

YOU GET WHAT YOU PAY FOR? CAN CERTIFICATION PROGRAMS CONTRIBUTE TO INCREASED PUBLIC SERVICE DELIVERY? *

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

APRIL 2019

A large recent literature emphasizes the importance of weak state capacity in explaining low public service delivery. However, less is known about the incentives and ability of governments to overcome such a constraint. We develop a theoretical model highlighting how the short-term unobservability of state capacity investments constrains such investments and ultimately limits service delivery, as well as the potential of corruptible certification programs to break this bad equilibrium. We test the model's empirical predictions in the context of a Mexican federal program designed to increase municipal public service delivery by certifying state capacity and public good provision. Certification is self-assessed and corroborated by corruptible local third-party institutions. Our difference-in-differences estimates indicate that certification had no average effect on overall public service delivery by municipal governments. However, consistent with the model, certification led to lower overall service delivery in municipalities where the certifying third party was more likely to be corrupt and municipal incumbents are more likely to be corrupt.

*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 and for excellent research assistance. Horacio Larreguy acknowledges financial support from the IQSS Undergraduate Research Scholars Program. All errors are our own.

[†]Department of Government, Harvard University. henn@g.harvard.edu.

[‡]Department of Government, Harvard University. hlarreguy@fas.harvard.edu.

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

1 Introduction

A large recent literature highlights the importance of weak states in explaining low public service delivery (Acemoglu 2005; Besley and Persson 2009; Dell, Lane and Querubin 2017; Michalopoulos and Papaioannou 2013). States with strong bureaucratic and fiscal capacities have the potential to serve significant shares of their societies with public service delivery. In contrast, fragile states lack these capacities (Acemoglu 2005; Herbst 2000; Michalopoulos and Papaioannou 2014). Given its importance, an understanding of the sources of state strength has demanded considerable recent academic attention (Acemoglu 2005; Besley and Persson 2009, 2010; Centeno 1997; Ferguson, Larreguy and Riaño 2018; Thies 2005, 2007; Sanchez de la Sierra 2017). However, little is still known about the incentives and ability of governments to invest in state capacity to overcome low service delivery.

We develop a theory that highlights the short-term unobservability of state-capacity investments as a constraint to those investments and thus service delivery. We consider an adverse selection, two-period model where municipal politicians are either honest or corrupt. Both types care about rents from holding office, but the former shares the voters' preferences and the latter enjoys funds appropriated from the public treasury. In our stylized model, municipal incumbents can allocate their budget into current public service delivery, state-capacity investments, or fund appropriation. While state-capacity investments entail lower current public service delivery, they lead to a more efficient future public service delivery. We consider the case where these investments are socially desirable. Voters decide whether to re-elect the incumbent between periods. While voters can observe if the budget was spent entirely into current public service delivery, otherwise they cannot distinguish at the time of the election whether the incumbent appropriated or invested the budget.

We show that, when the possibility of a corrupt incumbent in a municipality is sufficiently high, investments in state capacity are not feasible. This follows from the fact that, when voters do not see the budget being spent entirely on current service delivery, they are more likely to think

that they have a corrupt incumbent that has appropriated the budget than an honest incumbent that invested in state capacity, and thus vote the incumbent out of office. This result is consistent with the low baseline levels of state capacity and service delivery in contexts like Mexican municipal governments, where local corruption is substantial (Chong et al., 2015; Larreguy, Marshall and Snyder, 2018). The model thus highlights how the short-term unobservability of state capacity investments constrains such investments and ultimately limits service delivery.

We then analyze the effect of a program that aims to publicly certify the effects of state capacity investments on overall service delivery by municipalities. We model the certifier as a third party endowed with the unilateral right to certify that the municipal budget has been partly spent on state capacity investments. Third parties, however, are corruptible with a certain probability. We first show that the certification of state capacity investments, which naturally increases with program availability, is increasing in the level of corruptibility of the third-party. We then show that, while certification cannot reduce the overall service delivery of municipalities *on average*, it has important redistributive consequences across municipalities that adopt the program. Specifically, we show that the effect of certification on overall service delivery is lower in municipalities where incumbents are certified by third parties that are likely to be corrupt, and where municipal incumbents are likely to be corrupt.

We assess the empirical relevance of this theory by testing the model's empirical predictions in the context of the implementation of Mexico's *From the Local Agenda* (Agenda desde lo Local, ADL) program. The ADL program was first implemented by the federal government in 2004, in collaboration with state governments, following the United Nations Local Agenda 21—an action plan to promote sustainable development by strengthening the institutional capacity of local governments. The ADL program consists of four main stages: (1) self-diagnosis by municipal government officials across 39 indicators of state capacity and public service delivery, which can be designated a red, yellow, or green status; (2) third-party verification of this diagnosis by a local institution of higher education, which results in the municipality receiving a certificate for having

already achieved green status on any given indicator; (3) time for municipal government officials to invest in improving non-green areas; and (4) updated self-diagnosis and third-party verification, again resulting in the granting of certificates for each new indicator that receives green status in the municipality.

We identify the effects of the ADL certification program by using a generalized difference-in-differences design to leverage temporal variation in take-up of the program. To avoid comparing municipalities that took up the ADL program with those that never did it, we focus only on the sample of municipalities that ever decided to participate in the program. We proxy for the likelihood that a certifying third-party is corrupt with municipal political alignment with the state government, based on the fact that the state-dependent institutions of higher education that serve as third-party auditors have been involved in various cases of corruption . We further use low baseline levels of certification upon entry into the program to proxy for corrupt incumbents. As our theoretical model highlights, municipal incumbent corruption should be associated with low state capacity and public service delivery. Empirically, we cannot differentiate between state capacity investments and public service delivery, and thus focus on the combination of the two—as measured by ADL program indicators—as our primary outcome.

The results are largely in line with the empirical predictions of our theory. First, we observe increased certification in municipalities politically aligned with the state government, which validates our use of such an alignment as a proxy for the likelihood that a certifying third-party is corrupt. Second, we find no discernible effect of certification on overall public service delivery on average, which is consistent with a high fractions of corrupt certifying third parties and municipal incumbents. Third, the results indicate that certification led to overall lower public service delivery in municipalities aligned with state government, and where baseline certification levels were low.

These findings advance the literature on the determinants of state capacity by considering the effectiveness of programs designed to encourage long-dwindling investments in state capacity. This literature initially focused on whether and how inter- and intra-state conflicts, as well as pop-

ulation density, have contributed to fiscal state capacity in Europe (Tilly 1992; Gennaioli and Voth 2015), Africa (Herbst 2000; Sanchez de la Sierra 2017; Thies 2007), and Latin America (Centeno 1997; Thies 2005). In contrast with extant studies seeking to explain the impetus for investments in capacity, we study the role that certifying programs aimed at strengthening local bureaucratic capacity to improve service delivery can play in creating political incentives for investments in state capacity (see also Soifer 2015; Fergusson, Larreguy and Riaño 2018). Like Banerjee, Duflo and Glennerster (2008) and Raffler (2018), we also highlight the difficulty of designing incentive structures to prevent manipulation—in our case, of a political form.

Our paper also relates to previous work showing that varying the visibility of policy outcomes to voters can generate policy distortions (Mani and Mukand 2007; Marx 2018). Marx (2018) finds that incumbents in Sub-Saharan Africa are rewarded only for completing projects in visible sectors, namely projects providing basic infrastructure and social services, which leads incumbent only to focus on the completion of visible projects before elections. We highlight that programs aimed at strengthening local state capacity, whose implementation might be more visible than outcomes of policies that directly benefit voters, might actually lead to the crowd out of such policies, especially in contexts where significant program design flaws are present.

Lastly, our paper adds empirical rigor and theoretical clarity to extant research casting doubt on the effects of the ADL program on municipal development. In particular, Pérez Archudia and Arenas Aréchiga (2012) report the lack of a significant positive correlation between actual municipal development and ADL program assessments of municipal government capacity for good governance and municipal development. Pérez Archudia and Arenas Aréchiga (2012) conclude by questioning how well the ADL program assessments capture municipal government capacity and municipal development. We similarly find a weak association between program certification and state-capacity and public-service-delivery outcomes, when implementing a plausible identification strategy. We additionally find evidence supporting that this finding is driven by adverse effects of certification when certifying third parties and municipal incumbents are corrupt.

2 Theoretical model

We develop a theory that highlights the short-term unobservability of state capacity investments as an impediment to such investments and thus service delivery. We first characterize the equilibrium outcome in the absence of a state capacity investment certification program. Second, we characterize the equilibrium outcome when there is a program certifying state capacity investments. Third, we assess the effect of the certification program on overall service delivery by certified municipalities.

2.1 Agents, actions and preferences

We consider a stylized two-period model for a representative municipality. In every period t , the incumbent politician of type τ , which we discuss later, allocates a unit of budget entirely into: (i) current service delivery, $g_t^\tau \in \{0, 1\}$; (ii) investing in improving state capacity, $s_t^\tau \in \{0, 1\}$; or (iii) public-rent appropriation, $r_t^\tau \in \{0, 1\}$. Although such allocations are less stark in practice, they highlight the core theoretical logic.

A homogeneous set of voters decide whether to reelect their incumbent politician or vote for an alternative candidate in between period one and two, $v \in \{I, A\}$. In the baseline model, voters can only observe whether the incumbent spent the entire budget on current public service delivery. In other words, voters cannot tell whether their incumbent appropriated the budget funds ($r_1 = 1$) or invested in state capacity ($s_1 = 1$) if $g_1 = 0$.

Voters derive additive utility from public service delivery. Specifically,

$$U(g_t, s_t) = \sum_{t=1,2} s_t + \beta_\tau g_t,$$

where $\beta_t \in \{\bar{\beta}, \underline{\beta}\}$ determines the value of public services to voters, such that $\bar{\beta} > \underline{\beta} > 1$. While

voters start with $\beta_1 = \underline{\beta}$, the law of motion of β_2 is as follows:

$$\beta_2 = \begin{cases} \underline{\beta} & \text{if } s_1 = 0, \\ \bar{\beta} & \text{if } s_1 = 1. \end{cases}$$

In other words, while investments in state capacity entail a short term loss in public service delivery (since $\underline{\beta} > 1$), they increase the efficiency of future public service delivery (since $\bar{\beta} > \underline{\beta}$). Whether state capacity investments increase overall service delivery depends on whether the benefit, $\bar{\beta} - \underline{\beta}$, exceeds the opportunity cost $\underline{\beta} - 1$. For the model to be interesting, we impose the following condition:

Assumption 1. $1 + \bar{\beta} > 2\underline{\beta}$.

There are two types of incumbent politicians, $\tau = \{h, c\}$. With probability γ the incumbent is honest ($\tau = h$), and with probability $1 - \gamma$ the incumbent is corrupt ($\tau = c$). Both types receive per-period rents from office, $R > 0$. Honest politicians also share voters' preferences over public service delivery, but do not derive utility from public-rent appropriation. In contrast, corrupt politicians derive no utility from public service delivery, but do enjoy public funds appropriation.

2.2 Timing

The timing of the game is then as follows.

1. Nature draws the incumbent politician type, $\tau \in \{h, c\}$.
2. The incumbent politician of type τ selects policy $\{g_1^\tau, s_1^\tau, r_1^\tau\}$ in period $t = 1$.
3. Voters observe g_1 and decide whether to reelect their incumbent politician, $v \in \{I, A\}$.
4. If an incumbent politician is not reelected, nature draws a new incumbent politician type in period two.

5. The (possibly new) incumbent politician selects $\{g_2^\tau, s_2^\tau, r_2^\tau\}$ in period $t = 2$.
6. All utilities are realized and the game ends.

2.3 Equilibrium without a certification program

In the absence of a certification program, and assuming that voters play their welfare-maximizing equilibrium, there is a unique perfect Bayesian equilibrium, whose outcome depends on the probability γ that the municipal incumbent is honest. The following Proposition 1 characterizes this equilibrium.

Proposition 1. *Assume that $1 + \bar{\beta} > 2\underline{\beta}$ and that voters play their welfare-maximizing equilibrium, and denote $\gamma^* = \frac{\underline{\beta}}{1 + \bar{\beta} - \underline{\beta}}$. If $\gamma \geq \gamma^*$, corrupt incumbents choose policy vectors $\{g_1^c, s_1^c, r_1^c\} = \{g_2^c, s_2^c, r_2^c\} = \{0, 0, 1\}$, honest incumbents choose policy vectors $\{g_1^h, s_1^h, r_1^h\} = \{0, 1, 0\}$ and $\{g_2^c, s_2^c, r_2^c\} = \{1, 0, 0\}$, and voters reelect their incumbent politicians in between period one and two upon observing $g_1 = 0$. If instead $\gamma < \gamma^*$, both incumbent types choose policy vector $\{g_1, s_1, r_1\} = \{1, 0, 0\}$ in period one, corrupt and honest incumbents respectively choose policy vectors $\{g_2^c, s_2^c, r_2^c\} = \{0, 0, 1\}$ and $\{g_2^h, s_2^h, r_2^h\} = \{1, 0, 0\}$ in period two, and voters always reelect their incumbent politicians in between period one and two. Voters' expected utility, before the realization of τ , is then given by:*

$$E[U(g_t, s_t)] = \begin{cases} (1 + \gamma)\underline{\beta} & \text{if } \gamma < \gamma^*, \\ \gamma(1 + \bar{\beta}) & \text{if } \gamma \geq \gamma^*. \end{cases}$$

Proof. See Appendix, where we also define the off-equilibrium strategies and beliefs that complete the characterization of the Perfect Bayesian equilibrium. □

Proposition 1 show that voters are willing to risk allowing corrupt politicians to appropriate resources in the first period to generate investments in state capacity that would increase overall

public service delivery when the incumbent is honest. Intuitively, this occurs when the likelihood that their incumbent is honest is sufficiently high: $\gamma \geq \gamma^*$. Otherwise, voters only reelect incumbents that spend the entire budget into current public service delivery in period one. For ease of exposition, the following assumption restricts the characterization of the certification program to the empirically-relevant case where there is under provision of overall service delivery, because voters believe that the likelihood that their incumbent is honest is sufficiently low.

Assumption 2. $\gamma < \gamma^*$.

2.4 Equilibrium with a certification program

We next model the certification program as a third party that publicly certifies $a \in \{0, 1\}$ whether $s_1 = 1$ when $g_1 = 0$. Importantly, this third party is of two types, $\alpha \in \{H, C\}$. It is honest, $\alpha = H$, with probability ρ and corrupt, $\alpha = C$, with probability $1 - \rho$, and its type is only known to the municipal incumbent. Honest third parties report truthfully, $a = s_1$, but corruptible ones report $a = 1$ regardless.

Again assuming that voters play their welfare-maximizing equilibrium, there is a unique perfect Bayesian equilibrium where the outcome depends on the probability γ that the incumbent is honest and the probability ρ that the third-party that certifies state-capacity investments is honest. The following Proposition 2 characterizes this equilibrium.

Proposition 2. *Assume that $1 + \bar{\beta} > 2\underline{\beta}$, $\gamma < \gamma^*$, and that voters play their welfare-maximizing equilibrium, and denote $\gamma^{**}(\rho) = \frac{\underline{\beta} - \rho}{(1 + \bar{\beta} - \underline{\beta} - \rho)} < \gamma^*$, where $\frac{\partial \gamma^{**}}{\partial \rho} < 0$. If $\gamma \geq \gamma^{**}(\rho)$, municipal honest incumbents choose $\{g_1^h, s_1^h, r_1^h\} = \{0, 1, 0\}$ and $\{g_2^h, s_2^h, r_2^h\} = \{1, 0, 0\}$, corrupt politicians always chose $\{g_2^c, s_2^c, r_2^c\} = \{0, 0, 1\}$ in the second period and choose $\{g_1^c, s_1^c, r_1^c\} = \{0, 1, 0\}$ if the third-party is not corrupt ($\alpha = H$) and $\{g_1^c, s_1^c, r_1^c\} = \{0, 0, 1\}$ if the third-party is corrupt ($\alpha = C$), third parties always chose $a = 1$ and, only upon observing this, voters re-elect their incumbent politicians in between period one and two. If instead $\gamma < \gamma^{**}(\rho)$, both types of mu-*

nicipal incumbent choose policy vector $\{g_1, s_1, r_1\} = \{1, 0, 0\}$ in period one, corrupt and honest incumbents respectively choose policy vectors $\{g_2^c, s_2^c, r_2^c\} = \{0, 0, 1\}$ and $\{g_2^h, s_2^h, r_2^h\} = \{1, 0, 0\}$ in period two, and voters reelect their incumbent politicians in between period one and two, $v = 1$. Voters' expected utility, before the realization of τ , is then given by:

$$E[U(g_t, s_t)] = \begin{cases} (1 + \gamma) \underline{\beta} & \text{if } \gamma < \gamma^{**}(\rho), \\ \gamma(1 + \bar{\beta}) + (1 - \gamma)\rho & \text{if } \gamma \geq \gamma^{**}(\rho). \end{cases}$$

Proof. See Appendix, where we also define the off-equilibrium strategies and beliefs that complete the characterization of the Perfect Bayesian equilibrium. \square

Proposition 2 highlights that voters might benefit from the certification program, by transitioning to an equilibrium where both types make state capacity investments in the first period, only when the likelihood that both their incumbent and the third-party are honest is sufficiently high: $\gamma \geq \gamma^{**}(\rho)$. Otherwise, voters continue only to reelect incumbents that spend the entire budget into current public service delivery in period one, thereby selecting an equilibrium that discourages state capacity investments.

2.5 Effects of the certification program

We next turn to the empirically testable implications of the model, with Proposition 3 demonstrating that, while the average effect of the certification program of voter utility or overall service delivery in certified municipalities should be non negative, it should be decreasing in the likelihood that the certifying third parties are corrupt and lower for corrupt, as opposed to honest, incumbents.

Proposition 3. *The expected change in overall service delivery in municipalities where state capacity investments are certified, $a = 1$, is given by:*

1. *Non negative,*

$$E[U(g_t, s_t) | a = 1] - E[U(g_t, s_t) | a = 0] = \gamma(1 + \bar{\beta}) + (1 - \gamma)\rho - (1 + \gamma)\underline{\beta} \geq 0,$$

2. *Increasing with ρ ,*

$$\frac{\partial E[U(g_t, s_t) | a = 1] - E[U(g_t, s_t) | a = 0]}{\partial \rho} > 0,$$

and

3. *Lower for $\tau = c$, as opposed to $\tau = h$,*

$$E[U(g_t, s_t) | c = 1, \tau = c] - E[\Delta U(g_t, s_t) | c = 1, \tau = h] < 0.$$

Proof. See Appendix. □

The result in Proposition 3 indicates that certification by a third-party auditor leads to an increase in service delivery, on average. However, this effect masks the important redistributive consequences of certifying municipalities. In particular, its effect on overall service delivery is smaller in municipalities where third parties are likely to be corrupt, since third parties allow corrupt incumbents to claim that they are investing in state capacity but actually appropriate public funds; absent the certification program, they would have had to invest entirely in public service delivery in period one to get reelected. In contrast, third parties that are honest do not allow for this behavior by corrupt incumbents, who then do need to invest in state capacity to get re-elected. Similarly, the effect on overall service delivery is smaller in municipalities with corrupt incumbents than with honest incumbents. This result follows for various reasons. First, as just explained, the effect in municipalities with corrupt incumbents certified by corrupt third parties should be negative. Moreover, the effect is also negative in those municipalities, even when third parties are honest because

investments in state capacity are inefficient given that corrupt incumbents appropriate all public funds in the second period, and thus voters would have been better off with all public fund being invested in public service delivery in period one. Lastly, the effect on overall service delivery is positive in municipalities with honest incumbents since these are able to efficiently invest in state capacity to increase overall service delivery to voters.

3 Background

3.1 Mexican municipal governments and their poor institutional capacity

Mexico's federal system is divided into 31 states (and the Federal District of Mexico City), which contain around 2,500 municipalities. Municipalities are governed by mayors who are typically elected to three-year non-renewable terms.¹ Following major decentralization reforms in the 1990s (see *Wellenstein, Núñez and Andrés 2006*), municipal governments—the focus of this article—became the main responsible for the day-to-day provision of basic local infrastructure and local public services. These include local policing, roads, sewerage, and water. Municipalities also assist state and federal governments in the provision of other public services including elementary education, health services, and environmental protection, among others.

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

¹Re-election is possible for incumbents in some states starting in 2018.

In part due to the lack of capacity to generate revenues, Mexican municipalities often also lack the institutional capacity to effectively deliver basic public services and manage local infrastructure. With the exception of large urban municipalities, most lack procedures for the provision and management of local public services, have low tax-collection capacity, lack trained officials, and are reluctant to politicize their administrative functions, among other things (Pérez Archudia and Arenas Aréchiga 2012). This situation significantly deteriorated with the escalation of drug trade-related violence in Mexico since 2007 (Dell 2015; Durante and Guitérrez 2015). According to the head of legal affairs at the Interior Ministry (Secretaría de Gobernación, SEGOB), 75% of Mexican municipalities are susceptible to infiltration and corruption by organized crime since they have little or no tools to combat criminal influence.² The institutional weakness of municipal governments is highlighted by both government officials and researchers as the underlying cause for the spread of organized crime (Aguirre and Herrera 2013).³

3.2 The *From the Local Agenda* program

The *From the Local Agenda* (Agenda desde lo Local, ADL) program, which is now called the Municipal Development Agenda (Agenda para el Desarrollo Municipal, ADM),⁴ was motivated by the desire to strengthen municipal state capacity, improve service delivery, and ultimately facilitate local development. The ADL program specifically aimed to support sustained and inclusive municipal economic and environmental development. The program has been administered in Mexico since 2004 by the Interior Ministry through the National Institute for Federalism and Municipal Development (Instituto Nacional para el Federalismo y el Desarrollo Municipal, INAFED). The INAFED developed and implemented the ADL program in line with Agenda 21, an action plan of

²See InSight Crime, “75% of Mexico Municipalities Susceptible to Organized Crime: Official,” January 23rd 2015.

³According to the Minister of Interior this institutional weakness is behind the spread of organized crime. See SinEmbargo, “Osorio Chong reconoce que hay “debilidad” institucional en municipios,” February 13th 2015.

⁴In 2014 the ADL program was revised and modified to include more aspects related with the role of municipal authorities in economic and social development, as well as environmental sustainability. See <http://www.agendaparaeldesarrollomunicipal.gob.mx/>.

the United Nations designed to promote sustainable economic, social, and environmental development in the 21st Century.⁵

The INAFED has implemented the ADL program together with, and largely through, state governments. Participation by municipal governments is voluntary. The program consists of four stages: self-diagnosis, third-party verification, improvement in under-performing areas, and update and certification. In the first stage of the program, municipal governments—aided by state governments—self-diagnose their institutional capacity for service delivery and actual service delivery. Specifically, they assess their capacities and performance across 39 indicators, including 270 separate sub-indicators, which are grouped into four areas: (1) institutional capacity for good governance; (2) sustainable economic development; (3) inclusive social development; and (4) sustainable environmental development. During the self-diagnosis stage, municipal governments assign themselves a red (completely undesirable situation and dramatic room for improvement), yellow (some room for improvement), or green (acceptable situation) status for each indicator and sub-indicator based on the sub-indicator specific guidelines stipulated by INAFED.

In the second stage of the program, the self-diagnosis by municipal governments is subject to third-party verification, usually arranged by state governments. The third parties have to be institutions of higher education, usually public or private local universities or institution of tertiary education. The use of these institutions was supposed to ensure that verification was perceived as neutral and objective by government officials and citizens. However, the credibility of such institutions is challenged by the fact that institutions of higher education are largely funded by the federal and state governments. Especially when there is alignment between the incumbent parties at the municipal and state levels, governors may seek to manipulate third-party certification to benefit the reputation of their co-partisans and that of their party more generally. Moreover, there are countless

⁵This is a product of the United Nations Conference on Environment and Development, also known as Earth Summit, held in Rio de Janeiro, Brazil, in 1992. Section 3 and chapter 28 of the Agenda 21 embody the well-known Local Agenda 21 stating that local authorities are essential to promote sustainable development. See <https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf>.

instances of staff, including faculty and students, from institutions of higher education engaging in corruption.⁶ A recent corruption scandal resulting in the diversion of approximately USD 400 million of public funds involves the federal government and eleven institutions of higher education, four of which worked as third party verifiers: Universidad Autónoma del Estado de México, Universidad Autónoma del Estado de Morelos, Universidad Politécnica del Golfo de México, and Universidad Tecnológica de Tabasco.⁷

The faculty and students from those institutions who act as verifiers receive training on indicators and corresponding criteria to be examined. They are responsible for reviewing the supporting documentation provided by municipal governments and validating their self-diagnosis for each indicator.⁸ Where verification worked best, auditor accounts still frequently indicated concerning flaws in the process, including instances of municipalities receiving a high status along many indicators and municipal officials selectively providing evidence to support each subindicator status. In many cases, the verification team simply had to trust the information provided by officials without being able to examine the self-diagnosis in greater detail or even look at the original data (Turrubiates Flores, Vargas Cuéllar and Suárez Rodríguez 2014).

In 2017, 1,164 municipalities in 30 states participated in the program, but only 863 concluded the verification process.⁹ 1,827 individuals—including faculty and students—from 163 institutions of higher education verified the self diagnosis done by municipal governments. Out of these 163 institutions, 99 (61%) are universities, 57 (35%) technological institutes, 5 (3%) local colleges, and 2 (1%) higher education institutes. The mean institution conducted slightly more than 5 verifications, while the median conducted 3.¹⁰

⁶For an example of the latter, Educación Futura, “Corrupción en un instituto tecnológico,” February 11th 2016 shows evidence that acceptance to the Technological Institute of Querétaro (Instituto Tecnológico de Querétaro, ITQ) was sold for USD 1,000 and faculty modified grades for USD 50 or sexual favors.

⁷See NYT, “‘El dinero se iba a un agujero negro:’ el esquema de corrupción que compromete al gobierno de México,” September 5th 2017 for more details.

⁸See the INAFED website for more details.

⁹These numbers exclude the municipalities which had all their indicators certified, as well as those which did not re-register into the program.

¹⁰See the INAFED website for more details.

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

In the fourth stage of the program, the From the Local National Council (Consejo Nacional Desde lo Local)—which is formed by representatives from the federal and state governments, as well as representatives from institutions of higher education—grants certificates to municipal government for each indicator that is deemed as green. These certificates are handed out by federal and state officials in award ceremonies, which are widely publicized by municipal governments and local media. These usually highlight not only the great work of municipal government officials, but the fact that results are subject to third-party auditing by local higher education institutions. It is also often mentioned that the municipal governments are certified using international standards.¹¹

4 Empirical implications

We next apply the empirical implications of the theoretical model in section 2 to the context of the ADL program just described. We note two issues when bringing the model to the data. First, the ADL program certified the status of municipal government capacity and service delivery on a three-point scale ranging from a completely undesirable situation (red) to an acceptable situation (green). Our simplified model, however, only captures a red or yellow versus a green comparison. Second, and more importantly, while the model distinguishes between investments in state capacity from service delivery, this distinction is far from clear across the indicators certified within the ADL program. As a result, we focus on the empirical predictions on overall service delivery, as captured

¹¹Tribuna de la Bahía, “Aprueba Puerto Vallarta los 270 indicadores de la Agenda para el Desarrollo Municipal,” October 18th 2016.

by voter utility in the theoretical model.

We begin with two hypotheses regarding the impact of being certified by the ADL program—that is to say receiving a certification of any status—on the extent to which municipal governments receive a higher status certification. These predictions, captured by Hypotheses 1 and 2, follow intuitively from an extension of the simple theoretical model in Section 2 to allow for a scale-, as opposed to discrete-, status certification.

Hypothesis 1 naturally follows from the nature of the program.

Hypothesis 1. *There is an increase in the status certified over time of the municipalities certified within the program.*

As our Section 3 suggests, the dependence of local universities on state government funding makes them susceptible to corruption as certifying third parties. We thus proxy for the likelihood that the third party certifying a municipal government is corrupt with whether a given municipal government is politically aligned with the state government. We then expect certified municipalities that are politically aligned to receive a higher status certification over time than those unaligned.

Hypothesis 2. *The over-time increase in the status certified is larger for municipalities politically aligned with the state government.*

We next focus on three hypotheses concerning the impact of being certified by the ADL program on the overall service delivery of municipal governments. First, Hypothesis 3 follows from Proposition 3.1:

Hypothesis 3. *Being certified by the ADL program has a non-negative effect on a municipality's overall public service delivery.*

Then, again leveraging whether the municipal government is politically aligned with the state government as a proxy for the likelihood that the third party certifying a municipal government is corrupt, Hypothesis 4 follows from Proposition 3.2.

Hypothesis 4. *The effect of being certified by the ADL program on a municipality's overall public service delivery is lower for municipalities politically aligned with their state government.*

Lastly, since the theoretical model highlights that corrupt municipal incumbents should be associated with a lower overall public service delivery, we denote municipalities with corrupt incumbents as those with low baseline levels of certification upon entry into the program. While imperfect, this is a reasonable proxy for underlying corruption. Hypothesis 5 then follows from this definition and Proposition 3.3.

Hypothesis 5. *The effect of being certified by the ADL program on a municipality's overall public service delivery is lower for municipalities whose certification status upon entry into the program is low.*

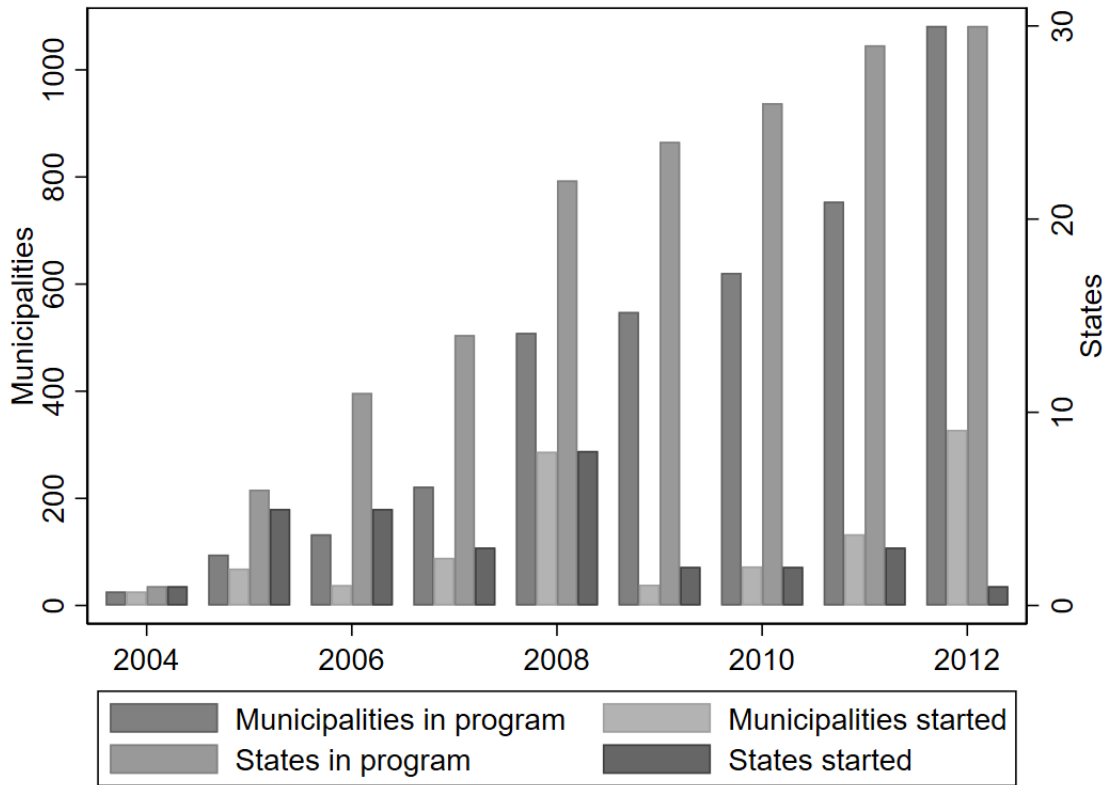
5 Research design

We next explain our identification strategy and how it allows us to test the empirical predictions of the model synthesized in the hypotheses enumerated in section 4, before describing the data that underpins it.

5.1 Identification strategy

To identify the effects of certification within the ADL program on state capacity outcomes, we exploit the staggered certification of Mexican municipalities that enroll in the ADL program using a generalized difference-in-differences design. Specifically, we examine the effects of a municipality being certified on certification status, as well as a range of state capacity and public service delivery measures, in the years prior to being certified to the years after, relative to municipalities that were certified at an earlier or later date. Given that many outcomes cannot be feasibly altered within months of joining the program, we define post-certification years starting with the

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



Notes: This plot shows the total number of municipalities participating in each year of the program as well as the number of municipalities that started the program in a given year.

year that certification results were first released for the municipality. Figure 1 shows the number of participating municipalities over the years of the program and the number of states that have municipalities participate in the program. Participation in the program seems to be driven to a large part on the state level with multiple municipalities from a given state joining at the same time. Figure 1 reveals that the number of municipalities that enter the program is correlated with states entering.

To test Hypothesis 1, which states that certified status should increase over time for the munic-

ipalities certified within the program, we estimate the following specification:

$$Y_{imt} = \beta_1 Program_{mt} + \eta_{its} + \theta_{im} + \epsilon_{imt}, \quad (1)$$

where Y_{imt} is the certified status for indicator i in municipality m in year t , and $Program_{mt}$ is an indicator for whether municipality m has been certified within the program by year t or not. We also include state-indicator-year fixed effects, η_{its} , and municipality-indicator fixed effects, θ_{im} , to capture the difference-in-differences design, and thus absorb all common shocks that could differ by state-indicator and time-invariant municipality-indicator characteristics. We cluster at the municipality level throughout. The main coefficient of interest is β_1 , which estimates the effect of program certification.

Hypothesis 2 predicts that the over-time increase in the status certified should be larger for municipalities politically aligned with the state government. We test this prediction by estimating the following specification:

$$Y_{imt} = \beta_1 Program_{mt} + \beta_2 (Program_{mt} \times 1[State Alignment_{tm} = 1]) + \eta_{its} + \eta'_{its} 1[State Alignment_{tm} = 1] + \theta_{im} + \theta'_{im} 1[State Alignment_{tm} = 1] + \epsilon_{imt} \quad (2)$$

where $State Alignment_{tm}$ indicates whether the party that governs the municipality also governs the state. The coefficient on β_1 estimates the effect of program certification on municipalities that are not aligned with the state government, while the coefficient on β_2 captures the differential effect of program certification for those municipalities aligned with the state government. The fixed effects are interacted with state alignment in order to exploit only variation within aligned and unaligned municipalities when estimating the effects of being certified.

To test Hypothesis 3, which predicts a non-negative effect of program certification on overall public service delivery, we estimate the same specification as in equation 1 but we consider actual

public service delivery, independently measured from the program (more details in Section 5.2, as an outcome instead. Similarly, we test Hypothesis 4, which predicts that such an effect should be lower for municipalities politically aligned with their state government, by estimating Equation 2 but actual public service deliver an outcome instead.

Hypothesis 5 predicts that the effect of being certified by the ADL program certification on overall public service delivery should be lower for municipalities whose certification status upon entry into the program is low. We test this prediction using the following specification:

$$Y_{imt} = \beta_1 Program_{mt} + \beta_2 (Program_{mt} \times 1[LowBaseline_m = 1]) + \eta_{its} + \eta'_{its} 1[LowBaseline_m = 1] + \theta_{im} + \epsilon_{imt}, \quad (3)$$

where $LowBaseline_m$ indicates whether the municipality received a low certification on a given indicator in the first year of being certified in the program. The benchmark coefficient, β_1 , estimates the effect of program on municipalities certified as having the highest status when entering the program. The coefficient on β_2 , in turn, captures the differential effect of program certification when municipal incumbents receive a low certification status upon entry into the program.

We require a parallel trends assumption to plausibly identify the effects of the certification program on outcomes. Specifically, this entails that municipalities that enter the program earlier would have otherwise followed the same trend as municipalities that entered the program later. Common approaches to test for this assumption are to: (i) conduct an event study plot by running the following specification; and (ii) formally test the parallel trends assumption by including k of leads of $Program_{mt}$ according to specifications of the following form

$$Y_{imt} = \sum_{\tau=0}^k \beta_{1\tau} Program_{mt+\tau} + \eta_{it} + \theta_{im} + \epsilon_{imt}, \quad (4)$$

where $Program_{mt+\tau}$ is an indicator for municipality m being certified within the program τ years

into the future. We use up to 3 leads (i.e. $k = 3$).

5.2 Data

Data on participation in the ADL program and the certifications come from the website of the National Institute for Federalism and Municipal Development (INAFED).¹² The data allows us to know 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. INAFED also provide the certified status that each municipality received in each of the 39 program indicators.

Data on state capacity and public service delivery outcomes, which are associated with the 39 indicators certified within the ADL program, comes from three sources, which are independent of the program and vary in the years in which they are available. First, we use an original data set on municipal state capacity and service delivery. Between 2000 and 2013, the National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía, INEGI) surveyed municipal governments 6 times. Using this data we were able to measure municipal institutional capacity and service delivery over time. For each survey wave, we have detailed measures of every municipality's personnel (by age, education, and department), resources (number of computers, vehicles, etc.), public good provision, existence of regulations, and more. As far as we are aware, this panel data set is the first of its kind, since local governments are rarely surveyed repeatedly in such detail. It is ideal for this study as it allows us to investigate changes in institutional capacity and public service delivery over time as a consequence of program certification. Importantly, the INEGI was not involved in the ADL program at any stage and surveyed municipal governments independently from ADL program activities. For each indicator, we study the sub-indicator used to evaluate the municipalities and link them to outcome variables in our data set. In many cases, we are able to find exact matches with the evaluation criteria of the program. We are then able to

¹²<https://www.gob.mx/inafed/acciones-y-programas/resultados-historicos-del-programa-agenda-desde-lo-local>

code whether a municipality actually satisfied the conditions to obtain red, yellow, or green for a given sub-indicator.

Second, some certification indicators relate to municipal finances (see 1.1, 1.5, 1.11 in Table 1). The municipal public finance data published by the INEGI provides detailed information on many of the outcomes of the finance sub-indicators.¹³ Third, since one indicator is related to housing (see 3.8 in Table 1) and several others to public service delivery to Mexican households, we exploit information from the quinquennial population census conducted by the INEGI.¹⁴ 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, which overlap closely with several sub-indicators contained in the ADL program. As with the municipal survey data, the municipal public finance and census data are collected independently from ADL program activities.

These three data sources produce 67 variables that link closely with 15 of the ADL program's 39 indicators. Table A3 in the appendix explains, for each indicator, which variables we use as outcomes, what data source they are from, and how they are coded. Table 1 shows the summary statistics for all outcome variables by indicator. For the analysis, the variables for each indicator are combined into a z -score.

6 Results

We begin by briefly showing that the certified status that municipalities receive is increasing over time and that this effect is larger when the municipal incumbent is politically aligned with the state government. Our main contribution, however, is to focus on the *actual metrics* of state capacity and public service delivery that we can measure independently from the program, which in principle should underpin the certified status but may in practice differ radically if the ADL

¹³See more details in the INEGI website.

¹⁴See more details in the INEGI website.

Table 1: Summary Statistics by Indicator

Indicator	Variable	Mean	Standard Deviation	Min	Max
1.1	Percentage women of total employees	0.27	0.10	0.00	0.87
	Does the mun. have an evaluation sector?	0.26	0.44	0.00	1.00
	Percentage of administrative personnel of total employees	0.67	0.18	0.00	1.00
	Own income / Current expenditure	0.47	0.50	0.00	11.87
	Current expenditure / Total expenses	0.71	0.10	0.11	1.00
	Public investment / Total income	0.30	0.16	0.00	1.00
	Personal services / Current expenditure	-1.20	0.72	-37.13	-0.00
1.2	Total expenses + Total revenue	0.13	0.27	-3.10	5.21
	Partnership with other municipalities	0.32	0.47	0.00	1.00
1.4	Contributions to intermunicipal agreements	0.40	0.96	0.00	6.00
	Sector for promoting social participation?	0.30	0.46	0.00	1.00
1.5	Participation of commissions and/or communal committees	0.33	0.47	0.00	1.00
	Index for regulations for participation	0.10	0.42	0.00	2.00
	Are there mechanisms for citizens participation?	0.84	0.37	0.00	1.00
	Whether the collection of property tax is done by mun.	0.94	0.20	0.00	1.00
1.6	Update of property values in the last 2 years	0.73	0.44	0.00	1.00
	Own income / Total income	0.12	0.12	0.00	1.00
	Public debt / Total expenses	0.96	0.05	0.19	1.00
	Changes in property tax	2.51	82.08	0.00	7,907.05
	Changes in Derechos + Productos + Aprovechamientos	1.68	12.17	0.00	970.34
1.7	Is there a plan for civil protection?	0.72	0.45	0.00	1.00
	Is there a map of risk zones?	0.49	0.50	0.00	1.00
	Index for regulations on civil protection	0.36	0.75	0.00	2.00
1.8	Total number of computers / total number of employees	0.19	0.17	0.00	1.00
	Does the municipal office have an Internet connection?	0.71	0.45	0.00	1.00
1.9	Is there a juridica?	0.69	0.46	0.00	1.00
	What percentage of sectors have regulations	0.40	0.30	0.00	1.00
	Index of regulations	0.50	0.77	0.00	2.00
1.10	Is there a justice office?	0.55	0.50	0.00	1.00
	Is there a juridica?	0.69	0.46	0.00	1.00
1.11	Institution responsible for transparency?	0.56	0.33	0.00	1.00
	Regulations to regulate access to public info.?	0.54	0.50	0.00	1.00
	Public servants responsible for public info.?	0.67	0.47	0.00	1.00
	A system of reception of and attention to public info. requests	0.53	0.50	0.00	1.00
	A system for archives	0.32	0.47	0.00	1.00
	Training program for public servants on public info.	0.33	0.47	0.00	1.00
	Is there open access?	0.85	0.36	0.00	1.00
	Are there regulations about transparency?	0.38	0.48	0.00	1.00
	Debt accumulated from previous years	-2.27	6.66	-21.43	2.30
	Percentage of budgeted contributions collected (Ordinal)	2.73	1.33	1.00	5.00
2.1	Federal participations / Total income	0.55	0.16	0.00	1.00
	Is there someone responsible for economic development?	0.49	0.50	0.00	1.00
3.1	Are there regulations for economic development?	0.25	0.44	0.00	1.00
	Percentage of mun. capital covered by drainage and sewage	0.79	0.27	0.00	1.00
	Percentage of rest of mun. covered by drainage and sewage	0.52	0.34	0.00	1.00
	Percentage of mun. capital covered by public lighting	0.82	0.25	0.00	1.00
	Percentage of rest of mun. covered by public lighting	0.63	0.32	0.00	1.00
	Percentage of mun. capital covered by street cleaning	0.81	0.27	0.00	1.00
	Percentage of rest of mun. covered by street cleaning	0.59	0.35	0.00	1.00
	Percentage of mun. capital covered by trash collection	0.85	0.24	0.00	1.00
	Percentage of rest of mun. covered by trash collection	0.62	0.34	0.00	1.00
	Are there grave regulations?	0.42	0.49	0.00	1.00
3.2	Are there market regulations?	0.41	0.49	0.00	1.00
	Do regulations on performance and sport exist?	0.35	0.48	0.00	1.00
3.6	Index for regulations on performance and sport	0.24	0.64	0.00	2.00
	Percentage of mun. capital covered by drinking water	0.84	0.22	0.00	1.00
3.8	Percentage of rest of mun. covered by drinking water	0.64	0.31	0.00	1.00
	Percentage of mun. capital covered by drainage and sewage	0.79	0.27	0.00	1.00
	Percentage of rest of mun. covered by drainage and sewage	0.52	0.34	0.00	1.00
	Percentage of mun. capital covered by drinking water	0.84	0.22	0.00	1.00
24	Percentage of rest of mun. covered by drinking water	0.64	0.31	0.00	1.00
	Percentage of mun. capital covered by drainage and sewage	0.79	0.27	0.00	1.00
	Percentage of rest of mun. covered by drainage and sewage	0.52	0.34	0.00	1.00
	Occupants in homes with drainage and / or toilet	0.89	0.13	0.20	1.00
	Occupants in homes with electric power	0.95	0.08	0.10	1.00
	Occupants in houses with dirt floor	0.85	0.16	0.04	1.00
	Housing without overcrowding	0.53	0.13	0.14	0.89
	Observations	14,898			

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

Table 2: Summary Statistics

	Mean	Standard Deviation	Min	Max
Program Start Year	2009	2.59	2004	2013
Program	0.35	0.48	0.00	1.00
Low baseline	0.78	0.42	0.00	1.00
Intensity	1.46	0.83	0.00	2.00
Aligned with State	0.58	0.49	0.00	1.00
Observations	125126			

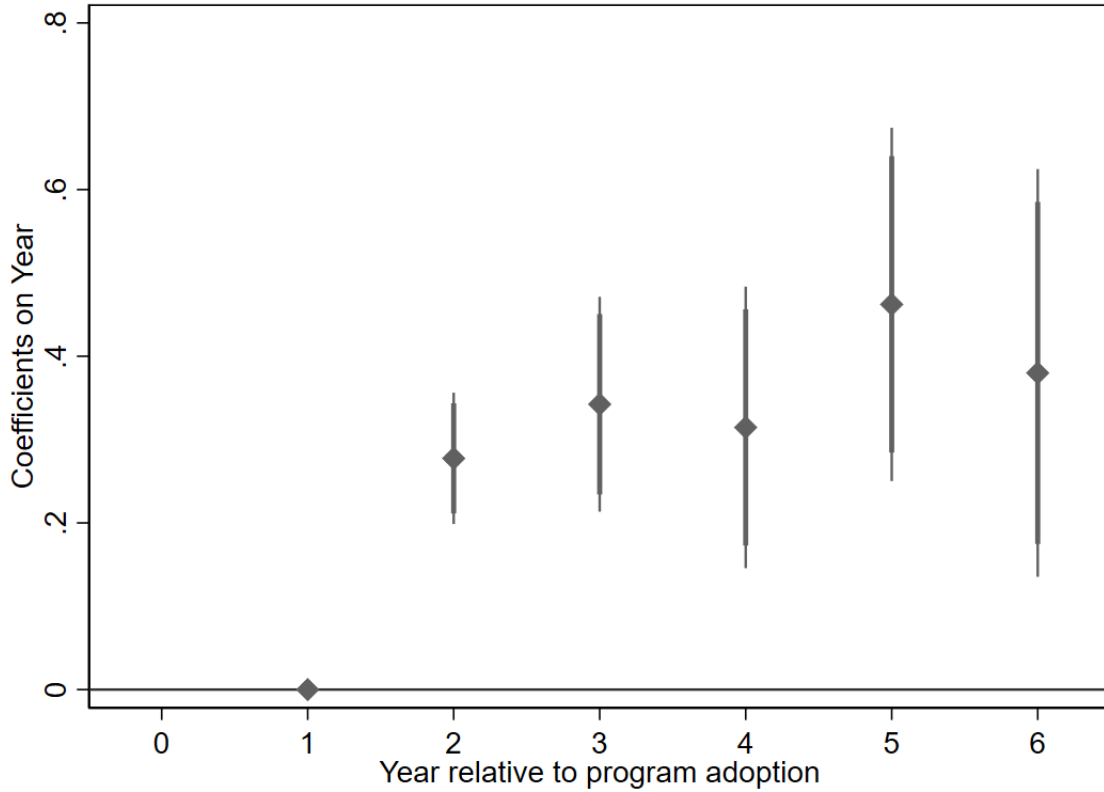
certification program is corrupted. When we examining these actual performance outcomes, we find that participation in the program generally does not increase municipalities' performance on governance outcomes related to program indicators. We then show that, consistent with the empirical predictions of our theoretical model, overall public service delivery of certified municipalities decreases with the political alignment of the municipal and state governments, and that the effect of the program on public service delivery of certified municipalities is weaker in places with low initial certified status.

6.1 Effect of the ADL program on certified status over time

We first examine Hypotheses 1 and 2 concerning the effect of being certified by the program on the certified status of participating municipalities. Table 3 shows the results of estimating equation (1), where the outcome is an ordinal scale for whether a municipality was certified as green (coded as 3), yellow (coded as 2), or red (coded as 1). Column (1) focuses on all program indicators, while columns (2) and (3) restrict to the indicators for which we have independent state capacity and public service delivery measures. Consistent with Hypothesis 1, columns (1) and (2) indicate that the program led to a substantial increase in the certified status over time among participating municipalities. The effect in column (2) is sizable, with the certification status increasing on average by 6% over the time of the program. Figure 2 illustrates this effect graphically over the duration of

a municipality's post-certification participation in the program.¹⁵

Figure 2: Average Effect on Certified Status



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

Column (3) in Table 3 shows heterogeneous effects by political alignment between the municipal and state governments, which recall we use as a proxy for the likelihood of corruptibility of the certifying third party. Consistent with Hypothesis 2, the results support that the effect estimated in column (2) is increasing with such an alignment. While the effect is marginally statistically insignificant, it is sizable. Namely, the effect of the program is 67% larger in municipalities politi-

¹⁵Note that we are unable to include the values prior to the municipal adoption of the program in Figure 2 since, naturally, there are no certification outcomes before that.

cally aligned with the state government. Figure 3 shows graphically the effects of the program on the certified status of the program indicator by year since the year of program adoption. The plot on the left depicts the effect for municipalities not politically aligned with the state government, while the plot on the right portrays the differential effect for municipalities politically aligned with the state government.

Table 3: Average Effect on Certified Status and Heterogeneous Effects by State Alignment

	(1)	(2)	(3)
	All Indicators	Our Indicators Only	Our Indicators Only
Program	0.0600*** (0.0136)	0.107*** (0.0183)	0.0646* (0.0377)
Program \times Aligned with State			0.0435 (0.0438)
Observations	544900	115625	109237
R^2	0.832	0.822	0.857
Mean of Outcome	1.725	1.732	1.732
SD of Outcome	0.913	0.914	0.913
Min of Outcome	1	1	1
Max of Outcome	3	3	3
Program + Program \times Aligned			0.108*** (0.0249)

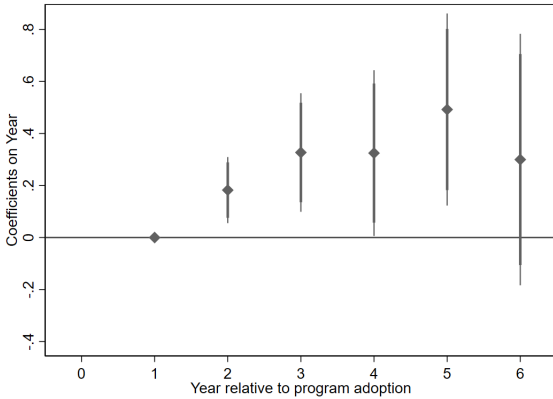
Notes: This table shows the regression results of the main specification using program certification as outcomes. An observation represents a program indicator in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-indicator and municipality-indicator fixed effects. Column (1) uses the certification status for all indicators of the program. Column (2) restricts to indicators for which we have independent measures. Column (3) also uses the restricted sample and adds an interaction term of whether the municipality is governed by the same party as its state. Note that the fixed effects in column (3) are also interacted with the state alignment. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.2 Effect of the ADL program on state capacity and public service delivery

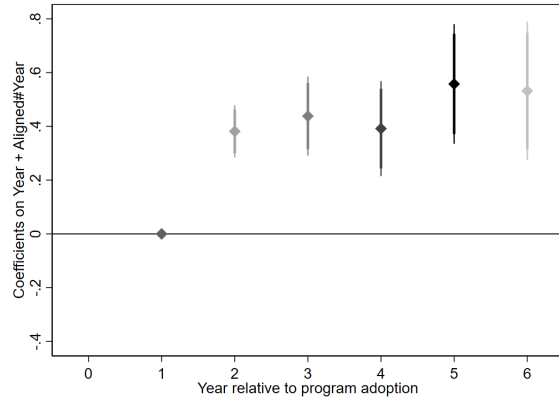
We next assess the effect of program certification on state capacity investments and public service delivery to test Hypotheses 3, 4, and 5. Table ?? considers as an outcome measures of program

Figure 3: Average Effect on Certified Status and Heterogeneous Effects by State Alignment

Panel A: Effect for Unaligned Municipalities



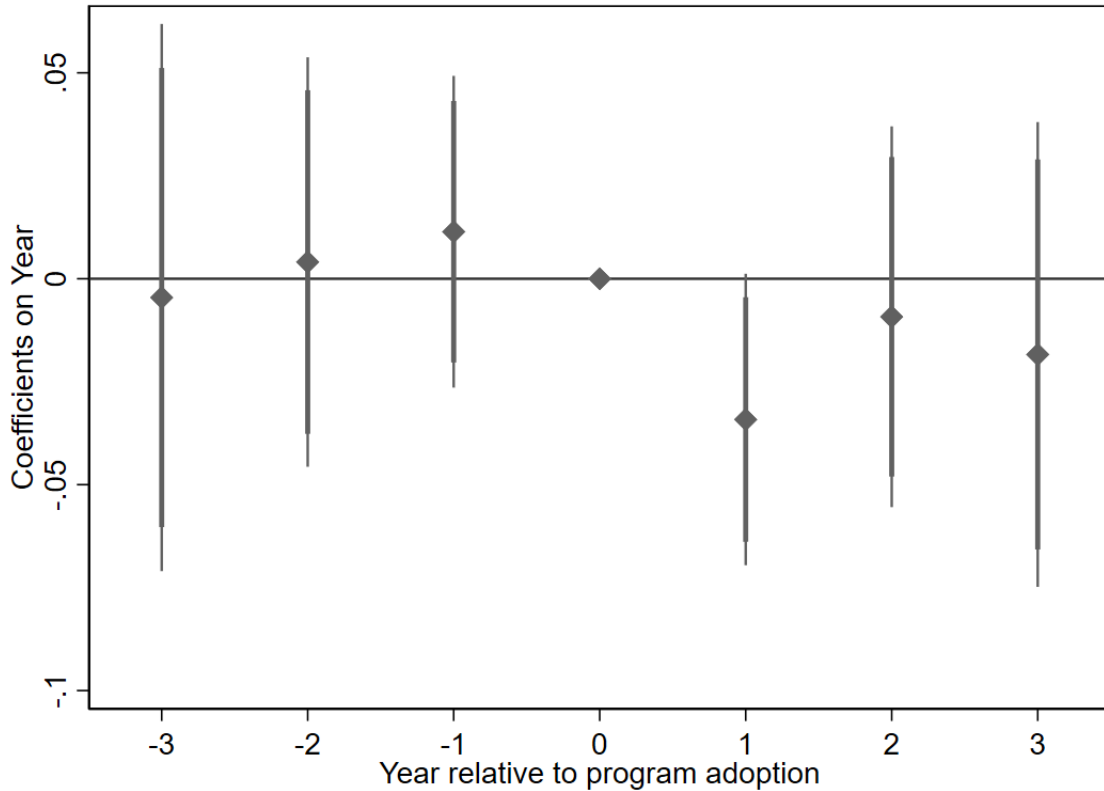
Panel B: Differential Effect for Aligned Municipalities



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

indicators constructed with data collected independently from the program. In contrast with Hypothesis 3, the null result in column (1) indicates no average positive effect of the program on the overall public service delivery of the certified municipalities. Columns (2) to (3) further show the absence of differential pre-trends in overall public service delivery of municipalities becoming certified earlier relative to those becoming certified later, thus lending support for the identification assumption of our generalized difference-in-differences design. These results are visually corroborated by Figure 4. *A priori*, the lack of an average positive effect could be considered inconsistent with Hypothesis 3, and consequently our theory. However, we note that the expected positive effect of being certified is expected to be small when the likelihoods of corrupt certifying third parties and municipal incumbents are high (i.e. ρ and γ are small, in the context of the model) and our specification includes state-year-indicator fixed effects for causal identification, which might capture part of the average effect of the program of certifying indicator. Moreover, all other empirical

Figure 4: Average Effect on Public Service Delivery



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

predictions of the model are supported by the data.

Table 4 focuses on the same outcome as Table ?? and considers two interactions with program certification. Column (1) reports the results on the interaction with political alignment between municipal and state governments, which Table 3 suggests is a reasonable proxy for the corruptibility of the certifying third parties. Consistent with Hypothesis 4, these results indicate that overall public service delivery of certified municipalities significantly decreases with state alignment. This effect can also be seen graphically in Figure 5, which reports the effects of the program on overall public service delivery of certified municipalities by year since the year of program adoption. The plot on the left depicts the effect for municipalities politically not aligned with the state govern-

ment, while the plot on the right shows the differential effect for municipalities politically aligned with the state government. Together with Table Appendix ??, Figure 5 also supports the absence of differential pre-trends in the overall public service delivery of municipalities that become certified earlier relative to those becoming certified later for both municipalities not aligned with the state government and municipality aligned.

Table 4: Interactions with State Alignment and Low Baseline Certification

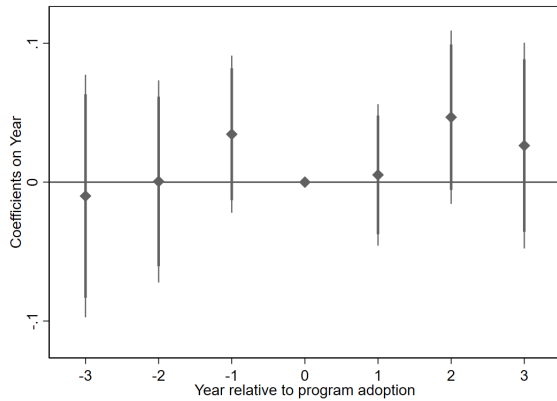
	(1)	(2)	(3)
	Z-Score	Z-Score	Z-Score
Program	0.0197 (0.0225)	0.0544 (0.0352)	0.0443 (0.0524)
Program \times Aligned with State	-0.0841*** (0.0317)		-0.0210 (0.0756)
Low baseline \times Program		-0.0837** (0.0366)	-0.0135 (0.0552)
Low baseline \times Program \times Aligned with State			-0.105 (0.0789)
Observations	119310	123910	116490
R^2	0.514	0.444	0.540
Mean of Outcome	0.00708	0.00331	0.00431
SD of Outcome	0.711	0.714	0.711
Min of Outcome	-7.252	-7.252	-7.252
Max of Outcome	11.70	11.70	11.70
Program + Program \times Aligned	-0.0644*** (0.0218)		0.0233 (0.0540)
Program + Program \times Low Baseline		-0.0292* (0.0153)	
Program \times Low Baseline + Program \times Low Baseline \times Aligned			-0.118** (0.0557)

Notes: This table shows the regression results of the main specification using program indicators of state capacity and overall public service delivery as outcomes. An observation represents a program indicator in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-indicator and municipality-indicator fixed effects. Column (1) interacts *Program* and the fixed effects with whether the municipality is governed by the same party as its state. Column (2) interacts them with whether the municipality received a low certification on the indicator in the first year of participating in the program instead. Column (3) considers the triple interaction instead. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

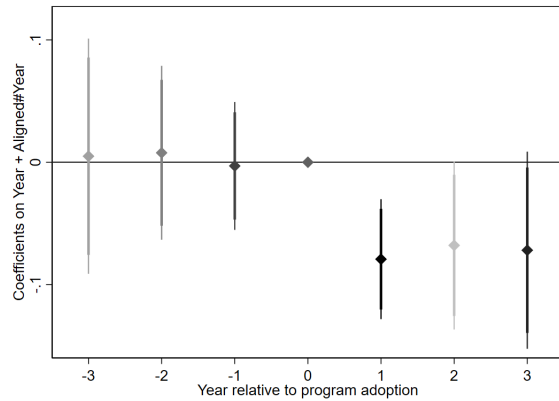
Column (2) in Table 4 reports the results for the interaction with a low initial certified status of

Figure 5: Average Effect on Service Delivery and Heterogeneous Effects by State Alignment

Panel A: Effect for Unaligned Municipalities



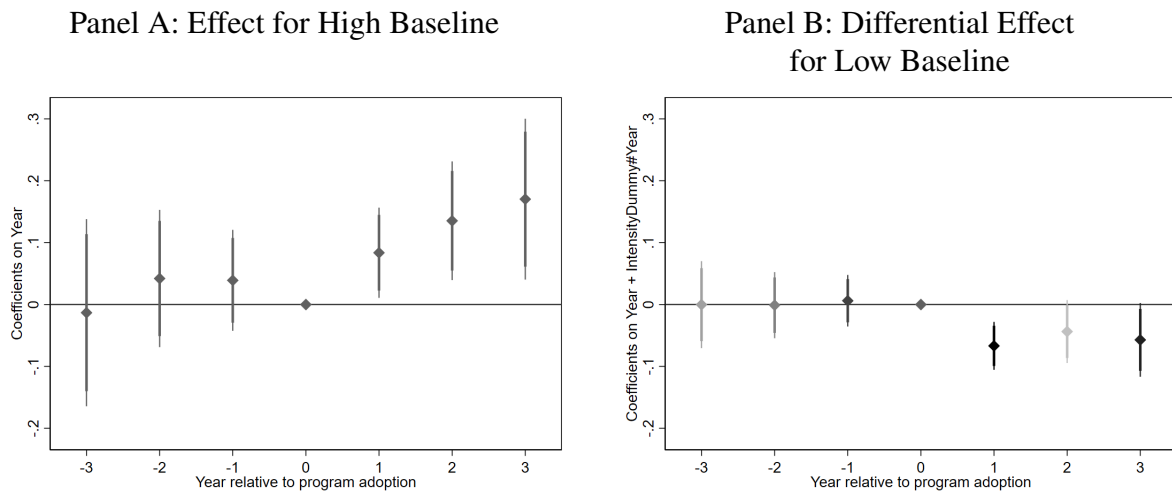
Panel B: Differential Effect for Aligned Municipalities



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

a given indicator, which we use as a proxy for the municipal incumbent being corrupt, as opposed to honest. Consistent with Hypothesis 5, these results indicate that the effect of being certified by the program on overall public service delivery significantly decreased in municipalities with a low baseline status certified. This effect can also be seen graphically in Figure 6, which reports the effects of the program on overall public service delivery of certified municipalities by year since the year of program adoption. The plot on the left depicts the effect for municipalities with a high baseline status certified, while the plot on the right shows the differential effect for municipalities with a low baseline status certified. The coefficients on those plots, together with those in Table ?? in the Appendix, also suggest that differential pre-trends in overall public service delivery of municipalities with a low baseline status certified cannot account for results in column (2) in Table 4.

Figure 6: Average Effect on Service Delivery and Heterogeneous Effects by Baseline Level of Certification



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

7 Robustness

We conduct several robustness exercises to ensure that our results are not driven by particular parameterizations of our regressors or and data quality issues. To begin with, we show in Table A1 that the results in Table 4, and therefore the empirical support for Hypotheses 4 and 5, are robust to using different measures of baseline level of certification status, namely a non-parametric and linear ones.

We then conduct several robustness exercises to ensure that our results are not influenced by coding decisions. We show in Table A2 that the result for column (1) of Table ?? and columns (1) and (2) of Table 4, and therefore the empirical support for Hypotheses 3, 4 and 5, are robust to alternative coding of the program indicators reflecting state capacity investments and public service delivery. Namely, we use four alternative recoding strategies. In column (1) we replicate our baseline measure that uses an index of the corresponding certification status in each of the subindicators following the program specifications when possible. Column (2) uses an index of an indicator for whether at least the first cutoff (from red to yellow) was passed for each of the subindicators. Column (3) uses an index of an indicator for whether the second cutoff (from yellow to green) was passed for each of the subindicators. Columns (4) instead uses the raw data used to code up the certification status in each of the subindicators winsorized at the 99th percentile. Column (5) winsorizes at the 95th percentile instead. These results are highly robust both in terms of significance and magnitude.

8 Conclusion

Our findings suggest that the short-term unobservability of state capacity investments represent an important impediment to investments in state capacity and service delivery. They also indicate that, at least in theory, the certification of state capacity investments and service delivery could help to

overcome such an impediment, but can also be undermined by existing institutional weaknesses. In particular, using a generalized difference-in-differences design, we show that a certification program of state capacity and service delivery had no effect on average overall service delivery. We find evidence suggesting that these negligible effects are largely explained by the corruption of both the third parties certifying state-capacity investments and service delivery and the municipal incumbents.

References

- Acemoglu, Daron. 2005. "Politics and Economics in Weak and Strong States." *Journal of Monetary Economics* 52:1199–1226.
- Aguirre, Jerjes and Hugo Amador Herrera. 2013. "Institutional weakness and organized crime in Mexico: the case of Michoacán." *Trends in Organized Crime* 16(2):221–238.
- Banerjee, Abhijit V., Esther Duflo and Rachel Glennerster. 2008. "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system." *Journal of the European Economic Association* 6(2-3):487–500.
- Besley, Timothy and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99(4):1218–1244.
- Besley, Timothy and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78(1):1–34.
- Castañeda, Luis César and Juan E. Pardini. 2012. "Sub-national Revenue Mobilization in Mexico." Inter-American Development Bank Working Papers, No. IDB-WP-354.
- Centeno, Miguel Angel. 1997. "Blood and Debt: War and Taxation in Nineteenth-Century Latin America." *American Journal of Sociology* 102(6):1565–1605.
- 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.
- Dell, Melissa. 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105(6):1738–79.

- Dell, Melissa, Nathan Lane and Pablo Querubin. 2017. “The Historical State, Local Collective Action, and Economic Development in Vietnam.”
- Durante, Ruben and Emilio Guitérrez. 2015. “Fighting Crime with a Little Help from My Friends: Party Affiliation, Inter-Jurisdictional Cooperation and Crime in Mexico.” Working Paper.
- Fergusson, Leopoldo, Horacio Larreguy and Juan Felipe Riaño. 2018. “Political Competition and State Capacity: Evidence from a Land Allocation Program in Mexico.” Working Paper.
- Gennaioli, Nicola and Hans-Joachim Voth. 2015. “State Capacity and Military Conflict.” *The Review of Economic Studies* 82(4):1409–1448.
- Herbst, J. 2000. *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton Studies in International History and Politics Princeton University Press.
- Larreguy, Horacio, John Marshall and Jr. Snyder, James M. 2018. “Publicizing Malfeasance: How Local Media Faciliates Electoral Sanctioning of Mayors in Mexico.” Working paper.
- 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.
- Michalopoulos, Stelios and Elias Papaioannou. 2013. “Pre-colonial Ethnic Institutions and Contemporary African Development.” *Econometrica* 81(1):113–152.
- Michalopoulos, Stelios and Elias Papaioannou. 2014. “National Institutions and Subnational Development in Africa.” *Quarterly Journal of Economics* 129(1):151–213.
- 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. 2018. "Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy." Working Paper.
- Sanchez de la Sierra, Raul. 2017. "On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo."
- Soifer, Hillel David. 2015. *State Building in Latin America*. New York: Cambridge University Press.
- Thies, Cameron G. 2005. "War, rivalry, and state building in Latin America." *American Journal of Political Science* 49(3):451–465.
- Thies, Cameron G. 2007. "The political economy of state building in Sub-Saharan Africa." *Journal of Politics* 69(3):716–731.
- Tilly, C. 1992. *Coercion, Capital and European States: AD 990 - 1992*. Studies in Social Discontinuity Wiley.
- Turrubiates Flores, Héctor Omar, Marco Iván Vargas Cuéllar and Irma Brígida Suárez Rodríguez. 2014. "Capacidades Institucionales y Problemas Recurrentes: Evidencias de las Verificaciones del Programa Agenda para el Desarrollo Municipal en San Luis Potosí." *Tlatemoani* .
- Wellenstein, Anna, Angélica Núñez and Luis Andrés. 2006. Social Infrastructure: Fondo de Aportaciones para la Infraestructura Social (FAIS). In *Decentralized service delivery for the poor, Volume II: Background papers*, ed. The World Bank. Mexico City: The World Bank pp. 167–222.

Table A1: Robustness of interaction

	(1) Z-Score	(2) Z-Score	(3) Z-Score	(4) Z-Score
Program	0.0544 (0.0352)	0.0452 (0.0334)	-0.156** (0.0789)	0.0437 (0.0503)
1.Intensity \times Program	-0.100* (0.0556)		0.0698 (0.0860)	
2.Intensity \times Program	-0.0824** (0.0368)		0.300*** (0.0892)	
Program \times c.Intensity		-0.0376** (0.0178)		-0.00657 (0.0273)
Program \times Aligned with State			0.101 (0.0837)	-0.0303 (0.0723)
1.Intensity \times Program \times Aligned with State			0.0103 (0.110)	
2.Intensity \times Program \times Aligned with State			-0.0893 (0.0984)	
Program \times Aligned with State \times c.Intensity				-0.0483 (0.0388)
Observations	123910	123910	106401	116490
R^2	0.444	0.444	0.903	0.540
Mean of Outcome	0.00331	0.00331	1.715	0.00431
SD of Outcome	0.714	0.714	0.912	0.711
Min of Outcome	-7.252	-7.252	1	-7.252
Max of Outcome	11.70	11.70	3	11.70
1.Intensity \times Post + 1.Intensity \times Post \times Aligned			0.0801 (0.0699)	
2.Intensity \times Post + 2.Intensity \times Post \times Aligned			0.210*** (0.0453)	
c.Intensity \times Post + c.Intensity \times Post \times Aligned				-0.0548** (0.0271)

Notes: This table shows the regression results of the main specification using program indicators of state capacity and overall public service delivery as outcomes. An observation represents a program indicator in a municipality in a year. The main independent variable, *Program*, is an indicator variable that equals one for each year after the municipality has entered the program. The specification includes state-year-indicator and municipality-indicator fixed effects. Column (1) interacts *Program* with a non-parametric version of the certified status of the indicator in the first year the municipality participated in the program, column (2) uses a linear version instead. Columns (3) and (4) use the same specifications as (1) and (2) but also include an interaction for whether the municipality is governed by the same party as the state. Standard errors, clustered at the municipality level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

Panel A: No Interaction		<i>Dependent variable: Z-Score</i>				
	(1)	(2)	(3)	(4)	(5)	
	Index	Dummy 1st cutoff	Dummy 2nd cutoff	Winsorized 99	Winsorized 95	
Program	-0.0135 (0.0146)	-0.0284** (0.0145)	-0.00756 (0.0143)	-0.0206 (0.0149)	-0.0207 (0.0149)	
Observations	124984	125000	125091	125024	125024	
R^2	0.425	0.408	0.425	0.461	0.462	
Panel B: Interaction with Alignment		<i>Dependent variable: Z-Score</i>				
	(1)	(2)	(3)	(4)	(5)	
	Index	Dummy 1st cutoff	Dummy 2nd cutoff	Winsorized 99	Winsorized 95	
Program	0.0197 (0.0225)	-0.00997 (0.0221)	0.0308 (0.0226)	-0.00263 (0.0231)	-0.00244 (0.0231)	
Program \times Aligned with State	-0.0841*** (0.0317)	-0.0651** (0.0312)	-0.0881*** (0.0314)	-0.0585* (0.0324)	-0.0590* (0.0324)	
Observations	119310	119327	119422	119352	119352	
R^2	0.514	0.500	0.514	0.547	0.547	
Panel C: Interaction with Low Baseline		<i>Dependent variable: Z-Score</i>				
	(1)	(2)	(3)	(4)	(5)	
	Index	Dummy 1st cutoff	Dummy 2nd cutoff	Winsorized 99	Winsorized 95	
Program	0.0544 (0.0352)	0.0503 (0.0352)	0.0565* (0.0338)	0.0496 (0.0377)	0.0500 (0.0376)	
Low baseline \times Program	-0.0837** (0.0366)	-0.0993*** (0.0370)	-0.0779** (0.0352)	-0.0878** (0.0391)	-0.0883** (0.0391)	
Observations	123910	123926	124018	123950	123950	
R^2	0.444	0.428	0.443	0.480	0.480	
FE	Year & Mun	Year & Mun	Year & Mun	Year & Mun	Year & Mun	
Cluster	Mun	Mun	Mun	Mun	Mun	
Mean of Outcome	0.00331	0.00273	0.00296	0.00446	0.00446	
SD of Outcome	0.714	0.705	0.714	0.732	0.732	
Min of Outcome	-7.252	-12.00	-4.719	-4.719	-4.719	
Max of Outcome	11.70	11.70	11.70	11.70	11.70	

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

Table A3: Linking indicators to outcomes

Indicator	Variable	Source	Coding
1.1 Administered responsibly	Percentage women of total employees	Municipal Census	Standardized (Mean 0, SD 1)
	Does the municipality have an evaluation sector?	Municipal Census	Standardized
	Percentage of administrative personnel of total employees	Municipal Census	Inversed and Standardized
	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%
1.2 Associated and linked	Whether the municipality has partnered with another municipality in the state for public good provision	Municipal Census	Standardized
	Has the municipality made contributions to intermunicipal agreements	Municipal Census	Standardized
1.4 Participatory	Whether the administrative structure has a sector for promoting social participation	Municipal Census	Standardized
	Whether the commissions and/or communal committees participate in the allocation	Municipal Census	Standardized
	Index for regulations for participation	Municipal Census	0 if no reg. 1 if updated 3-5 years ago, 2 if updated in last 3 years
	Are there mechanisms for citizens participation?	Municipal Census	Standardized
1.5 Fiscally responsible	Whether the collection of property tax is done by municipality	Municipal Census	Standardized
	Whether values used for the collection of property tax have been updated in the last 2 years	Municipal Census	Standardized
	Own income / Total income	Municipal Budget	Standardized
	Public debt / Total expenses	Municipal Budget	Inversed
	Property tax t / t-1	Municipal Budget	Winsorized at 99%
	Derechos + Productos + Aprovechamientos t/t-1	Municipal Budget	Winsorized at 99%

Linking indicators to outcomes

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

Linking indicators to outcomes

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