How unobserved investments inhibit public service delivery improvement in corrupt environments *

Soeren J. Henn † Horacio Larreguy ‡ John Marshall §

June 2024

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 politicians to make such investments. While audits that certify investments can mitigate this monitoring problem, the certification process's effectiveness can be undermined by corruptible 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 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.

*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.

†Newcastle University Business School. soeren.henn@gmail.com.

‡Department of Economics and Political Science, ITAM. horacio.larreguy@itam.mx.

§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 improve the state's capacity to provide services instead of misappropriating funds (Barro 1973; Fearon 1999; Ferejohn 1986). Unfortunately, there remains abundant evidence of limited or low-quality local public service delivery across democracies in the Global South (Habyarimana et al. 2007; Keefer 2007; Khemani et al. 2016). An active recent literature has examined whether voter information about incumbent performance can support electoral accountability by helping voters to select competent and honest politicians (Dunning et al. 2019; Pande 2011). 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 multi-tasking theory to highlight how, even for the best-intentioned incumbent parties, the short-term unobservability of investments by local governments to improve local service delivery can constrain such investments and thereby limit improvements in service delivery. A local government in our adverse selection model oversees the delivery of public services. Incumbent parties 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. Local political parties are of two types—clean or corruptible—in providing a given service, but these types are unobservable to voters. All parties care about rents from holding office, but only parties that are clean in providing a specific service share voters' preferences over, or lack the ability to extract rents from, that service. A representative rational voter decides whether to re-elect the incumbent party between periods. However, they can only observe 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 party is corruptible in the provision of a given service is sufficiently high, we show that investing in improving the delivery of that service is not feasible. Intuitively, this is because voters would infer that the incumbent party is likely to be corruptible in providing the service when they do not see the budget being spent on current service delivery. 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 high-corruption environments.

In principle, certification programs designed to publicly demonstrate that an incumbent party 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 that can certify whether 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 only negligibly improve the quality and quantity of a given public service.

We test key observable implications of the theory in the context of the From the Local Agenda (Agenda desde lo Local; ADL) program in Mexico, which followed the United Nations' Agenda 21 action plan for promoting better local governance. In this context, municipal governments often provide poor public service delivery and engage in corruption (e.g. Díaz-Cayeros, Estévez and Magaloni 2016; Diaz-Cayeros, González and Rojas 2006). The federal government first implemented the ADL program in 2004 in collaboration with state governments, which needed to enter the program before municipalities in their state could enroll. 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 municipal 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. Federal and state officials hand these certificates out in award ceremonies that are often publicized by local media and municipal governments' websites.

We estimate the effects of entering the ADL program on certification and actual service delivery outcomes using a twoway fixed effect difference-in-differences design that leverages within-state variation in when a municipality entered the program and addresses the issue of negative weights resulting from the staggered nature of the treatment (De Chaisemartin and d'Haultfoeuille, 2020). To avoid comparing municipalities that entered the ADL program with those that never did it, we focus on municipalities that entered the program between 2004 and 2013. 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. To examine the heterogeneous effects of the certification program implied by our model, we use municipal partisan alignment with the state governor to proxy for the likelihood that a certifying third party is corruptible, given strong ties between copartisan levels of government and the history of corruption in institutions of higher education that depend on state governments for resources. 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 party is corruptible in producing a given service.

The results show that municipalities were frequently and quickly awarded green statuses for certified improvements in local service delivery after entering the ADL program. However, data measured independently of the program largely support the model's more cynical empirical predictions. We find a positive effect of the certification program on public service delivery outcomes

on average. However, in line with our model, the results further indicate that the certification program only increased public service delivery in municipalities that were not aligned with the state governor and for services where baseline certification statuses were not low. These heterogeneous effects suggest that a significant number of corrupt incumbent governments used the ADL program as a shield to facilitate greater rent seeking.

Our theory and evidence make several main contributions. First, we extend multi-tasking models where politicians allocate effort between tasks whose outcomes are observed with more or less noise in a given period (e.g. Ashworth 2005; Mani and Mukand 2007; Marx 2018) to consider investments that only become visible after elections. We show that even well-intentioned politicians may only be induced to make efficient forward-looking investments when credible investment monitoring exists. By demonstrating this mechanism empirically, our findings complement extant studies showing that incumbents follow incentives to deliver services most clearly attributable to their actions (Harding 2015; Marx 2018), engage in corruption when public goods are insufficient visible to voters (Tavits 2005), and neglect investments that will not mature until after elections (Dal Bo and Rossi 2011). However, like Banerjee, Duflo and Glennerster (2008) and Raffler (forthcoming), we also highlight the difficulty of designing incentive structures to prevent manipulation or corruption of monitoring devices—in our case, of a political form.

Second, our theory adds nuance to when voter information improves public service delivery. While easily-collected indexes of performance can improve electoral selection (Enríquez et al. 2022; Ferraz and Finan 2008; Larreguy, Marshall and Snyder 2020) or reduce moral hazard (Avis, Ferraz and Finan 2018; Besley and Burgess 2002; Grossman and Michelitch 2018) when disseminated by mass media, we highlight the importance of *which* indexes are measured by illustrating how the short-term unobservability of investments to improve service delivery can reduce the incentive for clean politicians to make such investments. In this respect, our argument aligns with Gailmard and Patty's (2019) rational explanation for suboptimal investment in disaster prevention. Their model similarly shows that voter uncertainty about the benefits of particular disaster preven-

tion technologies causes voters to suspect politicians investing in disaster prevention of corrupt; honest politicians seeking re-election thus demonstrate their honesty by providing observable but inefficient relief only after a disaster.

Third, our identification strategy also provides a compelling empirical foundation for prior studies also documenting a limited correlation between ADL program assessments and actual municipal development outcomes in Mexico (Pérez Archudia and Arenas Aréchiga 2012). In addition to testing a key mechanism driving limited public service delivery more broadly across developing democracies, our findings thus show that an internationally-promoted program fails to improve the service delivery of local governments in an important developing context.

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 government public services. We first lay out a stylized dynamic model where a local government produces two services.[1] 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 in terms of certification and service delivery outcomes to generate testable implications of our underlying theory.

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 party (or politician, where re-election is feasible) is granted a unit of budget to produce the separable public services A and B. For each service $i \in \{A, B\}$, the

[1] We focus on two services to highlight differences across services over which expectations of corruption differ, although the same logic applies to one or many services.

incumbent party can allocate the budget in one of three ways: current service delivery $g_{i,t} \in \{0,1\}$, investment in future service delivery $G_{i,t} \in \{0,1\}$, or misappropriation $r_{i,t} \in \{0,1\}$.

A representative voter decides whether to re-elect the incumbent party or select an alternative candidate at the end of the first period $v \in \{0,1\}$. The voter can only observe whether the incumbent party 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 budget for service i ($r_{i,1} = 1$) or invested in future service delivery ($G_{i,1} = 1$).

The voter's payoff derives 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:

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

where $\beta > 1$ captures our assumption that investments in future public service delivery yield higher utility for the voter than current service delivery. 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 could may be extended to encompass more citizens.

All political parties receive per-period rents $R > 0$ when in office. However, parties that are corruptible when providing a specific service can also extract additional rents by misappropriating funds. With probability $\gamma_i \in [0,1]$, the incumbent party is not corruptible when producing service i ($\tau_i = h$); with probability $1 - \gamma_i$, the incumbent party can engage in corruption ($\tau_i = c$) when allocating the budget for service i. Clean parties politicians can be thought of as lacking the ability to misappropriate the budget for service i.[2]

An incumbent party 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 ser-

[2]Qualitatively similar results obtain if clean parties share the voter's preferences for service i.

vice 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 party 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 party's type (τ_A, τ_B) is known only to the politician. Without loss of generality, we assume that $\gamma_A > \gamma_B$. For 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 draws the incumbent's type, $\tau_i \in \{h, c\}$, for each service $i \in \{A, B\}$, which is revealed only to the incumbent.

2. At the beginning of period $t = 1$, the incumbent party of type (τ_A, τ_B) selects their policy $\left(g_{i,1}^{\tau_i}, G_{i,1}^{\tau_i}, r_{i,1}^{\tau_i}\right)$ for each service.

3. The voter observes $(g_{A,1}, g_{B,1})$ and decides whether to re-elect the incumbent party, $v \in \{0, 1\}$.

4. If the incumbent party is not re-elected, nature draws the victorious challenger party's type.

5. At the beginning of period $t = 2$, the (possibly new) incumbent party selects $\left(g_{i,2}^{\tau_i}, G_{i,2}^{\tau_i}, r_{i,2}^{\tau_i}\right)$ 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).[3] We start by characterizing this

[3]Two classes of equilibria always exist: (i) where the voter only re-elects the incumbent party 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 party 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.

equilibrium in the absence of a certification program ($p = 0$):

Proposition 1. *Denote $\gamma^* := \frac{1}{\beta}$. Then, for each service i:*

- *If $\gamma_i \geq \gamma^*$, only incumbent parties that are clean in producing service i invest in its future delivery: clean incumbents regarding service i choose the policy vectors $\left(g^h_{i,t}, G^h_{i,t}, r^h_{i,t}\right) = (0, 1, 0)$ in each period $t \in \{1, 2\}$, while corruptible incumbents regarding service i choose the policy vectors $\left(g^c_{i,t}, G^c_{i,t}, r^c_{i,t}\right) = (0, 0, 1)$ in each period $t \in \{1, 2\}$.*

- *If $\gamma_i < \gamma^*$, incumbent parties that are clean and corruptible 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 $\left(g^h_{i,2}, G^h_{i,2}, r^h_{i,2}\right) = (0, 1, 0)$ and $\left(g^c_{i,2}, G^c_{i,2}, r^c_{i,2}\right) = (0, 0, 1)$ in period $t = 2$.*

The voter's expected utility derived from public provision of service i is then given by:

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

Proof. See Appendix section A.1 for all proofs, where we also define the off-equilibrium strategies and beliefs of voters, political parties, and third-party verifiers that support this sequentially rational equilibrium. □

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. Voters induce such behavior by their incumbent party 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 corruptible is low. Otherwise, the voter only re-elects an incumbent party that allocates the budget for service i into current public service delivery, inducing both types of politicians not to make investments in making future services more efficient.

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 party could engage in corruption is sufficiently high. This entails assuming that $\gamma^* \in [\gamma_B, \gamma_A)$. 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. 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 only to the third party and the municipal incumbent party. 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.[4] To restrict attention to election motives, we assume that the incumbent party only incurs electoral costs when corruption is 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:

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

- *If $\gamma_B \geq \gamma^{**}$, incumbents parties that are clean and corruptible initially pool to invest in*

[4]For 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.

future delivery of both services when the third party is honest: for each service i, incumbent parties that are clean in producing that service choose $\left(g^h_{i,t}, G^h_{i,t}, r^h_{i,t}\right) = (0,1,0)$ in each period $t \in \{1,2\}$ and incumbent parties that are corruptible in producing that service choose $\left(g^c_{i,1}, G^c_{i,1}, r^c_{i,1}\right) = (0,1,0)$ in period $t = 1$ if the third-party is honest and $\left(g^c_{i,1}, G^c_{i,1}, r^c_{i,1}\right) = (0,0,1)$ if the third party is corruptible and chose $\left(g^c_{i,2}, G^c_{i,2}, r^c_{i,2}\right) = (0,0,1)$; each type of third party always reports $c_A = c_B = 1$.

- *If $\gamma_B < \gamma^{**}$, incumbent parties that are clean and corruptible 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 incumbent parties that are clean in producing service B choose $\left(g^h_{B,2}, G^h_{B,2}, r^h_{B,2}\right) = (0,1,0)$ and incumbent parties that are corruptible in producing service B choose $\left(g^c_{B,2}, G^c_{B,2}, r^c_{B,2}\right) = (0,0,1)$ in period $t = 2$, and each type of third party always reports $c_A = 1$.*

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

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

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

Proposition 2 shows that, when the probabilities that the third party is honest *and* the incumbent party is clean with respect to service i are sufficiently high, the representative 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 incumbent parties that are corruptible in producing service A to invest in the first period. This effect arises because third parties are effective monitors when they

are honest. Second, and more interestingly, the certification program induces clean incumbent parties to invest in future service delivery of service B. Intuitively, when $\gamma_B \geq \gamma^{**}$, this occurs because there is a sufficiently large probability that an honest third party will prevent corruptible incumbents from engaging in 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 political parties in the production of service B, and both incumbent party 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.[5] First, the following corollary establishes that the outcomes of certification reports by a third party pertaining to investments in future service delivery are independent of whether the certifying third party is corruptible:

Corollary 1. *Assume that $\gamma^* \in [\gamma_B, \gamma_A)$. Then:*

- *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.*

Under the certification program, this result unsurprisingly highlights that, whenever third parties have to certify future service delivery, corruptible third parties always certify that the incum-

[5] As we show below, the voter (or a social planner) would always want to implement the program, although some corruptible incumbent parties would prefer the program not to be implemented.

bent party invested in future delivery of each service i—regardless of whether they invested or not. 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 corollary assesses 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 political parties that are clean or corruptible in providing the service:

Corollary 2. *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:*

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

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

Provided that the share of political parties that are clean in producing service i is not too low, the certification program thus leads to an expected increase in total service delivery across periods. When $\gamma_i < \gamma^{**}$, there are too few clean incumbent parties in producing service i for the program to permit clean incumbents to invest in future delivery of that service and thereby induce corruptible

incumbents 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 because investments in future service delivery become feasible for clean incumbents and are effectively forced upon corruptible incumbents by honest third parties. When $\gamma_i \geq \gamma^*$, investments were already feasible absent the program, but the certification program enables honest third parties to enforce investments on corruptible incumbent parties 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 for two reasons.[6] First, incumbent parties that are corruptible in producing a given service can claim that they are investing in future service delivery while actually misappropriating public funds. Absent the certification program, these types of 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 incumbent parties into investing in future service delivery when $\gamma_i \geq \gamma^{**}$.

3 Empirical context

3.1 Mexican municipal governments and their limited institutional capacity

Mexico contains around 2,500 municipalities governed by mayors. Until a recent reform that permitted re-election in some states starting in 2018, municipal mayors were typically elected to three-year non-renewable terms. While this reform may strengthen accountability linkages, voters already held parties responsible for the performance of aligned mayors, given the importance of party labels and the role they play in candidate selection (Chong et al. 2015; Langston 2003; Larreguy, Marshall and Snyder 2020).

Following major decentralization reforms in the 1990s, municipal governments became the

[6] We 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 because a larger ρ expands the range of γ for which an equilibrium shift can be induced by the certification program.

main actors responsible for the local provision of basic infrastructure and 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. The decentralization reforms were not accompanied by a corresponding increase in tax collection responsibilities; in 2010, municipalities raised less than 20% of their revenues themselves (Castañeda and Pardinas 2012).

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).

3.2 The *From the Local Agenda* program

The *From the Local Agenda* (Agenda desde lo Local; ADL) program—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.[7] The program has been developed and implemented by Mexico's Interior Ministry 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 twenty-first century.[8] The Interior Ministry has administered the program together with, and largely through, state governments. Participation by municipal governments is voluntary, although most municipalities entered the program under state guidance.

[7] In 2014 the ADL program was revised and expanded to include the roles of municipal authorities in economic and social development, as well as environmental sustainability. See www.agendaparaeldesarrollomunicipal.gob.mx.

[8] This 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 www.sustainabledevelopment.un.org/content/documents/Agenda21.pdf.

The ADL 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, municipal governments—aided by state governments—self-assess their institutional capacity for service delivery and actual service delivery across 39 indexes, comprising 270 sub-indexes, that are grouped into four areas: (a) institutional capacity for good governance; (b) sustainable economic development; (c) inclusive social development; and (d) sustainable environmental development. For each index and sub-index, municipal governments assign themselves one of the following statuses: red (completely undesirable situation and dramatic room for improvement), yellow (some room for improvement), or green (acceptable situation). The program guidelines for each sub-index specify the quantitative indicators used to determine each status.

In the program's second stage, 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 they are largely funded by the federal and state governments. Especially when incumbent parties at the municipal and state levels are aligned, governors may seek to manipulate third-party certification to enhance the reputation of their co-partisans and that of their party more generally. Moreover, there are many instances of higher education institutions engaging in corruption. For example, a recent corruption scandal resulting in the diversion of approximately USD 400 million of public funds involved the federal government and eleven higher education institutions, 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.[9]

[9]See *New York Times*, "'El dinero se iba a un agujero negro:' el esquema de corrupción que compromete al gobierno de México," September 5th 2017 for more details.

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 government's assessment of each sub-index.[10] Where verification worked best, auditors could identify 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 other cases, verification teams 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, 863 municipalities in 30 states concluded the verification process. In total, 1,827 individuals—including faculty and students—from 163 higher education institutions verified the self diagnoses proposed by municipal governments. Out of these 163 institutions, 99 (61%) were universities, 57 (35%) were technological institutes, 5 (3%) were local colleges, and 2 (1%) were higher education institutes. The mean institution conducted slightly more than 5 verifications, while the median conducted 3.[11]

In the program's third stage, 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 measure 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, as well as representatives from higher education institutions—grants certificates to municipal governments for each index that receives a green certification. These certificates are handed out by federal

[10] See the www.gob.mx/inafed/articulos/cual-es-la-importancia-de-la-etapa-de-verificacion-del-programa-agenda-para-el-desarrollo-municipal for more details.

[11] See the program website for more details.

and state officials at award ceremonies, which are often publicized by municipal governments and local media. The ceremonies usually highlight the achievements of municipal government officials and note that the award was subjected to a third-party certification. It is also often mentioned that the municipal governments are certified using international standards.[12] This information could have important electoral consequences in a context where corruption and service delivery are salient concerns among the electorate (Chong et al. 2015; Enríquez et al. 2022; Larreguy, Marshall and Snyder 2020). Indeed, while voters are largely aware of the services currently available to them, they are poorly informed about mayoral responsibilities and budget allocations performance (Chong et al. 2015) and rely on local broadcast media to learn about less visible government actions such as mayoral malfeasance (e.g. Castañeda Sabido 2011; Larreguy, Marshall and Snyder 2020).

3.3 Mapping theory to empirics

To map the theoretical model's general implications to the specific empirical context of the ADL program, we first note two simplifying features of the model. First, while the ADL program certified indexes on a three-point scale from red to green, our model focused on whether an index is certified and thus does not distinguish improvements from red to yellow from improvements from yellow to green. However, the model can naturally be extended to match our empirical analysis treating all one-point increases on the scale equally. Second, while the utility that voters experience from current and future service delivery can be separated in the model, the indexes—and their underlying indicators—certified by the ADL program do not draw this distinction. We approximate the total utility received by voters from public services by focusing on the indexes that relate to the capacity to deliver services or metrics of actual service delivery, which broadly

[12] *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.

capture the current and future returns to investments in service delivery.

Our first hypothesis, which emerges from Corollary 1, relates to the effect of entering the ADL program on the certification status of a municipality's program indexes. Relative to the status of each index when a municipal government first entered the program (before it makes any improvements), Hypothesis 1 states 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 Corollary 2, we first hypothesize that:

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

As Corollary 2 demonstrates, the magnitude of the expected positive effect—and thus our ability to empirically detect it—depends on the probability that the mayor is corruptible in providing a specific public service: the ADL program only increases service delivery where this probability is sufficiently low.

Next, we consider how the ADL program's efficacy varies 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 is corrupt or encouraged by state officials to provide a generous certification (i.e. ρ) is whether a given municipal government is politically aligned with the state government. Numerous studies across Latin America have shown such alignment to facilitate the transfer of resources and facilitate corruption (Brollo and Nannicini 2012), while

Simpser et al. (2016) document similarly copartisan biases in a federal program in Mexico. Using this proxy, Corollary 2 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 where 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.

We further 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).[13] Our final hypothesis then follows from Corollary 2:

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 for which 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 part of that index.

4 Research design

We describe our data relating to the ADL program and service delivery, before then explaining and validating the identification strategy used to test the hypotheses just enumerated.

[13]The 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.

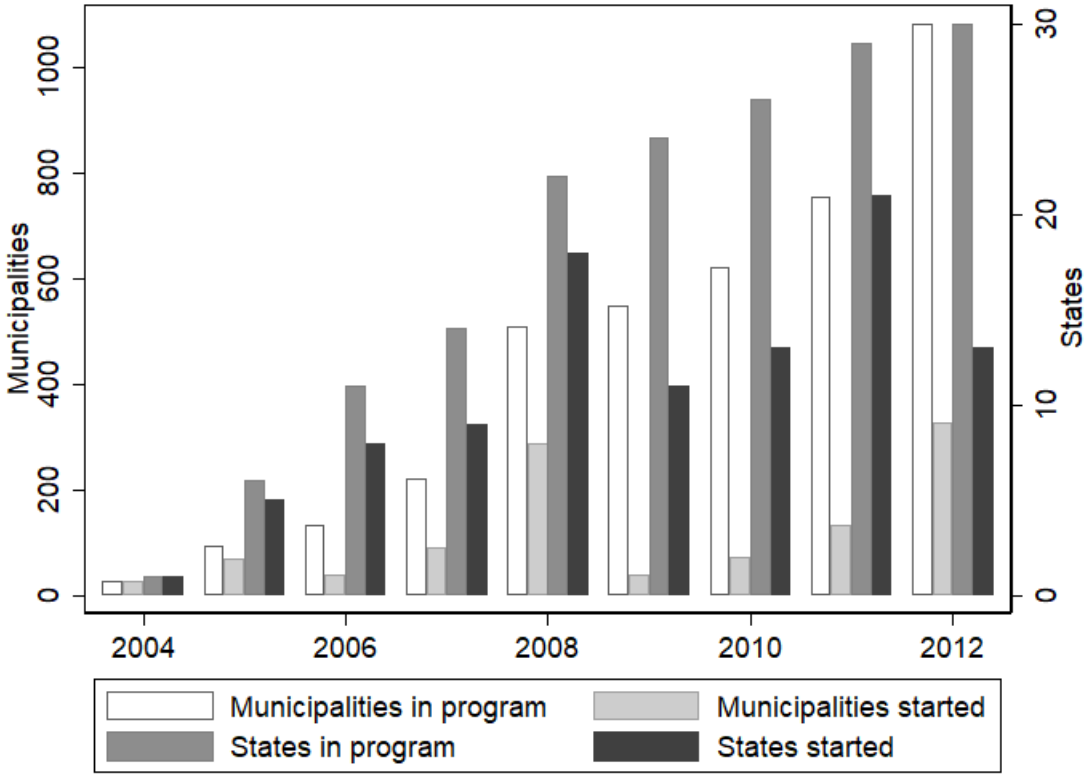
4.1 Data

Data on participation in the ADL program and the index certifications come from the National Institute for Federalism and Municipal Development (INAFED).[14] 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. Figure 1 shows the number of municipalities—and the number of distinct states in which participating municipalities were located—in the ADL program, as well as the corresponding number that entered the program in each year. Figure A6 in the Appendix report the distribution of entry date by state for all participating municipalities. INAFED also provides the certified status that every municipality received for each of the 39 program indexes while in the program.

To examine whether the certified statuses of the indexes track the quantitative criteria stipulated by the program guidelines, we independently collected the public service delivery data supposedly underlying the certifications. To do so, we collected data from three sources that are independent of the ADL program. First, our principal source is the census of municipal public service delivery conducted by the National Institute of Statistics and Geography (INEGI) in 2000, 2002, 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. Second, for the certification indexes relating to municipal finances, we use public finance data published by the INEGI between 2000 and 2013.[15] Third, we exploit information from the quinquennial population censuses conducted by the INEGI between 2000 and 2010.[16] 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. We

[14] www.gob.mx/inafed/acciones-y-programas/resultados-historicos-del-programa-agenda-desde-lo-local.
[15] See more details at www.inegi.org.mx/programas/finanzas.
[16] See more details at www.inegi.org.mx/programas/ccpv/cpvsh.

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



refer to the measures of service delivery derived from these sources, which vary in the years for which they are available, as indicators.

The indicators allow us to investigate actual changes in the quality and quantity of municipal public service delivery due to the ADL program. We restrict attention to 49 indicators that exactly or closely correspond to sub-indexes within 10 of the ADL program's 39 indexes; they are listed, together with summary statistics, in Table 1. These indicators enable us to assess whether public service delivery actually satisfied the guidelines for designating red, yellow, or green certification status for each sub-index. For example, the sub-index for federal transfers in index 1.11 takes the value 0 (red status) if more than 95% of total income comes from federal transfers, 1 (yellow) if more than 75% but less than 95% of total income comes from federal transfers, and 2 (green) if

Table 1: Summary statistics by service delivery index

Index	Variable	Mean	Std. dev.	Min	Max
1.1 Public accounts	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 Citizen participation	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 citizen participation?	0.86	0.34	0	1
1.6 Civil protection	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 Regulatory framework	Share of sectors that have regulations	0.41	0.30	0	1
	Index of regulations	0.52	0.77	0	2
1.10 Transparency	Institution responsible for transparency?	0.56	0.33	0	1
	Regulations to regulate access to public information?	0.55	0.50	0	1
	Public servants responsible for public information?	0.67	0.47	0	1
	A system of reception of and attention to public information requests	0.55	0.50	0	1
	A system for archives	0.34	0.47	0	1
	Training program for public servants on public information	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 Sustainable finances	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 Provision of public services	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 Sports and recreation	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 Public health	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 Housing	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 reports summary statistics for the outcome variables used for each index. Variables with extreme outliers are winsorized in the main analysis.

less than 75% of total income comes from federal transfers. The data source and coding of each sub-index is explained in Appendix Table A1.[17]

However, the program's official certifications are only systematically available at the index level. To match our independent coding of sub-indexes to the certification data, we aggregate our measures of sub-index status into a standardized public service delivery scale within each index.[18] The continuous nature of these variables is more sensitive to changes in actual service delivery than the coarse red/yellow/green categories that third parties assigned to each index. To the extent that service provision actually improves, our measures are thus well-placed to detect it.

4.2 Identification strategy

To identify the effects of the ADL program on both certification status and actual service delivery outcomes, we exploit the staggered entry of municipalities into the program using a twoway fixed effect difference-in-differences design. As Figure 1 shows, municipal involvement in the program grew significantly over time. While state uptake of the program permits municipal entry into the program, there nevertheless remains substantial within-state variation in the timing of a municipality's entry that we exploit to estimate effects of the program.

Because the first certification was conducted before municipalities could make investments to improve service delivery, we define this baseline certification stage as pre-program and all subsequent certifications as post-program periods. For the annual indicators collected independently from the program that capture the public service delivery indicators on which the program indexes are supposedly based, we define post-program years starting with the year that certification results were first released for the municipality. This definition ensures that no post-program outcomes are

[17]Table 1 shows the raw indicators and Appendix Table A1 shows how they are transformed into our sub-index outcome measures, which use the same 0/1/2 coding as the program's sub-indicators. Some variables, though they correspond to a specific sub-index, cannot be transformed in the 0/1/2 format because the program's coding instructions are incomplete or because they are not identical with the measure of the sub-index. In those cases, we use the standardized variable when aggregating to the index.

[18]For years in which some indicators of a standardized index are missing due to data availability constraints, we calculate the standardized index over all available indicators.

classified as pre-program, because improvements in service delivery could occur within the first certification year.

To test Hypotheses 1 and 2, we compare changes in outcomes across the years before and after a municipality entered the ADL program relative to municipalities that entered the ADL program at an earlier or later date using OLS to estimate the following regression specification:

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

where Y_{imt} is the certified status or a measure of actual service delivery for index i in municipality m (in 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. These fixed effects can, for example, flexibly adjust for uniform changes in how certifier standards or common constraints on municipality service delivery. We further include 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. Under certain conditions discussed below, β estimates the effect of being in the program for a year on index certification status. Standard errors are clustered by municipality throughout. Given the issue of negative weights induced by the staggered nature of the treatment, we implement De Chaisemartin and d'Haultfoeuille (2020)'s estimand.

To test Hypotheses 3 and 4, we further examine heterogeneity in the effects of the ADL program by subsetting the sample by whether municipality m is governed by a mayor that is a copartisan of the state governor in year t,[19] and $Low\ Baseline_{im}$ indicates whether the municipality received a red certification on the index associated with service delivery indicator i when the municipality entered the ADL program. Within our sample, the municipal incumbent was aligned with the state governor's party in 58% of years and 80% of indexes received a low certification upon entry into

[19]Where a 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.

the program. The state × year × index and municipality × index fixed effects are further interacted with state alignment and low baseline status in order to exploit only variation within each category when estimating the effects of entering the ADL program.[20]

4.3 Validating the identification strategy

The design identifies (variance-weighted) average and heterogeneous effects of the ADL program on service delivery outcomes under two assumptions (Goodman-Bacon 2021). First, a generalized common trends assumption requires that untreated potential outcomes in municipalities that entered the program earlier follow the same trend as untreated potential outcome in municipalities that entered the program later. Second, we further require that treatment effects are constant over time, in order for post-treatment trends in municipalities that entered the program earlier to serve as valid counterfactuals for municipalities that became treated later. To help validate these assumptions, we estimate the following event study-type regression:

$$Y_{imt} = \sum_{\tau=-k_0}^{k_1} \beta_\tau Enter_{mt+\tau} + \eta_{ist} + \theta_{im} + \varepsilon_{imt}, \tag{3}$$

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. This specification enables us to examine differences between municipalities that were within k_0 years of entering the program (or within k_1 years of having entered the program) and municipalities that would not enter the AFL program until more than k_0 years later (or municipalities that had entered the program more than k_1 years ago). If the common trends and constant treatment effect assumptions approximately hold, we should expect to observe—as we document below, for up to 3 leads (i.e. $k_0 = 3$)—no significant differences for the lead coefficients.

[20] Low baseline status is not interacted with the municipality × index fixed effects because it is subsumed by these fixed effects.

However, there is no theoretical reason to believe that treatment effects are constant across time. Where this not the case, the modern difference-in-differences literature has shown that twoway fixed effects estimators—like those in equations (2) and (3)—fail to recover an average treatment effect (e.g., Borusyak, Jaravel and Spiess, 2024; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). In our main specification we address this potential concern using the approach recommended by De Chaisemartin and d'Haultfoeuille (2020). Their estimator estimates the treatment effect in the groups that switch treatment, at the time when they switch and it does not rely on any treatment effect homogeneity condition. We also report results using the estimators proposed by Borusyak, Jaravel and Spiess (2024), (Callaway and Sant'Anna, 2021), and (Sun and Abraham, 2021).

Even when counterfactual trends in service delivery outcomes over the years before municipalities enter the program are valid and parallel, entering the program could still coincide with other events that might affect service delivery. Most plausibly, entering the program could coincide with copartisanship between governors or the election of different types of mayors, either of which could affect service delivery through other means—such as different policies or resources. To assess whether such compound treatments occur around the time that a municipality enters the program, we estimate equation (2) with copartisanship and new governing parties as outcomes. The results in panels A and B of Figure A1 and Table **??** in the Appendix show that entry into the program is not significantly correlated with such political changes.

5 Results

We begin by showing that the certified status of the indexes generally increased, relative to the point of entry, after municipalities entered the ADL program. We next assess the extent to which actual service delivery outcomes change, finding negligible effects of the program on average and significant heterogeneity by our proxies for the probability of corruptible incumbent parties and

Table 2: Effect of the ADL program on certified status

	Outcome: Index certification	
	(1)	(2)
	All indexes	Indexes with corresponding indicators
Program	0.486***	0.450***
	(0.015)	(0.024)
Observations	1,199,540	61,541
R^2	0.83	0.82
Outcome range	{0,1,2}	{0,1,2}
Outcome mean	0.73	0.68
Outcome std. dev.	0.91	0.90

Notes: This table reports OLS estimates of equation (2) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator. 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 are clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

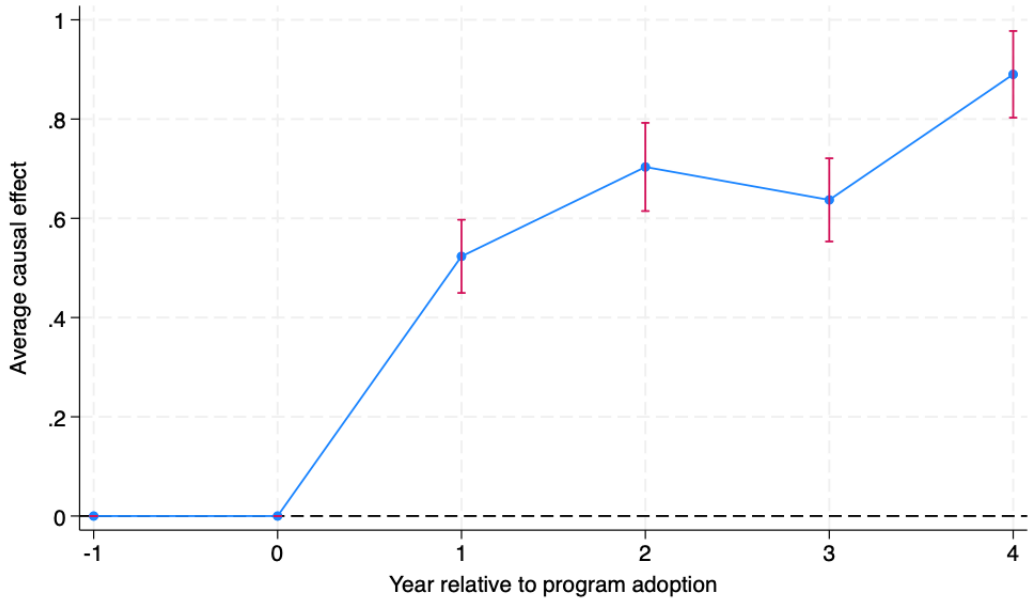
third party certifiers. These results are consistent with the empirical predictions of our theoretical model, and thus suggest that the limitations in the monitoring of investments in future service delivery can play an important role in hindering efforts to improve local public service delivery.

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 in participating municipalities. Considering all indexes, column (1) shows that certification increased

Figure 2: Effect of the ADL program on certified status



Notes: This figure reports coefficients and 95% confidence intervals from estimates of equation (3) with 4 lags using De Chaisemartin and d'Haultfoeuille (2020)'s estimator. The sample includes each of the indexes for which we have measures constructed with data collected independently from the program.

by 0.49 levels above the baseline status after a municipality entered the ADL program. Among the sub-indexes for which independent measures of service delivery outcomes are available, column (2) indicates that this effect slightly increased to 0.45 levels among the indexes that we focus on. Figure 2 illustrates this effect by year over the duration of a municipality's participation in the program.[21] These estimates show that, after 5 years in the program, the average index has increased by 0.8 levels. However, much of the apparent improvement occurs within the first year.

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 accordingly considers as its outcome the independently-constructed indicators that correspond to the indexes certified as part of the program. These indicators are measured for all

[21] We cannot include certification levels prior to the municipality entering the program because certifications only start upon entering the program.

Table 3: Effect of the ADL program on index-level public service delivery scales

	Outcome: Public service delivery scales (standardized)			
	(1)	(2)	(3)	(4)
Program	0.053**	0.053**	0.053**	0.053**
	(0.020)	(0.021)	(0.020)	(0.021)
Lead 1		0.021	0.021	0.021
		(0.016)	(0.020)	(0.020)
Lead 2			0.032	0.032
			(0.022)	(0.026)
Lead 3				-0.128***
				(0.038)
Observations	141,417	141,417	141,417	141,417
Outcome range	[-7.51,8.02]	[-7.51,8.02]	[-7.51,8.02]	[-7.51,8.02]
Outcome mean	0	0	0	0
Outcome std. dev.	1	1	1	1

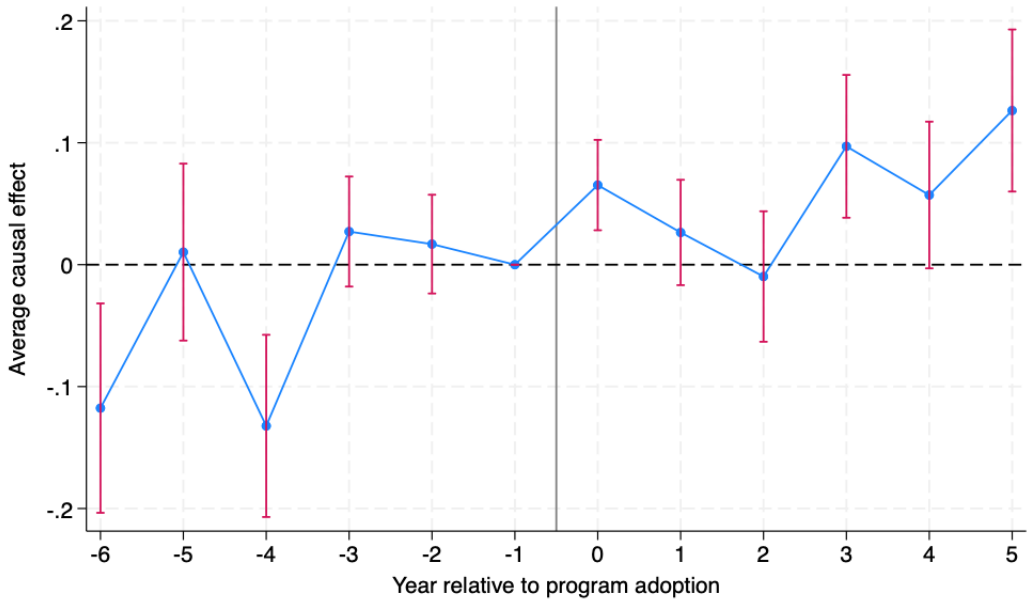
Notes: Column (1) reports OLS estimates of equation (2) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator, while columns (2)-(4) report OLS estimates of equation (3) with 1 to 3 leads. Standard errors are clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

available years before and after a municipality enters the program, ensuring that the results are not driven by selective attrition from the program.

In line with Hypothesis 2, column (1) indicates that there is, on average, a positive effect of the program on public service delivery. The estimate corresponds to a 0.05 standard deviation change in the indicators of service delivery supposed to underpin the certification process.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 design. These results are corroborated visually in Figure A2.

Table 4 focuses on the same outcome as Table 3, but now considers heterogeneity in the effect of a municipality entering the ADL program. Columns (2–3) first report the results when subsetting by the political alignment between municipal and state governments, which proxies for the

Figure 3: Effect of the Program on Public Service Delivery



Notes: This figure reports coefficients and 95% confidence intervals from estimates of equation (3) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator with 3 leads and 3 lags using public service delivery scales as the outcome.

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

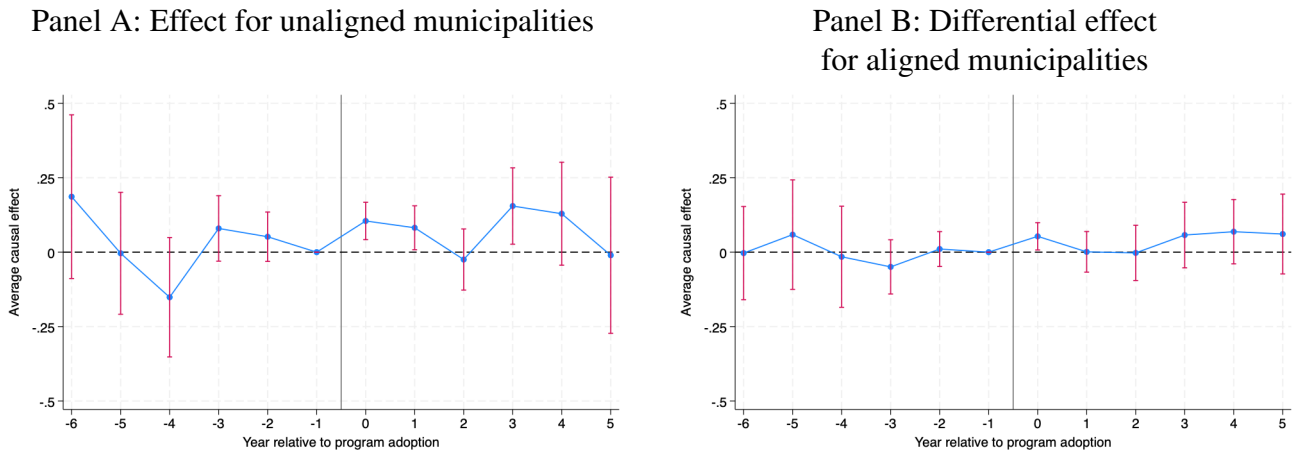
	Outcome: Public service delivery scales (standardized)				
	All	No State Alignment	State Alignment	High Baseline	Low Baseline Baseline
	(1)	(2)	(3)	(4)	(5)
Program	0.053**	0.080**	0.036	0.103**	0.052*
	(0.021)	(0.033)	(0.029)	(0.049)	(0.028)
Observations	141,417	30,820	49,211	10,155	56,864
Outcome range	[-7.51,8.02]	[-9.43,8.71]	[-9.43,8.71]	[-9.43,8.71]	[-9.43,8.71]
Outcome mean	0	0	0	0	0
Outcome std. dev.	1	1	1	1	1

Notes: This table reports OLS estimates of equation (2) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator. Column (1) uses the whole sample; columns (2–3) subsets by the political alignment between municipal and state governments; columns (4–5) subset by a low baseline certied status of a given index. Standard errors are clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

likelihood of corruptibility of the certifying third parties. Consistent with Hypothesis 3, we find statistically significant effects in municipalities that are not aligned with the state government and smaller, insignificant effects in aligned municipalities. Together, these results suggest that the ADL program provides cover for corruption in some municipalities. 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 by political alignment 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.

Columnss (4–5) of Table 4 further report the results of subsetting by a low baseline certified status of a given index, which proxies for the likelihood that a municipal incumbent is corruptible with respect to that index. In line with Hypothesis 4, the results indicate that the effect of the program on public service delivery is is considerably smaller for indicators corresponding to indexes with a low baseline status certified. However, as theorized, we observe a relatively sizable

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



Notes: This figure reports coefficients and 95% confidence intervals from estimates of equation (2) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator. Panel A reports the coefficients on year since program start for observations where the municipality is not governed by the same party as its state. Panel B shows the coefficients for observations where they are aligned.

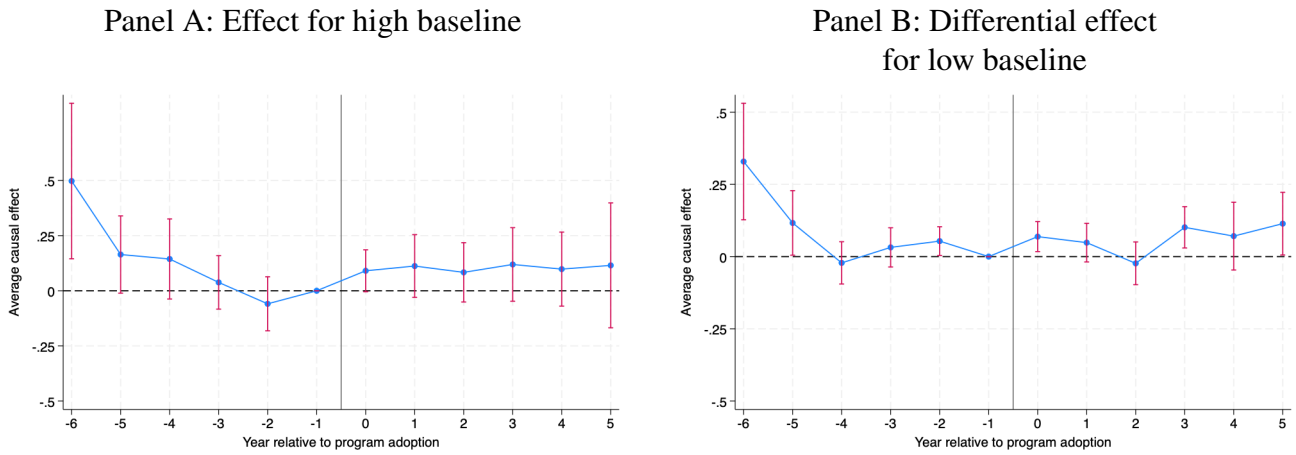
positive effect of the program in indexes with high baseline certification status. These effects are illustrated in Figure 5, which reports the effects of the program on public service delivery by year since entering the program by baseline certification status. The coefficients in those plots again suggest that differential pre-trends do not account for this finding.

To explore which indexes drive these results, Tables **??**-**??** report our main estimates for each outcome index separately.

5.3 Robustness checks

We conduct several robustness exercises to demonstrate that our results are not driven by particular parameterizations of our regressors, potential data quality concerns, or heterogeneous treatment effects over time. First, Appendix Table **??** shows that the results in Table 4—the core findings supporting our theoretical model—are robust to using several alternative operationalizations of baseline level of certification status. In particular, our conclusions are qualitatively similar if

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



Notes: This figure reports coefficients and 95% confidence intervals from estimates of equation (2) using De Chaisemartin and d'Haultfoeuille (2020)'s estimator. Panel A reports the coefficients on year since program start for observations where the municipality received a high baseline score in a given index. Panel B shows the coefficients for observations where the baseline score was low.

the probability of a mayor being corruptible in producing a given service is instead proxied by non-parametric or linearized measures of the baseline certified status of a given index. Table **??** similarly explores alternative ways of coding our proxy for the probability that an incumbent government is corrupt. Instead of using the first year certification as the baseline, it shows—at the cost of lost observations—that our results are robust to defining a low baseline score by the index outcome in the year before the program started. Column (2) uses a dummy indicating whether the municipality was in the bottom tercile for the index, while column (3) uses a dummy for whether the index score was below the median.

Second, we further conduct several exercises to ensure that our results are not specific to the way we coded our service delivery outcomes. Appendix Table **??** reports estimates using four alternative coding 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.

Third, Figures A3–A5 plots the coefficients when implementing standard two-way fixed effects OLS estimator, Borusyak, Jaravel and Spiess (2024)'s robust and efficient estimator, Callaway and Sant'Anna (2021)'s group-time average treatment effect estimator, the interaction weighted (IW) estimator proposed by Sun and Abraham (2021). Figure A3 estimates the ATE. Figure A4 show the results when subsetting by whether the municipal government is aligned with the state governor, using the last treated municipalities as the control group (because the moderator is not defined for never-treated municipalities). Figure A5 similarly show the results when interacting with an indicator for low baseline score. Note that due to the differences across the estimators a visual comparison of the pretrends is misleading as they are estimating different counterfactuals (Roth, 2024). However, across specifications a picture emerges. On average the program exhibits positive effects that are driven by unaligned municipalities and low baseline indicators. Throughout these robustness checks, our heterogeneous effects results remain qualitatively unchanged.

6 Conclusion

Our theory and evidence suggest that the short-term unobservability of investments to improve service delivery represents an important constraint on such critical investments. They also indicate that, at least in theory, effective monitoring of such investments could help to overcome this impediment by enabling good politicians to pursue such policies through the creation of electoral incentives that constrain corruptible politicians to follow suit. However, we also illustrate how the

effectiveness of monitoring technologies can be undermined by existing institutional weaknesses. In particular, while we find that such a certification program in Mexico had a positive effect on service delivery on average, we find no effects of the program on service delivery in politically connected municipalities. Rather, in line with our theoretical argument, the heterogenous effects on actual service delivery appear to reflect 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 that 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 broadly within 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 by 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. Our findings suggest 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 audit and monitoring programs 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

Ashworth, Scott. 2005. "Reputational dynamics and political careers." *Journal of Law, Economics, and Organization* 21(2):441–466.

Avis, Eric, Claudio Ferraz and Frederico Finan. 2018. "Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians." *Journal of Political Economy* 126(5):1912–1964.

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.

Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(1):19–42.

Besley, Timothy and Robin Burgess. 2002. "The political economy of government responsiveness: Theory and evidence from India." *Quarterly Journal of Economics* 117(4):1415–1451.

Borusyak, Kirill, Xavier Jaravel and Jann Spiess. 2024. "Revisiting event-study designs: robust and efficient estimation." *Review of Economic Studies* p. rdae007.

Brollo, Fernanda and Tommaso Nannicini. 2012. "Tying your enemy's hands in close races: the politics of federal transfers in Brazil." *American Political Science Review* pp. 742–761.

Callaway, Brantly and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225(2):200–230.

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.

Dal Bo, Ernesto and Martin A. Rossi. 2011. "Term Length and the Effort of Politicians." *Review of Economic Studies* 78(4):1237–1263.

De Chaisemartin, Clément and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110(9):2964–2996.

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 Hyde, Craig McIntosh and Gareth Nellis. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge University Press.

Enríquez, José Ramón, Horacio Larreguy, John Marshall and Alberto Simpser. 2022. "Mass Political Information on Social Media: Facebook Ads, Electorate Saturation, and Electoral Accountability in Mexico." Working paper, https://scholar.harvard.edu/files/jmarshall/files/mexico_saturation_intervention_v12.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.

Gailmard, Sean and John W. Patty. 2019. "Preventing prevention." *American Journal of Political Science* 63(2):342–352.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225(2):254–277.

Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112(2):280–301.

Habyarimana, James, Macartan Humphries, Daniel N. Posner and Jeremy M. Weinsten. 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.

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. 2020. "Publicising Malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico." *Economic Journal* 130(631):2291–2327.

Mani, Anandi and Sharun Mukand. 2007. "Democracy, Visibility and Public Good Provision." *Journal of Development Economics* 83(2):506–529.

Marx, Benjamin. 2018. "Elections as Incentives: Project Completion and Visibility in African Politics." Working Paper, https://drive.google.com/file/d/1aSDEhq-ZGlUcX5QxWbb-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.

Raffler, Pia. forthcoming. "Does Political Oversight of the Bureaucracy Increase Accountability?

Field Experimental Evidence from an Electoral Autocracy." *American Political Science Review* .

Roth, Jonathan. 2024. "Interpreting Event-Studies from Recent Difference-in-Differences Methods." *arXiv preprint arXiv:2401.12309* .

Simpser, Alberto, Lauren Duquette-Rury, José Antonio Hernández Company and Juan Fernando Ibarra. 2016. "The political economy of social spending by local government: A Study of the $3\times$ 1 Program in Mexico." *Latin American Research Review* pp. 62–83.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225(2):175–199.

Tavits, Margit. 2005. "The Development of Stable Party Support: Electoral Dynamics in Post-Communist Europe." *American Journal of Political Science* 49(2):283–298.

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* .

A Online appendix

Contents

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:

$$\left(g^*_{i,1}, G^*_{i,1}, r^*_{i,1}\right) = \begin{cases} (1,0,0) & \text{if} \quad \gamma_i < \gamma^*, \\ (0,1,0) & \text{if} \quad \gamma_i \geq \gamma^*, \tau_i = h, \\ (0,0,1) & \text{if} \quad \gamma_i \geq \gamma^*, \tau_i = c \end{cases}$$

and

$$\left(g^*_{i,2}, G^*_{i,2}, r^*_{i,2}\right) = \begin{cases} (0,1,0) & \text{if} \quad \tau_i = h, \\ (0,0,1) & \text{if} \quad \tau_i = c \end{cases}.$$

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

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

and

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

where $\widetilde{\gamma}_i = 0$ 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 $g_{i,1} = 1$, because $\tau_i = h$ 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:

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

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

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

and

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

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 $G^h_{i,1} = 1$ or from $g^h_{i,1} = 0$ 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 $r^c_{i,1} = 1$ in such an equilibrium; in an equilibrium where $g^c_{i,1} = 0$, 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 $g^h_{i,1} \neq g^c_{i,2}$.) 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:

$$\left(g^*_{i,1}, G^*_{i,1}, r^*_{i,1}\right) = \begin{cases} (1,0,0) & \text{if} \quad \gamma_i < \gamma^{**}(\rho), \\ (0,1,0) & \text{if} \quad \gamma_i \geq \gamma^{**}(\rho) \text{ and either } \tau_i = h \text{ or } \alpha = H \text{ and } \tau_i = c, \\ (0,0,1) & \text{if} \quad \alpha = C \text{ and } \tau_i = c, \end{cases}$$

and

$$\left(g^*_{i,2}, G^*_{i,2}, r^*_{i,2}\right) = \begin{cases} (0,1,0) & \text{if} \quad \tau_i = h, \\ (0,0,1) & \text{if} \quad \tau_i = c. \end{cases}$$

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

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

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:

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

and

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

$$\widetilde{\gamma}_B\left(g_{B,1}=0, c_i=0\right)=0,$$

and

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

where $\widetilde{\gamma}_i=0$ corresponds to an off-equilibrium belief. Given that only $\tau_i=c$ types can profitably deviate from playing $g_{i,1}=1$, because $\tau_i=h$ 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, $v\left(g_{A,1}, g_{B,1}, c_A, c_B\right)$, is as follows:

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

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

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

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

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

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 Corollary 1

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

Proof of Corollary 2

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:

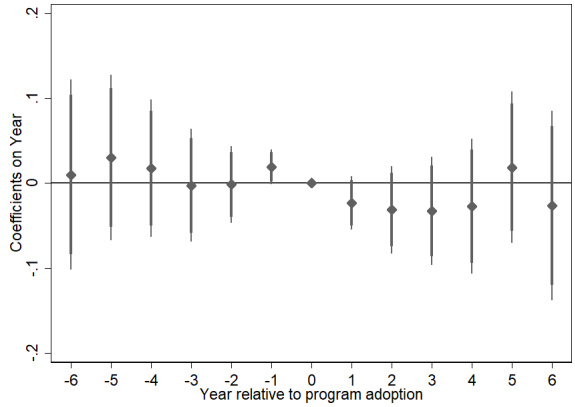
$$\begin{aligned}
\Delta_A &:= \mathbb{E}\left[\sum_{t=1,2}U\left(g_{A,t},G_{A,t}\right)\middle|p=1\right]-\mathbb{E}\left[\sum_{t=1,2}U\left(g_{A,t},G_{A,t}\right)\middle|p=0\right]\\
&= \left(1-\gamma_A\right)\rho\beta,\\
\Delta_B &:= \mathbb{E}\left[\sum_{t=1,2}U\left(g_{B,t},G_{B,t}\right)\middle|p=1\right]-\mathbb{E}\left[\sum_{t=1,2}U\left(g_{B,t},G_{B,t}\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

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 $\gamma_i<\gamma^{**}$.

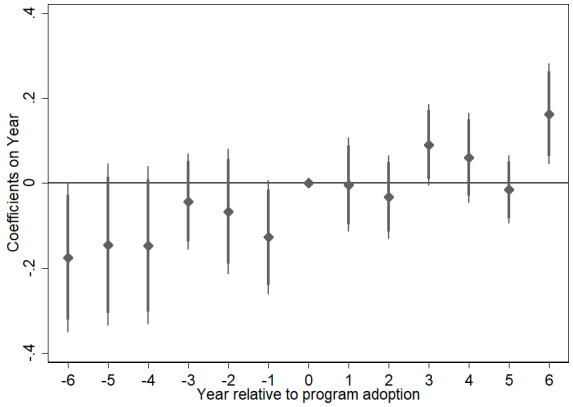
A.2 Additional figures

Figure A1: Effect of the ADL program on new government and state alignment

Panel A: Effect on alignment with state government

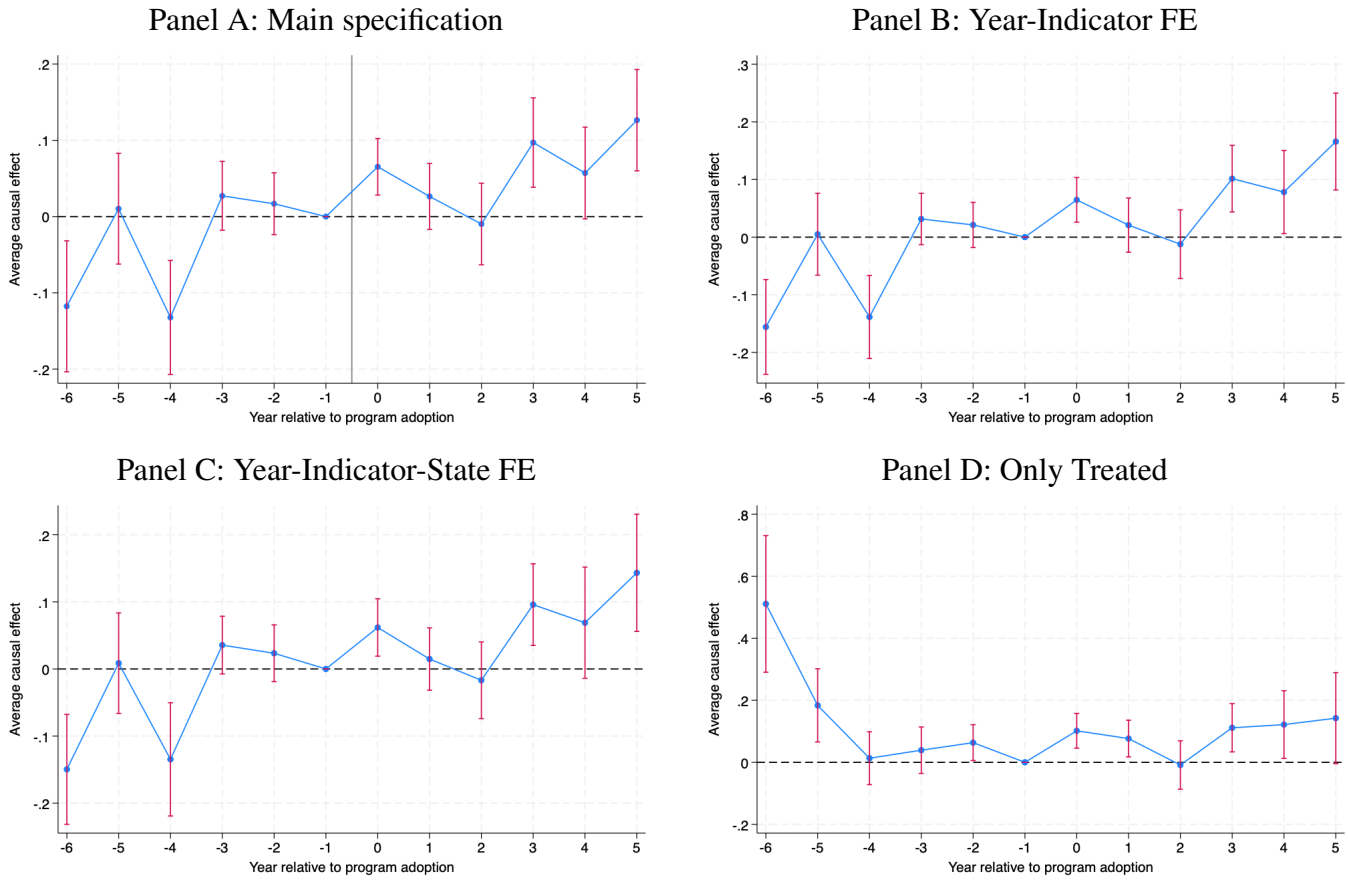


Panel B: Effect on having a new government



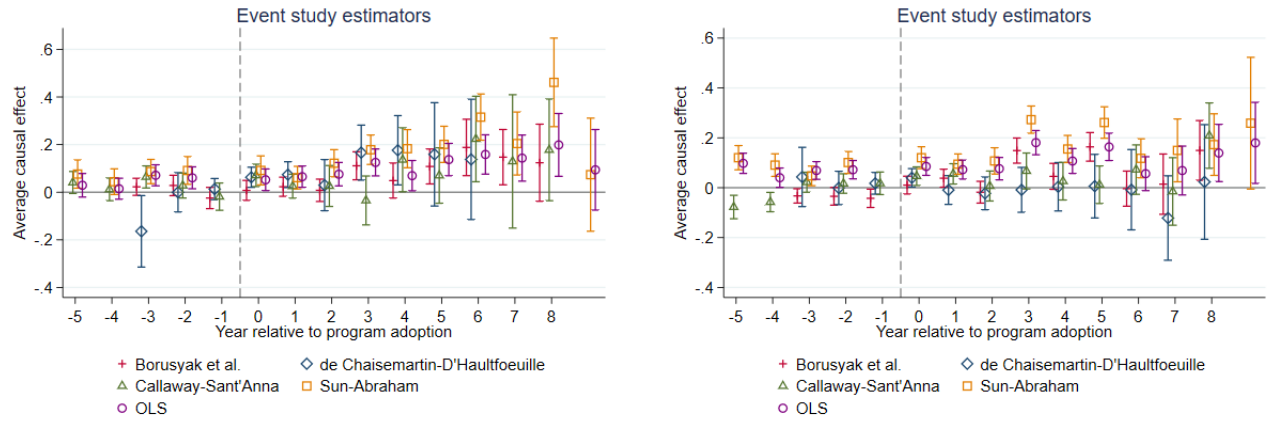
Notes:This figure reports coefficients and 90% and 95% confidence intervals from estimates of equation (2). Panel A reports the coefficients on year since program start with whether the municipal government is governed by a new party as the outcome. Panel B reports the coefficients on year since program start with whether the municipal government is aligned with the state government as the outcome variable.

Figure A2: Effect of the Program on Public Service Delivery



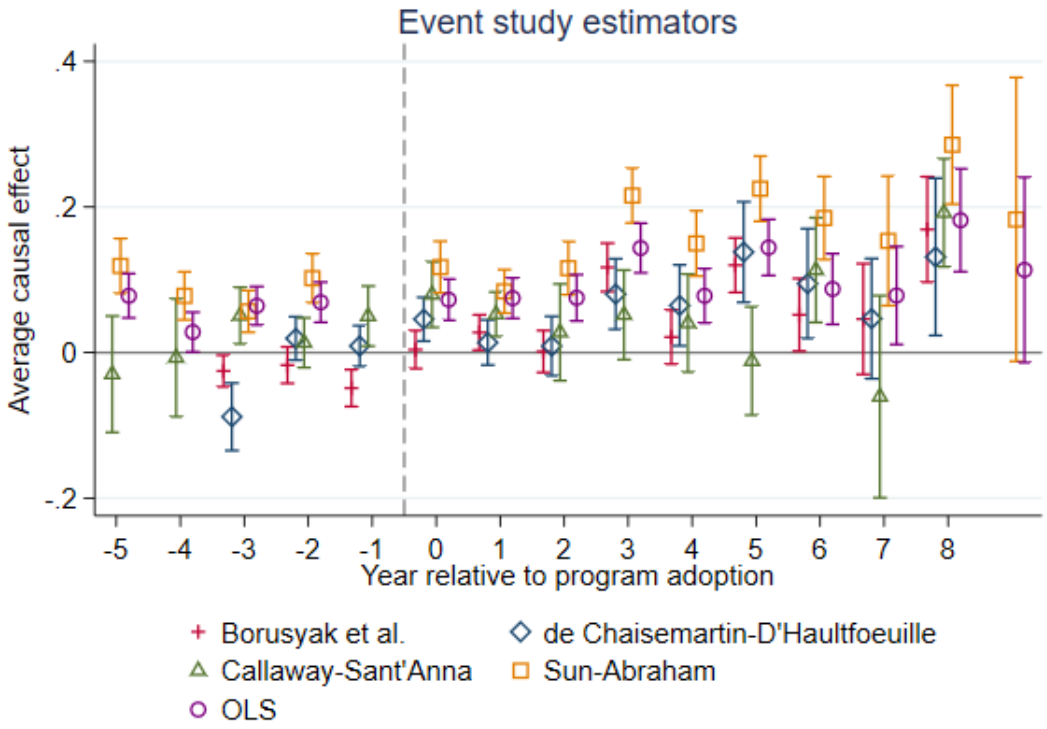
Notes:

Figure A4: Comparing DiD measures: Effects on State Capacity separately by state-match



Notes:

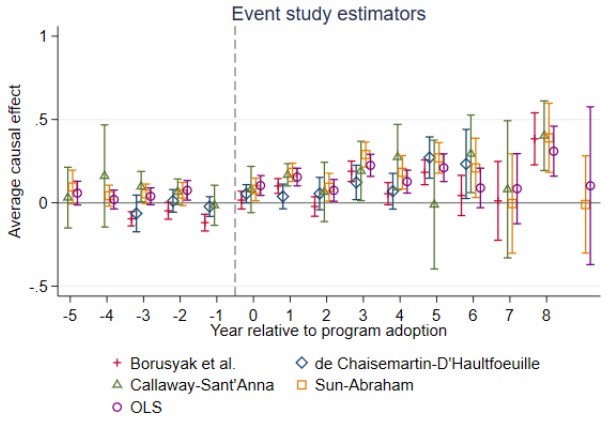
Figure A3: Comparing DiD measures: ATT



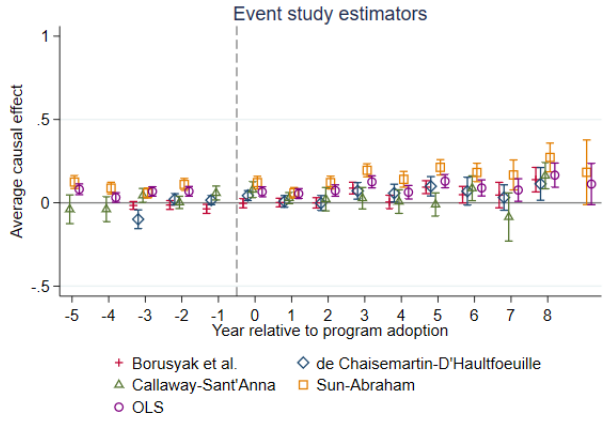
Notes:

Figure A5: Comparing DiD measures: Effects on State Capacity separately by intensity-dummy

Panel A: $intensitydummy = 0$



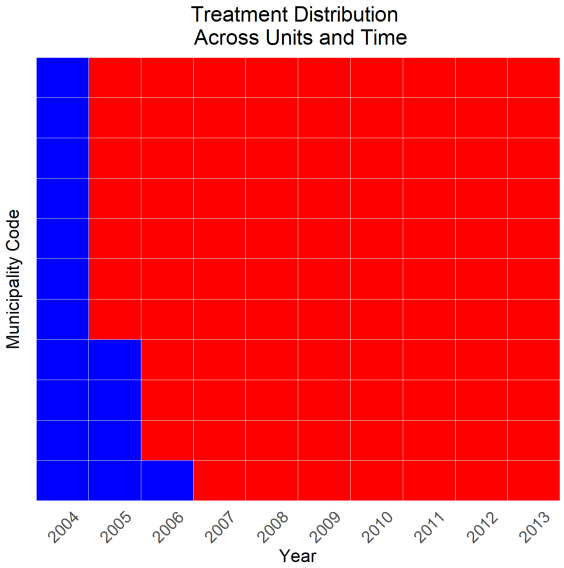
Panel B: $intensitydummy = 1$s



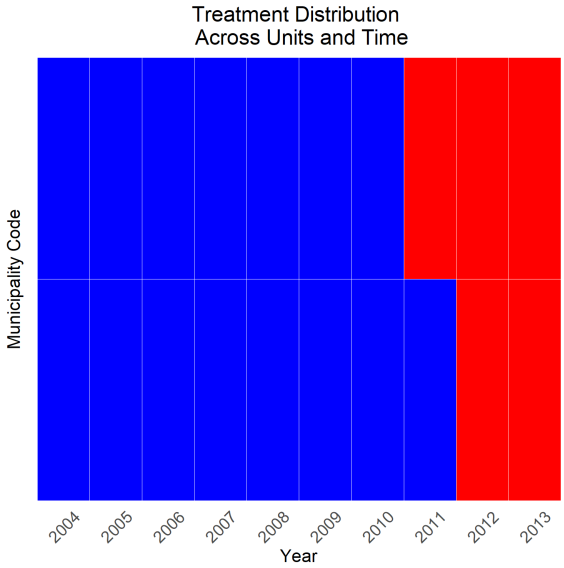
Notes:

Figure A6: Visualizing Treatment Distribution by States

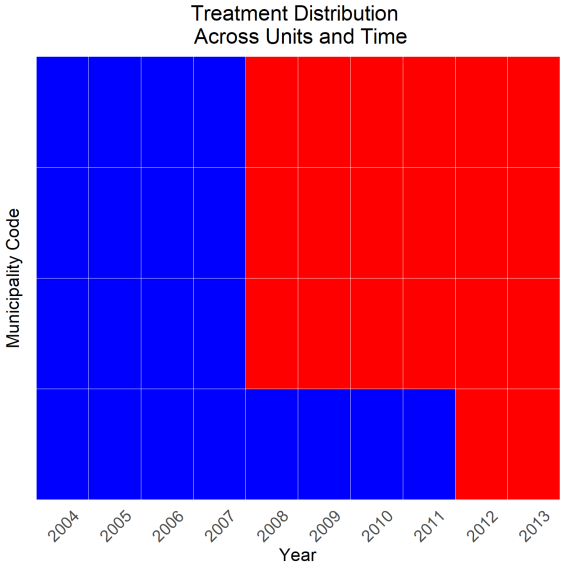
Panel A: Aguascalientes



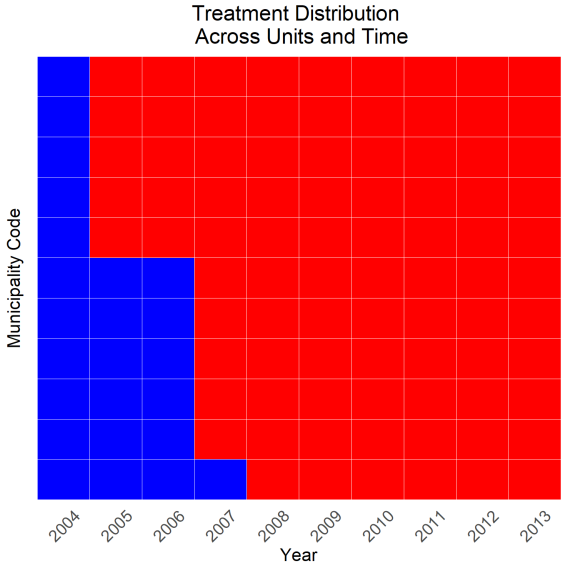
Panel B: Baja California



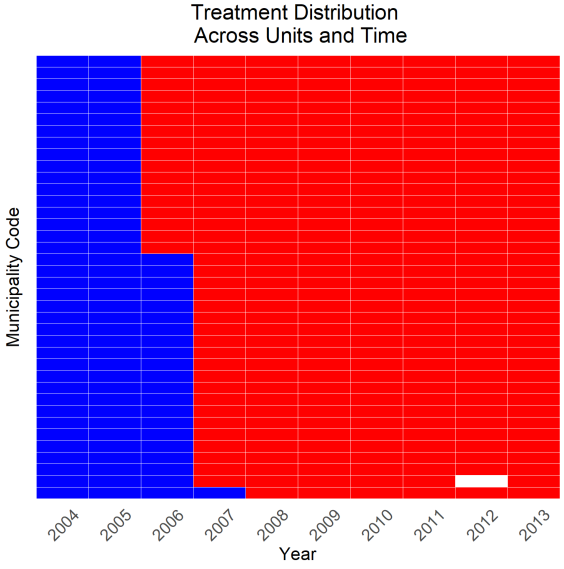
Panel C: Baja California Sur



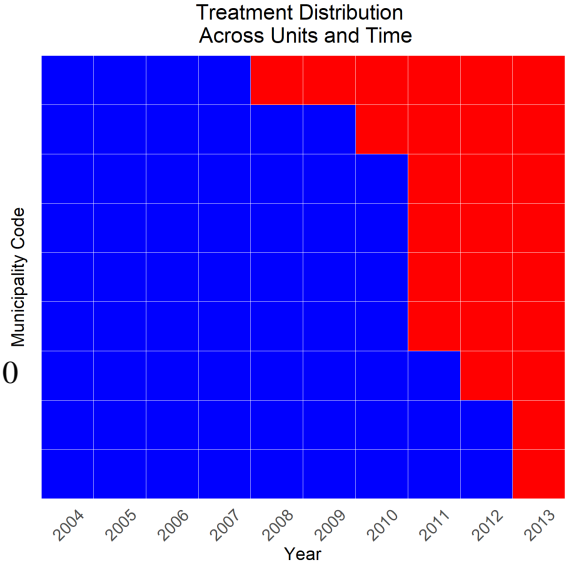
Panel D: Campeche



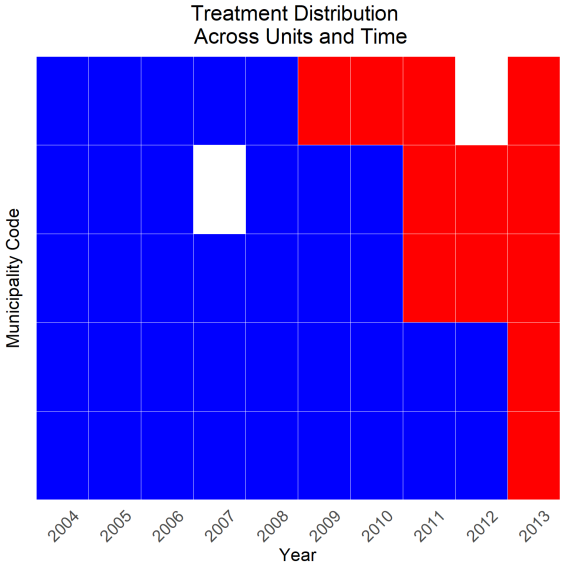
Panel E: Coahuila



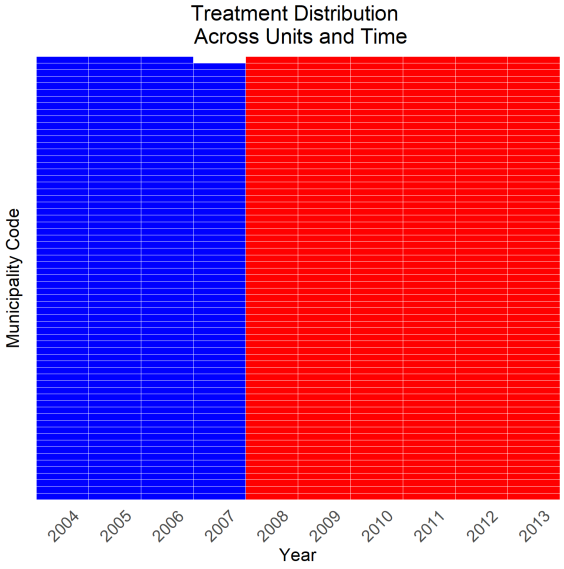
Panel F: Colima



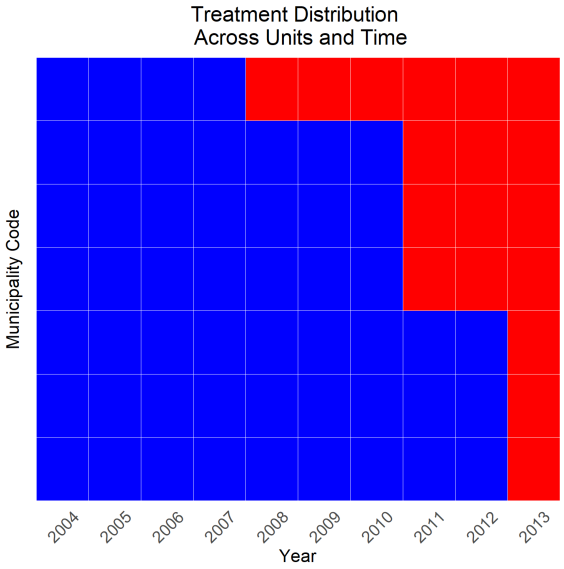
Panel G: Chiapas



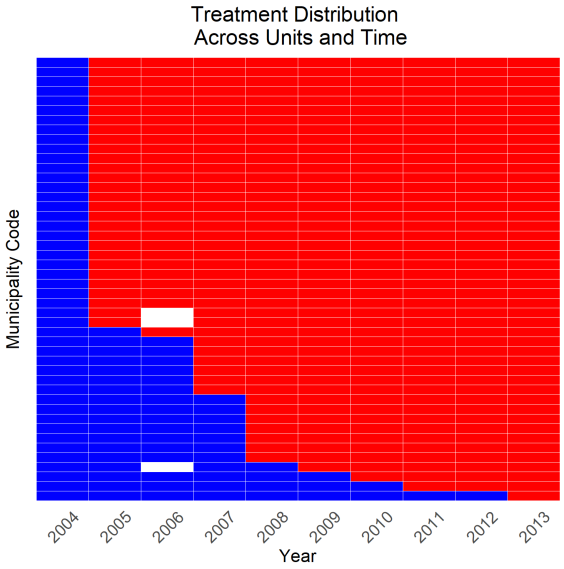
Panel H: Chihuahua



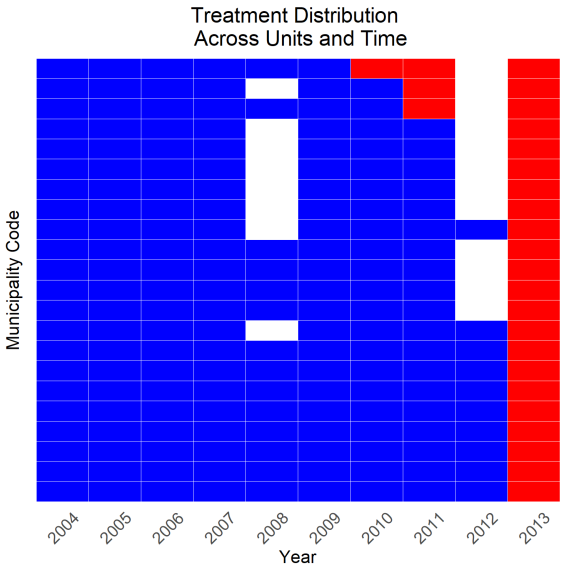
Panel I: Durango



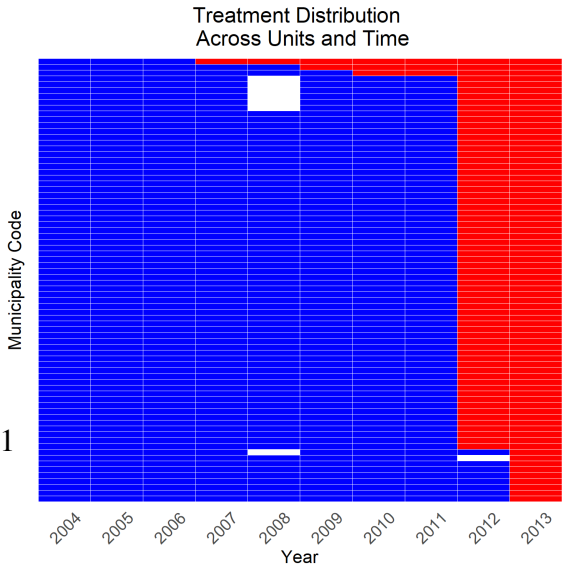
Panel J: Guanajuato



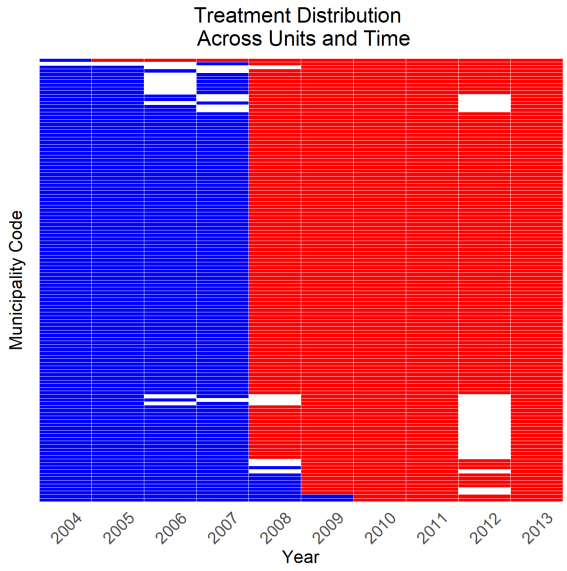
Panel K: Guerrero



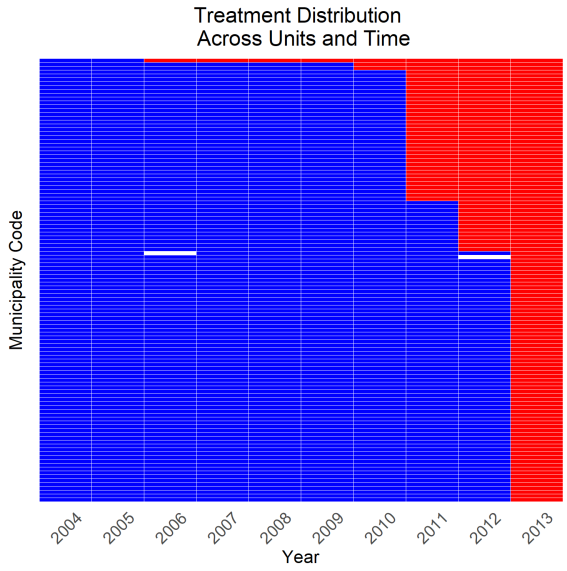
Panel L: Hidalgo



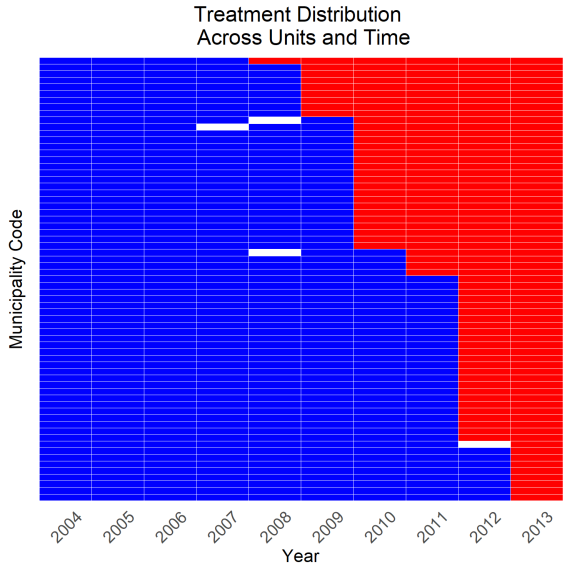
Panel M: Jalisco



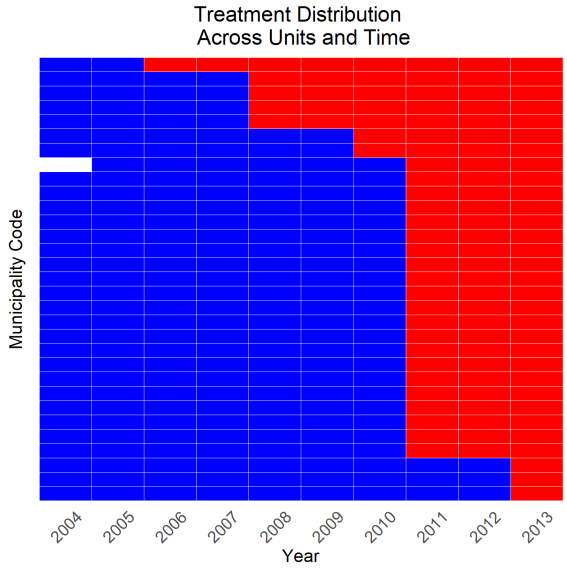
Panel N: Mexico



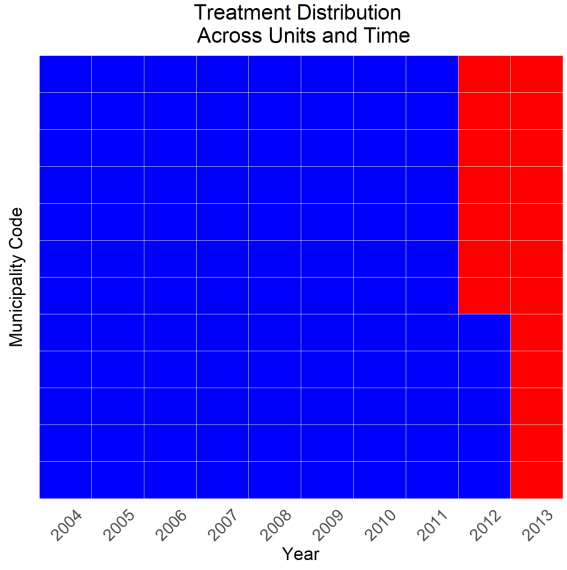
Panel O: Michoacan



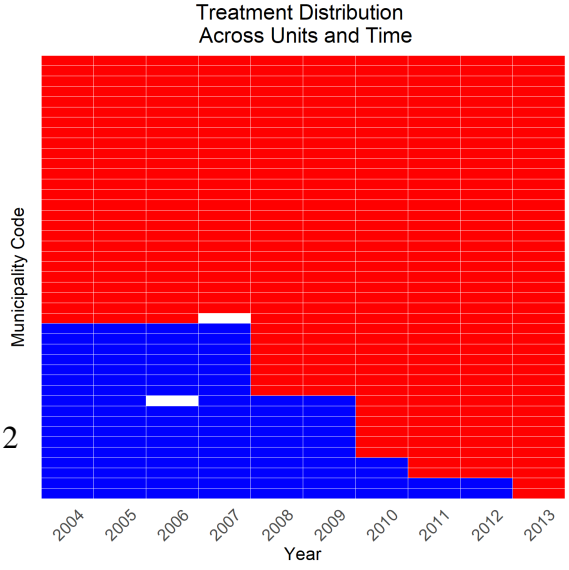
Panel P: Morelos



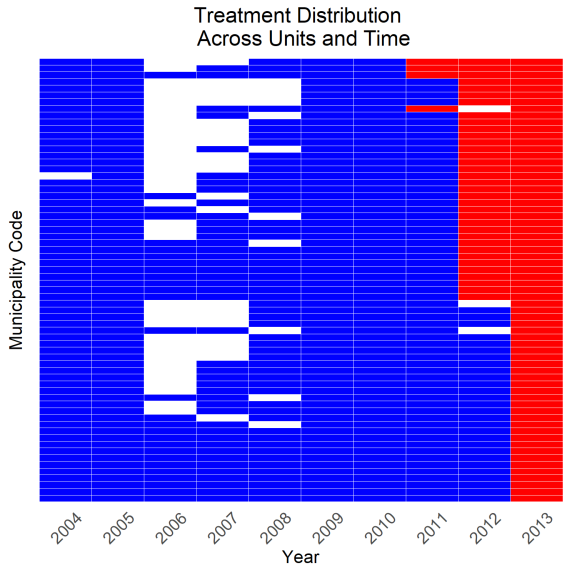
Panel Q: Nayarit



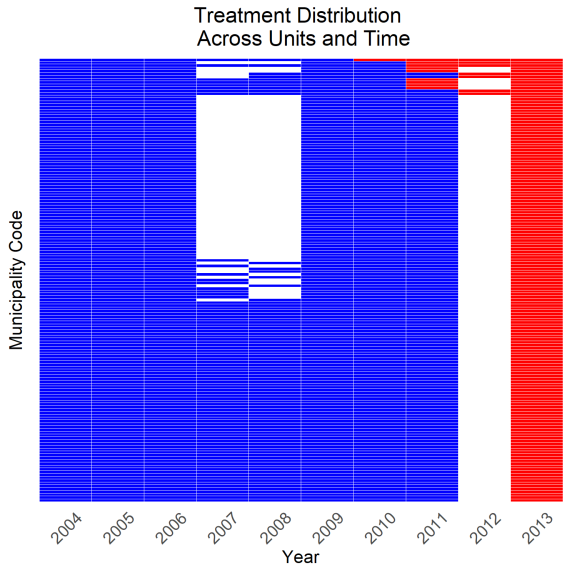
Panel R: Nuevo Leon



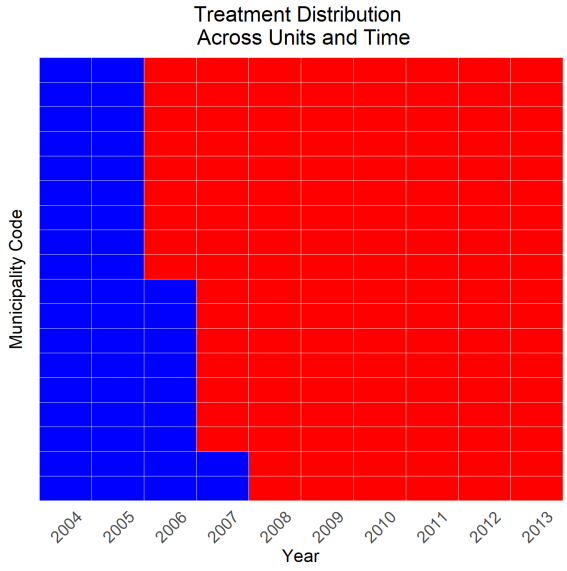
Panel S: Oaxaca



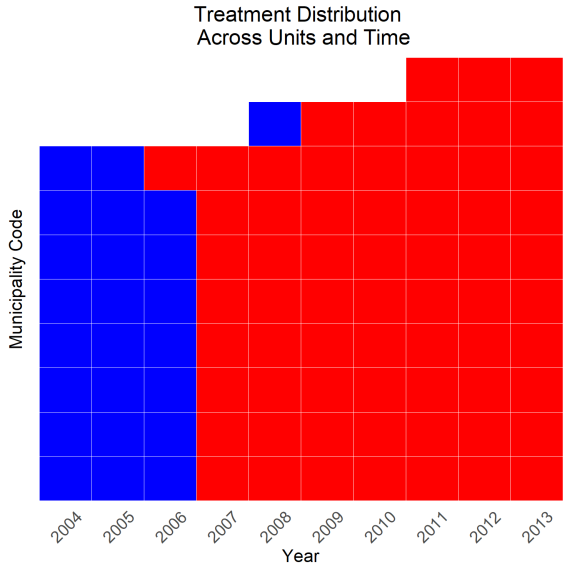
Panel T: Puebla



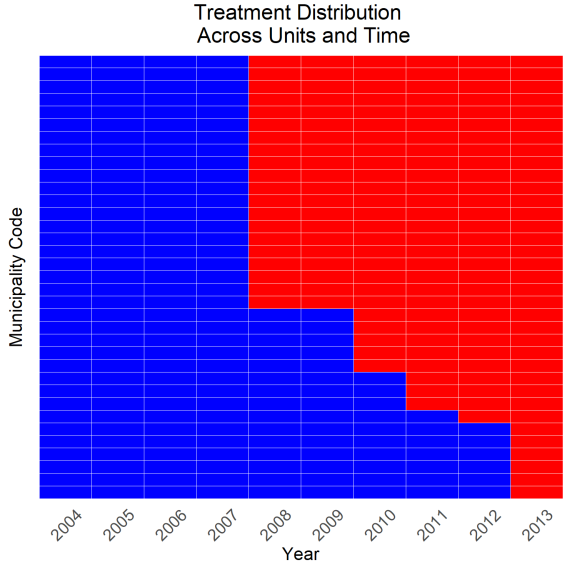
Panel U: Queretaro



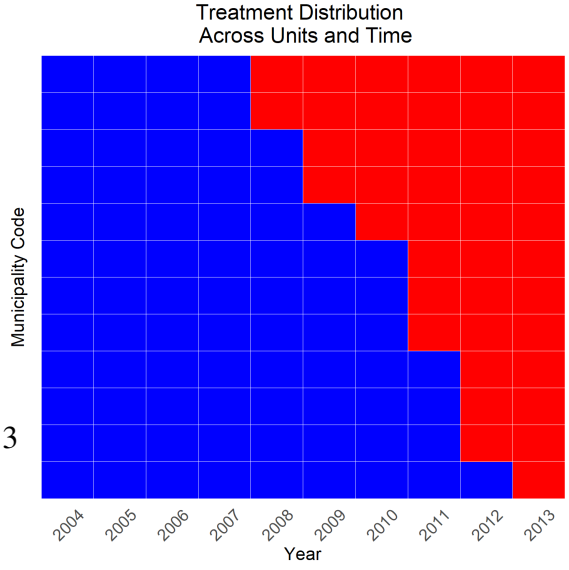
Panel V: Quintana Roo



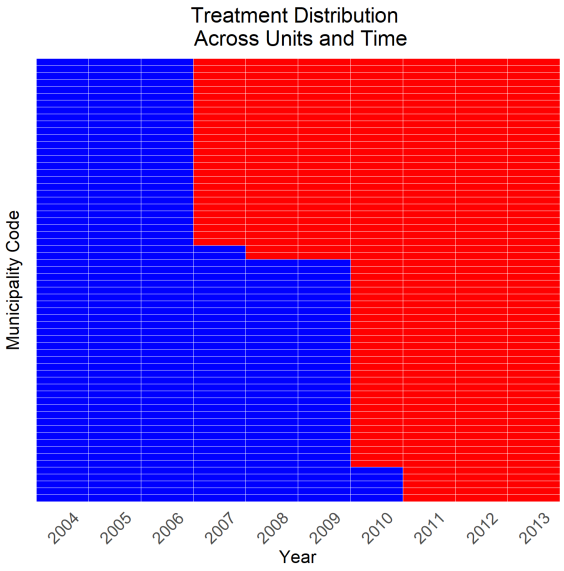
Panel W: San Luis Potosi



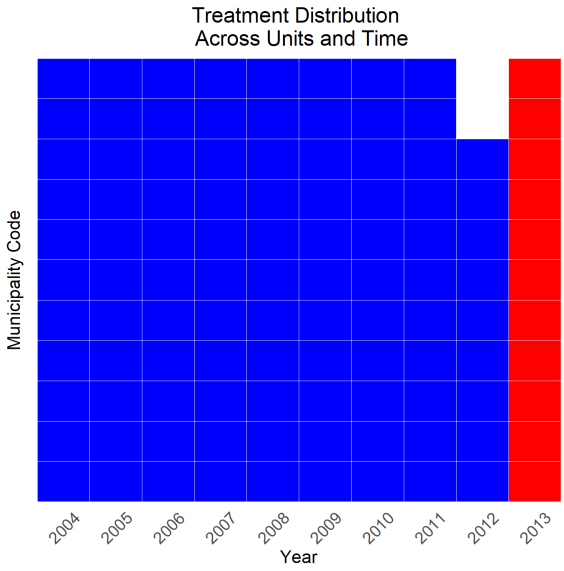
Panel X: Sinaloa



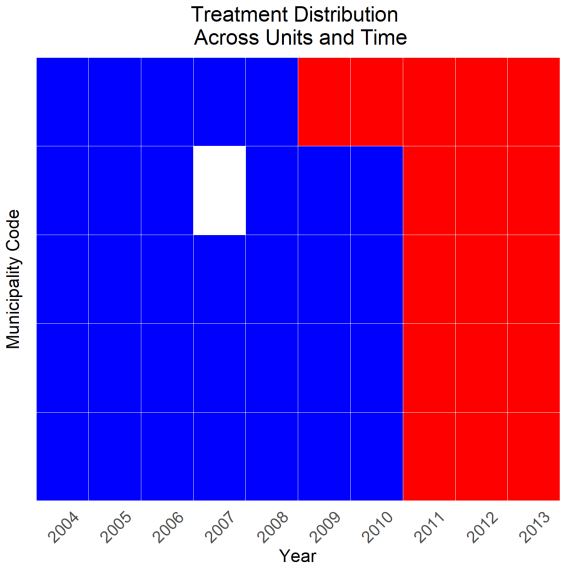
Panel Y: Sonora



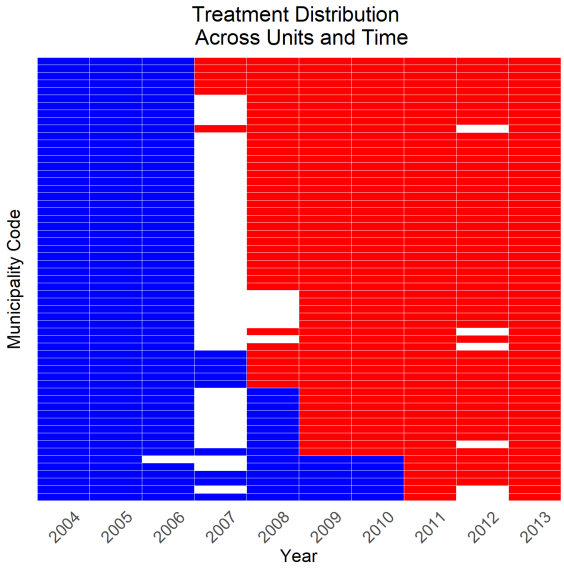
Panel Z: Tabasco



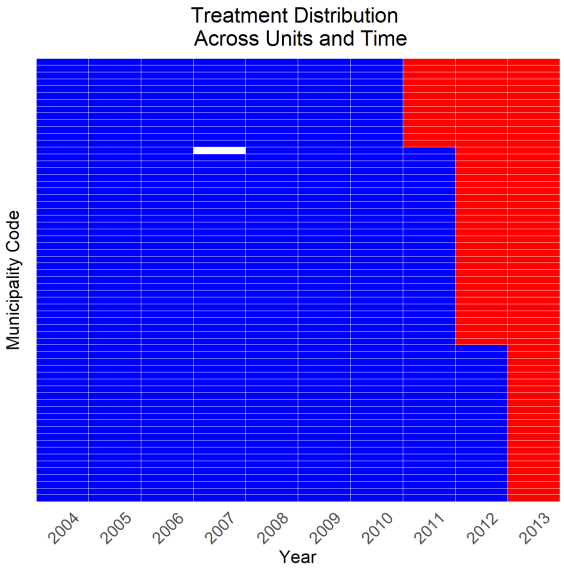
Panel AA: Tamaulipas



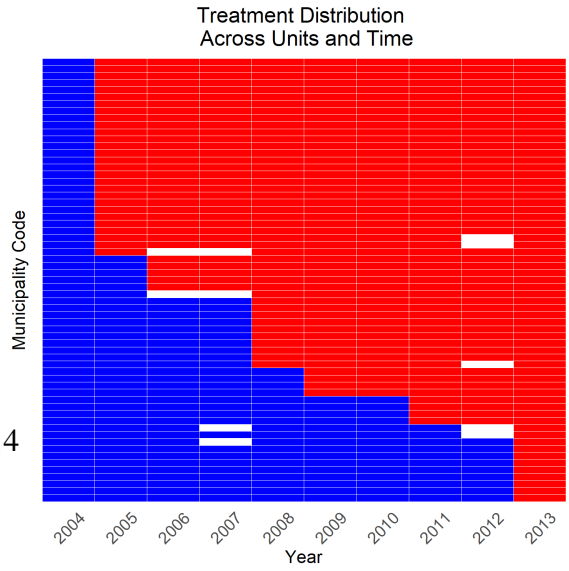
Panel AB: Tlaxcala



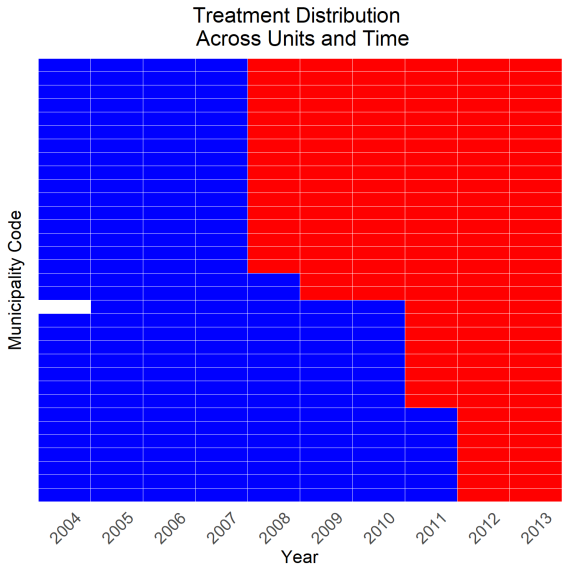
Panel AC: Veracruz



Panel AD: Yucatan



Panel AE: Zacatecas



Notes: Pre-treatment periods are shown in blue, while post-treatment periods are shown in red. Each row represents a municipality within a state.

A.3 Additional tables

Table A1: Linking indexes to outcomes

Index	Subindex	Data source	Coding
1.1 Public accounts	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 Citizen participation	Whether the administrative structure has a sector for promoting social participation	Municipal Census	Standardized[a]
	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 citizen participation?	Municipal Census	Standardized

[a] Some variables, such as this one, cannot be transformed in the 0/1/2 format because the program's coding instructions are incomplete or because they are not identical with the measure of the sub-index. In those cases we use the standardized variable.

Linking indexes to outcomes

Index	Sub-index	Data source	Coding
1.6 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 With regulatory framework	What percentage of transport, police, markets, graves, public works, cleaning, butchers, participation, civil protection, cadastre, fire, zoning have regulations	Municipal Census	Standardized
1.10 Transparency	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 Sustainable 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 Provision 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 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 Public health	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 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