

YOU GET WHAT YOU PAY FOR: WHEN DO CERTIFICATION PROGRAMS IMPROVE PUBLIC SERVICE DELIVERY? *

SOEREN J. HENN [†]
HORACIO LARREGUY [‡]
JOHN MARSHALL [§]

OCTOBER 2020

Poor local public service delivery is common across the Global South. We argue that the short-term unobservability of investments to improve service delivery combine with adverse selection to weaken incentives for governments to make such investments. While programs to certify investments can mitigate this monitoring problem, the certification process's effectiveness can be undermined by opportunistic politicians and certifiers. We test this argument using a Mexican program designed to certify service delivery investments, where certifications are self-assessed by municipal governments and validated by corruptible third-party institutions. Difference-in-differences estimates show that the program did not ultimately improve municipal public service delivery on average. Consistent with our model, this effect is only positive when the third party is unlikely to be corruptible and when the likelihood that the incumbent is not corruptible in producing the service is large. These findings highlight the challenges in improving service delivery and the importance of incentive-compatible monitoring.

JEL: D72, O12 .

Key words: limited public service delivery, unobservability of investments, certification.

*We thank Jonathan Rodden and Jaakko Pekka Merilainen for helpful comments, discussion, and suggestions. The paper also benefited from presentations at Stanford University and 2019 MPSA. We thank Oscar Elton and Francheska Loza for excellent research assistance. Horacio Larreguy acknowledges funding from the IQSS Undergraduate Research Scholars Program at Harvard University and the French Agence Nationale de la Recherche under the Investissement d'Avenir program ANR-17-EURE-0010. All errors are our own.

[†]Innovations for Poverty Action & Pearson Institute, University of Chicago. sjhenn@uchicago.edu.

[‡]Department of Government, Harvard University, and Visiting Researcher, Institute for Advanced Study in Toulouse and Toulouse School of Economics. hlarreguy@fas.harvard.edu and hlarreguy@tse-fr.eu.

[§]Department of Political Science, Columbia University. jm4401@columbia.edu.

1 Introduction

An accountable political system can enhance public service delivery by creating incentives for politicians to make investments that increase the efficiency of service delivery and not to misappropriate funds (Barro 1973; Fearon 1999; Ferejohn 1986). However, there is abundant evidence of limited or low quality local public service delivery across the Global South (Bardhan 2002; Cruz, Labonne and Querubin 2019; Habyarimana et al. 2007; Keefer 2007; Khemani et al. 2016; Pande 2011). An active recent literature has examined how voter information about incumbent performance and candidate policy platforms can help select competent and policy-aligned politicians.¹ However, little is yet known about if and when incentives then exist for elected local governments to invest in improving service delivery when voters are *partially* informed about the incumbent’s actions in office.

We develop a stylized theory to highlight how the short-term unobservability of investments by local governments to improve local service delivery can constrain such investments and thereby limit service delivery. We consider a two-period adverse selection model where a local government produces two kinds of services. Incumbents receive a unit of budget to produce each service, which they can allocate to current public service delivery, more efficient future service delivery, or misappropriation. Politicians are of two types—clean or corruptible—in producing each service, but voters are only aware of the service-specific distribution of these types in the population. All politicians care about rents from holding office, but only clean politicians regarding a certain service share voters’ preferences over, or lack the ability to extract rents from, that particular service. Voters decide whether to re-elect the incumbent between periods. However, they can only observe

¹Ferraz and Finan (2008), Enríquez et al. (2020), and Larreguy, Marshall and Snyder (forthcoming) show how the dissemination of information about incumbent malfeasance through broadcast and social media can promote electoral accountability. However, while recent studies show that information about campaign promises lead citizens to make more informed voting decisions (Bidwell, Casey and Glennerster 2020; Cruz et al. 2019; Platas and Raffler 2020), experimental evidence about the effect of localized incumbent information dissemination is mixed (Banerjee et al. 2011; Chong et al. 2015; Dunning et al. 2019).

if the budget was allocated to current public service delivery, and therefore cannot distinguish whether an incumbent that did not deliver a given service immediately misappropriated the budget or invested in future delivery of that service.

When the probability that the incumbent is corruptible in providing a given service is sufficiently high, we show that investing in improving delivery of that service is not feasible. Intuitively, this is because voters infer that the incumbent is likely to be corruptible in providing the service when they do not see the budget being spent on current service delivery in high-corruption settings, and thus vote the incumbent out of office. The model thus highlights how the short-term unobservability of investments to improve local service delivery constrains such investments and ultimately limits service delivery in environments defined by a significant risk of adverse selection.

In principle, certification programs designed to publicly demonstrate that an incumbent invested, rather than misappropriated, funds could improve service delivery by ameliorating voters' inability to monitor budgetary allocations. We model the certifier as a third party endowed with the unilateral right to certify that investments to improve future service delivery were undertaken. Third parties, however, are known to be corruptible with a certain probability. Consequently, while the program cannot reduce service delivery relative to the absence of the certification program in expectation, improvements in the delivery of a given service decreases in the probabilities that the third party and incumbent are corruptible. In environments where these probabilities of corruption are sufficiently high, certification programs may be ineffective in improving the quality and quantity of a given public service.

We assess this theory's relevance by testing its empirical predictions in the context of the From the Local Agenda (Agenda desde lo Local, ADL) program in Mexico. In this context, municipal governments often provide poor public service delivery and engage in corruption (Chong et al. 2015; Díaz-Cayeros, Estévez and Magaloni 2016; Diaz-Cayeros, González and Rojas 2006). The ADL program was first implemented by the federal government in 2004 in collaboration with state governments, which needed to enter the program before municipalities could participate in

it. The ADL program aims to promote investments in public service delivery and consists of four main stages: (i) self-assessment, by municipal government officials, across 39 indexes of the municipal government's public service delivery on a scale from red to yellow to green status; (ii) third-party verification of this diagnosis by a local institution of higher education, which results in the municipality receiving a certificate for achieving green status on any given index; (iii) time for municipal government officials to invest in improvements for non-green indexes; and (iv) updated self-assessment and third-party verification, again resulting in the granting of certificates for each new index that receives green status in the municipality. The ADL program thus allows incumbents to certify that investments to improve service delivery were made, whether due to actual investments or collusion with the third party.

We estimate the effects of entering the ADL program on certification and actual service delivery outcomes using a generalized difference-in-differences design that leverages temporal variation in when a municipality entered the program. To avoid comparing municipalities that entered the ADL program with those that never did it, we focus only on the sample of municipalities that entered the program between 2004 and 2013. Over this period, we examine certification status across the program's 39 indexes before assessing public service delivery for the 10 program indexes for which rich administrative data enables us to independently approximate the program's evaluation criteria. Empirical tests support the design's identifying parallel trends assumption. To examine the heterogeneity implied by our model, we proxy for the likelihood that a certifying third party is corruptible using municipal political alignment with the state governor, based on the fact that the state-dependent institutions of higher education that serve as third-party auditors have been involved in various cases of corruption. We further consider the baseline self-assessment of a given index upon entry into the program as a proxy for a high likelihood that the incumbent is corruptible in producing the specific service captured by the index.

The results indicate that, as expected, program participants were frequently awarded green statuses for certified improvements. Specifically, we show that the certification status of indexes of

local service delivery quickly increased after entering the ADL program. More importantly, the findings largely support the model's more cynical empirical predictions. Consistent with a high fraction of corruptible certifying third parties and municipal incumbents, we find no discernible average effect of the certification program on public service delivery outcomes that are measured independently from the program. Furthermore, and in line with the heterogeneity predicted by the model, our results indicate that the certification program led to an increase in public service delivery in municipalities that were not aligned with the state governor and in indexes where baseline certification statuses were not low. In contrast, the certification program had a negligible or even slightly negative effect in municipalities aligned with state governments and on indexes with low baseline certification levels. These heterogeneous effects suggest that a significant number of corrupt incumbents used the ADL program as a shield to facilitate greater rent seeking.

Our argument advances the debates surrounding informational constraints on electoral accountability and public service delivery. While a large body of work studies whether voter information about incumbent performance (Banerjee et al. 2011; Bhandari, Larreguy and Marshall forthcoming; Chong et al. 2015; Ferraz and Finan 2008; Larreguy, Marshall and Snyder forthcoming) and candidate policy platforms (Bidwell, Casey and Glennerster 2020; Bowles and Larreguy 2020; Cruz et al. 2019; Platas and Raffler 2020) contributes to electoral accountability, these studies generally assume that providing information about a few dimensions of government performance may be sufficient to improve governance outcomes by strengthening incentives for politicians or enhancing voter selection of politicians. Our findings suggest that easily-measured indexes of performance may not be sufficient where adverse selection challenges remain. Rather, we emphasize the importance of the *type of information asymmetry*, highlighting how the short-term unobservability of investments to improve service delivery can reduce the ability of clean politicians to serve their constituents by investing in service delivery. Furthermore, like Banerjee, Duflo and Glennerster (2008) and Raffler (2019), we also highlight the difficulty of designing incentive structures to prevent manipulation or corruption of monitoring devices—in our case, of a political form.

This article also contributes to the multitasking literature in developing contexts by arguing that the varying visibility of different public service investments can generate distortions in the incumbent's provision of such services and increase corruption. Several studies emphasize that, particularly in contexts of weak state capacity, incumbents have electoral incentives to deliver services that are clearly attributable to their actions (Harding and Stasavage 2013; Harding 2015; Mani and Mukand 2007; Marx 2018). Tavits (2007) similarly shows that when the visibility of public goods is sufficiently low, incumbent politicians face incentives to engage in corruption. In turn, Dal Bo and Rossi (2011) emphasize that, since politician terms are not sufficiently long for some investments in service delivery to mature before they go up for reelection, they are discouraged from making such investments. We instead show that, when both investments in future public service delivery are unobservable and the share of corruptible politicians in producing certain services is large enough, even clean politicians will not invest in improving the delivery of such services.

Finally, this article adds empirical rigor and theoretical clarity to extant research doubting the specific efficacy of the ADL program in facilitating municipal development. In particular, Pérez Archudia and Arenas Aréchiga (2012) report the lack of a significant positive correlation between actual municipal development outcomes and ADL program assessments of the purported drivers of such outcomes. They conclude by questioning how well the ADL program assessments capture municipal public service delivery. We similarly demonstrate a weak association between entering the ADL program and overall changes across an array of fine-grained public service delivery outcomes using a more compelling identification strategy. Illustrating how the theoretical channels underlying the program's ineffectiveness may generalize within many parts of the Global South, we provide evidence suggesting that the ADL program's relative ineffectiveness is driven by a high probability that third parties and municipal incumbents are corruptible.

2 A theory of investments in service delivery

We develop a theory that highlights how the short-term unobservability of investments in enhancing future service delivery can impede such investments and thereby limit or lower the quality of municipal public services. We first lay out a stylized dynamic model where a local government produces two services. We then characterize the equilibrium outcome in the absence of a potentially corruptible program that certifies investments in future service delivery, before characterizing the equilibrium outcome when such a program is implemented. Finally, we compare the equilibria to generate testable implications of the model for certification and service delivery outcomes that arise when a certification program is introduced.

2.1 The model

We consider a two-period, two-service model for a representative municipal government. In each period $t \in \{1, 2\}$, the incumbent politician is granted a unit of budget to produce the separable public services A and B . For each service $i \in \{A, B\}$, the incumbent politician can allocate the budget toward: (i) current service delivery, denoted by the decision $g_{i,t} \in \{0, 1\}$; (ii) investing in future service delivery, denoted by the decision $G_{i,t} \in \{0, 1\}$; or (iii) public-rent misappropriation, denoted by the decision $r_{i,t} \in \{0, 1\}$. Although such allocations are less stark in practice, our simplified setting highlights the core theoretical logic.

A representative voter decides whether to re-elect the incumbent politician or select an alternative candidate at the end of the first period.² The decision to re-elect the incumbent is denoted by $v \in \{0, 1\}$. In the baseline model, the voter can only observe whether the incumbent spent the budget for each service i on *current* service delivery. Consequently, if the voter observes no current delivery of service i ($g_{i,1} = 0$), then they cannot tell whether the incumbent misappropriated the

²By focusing on a representative voter, we abstract from voter coordination issues. A qualitatively similar equilibrium emerges with homogeneous voters playing weakly undominated strategies.

budget for service i ($r_{i,1} = 1$) or invested in future service delivery ($G_{i,1} = 1$).

The voter derives additive utility from the total level of service delivery—the sum of immediate and future service delivery—of each service that is realized at the end of the second period:

$$\sum_{i=A,B} \sum_{t=1,2} U(g_{i,t}, G_{i,t}) = \sum_{i=A,B} \sum_{t=1,2} (g_{i,t} + \beta G_{i,t}), \quad (1)$$

where $\beta > 1$ captures our assumption that investments in future public service delivery yield higher utility for the voter than current service delivery.³ In other words, investing in future public service delivery allows the government to provide higher-quality or more public services than it can provide immediately. For example, citizens may benefit more from having piped water in the future as opposed to water trucks now or the piped water network may be extended to encompass more citizens.

All politicians seek to remain in office, and receive per-period rents $R > 0$ when in office. However, corruptible politicians can also extract additional rents by misappropriating funds. With probability $\gamma_i \in [0, 1]$, the incumbent is always clean regarding their allocation of service i ($\tau_i = h$); with probability $1 - \gamma_i$, the incumbent possesses the ability to engage in corruption ($\tau_i = c$) when allocating the budget for service i . Clean politicians can be thought of as lacking the ability to misappropriate the budget for service i .⁴ An incumbent politician of type (τ_A, τ_B) in period $t = 1$ then chooses $(g_{i,1}^{\tau_i}, G_{i,1}^{\tau_i}, r_{i,1}^{\tau_i})$ for each service to maximize $\sum_{i=A,B} \mathbb{1}[\tau_i = c] r_{i,1}^{\tau_i} + \mathbb{1}[v = 1] \left(R + \sum_{i=A,B} \mathbb{1}[\tau_i = c] r_{i,2}^{\tau_i} \right)$, while an incumbent politician in period $t = 2$ chooses $(g_{i,2}^{\tau_i}, G_{i,2}^{\tau_i}, r_{i,2}^{\tau_i})$ for each service to maximize $\sum_{i=A,B} \mathbb{1}[\tau_i = c] r_{i,2}^{\tau_i}$. The distribution of types is independent across services and common knowledge, but the realization of a given politician's type (τ_A, τ_B) is known only to the politician. Without loss of generality, we assume that $\gamma_A > \gamma_B$. For

³The utility function in equation (1) can also be interpreted as the benefit of $G_{i,t}$ being deferred by one period. In this case, β captures the discounted value of future services, which still exceed one due to the higher quality or quantity of such services.

⁴Under slightly more restrictive conditions, qualitatively similar results obtain if clean politicians share the voter's preferences over service i .

example, this could reflect the relative ease of engaging in procurement fraud in the construction sector, relative to the health care sector.

The timing of the game is as follows:

1. Nature independently draws the incumbent politician's type, $\tau_i \in \{h, c\}$, for each service $i \in \{A, B\}$. This is private information revealed only to the incumbent.
2. At the beginning of period $t = 1$, the incumbent politician of type (τ_A, τ_B) selects their policy $(g_{i,1}^{\tau_i}, G_{i,1}^{\tau_i}, r_{i,1}^{\tau_i})$ for each service.
3. The voter observes $(g_{A,1}, g_{B,1})$ and decides whether to re-elect their incumbent politician, $v \in \{0, 1\}$.
4. If an incumbent politician is not re-elected, nature draws a new politician type for the victorious challenger.
5. At the beginning of period $t = 2$, the (possibly new) incumbent politician selects $(g_{i,2}^{\tau_i}, G_{i,2}^{\tau_i}, r_{i,2}^{\tau_i})$ for each service i .
6. All utilities are realized and the game ends.

2.2 Equilibrium without a certification program

Throughout our analysis, we restrict attention to the sequentially rational equilibrium most preferred by the voter (i.e. that yields the highest utility for the voter).⁵ We start by characterizing this equilibrium in the absence of a certification program ($p = 0$):

Proposition 1. *Let $\gamma_A > \gamma_B$ and denote $\gamma^* := \frac{1}{\beta}$. Then for each service i :*

⁵Two classes of equilibria always exist: (i) where the voter only re-elects the incumbent when they observe $g_{i,1} = 1$ and thus $G_{i,1} = 1$ is not possible on the equilibrium path; and (ii) where the voter only re-elects the incumbent when they observe $g_{i,1} = 0$, but only clean types choose $G_{i,1} = 1$. By restricting attention to the payoff-dominant equilibrium from the voter's perspective, we effectively assume that the voter is able to select among multiple equilibria.

- If $\gamma_i \geq \gamma^*$, only clean types in producing service i invest in its future delivery. Clean incumbents regarding service i choose the policy vectors $(g_{i,1}^h, G_{i,1}^h, r_{i,1}^h) = (g_{i,2}^h, G_{i,2}^h, r_{i,2}^h) = (0, 1, 0)$, while corruptible incumbents regarding service i choose the policy vectors $(g_{i,1}^c, G_{i,1}^c, r_{i,1}^c) = (g_{i,2}^c, G_{i,2}^c, r_{i,2}^c) = (0, 0, 1)$.
- If $\gamma_i < \gamma^*$, clean and corruptible types initially pool to deliver basic current services. Both incumbent types choose the policy vector $(g_{i,1}, G_{i,1}, r_{i,1}) = (1, 0, 0)$ in period $t = 1$, whereas clean and corruptible incumbents regarding service i respectively choose the policy vectors $(g_{i,2}^h, G_{i,2}^h, r_{i,2}^h) = (0, 1, 0)$ and $(g_{i,2}^c, G_{i,2}^c, r_{i,2}^c) = (0, 0, 1)$ in period $t = 2$.

If $\gamma_B \geq \gamma^*$, the voter re-elects the incumbent politician upon observing $g_{A,1} = g_{B,1} = 0$. If $\gamma^* \in [\gamma_B, \gamma_A)$, the voter re-elects the incumbent politician upon observing $g_{A,1} = 0$ and $g_{B,1} = 1$. If $\gamma_A < \gamma^*$, the voter re-elects the incumbent politician upon observing $g_{A,1} = g_{B,1} = 1$. The voter's expected utility derived from public provision of service i is then given by:

$$\mathbb{E} \left[\sum_{t=1,2} U(g_{i,t}, G_{i,t}) \right] = \begin{cases} 1 + \gamma_i \beta & \text{if } \gamma_i < \gamma^* \\ 2\gamma_i \beta & \text{if } \gamma_i \geq \gamma^*. \end{cases}$$

Proof. See Appendix section A.1 for all proofs, where we also define the off-equilibrium strategies and beliefs that support this sequentially rational equilibrium. \square

Proposition 1 shows that the voter may be willing to tolerate the risk that the incumbent misappropriates resources for service i in the first period to generate higher-return investments in future delivery of service i . Intuitively, this occurs when the likelihood that the incumbent is clean in producing service i is sufficiently high ($\gamma_i \geq \gamma^*$), and thus that the risk of the incumbent being corrupted is low. Otherwise, the voter only re-elects an incumbent that allocates the budget for service i into immediately observable, but less effective, current public service delivery in period one.

For the remaining analysis, we restrict attention to the part of the parameter space where only service B is inefficiently provided. This constitutes the interesting, and often germane, case where the likelihood that an incumbent could engage in corruption is sufficiently high. This entails imposing the following assumption: $\gamma^* \in [\gamma_B, \gamma_A)$. The case where $\gamma_B \geq \gamma^*$ is uninteresting because investments in future service delivery are already made for both services with positive probability, while the certification program that we next examine cannot shift the equilibrium where visible but inefficient immediate public services are always provided when $\gamma_A < \gamma^*$. We thus consider the parameter space where a certification program has greatest potential to be effective.

2.3 Equilibrium with a certification program

We now extend the model to incorporate a certification program ($p = 1$) aiming to address the under-investment that occurs due to the unobservability of investments in future public service delivery. Under the certification program, we assume that a third party publicly certifies, $c_i \in \{0, 1\}$, whether $G_{i,1} = 1$ occurred for each service i . When $g_{i,1} = 1$, certification is not required because the services provided are publicly observable. The third party is clean ($\alpha = H$) with probability $\rho \in (0, 1)$ and corruptible ($\alpha = C$) with probability $1 - \rho$. The third party's type α is known to the third party and the municipal incumbent, but is not known to the voter. Honest third parties always report truthfully, i.e. $c_i = G_{i,1}$, but corruptible ones report $c_i = 1$ regardless of investments in future service delivery.⁶ In keeping with the low probability that third party reports lead to meaningful legal sanctions, we assume that the incumbent only incurs electoral costs of corruption being revealed.

The voter's payoff-dominant sequentially rational equilibrium again depends on the probability γ_i that the municipal incumbent is clean in producing each type of service, but now also depends on the probability ρ that the third-party certifier is honest. Proposition 2 characterizes this equilibrium:

⁶For simplicity, we do not model the third party as a strategic actor. However, similar results would obtain if bargaining over $r_{i,1}$ was required for the third party to agree to hide a corruptible incumbent's corruption.

Proposition 2. Assume that $\gamma^* \in [\gamma_B, \gamma_A)$, and denote $\gamma^{**}(\rho) := \frac{1-\rho\beta}{\beta(1-\rho)} < \gamma^*$. Then:

- If $\gamma_B \geq \gamma^{**}$, clean and corruptible types initially pool to invest in future delivery of both services when the third party is honest. For each service i , incumbents that are clean in producing that service choose $(g_{i,1}^h, G_{i,1}^h, r_{i,1}^h) = (g_{i,2}^h, G_{i,2}^h, r_{i,2}^h) = (0, 1, 0)$ and incumbents that are corruptible in producing that service choose $(g_{i,1}^c, G_{i,1}^c, r_{i,1}^c) = (0, 1, 0)$ in period $t = 1$ if the third-party is honest and $(g_{i,1}^c, G_{i,1}^c, r_{i,1}^c) = (0, 0, 1)$ if the third party is corruptible and chose $(g_{i,2}^c, G_{i,2}^c, r_{i,2}^c) = (0, 0, 1)$. Each type of third party always reports $c_A = c_B = 1$. Only upon observing $c_A = c_B = 1$ does the voter re-elect the incumbent politician.
- If $\gamma_B < \gamma^{**}$, clean and corruptible types initially pool to invest in future delivery of only service A when the third party is honest. The strategies regarding service A are identical to the case where $\gamma_B \geq \gamma^{**}$. However, both types of incumbent choose $(g_{B,1}, G_{B,1}, r_{B,1}) = (1, 0, 0)$ in period $t = 1$, while incumbents that are clean in producing service B choose $(g_{B,2}^h, G_{B,2}^h, r_{B,2}^h) = (0, 1, 0)$ and incumbents that are corruptible in producing service B choose $(g_{B,2}^c, G_{B,2}^c, r_{B,2}^c) = (0, 0, 1)$ in period $t = 2$. Each type of third party always report $c_A = 1$. Upon observing $c_A = 1$ and $g_{B,1} = 1$ does the voter re-elect the incumbent politicians in between periods.

The voter's expected utilities derived from public provision of each service i are then given by:

$$\mathbb{E} \left[\sum_{t=1,2} U(g_{A,t}, G_{A,t}) \right] = 2\gamma_A\beta + (1 - \gamma_A)\rho\beta,$$

$$\mathbb{E} \left[\sum_{t=1,2} U(g_{B,t}, G_{B,t}) \right] = \begin{cases} 1 + \gamma_B\beta & \text{if } \gamma_B < \gamma^{**} \\ 2\gamma_B\beta + (1 - \gamma_B)\rho\beta & \text{if } \gamma_B \geq \gamma^{**}. \end{cases}$$

Proposition 2 shows that, when the probabilities that the third party is honest *and* the incumbent is clean with respect to service i are sufficiently high, the voter benefits from an increase in the provision of service i as a result of the certification program. This reflects two effects of certification.

First, in the case of service A , certification increases investment in future service delivery by forcing incumbents that are corruptible in producing service A to invest in the first period. Although incumbents are not induced to shift away from producing basic services when $\gamma_i \geq \gamma^*$, this effect arises simply because third parties are effective monitors when they are honest. Second, and more interestingly, the certification program makes an equilibrium where the incumbent invests in future service delivery of service B more attractive to the voter. Intuitively, when $\gamma_B \geq \gamma^{**}$, the certification program induces a shift to an equilibrium where incumbents that are clean in producing service B can also invest in future delivery of the service in the first period, regardless of whether the third party is honest, because there is a sufficiently large probability that an honest third party will prevent corruption. In effect, third-party certification (partially) compensates for the lower proportion of incumbents that are clean in producing service B . When $\gamma_B < \gamma^{**}$, the presence of a corruptible certifying third party cannot overcome the lack of clean types in the production of service B , and both types continue to produce less effective public services.

2.4 Effects of the certification program

We next turn to the empirically testable implications of the model regarding the effects of implementing the certification program on certification and service delivery outcomes.⁷ First, the following proposition establishes that the outcome of certification reports by a third party pertaining to investments in future service delivery is independent of whether the certifying third party is corruptible:

Proposition 3. *Assume that $\gamma^* \in [\gamma_B, \gamma_A)$. If $\gamma_B < \gamma^{**}$, the third party does not have to certify whether there has been investment in future delivery of service B ($c_B = \phi$), while it always reports $c_A = 1$ regardless of whether it is clean or corruptible. If $\gamma_B \geq \gamma^{**}$, the third party always reports $c_i = 1$ for each service $i = A, B$ regardless of whether it is clean or corruptible.*

⁷As we show below, the voter (or a social planner) would always want to implement the program, although some corruptible incumbents would prefer the program not to be implemented.

Under the certification program, this result unsurprisingly highlights that, whenever third parties have to certify future service delivery, corruptible third parties always certify that incumbents invested in future delivery of each service i regardless of whether they invested or not, and thus of the incumbent's type. More interestingly, because honest certifying third parties only certify actual investments, incumbents that are corruptible in producing service i must also invest in future delivery of that service to avoid their corruptible type being revealed to voters. As a result, honest third parties also always certify that investments occur because, in equilibrium, all incumbents invest in future service delivery when faced with an honest third party.

Second, our next proposition establishes the implications for the total delivery of each service i that the voter experiences across periods. (Total service delivery is equivalent to voter utility in this model.) The results demonstrate that, while the *expected* effect of the program on service delivery is non-negative for each service i , whether it is strictly positive depends on the share of politicians that are clean or corruptible in providing the service:

Proposition 4. *Assume that $\gamma^* \in [\gamma_B, \gamma_A)$. The expected effects of the certification program on total delivery of each service i are given by:*

$$\begin{aligned} \Delta_A &:= \mathbb{E} \left[\sum_{t=1,2} U(g_{A,t}, G_{A,t}) \middle| p = 1 \right] - \mathbb{E} \left[\sum_{t=1,2} U(g_{A,t}, G_{A,t}) \middle| p = 0 \right] \\ &= (1 - \gamma_A) \rho \beta, \\ \Delta_B &:= \mathbb{E} \left[\sum_{t=1,2} U(g_{B,t}, G_{B,t}) \middle| p = 1 \right] - \mathbb{E} \left[\sum_{t=1,2} U(g_{B,t}, G_{B,t}) \middle| p = 0 \right] \\ &= \begin{cases} 0 & \text{if } \gamma_B < \gamma^{**} \\ (\gamma_B + (1 - \gamma_B) \rho) \beta - 1 & \text{if } \gamma_B \geq \gamma^{**}, \end{cases} \end{aligned}$$

and are positive and increasing in ρ when $\gamma_i > \gamma^{**}$ and zero when $\gamma_i \leq \gamma^{**}$.

Provided that the share of politicians that are clean in producing service i is not too low, Proposition 4 shows that the certification program leads to an expected increase in service delivery across

periods. When $\gamma_i < \gamma^{**}$, there are too few clean incumbents in producing service i for the program permit clean types to invest in future delivery of that service and thereby induce corruptible types also to invest when the third party is honest. As a result, the program does not affect delivery of service i . Once $\gamma_i \in (\gamma^{**}, \gamma^*]$, the program increases expected service delivery since investments in future service delivery become feasible for clean incumbents and forced upon corruptible incumbents by honest third parties. When $\gamma_i \geq \gamma^*$, investments were already feasible absent the program, but the certification program induces honest third parties to enforce investments on corruptible incumbents in the first period.

The expected effect of the certification program on service delivery is decreasing in the likelihood that the certifying third party is corrupt.⁸ This is, first, because third parties that are corruptible in producing a given service claim that they are investing in future service delivery while actually misappropriating public funds. Absent the certification program, these incumbents would have instead allocated the budget to less effective delivery of service i when $\gamma_i < \gamma^*$. Second, corruptible third parties do not discipline corruptible incumbents into investing in future service delivery when $\gamma_i \geq \gamma^{**}$.

3 Background

3.1 Mexican municipal governments and their poor institutional capacity

Mexico's federal system is divided into 31 states (and the Federal District of Mexico City), which contain around 2,500 municipalities. Municipalities are governed by mayors, who lead a municipal council in which a co-partisan majority is guaranteed. Until a recent reform that allowed re-election in some states starting in 2018, municipal mayors were typically elected to three-year non-renewable terms. While this reform strengthened accountability linkages, voters already held

⁸We focus on the first-order effect of ρ , taking γ^{**} as fixed. The second-order effect of ρ through γ^{**} ($\frac{\partial \gamma^{**}}{\partial \rho} < 0$) yields qualitatively similar results as a larger ρ expands the range of γ for which an equilibrium shift can be induced by the certification program.

parties responsible for the performance of aligned mayors given the relative strength of the party label and the role they play in candidate selection (Chong et al. 2015; Langston 2003; Marshall 2019). In practice, relatively few voters can name their incumbent mayor, but the vast majority know their party.

Following major decentralization reforms in the 1990s (see Wellenstein, Núñez and Andrés 2006), municipal governments—the focus of this article—became the main responsible for the day-to-day provision of basic local infrastructure and local public services. These include local policing, roads, sewerage, and water. Municipalities also assist state and federal governments in the provision of other public services including elementary education, health services, and environmental protection.

While the decentralization reforms significantly increased the expenditure authority of municipalities, they were not accompanied by a corresponding increase in tax collection responsibilities. The fiscal capacity of municipal governments to raise their own resources has actually declined over time relative to federal and state transfers: whereas municipalities raised 39% of their own revenues in the 1990s, this had declined to less than 20% by 2010 (Castañeda and Pardinás 2012). Spending by municipal governments accounts for 20% of total government spending, yet municipalities continue to be funded primarily by formula-based transfers from federal and state governments.

In part due to their inability to generate revenues, Mexican municipalities often also lack the institutional capacity to effectively deliver public services and manage local infrastructure. With the exception of large urban municipalities, most municipalities lack procedures for the provision and management of local public services, have low tax-collection capacity, lack trained officials, and are reluctant to depoliticize their administrative functions (Pérez Archudia and Arenas Aréchiga 2012). This situation further deteriorated with the escalation of drug trafficking-related violence in Mexico since 2006 (Dell 2015; Pulido-Gómez 2018; Trejo and Ley 2019). According to the head of legal affairs at the Interior Ministry (Secretaría de Gobernación, SEGOB), 75% of Mexican mu-

municipalities are susceptible to infiltration and corruption by organized crime since they have little or no tools to combat criminal influence.⁹ The institutional weakness of municipal governments is highlighted by both government officials and researchers as the underlying cause for the spread of organized crime (Aguirre and Herrera 2013).¹⁰

3.2 The *From the Local Agenda* program

The *From the Local Agenda* (Agenda desde lo Local, ADL) program—which is now called the Municipal Development Agenda (Agenda para el Desarrollo Municipal, ADM)—was motivated by the desire to improve service delivery and facilitate local development.¹¹ The program has been administered by Mexico’s Interior Ministry through the National Institute for Federalism and Municipal Development (Instituto Nacional para el Federalismo y el Desarrollo Municipal, INAFED). The INAFED developed and implemented the ADL program in line with Agenda 21, an action plan designed by the United Nations to promote better governance and sustainable and inclusive economic, social, and environmental development in the 21st Century.¹²

The INAFED has implemented the ADL program together with, and largely through, state governments. Participation by municipal governments is voluntary, although most municipalities have entered the program under state guidance. The program consists of four stages: self-assessment, third-party verification, effort to improve in under-performing areas, and updated verification and certification.

In the first stage of the program, municipal governments—aided by state governments—self-

⁹See *InSight Crime*, “75% of Mexico Municipalities Susceptible to Organized Crime: Official,” January 23rd 2015.

¹⁰According to the Minister of Interior this institutional weakness is behind the spread of organized crime. See *Sin Embargo*, “Osorio Chong reconoce que hay “debilidad” institucional en municipios,” February 13th 2015.

¹¹In 2014 the ADL program was revised and modified to include more aspects related with the role of municipal authorities in economic and social development, as well as environmental sustainability. See www.agendaparaeldesarrollomunicipal.gob.mx.

¹²This is a product of the United Nations Conference on Environment and Development, also known as Earth Summit, held in Rio de Janeiro, Brazil, in 1992. Section 3 and chapter 28 of the Agenda 21 embody the well-known Local Agenda 21 stating that local authorities are essential to promote sustainable development. See <https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf>.

assess their institutional capacity for service delivery and actual service delivery. Specifically, they assess their capacities and performance across 39 indexes, including 270 separate sub-indexes, which are grouped into four areas: (a) institutional capacity for good governance; (b) sustainable economic development; (c) inclusive social development; and (d) sustainable environmental development. During the self-assessment stage, municipal governments assign themselves one of the following statuses for each index and sub-index based on the guidelines stipulated by INAFED: red (completely undesirable situation and dramatic room for improvement), yellow (some room for improvement), or green (acceptable situation). The guidelines for each sub-index specify the quantitative indicators required to achieve each status.

In the second stage of the program, municipal governments' self-assessments are subjected to third-party verification, usually arranged by state governments. The third parties must be institutions of higher education, usually public or private local universities or other institutions of tertiary education. The use of these institutions was intended to ensure that verification was perceived as neutral and objective by government officials and citizens. However, the credibility of such institutions is challenged by the fact that institutions of higher education are largely funded by the federal and state governments. Especially when there is alignment between the incumbent parties at the municipal and state levels, governors may seek to manipulate third-party certification to benefit the reputation of their co-partisans and that of their party more generally. Moreover, there are countless instances of staff, including faculty and students, from institutions of higher education engaging in corruption.¹³ A recent corruption scandal resulting in the diversion of approximately USD 400 million of public funds involved the federal government and eleven institutions of higher education, four of which worked as third-party verifiers: Universidad Autónoma del Estado de México, Universidad Autónoma del Estado de Morelos, Universidad Politécnica del Golfo de México, and Universidad Tecnológica de Tabasco.¹⁴

¹³For an example of the latter, *Educación Futura*, “Corrupción en un instituto tecnológico,” February 11th 2016 shows evidence that acceptance to the Instituto Tecnológico de Querétaro was sold for USD 1,000 and faculty modified grades for USD 50 or sexual favors.

¹⁴See *New York Times*, “‘El dinero se iba a un agujero negro:’ el esquema de corrupción que compromete al

The faculty and students from those institutions who act as the verification team receive training on the indexes and corresponding criteria to be examined. They are responsible for reviewing the supporting documentation provided by municipal governments and validating the governments; self-assessment of each sub-index.¹⁵ Even where verification worked best, auditor accounts still frequently indicated concerning flaws in the process, including instances of municipalities awarding themselves a high status along many indexes and municipal officials selectively providing evidence to support each sub-index status. In many cases, the verification team simply had to trust the information provided by officials without being able to scrutinize the self-diagnosis in greater detail or even examine the original data (Turrubiates Flores, Vargas Cuéllar and Suárez Rodríguez 2014).

In 2017, 1,164 municipalities in 30 states participated in the program, but only 863 concluded the verification process. In total, 1,827 individuals—including faculty and students—from 163 institutions of higher education verified the self diagnoses proposed by municipal governments. Out of these 163 institutions, 99 (61%) were universities, 57 (35%) technological institutes, 5 (3%) local colleges, and 2 (1%) higher education institutes. The mean institution conducted slightly more than 5 verifications, while the median conducted 3.¹⁶

In the third part of the program, after identifying areas for improvement in diverse aspects of their administration, municipal governments—again aided by state governments—produce and execute plans to strengthen municipal capacity to improve service delivery. These plans focus particularly on the indexes which were assigned a red status, and often include the training of municipal officials by state governments. To reflect improvement in these areas, municipal officials then reassess their self-diagnosis, which is again subject to third-party verification.

In the fourth stage of the program, the From the Local National Council (Consejo Nacional Desde lo Local)—which is formed by representatives from the federal and state governments,

gobierno de México,” September 5th 2017 for more details.

¹⁵See the [INAFED website](#) for more details.

¹⁶See the [INAFED website](#) for more details.

as well as representatives from institutions of higher education—grants certificates to municipal governments for each index that is certified as green. These certificates are handed out by federal and state officials in award ceremonies, which are widely publicized by municipal governments and local media. These usually highlight not only the great work of municipal government officials, but the fact that results are subject to third-party auditing by local higher education institutions. It is also often mentioned that the municipal governments are certified using international standards.¹⁷ This information is likely to have important electoral consequences in a context where voters are poorly informed about mayoral responsibilities and performance regarding public service delivery (Chong et al. 2015), and rely on local broadcast media to learn about mayoral malfeasance (e.g. Castañeda Sabido 2011).

3.3 Mapping theory to empirics

We next translate testable implications of our theoretical model to the context of the ADL program. We note two modeling simplifications when bringing the model to the data. First, the ADL program certified the status of municipal government capacity for service delivery and actual public service delivery on a three-point scale ranging from a completely undesirable situation (red) to an acceptable situation (green). Our stylized model, however, captures investments that enhance service delivery without distinguishing whether improvement was from a red to a yellow or green situation or from a yellow to a green situation. The model can naturally be extended to match our empirical analysis treating all one-point increases in the scale equally.

Second, voters experience utility from current and future service delivery in the model. In practice, the indexes certified by the ADL program do not distinguish between current and future service delivery. We approximate the utility received by voters from public services by focusing

¹⁷*Códice Informativo*, “Reconocen a Corregidora por resultados positivos en Agenda Para el Desarrollo Municipal,” January 17th 2018; *Línea de Contraste*, “Reconocen Alcaldía de Tlaxcala por implementación del programa Agenda para el Desarrollo Municipal 2018,” November 23rd 2018; *Moreli Activa*, “Reconocen a 11 municipios michoacanos por su participacin en el Programa Agenda para el Desarrollo Municipal,” November 21st 2018.

on the indexes that relate to the capacity to deliver services or metrics of actual service delivery, which broadly capture the current and future returns to investments in service delivery. In line with our model's focus on differences across types of municipality, these outcomes—which we will measure roughly biennially in many cases—reflect the accumulation of municipal governments actions over time.

We begin with our expectation, which emerges from Proposition 3, for the effect of entering the ADL program on the certified status of a municipality's program indexes. As noted above, the status of each index is first certified before municipal governments have the opportunity to invest in improving service delivery; the subsequent statuses that we then reflect whether government plans to strengthen service delivery are implemented. Hypothesis 1 captures the fact that—whether through investments in service delivery or collusion with the third party—participation in the ADL program is expected to improve a municipality's certification status on a given index:

Hypothesis 1. *The certification status of a program index increases from the assessment received upon entry into the ADL program.*

We next turn to our primary hypotheses concerning the impact of the ADL program on municipal service delivery. Given that certification statuses are susceptible to corruption, we focus on the program indexes whose underlying indicators we can measure independently of the program. Following Proposition 4, we first hypothesize that entering the ADL program should make voters better off in expectation:

Hypothesis 2. *On average, entering the ADL program increases the municipal public service delivery indicators corresponding to a given program index.*

Where the probability that the mayor can engage in corruption in providing a given service is sufficiently low, the effect of the program on service delivery is positive. If the probability is sufficiently high, voters will not permit a shift away from observable low-efficacy public service delivery. Consequently, as Proposition 4 demonstrates, the size of the average effect, and thus

our ability to empirically detect it—depends on the overall likelihood that mayors can engage in corruption.

Next, we consider how the ADL program’s efficacy is likely to vary with the degree to which the certification process can shield corruption. Based on how certifiers are chosen in practice, a plausible proxy for the likelihood that the third party certifying index of municipal government is corrupt (i.e. ρ) is whether a given municipal government is politically aligned with the state government. Using this proxy, Proposition 4 implies:

Hypothesis 3. *The average effect of entering the ADL program on the municipal public service delivery indicators corresponding to a given program index is lower in municipalities whose governments are copartisans of their state government.*

Where the probability that the third party will be corruptible is sufficiently high, the ADL program will not affect service delivery outcomes at all.

Second, we consider a low baseline certification status upon entry into the ADL program as a proxy for the probability that the incumbent is corruptible in producing services in a given index (i.e. γ_i).¹⁸ Our final hypothesis then follows from Proposition 4:

Hypothesis 4. *The average effect of entering the ADL program on the municipal public service delivery indicators corresponding to a given program index is greater for indexes whose certification status upon entry into the program is not low.*

If a low certification status corresponds to the case where the probability of corruptibility in producing a service is sufficiently high, then the ADL program will not affect services delivered as

¹⁸The theoretical model highlights that corruptible municipal incumbents in producing a given service should be associated with a lower public delivery of the service. It is possible that such corruptible incumbents could collude with corruptible certifying third parties to alter the baseline certification status of the service, which might bias our estimates. Two reasons ease this concern. First, municipal governments have little time to prepare for the first phase of the program. The short turnaround time between entering the program, self-assessment, and evaluation suggests that municipalities with lower public service delivery indexes received low baseline levels of certification of the indexes. Second, corruptible municipal governments in producing a given service have incentives to start with low certification statuses in order to be able to certify improvements. Accordingly, low baseline certification statuses may be indicative of corruptible incumbents in delivering the services measured by the indexes.

part of that index.

4 Research design

We next explain our identification strategy and how it allows us to test the hypotheses enumerated in section 3.3, before describing the data that underpins it.

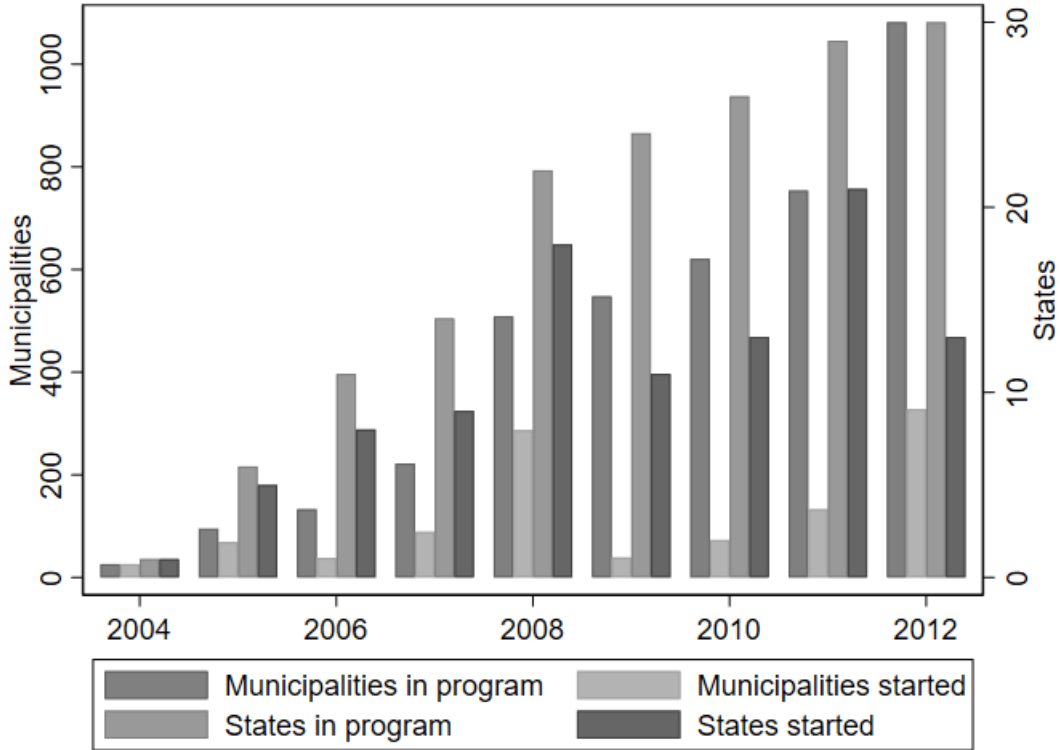
4.1 Identification strategy

To identify the effects of the ADL program on municipal public service delivery outcomes, we exploit the staggered entry of municipalities into the program using a generalized difference-in-differences design. Specifically, we estimate the effects of a municipality entering the program on certification status and a range of indicators collected independently from the program that capture public service delivery outcomes relating to the program indexes by comparing changes in outcomes across the years before and after a municipality entered the program relative to municipalities that entered the program at an earlier or later date.

Figure 1 shows the number of municipalities and states in the ADL program, as well as the corresponding number that started in a given year. The number of municipalities that enter the program is associated with state adoption of the program although there is also significant within-state variation in the timing of a municipality's entry. Since the first certification was conducted before municipalities could make investments to improve service delivery, we naturally define all certification periods after the first certification as post-certification program periods. For the annual indicators collected independently from the program that capture public service delivery outcomes relating to the program indexes, we define post-certification years starting with the year that certification results were first released for the municipality.

To test hypothesis 1, which states that the certified status should increase over time for the

Figure 1: Stocks and flows of municipalities and states participating in the ADL program



municipalities that entered the ADL program, we estimate the following specification:

$$Y_{imt} = \beta Program_{mt} + \eta_{ist} + \theta_{im} + \varepsilon_{imt}, \quad (2)$$

where Y_{imt} is the certified status for index i in municipality m (of state s) in year t , and $Program_{mt}$ is an indicator for whether municipality m had entered the program by year t or not. We include state \times year \times index fixed effects, η_{ist} , to account for any state-specific shocks that might affect the status of an index in a given year. This can, for example, flexibly adjust for uniform changes in how certifier standards or common constraints on municipality service delivery. We complete the difference-in-differences design by also including municipality \times index fixed effects, θ_{im} , to absorb all time-invariant factors influencing a municipality's production of the services corresponding

to a given index. Our coefficient of interest is β , which estimates the effect of being in the program for a year on index certification status.

To test hypothesis 2, which predicts a positive effect of entering the ADL program on public service delivery, we estimate the same specification as in equation (2) but using public service delivery outcomes. In particular, for the indexes for which the corresponding indicators could be independently obtained, we consider the indicators of public service delivery upon which the indexes statuses are supposedly based.

Hypothesis 3 predicts that such an effect will be lower for municipalities partisanship matches that of their state government, which we test by estimating the following specification:

$$Y_{imt} = \beta_1 Program_{mt} + \beta_2 (Program_{mt} \times State Alignment_{mt}) + \eta_{ist} + \eta'_{ist} State Alignment_{mt} + \theta_{im} + \theta'_{im} State Alignment_{mt} + \varepsilon_{imt}, \quad (3)$$

where $State Alignment_{mt}$ indicates whether the party that governs municipality m in year t is also the party that governs.¹⁹ The fixed effects are further interacted with state alignment in order to exploit only variation within the sets of aligned and unaligned municipalities when estimating the effects of entering the ADL program. The coefficient on β_1 estimates the effect of entering the program among municipalities that are not aligned with the state government, while the coefficient on β_2 captures the differential effect of the program among municipalities aligned with the state government. The effect of entering the program among aligned municipalities is thus given by $\beta_1 + \beta_2$.

Finally, hypothesis 4 predicts that the effect of the ADL program on public service delivery should be lower for cases where the certification status of the index at the time of entering the

¹⁹In case the municipality is governed by a coalition of multiple parties we consider the municipality aligned if any of the parties in the coalition are the same as the party of the state governor.

program is low. We test this prediction using the following specification:

$$Y_{imt} = \beta_1 Program_{mt} + \beta_2 (Program_{mt} \times Low\ Baseline_{im}) + \eta_{ist} + \eta'_{ist} Low\ Baseline_{im} + \theta_{im} + \theta'_{im} Low\ Baseline_{im} + \varepsilon_{imt}, \quad (4)$$

where $Low\ Baseline_{im}$ indicates whether the municipality received a low baseline certification on the index associated with service delivery indicator i . The state \times year \times index and municipality \times index fixed effects are similarly interacted with the $Low\ Baseline_{im}$ dummy to exploit variation only within the sets of low and higher baseline index statuses. The benchmark coefficient, β_1 , estimates the effect of entering the ADL program on indexes certified as having a yellow or green status when the municipality entered the program. The coefficient on β_2 , in turn, captures the differential effect of the program when the baseline certification status was low instead. The effect of entering the program among low baseline municipalities is thus given by $\beta_1 + \beta_2$.

4.2 Validating the identification strategy

This design principally relies on a parallel trends assumption to identify the effects of the ADL program on certification and service delivery outcomes. This entails that municipalities that entered the program earlier (later) would otherwise have followed the same trend as municipalities that entered the program later (earlier).²⁰ We follow two common approaches to validate the plausibility of this assumption. First, we conduct the following event study analysis:

$$Y_{imt} = \sum_{\tau=-k}^k \beta_{\tau} Enter_{mt+\tau} + \eta_{ist} + \theta_{im} + \varepsilon_{imt}, \quad (5)$$

²⁰Since municipalities that entered the program earlier also serve as counterfactuals for municipalities that become treated later, the parallel trends assumption also incorporates the assumption that treatment effects are constant over time.

where $Enter_{mt+\tau}$ is an indicator for the year τ relative to the year (normalized to 0) in which the municipality entered the program. Second, and close related, we focus on just the years preceding entry into the program, using the following specification to estimate differences in outcomes in the k years preceding a municipality’s entry into the program:

$$Y_{imt} = \sum_{\tau=0}^k \beta_{1\tau} Program_{mt+\tau} + \eta_{ist} + \theta_{im} + \varepsilon_{imt}, \quad (6)$$

where $Program_{mt+\tau}$ is now an indicator for municipality m being in the program τ years into the future. We consider up to 3 leads (i.e. $k = 3$). We conduct analogous exercises for the heterogeneous effects underpinning hypotheses 3 and 4.

4.3 Data

Data on participation in the ADL program and the index certifications come from the website of the National Institute for Federalism and Municipal Development (INAFED).²¹ This data allows us to identify which municipalities participated in a given year between 2004, the first year of the program, and 2013—the last year before the program switched name and slightly altered its implementation. The average municipality entered the program in 2009, and the municipal incumbent was aligned with the state governor’s party in 58% of years in our sample. INAFED also provides the certified status that each municipality received for each of the 39 program indexes. We focus on these indexes, since reliable independent data for sub-indexes is not available for the majority of the program.

Data on public service delivery outcomes associated with the 39 indexes certified within the ADL program comes from three sources, which are independent of the program and vary in the years in which they are available. First, we mainly use a census of municipal public service delivery conducted by the National Institute of Statistics and Geography (INEGI) in 2000, 2002,

²¹www.gob.mx/inafed/acciones-y-programas/resultados-historicos-del-programa-agenda-desde-lo-local.

2004, 2009, 2011, and 2013. For each wave of these municipal surveys, we have detailed measures of every municipality's personnel (by age, education, and department), resources (number of computers, vehicles, etc.), public service provision, active regulations, and more. Importantly, the INEGI was not involved in the ADL program at any stage and surveyed municipal governments independently from ADL program activities.

Second, for the certification indexes relating to municipal finances (see 1.1 and 1.11 in Table 1), we use public finance data published by the INEGI between 2000 and 2013. This provides detailed information for many of the outcomes of the financial sub-indexes.²²

Third, since one index is related to housing (see 3.8 in Table 1) and several others to public service delivery to Mexican households, we exploit information from the quinquennial population censuses conducted by the INEGI between 2000 and 2010.²³ Specifically, we use information on the extent to which households have access to the electricity, sewage, and water grids, as well as the quality of their dwelling (e.g., floor, roof, and wall material) and extent of overcrowding. These measures overlap closely with several sub-indexes contained in the ADL program.

These independent measures allow us to investigate actual changes in the quality and quantity of municipal public service delivery due to the ADL program. Such indicators exactly match the ADL's program's evaluation criteria for a given sub-index in most cases, enabling us to replicate whether public service delivery actually satisfied the conditions to obtain red, yellow, or green certification status for a given sub-index. We refer to the independent measures of service delivery that correspond to the ADL's sub-indexes as indicators. We restrict attention to 50 indicators that closely correspond to sub-indexes within 10 of the ADL program's 39 indexes. Of these, 80% received a red status initially upon entry into the ADL program.

Appendix Table A3 explains, for each of these 10 indexes, which sub-indexes we aggregate, their source, and how they are coded. For example, for index 1.11 measuring healthy finances,

²²See more details in the INEGI website.

²³See more details in the INEGI website.

Table 1: Summary statistics by service delivery index

Index	Variable	Mean	Standard Deviation	Min	Max
1.1	Own income / Current expenditure	0.48	0.50	0	11.87
	Current expenditure / Total expenses	0.71	0.10	0.11	1
	Public investment / Total income	0.30	0.16	0	1
	Personal services / Current expenditure	-1.20	0.73	-37.13	-0.00
	Total expenses + Total revenue	0.13	0.27	-3.10	5.21
1.4	Sector for promoting social participation?	0.32	0.47	0	1
	Participation of commissions and/or communal committees	0.34	0.47	0	1
	Index for regulations for participation	0.10	0.43	0	2
	Are there mechanisms for citizens participation?	0.86	0.34	0	1
1.6	Is there a plan for civil protection?	0.72	0.45	0	1
	Is there a map of risk zones?	0.48	0.50	0	1
	Index for regulations on civil protection	0.37	0.77	0	2
1.8	Share of sectors have regulations	0.41	0.30	0	1
	Index of regulations	0.52	0.77	0	2
1.10	Institution responsible for transparency?	0.56	0.33	0	1
	Regulations to regulate access to public info.?	0.55	0.50	0	1
	Public servants responsible for public info.?	0.67	0.47	0	1
	A system of reception of and attention to public info. requests	0.55	0.50	0	1
	A system for archives	0.34	0.47	0	1
	Training program for public servants on public info.	0.34	0.47	0	1
	Is there open access?	0.86	0.35	0	1
	Are there regulations about transparency?	0.38	0.49	0	1
1.11	Debt accumulated from previous years	-2.16	6.55	-21.43	2.30
	Share of budgeted contributions collected (ordinal)	2.77	1.33	1	5
	Federal transfers / Total income	0.55	0.16	0	1
3.1	Share of mun. capital covered by drainage and sewage	0.78	0.27	0	1
	Share rest of mun. covered by drainage and sewage	0.52	0.34	0	1
	Share of mun. capital covered by public lighting	0.82	0.24	0	1
	Share of rest of mun. covered by public lighting	0.64	0.31	0	1
	Share of mun. capital covered by street cleaning	0.81	0.27	0	1
	Share of rest of mun. covered by street cleaning	0.59	0.35	0	1
	Share of mun. capital covered by trash collection	0.84	0.24	0	1
	Share of rest of mun. covered by trash collection	0.62	0.34	0	1
	Are there grave regulations?	0.43	0.49	0	1
	Are there market regulations?	0.44	0.50	0	1
3.2	Do regulations on performance and sport exist?	0.36	0.48	0	1
	Index for regulations on performance and sport	0.25	0.65	0	2
3.6	Share of mun. capital covered by drinking water	0.84	0.22	0	1
	Share of rest of mun. covered by drinking water	0.64	0.30	0	1
	Share of mun. capital covered by drainage and sewage	0.78	0.27	0	1
	Share of rest of mun. covered by drainage and sewage	0.52	0.34	0	1
3.8	Share of mun. capital covered by drinking water	0.84	0.22	0	1
	Share of rest of mun. covered by drinking water	0.64	0.30	0	1
	Share of mun. capital covered by drainage and sewage	0.78	0.27	0	1
	Share of rest of mun. covered by drainage and sewage	0.52	0.34	0	1
	Share of occupants in homes with drainage and / or toilet	0.89	0.13	0.20	1
	Share of occupants in homes with electric power	0.95	0.08	0.10	1
	Share of occupants in houses with dirt floor	0.85	0.16	0.04	1
Share of households without overcrowding	0.53	0.13	0.14	0.89	
Observations		13,757			

Notes: This table shows summary statistics for the outcome variables used for each index. Variables with extreme outliers are winsorized in the main analysis.

we use municipal census data to construct an indicator corresponding to the sub-index capturing debt accumulated from previous years (inversed and logged) and municipal public finance data to construct an indicator corresponding to the sub-index capturing what percentage of the municipal income comes from federal transfers that takes the value 0 (red status) if more than 95% of the income comes from federal transfers, 1 (yellow) if more than 75% but less than 95% of income are federal transfers, and 2 (green) if less than 75% of income are federal transfers. For index 3.8 concerning housing, we use data from the population census to construct an indicator corresponding to a subindex that takes the value 0 (red status) if less than 50% of the inhabitants in the municipality have electricity, 1 (yellow) if more than 50% but less than 75% of inhabitants have electricity, and 2 (green) if more than 75% of inhabitants have electricity. Table 1 reports the summary statistics for each of our outcome variables by index.

For our analysis, the variables within each index are combined into a standardized index (z-score).²⁴ This approach is more fine-grained than the certification status coding, and is thus more sensitive to changes in actual service delivery than the coarse red/yellow/green categories that third parties were asked to assign. To the extent that service provision actually improves, such measures are well-placed to detect it.

5 Results

We begin by showing that the certified status of the indexes generally increased after municipalities entered the ADL program. We next assess the extent to which actual service delivery outcomes change. Our main contribution is to show that independent measures of the indexes that the ADL program is supposed to certify paint a far less sanguine picture in an environment with opportunistic politicians and third-party certifiers. In line with our model emphasizing the corruptibility of the certification process, we find that participation in the program does not increase munic-

²⁴For years in which some indicators of an standardized index are missing due to data availability constraints, we calculate the standardized index over the available indicators.

palties' performance on independently-measured indicators of the program indexes intended to certify public service delivery improvements on average. Moreover, and consistent with the empirical predictions of the theoretical model, the effect of the program on public service delivery decreases with the political alignment of the municipal and state governments, and is weaker for indexes with a low baseline certified status. These heterogeneous effects suggest that the ADL program's efficacy was largely undermined by Mexico's weak political institutions.

5.1 The effect of entering the ADL program on certified status over time

We first examine hypothesis 1 concerning the effect of the ADL program on the certified status of the program indexes of participating municipalities. Table 2 reports our estimates of equation (2), where the outcome is an ordinal scale for whether an index was certified as red (coded as 0), yellow (coded as 1), or green (coded as 2). Column (1) focuses on all program indexes, whereas column (2) restricts attention to the indexes for which we have independent measures of the service delivery indicators corresponding to sub-index outcomes.

Both columns (1) and (2) show that certification status significantly increased over time among participating municipalities. Among the sub-indexes for which independent measures of service delivery outcomes are available, column (2) indicates that the effect is more sizable: whereas certification increases by 0.4 levels above the baseline status when considering across all indexes, this effect slightly increases to 0.45 levels among the indexes that we focus on. Figure 2 illustrates this effect graphically over the duration of a municipality's participation in the program.²⁵ These estimates show that, after 5 years in the program, the average index has increased by 0.5 levels.

²⁵Note that we are unable to include the values prior to the municipality entering the program in Figure 2 since, naturally, there are no certification outcomes before that.

Table 2: Effect of the ADL program on certified status

	(1)	(2)
	All Indexes	Indexes with corresponding indicators
Program	0.408*** (0.024)	0.449*** (0.027)
Observations	550,241	83,552
R^2	0.83	0.82
Mean of Outcome	1.73	1.68
SD of Outcome	0.91	0.90
Min of Outcome	1	1
Max of Outcome	3	3

Notes: This table shows the regression results of the main specification using program certification as outcomes. An observation represents a program index in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-index and municipality-index fixed effects. Column (1) uses the certification status for all indicators of the program. Column (2) restricts to indexes for which we have independent measures. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

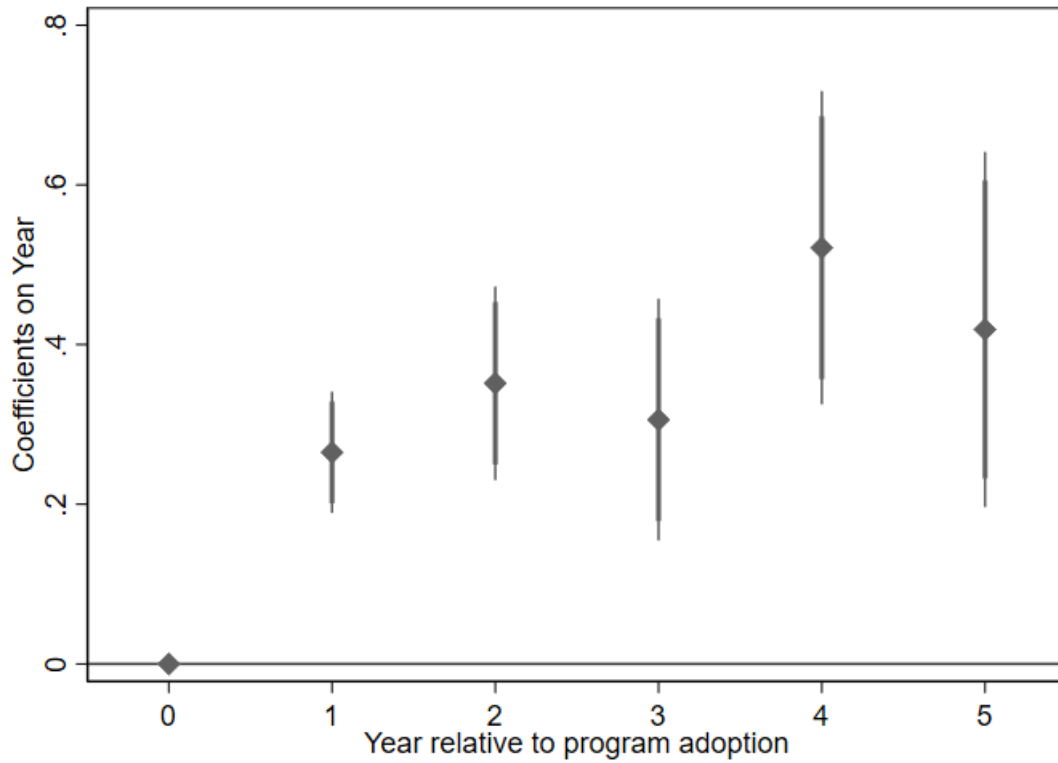
5.2 The effect of the ADL program on public service delivery

We next assess the effect of the ADL program on actual public service delivery to test hypotheses 2-4. Table 3 considers as its outcome the independently-constructed indicators that correspond to the indexes certified as part of the program.

In contrast with hypothesis 2, the null finding in column (1) indicates that there is, on average, no positive effect of the program on public service delivery. The leads included in columns (2)-(4) further show that this estimate is unlikely to be driven by differential pre-trends in public service delivery in municipalities that entered the program earlier than others. This lends support to the identification assumption underpinning our generalized difference-in-differences design. These results are corroborated visually in Figure 3.

A priori, the lack of a clear positive effect on average could be considered inconsistent with hypothesis 2, and consequently the theory. However, there are both theoretical and empirical

Figure 2: Effect of the ADL program on certified status



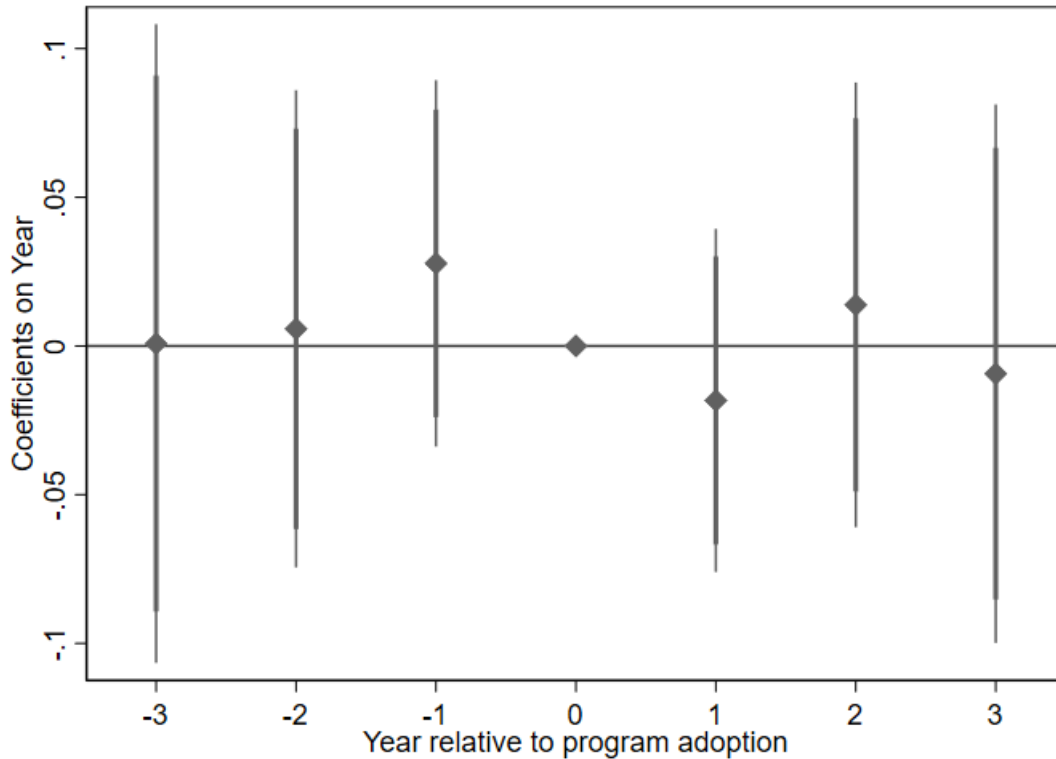
Notes: This figure shows the coefficients and confidence intervals of a regression of certified status on the year since a municipality entered the program. The sample includes each of the indexes for which we have measures constructed with data collected independently from the program. The specification includes state-year-index and municipality-index fixed effects.

Table 3: Effect of the ADL program on public service delivery

	(1)	(2)	(3)	(4)
	Z-score	Z-score	Z-score	Z-score
Program	-0.0016 (0.0236)	0.0071 (0.0260)	0.0117 (0.0263)	0.0117 (0.0264)
Lead 1		-0.0169 (0.0246)	-0.0306 (0.0271)	-0.0307 (0.0271)
Lead 2			0.0339 (0.0259)	0.0343 (0.0268)
Lead 3				-0.0008 (0.0243)
Observations	90,442	90,442	90,442	90,442
R^2	0.41	0.41	0.41	0.41
Mean of Outcome	0	0	0	0
SD of Outcome	1	1	1	1
Min of Outcome	-9.43	-9.43	-9.43	-9.43
Max of Outcome	8.71	8.71	8.71	8.71

Notes: This table shows the regression results of the main specification using program indexes of public service delivery as outcomes. An observation represents a program index in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-index and municipality-index fixed effects. Column (1) follows the main specification. Column (2) includes 1 year leads. Column (3) includes 2 year leads and Column (4) includes 3 year leads. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3: Effect of the Program on Public Service Delivery



Notes: This figure shows the coefficients and confidence intervals of a regression of public service delivery on the year since a municipality entered the program. The specification includes year-index-state and municipality-index fixed effects.

reasons to be unsurprised by this small effect. First, our model demonstrates that the anticipated positive effect of the program is expected to be small when the likelihood of corrupt certifying third parties and municipal incumbents are high (i.e. ρ and γ_i are small, in the context of the model). Second, to maximize the credibility of our estimation strategy, our specification includes state \times year \times index fixed effects. However, such fixed effects could also absorb part of the average effect of the program. We then turn to the other empirical predictions of the model for further tests.

Table 4 focuses on the same outcome as Table 3, but now considers two interactions with our indicator for a municipality entering the ADL program. Column (1) reports the results for the interaction with political alignment between municipal and state governments, which proxies for

the likelihood of corruptibility of the certifying third parties. Consistent with hypothesis 3, the negative interaction coefficient indicates that the effect of the program on public service delivery significantly decreases with state alignment. Moreover, while not statistically significant, we observe a 0.05 standard deviation positive effect of the program in municipalities that are not politically aligned with their state governments. As the coefficient test at the foot of column (1) shows, the effect is negative in aligned municipalities, although this is only statistically significant at the 10% level. Together, these results suggest that the ADL program can provide cover for corruption in some contexts. These effects are also depicted in Figure 4, which reports the effects of the program on public service delivery by year since the year of program adoption. The plot on the left shows the effect for municipalities that are not politically aligned with the state government, while the plot on the right shows the differential effect for municipalities politically aligned with the state government. Figure 4 also suggests that differential pre-trends in the public service delivery of municipalities that entered the program earlier rather than later are not driving the effects estimated in each subgroup.

Column (2) of Table 4 reports the results for the interaction with a low baseline certified status of a given index. This represents a proxy for the likelihood that a municipal incumbent is corruptible with respect to a given index. Consistent with hypothesis 4, these results indicate that the effect of the program on public service delivery is statistically indistinguishable from zero for indicators corresponding to indexes with a low baseline status certified. However, as theorized, we observe a statistically significant and relatively sizable positive effect of the program in indexes with high baseline certification status. The interaction term demonstrates that the difference between the effects in each type of index is statistically significant. These effects are illustrated in Figure 5, which reports the effects of the program on public service delivery by year since entering the program. The plot on the left depicts the effect for indexes with a high baseline status certified, while the plot on the right shows the differential effect for indexes with a low baseline status certified. The coefficients on those plots again suggest that differential pre-trends do not account for results in

Table 4: Effect of the ADL program on public service delivery, by state alignment and low baseline certification of the index

	(1)	(2)
	Z-score	Z-score
Program	0.053 (0.037)	0.100* (0.059)
Program \times Aligned with State	-0.120** (0.051)	
Low baseline \times Program		-0.124** (0.061)
Observations	87,329	89,646
R^2	0.50	0.43
Mean of Outcome	0	0
SD of Outcome	1	1
Min of Outcome	-9.43	-9.43
Max of Outcome	8.71	8.71
Program + Program \times Aligned	-0.067* (0.035)	
Program + Program \times Low Baseline		-0.024 (0.025)

Notes: This table shows the regression results of the main specification using program indexes of public service delivery as outcomes. An observation represents a program index in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-index and municipality-index fixed effects. Column (1) interacts *Program* and the fixed effects with whether the municipality is governed by the same party as its state. Column (2) interacts them with whether the municipality received a low certification on the index in the year it entered the program instead. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

column (2) of Table 4.

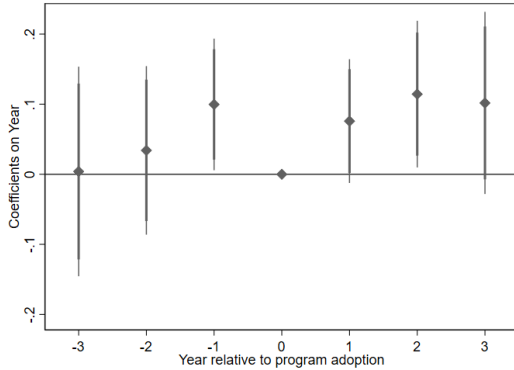
5.3 Robustness checks

We conduct several robustness exercises to demonstrate that our results are not driven by particular parameterizations of our regressors or data quality issues.

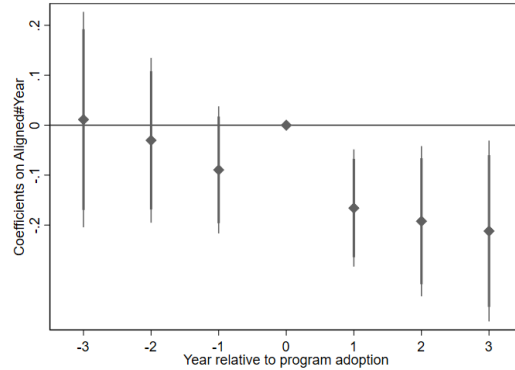
Appendix Table A1 shows that the results in Table 4—the core findings supporting our theoretical model, and explaining the ADL program’s negligible effect—are robust to using different

Figure 4: Effect of the ADL program on service delivery by state alignment

Panel A: Effect for unaligned municipalities



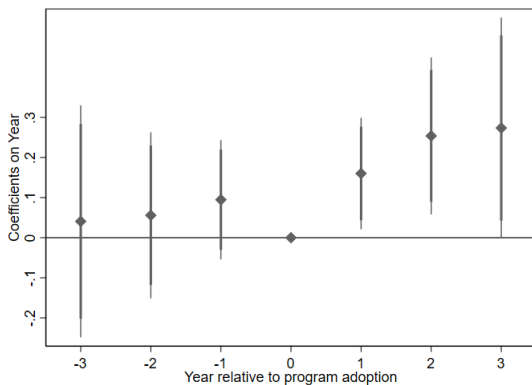
Panel B: Differential effect for aligned municipalities



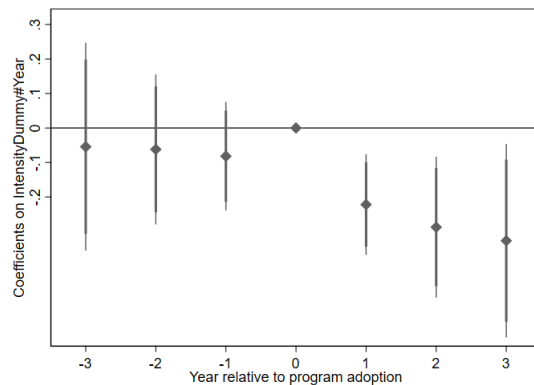
Notes: This figure shows the coefficients and confidence intervals of a regression of public service delivery on the year since a municipality entered the program interacted with state alignment. The sample includes each of the indexes for which we have measures constructed with data collected independently from the program. The specification includes state alignment-state-year-index and state alignment-municipality-index fixed effects. Panel A shows the coefficients on year since program start. Panel B shows the coefficients on year since program start *times* whether the municipality is governed by the same party as its state.

Figure 5: Effect of the ADL program on service delivery by baseline level of certification

Panel A: Effect for high baseline



Panel B: Differential effect for low baseline



Notes: This figure shows the coefficients and confidence intervals of the year since a municipality entered the program interacted with an indicator of low baseline level of certification status on the index upon entry into the program. The sample includes each of the indexes for which we have measures constructed with data collected independently from the program. The specification includes state-year-index and municipality-index fixed effects. Panel A shows the coefficients on year since program start. Panel B shows the coefficients on year since program start *times* an indicator for low baseline level of certification status.

measures of baseline level of certification status. In particular, our conclusions are qualitatively similar if the probability of a mayor being corruptible in producing a given service is instead proxied by a non-parametric or linearized operationalization of the baseline certified status of a given index.

We further conduct several exercises to ensure that our results are not influenced by the coding decisions used for our outcomes. Specifically, Appendix Table A2 reports the results using four alternative recoding strategies. Whereas the first column replicates our preferred estimation strategy, column (2) instead considers as our outcome a dummy for whether our indicators of service delivery show that at least the first cutoff—from red to yellow—was passed for each of the sub-indexes within an index. Column (3) further considers a dummy for whether our indicators show that the second cutoff—from yellow to green—was passed for each of the sub-indexes within an index. Columns (4) instead winsorizes the raw data used to code up the certification status in each of the subindexes at the 99th percentile. Column (5) winsorizes at the 95th percentile instead. These results are robust both in terms of significance and magnitude across these alternative specifications, suggesting that findings are not driven by our approach to mapping service delivery indicators to certified sub-indexes.

6 Conclusion

Our theory and evidence suggest that the short-term unobservability of investments to improve service delivery represent an important constraint on such critical investments. They also indicate that, at least in theory, the certification of investments to improve service delivery could help to overcome such an impediment. However, we show that such hopes may often be undermined by existing institutional weaknesses. In particular, we find that such a certification program in Mexico had no effect on service delivery on average, even though third-party certifiers announced significant improvements in indexes linked to these outcomes. Rather, in line with our theoretical

argument, the negligible effects on actual service delivery are largely explained by the corruptibility of both the third parties certifying investments in service delivery and the municipal incumbents in producing given public services.

While this study suggests the corruptible political institutions can stymie reforms aiming to improve service delivery in the context of Mexican municipalities, it is natural to ask how far such insights could extend. The scope conditions of our theory suggest several reasons to believe that our argument may apply pervasively across the Global South. First, possibly due to the high potential returns to elected office, corruption is common within the national and local governments in many developing contexts. As our model highlights, high levels of corruption discourage clean politicians from making investments that voters regard as signals of corruption. Second, information about politician behavior in office is limited in most contexts, both by a lack of timely, transparent, or accessible budgeting information and the potential political costs to media outlets of reporting such information when it exists. Third, many states in the Global South are weak in terms of their capacity to roll out credible transparency reforms and prevent collusion. We find that, at least for some politicians, this hindered the Mexican central government's efforts to incentivize investments in service delivery by providing a shield for corrupt activities. Sadly, this provides a logic as to why well-intentioned governments with limited local control might rationally avoid such reforms. Nevertheless, further research is required to explore the consequences of similar reforms in other contexts and study the conditions under which higher levels of government seek to introduce programs to audit government investments.

Conversely, our findings also have implications for bureaucratic and political reforms designed to make policies that incentivize initially-unobservable investments in improving service delivery feasible. In particular, it is critical that auditors and certifiers remain impartial. This requires efforts to insulate these agents from the actors they evaluate, potentially including external validation of reports, higher salaries, and greater training and professionalization. In addition to, or instead of, altering the incentives for third parties to collude with those that they monitor, governments might

seek to address the structural factors responsible for negative selection into politics or increase the risk of meaningful sanctions for transgression. Given the political and financial challenges of implementing such reform, a central message of this study is the importance of recognizing the complementarities between accountability dynamics and state capacity. In short, several structural factors may need to change simultaneously to facilitate long-term investments in enhancing the state's delivery of services.

References

- Aguirre, Jerjes and Hugo Amador Herrera. 2013. "Institutional weakness and organized crime in Mexico: the case of Michoacán." *Trends in Organized Crime* 16(2):221–238.
- Banerjee, Abhijit V., Esther Duflo and Rachel Glennerster. 2008. "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system." *Journal of the European Economic Association* 6(2-3):487–500.
- Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. 2011. "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India." Working paper, http://scholar.harvard.edu/files/rpande/files/doinformedvoters_nov11_0.pdf.
- Bardhan, Pranab. 2002. "Decentralization and Governance in Development." *Journal of Economic Perspectives* 16(4):185–205.
- Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(1):19–42.
- Bhandari, Abhit, Horacio Larreguy and John Marshall. forthcoming. "Able and mostly willing: An empirical anatomy of information's effect on voter-driven accountability in Senegal." *American Journal of Political Science* .
- Bidwell, Kelly, Katherine Casey and Rachel Glennerster. 2020. "Debates: Voting and Expenditure Responses to Political Communication." *Journal of Political Economy* 128(8):2880–2924.
- Bowles, Jeremy and Horacio Larreguy. 2020. "Who Debates, Who Wins? At-Scale Experimental Evidence on the Supply of Policy Information in a Liberian Election." Working paper, https://www.jbowles.net/s/draft_13.pdf.

- Castañeda, Luis César and Juan E. Pardinas. 2012. “Sub-national Revenue Mobilization in Mexico.” Inter-American Development Bank Working Papers, No. IDB-WP-354, <https://publications.iadb.org/publications/english/document/Sub-national-Revenue-Mobilization-in-Mexico.pdf>.
- Castañeda Sabido, Fernando. 2011. “Investigación sobre la percepción ciudadana acerca de la transparencia, rendición de cuentas y fiscalización en el uso de los recursos públicos en México.” *Reporte para la Auditoría Superior de la Federación*, https://www.asf.gob.mx/uploads/78_Estudios_sobre_la_percepcion_de_la_ASF/Percepcion_ciudadana_ASF_2010-UNAM.pdf.
- Chong, Alberto, Ana De La O, Dean Karlan and Leonard Wantchekon. 2015. “Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice and Party Identification.” *Journal of Politics* 77(1):55–71.
- Cruz, Cesi, Julien Labonne and Pablo Querubin. 2019. “Social Network Structures and the Politics of Public Goods Provision: Evidence from the Philippines.” *American Political Science Review* forthcoming.
- Cruz, Cesi, Philip Keefer, Julien Labonne and Francesco Trebbi. 2019. “Making Policies Matter: Voter Responses to Campaign Promises.” Working paper, <https://faculty.arts.ubc.ca/ftrebbi/research/cklt.pdf>.
- Dal Bo, Ernesto and Martin A. Rossi. 2011. “Term Length and the Effort of Politicians.” *Review of Economic Studies* 78(4):1237–1263.
- Dell, Melissa. 2015. “Trafficking Networks and the Mexican Drug War.” *American Economic Review* 105(6):1738–79.
- Díaz-Cayeros, Alberto, Federico Estévez and Beatriz Magaloni. 2016. *The Political Logic of Poverty Relief: Electoral Strategies and Social Policy in Mexico*. Cambridge University Press.

- Diaz-Cayeros, Alberto, José Antonio González and Fernando Rojas. 2006. Mexico's Decentralization at a Crossroads. In *Federal and Economic Reform: International Perspectives*, ed. Jessica S. Wallack and T. N. Srinivasan. Cambridge University Press pp. 364—406.
- Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, Gareth Nellis, Claire L. Adida, Eric Arias, Clara Bicalho, Taylor C. Boas, Mark T. Buntaine, Simon Chauchard, Anirvan Chowdhury, Jessica Gottlieb, F. Daniel Hidalgo, Marcus Holmlund, Ryan Jablonski, Eric Kramon, Horacio Larreguy, Malte Lierl, John Marshall, Gwyneth McClendon, Marcus A. Melo, Daniel L. Nielson, Paula M. Pickering, Melina R. Platas, Pablo Querubín, Pia Raffler and Neelanjan Sircar. 2019. "Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials." *Science Advances* 5(7).
- Enríquez, José Ramón, Horacio Larreguy, John Marshall and Alberto Simpser. 2020. "Online Political Information: Facebook Ads, Electoral Saturation, and Electoral Accountability in Mexico." Working paper, https://scholar.harvard.edu/files/jmarshall/files/mexico_saturation_intervention_v5.2.pdf.
- Fearon, James D. 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In *Democracy, Accountability, and Representation*, ed. Adam Przeworski, Susan Stokes and Bernard Manin. Cambridge University Press.
- Ferejohn, John. 1986. "Incumbent performance and electoral control." *Public Choice* 50(1):5–25.
- Ferraz, Claudio and Frederico Finan. 2008. "Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes." *Quarterly Journal of Economics* 123(2):703–745.
- Habyarimana, James, Macartan Humphries, Daniel N. Posner and Jeremy M. Weinstein. 2007.

- “Why Does Ethnic Diversity Undermine Public Goods Provision?” *American Political Science Review* 101(4):709–725.
- Harding, Robin. 2015. “Attribution and Accountability: Voting for Roads in Ghana.” *World Politics* 67(4):656–689.
- Harding, Robin and David Stasavage. 2013. “What Democracy Does (and Doesn’t) Do for Basic Services: School Fees, School Quality, and African Elections.” *Journal of Politics* 76(1):229–245.
- Keefer, Phillip. 2007. Seeing and Believing: Political Obstacles to Better Service Delivery. In *The Politics of Service Delivery in Democracies: Better Access for the Poor*, ed. Shantayanan Devarajan and Ingrid Widlund. Stockholm: Expert Group on Development Issues pp. 42–55.
- Khemani, Stuti, Ernesto Dal Bó, Claudio Ferraz, Frederico S. Finan, Johnson Stephenson, Louise Corinne, Adesinaola M. Odugbemi, Dikshya Thapa and Scott D. Abrahams. 2016. “Making Politics Work for Development: Harnessing Transparency and Citizen Engagement.” World Bank Policy Research Report 106337.
- Langston, Joy. 2003. “Rising from the ashes? Reorganizing and unifying the PRI’s state party organizations after electoral defeat.” *Comparative Political Studies* 36(3):293–318.
- Larreguy, Horacio, John Marshall and Jr. Snyder, James M. forthcoming. “Publicizing malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico.” *Economic Journal* .
- Mani, Anandi and Sharun Mukand. 2007. “Democracy, Visibility and Public Good Provision.” *Journal of Development Economics* 83(2):506–529.
- Marshall, John. 2019. “Tuning in, voting out: News consumption cycles, homicides, and electoral

- accountability in Mexico.” Working paper, https://scholar.harvard.edu/files/jmarshall/files/tuning_in_voting_out_v2_2.pdf.
- Marx, Benjamin. 2018. “Elections as Incentives: Project Completion and Visibility in African Politics.” Working Paper, <https://drive.google.com/file/d/1aSDEhq-ZG1UcX5QxWbb-0bAcAdSUXlKc/view>.
- Pande, Rohini. 2011. “Can informed voters enforce better governance? Experiments in low-income democracies.” *Annual Review of Economics* 3(1):215–237.
- Pérez Archudia, Eduardo and Ernesto de Jesús Arenas Aréchiga. 2012. “Agenda desde lo Local: Desarrollo Sostenible y Desarrollo Humano.” *Revista Austral de Ciencias Sociales* 2:43–54.
- Platas, Melina and Pia Raffler. 2020. “Closing the Gap: Information and Mass Support in a Dominant Party Regime.” *Journal of Politics* forthcoming.
- Pulido-Gómez, Amalia. 2018. “Drug-related Violence and Party Behavior: The Case of Candidate Selection in Mexico.” Ph.D. dissertation, University of North Texas.
- Raffler, Pia. 2019. “Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy.” Working Paper, http://piaraffler.com/wp-content/uploads/2019/06/Raffler_Political-Oversight_June2019.pdf.
- Tavits, Margit. 2007. “Clarity of Responsibility and Corruption.” *American Journal of Political Science* 51(1):218–229.
- Trejo, Guillermo and Sandra Ley. 2019. “Why Did Drug Cartels Go to War in Mexico? Subnational Party Alternation, the Breakdown of Criminal Protection, and the Onset of Large-Scale Violence.” *Comparative Political Studies* 7(51).

Turrubiates Flores, Héctor Omar, Marco Iván Vargas Cuéllar and Irma Brígida Suárez Rodríguez. 2014. “Capacidades Institucionales y Problemas Recurrentes: Evidencias de las Verificaciones del Programa Agenda para el Desarrollo Municipal en San Luis Potosí.” *Tlatemoani* .

Wellenstein, Anna, Angélica Núñez and Luis Andrés. 2006. Social Infrastructure: Fondo de Aportaciones para la Infraestructura Social (FAIS). In *Decentralized service delivery for the poor, Volume II: Background papers*, ed. The World Bank. Mexico City: The World Bank pp. 167–222.

A Online appendix

Contents

A.1 Proofs	A2
A.2 Additional tables	A7

A.1 Proofs

Proof of Proposition 1

Let $\gamma_A > \gamma_B$ and denote $\gamma^* := \frac{1}{\beta}$. We establish equilibrium existence by demonstrating that the following set of strategies and beliefs constitute a sequentially rational equilibrium in the absence of a certification program ($p = 0$).

Incumbent politicians. For each service i , the incumbent politician's strategies are the following functions of γ_i and τ_i :

$$(g_{i,1}^*, G_{i,1}^*, r_{i,1}^*) = \begin{cases} (1, 0, 0) & \text{if } \gamma_i < \gamma^*, \\ (0, 1, 0) & \text{if } \gamma_i \geq \gamma^*, \tau_i = h, \\ (0, 0, 1) & \text{if } \gamma_i \geq \gamma^*, \tau_i = c \end{cases}$$

and

$$(g_{i,2}^*, G_{i,2}^*, r_{i,2}^*) = \begin{cases} (0, 1, 0) & \text{if } \tau_i = h, \\ (0, 0, 1) & \text{if } \tau_i = c \end{cases}.$$

Voter. The voter's posterior beliefs that the incumbent is $\tau_i = h$ type in producing service i are as follows:

$$\tilde{\gamma}_i(g_{i,1} = 0) = \begin{cases} 0 & \text{if } \gamma_i < \gamma^*, \\ \gamma_i & \text{if } \gamma_i \geq \gamma^* \end{cases}$$

and

$$\tilde{\gamma}_i(g_{i,1} = 1) = \begin{cases} \gamma_i & \text{if } \gamma_i < \gamma^*, \\ 0 & \text{if } \gamma_i \geq \gamma^*, \end{cases}$$

where $\tilde{\gamma}_i = 0$ corresponds to an off-equilibrium belief, where we suppose that voters believe that the incumbent is a $\tau_i = c$ type in the case of a deviation from the equilibrium strategy. Given that only $\tau_i = c$ types can profitably deviate from playing $g_{i,1} = 1$, since $\tau_i = h$ types do not experience utility from policy and cannot engage in corruption, such a restriction on beliefs satisfies

the intuitive criterion.

Given (γ_A, γ_B) and upon observing $(g_{A,1}, g_{B,1})$, the voter's reelection rule, $v(g_{A,1}, g_{B,1}) \in \{0, 1\}$, mechanically follows from the voter's posterior beliefs as follows:

$$v(g_{A,1} = 0, g_{B,1} = 0) = \begin{cases} 1 & \text{if } \gamma_B \geq \gamma^* \\ 0 & \text{if } \gamma_B < \gamma^*, \end{cases}$$

$$v(1, 0) = 0,$$

$$v(g_{A,1} = 0, g_{B,1} = 1) = \begin{cases} 0 & \text{if } \gamma_B \geq \gamma^* \\ 1 & \text{if } \gamma^* \in [\gamma_B, \gamma_A) \\ 0 & \text{if } \gamma_A < \gamma^* \end{cases}$$

and

$$v(g_{A,1} = 1, g_{B,1} = 1) = \begin{cases} 0 & \text{if } \gamma_A \geq \gamma^*, \\ 1 & \text{if } \gamma_A < \gamma^*, \end{cases}$$

where any $v = 0$ corresponds to an off-equilibrium strategy.

It is straightforward to verify that the the voter's strategies are optimal given the voter's updated beliefs and that the voter's updated beliefs are confirmed on the equilibrium path. For incumbents that are clean in producing service i , there is no profitable deviation from $G_{i,1}^h = 1$ or from $g_{i,1}^h = 0$ in such an equilibrium because the incumbent loses the election. For incumbents that are corruptible in producing service i , there is obviously no profitable deviation from their first best outcome of $r_{i,1}^c = 1$ in such an equilibrium; in an equilibrium where $g_{i,1}^c = 0$, corruptible types do not deviate because they lose the election. (Note also that c types would always have an incentive to deviate in a separating equilibrium where $g_{i,1}^h \neq g_{i,2}^c$.) As a result, the set of strategies and beliefs constitute a sequentially rational equilibrium, where the voter prefers the equilibrium where $g_{i,1} = 1$ when $\gamma_i < \gamma^*$.

Proof of Proposition 2

Assume that $\gamma^* \in [\gamma_B, \gamma_A)$, and denote $\gamma^{**}(\rho) := \frac{1-\rho\beta}{\beta(1-\rho)} < \gamma^*$. We again establish equilibrium existence by demonstrating that the following set of strategies and beliefs constitute a sequentially rational equilibrium in the presence of a certification program ($p = 1$).

Incumbent politicians. For each service i , the incumbent politician's strategies are the following functions of γ_i , τ_i , ρ , and α :

$$(g_{i,1}^*, G_{i,1}^*, r_{i,1}^*) = \begin{cases} (1, 0, 0) & \text{if } \gamma_i < \gamma^{**}(\rho), \\ (0, 1, 0) & \text{if } \gamma_i \geq \gamma^{**}(\rho) \text{ and either } \tau_i = h \text{ or } \alpha = H \text{ and } \tau_i = c, \\ (0, 0, 1) & \text{if } \alpha = C \text{ and } \tau_i = c, \end{cases}$$

and

$$(g_{i,2}^*, G_{i,2}^*, r_{i,2}^*) = \begin{cases} (0, 1, 0) & \text{if } \tau_i = h, \\ (0, 0, 1) & \text{if } \tau_i = c. \end{cases}$$

Third-party certifier. For each service i , and given $(g_{i,1}, G_{i,1}, r_{i,1})$ and α , the third-party certifier's strategies are a mechanical function of the certifier's type:

$$c_i^*(g_{i,1} = 1, G_{i,1} = 0, r_{i,1} = 0) = \phi,$$

$$c_i^*(g_{i,1} = 0, G_{i,1} = 1, r_{i,1} = 0) = 1,$$

and

$$c_i^*(g_{i,1} = 0, G_{i,1} = 0, r_{i,1} = 1) = \begin{cases} 0 & \text{if } \alpha = H \\ 1 & \text{if } \alpha = C. \end{cases}$$

Voter. Upon observing $g_{i,1}$ and c_i , the voter's posterior beliefs that the incumbent is $\tau_i = h$ in

producing service i are as follows:

$$\tilde{\gamma}_A (g_{A,1} = 1, c_i = \phi) = 0, \tilde{\gamma}_A (g_{A,1} = 0, c_i = 0) = 0, \text{ and } \tilde{\gamma}_A (g_{A,1} = 0, c_i = 1) = \gamma_A,$$

and

$$\tilde{\gamma}_B (g_{B,1} = 1, c_i = \phi) = \begin{cases} \gamma_B & \gamma_B < \gamma^{**}(\rho), \\ 0 & \gamma_B \geq \gamma^{**}(\rho), \end{cases}$$

$$\tilde{\gamma}_B (g_{B,1} = 0, c_i = 0) = 0,$$

and

$$\tilde{\gamma}_B (g_{B,1} = 0, c_i = 1) = \begin{cases} 0 & \gamma_B < \gamma^{**}(\rho), \\ \gamma_B & \gamma_B \geq \gamma^{**}(\rho). \end{cases}$$

where $\tilde{\gamma}_i = 0$ corresponds to an off-equilibrium belief. Given that only $\tau_i = c$ types can profitably deviate from playing $g_{i,1} = 1$, since $\tau_i = h$ types do not experience utility from policy and cannot engage in corruption, such a restriction on beliefs satisfies the intuitive criterion. f

Upon observing $g_{A,1}$, $g_{B,1}$, c_A , and c_B , the voter's reelection rule, $v(g_{A,1}, g_{B,1}, c_A, c_B)$, is as follows:

$$v(0, 0, 1, 0) = v(0, 0, 0, 1) = v(0, 0, 0, 0) = 0,$$

$$v(0, 0, 1, 1) = \begin{cases} 0 & \gamma_B < \gamma^{**}(\rho), \\ 1 & \gamma_B \geq \gamma^{**}(\rho), \end{cases}$$

$$v(1, 0, \phi, 0) = v(1, 0, \phi, 1) = v(0, 1, 0, \phi) = 0,$$

$$v(0, 1, 1, \phi) = \begin{cases} 1 & \gamma_B < \gamma^{**}(\rho), \\ 0 & \gamma_B \geq \gamma^{**}(\rho), \end{cases}$$

$$v(1, 1, \phi, \phi) = 0.$$

where any $v = 0$ corresponds to an off-equilibrium strategy.

Following the same logic as the proof of the previous proposition, it is straightforward to verify that the incumbent politician and the voter's strategies are optimal, given the third-party verifier's strategies and the voter's updated beliefs, and that the voter's updated beliefs are confirmed on the equilibrium path. As a result, the set of strategies and beliefs constitute a sequentially rational equilibrium, where the voter prefers the equilibrium where $g_{i,1} = 1$ when $\gamma_i < \gamma^{**}$.

Proof of Proposition 3

Follows directly from Proposition 2 when $\gamma^* \in [\gamma_B, \gamma_A)$.

Proof of Proposition 4

Assuming that $\gamma^* \in [\gamma_B, \gamma_A)$ and using the results from Propositions 1 and 2, the expected effects of the certification program on equilibrium utility of each service i are then given by:

$$\begin{aligned} \Delta_A &:= \mathbb{E} \left[\sum_{t=1,2} U(g_{A,t}, G_{A,t}) \middle| p = 1 \right] - \mathbb{E} \left[\sum_{t=1,2} U(g_{A,t}, G_{A,t}) \middle| p = 0 \right] \\ &= (1 - \gamma_A) \rho \beta, \\ \Delta_B &:= \mathbb{E} \left[\sum_{t=1,2} U(g_{B,t}, G_{B,t}) \middle| p = 1 \right] - \mathbb{E} \left[\sum_{t=1,2} U(g_{B,t}, G_{B,t}) \middle| p = 0 \right] \\ &= \begin{cases} 0 & \text{if } \gamma_B < \gamma^{**} \\ (\gamma_B + (1 - \gamma_B) \rho) \beta - 1 & \text{if } \gamma_B \geq \gamma^{**}, \end{cases} \end{aligned}$$

and are weakly positive because $\beta > 0$, $\rho \geq 0$, and $1 - \gamma_A \geq 0$ and evidently increasing in ρ when $\gamma_i \geq \gamma^{**}$ and invariant to ρ when $\gamma_i < \gamma^{**}$.

A.2 Additional tables

Table A1: Robustness of baseline certification interaction operationalization

	(1)	(2)
	Z-score	Z-score
Program	0.100*	0.091
	(0.059)	(0.056)
Program × Yellow baseline	-0.121	
	(0.095)	
Program × Red baseline	-0.125**	
	(0.061)	
Program × Baseline (ordinal)		-0.058**
		(0.029)
Observations	89,646	89,646
R^2	0.43	0.43
Mean of Outcome	0	0
SD of Outcome	1	1
Min of Outcome	-9.43	-9.43
Max of Outcome	8.71	8.71

Notes: This table shows the regression results of the main specification using program indexes overall public service delivery as outcomes. An observation represents a program index in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-index and municipality-index fixed effects. Column (1) interacts *Program* with a non-parametric version of the certified status of the index in the first year the municipality participated in the program, column (2) uses a linear version instead. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Regression results using state capacity as outcomes and different coding

Panel A: No Interaction	Dependent variable: Z-score				
	(1) Index	(2) Dummy 1st cutoff	(3) Dummy 2nd cutoff	(4) Winsorized 99	(5) Winsorized 95
Program	-0.0016 (0.0236)	-0.0158 (0.0237)	0.0085 (0.0232)	-0.0095 (0.0240)	-0.0096 (0.0240)
Observations	90,442	90,459	90,551	90,484	90,484
R ²	0.41	0.40	0.41	0.43	0.43
Panel B: Interaction with Alignment	Dependent variable: Z-score				
	(1) Index	(2) Dummy 1st cutoff	(3) Dummy 2nd cutoff	(4) Winsorized 99	(5) Winsorized 95
Program	0.053 (0.037)	0.033 (0.037)	0.068 * (0.037)	0.040 (0.037)	0.041 (0.037)
Program × Aligned with State	-0.120** (0.051)	-0.109** (0.052)	-0.125** (0.051)	-0.103** (0.052)	-0.105** (0.052)
Observations	87,329	87,347	87,443	87,373	87,373
R ²	0.50	0.50	0.50	0.52	0.53
Panel C: Interaction with Low Baseline	Dependent variable: Z-score				
	(1) Index	(2) Dummy 1st cutoff	(3) Dummy 2nd cutoff	(4) Winsorized 99	(5) Winsorized 95
Program	0.100* (0.059)	0.098 (0.061)	0.108* (0.056)	0.111* (0.063)	0.112* (0.063)
Program × Low baseline	-0.124** (0.061)	-0.137** (0.064)	-0.122** (0.058)	-0.146** (0.065)	-0.148** (0.065)
Observations	89,646	89,663	89,756	89,688	89,688
R ²	0.43	0.42	0.43	0.45	0.45
Mean of Outcome	0	0	0	0	0
SD of Outcome	1	1	1	1	1
Min of Outcome	-9.43	-11.48	-9.43	-9.43	-9.43
Max of Outcome	8.71	8.19	9.56	8.32	8.35

Notes: This table shows the main regression results investigating the effect of the certification program on public service delivery outcomes while using different coding strategies of the outcome variables. An observation represents a program index in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-indicator and municipality-index fixed effects. Panel (A) uses the main specification. Panel (B) interacts with whether the municipality is governed by the same party as its state. Panel (C) interacts with whether the municipality received a low certification on a given index in the year it entered the program. Column (1) uses indexes of the corresponding certification status in each of the subindexes following the program specifications when possible. Column (2) uses an indicator of an index for whether at least the first cutoff (from red to yellow) was passed for each of the subindexes. Column (3) uses an indicator of an index for whether the second cutoff (from yellow to green) was passed for each of the subindexes. Columns (4) instead uses the raw data used to code up the certification status in each of the subindexes winsorized at the 99th percentile. Column (5) winsorizes at the 95th percentile instead. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Linking indexes to outcomes

Index	Subindex	Data source	Coding
1.1 Administered responsibly	Own income / Current expenditure	Municipal Budget	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Current expenditure / Total expenses	Municipal Budget	0 if >70%, 1 if >50% and <70%, 2 if <50%
	Public investment / Total income	Municipal Budget	0 if <25%, 1 if >25% and <50%, 2 if >50%
	Personal services / Current expenditure	Municipal Budget	0 if >70%, 1 if >50% and <70%, 2 if <50%
	Total expenses + Total revenue	Municipal Budget	Winsorized at 99%
1.4 Participatory	Whether the administrative structure has a sector for promoting social participation	Municipal Census	Standardized
	Whether the commissions and/or communal committees participate in the allocation	Municipal Census	Standardized
	Index for regulations for participation	Municipal Census	0 if no reg. 1 if updated 3-5 years ago, 2 if updated in last 3 years
	Are there mechanisms for citizens participation?	Municipal Census	Standardized

Linking indexes to outcomes

Index	Sub-index	Data source	Coding
1.6 Leader in civil protection	Is there a plan for civil protection?	Municipal Census	Standardized
	Is there a map of risk zones?	Municipal Census	Standardized
	Index for regulations on civil protection	Municipal Census	0 if no regulations, 1 if updated 3-5 years ago, 2 if updated in last 3 years
1.8 Legally ordered	What percentage of transport, police, markets, graves, public works, cleaning, butchers, participation, civil protection, cadastre, fire, zoning have regulations	Municipal Census	Standardized
1.10 Transparent	Is there a institution in the municipality responsible for transparency?	Municipal Census	Standardized
	Does the municipality currently have regulations to regulate access to public information?	Municipal Census	Standardized
	Does the municipality currently have a public servant responsible for dealing with requests for public information in each of the institutions?	Municipal Census	Standardized
	A system of reception of and attention to public information requests	Municipal Census	Standardized
	A system or procedures of organization, protection, and maintenance of archives	Municipal Census	Standardized
	Training program for public servants on the rights and obligations of access to public information	Municipal Census	Standardized
	Is there open access?	Municipal Census	Standardized
	Are there regulations about transparency?	Municipal Census	Standardized
1.11 Healthy finances	Debt accumulated from previous years	Municipal Census	Inversed, Logged and Standardized
	Percentage of budgeted contributions collected	Municipal Census	Standardized
	Federal transfers / Total income	Municipal Budget	0 if >95%, 1 if <95% and >75%, 2 if <75%

Linking indexes to outcomes

Index	Sub-index	Data source	Coding
3.1 Provider of public services	Percentage of municipal capital covered by drainage and sewage system	Municipal Census	Standardized
	Percentage of rest of municipality covered by drainage and sewage system	Municipal Census	Standardized
	Percentage of municipal capital covered by public lighting	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of rest of municipality covered by public lighting	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of municipal capital covered by street cleaning	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of rest of municipality covered by street cleaning	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of municipal capital covered by trash collection	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of rest of municipality covered by trash collection	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Are there grave regulations?	Municipal Census	Standardized
	Are there market regulations?	Municipal Census	Standardized
3.2 Promoter of sport and recreation	Do regulations on performance and sport exist?	Municipal Census	Standardized
	Index for regulations on performance and sport	Municipal Census	0 if no regulations, 1 if updated 3-5 years ago, 2 if updated in last 3 years
3.6 Healthy Municipality	Percentage of municipal capital covered by drinking water	Municipal Census	0 if 0%, 1 if >0% and <100%, 2 if 100%
	Percentage of rest of municipality covered by drinking water	Municipal Census	0 if 0%, 1 if >0% and <100%, 2 if 100%
	Percentage of municipal capital covered by drainage and sewage system	Municipal Census	0 if 0%, 1 if >0% and <90%, 2 if >90%
	Percentage of rest of municipality covered by drainage and sewage system	Municipal Census	0 if 0%, 1 if >0% and <90%, 2 if >90%
3.8 With decent housing	Percentage of municipal capital covered by drinking water	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of rest of municipality covered by drinking water	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of municipal capital covered by drainage and sewage system	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Percentage of rest of municipality covered by drainage and sewage system	Municipal Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Occupants in homes with drainage and / or toilet	Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Occupants in homes with electric power	Census	0 if <50%, 1 if >50% and <75%, 2 if >75%
	Occupants in houses with dirt floor	Census	0 if >30%, 1 if >14% and <30%, 2 if <14%
	Housing without overcrowding	Census	Standardized