-study specification suggests the effect may reach up to 18 percentage points approximately five years after the start of the intervention, though this longer-run estimate is less precisely estimated under conservative inference methods.

These results suggest that organized, sustained civil society engagement—through repeated meetings, follow-ups, and site visits—achieves better results than information-based interventions aimed at individual citizens. Our findings also contribute to Brazilian scholarship showing that institutionalized civil society participation enhances democratic legitimacy and policy responsiveness (Adrián Gurza Lavalle, Houtzager, and Castello 2006; Isunza Vera and Gurza Lavalle 2012; Pogrebinschi and Santos 2011; Adrian Gurza Lavalle, Voigt, and Serafim 2016).

Moreover, Brazil's relatively strong vertical and horizontal accountability institutions make it a relevant case for broader theorization. As a country with accountability levels comparable to Western democracies like France and Spain (Lührmann, Marquardt, and Mechkova 2020), Brazil provides insight into the conditions under which bottom-up accountability can work. This is especially pertinent in an era of democratic backsliding, where civil society plays a critical role in safeguarding institutional performance (Waldner and Lust 2018; Little and Meng 2024; Smidt, Johansson, and Richter 2025).

The paper proceeds as follows. Section 2 introduces the Obra Transparente project. Section 3 lays out the theoretical framework. Section 4 describes the data and empirical strategy. Section 5 presents the quantitative results. Section 6 discusses the mechanisms using qualitative evidence. Section 7 discusses broader implications. Section 8 concludes.

2 Project Obra Transparente

Transparência Brasil (TB) launched the Obra Transparente project in May 2017 to tackle persistent delays and inefficiencies in federally funded school and nursery construction overseen by municipal governments. Partnering with Observatório Social do Brasil, an NGO that coordinates a network of local organizations, TB developed and promoted a standardized methodology for local government oversight. The project concluded on June 30, 2019. The first year (May 2017 to mid-2018) was dedicated to training the local CSOs and building the monitoring infrastructure. Active field monitoring of construction sites began in the second half of 2018, which is why we code the treatment as starting in September 2018 in our quantitative analysis.

During the project, TB monitored construction plans, resource allocation, procurement processes, and work execution. All data and information gathered were shared with the local CSOs in a structured and scheduled manner.

Transparência Brasil collaborated with the Controladoria Geral da União, a federal government agency responsible for overseeing federal public spending, to develop training materials and provide courses to the network of local CSOs. The courses reached 270 participants, featuring a mix of online and in-person modules over several weeks during the project's duration. These modules covered topics related to public bidding, construction monitoring, and contracting procedures.

Armed with the training and information, the local Observatórios Sociais do Brasil (OSBs) were assigned to monitor all schools and nursery construction works funded by the Federal Government as part of the ProInfância program. For works that had already started, they conducted on-site visits to the construction sites as a way of monitoring whether the planned schedules were on time or if there were signs of delay in the completion of the work. For works in the bidding phase, they monitored the contracting process to ensure it followed laws and regulations on public procurement. Upon completion, they learned how to audit the delivery to ensure it followed contracts and rules.

To address any suspicion of lawbreaking, TB developed a "technical chamber" that would provide legal and technical assistance from lawyers and engineers. It was designed to ensure that the monitoring was technically sound and to avoid erroneous accusations of wrongdoing.

3 Theoretical Framework

Effective social accountability in the context of the present study means that the government is responsive to civil society. The punishment dimension inherent in the concept of accountability is absent in the context of social accountability. In contrast to horizontal accountability and vertical accountability, civil society has no means to punish bureaucrats and politicians. At most, it can signal or prompt either audit and judicial institutions or voters to hold office holders accountable.

As a result, there are two primary mechanisms by which direct monitoring of public services and officials by citizens or community groups can influence public policies. The first approach is to utilize what Hirschman called "voice" (Hirschman 1972), in which a group complains, protests, or engages with principals to change or improve practices and policies. Small civil society groups cannot exit, as they will remain part of the community in which they participate as long as they exist as organized groups. As mentioned, they can at most signal to voters to exit, i.e., to vote out of office political representatives, or to comply with or demand from audit and judicial institutions to punish office holders, effectively establishing that relationship (if the punishment occurs).

Civil society groups that are influential and have a broad audience, either directly or indirectly through the media, may credibly claim to have the ability to influence voters and, as such, to have an effect on public policies by threatening to mobilize voters to "opt out". Individual citizens, acting in a disorganized manner, as seen in the numerous failed replications of the *Power to the People* experiment (Freire, Galdino, and Mignozzetti 2020; Ankamah 2019; Fox 2015), have significantly less credible claims to convince voters to exit. That is one possible reason why many of the studies failed to find an effective result.

Nonetheless, it is possible for this mechanism to have an effect, since even disorganized citizens can use their voice to demand improvement in public policies. When effective, the mechanism of voice explains how "bolstering citizen engagement" through monitoring can create a "short route" of accountability between users and providers, circumventing bureaucratic hierarchy.

Professionalized civil society groups, such as anti-corruption NGOs, human rights organizations, and budget transparency initiatives, serve as watchdogs that expose wrongdoing and lobby for reform. They often gather evidence (through audits, reports, and investigative journalism) and mobi-

lize public opinion against abuses of power. These forms of advocacy-oriented civil society action are sometimes termed "diagonal accountability," as they involve citizens and associations collaborating with media or oversight agencies to hold the government accountable (Lührmann, Marquardt, and Mechkova 2020). Thus, in theory, CSOs are expected to be more effective in pressing for change, given their greater capacity to use both voice and exit.

In the case of Project Obra Transparente, the exit mechanism via mobilization of voters is absent, as it is standard practice for Observatórios Sociais do Brasil not to engage the media when identifying governmental problems. Their approach relies instead on engaging directly with local government (use of voice) and, as a resource of last resort, mobilizing audit and judicial institutions. As a result, the hypothesis of the present study is that, in the case of Project Obra Transparente, it is mainly the voice that explains how CSOs can improve public policies.

4 Research Design and Methods

To answer our research question and show that it is the use of voice that explains how CSOs improved the construction of schools and nurseries, we employ a mixed-methods approach, combining qualitative and quantitative evidence. To credibly establish causal evidence linking the monitoring by the CSOs and the delivery of construction works, we use a difference-in-differences method.

To provide evidence about the mechanism (usage of voice), we complement the quantitative analysis with illustrative case studies drawn from project documentation, semi-structured interviews, and an independent evaluation report (Medeiros 2019). These cases provide suggestive evidence about the mechanisms through which CSO monitoring improved construction outcomes.

In this study, we investigate the impact of Civil Society Organization (CSO) monitoring on two aspects: the completion of construction works and the rectification of issues and irregularities found in construction facilities.

4.1 Data

We collected data on school and nursery construction projects across the entire country, encompassing over 2,000 municipalities and more than 14,000 construction projects, spanning all phases from

planning to execution and completion. Here, we focus on construction projects in the five states covered by the project, where municipalities are more comparable. The data come from SIMEC (Sistema Integrado de Monitoramento, Execução e Controle), managed by FNDE (Fundo Nacional de Desenvolvimento da Educação). SIMEC provides up-to-date information about the current status of all school and nursery construction works funded by the federal government. However, it does not keep records of previous statuses. Fortunately, Transparência Brasil kept a record with snapshots of the database from time to time, which we use to track the evolution of completion rates over time. Table 1 below provides descriptive statistics for these construction projects.

Table 1: Summary Statistics by Treatment Status

	Control		Treatment	
Variable	Mean	SD	Mean	SD
Human Development Index (HDI)	0.71	0.05	0.76	0.02
Extreme Poverty Rate (%)	3.72	4.81	1.22	0.57
Per Capita Income (R$/month)	650.74	217.55	895.41	142.74
Poverty Rate (%)	10.76	9.51	5.00	2.06
Income Share of Top 10% (%)	37.04	5.58	39.67	3.94
School Enrollment (Age 15-17) (%)	81.89	6.38	83.43	3.67
Illiteracy Rate (%)	9.42	5.43	4.57	1.85
Population (thousands)	38.31	269.57	217.92	164.39

Note: Descriptive statistics for municipalities participating and not participating in the Obra Transparente project. Treated municipalities tend to have higher income and lower poverty, reflecting the presence of more structured local CSOs.

The qualitative evidence is primarily based on documented reports produced by the local OSBs and collected by Transparência Brasil over the course of the project. We also use evidence collected from semi-structured interviews and an online survey with representatives of the 21 CSOs conducted by an independent evaluator of the project. For the quantitative analysis, we use administrative data about construction works by municipalities and socio-demographic data on municipalities. A detailed description of the data is available from the authors upon request.

5 Results

We estimate the causal effect of the Obra Transparente intervention on the completion rate of construction works using a difference-in-differences design. Since participation in the program is con-

tingent on the presence of local CSOs, treated and control municipalities differ in baseline characteristics, as can be seen in Table 1. They also differ in their completion rates before the start of the project and when it ended (see Table 2). If the municipalities in the treated and control groups were similar, we would expect that they would have similar completion rates years before the project.

A key identifying assumption of DiD is that in the absence of treatment, treated and control groups would have followed parallel trends[1]. This assumption is partially tested, as we can look at pre-trends. Our data support this assumption: with 6 periods of data (2015-2023), we have three pre-treatment periods to test parallel trends. Prior to the intervention (periods -3, -2, and -1), treatment and control groups exhibited nearly identical trends in completion rates, as seen in both the event-study coefficients and graphical analysis. A joint F-test for the pre-treatment coefficients yields $p > 0.90$, consistent with the parallel trends assumption, though we note the test has limited power given the small number of treated clusters (Roth and Sant'Anna 2023). While recent advances in DiD methodology have addressed concerns about staggered treatment adoption and heterogeneous treatment effects (Callaway and Sant'Anna 2021; de Chaisemartin and D'Haultfœuille 2020), these concerns do not apply to our setting because all 21 treated municipalities receive treatment simultaneously in the same period. With a single treatment cohort and common timing, the standard TWFE estimator is numerically equivalent to the modern heterogeneity-robust estimators, and no "forbidden comparisons" arise (Goodman-Bacon 2021). We thus interpret the estimated post-treatment effects as causal impacts of the intervention.

We also assume no interference between units (SUTVA): monitoring in treated municipalities does not affect outcomes in control municipalities. This is plausible given the geographic dispersion of treated municipalities across five states and the fact that there were no official changes in federal oversight behavior.

Importantly, this design does not require treated and control municipalities to be similar in levels (which they are not), but only in (average) trends, which is a weaker and partially empirically testable assumption in our context.

[1]The other identifying assumption is no anticipation: the treatment has no causal effect before its implementation. Although the project launched in May 2017, the first year was devoted to training and infrastructure—active field monitoring began only in the second half of 2018 (coded as September 2018). Since construction projects require a minimum of approximately nine months to complete, it imposes a physical constraint that limits how quickly any behavioral change could translate into higher completion rates.

Table 2: Completion Rates (%) by Period (6 periods, 2015-2023)

Period	Control	Treated	Diff	DiD
Aug 2015	44.6	17.6	-27.1	—
May 2017	55.4	29.8	-25.6	+1.5
Mar 2018	57.2	31.4	-25.8	-0.2
Sep 2018*	59.1	35.1	-24.0	+1.8
Aug 2019*	62.0	40.4	-21.6	+2.4
Oct 2023*	86.5	78.7	-7.8	+13.9

* Post-treatment periods. Diff = Treated − Control. DiD = Diff − lag(Diff).

Our first specification model used the following regression equation:

$$Y_{it} = \alpha_i + \delta_t + \beta\,(D_i \times Post_t) + \varepsilon_{it}$$

Since Y_{it} is binary, this is a Linear Probability Model (LPM), which provides directly interpretable estimates of marginal effects on the probability scale (Angrist and Pischke 2009). Here i indexes construction projects and t the time period. Y_{it} equals 1 if construction project i is completed at time t and 0 otherwise. α_i and δ_t denote project and period fixed effects, respectively. β estimates the average treatment effect on the treated (ATT), where D_i is a binary indicator for whether the municipality of project i participated in Obra Transparente and $Post_t$ indicates the post-treatment period. Note that D_i is absorbed by α_i and $Post_t$ is absorbed by δ_t, so only their interaction is estimated. The error term is ε_{it}, with standard errors clustered at the municipality level.

Table 2 summarizes the completion rates across the 6 periods (2015-2023). In our preferred specification, the effect of the project on the treated units is approximately 8 percentage points (95% CI: 2 to 15 pp), i.e., a municipality monitored by the Obra Transparente project had, on average, 8 pp more constructions completed than what we would observe in absence of the treatment. These results support our argument that sustained and costly bottom-up accountability improves public policies, a point we elaborate with qualitative evidence in Section 6.

As Figure 1 below shows, the control group had a higher construction completion rate than the treated cities, even before the project began. This highlights how any simple comparison between the two groups would be confounded. There is considerable imbalance between treatment and

control groups across many observable covariates. Nonetheless, imbalance in levels does not invalidate a difference-in-differences design, as long as the average temporal evolution of both groups is parallel in the absence of treatment. That is the key identification assumption to be able to establish the causal effect of the treatment on the treated. It is partially testable by examining whether pre-treatment trends are parallel, which they are in our data. As shown above, the pre-treatment trends are parallel, supporting the identifying assumption, though we acknowledge the limitations discussed below.

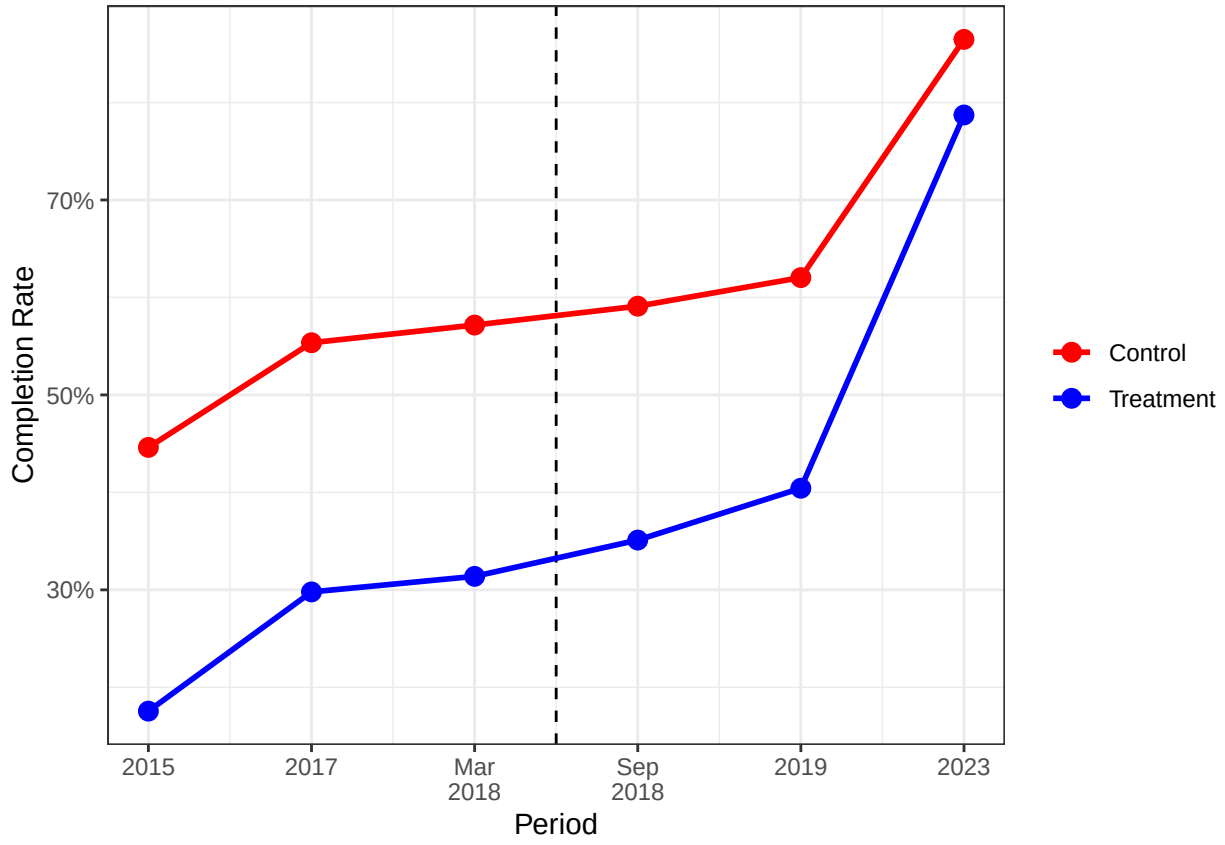

Figure 1: Completion Rates Over Time by Treatment Status (6 periods, 2015-2023). The vertical line indicates the start of treatment (September 2018).

The treatment starts at period 3, and by period 4 it may have begun to produce effects. What we observe is that the effect is small in the initial post-treatment period and becomes visible only after period 4. This suggests dynamic effects (i.e., effects that vary over time), which makes an event-study specification more appropriate.

With a dynamic DiD, the estimated model is:

$$Y_{it} = \alpha_i + \delta_t + \sum_{\substack{k=-3 \\ k \neq -1}}^{2} \beta_k \, \mathbf{1}\{t - T^* = k\} + \varepsilon_{it}$$

where T^* is the treatment period (common to all treated municipalities) and we omit $k = -1$ as the reference point. $\mathbf{1}\{t - T^* = k\}$ is an indicator function that equals 1 when the observation is k periods before treatment $(k < 0)$ or after $(k > 0)$, and α_i and δ_t denote project and period fixed effects, respectively. With 6 periods of data (2015-2023), we have three pre-treatment periods ($k \in \{-3, -2, -1\}$) and three post-treatment periods ($k \in \{0, 1, 2\}$), providing a more robust test of parallel trends.

The results for the static DiD model are presented in Table 3.

Table 3: Static DiD Estimates: Effect of CSO Monitoring on Construction Completion

	(1)
treat_post	0.083
	se = 0.034
	[0.017, 0.150]
Num.Obs.	20 964
R2	0.794
R2 Adj.	0.753
Std.Errors	by: municipio
FE: id	X
FE: periodo	X

Cluster-robust standard errors in parentheses.
95\% confidence intervals in brackets.
Fixed effects by project and period included.

The dynamic DiD specification allows us to trace the evolution of the treatment effect over time. We find no significant differences in the three periods before treatment (placebo test), with coefficients close to zero and statistically insignificant across all pre-treatment periods, consistent with the parallel trends assumption. Post-treatment, the effect emerges gradually and becomes statistically significant at the 5% level in the second period after the intervention (Period +2, p = 0.039) under cluster-robust inference (the wild cluster bootstrap p-value is 0.0498), with an estimated increase in completion rates of approximately 18 percentage points. The permutation test yields p = 0.073.

We therefore present this dynamic effect as suggestive rather than definitive evidence of growing treatment effects. This delay is substantively plausible: construction projects require a minimum of approximately nine months to complete, so even effective monitoring cannot produce measurable increases in completion rates immediately after the intervention begins. Rather, the intervention strengthens oversight that progressively improves implementation over subsequent periods.

The estimated effect in the last period is 0.18, suggesting that treated municipalities may have had approximately 18 percentage points more completed construction works than the control. Since we are estimating the ATT, this is the average causal effect of project Obra Transparente on the 21 treated municipalities in comparison to the control. This is a substantial gain, especially in the context of frequent construction delays and chronic incompletion.

The event study plot provides a clearer visualization of the dynamic effects. It confirms that the pre-trends are parallel, as the pre-treatment coefficients are close to zero and statistically insignificant.

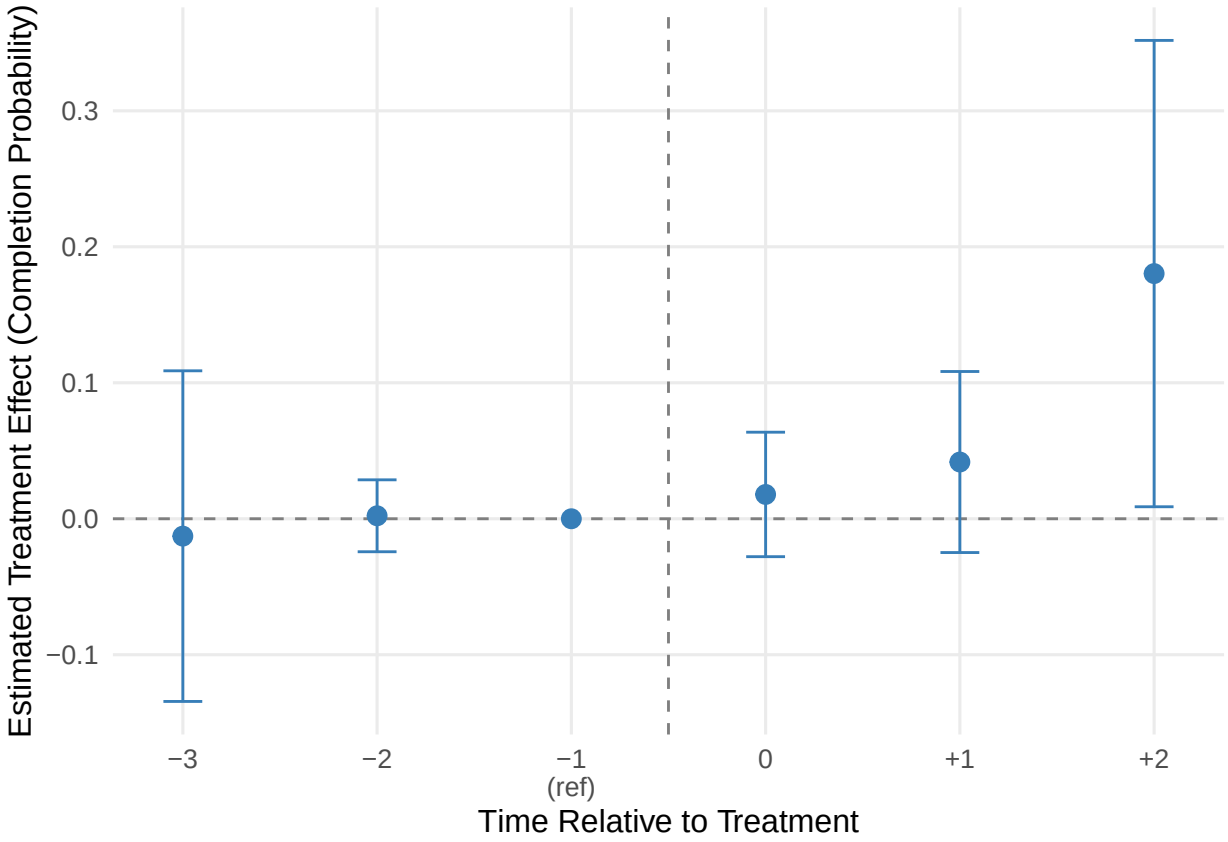

Figure 2: Event Study: Dynamic Treatment Effects of CSO Monitoring (6 periods, 2015-2023). Bars represent 95% confidence intervals. Reference period is t = -1.

Table 4: Dynamic Event-Study DiD: Effects of CSO Monitoring on Construction Completion

	(1)
t = -3 (Pre-treatment)	−0.013
	(0.062)
	[−0.134, 0.109]
t = -2 (Pre-treatment)	0.002
	(0.013)
	[−0.024, 0.029]
t = 0 (Treatment start)	0.018
	(0.023)
	[−0.028, 0.064]
t = +1 (Post-treatment)	0.042
	(0.034)
	[−0.025, 0.108]
t = +2 (Post-treatment)	0.180**
	(0.087)
	[0.009, 0.352]
Num.Obs.	20 964
R2	0.795
R2 Adj.	0.754
Std.Errors	by: municipio
FE: id	X
FE: periodo	X

* p <0.1, ** p <0.05, *** p <0.01
Significance: * p < 0.1, ** p < 0.05, *** p < 0.01

We assess the sensitivity of our estimates to political covariates that might confound the treatment–outcome relationship. Adding the mayor's party ideology (a left–right scale based on party affiliation) and the mayor's electoral margin of victory—individually or jointly—leaves the ATT virtually unchanged (0.082, p = 0.015 in the joint specification), confirming that our results are robust to observable political heterogeneity.

To verify that the results are not driven by any single treated municipality, we conduct a leave-one-out exercise, re-estimating the static ATT while dropping each of the 21 treated municipalities in turn. The estimates range from 0.074 to 0.100 and remain statistically significant at the 5% level in all 21 specifications, confirming that no individual municipality drives the main finding.

A limitation of our analysis is the small number of treated clusters (21 municipalities). With few clusters, standard cluster-robust standard errors may be downward-biased (A. C. Cameron, Gelbach, and Miller 2008). To assess the robustness of our inference, we implement a wild cluster bootstrap with Rademacher weights using the `fwildclusterboot` package (Fischer and Roodman 2021), which implements the fast algorithm of Roodman et al. (2019), as well as a randomization (permutation) inference test. The wild bootstrap p-value for the static ATT is 0.017, confirming statistical significance at the 5% level. The event study coefficient at t=+2 yields a wild bootstrap p-value of 0.0498, significant at the 5% level. A permutation test that randomly reassigns treatment across municipalities (4,999 permutations) produces a p-value of 0.073, significant at the 10% level. These results suggest that our findings are robust to the few-cluster concern, though the permutation-based evidence is less definitive.

Additionally, the four-year gap between our penultimate observation (August 2019) and the final period (October 2023) includes the COVID-19 pandemic, which may have differentially affected construction timelines across municipalities. While this concern is mitigated by the fact that the pandemic would likely have delayed construction across all municipalities similarly, we cannot rule out differential impacts. Readers should interpret the t=+2 estimate with this caveat in mind.

Furthermore, treated municipalities start with substantially lower completion rates than controls (17.6% vs. 44.6% in 2015), which raises concerns about regression to the mean. Since construction completion is an absorbing state—once a project is completed, it remains completed—there is a natural tendency for municipalities with lower initial completion rates to show larger increases over

time, even absent any intervention. This mechanical convergence could contribute to a positive DiD estimate. Our event-study specification partially addresses this concern by showing no differential pre-trends over three periods prior to treatment, and the static ATT (approximately 8 percentage points) is modest relative to the level difference. Nevertheless, we cannot fully rule out that some portion of the estimated effect, particularly the larger t=+2 coefficient, reflects catch-up dynamics rather than the intervention itself.

6 Mechanisms

The quantitative results show that CSO monitoring increased construction completion rates, but they do not reveal *how* this effect was produced. Our theoretical framework identified voice—repeated, technically informed engagement with local authorities—as the primary mechanism distinguishing CSO-led monitoring from information-only interventions. To provide suggestive evidence on this mechanism, we draw on illustrative case studies from project documentation, semi-structured interviews with CSO volunteers, and an independent evaluation report (Medeiros 2019).

The cases were selected to illustrate the distinct channels through which voice operated in practice: (a) persistent pressure that accelerated project delivery (Taubaté), (b) technical expertise that prevented wasteful procurement before construction began (Araucária), (c) frequent on-site visits that detected and corrected defects during execution (Taubaté and Foz do Iguaçu), and (d) post-completion auditing that revealed quality failures (Palhoça and Goioerê). Together, these cases suggest that the mechanism requires not just information, but sustained engagement, technical knowledge, and institutional support—resources that organized CSOs possess but individual citizens typically do not.

6.1 Delivery of the construction works

In our first case study, we document a notable achievement in ensuring the completion of construction work. In Taubaté, four construction projects were underway, which the local OSB began monitoring as part of Obra Transparente. By the project's conclusion, their efforts played a crucial role in facilitating the successful delivery of these schools and nurseries to the city, ensuring they

were ready to serve students.

To put the completion of all four projects in perspective: of 135 construction works planned across 21 municipalities, only 25 were delivered by the end of the project—an 18.5% delivery rate. In Taubaté, before the monitoring by the local CSO, the delivery rate was zero. After monitoring began, the delivery rate reached 100% and one of the construction works was a previously stalled construction. According to a report by Transparência Brasil (Coelho, Galdino, and Sakai 2021), out of 771 stalled construction projects that were eventually completed, over 50% of them took more than two years to be delivered and 25% took three years or more. As we can see, the odds were against finishing the construction works.

The local CSO monitoring construction in Taubaté performed among the most effectively of all 21 municipalities covered by the project. In fact, it was mostly the work of a single volunteer with support from the local OSB and Transparência Brasil throughout the project. She monitored four ongoing construction works on a periodic basis. She was very knowledgeable and insistent in demanding responses from contractors and the local government. To illustrate the quality of the monitoring, in an email sent on 17 January 2018 to the local government, the local OSB volunteer wrote: "Fazendinha Daycare: a. construction log verified. There have been a few more construction workers on the site since December 15, 2017, but not enough, especially considering the extensive amount of work to be completed by the new delivery date in April 2018. b. (…). c. Lack of roof sealing, leading to water entering the classrooms (a situation that already existed on November 14, 2017)…". She demonstrated exceptional knowledge and dedication throughout the project.

All of the construction works were effectively finished and delivered by the end of the project, although overdue. One of them, a resumption of a then halted construction work, was delivered two months overdue. The remaining three were further delayed, but no amendments were made on prices, only deadlines.

6.2 Preventing problems before projects started

The Araucária case illustrates what can happen when oversight arrives at the right moment to prevent a costly government procurement.

In 2017, the Brazilian municipality of Araucária conducted bidding processes seeking to hire companies to build three nurseries. The project sent technical experts to the three planned construction sites, together with volunteers from the local Observatório, to check whether plans were consistent with the physical conditions observed on site. The experts' assessment was that high-cost retaining walls included in the construction plans were unnecessary or could be replaced by low-cost solutions through adjustments. Originally, the retaining walls would cost R$ 1,577,338.57 (US$ 419,504.90, at the nominal exchange rate in 2019). The analysis was first directed to the local administration, suggesting changes to the construction plans, which were rejected. The bidding processes proceeded as initially planned. Subsequently, the experts' findings were submitted to the Brazilian Supreme Audit Institution, Tribunal de Contas da União (TCU), by Transparência Brasil and Observatório Social do Brasil (documentation submitted to TCU; details available from the authors upon request). TCU recommended that the bidding process should be redone, and construction plans were to be redesigned, with more cost-effective solutions for the retaining walls. In the new bidding process, their cost plummeted to R$ 416,883.17 (US$ 110,873.20), a reduction of US$ 308,631.70 — or 74% less. To give a sense of the impact on the city budget, the amount saved on these three constructions alone represents about 3% of all capital investments of Araucária in 2018. In short, in a single municipality, the project resulted in savings higher than the cost of the entire project funded by the grantor UNDEF (US$ 220,000.00).

6.3 Fixing defects in ongoing projects

Another impact of the project was the correction of irregularities and construction failures, improving the quality of the delivered construction work. During their monitoring activities, the OSBs found defects and issues, promptly reported them to the municipal administration, which then acted to fix them. The project partners effectively substituted for government inspectors, compensating for their shortcomings in ensuring contract compliance through supervision. This was the case in both the four constructions in Taubaté (SP) and three that were going on in Foz do Iguaçu (PR). In both cases volunteers performed monitoring activities very frequently, which allowed for a thorough investigation and revealed a series of defects in projects and execution.

6.4 Fixing defects in finished projects

Palhoça (SC) is an example of when oversight came too late. One of the daycares completed before the monitoring period started was delivered with a wall of perforated elements without proper grouting, with pieces that were already coming loose. The daycare was put into operation without addressing the issue, which led the staff to isolate the area to minimize the risk of injuries to the children. In other words, part of the building cannot be used due to a defect that should have been addressed by the responsible contractor.

In Goioerê, the story had a happier ending. Volunteers visited an already delivered school building and noticed improperly installed windows, rendering them non-functional. A series of finishes were also lacking, such as grab bars and handles in accessible restrooms for persons with disabilities or reduced mobility. These shortcomings compromised adequate use of the building and exposed its users—mainly children—to potentially dangerous situations. These flaws were mostly corrected by the contractor after being reported by the local Observatório.

We note two important caveats regarding the mechanism evidence. First, the cases presented above were selected to illustrate successful monitoring outcomes; cases where CSO monitoring was less effective are not represented, which limits the representativeness of these examples. Second, the intervention bundled multiple components—training, technical support from Transparência Brasil, information provision, and engagement with horizontal accountability institutions such as TCU—making it difficult to isolate voice as the sole operative mechanism. The Araucária case, for instance, involved escalation to TCU, which represents diagonal accountability rather than pure voice. We therefore interpret the qualitative evidence as consistent with, but not exclusively demonstrating, a voice-based mechanism.

7 Discussion of Results

One of the reasons identified for problems with construction works is a lack of qualified government personnel from the government to monitor the contractor work. As a result, the local government does not properly oversee the work of contractors. This is a task that requires technical expertise and is time-consuming, involving going multiple times to the construction site, comparing the construc-

tion log with the schedule to assess if there is potential delay and what is specified in the contract, to assess if the materials used are correct, among other tasks.

In the broader literature on bottom-up accountability, Freire, Galdino, and Mignozzetti (2020) evaluated Tá de Pé, a related intervention that provided information to citizens about construction work in Brazil. The contrast between information-only interventions and the sustained, organized monitoring by CSOs examined in this paper highlights the importance of distinguishing between different types of bottom-up approaches. The evidence we gathered here, assessed with multiple methods, demonstrates that sustained and rigorous oversight by civil society can produce substantial results, whether through increased project completion rates, improved construction quality, or reduced government spending.

Our findings also relate to Olken (2007), who found that top-down government audits were more effective than grassroots participation in reducing corruption in Indonesian road construction. While Olken's results might seem to contradict ours, the two studies examine different types of bottom-up intervention: Olken evaluated open village meetings (essentially providing information to disorganized citizens), whereas our study examines sustained monitoring by organized, technically trained CSOs. This distinction aligns with our theoretical framework: it is not bottom-up participation per se that matters, but the organizational capacity and persistence of the monitoring agents.

The qualitative evidence gathered is important to elucidate the mechanism. It is important to contact local authorities repeatedly, making it harder for them to disregard citizens' concerns. Another key aspect is to have a deep knowledge of the policy in question. In our case, we are talking about understanding a construction timetable, knowing what the law and the contract mandate, being able to assess whether the number of workers on a construction site is sufficient to meet the deadline, and monitoring progress repeatedly over time.

Regarding external validity, our findings are most directly applicable to contexts where three conditions hold: (i) an organized and technically capable civil society exists at the local level, (ii) the policy outcome is observable and verifiable through on-site monitoring (as with construction works), and (iii) government responsiveness to civil society pressure is feasible—i.e., authorities are not fully insulated from social demands. These conditions are met in much of urban Brazil and in similar middle-income democracies with active civic sectors, but may not generalize to settings

where CSOs are absent, repressed, or where outcomes are harder to monitor (e.g., service quality rather than infrastructure delivery).

The detection of evidence of procurement fraud cases also shows local administrations are not diligent enough in conducting procurement procedures, and the local partners of TB contributed to addressing that gap as well. Moreover, as federal audit agencies lack the capacity to adequately oversee and analyze the immense volume of contracting procedures conducted at the local level, the social monitoring resulted in the identification of high-risk cases that otherwise might not have been assessed by the competent bodies.

8 Concluding Remarks

Accountability, as a fundamental element in democratic systems, plays a pivotal role in ensuring the efficient provision of public services. Recent studies investigating tactical interventions aimed at providing information to citizens have shown their limited impact. However, it would be a mistake to dismiss entirely the potential of citizen voices in governance. Our research has differentiated between information-led tactical interventions and active mobilization and monitoring by Civil Society Organizations (CSOs). We argue that organized CSOs, deeply rooted in society and politics, hold the capacity to bring about substantial policy change.

This paper provides evidence in support of this argument. By assessing the effectiveness of the "Obra Transparente" project developed by Transparência Brasil in Brazil, we illustrate the significant role that organized CSOs can play in shaping government behavior and improving public service delivery. Unlike interventions solely focused on providing information, this project aimed to create a network of local CSOs, offer technical support, and provide training on monitoring public school and nursery construction.

Our findings underscore the importance of sustained, concerted efforts by organized civil society, including direct engagement with municipal officials, persistent follow-ups, and on-site visits. These endeavors yield more substantial outcomes compared to interventions that merely deliver information and generate sporadic citizen actions. Our research contributes to a more comprehensive understanding of how bottom-up accountability can be achieved through the mobilization of

organized civil society.

9 Data Availability

The data and code required to replicate the quantitative results in this paper are available at https://github.com/mgaldino/did-obra-transparente. The replication package includes: (i) administrative data on construction project status from SIMEC (Sistema Integrado de Monitoramento, Execução e Controle), maintained by FNDE, collected at six time points between 2015 and 2023; (ii) municipal socioeconomic characteristics from IPEA and IBGE; (iii) mayoral election results from TSE; and (iv) all R scripts needed to reproduce the analysis. The complete pipeline can be executed with a single command (`Rscript code/99_run_all.R`). See `REPLICATION.md` in the repository for detailed instructions. Qualitative evidence (project reports, interview transcripts, and evaluation documents) is available from the corresponding author upon reasonable request.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Ankamah, Samuel Siebie. 2019. "Why Do 'Teeth' Need 'Voice'? The Case of Anti-Corruption Agencies in Three Australian States." *Australian Journal of Public Administration* 78 (4): 481–96.

Bertelli, Anthony M., and Gregg G. Van Ryzin. 2020. "Heuristics and Political Accountability in Complex Governance: An Experimental Test." *Research & Politics* 7 (3): 2053168020950080. https://doi.org/10.1177/2053168020950080.

Besley, Timothy, and Maitreesh Ghatak. 2003. "Incentives, Choice, and Accountability in the Provision of Public Services." *Oxford Review of Economic Policy* 19 (2): 235–49.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–69.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple

Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90 (3): 414–27.

Cameron, Wayne. 2004. "Public Accountability: Effectiveness, Equity, Ethics." *Australian Journal of Public Administration* 63 (4): 59–67.

Coelho, Jonas, Manoel Galdino, and Juliana Sakai. 2021. "Levantamento: Obras federais paralisadas de creches e escolas com maiores chances de serem concluídas."

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–96.

Fischer, Alexander, and David Roodman. 2021. *Fwildclusterboot: Fast Wild Cluster Bootstrap Inference for Linear Models*. https://github.com/s3alfisc/fwildclusterboot.

Fox, Jonathan A. 2015. "Social Accountability: What Does the Evidence Really Say?" *World Development* 72: 346–61.

Freire, Danilo, Manoel Galdino, and Umberto Mignozzetti. 2020. "Bottom-up Accountability and Public Service Provision: Evidence from a Field Experiment in Brazil." *Research & Politics* 7 (2): 1–8. https://doi.org/10.1177/2053168020914444.

Gong, Ting, and Hanyu Xiao. 2017. "Socially Embedded Anti-Corruption Governance: Evidence from Hong Kong." *Public Administration and Development* 37 (3): 176–90.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77. https://doi.org/10.1016/j.jeconom.2021.03.014.

Hirschman, Albert O. 1972. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard university press.

Isunza Vera, Ernesto, and Adrian Gurza Lavalle. 2012. "Arquitetura Da Participação e Controles Democráticos No Brasil e No México." *Novos Estudos CEBRAP*, 105–21.

Joshi, Anuradha, and Peter P. Houtzager. 2012. "Widgets or Watchdogs? Conceptual Explorations in Social Accountability." *Public Management Review* 14 (2): 145–62.

Kenney, Charles D. 2003. "Horizontal Accountability: Concepts and Conflicts." In *Democratic Accountability in Latin America*, 55–76. Scott Mainwaring & Christopher Welna.

Lavalle, Adrian Gurza, Jessica Voigt, and Lizandra Serafim. 2016. "O Que Fazem Os Conselhos e

Quando o Fazem? Padrões Decisórios e o Debate Dos Efeitos Das Instituições Participativas." *Dados* 59: 609–50.

Lavalle, Adrián Gurza, Peter P. Houtzager, and Graziela Castello. 2006. "Representação Política e Organizações Civis: Novas Instâncias de Mediação e Os Desafios Da Legitimidade." *Revista Brasileira de Ciências Sociais* 21 (60): 43–66. https://doi.org/10.1590/s0102-69092006000100003.

Little, Andrew T, and Anne Meng. 2024. "Measuring Democratic Backsliding." *PS: Political Science & Politics* 57 (2): 149–61.

Lührmann, Anna, Kyle L. Marquardt, and Valeriya Mechkova. 2020. "Constraining Governments: New Indices of Vertical, Horizontal, and Diagonal Accountability." *American Political Science Review* 114 (3): 811–20. https://doi.org/10.1017/S0003055420000222.

Mattoni, Alice, and Fernanda Odilla. 2021. "Digital Media, Activism, and Social Movements' Outcomes in the Policy Arena. The Case of Two Anti-Corruption Mobilizations in Brazil." *Partecipazione e Conflitto* 14 (3): 1127–50.

Medeiros, Vanda. 2019. "Post Project Evaluation for the United Nations Democracy Fund." Independent {{Report}} Submitted to {{UNDEF}}.

Mulgan, Richard. 2000. "'Accountability': An Ever-Expanding Concept?" *Public Administration* 78 (3): 555–73.

O'donnell, Guillermo. 1998. "Horizontal Accountability in New Democracies." *J. Democracy* 9: 112.

———. 2003. "Horizontal Accountability: The Legal Institutionalization of Mistrust." In *Democratic Accountability in Latin America*, 34:34–54. Scott Mainwaring & Christopher Welna.

O'Loughlin, Michael G. 1990. "What Is Bureaucratic Accountability and How Can We Measure It?" *Administration & Society* 22 (3): 275–302.

Odilla, Fernanda. 2023. "Bots Against Corruption: Exploring the Benefits and Limitations of AI-based Anti-Corruption Technology." *Crime, Law and Social Change*, 1–44.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.

Paul, Samuel. 1992. "Accountability in Public Services: Exit, Voice and Control." *World Development* 20 (7): 1047–60.

Pogrebinschi, Thamy, and Fabiano Santos. 2011. "Participação Como Representação: O Impacto Das Conferências Nacionais de Políticas públicas No Congresso Nacional." *Dados* 54 (3): 259–305. https://doi.org/10.1590/s0011-52582011000300002.

Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2019. "The Weakness of Bottom-up Accountability: Experimental Evidence from the Ugandan Health Sector." *Los Angeles: Innovations for Poverty Action*.

Relly, Jeannine E. 2012. "Examining a Model of Vertical Accountability: A Cross-National Study of the Influence of Information Access on the Control of Corruption." *Government Information Quarterly* 29 (3): 335–45.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using Boottest." *The Stata Journal* 19 (1): 4–60. https://doi.org/10.1177/1536867X19830877.

Roth, Jonathan, and Pedro HC Sant'Anna. 2023. "When Is Parallel Trends Sensitive to Functional Form?" *Econometrica : Journal of the Econometric Society* 91 (2): 737–47.

Smidt, Hannah, Jessica Johansson, and Thomas Richter. 2025. "Civil Society Under Attack: The Consequences for Horizontal Accountability Institutions." *Studies in Comparative International Development* 60 (1): 81–110. https://doi.org/10.1007/s12116-023-09423-x.

Waldner, David, and Ellen Lust. 2018. "Unwelcome Change: Coming to Terms with Democratic Backsliding." *Annual Review of Political Science* 21 (1): 93–113.