The Limits of Government Outsourcing: Property Tax Re-Assessments in India

Esther Duflo, Rema Hanna, Benjamin A. Olken, and Shreya Tandon*

November 2025

Abstract

Many theories of outsourcing presume private firms are more efficient producers of goods and services, but that contracting frictions mean that firms do not internalize the government's objective function. But perhaps, for some tasks, the government can actually be more productively efficient? We study this question through an experiment in Chennai, India, where the government randomly selected neighborhoods to have property tax assessments determined by private firms contracted by the government, rather than government tax inspectors. We show that inspections by government tax inspectors yielded almost double the increase in tax revenue as those done by private firms. They were also more accurate, as judged by comparing new assessments to independent assessments done by third-party surveyors we hired. Difficulties contracting with the government discouraged many competent firms from bidding. Perhaps as a result, government inspectors had higher skills and put in more effort than the workers at the private firms. Importantly, the government inspectors had more authority to enter properties, and soft knowledge on how to find properties, both of which may be hard to transfer even to higher quality firms. The results suggest that the limits to government outsourcing may go beyond multitasking issues.

*This project was a collaboration involving many people. We thank our many colleagues in the Government of Tamil Nadu and the Greater Chennai Corporation, particularly, Minister Palanivel Thiaga Rajan, J. Radhakrishnan (then chairman of GCC), Gagandeep Singh Bedi, Vishu Mahajan and R. Lalitha, as well as many colleagues from the Revenue Department and Information Technology Department of the Greater Chennai Corporation for their support with project implementation, data collection, and data transfers. We thank Youssef Assarssah, Clotaire Boyer, Smitha Cunigal, Abhinav Kumar, Karan Makkar, Murali Mallikarjunan, Hasto Narendra, Santiago Andres Valdivieso Puertas, Shrddha Rajesh, and Varshini Sridhar for outstanding research assistance in Chennai and in Cambridge. We thank the Government of Tamil Nadu / J-PAL partnership fund, and the National Science Foundation (award number 2315488) for financial support.

1 Introduction

Governments face a tradeoff between performing tasks in-house with their own employees, or procuring these services from private firms. Classic theories, such as Hart et al. (1997), frame this as a tradeoff between efficiency and multitasking concerns: even though private firms may be more efficient, have strong motives to cut costs, and have a greater ability to impose high powered incentives on their employees, contracting frictions imply that they do not face the full set of multi-dimensional incentives as the government, generating problems.[1]

This setup, however, may miss an important set of issues. First, there may be informational differences between the government and private firms – and while the government can more easily share information that is recorded in official data, sharing 'soft' information, i.e., knowledge embedded with government actors, with private firms may be much harder. Second, delegating government authority to private sector firms to interact with citizens may be even more important, but challenging: private firms may lack authority and legitimacy. Third, there are a wide range of private firms with varying quality, and it is not clear that all private firms are more efficient than the government. Identifying and contracting the right firms may be very difficult for the government. Moreover, even if there exists a private firm somewhere that could be more productively efficient, it is not clear that they would want to deal with the myriad complexities of government procurement and shifting priorities. Given all this, government outsourcing may not work well for a different set of reasons than the tension outlined by Hart et al. (1997): for both direct reasons and selection reasons, outsourcing may actually not be more efficient than the government just completing the task themselves.

We study these issues in the context of outsourcing property tax assessments in Chennai, India, a common function that many local governments worldwide perform.[2] At a basic level, determining the characteristics of a property for valuation is a common task for both the public and the private sector: in the public sector, tax inspectors assess property charac-

[1] The classic example is prisons, where the private sector may be more efficient than the government, but it may be hard to contract over the full set of human rights outcomes the government may want to achieve.

[2] Outsourcing aspects of tax collection is common. For example, municipal governments in Nigeria, Tanzania, Uganda and have all outsourced various aspects of tax collection; for a qualitative review of the experience of outsourcing tax in poor countries, see Haas and Manwaring 2017. Moreover, outsourcing property tax assessments is not limited to lower income countries; in rich countries as well, including the US, Australia and New Zealand, some municipalities outsource property tax assessment to private firms. For examples of outsourcing firms, see, for example, vsgi.com (a vendor based in Massachusetts) and www.qv.co.nz (a vendor based in New Zealand); see also the New South Wales, Australia propety tax outsourcing website (https://www.nsw.gov.au/housing-and-construction/land-values-nsw/why-land-values/service-contracts).

teristics to determine property taxes, while even outside of tax contexts, private sector actors complete the equivalent task to facilitate commercial transactions and bank underwriting. At its core, the task is relatively simple and formulaic: measure the property and determine its characteristics and usage – how many square feet, how many bathrooms, how much retail space, how much residential space, and so on. In practice, however, it is often more complex because the task also involves high-stakes interactions with owners - owners want to keep their taxes low and may try a variety of strategies to do so, from denying access to properties to assessors to attempting to bribe them (Khan et al., 2016).

In the setting that we study, the Chennai city government was interested in reassessing property taxes for about 280,000 properties, which they had identified through a coarser GIS-based assessment as likely substantially undervalued. The government was concerned, however, that doing so would take a long time with their existing tax staff because their normal staffing levels are set based on the typical year-to-year changes in properties that need to be checked, rather than a wholesale survey of every property. For this reason, they decided to experiment with outsourcing some of the re-assessments to private property surveying firms to see whether this would be effective. They procured firms using two types of contracts: a fixed-fee contract, in which firms would be paid a fixed fee per property assessed, and an incentive contract, in which firms would be paid a combination of a fixed fee and a percentage of the increase in property tax found. We worked with the government to randomize which of the 199 areas of the city (wards) would be reassessed by these firms and which would be done by regular government assessors, and among wards randomized to be outsourced, which wards would be done by firms with fixed vs. incentive-based contracts.

We first examine who bids for the contracts. Consistent with the idea that this is a common task, more than 20 firms initially expressed interest in the project and attended informational meetings. The vast majority of them, however, did not submit bids. Challenges with the procurement process may suggest why: qualitatively, the firms reported that the project information was complicated and that the government often did not respond to their queries. In our firm survey, firms also reported frequent payment delays for past government projects. In addition, the government demanded low prices: firms that did not bid reported that the price at which they would have been willing to accept the contract was anywhere from 2 to 10 times what the government was willing to pay. Moreover, even for the firms that bid, the government negotiated a final contract price that was about 75 percent less than their initial bid. In the end, despite the government aiming to select 6 firms, only 4 qualified firms bid on the contract, and only 3 firms were ultimately willing to accept the

prices offered by the government and agreed to the contract.

As it turned out, the firms that were skeptical of working with the government had good reasons. Partway through the contract period, after some citizen protests about new tax assessments, the government changed course and abruptly decided to cancel the re-assessment project entirely. The contracts with the outsourcing firms were suspended and all reassessment work, by both government and private firms, was halted. Payments for the work were then substantially delayed, with private firms only being paid 16 months later.

We examine the impact of being randomly assigned to be assessed by either the government agency or one of the private firms in the period where reassessments were happening. We test whether the government and private firms had different levels of productivity in the number of properties that they re-assessed prior to the work stoppage, and we estimate (through an instrumental variables approah) the impact on re-assessed property characteristics (and hence property values) conditional on being re-assessed.

Overall, the government department was much more effective than the private firms. During the same time period (from the time the re-assessment started until they were abruptly halted), and despite on average spending fewer man-days on the reassessment project and having other responsibilities as well, the government inspectors re-assessed 25 percent more properties than the private firms. In fact, the GCC dedicated 449 man-days to re-assessments, while private firms collectively invested nearly twice that amount (808 man-days). Perhaps even more importantly, they also raised more revenue from the properties that they did re-assess: On average, the properties re-assessed by the government surveyors saw their property valuations increase by 84 percent, whereas those re-assessed by the private firms saw valuations increase by only 53 percent (p-value < 0.01). This translated into more revenues collected – tax collection increased by 69 percent for properties re-assessed by the government inspectors, but by only 36 percent from those randomized to be surveyed by the private firms (p-value (p-value < 0.01)). Put another way, the increase in tax revenues – per property inspected – from inspections by government firms was almost double (92 percent higher) that from inspections done by private firms.

While these results represent average effects, we also explore whether there is substantial treatment effect heterogeneity based on properties' baseline tax characteristics. We find that there is – treatment effects for government assessors are almost twice as large for the largest properties as for smaller properties, whereas treatment effects for private firms are essentially flat with respect to baseline property size. Since which properties were assessed by government and private firms were fairly similar (government firms were slightly more

likely to reassess large properties but not overwhelmingly so), we show that differential composition of which properties were reassessed explains at most 4 percent of the difference between government and private reassessments. We also show the fact that government assessors have a comparative advantage for larger properties implies that, if government assessors time was limited, optimally allocating small properties to private reassessment and large properties to government could increase average treatment effects by 7 percent compared to randomly assigning properties for reassessment.

We then decompose the changes that we find in assessments into both changes in area and changes in usage. Property measurement requires both full access to the property and technical skill. Specifically, a team of two inspectors must enter the property, access the roof, and then measure the overall property footprint by measuring the size of the roof with a tape measure.[3] For buildings that are just rectangular boxes, one can estimate the total floor area by multiplying the roof measurement with the number of floors. However, to the extent that the building is irregularly shaped, with different floors of different sizes, or has a mix of residential and commercial properties (which are differentially taxed), tax inspectors need to manually measure each room, drawing a sketch of each floor of the property, and then adding up the area of each room on each floor of each usage type. Property usage, however, is more discrete: for example, discovering that a property that is taxed at the lower residential rate is actually being used for commercial purposes. We find that most of the changes – as well as most of the additional differences between government and private inspectors – come from differences in area, particularly non-residential area, which is taxed at a higher rate. Moreover, we find that government assessors find substantially more buildings with irregular measurements – i.e. measurements where each floor is a different size. These 'irregular-shaped properties' require carefully measuring each room separately, rather than just overall footprint, which can be more simply observed from a single roof measurement. This suggests that the differences require some skilled work – a theme we will return to below.

The finding that government workers do better than private firms is striking, given most economists' priors, but not unheard of: in a very different setting (placement services for hard to place unemployed workers), Behaghel et al. (2014) also find that government workers are more effective and cost effective than the private job placement firms that they hired. A nice feature of our setting and design is that we have rich data to explore why it is the case. We explore four broad reasons.

[3]Buildings in Chennai typically have flat, accessible roofs, which makes this possible.

The first is *soft knowledge* - knowledge that inspectors have about local properties that is hard to pass on to the inspectors. We explore this in two ways. First, we examine whether it was harder for private assessors to find the properties from the official tax rolls than the government, and indeed, find this is the case. This is despite the fact that the firms were given GPS coordinates – nevertheless, in a setting with often unclear or unposted street numbers, and many properties close together (and potentially within measurement error of GPS), this can be a major challenge, and the detailed address information is hard to pass on. Indeed, while about 75% of the private firm assessors reported using GPS coordinates to locate properties, only 17% of government surveyors did – most likely because they already knew the locations of the properties. Combined, this suggests that government inspectors had increased soft knowledge – both about which properties were likely undervalued, and how to find properties that improved their productivity.

Second, we look at whether they had additional soft information about under-valuations - i.e. they may have known which properties were undervalued in ways not captured in the formal data. To examine this, we use the fact that the government had previously commissioned an independent drone-based survey of properties. For this, a separate firm conducted an aerial survey of all of Chennai, demarcated properties, and created estimates of property demand for each property.[4] In particular, we examine properties where the drone survey indicated the property was undervalued, and ask whether those properties are more likely to be re-assessed by either government surveyors or private firms (neither of whom had access to data from the drone survey, as it was still being finalized). We do not find that properties that the drone survey suggests were under-valued were more likely to be re-assessed. However, we find that government assessors find more demand in very undervalued properties compared to the firms; this suggests that they have some soft information on which properties to probe more (or just expended greater effort fully examining the properties).

The second class of reasons we explore is measurement *skill*. Importantly, the fact that property assessments by government surveyors were higher does not necessarily mean that government workers actually were better at assessing properties; they could have been inaccurate and systematically over-valuing properties. To examine this, following Khan et al. (2016), for a random subset of properties that were inspected, we conducted our own

[4]Note that these surveys were intended to be a diagnostic exercise at the aggregate level – and indeed, they informed the Chennai government's interest in city-wide re-valuations – but at the property level, they are too noisy to be used for actual legal tax assessment. While we can use them here as a statistical proxy for soft information - i.e. likely increases in tax assessment - the government could not use them as a substitute for formal tax assessments we study here.

independent assessment. Comparing our independent measure and the official valuation, we find that the assessments by government firms are more accurate than those by independent firms, and that there are no systematic differences in over or under-valuation. This suggests the higher valuations that we observe from government workers are not due to incompetence or systematic overvaluations by government staff; rather, government staff are more-accurate.

Consistent with government workers having more skill, we find that government workers have more prior property measurement experience – an average of 11 years of experience, compared to just 2 years for private workers. They are also more educated – 43 percent percent of the outsourced inspectors do not have a college degree, compared with just 15 percent of government workers. Given that calculating property areas of different types requires some mathematical ability and spatial training (i.e. making sketches of complex property layouts), this may be related to the accuracy results.

The third class of reasons we explore is *effort*. Using our survey data, we find that government workers exert substantially more effort. In particular, conditional on entering a property, they spend about 35 percent more time measuring a property – about 31 minutes for a government worker, compared about about 23 minutes for a private firm workers. This may be capturing measuring each room separately, for example, and may be related to why government workers are more accurate.

Interestingly, we show that this increased effort is picking up effort at the measuring task, rather than an intrinsic 'capacity' or willingness for exerting effort in general. To examine this, in our survey, we asked inspectors to perform the "A/B" task, from DellaVigna and Pope (2018). This task involves asking the respondent to press the 'A' and 'B' keys on a computer keyboard, in sequences, as many times as possible in 1 minute; the key measure of effort is simply the number of times the respondent can do this in 1 minute.[5] We find that, if anything, the private sector workers score higher on the "A/B" effort task than government workers 27.2 more presses (if each press means one point), or 34.5% more ($p = 0.02$) – suggesting that we are picking up effort at the task of measuring properties rather than a pure capacity for effort or generalized motivation per se.

The final class of explanations we explore is *authority*. Tax assessment is an official government function, which requires obtaining access to citizens' property. Private firms were issued official letters from the government authorizing them to access properties, as well as formal identification cards and appropriate attire (caps, jackets). However, the government workers may have more confidence exercising this authority, and this may be

[5]The original version in DellaVigna and Pope (2018) asks respondents to do the task for 10 minutes.

hard to delegate – particularly in a potentially adversarial context such as tax collection, where property owners typically want to minimize their tax bills. We find that outsourced firms report being refused entry to properties substantially more than government workers. And, related to the idea of effort above, conditional on being refused, government workers do not give up – they do 4 times as many revisits of refused properties compared to private sector workers.

Given the performance of the private firms, an interesting question to ask is whether they could have done better with greater *incentives* to increase property tax assessments. Of course, government workers – who are paid fixed salaries with essentially guaranteed life tenure – may not face strong incentives either, but it is possible that there are soft incentives in the government to increase revenue that were not passed on private firms. In contrast, based on being paid for each audit they conduct, a firm's incentive may have been to instruct their workers to do the job as quickly as they can, which would potentially explain the low accuracy (whereas government inspectors may feel they can take all the time they want). Moreover, private firms may also find it expedient to collude with the taxpayer when they protest, which would reduce accuracy as well. To examine how incentives could affect the firm behavior of the firms, the government recruited firms under two types of contracts – a fixed fee per-property contract, and incentive contract, where the firms were also paid a percentage of the increase in revenue collected. Wards were randomly assigned to either the fixed or incentive firms. In general, we find no statistically significant differences between the fixed and incentive firms on the number of properties re-assessed, the amount of revenue raised, the accuracy of assessments, or the knowledge or effort dimensions that we observe. This suggests that to the extent that outsourcing is less effective at collecting revenue, it is unlikely driven by differences in incentives to do so in this setting.

We show that, in general, these tax-reassessments – of all types – are astoundingly cost-effective. For private firms, we know the monetary costs the government paid the firm; for government inspectors, who have responsibilities beyond assessing properties (i.e. issuing notices and collecting revenue), we know their wage and, and for a subsample, how many properties they can do in a day.Even assuming that the increase in revenues from the re-assessments last only for a single year, we calculate between a 1,578% and 2,033% return-on-investment from re-assessments. Since higher tax assessments likely last for years if not decades, returns on investment are commensurately higher – if the higher assessments last for a decade, this implies returns-on-investment of 16,675% to 21,233%. This is because a tax re-assessment takes about 40 minutes on average of an inspector's time, plus travel

time – but increases tax payments by 30-50 percent for years. In short, the government is more cost effective.

All told, the experience here suggests a different set of constraints on outsourcing. The classic tradeoff set up by Hart et al. (1997) is a principle-agent problem: private firms are more efficient, but shirk on quality. Here, we find a different set of issues. Government workers appear more efficient than the private sector – in part, because they have greater authority and some knowledge that is hard to transmit to the private firms, and because they have more skill and experience than the firms that ultimately bid on the contracts. These frictions may be exacerbated by the government as an unreliable partner – which firms anticipated ex-ante and which was true ex-post – which may also lead to lower quality firms being the ones to take up outsourcing. The results suggest that for some types of services, the government may actually better deliver than the private firms.

This project contributes to the both the literature on outsourcing public service provision as well as the literature on tax administration in developing countries. Prior empirical research on outsourcing public service delivery has largely focused on understanding which services are likely to be outsourced and where (Levin and Tadelis, 2010) and comparing the performance of outsourcing to government provision of a service (Galiani et al., 2005; Behaghel et al., 2014; Romero et al., 2020; Banerjee et al., 2019). However, less is known about contexts that may be more adversarial such as taxation, where the tax assessor is tasked by the government with fairly assessing tax liabilities among many parties (who may want to minimize their tax payments). Theory suggests such 'adversarial' cases could be quite different (Banerjee et al., 2012). For example, in the context of pollution audits, (Duflo et al., 2013) find that the performance of private auditors is catastrophic, due to the perverse incentive they face to collude with the firms they audit, but there is no direct comparison of their performance with government workers.

With regard to taxation, much of the previous economics literature on property tax collection in developing countries has focused on incentivizing government tax collectors (Khan et al., 2016, 2019), optimal assignment and team composition of tax collectors (Bergeron et al., 2022), improving tax compliance by engaging local elites (Balán et al., 2022), and the effects of varying property tax and enforcement rates (Brockmeyer et al., 2021; Bergeron et al., 2024), but little is known about outsourcing parts of the tax assessment and collection system entirely. Yang (2008)'s quasi-experimental study of pre-shipment inspection in customs suggests the potential for outsourcing to be effective, but it is not clear how the results from that context-—an international, high-reputation firm being hired by a domestic

customs agency to perform customs inspections in a different country – inform the question of whether a government should outsource its own regular tax assessments to domestic firms. There is also a small qualitative literature, outside of economics, which focuses specifically on government outsourcing of tax collection in Sub Saharian Africa (see (Fjeldstad et al., 2008), (Iversen et al., 2006), (Cheeseman and de Gramont, 2017)). Outside of Lagos, Nigeria, which is generally held as a success (but where the subcontracted firm was associated with the mayor), these experiences have also been shortlived and relatively unsuccessful, for reasons having to do with contracting, legitimacy, and excessive prices ((Haas and Manwaring, 2017)). Our results are consistent with these insights, but establish the results much more systematically, and in a context with higher government capacity.

The remainder of the paper is organized as follows. Section 2 describes the settings, the experimental design, and the data we use. Section 3 describes the outcomes of the procurement process. Section 4 explore the impact of outsourcing on tax assessments and tax revenues, and Section 5 explores mechanisms in more detail. Section 6 discusses cost-effectiveness. Section 7 concludes.

2 Setting, Experimental Design, and Data

2.1 Setting

We study urban property taxes in the city of Chennai, India. Property taxes are assessed and collected by the municipal government, officially known as the Greater Chennai Corporation (GCC). The GCC is part of the Government of the State of Tamil Nadu, and therefore their procurement practices and regulations follow local state law and procedures.

Property taxes are determined by a formula that considers property characteristics – particularly the floor area and land area – and usage. Each square foot of property in a given usage category (residential or non-residential) is assessed at a fixed street rate. Since the per-square-foot street rate is fixed exogenously by the government, the main determinant of a given property's tax is therefore its area and usage type.[6]

Tax inspectors measure property areas manually. Since most buildings in Chennai have flat, accessible roofs, a team of two inspectors usually begins by measuring the roof's dimensions with a tape measure. If the building's footprint is identical on all floors, inspectors

[6]Small discounts are also given based on property age (3% for buildings built between 2007 – 2017, 6% for buildings built between 1997 – 2007, and 10% for older buildings) and for 'semi-permanent' structures, such as areas covered by awnings.

can multiply the roof area by the number of floors to compute the total covered area.

While this process sounds straightforward, it has a number of complexities in practice. First, this procedure requires access to the property, and while inspectors are legally allowed to enter a property, the owners may not be home, or they may be resistant. Second, the properties may be irregularly shaped, requiring detailed measurements room by room on each floor. Third, the usage may not be obvious and may also be concealed since commercial areas are taxed at a higher rate. For example, in one visit that we conducted, inspectors discovered a workshop behind a closed door on the second floor of an otherwise residential building; a lax inspector may miss this. Fourth, there are a variety of edge cases that require detailed knowledge of the law - at what point does an awning or non-permanent roof become 'covered area,' how does one account for light shafts in the middle of buildings, or courtyards on an intermediate floor that are open to the sky, and so on. Thus, a complete property assessment takes some combination of authority to enter the property, effort, and skill.

Prior to our project, the GCC had concerns that many properties may have been under-assessed. The GCC does not re-asses properties at regular intervals to update its database, largely relying on tax owners to report any changes to their properties.[7] In 2018, it commissioned an external firm to assess the degree of under-reporting and identify any properties that had not been assessed yet. To do so, this firm used drones to take areal photographs of the city and then used GIS tools to demarcate properties (henceforth the 'drone survey'). Staff from the drone survey company conducted some field visits to supplement this information with the number of floors for each property and the property usage. While this survey is not usable for assessing taxes – the area of the property is based on an approximation of the overall footprint, rather than a detailed hand-measurement accounting for irregularities – it revealed that 340,000 properties were likely substantially undervalued or not valued.

2.2 Experimental Design and Sample Selection

The large number of properties to be re-assessed meant that only using regular government surveyors to conduct the assessment could take a long time. Thus, the GCC was keen to test whether outsourcing the assessment to private firms could be an effective alternative.

We worked with the Chennai government to experimentally test this idea. The government identified firms through its typical procurement process (see details in Section 3).

[7] In 2016, the GCC was contemplating a switch to a self-assessment based tax system, and offered taxpayers an opportunity to declare changes to their properties. However, the data were not used to update any of the property records as the GCC decided not to adopt the self-assessment based system.

Firms would be responsible for visiting the list of properties supplied by the government and measuring the area and usage of each property. GCC assessors and private firm surveyors were provided the list of properties that they were supposed to re-assess, including the street address and GPS coordinates collected during the drone survey. The GCC's decision to outsource the re-assessment to private firms was announced in major newspapers as well as on the GCC's Instagram page. Surveyors hired by private firms were provided caps and jackets with the GCC logo, ID cards, as well as an official letter confirming they had been authorized by the GCC to enter and measure properties as part of a re-assessment drive.

The government procured firms under two different payment mechanisms. In the *fixed fee* version, the government would pay the firm a fixed fee based on each successfully surveyed property. For this contract, firms' financial bids consisted of the per-property fee that they would charge the government. In contrast, in the *incentive* version, the government would pay the firm a fixed fee of 50 rupees per property, plus a fixed percentage of the increase of the tax liability uncovered. For this contract, firms' financial bids consisted of the percentage of the revenue increase that they would be paid. Firms could bid under either or both contracts.

On net, as described in Section 3, the government selected two firms under the fixed fee contract and one firm under the incentive contract. Once chosen, the government conducted a training for the private firms' assessors in the second week of April 2023, consisting of an in-class part where the measurement process was explained in full by the government staff and a practical training, where the private firm assessors visited specific properties with the city government staff and practiced taking measurements. The contracts were finalized in the second week of February 2023, and firms began work in the second week of April.

The government prioritized the properties to be re-assessed in two ways. First, as described above, it used the drone survey to flag properties where there were differences in covered area of at least 20 percent, or differences in usage. Second, it used a self-declaration exercise, where property owners were asked to declare the official size of their property; any deviations of 20 percent or more between self-declarations and official tax liabilities were also flagged. The government then removed government buildings, large apartment buildings, and several other categories that it deemed were too complex for reassessment by private firms.[8] This resulted in about 279,000 properties for re-assessment.

The project was designed to have a 'study phase' – to determine whether private outsourcing worked, and if so, which type of contract was more effective – and a eventually

[8]The reassessment was supposed to also include un-assessed properties identified by the GIS. However, spot checks identified by the city assessors concluded that many of these were vacant lots, or they could not be easily linked to the databases of property owners, so they decided not to include them.

scale-up phase, whereby the most effective approach was scaled up. We examine the 'study phase' in this paper.

We worked with the government to randomize which properties would be re-assessed during the study phase, and whether by government surveyors or private firms. The randomization was conducted in two steps (Figure 1). First, at the ward level (i.e. geographic unit), we randomized each of the city's 199 wards to be reassessed by one of the three private surveying firms or by government surveyors, stratified by the 3 regions of the city (north, central, south); hence, 50% of wards were assigned to the fixed-fee private reassessment, 25% to the incentive-based private reassessment, and 25% to the government surveyors (see Appendix Figure 1). Second, within each ward, we randomized individual street segments to be assessed or not.[9] Within each selected ward, 40% of the street segments were randomly selected to be assessed during the study phase, with the remaining 60% of properties serving as a control group. These treatment street segments were further randomly divided, within each ward, into three lists, lists A, B, and C. Firms and GCC surveyors were both instructed to begin with the first randomized list of treatment streets provided to them, 'list A'.

The government unexpectedly halted the reassessment process – by both private firms and government surveyors – on June 3, 2023, due to political sensitivities. At this time, both the firms and the government surveyors had only seen treatment list A. In our analysis, we therefore compare the properties on list A – which were given to the firms to reassess during the period the study was underway – to the 60% of streets in the control group, yielding a total sample of approximately 206,000 properties.[10] A small number of properties were reassessed by the government after the project halted; we discuss these below and account for them in the analysis using an intent-to-treat design.

2.3 Data

We use several data sources. First, we use *administrative data* on property tax assessments and payments. These data have information on tax assessments, including the detailed property characteristics (i.e. floor area by usage type, street rate) used in the tax formula, as well as the overall tax assessment (known as the 'tax demand'). We have data on property characteristics for both the baseline year and the year following the experiment. We also

[9]Note that if a street contained more than 20 properties, we sequentially subdivided the streets geographically until we were left with street segments of 20 properties or fewer each.

[10]In our main analysis, we drop streets on study phase lists B and C, which would have been treated but were not due to the unexpected study halt, which yields a total sample of about 206,000 properties. In the appendix, we report alternative analysis including lists B and C as well.

have administrative data on the actual tax demand and tax payments between 1999 and 2024, the year after the intervention. [11]

Second, we use data from three different surveys that we conducted. First, to assess tax assessment accuracy, for a random subsample of 2,150 properties that had been re-assessed by either the private firms or government enumerators, we conducted a *property survey*. In this survey, our survey team (two people per property) obtained an independent assessment of the property's valuation by measuring each property using the same procedures as the tax staff (i.e. start on the roof with a tape measure, then for irregularly shaped properties, measure each room separately, and so on). They also noted property usage (and the square feet of each usage type for mixed properties) and collected some demographic characteristics of the property occupant. We were able to find and survey about 70 percent of these properties, for a total of 1,456 properties; as shown in Appendix Table 1, we observe no evidence of differential attrition. These data allow us to compare the accuracy of private and government tax assessments.[12]

Second, we conducted an *assessor survey*. This survey was administered to both the government enumerators and the assessors employed by private firms. The survey includes demographic characteristics (e.g. age, education, prior experience), and information about the surveying process itself (e.g., the number of minutes spent on each property, refusals, repeated tries after refusals). The survey also collected an experimental measure of effort – the 'A/B' task from from DellaVigna and Pope (2018) – in which respondents are asked to alternately press the A and B keys on a computer as many times as possible during 1 minute. The survey was conducted between June and September 2023, after the project was halted. The GCC and private firms provided the names and contact information of employees who had been involved in re-assessing properties. We were able to successfully survey 82% of GCC assessors and 66% of firm employees. When we were unable to contact firm employees, it was largely due to incorrect phone numbers provided by the private firms.

[11] We observe tax payments in two datasets. The Receipts Dataset records each individual tax payment, including the amount, payment date, and the tax year to which it is applied. The Demand, Collection, and Balance (DCB) Dataset is dynamically updated and reflects the current levels of tax demand, tax collection, and unpaid balances for each tax year as of the time the data extract is generated. For example, if a taxpayer makes a large advance payment in 2024, the DCB dataset may show positive tax collection for 2025, even though the Receipts Dataset would show no payment made in that year. Since we aim to measure payments made after the intervention, we rely on the Receipts Dataset for our analysis.

[12] We had initially aimed to collect data on satisfaction with the tax authority and city services, as well as any extra payments made by citizens during the re-assessments. However, we, given the sensitivity around tax collection by the end of our endline survey, were not given approval to ask these questions. Nonetheless, we have some information on these topics based on a survey of 197 pilot properties that we conducted.

Third, we conducted a *firm survey*. We attempted to survey all 34 firms that the GCC actively invited to bid on the tender – both those that ultimately bid on the project and those that did not. We surveyed 16 firms, including the 4 submitted qualified bids. This allows us to capture additional information about the background of the firms, as well as to understand why those firms that chose not to bid did not do so.

Finally, we have the approximate tax assessments from the *drone survey*. While these are not intended to be as accurate as an official firm assessment, on average, the difference between valuations on the drone survey and official property valuations may be indicative of which properties were most undervalued, and thus the highest priority for re-assessment. Properties were short-listed for re-assessment by comparing the property characteristics in the drone survey to the official tax records. Properties were categorized for reassessment (and hence included in our sample) if the property usage recorded during the drone survey placed them in a higher tax category than the usage recorded in the official data (mixed-use or non-residential v.s. residential and non-residential v.s. mixed), or if the estimated built up area in the drone survey was larger than the official records.

3 Procurement Process, Challenges and Frictions

The procurement process followed the Tamil Nadu government's standard outsourcing practices. In this section, we describe how this process proceeded – both in terms of the formal procurement design and in reality – to illustrate the contracting frictions and the subsequent challenges that they can create for outsourcing.

3.1 Procurement Process

The government aimed to hire six firms, three with the fixed contract and three with the incentive-based contract. To do so, the GCC followed its standard procurement practices. This began by listing a preliminary notice for tender on the Government of Tamil Nadu's website on November 2, 2022. This preliminary notice had limited information, such as the earnest money deposit required to submit a bid, and stated a bid submission date of December 20, 2022. The tender document itself, which specified the full set of requirements, was supposed to be posted one week later, but it was delayed to the last week of November 2022, as city officials were engaged in flood control due to heavy rains.

For each contract, firms were to be selected through a two-stage process. First, the government would conduct a technical review of each firm's bid to ensure they were qualified.

Second, the government would choose firms with the lowest financial bid from among qualified bidders. Bidders were required to submit a technical proposal, including balance sheets, evidence of prior experience, and CVs of key personnel. In addition, they had to present their proposal in an in-person presentation to GCC officials. Only firms that had completed at least two similar assignments in the last ten years were eligible to submit bids.[13] Technical proposals were scored out of 100 based on the evaluation criteria in Appendix Table 2, and the GCC only opened financial bids submitted by technically-qualified bidders that received a technical score of 80%. The financial bid depended on the contract type: for the fixed-price tender, the contract template asked firms to bid a fixed price per property assessed; for the incentive-based tender, applicants were required to bid a percentage of the additional tax demand (i.e. assessed values) that they would be paid, up to a cap of 6%. As per the GCC's tender rules, the GCC begins by negotiating with the firm with the lowest financial bid to reduce the price further. Once the final price was determined, all technically-qualified bidders were offered the opportunity to match the bid.

The government advertised the tender in several ways. In addition to posting it on the city government's e-tendering website, the government publicized it in one Tamil and one English newspaper. They also identified (through the internet) and directly contacted firms that had specialized in property measurement, topographical surveys, GIS surveys, and property valuation about the tender.

3.2 Bidding Process and Outcomes

The government held two pre-bidding meetings in December 2022 with prospective firms to provide details of the project deliverables and to answer any of the firms' questions. A total of 23 firms attended one or both of these two meetings. The firms asked questions about the pre-qualification criteria, the techniques for measuring properties, the difference between the fixed and incentive-based contracts, and the support that the chosen firms would receive from the government to enter properties.

The meetings revealed that incentive contract created additional uncertainty for firms. Since firms in this contract would be paid a percentage of increased tax demand, the amount they would be paid in rupee terms would depend on the extent of under-taxation. Firms accordingly asked various questions during the bidding meetings about the expected revenues from this contract; firms were directed to conduct their "own research" to determine the

[13] Similar projects included enumeration, assessment, or collection of property tax for municipal governments, topographical, cadastral, land surveys for public- or private-sector clients etc.

expected gains they could make.

Despite the fact that 23 firms attended at least one of the bidding meetings, in the end, few firms bid on the contract. The stated reasons for this varied. Qualitatively, some firms that we talked to suggested that the prices per property that the government would pay were too low. Past experiences was another reason: one firm that we talked with declined to bid because they had outstanding payments from the city government for another project. To further facilitate the bidding process, and allow for more firms to enter, the government extended the deadline from December 20, 2022 to January 2, 2023. In the end, only five firms completed the bidding process.[14] Three firms submitted the documents to bid for the fixed contract only, while two firms bid for both the fixed and the incentive contract. One of the firms that bid for both contacts did not meet the pre-qualification criteria, and was deemed ineligible, leaving four firms remaining.[15]

After submitting the bids, the four firms presented their work plans to the city government in person. One firm (the firm that bid for both contacts) demonstrated a clear understanding of the project requirements; the other three firms presented workplans that exceeded the scope of work, including only hiring engineers (while the task required measuring properties and computing room areas, a fully trained engineer was not needed) or conducting surveys to understand taxpayers' "socio-economic behavior." Thus, the GCC panel re-emphasized the project scope of work, and acknowledged a need for a price negotiation so that the prices paid were consistent with the reassessment task.

For the four firms that bid for the fixed contract, the firm that also applied to the incentive contract was the lowest bid at INR 225 (USD 2.70) per property; the other three bids were INR 960, INR 975 and INR 1000.[16] For the incentive contract, the one eligible firm bid 4.8 percent (i.e., it would be paid 4.8 percent of the change in assessed property demand); it is worth noting that this firm, despite bidding for the incentive contract, also stated a preference for the fixed contract.

The government then negotiated the prices with the bidding firms. Essentially, the government used the lowest-priced bid on the fixed price contract as the reference bid for all

[14]Other firms started the process of applying, but then did not submit all the required documents (despite continued communication from city officials to provide assistance). Given the low number of firms that bid, there was a discussion about whether to re-open the tender to allow more firms to bid, but the government ultimately decided against this.

[15]The firm was unable to provide evidence of having completed two property measurement projects of a similar nature within the past ten years.

[16]The high bids may have also been because many of the valuation companies had experience dealing with very high-value, commercial properties (e.g. ports, industrial estates), which require highly skilled staff, like engineers, and take quite a bit of time.

firms, and demanded that the other firms meet this price. In the end, the one firm that bid for the incentive contract agreed to the contract at 4.8 percent, while two firms agreed to the fixed contract at INR 224 (almost exactly the lowest bid obtained by the government, but only about 25 percent of what these two firms had initially bid themselves). The final firm felt that the proposed price was too low and dropped out.

There were thus several *de facto* challenges with the bidding process that may have deterred firms from bidding. First, as indicated above, the bid submitted ended up being a reference price and the opening to a subsequent negotiation, rather than the actual price that would be paid as in a typical first-price auction. To the extent that firms may have anticipated this (perhaps from previous dealings with the government), but it was not formally stated, this may have created additional uncertainty during the bidding process. Second, there was a lack of clarity on some aspects of the contract, which also could have been a deterrent to the degree that firms anticipated this ex-ante. For example, in 2022, the government raised property tax *rates* – separately from the assessment exercise here, which focused on the underlying characteristics (e.g., covered area) that would then be multiplied by these rates. The government did not specify ex-ante that the incentive firm would be renumerated with a percentage of the additional demand generated through the re-assessment, calculated based on the tax rates existing *before* a 2022 tax hike. This caused issues down the road, when the firm expected to be compensated on the change in revenue calculated using the higher, post 2022 tax hike rates. Similarly, the government did not provide projections of firms' expected revenues from the incentive contract, despite being asked for this information during the pre-bid information sessions. Uncertainty about potential earnings and lack of familiarity with performance-based contracts likely discouraged firms from bidding on the incentive contract.

3.3 Comparison of Bidding and Non-Bidding Firms

Our main analysis compares the performance of government workers to those hired by the firms that have survived the bidding process. This is in itself an outcome of the way the outsourcing was conducted, and is thus part of the treatment. Therefore, it is also worth comparing the bidders to the set of firms who had initially expressed interest.

Table 1 provides descriptive statistics to compare the firms that ultimately bid on the project with those that did not based on a survey that we conducted. Column 1 provides means for bidders, while Column 2 provides means for non-bidders.

Firm characteristics are presented in Panel A. Both the bidding and non-bidding firms

were large, established firms, with an average firm age of 22 years. However, in general, those that chose not to bid were smaller, both in terms of employees (mean of 42 employees for non-bidding firms, compared with 73 employees for the bidding firms) and revenue.

Many of the firms had experiences working with government, but their experiences were mixed (Panel B). Those that bid had more government experience: Seventy-five percent of bidders had a government project before, while 43 percent of non-bidders did. Interestingly, virtually all firms that had prior government experience - bidding and non-bidding firms alike - reported delays in getting paid by the government. It is plausible that the delays in payment may be more costly for smaller firm (who may be less able to absorb these delays), which may be related to why smaller firms were less likely to ultimately bid on the project.

We next asked firms about what their proposed bids would be for the reassessment contracts, had they bid. Figure 2, Panel A graphs the distribution of bids (i.e. the price per property proposed) for the fixed contract; for comparison, we also graph the actual bids and demark the final contract price in a dotted line. In Panel B, we do the same for the incentive contract where a bid is the percentage increase in tax demand given as a bonus payment. Both figures show that the proposed bids would have been much higher than the actual price chosen. In fact, 64% percent of firms (including three of the contracted firms) would have bid higher than the actual price in the fixed contract, and almost all of the firms would have done so in the incentive contract. Thus, perhaps few firms bid because they believes it would not have been profitable for them to do so.

3.4 Project Halt

The firms that were skeptical of working with the government perhaps had good reason. Partway through the contract period, in early June 2023, the government abruptly canceled all re-assessments, both by government assessors and private firms. Some taxpayers complained to their ward councilors (elected representatives of each ward) that only small properties were being targeted, and that large property owners were being let off the hook. The ward councilors raised these issues to the relevant state government officials, who, in turn, instructed the city to halt the re-assessments. It is worth noting that general elections were coming up in 2024, so the government may have been concerned about political pushback prior to the elections.

The firms, left in limbo, lacked clarity on the possibility of resuming the exercise and the duration of the hiatus. They also had not been paid at all at this point. Firm surveyors, who had been actively engaged in fieldwork at the time of the halt, were instructed to

cease operations. The firms attempted to discuss this issue with the city government, but the relevant official was on leave. Firms were therefore uncertain about what to do with the staff they had hired for the project – terminating them was undesirable if they were restarting anytime soon, but they also could not absorb them temporarily in other projects given that they had been hired for this project.

After the reassessment happened, we observe a small number of assessments by the government in the post-period (discussed below). However, ultimately, the outsourcing never resumed, and the firms were not paid till 16 months later (October 2024).

4 Impacts on Tax Assessments

4.1 Estimation

To estimate the impact of outsourcing, we use the fact that we worked with the government to randomly assign a) wards to be re-assessed by private firms (either with fixed or incentive-based contracts), to be reassessed by government assessors, or to not be reassessed, and b) within wards, to randomly choose which streets would be assessed.

Since the re-assessment process was unexpectedly halted partway through the implementation, not all properties on streets that were randomly assigned to be re-assessed were completed.

We first assess whether government worker and private firms differentially progressed through their assigned list for re-assessment:

$$Reassessed_{isw} = \alpha_w^F + \beta_G^F AssignedGovernment_{isw} + \beta_P^F AssignedPrivate_{isw} + \gamma^F y_{isw}^0 + \upsilon_{isw} \tag{1}$$

Then we compute the intention to treat on tax collection and tax demand:

$$y_{isw} = \alpha_w^R + \beta_G^R AssignedGovernment_{isw} + \beta_P^R RaPrivate_{isw} + \gamma^R y_{isw}^0 + \omega_{isw} \tag{2}$$

Since only some streets within each ward were randomized to be re-assessed, we can include ward fixed effects α_w^R. We control for the pre-randomization value of the outcome variable, indicated by y_{isw}^0 when possible. Since the randomization to the government or private firms occurred at the ward level, we cluster the standard errors by ward.

Equations (1) and (2) are of independent interest: Equation (1) will tell us whether the government workers were more or less efficient at conducting the reassessment tax than the private companies that were contracted specifically for this purpose. Equation (2) tell us how much the government and the private firms manage to increase demand and collection, in a given period of time.

In addition, they form the first stage and reduced form for an instrumental variable strategy of the effect, for a particular property, of being re-asssed.

We estimate the following equation:

$$y_{isw} = \alpha_w + \beta_G ReassessedGovernment_{isw} + \beta_P ReassessedPrivate_{isw} + \gamma y^0_{isw} + \epsilon_{isw} \quad (3)$$

where $ReassessedGovernment_{isw}$ indicates that property i located on street s in ward w was re-assessed by the government surveyors and $ReassessedPrivate_{isw}$ indicates that the property was re-assessed by the private surveyors. We instrument for $ReassessedGovernment_{isw}$ and $ReassessedPrivate_{isw}$ with dummies for whether the street was randomly chosen to be re-assessed by a government or a private firm, i.e. $AssignedGovernment_{sw}$ and $AssignedPrivate_{sw}$. The control variables are the same as for the reduced form and first stages.

The results from Equation (3) show the impact of being reassessed by a government or private firm, compared to not being reassessed at all. However, we are also interested in whether government or private firms are more effective. We, therefore, report statistical tests of whether $\beta_G = \beta_P$ in all specifications.

To estimate the impacts of the two different types of private contracts (fixed vs. incentive), we also re-estimate equation (3) with separate variables for reassessment by fixed fee and incentive firms, instrumented by the randomization of wards and streets into private fixed, private incentive, or government reassessment. For these regressions, we report statistical tests for the difference between fixed and incentive reassessment.

We report tests for statistical balance based on baseline administrative variables in Appendix Table 3. We examine 10 variables, and regress each baseline variable on the government and private firm treatment indicators and ward fixed effects, clustering the standard errors by ward. We also report the p-value of the difference between the treatment effect for the private and government firms. Out of the 30 differences computed (10 for government versus control, 10 for private versus control and 10 for government versus private), we find that 4 of them are significant at the 10 percent level, as expected by chance.

4.2 Completion Rates

We begin by examining what share of properties were re-assessed, and whether this varied between the government and private firms. Not only is this is one measure of productivity, but this also serves as the first stage when we estimate the differential impact of a re-assessment using equation (3).

We examine two measures of reassessments in Table 2. In Panel A, Column 1, we regress the reassessments collected by August 2023, right after the experiment was halted, on treatment status. As show in the Table, the government reassessed 47 percent of the properties assigned as compared to 34 percent of properties assigned to private firms (p-value of the difference 0.006).

In Panel A Column 2, we replicate Column 1 using data updated as of April 2024, which corresponds to when we observe updated tax demand and collections. We do so for several reasons. First, the government suggested that there was a delay in the reassessment data entry from the private firms, and so that some of the observed differences as of August 2023 may have been driven by lags; all of this lagged data entry was completed by April 2024. Second, as discussed above, the government completed a small number of reassessments after the outsourcing was halted; this appears to be very small (at most 313 reassessments, or less than 6.5 percent of the total).[17] Finally, since April 2024 is when we observe tax demand and tax payments, this serves as our first stage when we estimate equation (3) using instrumental variables. Using the updated April 2024 data, we find that the government surveyors re-assessed 47 percent of the properties they had been assigned, compared to 38 percent for private firms (column 2). This implies that the government was able to complete about 25 percent more properties (p-value 0.050), despite also having other responsibilities.[18] This provides some initial suggestive evidence that the government workers may have been more productive; we will explore this in more detail below.

4.3 Overall Impacts on Tax Assessments and Collections

We next investigate whether this led to differences in actual tax assessments. Panel B of Table 2 shows the instrumental variable estimates from estimating equation (3). We focus on

[17] The April 2024 data reflects 313 additional properties measured by government assessors. Although our current data do not allow us to distinguish between reassessments completed by the government after the project halt and lagged data entry, this serves as an upper bound.

[18] From our pilot endline survey, we find that about 90 percent of property owners/renters report that a tax official recently came to their property. It is quite reasonable that some may not know, especially if someone else was home, or they forgot. Therefore, we believe that most properties were actually visited.

two key outcome variables, tax demand for the 2023-2024 year (measured as of May 2024), i.e., the annual tax assessment of a property (column 3), and actual tax collections for the same period (column 4). This IV specification uses the randomization to the government, private firms, or the control to estimate the change in taxes for those properties that were actually re-assessed, using the first stage from Panel A, column 2. The equivalent reduced form (where we regress tax demand and collection on treatment assignment) are shown in columns 3 and 4, respectively, of Panel A.[19]

Overall, reassessment – by either government or private firms – leads to substantial increases in tax assessments and collections. For those properties re-assessed by government workers, tax demand increases by an average of 6,080 rupees, or about 84 percent. For those properties re-assessed by private firms, tax demand increases by an average of 3,859 rupees, or about 53 percent. Thus, government workers increase tax assessments by an additional 31 percentage points (p-value 0.001). These differential increases in tax demand, in turn, translate into differential increases in tax revenue collected by the government (Column 4 of Table 2). Tax collections increase by 69 percent when re-assessed by government surveyors, but by only 36 percent when re-assessed by private firms. Thus, in short, a reassessment increased government collections in the 2023-2024 tax year by 92 percent more than if a private firm reassessed the property.[20]

What drives the differences in re-assessments? Recall that tax assessments in this context are determined by a formula that essentially multiplies the floor area in square feet for each usage type by a fixed multiplier based on the property location. For re-assessments to change valuations, they therefore must change area, usage, or some combination. Some of these changes are relatively easy to observe – walking up to a property, one can easily count the number of floors the property has, for example. Some changes are harder – capturing differences in area requires carefully measuring each room one by one, especially if there are different use types in the building which must be measured separately, or if building is irregularly shaped (particularly with different shapes on different floors).

[19]We replicate Table 2 omitting the baseline controls in Appendix Table 4, and with adding the B and C treatment lists as controls in Appendix Table 5. The conclusions are similar with these alternative specifications.

[20]One concern is that the government officials or private sector employees would ask for informal payments to change the level of the tax collection. While we could not ask this in our endline, we did ask questions about unofficial payments during reassessments in our pilot endine survey for 191 properties. Interesting, we find that bribes are low in this setting. Only 8.1 percent of respondents know of any asked for unofficial payments made in this setting, and only 7.6 percent of respondents knew of anyone paying an unofficial payment in this setting).

Table 3 therefore decomposes the changes in tax assessments.[21] We first examine total area, and find that it increases by 573 sq feet if reassessed by government, whereas it increases by 476 sq feet if reassessed by private firms (p-value of difference = 0.071).

We then disaggregate area by residential area (which is taxed at a lower rate) and non-residential area (which is taxed at the higher rate). We find that residential area property increases for both government assessors (507 sq feet) and the private firms (439 sq feet), with a p-value of the difference of 0.154 (Column 2). However, as shown in Column 3, we observe a large difference for the higher taxed non-residential property: government assessors find about 40 more square feet, while we observe no effect for the private firms (we can reject that these are the same with a p-value of 0.008). Since non-residential property is taxed at a much higher rate, this difference is an important driver of the difference in tax assessments between government and private assessors.

In addition, while both the government and private firms both find more floors–0.57 floors for government vs 0.49 floors for private assessed properties–the government appears to find about 16 percent more (p-value of difference 0.15; Column 4).

Consistent with these findings, it is also worth noting that changes in the type of usage within a given property are more likely with government assessors. Column 5 and 6 show if usage type is 'upgraded' or 'downgraded,' respectively, where 'upgraded' indicates moving to a usage type with a higher tax rate (typically commercial or mixed us) and downgraded means moving to a usage type with a lower tax rate (typically residential). Column (5) shows that while reassessments of both types lead to properties being reclassified at higher rates, this is almost twice as likely for government compared to private sector assessors (4.2 percent vs. 2.5 percent; p-value 0.04). Downgrades also happen with re-assessments, but these are less likely with government than private assessors (0.9 percent vs. 1.7 percent; p-value 0.021).

Together, these facts suggest that some of the differences between government and private assessors come from discovering features that are more highly taxed. This typically requires a more in-depth assessment than single-use buildings, as each room needs to be measured separately.

To probe this idea further, we define a "irregular-shaped assessment" as one in which different floors of the building are recorded with different areas (see Column 7 of Table 3). Recall that for a standard, rectangular building with the same size floors, assessors just mea-

[21] Appendix Table 6 provides the reduced form estimates, while Appendix Table 7 provides the estimates using the B and C lists additionally as controls. The conclusions are the same.

sure the area of the rooftop, and use that area for each floor; by contrast, a "irregular-shaped assessment" requires measuring each room separately and constructing the area of each floor from the room-by-room assessments. In the control group (i.e. prior to reassessments), only 45 percent of properties have irregular-shaped, i.e. non-uniform-by-floor, assessments. This increases substantially with reassessments, but more so when done by government surveyors (24pp vs. 19pp; p-value 0.056). Combined, this suggests that government surveyors are not only finding more area, but they are also doing more sophisticated assessments that result in higher taxed features.

Finally, we check whether government assessors and private firms applied the tax formulas differently conditional on the areas they measured. Recall that the tax is supposed to be the residential and non-residential areas measured multiplied by a fixed street rate for each. We therefore recalculate the tax demand implied by applying the tax formula to measured areas in the administrative data, and re-estimate equation (3). The results are shown in column 2 of Appendix Table 8; column 1 repeats the actual tax demand results for ease of comparison. The results are broadly similar, but show slightly higher demand for government assessors (93% increase in demand when we apply the formula, vs. 84% for official tax demand) and a more substantial increase for private assessors (72% when we apply the formula vs. 53% for official tax demand), suggesting that private assessors may have slightly under-applied the formulas even conditional on the areas measured. Nevertheless, even when we recalculate the tax demand using measured areas, the difference between government and private assessments remains large and statistically significant (p-value 0.03), so this is not the primary factor driving the differences.[22]

4.4 Heterogeneity in Targeting and Treatment Effects

The results presented thus far show that government assessors appear to generate larger treatment effects from reassessments than private assessors, based on the random assignment

[22]We also check whether each type of tax inspector differentially applied a change in tax formulas. Prior to our experiment, the government used to give a 25% reduction in tax for owner-occupied residential areas and a 10% reduction for owner-occupied non-residential areas. These reductions were canceled prior to the start of this experiment, but anecdotally may only have applied once reassessments took place. If this change in tax formula was applied equally by government and private assessors, as it should be, it would not explain the difference between them, though it may be related to why the overall increase in tax demand is larger in percent terms than the increases in area in Table 3. To examine this, in column 3 of Appendix Table 8, we recalculate the tax demand given measured areas, but this time we include these now-canceled rebates, so that un-assessed and re-assessed properties both receive rebates. Again government assessors increase demand much more than private assessors (p-value 0.02), but this brings the percent increase down from 84% to 64% for government assessors and from 53% to 45% for private assessors.

of wards to government vs. private assessors. To understand the sources of this difference, it is important to understand whether there is important heterogeneity in treatment effects across different types of properties. If government and private firms chose to focus on different sets of properties within a given ward, and there are heterogeneous treatment effects based on which type of property were chosen for reassessment, then part of the difference we are capturing could reflect the fact that government vs. private assessors could lead to different properties being assessed, and therefore different local average treatment effects. Moreover, if there are important heterogeneous treatment effects, that could suggest that an ideal assignment mechanism could consider using each type of surveyor where their comparative advantage is highest. This section explores these issues.

To begin, we first examine whether, in fact, government and private assessors differentially focus on different types of properties. Figure 3, Panel A plots the results from a series of locally-weighted regressions of the probability of reassessment by both government and private firms against log baseline tax demand; the right hand column also plots the difference between the two. To do so, we divide the x-axis into 100 equally sized points. For each point, we run a separate weighted regression, with weights given from a triangular kernel with width $\frac{1}{4}$ of the total width of the x-axis centered at that point. We then plot the results of each of these regressions as a function of the x point. We overlay a histogram of the data in the background.

These non-parametric regressions show that the government surveyors do appear to differentially target larger properties, though the differences are modest. Specifically, the difference in the probability of re-assessment between government and private firms is negligible for the smallest properties at baseline, but rises to about 12 percentage points for the largest properties.

To see whether this differential targeting explains part of the treatment effects we saw above, we then estimate whether there are also heterogeneous treatment effects for different types of properties. If treatment effects were homogeneous, then the fact that different types of properties were targeted would not matter, but if the difference in treatment effects between government and private surveyors is also heterogeneous, then we could be capturing different local-average treatment effects.

To investigate this, Panel B of Figure 3 repeats the same locally-weighted regression procedure for our main instrumental variables treatment regressions. That is, for each point on the x axis, we re-run our regression from equation (3), using the same local kernels as in Panel A. In this regression, one can think of the non-parametric regression results shown

in Panel A as capturing the differential first stage as we move across different values of baseline log tax demand, ensuring that we are capturing the different treatment-on-treated estimates at each point on the x-axis. As before the left panel shows the results separately for government and private reassessment compared with the control group; the right panel shows the difference between treatment effects for government vs. private firms.

Panel B shows that indeed, there are substantial heterogeneous treatment effects. While both government and private sector assessors find substantially more tax demand everywhere, their effects are similar for the the smallest properties. But while treatment effects for private re-assessment are essentially flat with respect to baseline tax demand, they rise dramatically for government assessors, so that the treatment effects of government reassessment are more than double for the largest properties. Note that since these are all locally weighted IV regressions, these are capturing heterogeneous treatment-on-treated effects at different levels of baseline demand levels, and do not reflect the differential reassessment rates shown in Panel A.

These heterogeneous treatment effects have several implications. Qualitatively, it is important to note that Figure 3 Panel B shows that the treatment effects for government assessors lie substantially (and statistically significantly) above the treatment effects for private assessors for almost the entire mass of the distribution. This means that qualitatively, the result that government assessors are more effective holds true essentially everywhere, and hence also holds true on average as well. To calculate this quantitatively, we can compute average treatment effects by integrating the heterogeneous treatment effects from Panel B of Figure 3 over the distribution of properties show in in the histogram. This moves us from estimating two different local average treatment effects over potentially two different complier sets, which we have in Table 2, to estimating average treatment effects holding the set of compliers fixed. When we do this, we find impacts that are very similar to the LATEs we estimated in Table 2 – we find that the average impact of a government re-assessment is Rs 6,092 and a private impact of Rs. 3,993. This shows that while there are heterogeneous treatment effects, because the differential probabilities of re-assessment in Panel A of Figure 3 are not that large, on net, differences in selection of who is reassessed do not substantially affect our results.[23]

[23]Note that this calculation in fact changes two things compared to the LATEs estimated by the IV regressions: a) it holds the set of properties assessed constant across both treatment groups and b) it considers all properties, not just those that were reassessed. If we consider (a) alone – i.e. holding the set of properties constant but not switching from actually reassessed properties to all properties, we find that the differential selection shown in Figure 3 Panel A combined with the heterogeneous treatment effects shown in Figure 3 Panel B explains only about 4 percent of the difference between government and private assessors.

These results are broadly similar because the probabilities of reassessment are fairly similar across the two treatment groups. But the heterogeneous treatment effects also suggest that the government assessors have a *comparative advantage* at reassessing high value properties, and so an intentional assignment rule could take advantage of this to increase average treatment effects overall. That is, if the government has a limited number of government assessors available and wants to supplement with private firms, the results suggest having the private firms be assigned to take on properties with low baseline values, and government assessors be assigned take on those with higher baseline values, could lead to much higher revenue allocations.

To investigate this, we consider instead what would happen if we assigned the smallest 50% of properties (based on baseline value) to the private firms, and the largest 50% of properties (again based on baseline value) to government re-assessors. This 'optimal assignment' would yield a combined average treatment effect of Rs. 5,393 per property, compared with an average treatment effect of Rs. 5,042 if we randomly assigned half of the properties to government and half to private firms. Thus, if the government is constrained in how many government assessors it has and needs to use some private firms to supplement, it could obtain an additional 7 percent more revenue (5382/5042) by optimally assigning the larger properties to government assessors.

4.5 Accuracy

Just because the government assessors have higher assessments and raise more revenues does not necessarily mean that they are better. After all, the goal is that these higher assessments should be the result of government surveyors assessing properties more accurately, rather than just systematically overcharging.

To investigate this, we use the fact that we have independent measures of a random subsample of re-assessed properties that we obtained from our own survey. Following Khan et al. (2016), we define *accuracy* as

$$Innaccuracy_i = \frac{|\hat{y}_i - y_i^*|}{\hat{y}_i + y_i^*} \times 2 \tag{4}$$

where y_i^* is our own independent assessment of a property's valuation from our survey and $\hat{y}_i$ is the official assessment by the government or private sector surveyors. This captures the percent difference between the official valuation and our independent valuation. Note that $Accuracy_i$ measures the absolute difference between the independent property valuation and

the official valuation, but it is non-directional. We also compute the directional equivalent, which measures systematic over- or under-valuation, as follows:

$$TaxGap_i = \frac{\hat{y}_i - y_i^*}{\hat{y}_i + y_i^*} \times 2 \tag{5}$$

This captures the average directional percent difference between the official valuation and our independent valuation.

The results are presented in Table 4. Note that since we only surveyed re-assessed properties, in Table 4 (unlike in previous analysis), we are comparing properties reassessed by the private firms vs. those re-assessed by the government workers, with government re-assessments as the omitted category, rather than control properties.

We find that the government assessors are substantially *more* accurate than assessments done by private firms, not less. Specifically, inaccuracy is about 16 percentage points higher for assessments done by private sector firms compared to government officials (p-value $<$ 0.01). On net, the key point here is that the differential increases in tax demand and government revenue we saw from government assessors compared to private assessors are associated with a higher rate of accuracy of assessments, not less.

It is worth noting that private firms were less accurate despite having explicit incentives to maintain the accuracy of assessments. Contracts signed by firms stated that at each payment milestone, government assessors would cross-check a 10% random sample of properties re-assessed by firms. If the area deviation between government and firm estimates exceeded 10%, the firm would have to pay a penalty, determined by the share of properties with substantial deviations.[24] Moreover, the government reserved the right to terminate the contract if a firm's assessment quality fell below the required standard.

5 Mechanisms

The results thus far show that reassessments by government workers appear to dominate those by private firms: they complete more properties, identify more complex cases, and generate higher tax revenues – and are more accurate, not less, when judged by an indepen-

[24]Penalties were based on the share of randomly sampled properties with a discrepancy >10% at each payment milestone. For each milestone, the GCC would deduct: 2% of the value of the payment due if 1–2% of properties had a discrepancy; 4% of the value if the share was 2–4%; 6% if the share was between 4–6%, and so on. If over 10% of randomly sampled properties had discrepancies, the government would deduct 20% of the payment owed to the firm.

dent firm.

In this section, we explore a series of mechanisms to understand why government surveyors appear to have been more effective. We explore a number categories of explanations: better *soft information*, greater *skills*, higher *effort*, and more *authority*. We find evidence that all of these channels – information, skills, effort, and authority – seem to be higher for government workers, suggesting that all of these mechanisms may be at play. We then use the experimental variation in contract types to ask whether stronger *incentives* for private firms would lead to better outcomes for private firms, but find that they do not - even paying firms based on the additional demand they uncover does not appear to meaningfully increase private firm performance.

5.1 Knowledge of Property Locations and Differential Property Targeting

We begin by exploring the role of knowledge. Local information may be important both to find where properties are located, and also decide which properties could have the highest returns to target and spend time on.

From our qualitative research, we learned that the firms encountered many challenges both finding properties and conducting the property measurements. The private sector assessors noted inaccuracies in both the property centroids and their door numbers, leading them to spend considerable time trying to find them, as well as leading them to visit the same streets multiple times (rather than complete all properties on a street at once).[25] Indeed, as shown in Figure 4, the private firms were more likely to state that they had difficulty finding properties than the government assessors (p-value < 0.001).

In addition, the private assessors had to spend more time explaining the visit's purposes to citizens relative to the city government assessors, and complained that they faced many refusals. This may have been worsened by the fact that private sector assessors did not always wear the city government-branded items. The private sector assessors expressed frustration at wandering around trying to find properties in the summer heat, sometimes encountering unleashed dogs, only to be not let in and abused by the citizens when they found the property.

It is also possible that the government had more knowledge than the private assessors of what properties would have the highest returns. Knowledge could reflect 'soft knowledge'

[25]This may also have been due to poor planning, where they would have to streets multiple times since they had not always planned to visit all properties on a street at once.

that comes from knowing the neighborhood if certain inspectors were able to figure out which properties had higher expected returns and to target their efforts accordingly.

To examine this, we examine whether whether the assessors are differentially assessing those with a higher potential *gain* in demand. We can construct this using the estimate of property demand from the drone survey. To isolate the role of 'soft information' we regress the this estimate of demand from the GIS drone survey on baseline demand consider the residuals; this captures that portion of expected increase in demand that cannot be predicted based on the 'hard information' available in the tax information. Recall that neither the GCC staff or private companies had access to the predicted demand from the GIS survey, so this is proxy for soft information that assessors may have had from, for example, knowing the neighborhoods and hence knowing that a property listed as residential actually had some commercial activities.

The results are shown in Figure 5, which shows the results from the type of non-parametric regressions used in Figure 3, but where we consider heterogeneity in soft information (residualized demand from the GIS drone surveys) as the key dimension of heterogeneity. As shown in Panel A, we do not find any evidence that properties that the drone survey suggests were under-valued were more likely to be re-assessed by either the government or private surveyors. We do find, however, that government assessors increase demand more in very undervalued properties (i.e, those properties with high expected gain in tax demand) compared to the private firms. This could suggest that they have some soft information on which properties to probe more, or perhaps just expended greater effort fully examining the properties.

5.2 Skill and Experience

The analysis in Section 4 provided several pieces of evidence of that suggest that higher *skill* may be an important difference between government assessors and private sector assessors: government assessors were more accurate in their assessments, and in particular, they were more likely to find irregular-shaped properties, which require measuring each floor separately rather than just taking a single measurement from the roof.

In Table 5, we use the assessors survey to explore whether there are observable characteristics that are consistent with the idea that government assessors are higher skill. First, in Column 1, we show that government assessors are much more experienced than the workers hired by private firms to perform the same task. Government assessors have an average of 11 years of work experience surveying properties; private sector assessors, by contrast, have

been doing this task for an average of only 2 years (p-value < 0.001). While work experience can accumulate over time, and hence would only be transitory, government workers also have more education. In particular, 85 percent of government assessors are college graduates (with 35 percent having a masters), compared to only 57 and 6 percent for private assessors, respectively (p-values < 0.001). The combination of higher education, and more experience, may be related to why they are more accurate in their assessments.

5.3 Effort

Government workers may also differ in the degree to which they put effort into the reassessment task. We measure this using survey data in two ways. First, we ask the surveyors the number of minutes they spent on each property. Table 5, column 4 shows that government surveyors spend 35 percent longer per property than private surveyors – 31 vs 22 minutes, respectively (p-value < 0.001). This is again consistent with our administrative data, which showed more "complicated" assessments (with different area values per floor) and more accurate assessments. Second, We examine the number of times surveyors revisit a property when they are not able to access it on the first visit. Both government and private surveyors are similar in that they each do 2 revisits, on average, when no one is home to let them in (Column 3).

Interestingly, the differences in the effort in measuring properties appear to be task-specific effort, rather than reflecting a general effort propensity. To test generalized effort, in our survey, we included the "A/B" task, from DellaVigna and Pope (2018). This task measures effort by asking the respondent to press the 'A' and 'B' keys on a computer keyboard, in sequence, as many times as possible for 1 minute. Figure 6 shows the results of the A/B task – which shows that actually private assessors do 34.5% percent more key presses than private sector assessors (p-value of the difference is 0.02). The effort differences we are detecting here, then, appears to be specific to the task in question rather than a pure capacity for effort or generalized motivation per se.

5.4 Authority

An important challenge in outsourcing tax assessments is authority - making sure the private workers had the same authority to enter and assess properties as government assessors. The government took several steps to accomplish this. First, ID cards, jackets and caps with the city government logos were procured and distributed to the assessors, so that they could

officially identify themselves – although in practice, not all the reported receiving all these items, leading to some firms printing some of them at their own expense. Second, all private surveyors were provided a letter on the official GCC letterhead stating that they had been authorized to enter and measure properties as part of the GCC's re-assessment drive. Third, the government posted photos of the the private inspectors alongside government inspectors on their Instagram page, so that people could verify that the private inspection process was legitimate. All properties identified for the reassessment were also published on the city government website so that taxpayers could check that their property was indeed scheduled to be reassessed.[26]

Nevertheless, the lack of perceived authority still appears to have been a constraint for private sector workers. Table 6, Column 1, shows that refusals – whereby a property owner refused access to the property – were 25 percent more likely for private assessors compared to government assessors – 19 percent vs. 15 percent of properties (p-value 0.07), even though both had the legal authorization to demand entry to a property for tax assessment purposes.

Perhaps related to this, private surveyors also reported more pushback from property owners about their assessments. Private surveyors reported that owners were 28 percentage points (72 percent; p-value < 0.001) more likely to request a change to the recorded area that the surveyor found, and 12 percentage points (60 percent; p–value < 0.09) more likely to request a change to usage type. While we do not know whether or not assessors actually made changes in response to these requests, it is possible they were more susceptible to this type of pushback than government assessors, leading perhaps in part to the overall lower assessed values.

Finally, as shown in Column 5, the government surveyors appear to try substantially harder if denied entry on the first visit – they revisit 2 times on average then as well, compared to only 0.59 times for private surveyors (p-value < 0.01). They also are less likely to close a file as non-collectible and take an alternative action (column 6) when refused entry. Importantly, whereas both the government and private surveyors both revisit if none one is home, the government employees may have felt more authority to go back and revisit if denied entry whereas the private sector employees may have felt that they did not have a right to do so.

[26]The government initially also planned to send text messages to the taxpayers that were identified for reassessment, followed by auto-rickshaw announcements and paper notices. The government did not ultimately do not this; instead, the fixed contract firms independently purchased speakers and initiated their own announcements in some areas.

5.5 Incentives

Given the performance of the private firms thaht we have seen, an important question is whether stronger *incentives* would meaningfully increase their performance. In general, government workers do not face strong explicit financial incentives – they have de facto tenure until retirement in their jobs, and promotions are based on seniority, degree qualifications, and qualifying exams rather than performance on assessments. They receive fixed salaries which are not explicitly tied to performance, and they do not have monthly assessment targets.[27] While there can be incentives to perform in the form of softer incentives (better postings within the city, perhaps, as in Khan et al. (2019), or informal recognition by leadership), a priori, one does not tend to think of government workers in this context as highly incentivized.[28] So given this, it does not seem a priori likely that incentives are stronger for the government than for private firms.

on the other hand, firms may face perverse incentives: if their bid was per property assessed, they could have an interest to go fast from property to property, which may be a reason for decreasing their accuracy (and for avoiding the complex cases). Indeed, the fastest way to perform re-assessment would be to not even go visit the property, but apply a mechanical increase in valuation. They may also be more receptive to complaints by the taxpayers to speed up the process. For example, (Duflo et al., 2013) find that, in Gujarat, private pollution auditors (who were hired as a court ordered third party audit scheme because the court found the government to inefficient) mostly declared the firms to be in compliance, without performing the audits.[29]

To explore whether stronger output based incentives for private firms would make a difference, we use the fact that, within the set of properties which were being assessed by private firms, the government randomized properties be assessed either by firms with a fixed fee contract or with a linear incentive contract, wherein firms were paid 4.8 percent of the change in tax demand for all properties re-assessed.[30] The size of these incentives meant that firms were paid about the same on average per property re-asssesed, but the fixed-fee

[27] The GCC imposes monthly revenue targets for tax collectors who are supervised by assessors.

[28] Non-monetary incentives such as medals and awards are limited, and typically only awarded to one or two assessors per year.

[29] In that instance, the incentives were more perverse, since the auditors were to be hired, selected and paid by the firm.

[30] Note that the incentives to firms translated into different incentives faced by employees, but somewhat strangely, they they did not bring the revenue-percentage sharing down to the employee level: The fixed firms paid a flat salary to their surveyors, while 89% of incentive firm surveyors reported that they were being paid a fee for each property they re-assessed (between 50-100 rupees per property).

firms were paid a fixed rate per property re-assessed, whereas the incentive firms received almost all of their pay from the linear incentives based on tax demand.[31]

Table 7 presents the results on the key outcomes we examine – number of properties re-assessed, tax demand, tax collections, and accuracy – where separate out assessments by fixed-fee private assessors vs. incentive-based private assessors based on the ward-level randomization.[32] We find no statistically significant differences in performance between fixed fee and incentive-based firms on any of the dimensions we examine. As shown in Column 2 and 3, while the incentive firm has slightly higher demand and collection than the fixed firm in magnitude, we cannot reject that they are different from one another, and we can reject that they are both different than the government.

This result suggests that differences in financial incentives are unlikely to be driving the differences between the government and the private firms that we observe. While it is possible of course that larger incentives could have generated larger effects, the incentives here were such that the total cost to the government was about the same in both treatments, and at that level of incentive, we find no differences between the two approaches.

6 Cost effectiveness and implications

We finally turn to considering the cost-effectiveness of these two alternative approaches. For private firms, we know the monetary costs the government paid the firm, so determining costs is straightforward. For government inspectors, from our survey data, we can calculate an approximate amount of time that they spend on conducting reassessments, and then multiple it by their salary. [33]. We report estimates under different assumptions for how long the increases in revenue last.

For each alternative, we calculate the net benefits to the government, i.e. $NetBenefits = E[\Delta Revenue] - E[Costs]$. As long as $NetBenefits > 0$, the re-assessments are expected

[31] To see this, note from Table 2 that private firms on average raised tax demand by Rs. 3859. Their average pay per property was therefore Rs. $50 + 0.05 \times 3859 = 245$, very similar to the Rs. 224 per property paid to the fixed fee firms.

[32] In Appendix Table 9, we also examine the specific property characteristics.

[33] Specifically, we first take the number of hours they work in a day and multiple by the fraction of days that they work in a week on reassessments to have the total time in a day conducting reassessments. We divide that by the average number of reassessments that they complete in the day, and multiple it by their wage. We had many refusals for time spent per day working, so we have fewer observations to calculate the cost. Alternatively, we also calculated costs and net benefits using the time spent on the assessment, and we find similar costs (about 199) Given that this measure omits the time spending finding property, despite having less data, we use the full calculation above.

to be profitable for the government. If the government is not budget constrained in how much it can allocate to this task – which presumably it is not, as long as it produces positive expected revenues – it should choose whichever approach has the highest $NetBenefits$.

Table 8 presents the results. Under all assumptions for the duration of expected benefits, the expected net benefits are substantially larger using government assessors than private firms. For example, if results last 5 years, the net benefits from using government enumerators are 58% larger than from using private firms.

More generally, it is worth noting that both approaches have extremely positive returns. For example, if results last 5 years, the returns to the government are 8,288% to 10,566%. If the higher assessments last for 10 years – plausible given that they had not been systematically reassessed in years – returns are between 16,675% to 21,233%.[34] Thus, while the returns are substantially larger from using government surveyors, both approaches are staggeringly cost effective.

Given that they are much better at the task, one can argue that the government should perform it. Moreover, it is not clear that they would have to hire more workers to do so. When they are not helping to collect on tax bills twice a year, the government workers often have downtime in which they engage in other government activities, e.g. flood control. Thus, the assessments could be done during these non-collection periods. On the other hand, given than both the government and the private firms have very high net benefits, one could argue that if the government is constrained by manpower, it should devote as many people to the task as possible, and then hire a private firm to do the rest, potentially planning to conduct a re-assessment faster for the firms that have been allocated to private firms. Another option would be for the firms to be assigned "easier" properties (if they could be identified), or smaller properties where mistakes are less costly in terms of future revenue stream to the private firm, and let the government employees handle the larger or harder properties.

7 Conclusion

Economic theories of outsourcing often highlight the tension between a more efficient private sector that can "get things done" much better than the government, but that is stymied by contracting failures that end up providing the wrong incentives to firms. However, it is not

[34]Given that this was the first large-scale re-assessment drive implemented by the GCC in history, it is certainly plausible that these higher valuations would last for 10 years or more. Previous exercises, conducted sporadically between 2007 and 2022, have focused exclusively on commercial properties and were limited to approximately 10,000 properties.

clear that this tradeoff paints a full and nuanced picture of the challenges of outsourcing.

In this paper, we experimentally compare in-house property tax assessments conducted by the government to assessments that had been outsourced to private firms. We find that that government was more effective than the private firms: not only did the government employees assess more properties over the same time period — despite having other responsibilities as well – but they also raised more money from each reassessed property. In fact, the government employees were more accurate with each measurement and find substantially more area with complex, irregular measurements.

Prima facie, the private firms should have had incentives to be as efficient as possible (possibly at the cost of cutting corners and thus being less accurate); the government's solution to this concern was therefore to incentivize the firms based on the amount of revenue they collected. However, interestingly, the specific contract incentives did not matter much: the fixed fee contracts for private firms led to similar revenue gains as an incentive contract which paid out based on the increased revenue from the reassessments. The fact that moving from flat fee payments to revenue-based incentives did not change much suggests that these types of incentive issues may not have been at the heart of the outsourcing challenge.

Rather, we find two other classes of potential explanations for the differences between the in-house and private assessments. First, some of this difference may have been driven by the types of firms that opted to bid on the contracts. Despite the initial interest, very few firms actually bidding. Many of the firms were frustrated by frequent payment delays for past government projects. In fact, those that did not bid also believed that the prices that the government demanded were not worth the effort. In retrospect, this was perhaps practical foresight, as the government unexpectedly halted the project early on in the contract due to changing the political situation, and the firms that were contracted by the government ended up facing long delays on being paid for the work they did complete. As such, given this selection of firms, it may not have been surprising that the government employees were more educated, had more experience, and exerted more effort in the tasks than the temporary workers that the firms recruited to complete the contract.

Second, government officials also had more authority with citizens to enter properties and interact with citizens for the inspections, despite the government trying to confer authority onto the private firms' workers. Moreover, it may be hard for the government to equip the firms with the 'soft' information they need to complete the tasks. For example, the government officials reported having less trouble finding properties from the same information than the firms.

Taken together, this suggest that the limits to government outsourcing may go beyond multitasking issues, and instead that there may be certain tasks – such as tax assessments as we study here – where government provision is actually more efficient than than the private sector. Developing a clear understanding of for which sets of tasks this is likely to be true, and which it is not, is an important direction for future research.

These results also should lead us to question the frequent conclusion drawn by academics, policy makers, or citizens, when they observe that the government appears to be very inefficient at some of the tasks it does, compared to private firms. Many of the tasks the government are particularly difficult to carry out (because after all, no one wants to pay taxes, or to follow pollution or safety regulations). And we see that at least in this context, private firms are worse than the government when trying to carry out the same tasks.

In fact, the main difficulty in collecting taxes in this setting was perhaps not the technical ability to collect the relevant information, but rather the political willingness for the government to enforce its own tax code. While we show that reassessments by government employees are substantially more more effective than those done by private firms, both types of reassessments were astoundingly cost-effective. Yet several months in to the re-assessment exercise, after some taxpayers complained about the re-assessment and in advance of an election, the entire exercise was stopped, even though it was very clear by then that this would be a very significant source of revenues. While most political economy analyses focus on changes to the tax code, this suggests that understanding the political economy of tax enforcement itself may be an important direction for future work.

References

Balán, P., Bergeron, A., Tourek, G., and Weigel, J. L. (2022). Local elites as state capacity: How city chiefs use local information to increase tax compliance in the democratic republic of the congo. *American Economic Review*, 112(3):762–797.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., and Sumarto, S. (2019). Private outsourcing and competition: Subsidized food distribution in indonesia. *Journal of Political Economy*, 127(1):101–137.

Banerjee, A., Hanna, R., and Mullainathan, S. (2012). Corruption. In Gibbons, R. and Roberts, J., editors, *The Handbook of Organizational Economics*, pages 1109–1147. Princeton University Press.

Behaghel, L., Crépon, B., and Gurgand, M. (2014). Private and public provision of counseling to job seekers: Evidence from a large controlled experiment. *American Economic Journal: Applied Economics*, 6(4):142–174.

Bergeron, A., Bessone, P., Kabeya, J. K., Tourek, G. Z., and Weigel, J. L. (2022). Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the drc. Technical report, National Bureau of Economic Research.

Bergeron, A., Tourek, G., and Weigel, J. L. (2024). The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the drc. *Econometrica*, 92(4):1163–1193.

Brockmeyer, A., Estefan, A., Arras, K. R., and Serrato, J. C. S. (2021). Taxing property in developing countries: Theory and evidence from mexico. Technical report, National Bureau of Economic Research.

Cheeseman, N. and de Gramont, D. (2017). Managing a mega-city: learning the lessons from lagos. *Oxford Review of Economic Policy*, 33(3):457–477.

DellaVigna, S. and Pope, D. (2018). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2013). Truth telling by third-party auditors and the response of polluting firms. *The Quarterly Journal of Economics*, 128(4):1499–1545.

Fjeldstad, O.-H., Katera, L., and Ngalewa, E. (2008). Outsourcing revenue collection: Experiences from local government authorities in tanzania. Technical report, Research on Poverty Alleviation.

Galiani, S., Gertler, P., and Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1):83–120.

Haas, A. and Manwaring, P. (2017). Enhancing local tax revenues: private or public collection? Technical report, IGC Cities that Work Policy brief.

Hart, O., Shleifer, A., and Vishny, R. W. (1997). The proper scope of government: theory and an application to prisons. *The Quarterly Journal of Economics*, 112(4):1127–1161.

Iversen, V., Fjeldstad, O.-H., Bahiigwa, G., Ellis, F., and James, R. (2006). Private tax collection—remnant of the past or a way forward? evidence from rural uganda. *Public Administration and Development*, 26(4):317–328.

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics*, 131(1):219–271.

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2019). Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review*, 109(1):237–270.

Levin, J. and Tadelis, S. (2010). Contracting for government services: Theory and evidence from us cities. *The Journal of Industrial Economics*, 58(3):507–541.

Romero, M., Sandefur, J., and Sandholtz, W. A. (2020). Outsourcing education: Experimental evidence from liberia. *American Economic Review*, 110(2):364–400.

Yang, D. (2008). Integrity for hire: an analysis of a widespread customs reform. *The Journal of Law and Economics*, 51(1):25–57.

Table 1: Descriptive Statistics: Bidders vs Non-bidders

	Bidder	Non-bidder
	(1)	(2)
Panel A: Firm Characteristics		
Firm Age	22.75	22.33
	(4.92)	(26.82)
Permanent Staff Count	72.50	42.18
	(85.39)	(64.05)
Temporary Staff Count	55.00	24.30
	(63.51)	(44.51)
Turnover: INR 10,00,000 to INR 50,00,000	0.00	0.33
	(0.00)	(0.49)
Turnover: INR 50,00,000 to INR 1,00,00,000	0.00	0.33
	(0.00)	(0.49)
Turnover: INR 1,00,00,000 to INR 5,00,00,000	0.50	0.00
	(0.58)	(0.00)
Turnover: More than INR 5,00,00,000	0.25	0.25
	(0.50)	(0.45)
Panel B: Past Experience with Government		
Past Tender Bids	1.00	0.64
	(0.00)	(0.50)
Past Tender Win	0.75	0.43
	(0.50)	(0.53)
Government Temporarily Halted Past Projects	0.25	0.00
	(0.50)	(0.00)
Government Delayed Payments in the Past	1.00	0.83
	(0.00)	(0.41)
Panel C: Contract Preferences		
Contract Preferred: Incentive	0.50	0.25
	(0.58)	(0.45)
Contract Preferred: Fixed	0.50	0.58
	(0.58)	(0.51)
Contract Preferred: Indifferent	0.00	0.17
	(0.00)	(0.39)
Observations	4	12

Note: This table reports descriptive statistics comparing bidder and non-bidder firms. The data come from a survey of firms that we conducted.

Table 2: Difference between Properties Re-Assessed by Government Assessors and Firms on Tax Demand and Collection

	Reassessed (Apr. 2023)	Reassessed (Apr. 2024)	Tax Demand (Rs.)	Tax Collection (Rs.)
	(1)	(2)	(3)	(4)
Panel A: Reduced Form Estimates				
Assigned to government	0.473***	0.471***	2867.13***	1869.38***
	(0.043)	(0.042)	(366.79)	(272.64)
Assigned to private	0.338***	0.377***	1454.73***	797.01***
	(0.024)	(0.022)	(190.43)	(100.57)
Government = Private	0.006	0.050	0.00	0.00
Panel B: Instrumental Variables Estimates				
Reassessed by government			6080.31***	3964.66***
			(540.06)	(431.04)
Reassessed by private			3859.76***	2116.19***
			(378.13)	(230.27)
Government = Private			0.00	0.00
Control Mean	0.000	0.001	7218.18	5721.16
Observations	206257	206257	206257	206257

Note: This table experimentally estimates the difference in the likelihood of re-assessment, tax demand, and tax collection for properties assigned to government tax assessors and private firms. Reassessed (Aug 2023) is an indicator variable capturing whether the property was re-assessed by August 2023. Reassessed (Apr 2024) is an indicator variable capturing whether the property was re-assessed by April 2024. Tax Demand (Rs.) is the property tax liability in the 2023-24 tax year. Tax Collection (Rs.) is the property tax paid in the 2023-24 tax year. Panel A reports reduced-form effects of random assignment to government vs. private assessors on each outcome. Panel B reports 2SLS estimates of the effect of reassessment on tax outcomes, instrumenting reassessment with treatment assignment. All specifications include ward fixed effects and columns (3) and (4) include controls for the outcome variable at baseline. Tax demand data come from the DCB dataset and Tax Collection data come from the Receipts dataset. We impute tax collection as zero if there is no receipt associated with that property in 2023-24 in the Receipts Data. The sample is limited to the first list of properties the agents were instructed to re-assess and all control properties. Standard errors are clustered at the ward level and displayed in parentheses. We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Impacts of Reassessment on Property Characteristics (IV Estimates)

	Area (sq. ft.)				Property Usage		
	Total	Residential	Non-Residential	Floor Count	Usage Upgraded	Usage Downgraded	Irregular-shaped Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reassessed by government	573.690***	507.209***	41.884***	0.572***	0.042***	0.009***	0.241***
	(44.19)	(39.59)	(9.06)	(0.04)	(0.01)	(0.00)	(0.02)
Reassessed by private	476.591***	439.087***	11.025	0.494***	0.025***	0.017***	0.188***
	(30.31)	(26.97)	(7.19)	(0.03)	(0.01)	(0.00)	(0.02)
Government = Private	0.07	0.154	0.008	0.147	0.038	0.021	0.056
Control Mean	1338.762	1108.937	206.424	1.869	0.015	0.004	0.445
Observations	206257	203583	203583	203583	206257	206257	203572

Note: This table decomposes the experimental estimate of the difference in tax demand for properties re-assessed by government tax assessors and private firms to determine whether the increase in demand was driven by increases in built-up area or property usage types. Property Usage Upgraded is an indicator recording whether the property usage was upgraded relative to baseline as a result of the re-assessment, including switches from residential to mixed-use and from mixed-use to non-residential. Property Usage Downgraded is an indicator recording whether the property usage was downgraded relative to baseline, including switches from non-residential to mixed-use and from mixed-use to residential. Irregular-shaped Property identifies multifloor properties where the floors have different areas. All specifications include ward fixed effects and controls for the respective outcome variable at baseline. Area and Property Usage data come from the ERP Master Table. Area data at baseline comes from the Floor Table. The sample is limited to the first list of properties the agents were instructed to re-assess and all control properties. Standard errors clustered by ward. We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Accuracy of Reassessed Properties (OLS)

	Tax Gap	Inaccuracy
	(1)	(2)
Assigned to private	-0.119^{**}	0.156^{***}
	(0.051)	(0.037)
Government = Private	0.02	0.00
Government Mean	-0.11	0.25
Observations	1,481	1,481

Note: This table experimentally estimates the difference in non-revenue outcomes for properties re-assessed by government tax assessors versus private firms. Tax Gap is the difference in the official Tax Demand for the 2023-24 tax year minus our estimated Tax Demand, divided by the sum of the two, and multiplied by 2. Estimated Tax Demand was calculated based on property characteristics recorded during the Property Survey and the government's property tax formula. Inaccuracy is the absolute value of Tax Gap. Tax Demand data come from the DCB dataset. We estimate OLS regressions where the omitted category is properties re-assessed by the government. All specifications include controls for the outcome variable at baseline, i.e. the tax gap or Inaccuracy based on the official tax demand in 2022-23. Standard errors are clustered at the ward level. We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Differences in Skill and Effort between Government and Private Firms

	Skill			Effort	
	Total years experience surveying properties	Graduate degree or above	Master's degree or above	Average time spent on measuring one property	Number of revisits when no-one was present at the property
	(1)	(2)	(3)	(4)	(5)
Assigned to private	-9.60***	-0.28***	-0.29***	-8.01***	-0.42
	(1.09)	(0.07)	(0.06)	(1.97)	(0.33)
Government = Private	< 0.001	< 0.001	< 0.001	< 0.001	0.206
Government Mean	11.03	0.85	0.35	30.72	1.96
Observations	147	147	147	139	147

Note: This table presents differences in employee authority between government assessors and surveyors hired by private firms. We estimate OLS regressions of the outcome variable on an indicator for assigned to private firms (compared to government) with robust standard errors We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Differences in Authority between Government and Private Firms

	Authority				Revisits/Action	
		Owner requested change to...				
	Share of refusals	Property area	Property usage	Number of floors	Number of revisits when owner denied permission for re-assessment	Action taken after refusal/ owner absence
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned to private	0.04*	0.28***	0.12*	0.06	-1.42***	-0.39***
	(0.02)	(0.08)	(0.07)	(0.07)	(0.30)	(0.08)
Government = Private	0.069	< 0.001	0.100	0.340	< 0.001	< 0.001
Government Mean	0.15	0.35	0.20	0.16	2.01	0.70
Observations	137	147	147	147	147	144

Note: This table presents differences in employee authority between government assessors and surveyors hired by private firms. We estimate OLS regressions of the outcome variable on an indicator for assigned to private firms (compared to government) with robust standard errors We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Instrumental Variables Estimates of Reassessment by Contract Type

	Reassessed (Apr 2024)	Tax Demand (Rs.)	Tax Collection (Rs.)	Tax Gap	Inaccuracy
	(1)	(2)	(3)	(4)	(5)
Reassessed by government	0.471***	6078.154***	3965.899***		
	(0.04)	(540.70)	(431.24)		
Reassessed by fixed-contract firms	0.369***	3465.956***	2041.353***	-0.087	0.149***
	(0.02)	(462.51)	(295.49)	(0.06)	(0.04)
Reassessed by incentive-contract firms	0.392***	4555.006***	2248.323***	-0.148**	0.162***
	(0.05)	(620.68)	(363.80)	(0.07)	(0.05)
Government = Fixed	0.035	<0.01	<0.01	0.150	<0.01
Government = Incentive	0.211	0.064	0.002	0.042	0.002
Fixed = Incentive	0.667	0.160	0.659	0.493	0.826
Control Mean	0.00	7250.66	5804.43		
Government Mean				-0.11	0.25
Observations	206255	206255	206255	1481	1481

Note: This table shows experimental effects of property reassessment by government assessors and private firms. Column 1 presents the first stage estimates. Columns 2 and 3 show treatment effects using 2SLS estimation with treatment assignment as instruments. Columns 4-5 show OLS estimates comparing private firms to government assessors (reference category). Tax Demand and Collection are for 2023-24. Tax Gap is the difference in the official Tax Demand for the 2023-24 tax year minus our estimated Tax Demand, divided by the sum of the two, and multiplied by 2. Estimated Tax Demand was calculated based on property characteristics recorded during the Property Survey and the government's property tax formula. Inaccuracy is the absolute value of Tax Gap. All specifications include ward fixed effects and baseline controls. Standard errors clustered by ward. We provide p-values for the tests of equality between government and fixed-contract firms, government and incentive-contract firms, and and fixed-contract and incentive-contract firms. Control Mean shows means for untreated properties (columns 1-3). Government Mean shows means for government-assessed properties (columns 4-5). *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Net Benefit Analysis (Demand-Based Estimates)

	Duration (in years)		
	1	5	10
Panel A: Expected Benefit Per Property			
Government	6080.31	30401.53	60803.05
Private Firms	3859.76	19298.81	38597.61
Panel B: Costs Per Property			
Government	285.02	285.02	285.02
Private Firms	230.09	230.09	230.09
Panel C: Net Benefit Per Property			
Government	5795.28	30116.50	60518.03
Private Firms	3629.67	19068.71	38367.52

Note: Net benefit equals the average treatment effect on Tax Demand minus the average cost of reassessment. The government per-property cost is estimated using employee survey data as total salary per employee (monthly salary × project duration), divided by the number of properties handled per employee and adjusted by the fraction of total hours spent on reassessment. We assume constant annual benefits over the specified duration periods.

Figure 1: Experimental Design

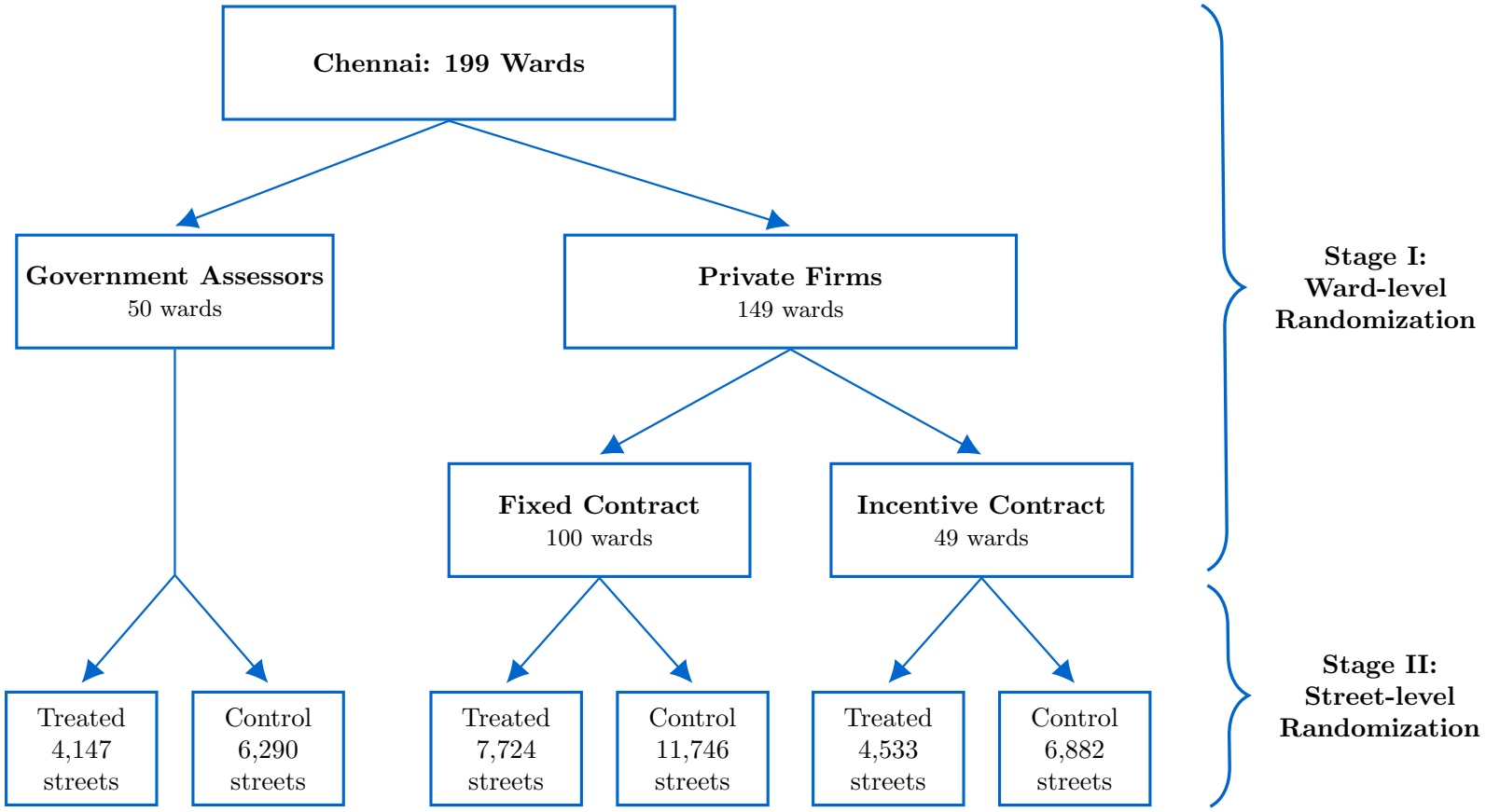
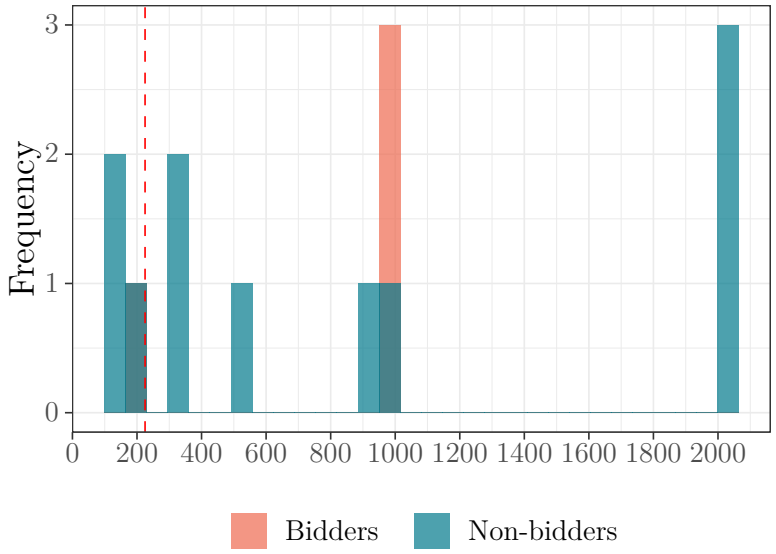
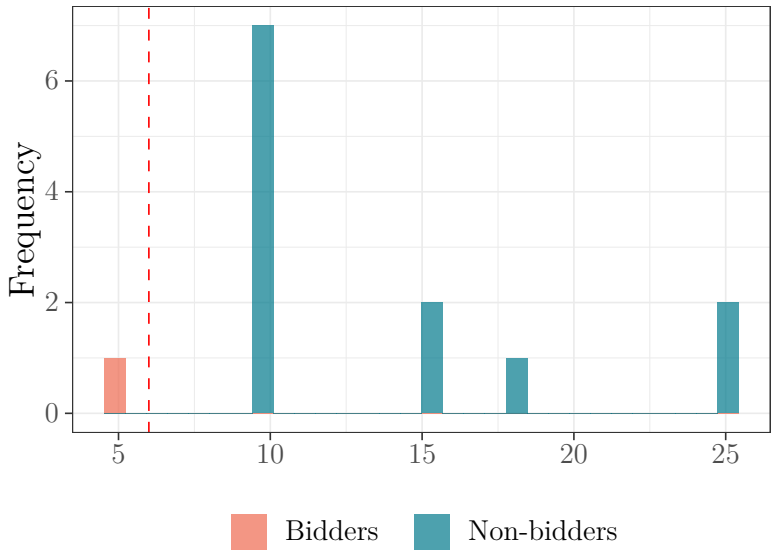


Figure 2: Proposed Bids for Re-assessment Contracts

Panel A: *Fixed firms*



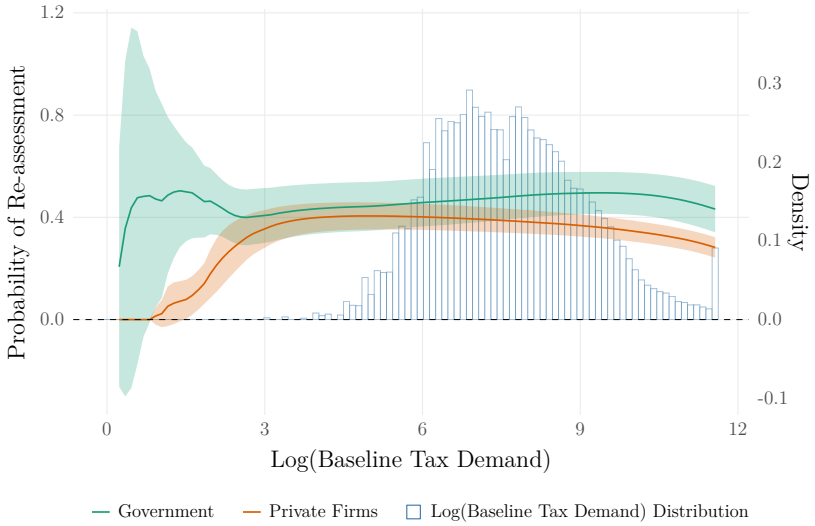
Panel B: *Incentive Firms*

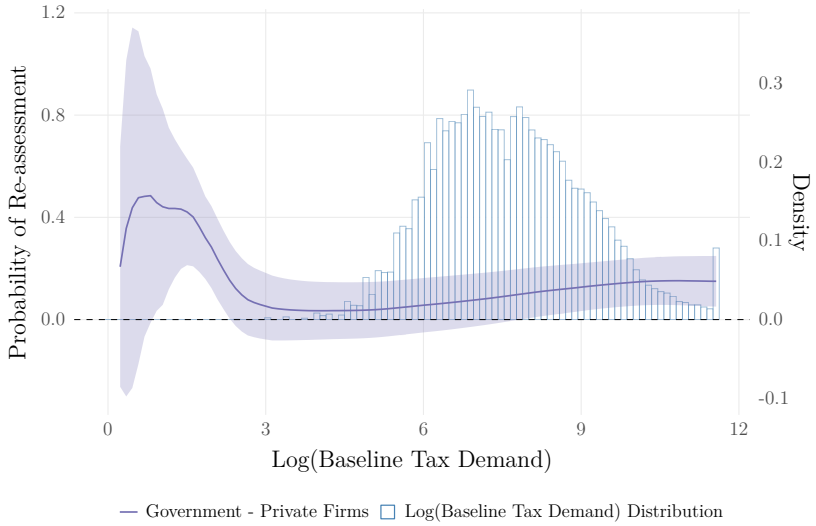


Note: The figure plots the frequency distribution of bids that bidders submitted and those that non-bidders would have submitted for the fixed (Panel A) and incentive (Panel B) contracts, respectively. A "bid" corresponds to the price charged for re-assessing one property (Panel A), and the percentage of increase in Tax Demand (Panel B). Non-bidders were asked hypothetical questions about how much they would have charged to reassess a property under the fixed and incentive schemes. The dashed line in Panel A depicts the final price per property under the Fixed Contract (Rs. 224). The dashed line in Panel B depicts the maximum incentive payment bid specified by the government during the bidding process (6% of the increase in Tax Demand).

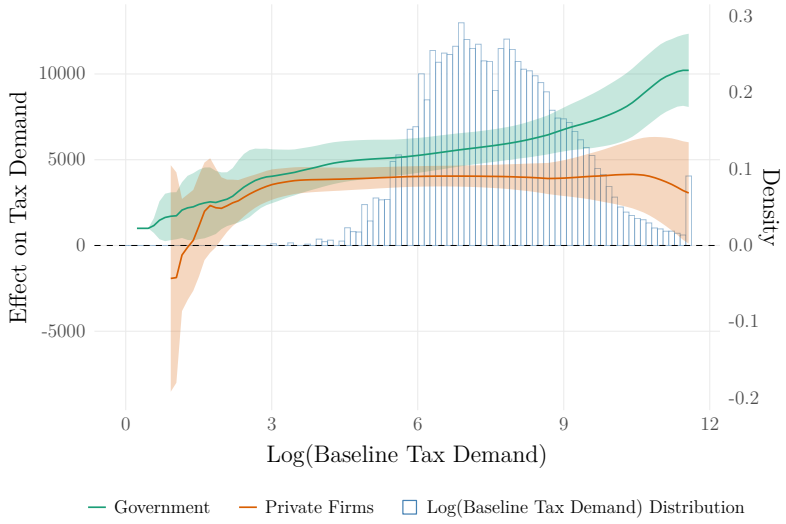
Figure 3: Heterogeneity in Reassessments and Demand, by Log(Baseline Tax Demand)

Panel A: First-stage Impacts on Probability of Reassessment





Panel B: IV Estimates of Heterogeneous Treatment Effects on Tax Demand



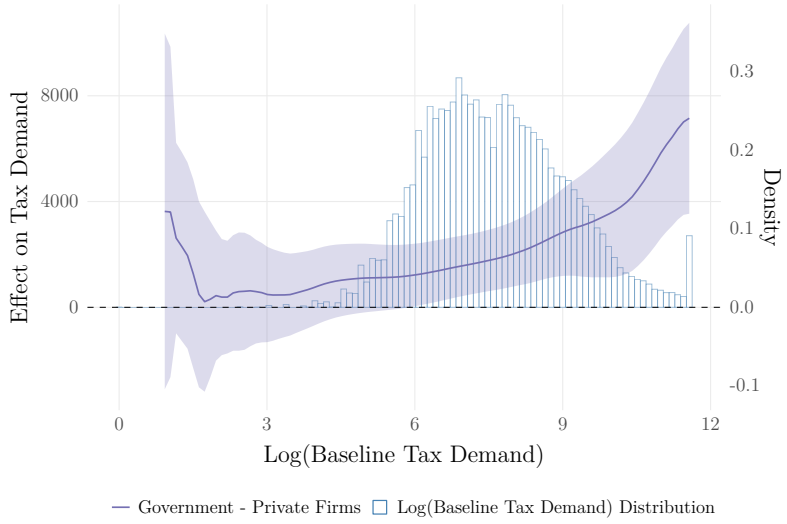
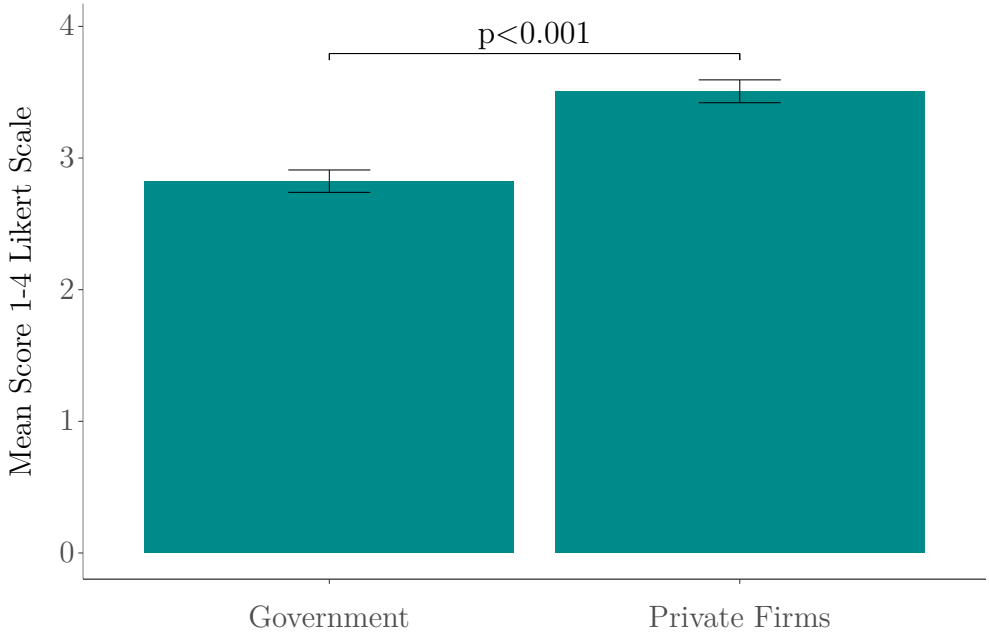


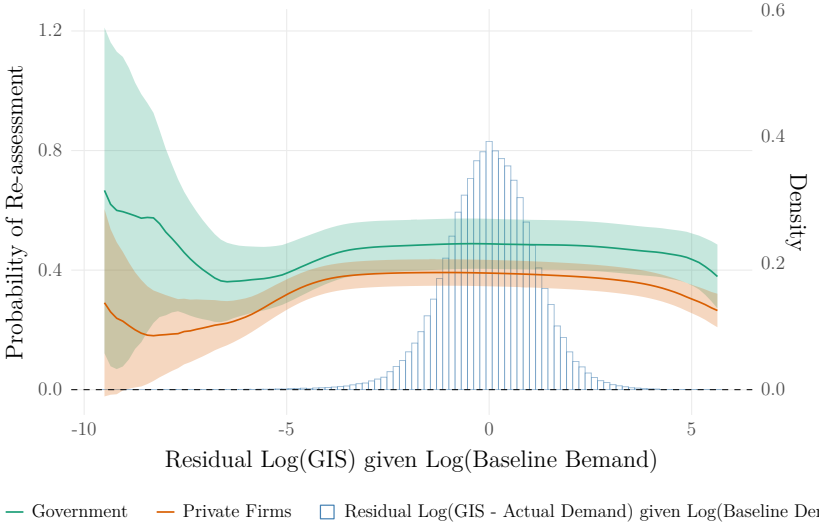
Figure 4: Difficulty in Locating Property

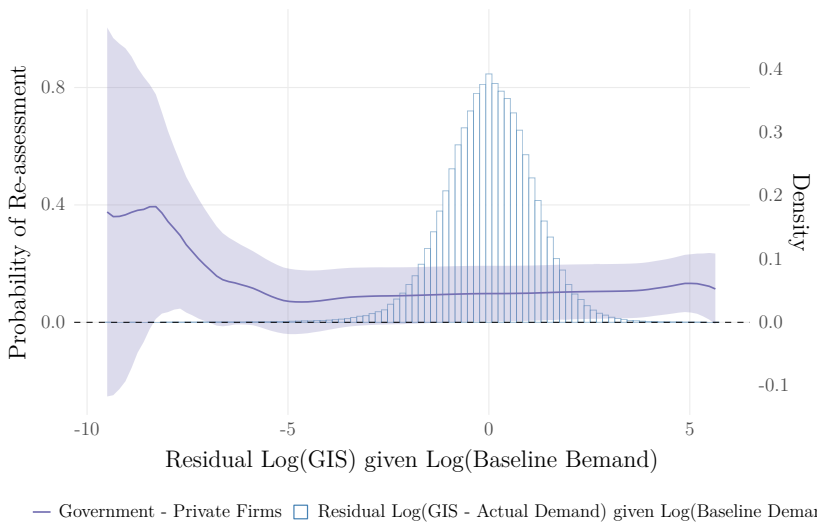


Note: The figure depicts the mean rating of self-reported difficulty in finding property locations of government assessors and private firms. Difficulty was reported on a 1–4 Likert scale, where 1 is labeled as 'Big Challenge', 2 as 'Moderate Challenge', 3 as 'Easy', and 4 as 'Very Easy'. The length of the error bars corresponds to two times the robust standard errors obtained when regressing the likert scale difficulty ratings on dummy variables for private firms and government without including an intercept.

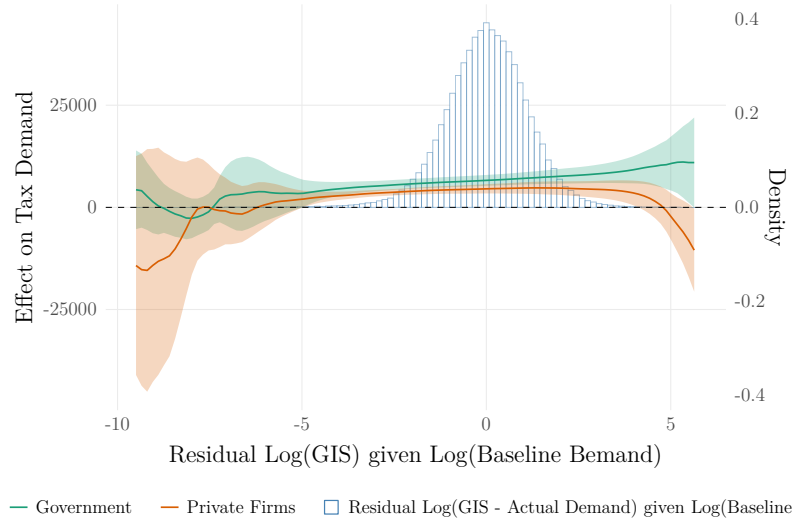
Figure 5: Targeting Undervalued Properties

Panel A: First-stage Impacts on Probability of Reassessment





Panel B: IV Estimates of Heterogeneous Treatment Effects on Tax Demand



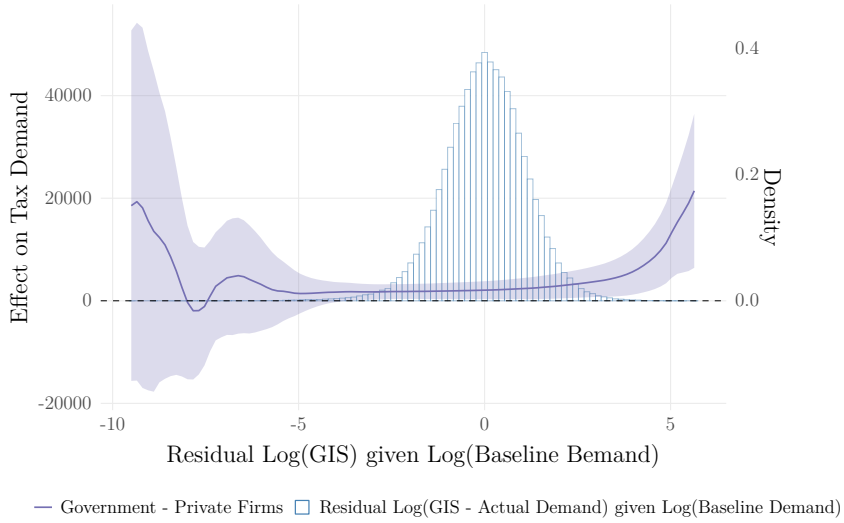
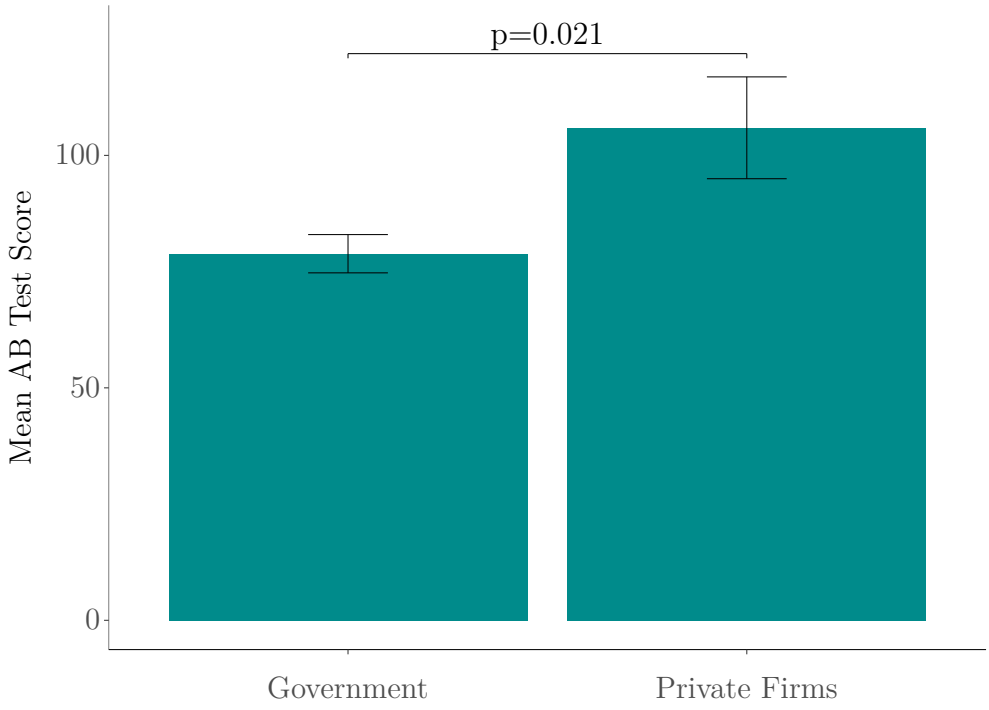


Figure 6: AB Test Score of Government vs Private Workers



Note: The figure depicts average scores on the AB effort test administered to government assessors and surveyors hired by private firms, and its standard errors In the AB effort test, respondents were asked to alternately press the 'a' and 'b' keys on their keyboard as quickly as possible for 1 minute to score points, and could score as many points as they liked. A higher Test Score is indicative of more effort. The length of the error bars corresponds to two times the robust standard errors obtained when regressing the AB Test score on dummy variables for private firms and government without including an intercept.

Appendix Tables and Figures

Appendix Table 1: Property Survey Response Rates, by Treatment Arm

	P(Reassessed)	P(Reassessed)
	(1)	(2)
Assigned to private firms	0.005	
	(0.021)	
Assigned to fixed-contract firms		0.034
		(0.025)
Assigned to incentive-contract firms		−0.019
		(0.024)
Government Mean	0.690	0.690
Observations	2,150	2,150

Note: This table presents results on response rates in the Property Survey, by treatment groups. The outcome is an indicator for whether the property was successfully surveyed. The sample includes treated properties that were randomly selected for the Property Survey. We estimate OLS regressions where the omitted category is properties assigned to the government. Standard errors are clustered at the ward level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 2: Technical Proposal Evaluation Criteria

Criteria	Marking System	Max Marks
A. Bidder's Prior Experience		
Experience conducting Property Enumeration / Assessment or similar tasks	• Relevant projects > 20,000 survey points: **30 marks** • Relevant projects > 10,000 assessment: **20 marks** • Relevant projects > 5,000 survey points: **10 marks**	30
B. Approach & Methodology		
Approach & Methodology	• Presentation on project understanding and value proposition • Adequacy of work plan responding to Terms of Reference • Project Team & resource-wise KPIs	40
C. Project Key Resource Profiles		
Project Manager	**Qualification:** Bachelors or equivalent in Urban Planning / Engineering / Geography / Management / Public admin / Commerce / Economics: **4 marks** **Experience as Project Manager:** • More than 7 years: **5 marks** • 5-7 years: **4 marks** • 3-5 years: **2 marks** **Experience as Project Manager in Municipal admin/Finance:** • More than 5 projects: **6 marks** • 3-5 projects: **4 marks** • 1-3 projects: **2 marks**	15
Team Lead	**Qualification:** Bachelors or equivalent in Urban Planning / Engineering / Geography / Management / Economics / Public admin / Commerce disciplines: **4 marks** **Experience as Team Lead:** • More than 7 years: **5 marks** • 5-7 years: **4 marks** • 3-5 years: **2 marks** **Experience as Team Lead in Municipal admin/Finance/Property tax:** • More than 5 projects: **6 marks** • 3-5 projects: **4 marks** • 1-3 projects: **2 marks**	15
	Total Maximum Marks	**100**

Appendix Table 3: Randomization Check (OLS estimates)

	Tax Demand (2022-23)	Tax Collection (2022-23)	Property Area (square feet)	Share Mixed-Use Properties (Government Records)	Share Non-Residential Properties (Government Records)	Share Mixed-Use Properties (GIS Survey)	Share Non-Residential Properties (GIS Survey)	Share Rental Properties	Share Owner-Occupied Properties	Share Mixed Tenancy Properties
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Assigned to government	276.868	182.303	17.222	0.004	0.000	0.002	0.000	-0.006*	-0.005	0.011**
	(187.99)	(175.80)	(20.43)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
Assigned to private	229.182	230.007	7.035	-0.002	0.006*	0.001	0.006	0.000	0.002	-0.002
	(181.64)	(154.74)	(14.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Government = Private	0.855	0.839	0.681	0.316	0.261	0.885	0.335	0.138	0.373	0.029
Control Mean	6454.78	5639.72	1295.44	0.10	0.06	0.13	0.08	0.07	0.74	0.18
Observations	206255	206255	206255	206255	206255	206255	206255	203581	203581	203581

Note: This table provides a baseline balance check. Each column presents the coefficients from running an OLS regression of a baseline variable on the treatment arms variables. Regressions include ward fixed effects and standard errors are clustered are the ward level. Data on tax demand come from the DCB Table. Data on tax collection come from the Receipts Table. We provide p-values for the test of equality between the government and private firms. Standard errors clustered by ward. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 4: Difference between Properties Re-Assessed by Government Assessors and Firms on Tax Demand and Collection without baseline controls

	Reassessed (Apr. 2023)	Reassessed (Apr. 2024)	Tax Demand (Rs.)	Tax Collection (Rs.)
	(1)	(2)	(3)	(4)
Panel A: Reduced Form Estimates				
Assigned to government	0.473***	0.471***	3162.19***	2025.73***
	(0.043)	(0.042)	(460.37)	(311.77)
Assigned to private	0.338***	0.377***	1698.10***	992.52***
	(0.024)	(0.022)	(279.57)	(188.56)
Government = Private	0.006	0.050	0.01	0.00
Panel B: Instrumental Variables Estimates				
Reassessed by government			6706.87***	4296.50***
			(748.94)	(514.58)
Reassessed by private			4505.53***	2634.28***
			(651.45)	(479.47)
Government = Private			0.03	0.02
Control Mean	0.000	0.001	7218.18	5721.16
Observations	206257	206257	206257	206257

Note: This table replicates Table 2, but omits baseline control variables in columns (3) and (4). *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table 5: Reduced Form of the Difference between Properties Re-Assessed by Government Assessors and Firms on Property Characteristics with lists B and C as controls

	Reassessed (Apr. 2023)	Reassessed (Apr. 2024)	Tax Demand (Rs.)	Tax Collection (Rs.)
	(1)	(2)	(3)	(4)
Panel A: Reduced Form Estimates				
Assigned to government	0.474***	0.473***	3034.66***	1880.56***
	(0.043)	(0.043)	(377.77)	(275.46)
Assigned to private	0.338***	0.378***	1639.33***	781.64***
	(0.024)	(0.022)	(191.88)	(98.10)
Government = Private	0.006	0.050	0.00	0.00
Panel B: Instrumental Variables Estimates				
Reassessed by government			6417.27***	3977.01***
			(552.84)	(430.69)
Reassessed by private			4343.83***	2072.61***
			(371.00)	(222.13)
Government = Private			0.00	0.00
Control Mean	0.000	0.001	7047.89	5736.04
Observations	279057	279057	279057	279057

Note: This table replicates Table 2, but includes properties in lists B and C as control properties. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table 6: Reduced-Form Effects of Reassessment on Property Characteristics

		Area (sq. ft.)			Property Usage		
	Total	Residential	Non-Residential	Floor Count	Usage Upgraded	Usage Downgraded	Irregular-shaped Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Assigned to government	270.503***	242.197***	20.000***	0.273***	0.020***	0.004***	0.115***
	(30.98)	(27.09)	(4.79)	(0.03)	(0.00)	(0.00)	(0.01)
Assigned to private	179.746***	166.517***	4.156	0.187***	0.010***	0.006***	0.071***
	(17.68)	(15.67)	(2.84)	(0.02)	(0.00)	(0.00)	(0.01)
Government = Private	0.011	0.016	0.004	0.02	0.014	0.177	0.009
Control Mean	1338.762	1108.937	206.424	1.869	0.015	0.004	0.445
Observations	206257	203583	203583	203583	206257	206257	203572

Note: This table provides the reduced form estimates for Table 3. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 7: Difference between Properties Re-Assessed by Government Assessors and Firms on Property Characteristics with lists B and C as controls

		Area (sq. ft.)			Property Usage		
	Total	Residential	Non-Residential	Floor Count	Usage Upgraded	Usage Downgraded	Irregular-shaped Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reassessed by government	586.572***	514.900***	47.850***	0.587***	0.049***	0.010***	0.248***
	(44.06)	(39.61)	(8.96)	(0.04)	(0.01)	(0.00)	(0.02)
Reassessed by private	497.707***	451.043***	21.595***	0.516***	0.034***	0.019***	0.196***
	(30.02)	(26.94)	(6.76)	(0.03)	(0.01)	(0.00)	(0.01)
Government = Private	0.095	0.182	0.019	0.186	0.059	0.008	0.063
Control Mean	1332.143	1104.476	203.740	1.862	0.012	0.003	0.442
Observations	279057	275408	275408	275408	279057	279057	275397

Note: This table provides the reduced form estimates for Table 3, but includes properties in lists B and C as control properties. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 8: Difference between Properties Re-Assessed by Government Assessors and Firms on Actual and Imputed Tax demand

		Re-calculated demand for Reassessed Properties	
	Actual Tax Demand	Tax Demand (No Rebate for Reassessed)	Tax Demand (Rebate for Reassessed)
	(1)	(2)	(3)
Reassessed by government	6080.31***	6795.71***	4666.73***
	(540.06)	(619.87)	(493.89)
Reassessed by private	3859.76***	5259.72***	3283.80***
	(378.13)	(370.07)	(315.83)
Government = Private	0.00	0.03	0.02
Control Mean	7218.18	7229.96	7225.20
Observations	206257	206168	206168

Note: This table experimentally estimates the difference in actual and imputed tax demand for properties re-assessed by government tax assessors and private firms. Tax Demand (Rs.) is the property tax liability in the 2023-24 tax year. Tax Demand (No Rebate for Reassessed) represents the estimated tax demand for reassessed properties by applying the government's tax formula to the property's characteristics without any rebate. For properties that have not been reassessed, the actual tax demand is assigned. Tax Demand (Rebate for Reassessed) estimates the tax demand for reassessed properties by applying the government's tax formula with rebate. As with the previous measure, the actual tax demand is assigned to properties that have not been reassessed. All specifications include ward fixed effects and control for baseline annual demand (2022). Standard errors clustered by ward. We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 9: Impacts of Reassessment on Property Characteristics (Fixed vs Incentive Contracts)

	Area (sq. ft.)			Property Usage			
	Total	Residential	Non-Residential	Floor Count	Usage Upgraded	Usage Downgraded	Irregular-shaped Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reassessed by government	573.688***	507.209***	41.884***	0.572***	0.042***	0.009***	0.241***
	(44.19)	(39.59)	(9.06)	(0.04)	(0.01)	(0.00)	(0.02)
Reassessed by fixed-contract firms	432.192***	400.631***	2.145	0.480***	0.026***	0.016***	0.175***
	(35.45)	(32.55)	(8.33)	(0.04)	(0.01)	(0.00)	(0.02)
Reassessed by incentive-contract firms	554.974***	506.697***	26.636**	0.519***	0.025***	0.017***	0.211***
	(50.63)	(42.99)	(12.66)	(0.05)	(0.01)	(0.00)	(0.02)
Government = Fixed	0.012	0.037	0.001	0.126	0.083	0.055	0.03
Government = Incentive	0.781	0.993	0.327	0.41	0.09	0.084	0.341
Fixed = Incentive	0.047	0.049	0.106	0.539	0.976	0.845	0.226
Control Mean	1338.762	1108.937	206.424	1.869	0.015	0.004	0.445
Observations	206257	203583	203583	203583	206257	206257	203572

Note: This table replicates Table 3, decomposing the private category into properties reassessed by fixed-contract and incentive-contract firms. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 10: Accuracy of Reassessed Properties (OLS)

	Tax Gap	Inaccuracy
	(1)	(2)
Assigned to private	0.001	0.185***
	(0.056)	(0.039)
Government = Private	0.99	0.00
Government Mean	-0.04	0.24
Observations	1,481	1,481

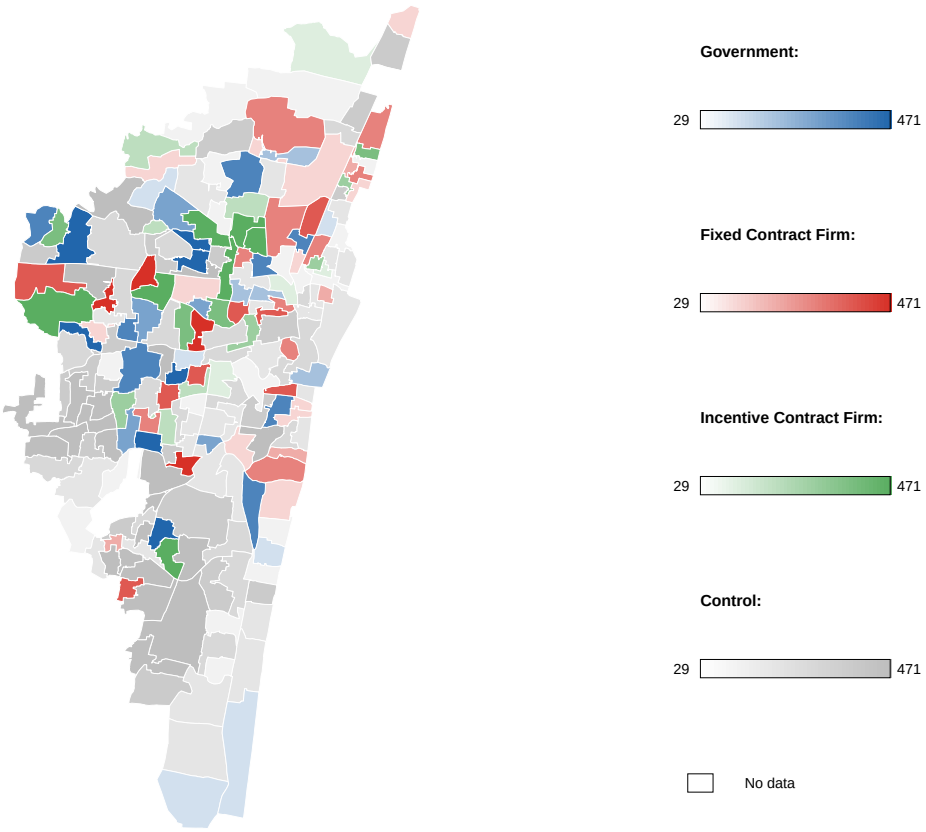
Note: This table experimentally estimates the difference in non-revenue outcomes for properties re-assessed by government tax assessors versus private firms. Tax Gap is the difference in the official and estimated demand, divided by the estimated demand, while inaccuracy is the absolute value of this measure. Official Tax Demand data come from the DCB dataset. We estimate OLS regressions where the omitted category is properties re-assessed by the government. All specifications include controls for the outcome variable at baseline, i.e. the tax gap or Inaccuracy based on the official tax demand in 2022-23. Standard errors are clustered at the ward level. We provide p-values for the test of equality between the government and private firms. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 11: Net Benefit Analysis (Collection-Based Estimates)

	Duration (in years)		
	1	5	10
Panel A: Expected Benefit Per Property			
Government	3964.66	19823.30	39646.59
Private Firms	2116.19	10580.94	21161.89
Panel B: Costs Per Property			
Government	285.02	285.02	285.02
Private Firms	230.09	230.09	230.09
Panel C: Net Benefit Per Property			
Government	3679.64	19538.27	39361.57
Private Firms	1886.10	10350.85	20931.80

Note: This table replicates Table 8, but for tax collection.

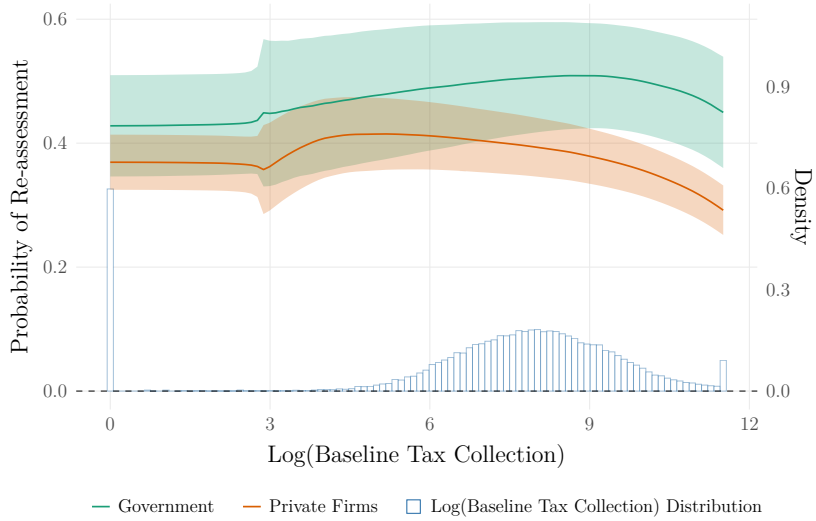
Appendix Figure 1: Properties Assigned for Re-Assessment by Ward and by Treatment

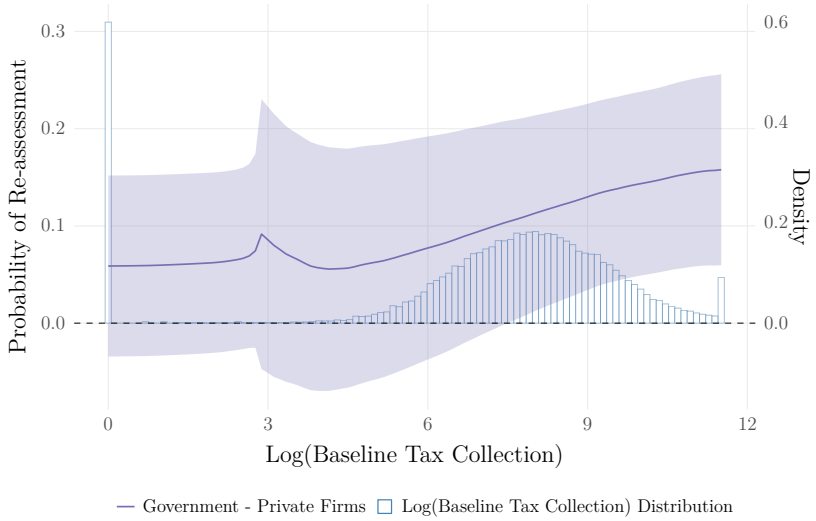


Note: This figure depicts the geographic distribution of properties selected for re-assessment. The sample is limited to the first list of properties the agents were instructed to re-assess. The color indicates the treatment arm and the color intensity the number of properties in each ward.

Appendix Figure 2: Heterogeneity in Reassessments and Tax Collection, by Log(Baseline Tax Collection)

Panel A: First-Stage Effects on the Probability of Reassessment





Panel B: IV Estimates of Heterogeneous Treatment Effects on Tax Collection

