Spillovers in State Capacity Building: Evidence from the Digitization of Land Records in Pakistan*

Shan Aman-Rana†, Clement Minaudier‡

January 2024

Abstract

Digitization reforms have been hailed as an effective way to strengthen state capacity. However, digitization can also disrupt the organization of bureaucracies. Using a unique administrative dataset on agricultural taxation and surveys of local bureaucrats from Punjab, Pakistan, we show that digitization reforms can have unintended consequences for state capacity. We exploit the staggered roll-out of the digitization of land records in Punjab to show that digitization had a negative effect on tax collection. The fall in taxes was not due to a decrease in the tax base. Instead, digitization affected the bureaucracy's capacity to collect taxes. The paper thus sheds light on the importance of understanding state capacity development from an organisational perspective.

JEL codes: D73, H11, H71, O12, O17

Keywords: Digitization, property rights, taxation, bureaucratic agency, bureaucracy, civil service, corruption, Punjab, Pakistan

*We are grateful to Gaurab Aryal, Oriana Bandiera, Sabrin Beg, Lauren Falcao Bergquist, Claudio Ferraz, Ali Hasanain, Namrata Kala, Andrew Kloosterman, David Lagakos, Caroline Le Pennec-Çaldichoury, Meera Mahadevan, John McLaren, Dilip Mookherjee, Anant Nyshadham, Paulina Oliva, Andrea Prat, Imran Rasul, Sheetal Seekhri, Kartini Shastry, Sandip Sukhtankar, Sebastian Tello-Trillo, Stephanos Vlachos, Leonard Wantchekon, Pinar Yildirim and seminar participants at workshops for comments and suggestions. Dylan Kaplan, Ryan Keller, Yanwen Wang, Yuanzhan Gao, and Moon Kim at UVA provided excellent research assistance. Financial support from the UVA is gratefully acknowledged. All mistakes are our own.

†Department of Economics, University of Virginia. Email: sa8ey@virginia.edu

‡Department of Economics, City University of London. Email: clement.minaudier@city.ac.uk

1 Introduction

Strong state capacity is essential for economic development. An effective approach to strengthening it is to introduce technology in bureaucracies. Technology can improve the productivity of bureaucrats and address a range of asymmetric information issues. It has helped to reduce agency problems between bureaucrats and their principals (Duflo et al., 2012; Lewis-Faupel et al., 2016; Dal Bó et al., 2021; Callen et al., 2020a; Debnath et al., 2023), to improve the reliability of information on tax payers (Okunogbe and Pouliquen, 2022; Ali et al., 2021; Brockmeyer and Sáenz Somarriba, 2022; Dzansi et al., 2022), or the identification of welfare recipients (Muralidharan et al., 2016a).

However, the introduction of technology can also disrupt the organization of bureaucracies. As public administration scholars have noted, digitization "reconfigures public sector organizations in fundamental, although uneven, ways" (Plesner et al., 2018). Technology can change the relationships between different bureaucratic agencies (Di Giulio and Vecchi, 2023) and increase specialization (Gundhus et al., 2022). This in turn can affect bureaucrats' sense of autonomy and their relationship with the public (Pors and Pallesen, 2021). While technology can lead to less corruption by reducing bureaucrat discretion (Muralidharan et al., 2016b), it can also result in the displacement of corruption onto other activities (Yang, 2008; Muralidharan et al., Forthcoming) or onto different individuals (Dzansi et al., 2022). Whether these changes are sufficiently disruptive to limit the benefits of digitization remains an open empirical question.

In this paper, we seek to understand whether the introduction of technology in bureaucracies can reduce their effectiveness and weaken state capacity. We study this question in the context of the digitization of land records in Punjab, Pakistan and show that the reform had a negative impact on the ability of the state to collect taxes. This negative relationship is not due to the effect of the digitization reform on the tax base but instead is due to its effect on the capacity of bureaucrats to collect taxes.

Digitizing land records has been a particularly popular way of leveraging technology to strengthen state capacity. From 2010 to 2019, fifty-two economies

computerized their land registries both in developing and developed countries, mobilising significant resources in the process.[1] These digitization reforms are particularly promising as they have the potential to spill over into other areas of state capacity. By strengthening property rights, they can encourages investment into productive activities (Field and Torero, 2006; Field, 2007; Beg, 2020; Chari et al., 2021), and in turn affect the capacity of the state to collect fiscal revenues (Besley and Persson, 2009, 2010, 2014a).

We exploit the staggered roll-out of the digitization of land records across districts of Punjab. Since this reform was carried out in three phases between 2011 and 2014, we can use a difference-in-difference design to identify the causal effect of the digitization reform on the amount of tax collected by the state. We digitized a novel administrative data set of rural agricultural taxes and combined it with data on the roll-out of the reform to test the effect of the reform on tax collection. We complement this data with satellite data on vegetation cover, survey data from local farmers, and unique data on the career trajectory of individual bureaucrats to separate the effect of the reform on the tax base from its effect on the bureaucrats' capacity to collect taxes.

We begin by documenting how the digitization reform affected the bureaucracy. First, bureaucrats who were in charge of both tax collection and the management of land records before the reform were no longer responsible for land records after the reform. Second, a large portion of bureaucrats (46%) reported that digitization negatively impacted tax collection. Of those, 57% reported that this was due to lower influence on taxpayers. Finally, bureaucrats lost a lucrative source of bribes: the proportion of bureaucrats who agreed that citizens bribed officials for land titles dropped from 49% to 34% after digitization. This effect aligns with the broader trend observed in technology-driven reforms (Muralidharan et al., 2016b; Debnath et al., 2023).

We then show our main result: that this reform had a significant impact on the state's ability to collect taxes. Digitization of land records led to an 84% decrease in tax collection in districts in the first two phases of the program relative to those

[1]Source: Doing Business Database, World Bank, https://subnational.doingbusiness.org/en/data/exploretopics/registering-property/good-practices.

in the third phase, which were not yet digitized. The modernization of state capacity not only failed to translate in higher tax revenues for the state, it actually reduced them. These results are robust to using different definitions of the timing of digitization, to different assumptions about differential pre-treatment trends (Rambachan and Roth, 2023), and to using alternative estimators from the recent literature on staggered treatment differences-in-differences designs (Callaway and Sant'Anna, 2021).

A decrease in tax revenue does not necessarily indicate a decrease in fiscal capacity. It is possible that the tax base decreased while the ability to collect taxes remained unchanged. We show that this was not the case here. The tax we study is levied on farmers based on the maximum of the tax due on cultivated area and profits generated. We find that the reform had a positive, but not statistically significant, effect on vegetation cover and irrigated cultivated area and a large and significant positive effect on farm profits.

Instead, we find a significant reduction in the capacity of bureaucrats to collect taxes. First, we find that bureaucrats in digitized districts reported lower cultivated areas in their tax assessments and therefore issued lower tax demands after digitization. This is despite the fact that we find no significant decrease in the tax base using satellite and household survey data. Second, bureaucrats in the digitized districts collected 6.6 percentage points lower taxes as a percentage of the tax demands after digitization. The magnitude of the effect is almost the same as the average tax collected in non-digitized districts. The proportion of bureaucrats collecting at least 50% of their target and the proportion collecting at least 75% both fell. We also find a positive but not statistically significant increase in the number of months in which zero taxes were collected. In sum, the digitization reform both led bureaucrats to issue lower tax demands and to collect a smaller portion of these demands.

The bureaucrats we surveyed reported that digitization led to a decrease in the influence of bureaucrats over citizens, which they identified as having a negative effect on tax collection. We propose two channels through which this loss of influence can translate into both lower tax assessments and lower tax collection. First, by

losing responsibility over land record management, bureaucrats lost leverage over tax payers. Before digitization, tax collectors could offer to process land permits or resolve land disputes in exchange for tax payments. Second, bureaucrats lost a lucrative source of bribe from land permits. If some of these bribes are instead exchanged against lower tax assessment after digitization, then this displacement could explain the reduction in tax demands we observe.

Our results therefore highlight a novel channel through which digitization reforms can affect state revenues: while modernizing state capacity can positively affect the tax base[2], it can also affect the interaction between bureaucrats and the local population and, as a result, reduce their performance. We find that the second effect can be sufficiently strong to generate an overall decline in tax collection.

These findings have important implications for the design of state capacity reforms. First, reforms to different dimensions of state capacity cannot be studied in isolation. In our context, reforming legal capacity, by digitizing land rights, had a negative effect on fiscal capacity because it removed existing complementarities between the two. In other contexts, reforms could harness these existing complementarities to increase the returns on investments in state capacity.

Second, investments in technology alone may not be sufficient to improve overall state capacity. In our context, technology did have a positive impact on the tax base and empowered the citizens, in line with the existing literature (Muralidharan et al., 2016b; Beg, 2020; Dzansi et al., 2022; Okunogbe and Pouliquen, 2022). However, technology also disrupted the relationship between the bureaucracy and its users, which reduced its ability to collect taxes. This suggests that the human and social dimensions of the bureaucracy are important and can be affected by investment in technology. Digitization reforms should therefore consider alternative means for bureaucrats to interact with citizens to maintain important social connections, as well as changes in human resources policies – such as incentive schemes – to address any loss of incentives that can arise with the loss of social

[2]Previous work has shown that improvements in legal capacity can result in higher taxable income for firms and households through an improvement in property rights. This can be due to better access to credit (Field and Torero, 2006) or better allocation of inputs, such as land, to more productive agents (Chari et al., 2021; Beg, 2020). Second, strong legal capacity allows tax policies to be enforced through a more effective legal system (Carrillo et al., 2017).

connections.[3]

Our results contribute to three strands of literature: the literature on digitization and development, the literature on state capacity and bureaucracies, and the literature on public finance in developing countries.

We contribute to the rapidly growing literature that examines the effects of digitization on economic development by showing that digitization can have unintended consequences on state capacity. A strand of that literature has focused on the direct effect of technology on the productivity or accountability of bureaucrats and front line providers (Duflo et al., 2012; Lewis-Faupel et al., 2016; Barnwal, 2018; Dal Bó et al., 2021; Callen et al., 2020a, 2023; Muralidharan et al., 2021, Forthcoming; Debnath et al., 2023). Other studies have found beneficial effects of introducing technology on tax collection. The technology studied either helped improve tax filing (Okunogbe and Pouliquen, 2022) or VAT records (Ali et al., 2021; Brockmeyer and Sáenz Somarriba, 2022), helped identifying tax payers and welfare recipients (Muralidharan et al., 2016a), or helped tax collectors geo-locate tax payers (Dzansi et al., 2022). Our work is most closely related to the studies that highlight the importance of organizational or management practices in the success of technological reforms (Milgrom and Roberts, 1990; Banerjee et al., 2008). Garicano and Heaton (2010) show that the introduction of IT systems in police stations only resulted in higher productivity when coupled with other organizational changes such as resource allocation and management practices. Our results are consistent with a similar 'complementarity' hypothesis: fiscal capacity can suffer from digitizing land rights if no further organizational changes are introduced.

We contribute to the literature on state capacity building (Besley and Persson, 2009, 2010; Bardhan, 2016) by presenting micro evidence on the negative spillover effects of an improvement in property rights on tax collection. Because the reform we study reduced the scope of the bureaucrats' work, our paper is most closely related to studies focusing on task design, particularly multitasking in public organizations (Holmstrom and Milgrom, 1991; Dewatripont et al., 1999a; Rasul and Rogger, 2018; Chen et al., 2018). We contribute to that literature by showing

[3]See Ashraf and Bandiera (2018) for a review of the importance of social connections at work.

that reducing the number of tasks can reduce the performance of bureaucrats. Contrary to the existing literature, we also show that changes in the scope of tasks do not just affect the relationship between bureaucrats and their supervisor (Dewatripont et al., 1999b), but also between bureaucrats and the population. Our paper therefore also contributes to understanding how the "embeddedness of the bureaucrat" - the social connections and interactions of the bureaucrats with the local population - affects the functions of the state.[4] Together, these results contribute to a growing literature on the organizational economics of the state that highlights organization design as a determinant of state capacity building in developing countries.[5]

Finally, we also contribute to the large literature on public finance in developing countries that seeks to identify the obstacles that developing countries face in collecting taxes (Besley and Persson, 2014b; Gadenne and Singhal, 2014). This literature has identified issues with tax enforcement, either due to lack of formal records (Pomeranz, 2015; Okunogbe et al., 2021; Jensen, 2022), the choice of tax rates or the design of the tax code (Best et al., 2015; Brockmeyer et al., 2021; Bergeron et al., 2021; Basri et al., 2021), corruption (Le et al., 2020), and citizen's misreporting or their lack of trust in the state (Carrillo et al., 2017; Naritomi, 2019; Besley and Dray, 2022). Our work is most closely related to papers that highlight the incentives and the ability of tax collectors as important determinants of fiscal capacity (Khan et al., 2016, 2019; Bergeron et al., 2021, 2022). We contribute to this literature by showing that introducing technology through piecemeal state capacity building can have unintended consequences for fiscal capacity because of its effect on tax collectors.

[4]Frequent interactions between bureaucrats and citizens can reduce transaction costs and increase trust (Pepinsky et al., 2017), increase the flow of information and influence of the bureaucrat (Evans, 1995; Bhavnani and Lee, 2018), and allow communities to hold bureaucrats to account through informal means (Tsai, 2007).

[5]Several studies show that the incentives of bureaucrats matter for public service delivery. These can be in the form of explicit incentive schemes (Khan et al., 2016, 2019), career concerns (Bertrand et al., 2020), monitoring (Callen et al., 2013), or autonomy in decision making (Rasul and Rogger, 2018; Duflo et al., 2018; Bandiera et al., 2021). Others show that the selection of bureaucrats is an important determinant of state effectiveness (Callen et al., 2020b), where selection can be affected either at the recruitment stage (Dal Bó et al., 2013; Bai and Jia, 2016; Deserranno, 2019; Ashraf et al., 2020; Colonnelli et al., 2020; Moreira and Pérez, 2022), or through the assignment of bureaucrats across jobs or promotions (Iyer and Mani, 2012; Jia et al., 2015; Bergeron et al., 2022; Best et al., 2023; Aman-Rana, 2023).

2 Background and data

2.1 Background

Agricultural Income Tax. We focus on the collection of a tax, the Agricultural Income Tax (AIT), which is levied on landowners in rural areas of the province of Punjab. This tax is one of the main sources of revenues to the government from agriculture. The amount of tax due is based on either the area of cultivated land or the profits of the farm. Specifically, farmers owe whichever of the cultivated area-based tax and the profit-based tax is largest (*Punjab Agricultural Income Tax Act*, 1997, section 3.4). When land is rented out by landowners to farmers, the landowner is liable for the tax (*Punjab Agricultural Income Tax Act*, 1997, sections 2.1 and 3.1).

The cultivated area-based tax is progressive and varies based on factors such as the size of the cultivated area, whether it is irrigated, and whether it functions as an orchard. The tax ranges from 300 to 600 PKR per acre, with irrigated areas and orchards subject to a higher tax rate (*Punjab Agricultural Income Tax Act*, 1997, section 3.1). The profit-based tax is also progressive and starts with a flat amount of 1,000 PKR for the first tranche (profits between 400,000 and 800,000 PKR), progressively increasing to 300,000 PKR plus 15% of the amount of profits exceeding 48,00,000 PKR (*Punjab Agricultural Income Tax Act*, 1997, section 3.3).[6] In practice, due to the difficulty of measuring income, the profit-based tax is restricted to large land owners who own more than 50 acres of land, which only applies to 12% of farms (The Agricultural Census, 2010).

The tax is collected by a team of local bureaucrats called *revenue officers*. Each team covers a jurisdiction comprising 2 to 3 villages called a *revenue circle*. The taxable amount in a fiscal year, which runs from the 1st of July to 30th of June the following year, is assessed by the same bureaucrats who collect the tax. At the start of a fiscal year, in the fall and spring, bureaucrats assess whether an agricultural land has been cultivated and also note its characteristics (irrigated or not, type of crops) during crop inspections (*Girdawari*). This is done in order to calculate the

[6] Income below 400,000 PKR is exempt.

cultivated area-based tax which varies based on these characteristics. Once tax is assessed, the bureaucrats issue tax demands around November and collect taxes over the remaining course of the fiscal year.[7] Income-based tax is calculated using self-reported profits.[8]

Bureaucrats do not receive any performance-based compensation for tax assessment or tax collection. Senior officials in the revenue department are required to oversee crop inspections conducted by junior officials (the unit of analysis in this paper) and are expected to conduct random checks on a minimum of 25% of the land under their jurisdiction. In cases where a junior official is found to be underperforming, they may face suspension from their position.[9] In the case of tax collection, managers monitor the progress of the team. Similar disciplinary action can be taken if the official systematically fails to collect enough taxes. The bureaucrats' promotions are based on tenure in the bureaucracy according to a pre-determined schedule. However, senior officials and politicians can informally influence the timing of promotions to allow high-performing bureaucrats to be fast-tracked. As in other bureaucracies (Iyer and Mani, 2012; Khan et al., 2019), transfers of bureaucrats across different revenue circles serve as an additional means of recognizing and incentivizing performance. These mechanisms introduce some potential career incentives for bureaucrats to achieve a high performance in their different tasks.

Digitization of land records. In 2005, the government of Punjab began a reform of the land record management system to digitize the records with the support of the World Bank. The main objective of the reform was to increase the reliability and the transparency of a system that was prone to errors and corruption.

[7] Bureaucrats are expected to carry out two crop inspections: during the fall (*Kharif*) and spring (*Rabi*) cropping season. In principle, taxes should be levied separately based on the two cropping seasons. However, in practice, tax demand for the entire year is created for each landowner based on the first cropping season. Bureaucrats do this to allow themselves enough time to collect taxes by the end of the fiscal year. A community representative (*Lambardar*) helps coordinate tax collection and works closely with the bureaucrats on tax collection.

[8] Landowners are expected to file their tax returns by the end of December for the respective fiscal year.

[9] A suspension does not involve a wage cut but the removal of the revenue official from the job and its associated perks.

The government planned to roll-out the digitization program in three phases. In each phase, 10-12 districts were selected to be digitized. This staggered design was driven by the financial difficulty of rolling out a reform of this size across the whole province at once. Figure 2 shows that there were no statistically significant differences in baseline characteristics between districts digitized in the first two phases of the roll-out and those digitized in the last phase.[10] This lends support to the use of the *planned* roll-out of digitization reform as an exogenous source of variation in legal capacity. The initial schedule was to roll out the digitized system in 2009 for phase 1, 2010 for phase 2, and 2011 for phase 3. The actual roll out was delayed and Figure 1 shows the proportion of villages that were digitized in each phase over time. We use the actual dates of the start of the roll out in each phase in our estimation, as discussed in the next section.

The reform had two effects. First, it secured property rights and decreased disputes between land owners.[11] Second, it changed the type of tasks carried out by the local bureaucrats that we study. Prior to the digitization reform, bureaucrats were responsible for recording sales or exchanges of land and properties and for issuing land titles, as well as for assessing and collecting taxes, as described above.

The main effect of the reform on the bureaucrat was to relieve them of one of their duties, namely, managing land records. Since this activity required meeting in person with property owners, the reform affected both the set of tasks they carried out and their interaction with the local population. Figure B.5 shows that 69% of bureaucrats reported that the reform changed their tasks, of which 75% said that some tasks were removed but 59% indicated that some tasks were also added. The tasks added were mostly about record correction and additional paper work, which was part of the transition from manual to digitized land records (see Figure B.6 in appendix). Therefore, the new tasks were mostly relevant in the short-term. Overall, the number of hours worked reported by the bureaucrats

[10] Figure B.18 shows a similar balance test but splits digitized districts into phase 1 and 2 districts. We can see that phase 1 districts were statistically significantly smaller in population size and have less factories. This suggests that the government planned to start out the program in smaller less developed districts, but did not otherwise select districts in systematic ways.

[11] Beg (2020) shows that the reform led to an improvement in the security of property rights and resulted in some re-allocation of land to more productive farmers and, in turn, increased profits of farmers.

did not increase. Figure B.8 shows that 75% of bureaucrats reported no change in hours worked, 5% reported a decrease, and 20% reported an increase. This suggests that the majority of bureaucrats either used the time freed up from land records to work on other tasks or simply worked less after the reform. Figure B.9 confirms that the net reported decrease in hours is not significantly different from the net reported increase.

The bureaucrats also reported two interesting changes following the reform. First, they indicated that digitization negatively affected their ability to collect taxes (Figure B.10). The main reason cited for this is a loss of influence on the local population as shown in Figure B.11. Second, the bureaucrats lost an important source of bribe income. Obtaining bribes or 'tips' in exchange for a speedy processing of land records was common before the reform. Because the bureaucrats no longer had control over the land record process, they lost this lucrative source of bribe. Figure B.17 shows that 49% of bureaucrats agreed that citizens want to tip to get land titles in revenue circles that there not digitized compared to 34% in revenue circles that were digitized.

2.2 Data sources and key variables

Digitization roll-out. The data on digitization includes both the planned and actual roll-out of the digitization reform. We obtained data on the planned roll-out of the program from the Land Record Management Information System (LRMIS) project office in Lahore. This data indicates which districts were intended to be digitized in phase 1, 2 or 3 of the program. Second, we obtained data on the actual progress of the digitization program from the Punjab Land Records Authority (PLRA) in February 2018. This data describes the digitization status for each revenue circle: whether and on which date the land records for each of the villages in the revenue circle have been digitized.

We define a phase as the set of districts that were intended to be digitized at the same time as each other in the roll-out plan. Based on that, we define digitization of land records as a dummy variable that takes value 1 in a year and district if the district belongs to a phase that was digitized in that year, and remains zero

otherwise. We define a phase as being digitized in a given year if at least 5% of villages that are located in districts belonging to that phase have been digitized by that year.[12] In Appendix Table A.4, we show that our results are robust to using alternative thresholds of proportion of villages to define a phase as digitized. Based on this definition, the treatment years are defined as follows: Phase 1 is first treated in fiscal year 2012, phase 2 in fiscal year 2013, and phase 3 in fiscal year 2014.

Agricultural tax collection. We carried out a large-scale data digitization exercise of the agricultural tax collection records of the Board of Revenue (BOR), the agency in charge of tax collection (see Appendix C for the record room and the proforma on which this information is collected). In addition to the total amount of taxes collected,[13] the data also contains information on the tax demands that bureaucrats issue to tax payers.[14] This is based on the tax assessment carried out by the bureaucrat and serves as the target amount of taxes that bureaucrats aim to collect.[15]

The records have monthly information at the revenue circle level from 2006-2013 (28,572 revenue circle-months).[16] Since the digitization reform was planned at the district level, we aggregated the taxation data at that level. We also aggregate the data at the year level since tax assessments are issued annually and the monthly tax collection data is therefore noisier.The resulting data is a balanced panel of 219

[12]We combine the planned roll-out (for the definition of phases) and the actual roll-out (for the definition of dates) for two reasons. First, because the actual roll-out was significantly delayed relative to the plan. Second, the delays in the roll-out can depend on characteristics of the district and therefore be endogenous. We therefore use the planned set of districts to reduce these concerns.

[13]Tax collection is available for both the ongoing fiscal year, as well as arrears from past years. We only use the current year's tax collection, as that record is most up-to-date.

[14]This information is at the revenue circle level.

[15]The data also contains information on remissions, suspensions and notifications of irrecoverable tax, but those are rare events in the data.

[16]This data is an unbalanced panel of revenue-circles and months since some of the tax files were destroyed in flooding of the archives. To ensure that the data is representative at a district level, we created inverse probability-weighted sums. For each time period, the weights are calculated based on the number of revenue circles for which we have data, relative to the total set of revenue circles in a Tehsil and district. Due to the presence of outliers, we dropped a revenue circle-fiscal year if the annual tax targets were at least two standard deviations or higher for that revenue circle. This resulted in a drop of 65 revenue circle-fiscal years out of 3,492 (1.9%) and one observation at the district-fiscal year level out of 220 (0.5%).

districts-fiscal years. This data forms the basis of the main analysis.

Actual tax base. To evaluate the effect of the reform on the tax base (cultivated area or farm income), we rely on three sources of data. First, to observe cultivated area we compiled satellite data measuring the vegetation cover of the area we study. Specifically, we used the Normalized Difference Vegetation Index (NDVI) to measure the yearly vegetation cover between 2006-2013 (see Appendix F for details of our methodology to compile the data for our analysis). The NDVI data is a commonly-used proxy for crop yield in developing countries (Rasmussen, 1992; Vrieling et al., 2011; Beg, 2020) and has been shown to be sufficiently stable to permit meaningful comparisons of seasonal and year-to-year changes in vegetation growth and activity (Huete et al., 2002).

We complement the satellite data with additional survey data from the Pakistan Living Standards Measurement Survey (PSLM) which collects data on agricultural land owned (in acres) as well as agricultural land that is irrigated from a repeated cross-section of households across districts of Punjab.[17] The data is representative at the district level. The waves used in this study are for the years 2006, 2008, 2010 and 2012.

We use Household Income and Expenditure Surveys (HIES) data provided by Beg (2020) to investigate the effects of digitization on agricultural profits, the other possible element of the tax base. This data collects demographics, employment, expenditure, and saving information from a repeated cross-section of households across rural areas in districts of Punjab. We use the 2005, 2007, 2011, and 2013 waves of the survey and, as in Beg (2020), restrict attention to the farm level data provided by cultivating households.

Bureaucrat characteristics and performance. For the last part of our analysis, we complement the administrative tax collection data with a retrospective survey of the bureaucrats involved in tax collection around the time of the reform.[18]

[17] Since we were interested in agricultural outcomes, we restricted the analysis to rural households in PSLM.

[18] The survey was first carried out in-person in September 2020. To obtain more details on their career paths while ensuring a high response rate, we carried out a separate telephonic survey

These surveys give us the career history of these bureaucrats across different revenue circles, which allows us to test whether digitization changed the tax collection performance of these bureaucrats. Appendix D describes the details of the sampling methodology. In addition, we used the survey to measure a number of individual characteristics. These include measures of the ability of bureaucrats (results on entrance exams, knowledge of the law, memory tests), of their relationship with citizens and superiors, and personal traits (pro-sociality, honesty). Finally, the survey also asked the respondents about their perceptions of the reform and how it affected their interactions with superiors and with the population.

There were no unique revenue circle identifiers in any of our data. We therefore, carried out an extensive string matching exercise to merge the tax, digitization and bureaucrats' careers datasets. Merging tax and digitization data with the bureaucrats' career data helped us locate whether a bureaucrat worked in a revenue circle that was digitized including their tax collection performance over time. Appendix E describes how we string-matched the tax and bureaucrats' survey data to create a panel of bureaucrats-revenue circles-fiscal year and how we matched the tax and digitization data. In summary, we string-matched the two data sets using revenue circle, Tehsil and district names and cleaned them manually. This data was further verified through other government documents from the relevant local offices.

Tax base reported by bureaucrats. To measure the bureaucrat's assessment of the tax base, carried out during crop inspections, we use data compiled by the Directorate of Agriculture (Economics and Marketing) of Punjab who used the cultivated area reported by these bureaucrats to report average cultivation across districts for 2007-2013.[19] We complement this data with tax demands issued by the bureaucrats to taxpayers from the tax data described above. This data is also

focusing on their career history in November 2020. For a random subset of the data, we confirm the accuracy of the responses by comparing them to official records of bureaucratic transfers.

[19]This data is available at http://www.amis.pk/Agristatistics/DistrictWise/DistrictWiseData.aspx. There is no data available for the year 2006. Obtaining the bureaucrat-level data directly is difficult since it is not digitized and is only available from local offices across Punjab.

based on the assessment of cultivated area made by the bureaucrats at the revenue circle-fiscal year level. We aggregated it at the district-fiscal year level to obtain the bureaucrats' tax assessments in each district.

3 Did the digitization reform affect tax collection?

We now turn to testing our main question: how did the digitization reform affect tax collection?

3.1 Identification strategy

There are several difficulties in measuring the effect of digitization reforms on fiscal capacity. Policy makers could introduce digitization reforms at times where bureaucracies are underperforming due to structural issues. Alternatively, some districts might be targeted for the implementation of the reform because bureaucrats in these districts face difficulties collecting taxes and need technological support in other tasks.

Our differences-in-differences strategy based on planned roll-out of the reform allows us to address these concerns. We exploit the *planned* staggered roll-out of the digitization of land records in our difference-in-differences strategy to estimate the effect of the digitization of land records on tax collection. Recall from Section 2.2 that we define a set of districts in a given phase as digitized (treated) in a given year if at least 5% of villages in these districts have been digitized by the end of that year. Once a district is treated, it therefore remains treated for the remaining of the sample period. Figure 1 shows the cumulative proportion of villages digitized by month in each set of districts. Appendix Table A.1 shows the correlation of digitization and phase 1 and 2 dummy variables with the number of villages with digitized land records. We see that being in a planned digitized district results in 173 more villages with digitized land records than the phase 3 districts.

Our strategy compares the difference in tax collection before and after digitization between districts where digitization was planned to be introduced and those

where land records remained manual. The identification assumption motivating this estimation strategy is that early digitized districts and later digitized districts have parallel trends: districts in phases 1 and 2 of the reform would have experienced, on average, the same changes in tax collection over time as those in phase 3, were it not for the digitization of their land records. Throughout the paper we present intent-to-treat analysis, which estimates the average return to as-is implementation of the digitization reform following the "intent" to implement the new digitized land record system. These estimates yield the relevant policy parameter of interest, because they reflect the impacts that followed a decision by the government to invest in the digitized land records and are net of all the logistical and political economy challenges inherent in implementing such a project in practice.

Following the existing literature, we assess the evidence in support of this assumption using an event-study plot prior to conducting the main analysis. Specifically, for district d in a digitization phase g and fiscal year t between 2006-2013, we estimate the following regression:

$$y_{dgt} = \alpha_d + \alpha_t + \sum_{k=-6}^{k=2} \rho_k D_{k(dgt)} + \epsilon_{dgt} \tag{1}$$

where y_{dgt} is the inverse hyperbolic sine (arcsinh) of the total tax collected in district d, belonging to digitization phase g, during fiscal year t, that is, $y_{dgt} = \ln(tax_{dgt} + \sqrt{1 + (tax_{dgt})^2})$. α_d are district fixed effects, α_t are fiscal year fixed effects, and $D_{k(dgt)}$ is a set of indicator variables that takes value one if districts d in a digitization phase g in fiscal year t were k years away from being digitized. The error terms are clustered at the district level as that is the level of the treatment (Abadie et al., 2023).

The inverse hyperbolic sine (arcsinh) transformation is used when the dependent variables follow right-skewed distributions, and have large probability mass at zero (Chen and Roth, 2023). We use it because of the former rather than the latter reason. A large probability mass at zero is less of an issue in our context: the proportion of observations with zeros are just 5% - 11 out of 212 district-year

observations.[20]

The coefficients ρ_k estimate the effect of being treated k years before and after digitization.[21] The omitted time period is the one right before the digitization year. If ρ is statistically insignificant for all the years before treatment, then this lends support to the validity of the parallel trends assumption. Figure 4 plots ρ_k for each period k and their corresponding 95% confidence intervals. The figure shows that before the digitization of land records there are no statistically significant differences in tax collection between districts that were digitized and those that were not. This lends support to the parallel trends assumption. At the end of this section we discuss other checks on the validity of our identifying assumptions.

3.2 Estimation and results

To obtain the causal effect of digitization on tax collection, we use a staggered difference-in-differences estimation by running the following regression at the district-year level:

$$y_{dt} = \eta_d + \eta_t + \beta_T \text{Digitization}_{dt} + \varepsilon_{dt} \tag{2}$$

Our outcome variable, y_{dt} is the inverse hyperbolic sine (IHS) of the total tax collected in district d during fiscal year t, in thousand of Rupees. That is, $y_{dt} = \ln\left(tax_{dt} + \sqrt{1 + (tax_{dt})^2}\right)$. Our treatment variable, Digitization_{dt}, is a dummy that takes the value of one if a district d belongs to a phase that has been digitized by year t.[22] Finally, η_d and η_t are district and fiscal year fixed effects, respectively.

[20] Treatment effect estimates using arcsinh transformations are sensitive to changes in the units of the outcome (De Brauw and Herskowitz, 2021; Aihounton and Henningsen, 2021). The problems identified with log-like transformations are more prominent when there is a large probability mass at zero and the extensive margin effect is prominent (Mullahy and Norton, 2022; Chen and Roth, 2023). In Appendix Table A.3, we present separate estimates for the extensive and intensive margin effects, as suggested in Chen and Roth (2023). Our results are mainly driven by the intensive margin effect. The extensive margin effect is very small and statistically insignificant. We also show that our main results are comparable when running a median regression and when the outcome variable is expressed in levels in thousands of rupees.

[21] Phase 1 districts were digitized in 2011 while phase 2 districts were digitized in 2012. Therefore, we are able to estimate the effects of digitization for two years after the reform (2012-2013) for treated districts in phase 1, however, we are only be able to estimate effects of the reform for one year after digitization (2013) for treated districts in phase 2. The control districts are phase 3 districts that were digitized starting from 2013.

[22] Recall from Section 2.2 that we define a phase as digitized in a given year if at least 5% of

Standard errors are clustered at the district level as that is the level of the treatment (Abadie et al., 2023).

Table 1 shows our main result: digitization led to an 84% decrease in tax collection relative to districts that were not digitized.[23] Instead of increasing fiscal revenues, this state capacity reform therefore caused a large and statistically significant decrease in tax collection (*Wild* clustered bootstrapped p-values<0.05).

The decrease in tax collection due to the digitization reform is substantial. Even when measured in Rupee terms, the decrease of 8,053 thousand Rupees following the reform in treated districts relative to untreated ones represents a 49% decrease relative to the mean of tax collection in untreated districts. The magnitudes of these effects is in line with other findings in the literature. For instance, Okunogbe and Pouliquen (2022) find that corporate tax payments decreased by 40% following the introduction of electronic filings among firms classified as presenting a lower risk of tax avoidance.[24] Knebelmann et al. (2023) find that the tax base for property tax is 83% lower when the assessment is carried out at the discretion of bureaucrats relative to when it follows an algorithm. The magnitude of the decrease should also be assessed in the context of the tax we study. Given that there can be a lot of variations in agricultural tax collection, the decrease of 8,053,000 Rupees represents 40% of the historical standard deviation in tax collection prior to the reform.[25]

Finally, the decrease in tax collected can have important economic consequences. While the tax that we study is not the largest source of revenue for the government, the loss of 8,053,000 Rupees due to the reform still represents a significant shortfall. Extrapolated across all 36 districts, the amount of lost taxes could have funded cash transfers for an additional 15,428 families on the government's main social welfare program (Benazir Income Support Programme

villages located in districts belonging to that phase have been digitized by that year.

[23]The percentage change is approximated as usual by $\exp(\hat{\beta}) - 1 = \exp(-1.826) - 1 = -0.839$.

[24]The decrease they measure, however, is not statistically different from zero at conventional levels. The intuition they suggest for this result is that these firms used to pay more in the physical presence of the tax collector.

[25]The standard deviation of our outcome variable, tax collection, across all districts between FY2006 and FY2011 is 20,124 thousand Rupees.

(BISP)).[26]

This decrease could be due to two reasons. First, the tax base (the cultivated area and the farmers' profits) could have decreased as a result of digitization. Second, digitization could have decreased the bureaucrats' capacity to collect taxes. Before turning to these mechanisms in the next section, we discuss a number of robustness tests we carried out.

Robustness using the "more credible approach". We follow Rambachan and Roth (2023) and use the "more credible approach" to parallel trends: instead of requiring that parallel trends hold exactly, we investigate the robustness of our results to alternative assumptions about differences in trends in tax collection between digitized and non-digitized districts.[27]

Rambachan and Roth (2023) suggest that restrictions on the possible violations of parallel trends must be specified by the researcher depending on the economic context. To determine these restrictions, we use pre-digitization data to create a linear trend that is extrapolated to post-digitization. Appendix Figure B.19 plots the linear trend (in black), along-with the event study coefficients plot based on Equation 1 (in blue). Since in the pre-treatment period the coefficients are close to the linear trend, we impose that the differential trends evolve smoothly over time by bounding the extent to which its slope may change across consecutive periods. We therefore use the following formula suggested by Rambachan and Roth (2023):

$$\Delta^{SD} := \{\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \forall t\}$$

where δ_t refers to the difference in trends between the districts with digitized and manual land records at time t. M governs the amount by which the slope of δ can change between consecutive periods. If M=0 the difference in trends between

[26]The annual transfer for families eligible to the BISP was 18,792 Rupees in 2015 (Cheema et al., 2016). The loss of 289,922 thousand Rupees (8,053 thousand Rupees multiplied by 36 districts) would therefore cover $\frac{289,922}{18,792} = 15,428$ famillies.

[27]See Roth et al. (2023) for a review of the recent developments in differences-in-differences estimation and Manski and Pepper (2018) for a move from invariance assumptions to assumptions of bounded variation that can be used to partially identify treatment effects yielding bounds rather than point estimates.

digitized and manual districts would be exactly linear, while $M > 0$ relaxes the assumption of exact linearity.

The results are summarized in Figure 5. This figure shows that the main results remain robust up to a value of M=0.14. As discussed by Rambachan and Roth (2023), nothing in the data itself can place an upper bound on the parameter M. We therefore follow the methodology in Dustmann et al. (2021) and Aman-Rana et al. (2023) and use the data in pre-treatment periods to analyse and benchmark the value of M. We calculate the median of the absolute deviations of the coefficient ρ (see Equation 1) from the linear trend using only pre-digitization data. This method leads to a median value of M=0.17, with the 45^{th} percentile as M= 0.07. Our results therefore remain robust to a deviation from a linear trend as large as the 45^{th} percentile but not the median of the pre-treatment deviations from the linear trends.

Two-way fixed effects bias due to the staggered roll-out of digitization. Two-way fixed effects (TWFE) regressions like the one we use (Equation 2) have been shown to deliver consistent estimates only under relatively strong assumptions about homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2023).[28] We address concerns about the reliability of TWFE estimator by showing two sets of results. First, Table A.5 replicates Equation 2 but only uses the never-treated phase 3 districts (where land records were not digitized in the sample period) as the control group. This avoids the 2 × 2 difference-in-differences comparisons between newly treated and already treated units and gives us consistent estimates even in the presence of heterogeneous treatment effects across time and/or treated units. Column (1) presents results using the sample of phase 1 and 3 districts, while Column (2) presents the results restricting the sample to phase 2 and 3. Table A.5 in appendix shows that the effects are of similar

[28]Since the estimate of the treatment effect obtained from a TWFE model is a weighted average of all possible 2 × 2 difference-in-differences comparisons between groups of units treated at different points in time, some of these compare units treated at two different times, using the later-treated group as a control before its treatment begins and then the earlier-treated group as a control after its treatment begins (Goodman-Bacon, 2021). If treatment effects are heterogeneous across groups or time, the TWFE estimator does not deliver consistent estimates for the ATT.

magnitude to Table 1 and significant at conventional levels when comparing the treatment phases separately.

Second, we use the estimator proposed by Callaway and Sant'Anna (2021). This estimator measures weighted averages of group-specific treatment effects to allow for heterogeneous treatment effects across treatment phases. Table A.6a and A.6b in appendix shows that the effect is robust to using their estimator.

Other checks. An additional worry is that the digitization reform drove bureaucrats to move out of digitized districts to non-digitized ones or vice versa. These transfers could explain the decrease in tax collection if bureaucrats that were systematically collecting less taxes were relocating to digitized districts or bureaucrats collecting more taxes to non-digitized ones. While such transfers are not allowed by law, we also verify that this was indeed the case by using our data on careers of the bureaucrats. This data confirms that only 2 out of the 118 senior bureaucrats have ever been posted outside the districts where they started their careers (see Figure B.4). The situation was similar for their subordinates: only 2 out of 440 subordinates were ever transferred out of their original districts. This helps rule out that this type of spillovers could threaten our identification strategy.

Another concern is that either the bureaucrats or the citizens could have anticipated the digitization reform and changed their behaviour as a result. These anticipation effects could bias our results. Anticipation effects matter if they systematically impact tax collection more in phases 1 and 2 districts relative to phase 3 districts. Appendix Table A.7 uses two alternative definitions of the timing of digitization as a placebo test. In Column (1) the digitization reform is defined as starting in 2006 for phase 1 districts and 2007 for phase 2 districts, while in Column (2) these timings are defined as 2009 and 2010, respectively. In both Columns we can see that there are no statistically significant effects, (*Wild* clustered bootstrapped p-value is 0.79 and 0.59). The magnitudes of the effects are also much smaller than the main effects in Table 1.

Finally, in Appendix Table A.8 we replicate Table 1, but report the p-values from permutation test similar to the randomization-based inference test (Athey

and Imbens, 2017; Young, 2019). This tests whether the effects of digitization are simply due to chance owing to the selection of districts that were assigned to be digitized in phase 1 and 2 relative to phase 3. We re-assign digitization over districts and fiscal years 1000 times and compute the estimates under the null hypothesis that the treatment has no effect. We then look up the point estimates coming from our real data in the distribution of the 1000 treatment assignment simulations. The *p*-value we report at the bottom of the table comes from the share of estimates from the 1000 reassignments based on simulations that are higher in absolute value than our point estimates in Table 1. The *p*-value is zero, increasing confidence in our main analysis.

4 Why did tax collection decline?

We now turn to investigating two possible channels behind the decrease in tax collection: that the tax base decreased and that the capacity of bureaucrats to collect tax decreased. Our analysis suggests that the effect is more likely to come from the bureaucrat's performance.

4.1 Changes in the tax base

As described in Section 2, the tax collected by bureaucrats is based on two measures: the area cultivated by farmers and the profits of the farmers. In particular, the amount of tax due is calculated based on the maximum of the tax due on cultivated area and the tax due on profit. The digitization reform could have directly impacted both of these dimensions of the tax base. As digitization makes property rights more secure and transparent, it can lead farmers to start cultivating plots of lands whose ownership was previously disputed or encourage land owners to rent out uncultivated land to more productive farmers. These effects should translate into larger cultivated-area. Finally, more secure property rights reduce risks of expropriation and can encourage farmers to invest in better technology. This effect would translate into higher profit. We show in this section that digitization increased farmers' profits and had no significant effect on cultivated area.

To show this, we use four different outcome variables: the log of farm-level profits, the satellite vegetation cover index, a measure of whether land owned was irrigated or not, and the log of agricultural land owned.[29] For each measure, we run the following regression:

$$y_{dt} = \mu_d + \mu_t + \beta_B \text{Digitization}_{dt} + e_{dt} \tag{3}$$

where y_{dt} is either a measure of cultivated area, or farm profits in Pakistani Rupees, $Digitization_{dt}$ is a dummy that takes value 1 if district d was digitized in year t, μ_d and μ_t are district and year fixed-effects, respectively. Standard errors are clustered at the district level.

Table 2 shows the results. Column (1) estimates the effect of digitization on profits, while column (2)-(4) estimates the effect on cultivated land by using the satellite vegetation cover index and the survey data on land owned and irrigation as proxies. The coefficient in column (1) shows that digitization had a statistically significant positive effect on profits. Column (2) to (4) show instead that digitization had a positive effect on cultivated land but that effect is not statistically different from zero. These results are consistent with the findings of Beg (2020) who exploits the same reform to understand the effect of stronger property rights on land and labor makers. Beg (2020) shows that digitization increased the productivity of farmers due to two effects: a re-allocation of land to more productive farmers and an improvement in the use of inputs and investments. This productivity mechanism could explain why we detect an effect of digitization on profits but not cultivated area, at the district level.

We therefore conclude that digitization did not lead to a decrease in the tax base: cultivated area remained unchanged while profits went up significantly as a result of it. This suggests that the decrease in tax collected we observe cannot be driven by a decrease in the tax base.

[29] As described in Section 2.2 the farm-level profit data comes from a survey of farmers which is only available for 5 waves (2007–2008, 2011–2012, 2013–2014, and 2015–2016) and not yearly. We therefore need to modify our definition of the treatment year for the estimation based on this outcome: we pool phase 1 and phase 2 districts and define them as digitized from the 2013-2014 wave onward while phase 3 districts remain the control group. For the rest of the outcomes the definition of the treatment year remains the same as in Section 3.1.

4.2 Effect on performance of bureaucrats

If tax collection decreased, as shown in Section 3, but the tax base did not, as shown in Section 4.1, another possibility is that the digitization reform reduced the bureaucrats' capacity to collect taxes. That is, the reform reduced fiscal capacity. In this section, we provide evidence that this was the case by exploring how the reform affected the bureaucrats' performance in two of their tasks.

4.2.1 Changes in tax assessment

One reason the reform could have reduced the amount of tax collected is if it induced the bureaucrats to change how they assess the tax base. Recall that bureaucrats determine the size of the cultivated area and its characteristics (irrigation, crops) during their crop inspection in fall. This assessment is then used to determine the tax demands that are issued to farmers. The bureaucrats' discretion over this assessment creates an opportunity for collusion: farmers could offer to pay a bribe in exchange for a lower assessment and therefore a lower tax bill. Bureaucrats would accept this arrangement if the bribe is sufficiently valuable. The loss of bribe income from land records due to the digitization reform could encourage bureaucrats to shift bribe taking to the assessment process. As a result, we could expect the digitization reform to result in lower tax demands.

To address this question, we use two sets of data: the data from the Directorate of Agriculture on cultivated areas which is based on reports from the bureaucrats, and the district level tax demands that aggregate the tax demands that the bureaucrats issue to tax payers based on their assessment of cultivated areas. We use these two outcome variables to run the following regression:

$$y_{dt} = \nu_d + \nu_t + \beta_A \text{Digitization}_{dt} + v_{dt} \tag{4}$$

where y_{dt} is one of the two outcome variable described above in district d and year t, $Digitization_{dt}$ is a dummy that takes value 1 if district d was digitized in year t, ν_d and ν_t are district and year fixed effects. Standard errors are clustered at the district level.

Table 3 shows the results: after the digitization reform, districts with digitized land records had a 10% lower reported cultivated area (Column 1), as well as 45% lower tax demands (Column 2), relative to districts with manual land records.[30] This is despite the fact that neither the vegetation cover index nor the agricultural land irrigated or owned decreased significantly, as shown in Column (2)-(4) of Table 2. These results therefore indicate that the digitization reform led bureaucrats to under-report the tax base, which in turn reduced the tax demands they issued to farmers.

These results are consistent with the digitization reform increasing collusion between bureaucrats and tax payers. Figure B.17 shows that bribes were common before the reform, as 49% of respondents in our survey either agreed or completely agreed that citizens wanted to 'tip' bureaucrats to obtain land titles. Only 34% agreed or completely agreed with that statement when considering areas were the reform had taken place.[31] The digitization reform could have therefore led bureaucrats to try and make up for this lost income by increasing collusion on tax assessment. This is in line with other instances of 'bribe' displacement found in the literature (Yang, 2008; Sequeira, 2011, 2016; Dávid-Barrett and Fazekas, 2020).[32]

The decrease in the *reported* tax base and the corresponding lower tax demand can explain part of the decrease in tax income shown in Table 1. However, we show in the next subsection that tax collection decreased even relative to this reduced tax demand. Collusion between bureaucrats and taxpayers in setting tax targets can therefore not explain all of the decrease in tax collection.

[30] These effects are approximated using the transformations $\exp(-0.106) - 1 = -0.10$ and $\exp(-0.600) - 1 = -0.45$ respectively, as we did for the main result in Table 1

[31] Given that admitting to this behavior reflects badly on the bureaucracy, it is likely that these responses underestimate the true magnitude of bribery. We expect the under-reporting to be similar before or after the reform. Appendix Figure B.16 provides support that this might be the case. It shows the proportion of respondents that refused to answer the question on citizens' tipping behavior before and after the reform, and these are roughly similar.

[32] Beg (2020) shows that digitization led to more land being rented out to farmers. One concern might be that the reform led tax collectors to interact less directly with land owners (the tax payer) as they would instead meet with tenants on their farm visits. However, this lack of direct interaction should result in less opportunities to collude on assessment, rather than more, and therefore should translate into higher rather than lower assessments post digitization.

4.2.2 Change in performance relative to tax demand

While a decrease in the reported tax base and the associated tax demand can explain the decrease in taxes, it is also possible that the digitization reform led bureaucrats to *collect* less taxes. That is, it is possible that the *performance* of bureaucrats decreased. As shown in Figure B.11, a large portion of bureaucrats reported that the digitization reform made tax collection more difficult because it reduced their influence over the population. One possibility is that bureaucrats were able to use the prospect of facilitating land titles issuance or the resolution of land disputes to encourage farmers to pay their taxes. The loss of this source of leverage disappeared following the digitization reform should make it more difficult for bureaucrats to collect a given proportion of tax demands. We confirm that this was the case by looking at three different measures of performance.

First, we look at the effect of the digitization reform on the tax collected by bureaucrats as a percentage of the tax demands they need to collect. The tax demands issued by bureaucrats correspond to the tax due by farmers and therefore the target that bureaucrats are expected to collect by their superiors.[33] We confirm this measure with two alternative variables: whether bureaucrats achieved at least 50% of their targets, and whether they achieved at least 75% of their target. These binary variables allow us to confirm that the results from the continuous measure are not driven by outliers with particularly low performance. Finally, we also analyze whether the reform affected the ability to collect taxes at the bottom-end of the performance distribution by looking at the share of months per year in which the bureaucrats collected no taxes at all.

We use the sample restricted to revenue circles in which we have survey data available for individual bureaucrats and carry out the analysis at the individual bureaucrat level instead of the district-level analysis in the previous section. This allows us to include bureaucrat fixed effects and thus estimate the within-

[33] Since the bureaucrat's objective is to collect the tax due, it is not possible for bureaucrats to over perform based on our measure of performance: once the tax due is collected, bureaucrats cannot collect more tax from that farmer.

bureaucrat variation in tax collection. Specifically, we run the following regression:

$$y_{idt} = \gamma_i + \gamma_t + \beta_P \text{Digitization}_{dt} + u_{idt} \tag{5}$$

where y_{idt} is one of the measures described above for bureaucrat i in district d and year t, $Digitization_{dt}$ is a dummy that takes value 1 if district d was digitized in year t, γ_i are bureaucrat fixed effects and γ_t are fiscal year fixed effects. Standard errors are clustered at the district level.

Table 4 shows the results of this regression. Column (1) shows that the digitization reform led to a substantial decrease in the bureaucrat's performance. Bureaucrats in digitized districts collected almost 7 percentage points less of their collection target after digitization, relative to non-digitized districts (*Wild* clustered bootstrapped p-value<0.05). We can exclude the possibility that this is due to the denominator increasing since Table 2 shows that tax demands decreased, if anything, as a result of the digitization reform. In other words, tax collection decreased even more than the tax demands did, implying that the capacity of bureaucrats to collect taxes went down.

One could worry that this is driven by bureaucrats whose tax collection dropped completely, given that targets (i.e. collecting all the tax due) were rarely met even before digitization and that the percentage of tax demand collected was quite low. However, columns (2) and (3) show that the digitization reform also affected the ability of bureaucrat to achieve higher levels of tax demands: bureaucrats were 33 percentage points less likely to collect at least 50% of the tax demands in their area, and 31 percentage points less likely to collect at least 75% of these tax demands (*Wild* clustered bootstrapped p-values<0.05). Finally, column (4) shows that digitization also affected the bottom of the performance distribution: the share of months in which no tax was collected at all increased in digitized districts after the reform, although the effect is not statistically significant.

These results, together with the analysis of the tax base presented in Section 4.1 indicate that the responsibility for the decrease in tax collection lies with the performance of bureaucrats rather than changes in the actual tax base. This decrease in performance can be attributed to both under-reporting of the tax base

and lower tax collection relative to collection targets. As discussed in Section 3 and shown in Figure B.4, the results are unlikely to be driven by a change in the composition of the bureaucracy as there were extremely few transfers across districts. Instead, the results are likely to capture a change in the behavior of bureaucrats.

5 Conclusion

Building strong state capacity is a prerequisite for sustainable economic development. However, state capacity is not simply the sum of the technologies and processes that governments invest in. The capacity of states to raise taxes and protect property rights also depends on the behavior of state officials.

We show that reforms introducing technology to improve the effectiveness of bureaucracies can have unintended consequences by disrupting the relationship between bureaucrats and tax payers. Despite the positive effect of digitization on property rights and agricultural productivity, we find that the collection of agricultural tax decreased as a result of the reform.

Because the reform affected the organization of the bureaucracy and changed the interactions between local bureaucrats and citizens our results highlight the need to consider informal social interactions in the bureaucracy as key dimensions of state capacity (Besley and Dray, 2022; Best et al., 2023).

Our results also have important implications for the design of state capacity reforms. These reforms should be accompanied by appropriate changes to the incentives of bureaucrats, pay close attention to complementarities between different forms of state capacity and whether the reform enhances or reduces these complementarities, and find ways to replace interactions with the local population that are removed by the introduction of technology.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, "Should You Adjust Standard Errors for Clustering?," *The Quarterly Journal of Economics*, September 2023, *138* (1), 1–35.

Aihounton, Ghislain BD and Arne Henningsen, "Units of measurement and the inverse hyperbolic sine transformation," *The Econometrics Journal*, November 2021, *24* (2), 334–351.

Ali, Merima, Abdulaziz B Shifa, Abebe Shimeles, and Firew Woldeyes, "Building fiscal capacity in developing countries: Evidence on the role of information technology," *National Tax Journal*, 2021, *74* (3), 591–620.

Aman-Rana, Shan, "Does Ability Matter for Discretionary Promotions in Bureaucracies? Evidence from Pakistan," *Available at SSRN: https://ssrn.com/abstract=4390268 or http://dx.doi.org/10.2139/ssrn.4390268*, 2023.

_ , Daniel Gingerich, and Sandip Sukhtankar, "Screen Now, Save Later? The Trade-Off between Administrative Ordeals and Fraud," Technical Report w31364, National Bureau of Economic Research June 2023.

Aron, Arthur, Tracy McLaughlin-Volpe, Debra Mashek, Gary Lewandowski, Stephen C Wright, and Elaine N Aron, "Including others in the self," *European Review of Social Psychology*, 2004, *15* (1), 101–132.

Ashraf, Nava and Oriana Bandiera, "Social incentives in organizations," *Annual Review of Economics*, 2018, *10*, 439–463.

_ , _ , Edward Davenport, Scott Lee et al., "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," *American Economic Review*, 2020, *110* (5), 1355–94.

Athey, Susan and Guido W Imbens, "The econometrics of randomized experiments," in "Handbook of economic field experiments," Vol. 1, Elsevier, 2017, pp. 73–140.

Bai, Ying and Ruixue Jia, "Elite recruitment and political stability: the impact of the abolition of China's civil service exam," *Econometrica*, March 2016, *84* (2), 677–733.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat, "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," *The Quarterly Journal of Economics*, August 2021, *136* (4), 2195–2242.

Banerjee, Abhijit V, Esther Duflo, and Rachel Glennerster, "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system," *Journal of the European Economic Association*, May 2008, *6* (2-3), 487–500.

Bardhan, Pranab, "State and development: The need for a reappraisal of the current literature," *Journal of Economic Literature*, 2016, *54* (3), 862–892.

Barnwal, Prabhat, "Curbing leakage in public programs: Evidence from India's direct benefit transfer policy," 2018.

Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken, "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia," *American Economic Review*, December 2021, *111* (12), 3827–71.

Beg, Sabrin, "Digitization and Development: Property Rights Security, and Land and Labor Markets," *Journal of European Economic Association*, February 2020, *20* (1), 395–429.

Bergeron, Augustin, Gabriel Tourek, and Jonathan Weigel, "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC," May 2021, (DP16116).

_ , Pedro Bessone, John Kabeya Kabeya, Gabriel Z Tourek, and Jonathan L Weigel, "Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the DRC," Technical Report 30413, National Bureau of Economic Research September 2022.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu, "The Glittering Prizes: Career Incentives and Bureaucrat Performance," *The Review of Economic Studies*, May 2020, *87* (2), 626–655.

Besley, Timothy and Sacha Dray, "Trust as state capacity: The political economy of compliance," WIDER Working Paper 2022/135, Helsinki 2022.

_ **and Torsten Persson**, "The origins of state capacity: Property rights, taxation, and politics," *American Economic Review*, September 2009, *99* (4), 1218–44.

_ **and** _ , "State capacity, conflict, and development," *Econometrica*, 2010, *78* (1), 1–34.

_ **and** _ , "The causes and consequences of development clusters: State capacity, peace, and income," *Annual Review of Economics*, 2014, *6* (1), 927–949.

_ **and** _ , "Why do developing countries tax so little?," *Journal of economic perspectives*, November 2014, *28* (4), 99–120.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus revenue efficiency with limited tax capacity: theory and evidence from Pakistan," *Journal of political Economy*, 2015, *123* (6), 1311–1355.

_ **, Jonas Hjort, and David Szakonyi**, "Individuals and organizations as sources of state effectiveness," *American Economic Review*, August 2023, *113* (8), 2121–67.

Bhavnani, Rikhil R and Alexander Lee, "Local embeddedness and bureaucratic performance: evidence from India," *The Journal of Politics*, 2018, *80* (1), 71–87.

Bó, Ernesto Dal, Frederico Finan, and Martín A Rossi, "Strengthening state capabilities: The role of financial incentives in the call to public service," *The Quarterly Journal of Economics*, April 2013, *128* (3), 1169–1218.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, April 2023.

Brauw, Alan De and Sylvan Herskowitz, "Income variability, evolving diets, and elasticity estimation of demand for processed foods in Nigeria," *American Journal of Agricultural Economics*, 2021, *103* (4), 1294–1313.

Brockmeyer, Anne, Alejandro Estefan, Karina Ramírez Arras, and Juan Carlos Suárez Serrato, "Taxing property in developing countries: theory and evidence from Mexico," Technical Report 28637, National Bureau of Economic Research April 2021.

_ **and Magaly Sáenz Somarriba**, "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform," March 2022.

Bó, Ernesto Dal, Frederico Finan, Nicholas Y. Li, and Laura Schechter, "Information Technology and Government Decentralization: Experimental Evidence From Paraguay," *Econometrica*, 2021, *89* (2), 677–701.

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

Callen, Michael Joseph, Saad Gulzar, Syed Ali Hasanain, and Muhammad Yasir Yasir Khan, "The political economy of public employee absence: Experimental evidence from Pakistan," August 2013.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Khan, and Arman Rezaee, "Data and policy decisions: Experimental evidence from Pakistan," *Journal of Development Economics*, 2020, *146* (C), S0304387820300985.

_ **,** _ **,** _ **, Muhammad Yasir Khan, and Arman Rezaee**, "Personalities and public sector performance: Evidence from a health experiment in Pakistan," Technical Report 21180, National Bureau of Economic Research August 2020.

_ **,** _ **,** _ **,** _ **, and** _ , "The political economy of public sector absence," *Journal of Public Economics*, 2023, *218*, 104787.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, April 2017, *9* (2), 144–64.

Census, 2010 The Agricultural, "The Agricultural Census 2010," 2010. [Online; accessed 30-July-2023].

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

Chari, Amalavoyal, Elaine M Liu, Shing-Yi Wang, and Yongxiang Wang, "Property rights, land misallocation, and agricultural efficiency in China," *The Review of Economic Studies*, November 2021, *88* (4), 1831–1862.

Cheema, Iftikhar, Simon Hunt, Sarah Javeed, Tanya Lone, and Sean O'Leary, "Benazir income support programme: Final impact evaluation report," *UK: Oxford Policy Management*, 2016.

Chen, Jiafeng and Jonathan Roth, "Log-like? ATEs defined with zero outcomes are (arbitrarily) scale-dependent," *arXiv preprint arXiv:2212.06080*, 2023.

Chen, Yvonne Jie, Pei Li, and Yi Lu, "Career concerns and multitasking local bureaucrats: Evidence of a target-based performance evaluation system in China," *Journal of Development Economics*, 2018, *133*, 84–101.

Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso, "Patronage and selection in public sector organizations," *American Economic Review*, October 2020, *110* (10), 3071–3099.

Debnath, Sisir, Mrithyunjayan Nilayamgode, and Sheetal Sekhri, "Information Bypass: Using Low-cost technological innovations to curb leakages in welfare programs," *Journal of Development Economics*, 2023, *164*, 103137.

Deserranno, Erika, "Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 277–317.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The economics of career concerns, part I: Comparing information structures," *The Review of Economic Studies*, January 1999, *66* (1), 183–198.

_ , _ , and _ , "The economics of career concerns, part II: Application to missions and accountability of government agencies," *The Review of Economic Studies*, January 1999, *66* (1), 199–217.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "The value of regulatory discretion: Estimates from environmental inspections in India," *Econometrica*, 2018, *86* (6), 2123–2160.

_ , Rema Hanna, and Stephen P Ryan, "Incentives work: Getting teachers to come to school," *American Economic Review*, June 2012, *102* (4), 1241–78.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge, "Reallocation effects of the minimum wage," *Quarterly Journal of Economics*, August 2021, *137* (1), 267–328.

Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli, *Technology and Tax Capacity: Evidence from Local Governments in Ghana* number 29923, National Bureau of Economic Research, September 2022.

Dávid-Barrett, Elizabeth and Mihály Fazekas, "Anti-corruption in aid-funded procurement: Is corruption reduced or merely displaced?," *World Development*, 2020, *132*, 105000.

Evans, Peter B, *Embedded autonomy: States and industrial transformation*, Princeton University Press, 1995.

Field, Erica, "Entitled to work: Urban property rights and labor supply in Peru," *The Quarterly Journal of Economics*, November 2007, *122* (4), 1561–1602.

_ and Maximo Torero, "Do property titles increase credit access among the urban poor? Evidence from a nationwide titling program," Working paper, Department of Economics, Harvard University, Cambridge, MA 2006.

Gadenne, Lucie and Monica Singhal, "Decentralization in developing economies," *Annual Review of Economics*, 2014, *6* (1), 581–604.

Garicano, Luis and Paul Heaton, "Information technology, organization, and productivity in the public sector: Evidence from police departments," *Journal of Labor Economics*, 2010, *28* (1), 167–201.

Giulio, Marco Di and Giancarlo Vecchi, "Implementing digitalization in the public sector. Technologies, agency, and governance," *Public Policy and Administration*, 2023, *38* (2), 133–158.

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

Gundhus, Helene OI, Niri Talberg, and Christin T Wathne, "From discretion to standardization: Digitalization of the police organization," *International journal of police science & management*, 2022, *24* (1), 27–41.

Hanna, Rema and Shing-Yi Wang, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, August 2017, *9* (3), 262–90.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, January 1991, *7* (special_issue), 24–52.

Huete, Alfredo, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao, and Laerte G Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote sensing of environment*, 2002, *83* (1-2), 195–213.

Iyer, Lakshmi and Anandi Mani, "Traveling agents: political change and bureaucratic turnover in India," *Review of Economics and Statistics*, 2012, *94* (3), 723–739.

Jensen, Anders, "Employment structure and the rise of the modern tax system," *American Economic Review*, January 2022, *112* (1), 213–234.

Jia, Ruixue, Masayuki Kudamatsu, and David Seim, "Political selection in China: The complementary roles of connections and performance," *Journal of the European Economic Association*, August 2015, *13* (4), 631–668.

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken, "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics*, October 2016, *131* (1), 219–271.

_ , **Asim Ijaz Khwaja, and Benjamin A Olken**, "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings," *American Economic Review*, January 2019, *109* (1), 237–270.

Knebelmann, Justine, Victor Pouliquen, and Bassirou Sarr, "Discretion versus algorithms: Bureaucrats and tax equity in senegal," Technical Report, Working Paper.[6] 2023.

Le, Duong Trung, Edmund Malesky, and Anh Pham, "The impact of local corruption on business tax registration and compliance: Evidence from Vietnam," *Journal of Economic Behavior & Organization*, 2020, *177*, 762–786.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A Olken, and Rohini Pande, "Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia," *American Economic Journal: Economic Policy*, August 2016, *8* (3), 258–283.

Manski, Charles F and John V Pepper, "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions," *Review of Economics and Statistics*, May 2018, *100* (2), 232–244.

Milgrom, Paul and John Roberts, "The economics of modern manufacturing: Technology, strategy, and organization," *American Economic Review*, 1990, pp. 511–528.

Moreira, Diana and Santiago Pérez, "Civil service exams and organizational performance: Evidence from the pendleton act," Technical Report 28665, National Bureau of Economic Research January 2022.

Mullahy, John and Edward C Norton, "Why transform Y? A critical assessment of dependent-variable transformations in regression models for skewed and sometimes-zero outcomes," Technical Report 30735, National Bureau of Economic Research December 2022.

Muralidharan, Karthik, Jishnu Das, Alaka Holla, and Aakash Mohpal, *The fiscal cost of weak governance: Evidence from teacher absence in India*, The World Bank, 2016.

_ , **Paul Niehaus, and Sandip Sukhtankar**, "Building state capacity: Evidence from biometric smartcards in India," *American Economic Review*, October 2016, *106* (10), 2895–2929.

_ , _ , **and** _ , "Identity verification standards in welfare programs: Experimental evidence from India," *Review of Economics and Statistics*, February Forthcoming.

_ , _ , _ , **and Jeffrey Weaver**, "Improving last-mile service delivery using phone-based monitoring," *American Economic Journal: Applied Economics*, April 2021, *13* (2), 52–82.

Naritomi, Joana, "Consumers as Tax Auditors," *American Economic Review*, September 2019, *109* (9), 3031–72.

Okunogbe, Oyebola and Victor Pouliquen, "Technology, taxation, and corruption: evidence from the introduction of electronic tax filing," *American Economic Journal: Economic Policy*, February 2022, *14* (1), 341–72.

_ **et al.**, *Becoming legible to the state: the role of detection and enforcement capacity in tax compliance*, World Bank, 2021.

Pepinsky, Thomas B, Jan H Pierskalla, and Audrey Sacks, "Bureaucracy and service delivery," *Annual Review of Political Science*, 2017, *20*, 249–268.

Plesner, Ursula, Lise Justesen, and Cecilie Glerup, "The transformation of work in digitized public sector organizations," *Journal of Organizational Change Management*, 2018, *31* (5), 1176–1190.

Pomeranz, Dina, "No taxation without information: Deterrence and self-enforcement in the value added tax," *American Economic Review*, August 2015, *105* (8), 2539–2569.

Pors, Anja Svejgaard and Eva Pallesen, "The reorganization of the bureaucratic encounter in a digitized public administration.," *ephemera: theory & politics in organization*, 2021, *21* (3).

Punjab Agricultural Income Tax Act

Punjab Agricultural Income Tax Act, http://punjablaws.gov.pk/laws/398.html *(Accessed: 2023-07-12) 1997.*

Rambachan, Ashesh and Jonathan Roth, *"A More Credible Approach to Parallel Trends," The Review of Economic Studies, 02 2023. rdad018.*

Rasmussen, M. S., *"Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR.," International Journal of Remote Sensing, 1992, 13 (18), 3431–3442.*

Rasul, Imran and Daniel Rogger, *"Management of bureaucrats and public service delivery: Evidence from the nigerian civil service," The Economic Journal, 2018, 128 (608), 413–446.*

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, *"What's trending in difference-in-differences? A synthesis of the recent econometrics literature," Journal of Econometrics, 2023.*

Sequeira, Sandra, *"Displacing corruption," Technical Report, Mimeo. London School of Economics 2011.*

_ , *"Corruption, trade costs, and gains from tariff liberalization: Evidence from Southern Africa," American Economic Review, October 2016, 106 (10), 3029–3063.*

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

Tsai, Lily L, *"Solidary groups, informal accountability, and local public goods provision in rural China," American Political Science Review, 2007, 101 (2), 355–372.*

Vrieling, Anton, Kirsten M de Beurs, and Molly E Brown, *"Variability of African farming systems from phenological analysis of NDVI time series," Climatic change, December 2011, 109, 455–477.*

Yang, Dean, *"Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines," The Review of Economics and Statistics, 02 2008, 90 (1), 1–14.*

Young, Alwyn, *"Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," The Quarterly Journal of Economics, November 2019, 134 (2), 557–598.*

Tables

Table 1: Does digitization of land records affect district level tax collection?

Dependent variables:	IHS Tax collection
	(1)
Digitization of land records	-1.826***
	(0.634)
	[0.01]
Control mean of outcome	9.05
District fixed effects	Yes
Fiscal year fixed effects	Yes
Observations	212

Notes: The unit of observation is a district-fiscal year. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. IHS tax collection is the inverse hyperbolic sine of the amount of taxes in (000) PKR that were collected by the government in a district-fiscal year. Standard errors are clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Does digitization of land records affect agricultural tax base?

	Log farm level profit per acre (1)	Satellite vegetation cover index (2)	Whether agri land irrigated? (3)	Log agricultural land owned (4)
Digitization of land records	0.923***	0.00597	0.00945	0.0645
	(0.0591)	(0.00571)	(0.0510)	(0.0445)
	[0.000]	[0.262]	[0.789]	[0.174]
Control mean of outcome	2.43	0.53	7.69	1.24
District fixed effects	Yes	Yes	Yes	Yes
Fiscal year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	No	Yes	Yes
Observations	5,958	288	161,796	161,836

Notes: The unit of observation is a district-fiscal year. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. Farm level profit per acre (PKR) is from Beg (2020) and it is calculated as the difference in the value of output per acre and the total expenses per acre. We take the natural log of that meausre. This measure uses HIES data from cultivating households for the 2005, 2007, 2011, and 2013 waves of the survey. Satellite vegetation cover index is the Normalized Difference Vegetation Index (NDVI) measured between -1 and 1, This data is from NASA's MODIS land products. Whether agricultural land is irrigated is from the PSLM survey data and it is an dummy variable that takes the value 1 when the agricultural land of the household is irrigated. Agricultural land owned is the acres of agricultural land that is owned by households. We take the natural log of agricultural land owned. This measure is also compiled from the PSLM survey. The data is representative at the district level. The waves used in this study are for the years 2006, 2008, 2010 and 2012. Standard errors are clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) p-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 3: District level tax base as assessed by bureaucrats

	Log reported cultivated area	Log admin tax targets
	(1)	(2)
Digitization of land records	-0.106***	-0.600**
	(0.0379)	(0.232)
	[0.004]	[0.015]
Control mean of outcome	6.88	9.74
District fixed effects	Yes	Yes
Fiscal year fixed effects	Yes	Yes
Controls	No	No
Observations	214	203

Notes: The unit of observation is a district-fiscal year. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. The reported cultivated area is measured in (000) acres, while the administrative tax targets is in (000) PKR. We take the natural logs of both measures. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

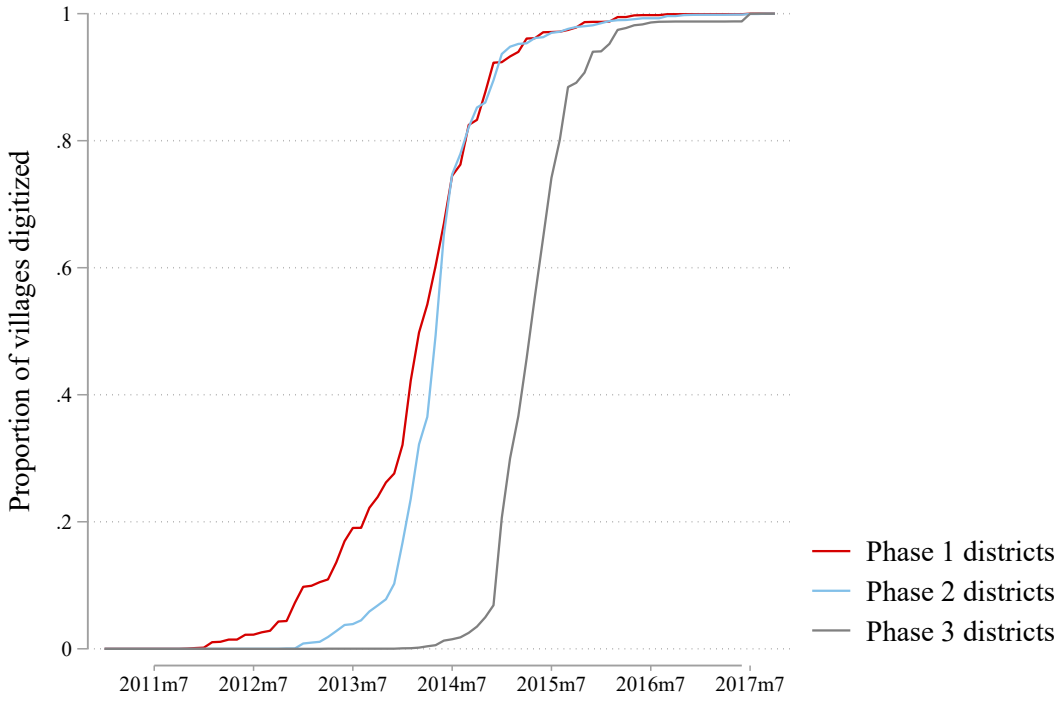
Table 4: Does digitization of land records affect tax collection performance of local officials?

Dependent Variables:	**Performance of bureaucrats**			
	$\frac{Tax\ collected}{Tax\ target}$(%)	Whether at least 50% tax target was achieved	Whether at least 75% tax target was achieved	Share of months with zero collection
	(1)	(2)	(3)	(4)
Digitization of land records	-6.572**	-0.327*	-0.309*	0.225
	(2.708)	(0.161)	(0.158)	(0.139)
	[0.011]	[0.038]	[0.045]	[0.111]
Control mean of outcome	6.72	0.61	0.52	0.19
Bureaucrat fixed effects	Yes	Yes	Yes	Yes
Fiscal year fixed effects	Yes	Yes	Yes	Yes
Controls	No	No	No	No
Observations	304	304	304	304

Notes: The unit of observation is a bureaucrat-revenue circle-fiscal year. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. The first measure compares the tax they collected to the targets they were given by the Board of Revenue. The second and third measure are dummy variables that takes value 1 if at least 50% (75%) of the annual tax target was achieved, and remains zero otherwise. The final measure is the share of months in the fiscal year in which no tax was collected. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

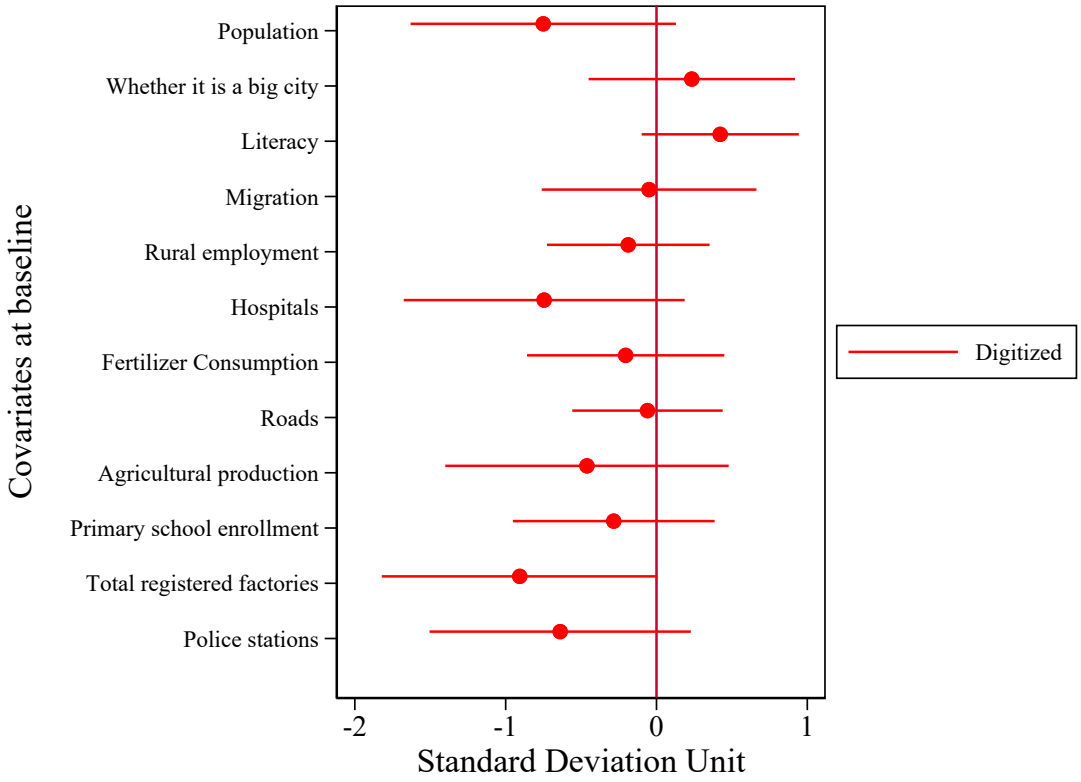
Figures

Figure 1: Phase wise roll-out of digitization reform over time



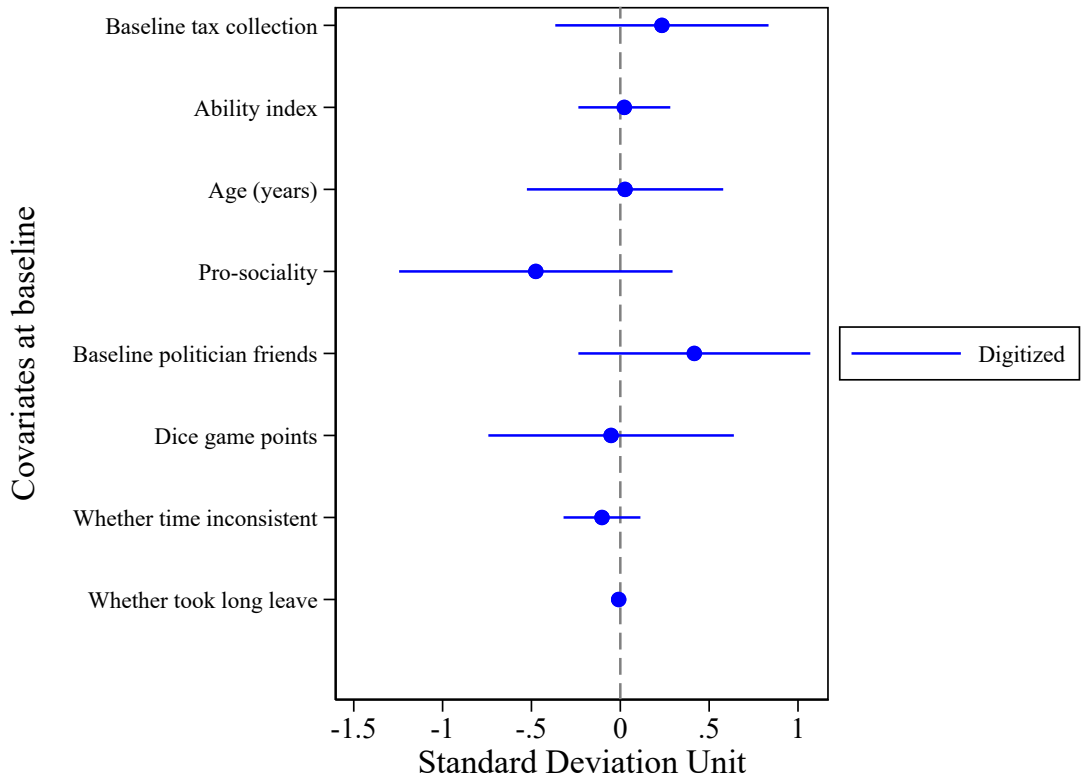
Notes. Districts that were planned to be digitized in phase 1 are Lahore, Lodhran, Hafizabad, Mandi Bahauddin, Nankana Sahib, Jhelum, Gujrat, Sialkot, Chakwal, Attock, Rawalpindi. Districts that were planned to be digitized in phase 2 are Bahawalpur, Gujranwala, Jhang, Layyah, Kasur, Multan, Muzaffargarh, Narowal, Okara, Rahim Yar Khan, Sargodha, Sheikhupura, Toba Tek Singh. Districts that were planned for phase 3 were Bahawalnagar, Bhakkar, Chiniot, Dera Ghazi Khan, Faisalabad, Mianwali, Khanewal, Khushab, Pakpattan, Rajanpur, Sahiwal, Vehari.

Figure 2: Balance test of baseline characteristics of districts with digitized and manual land records



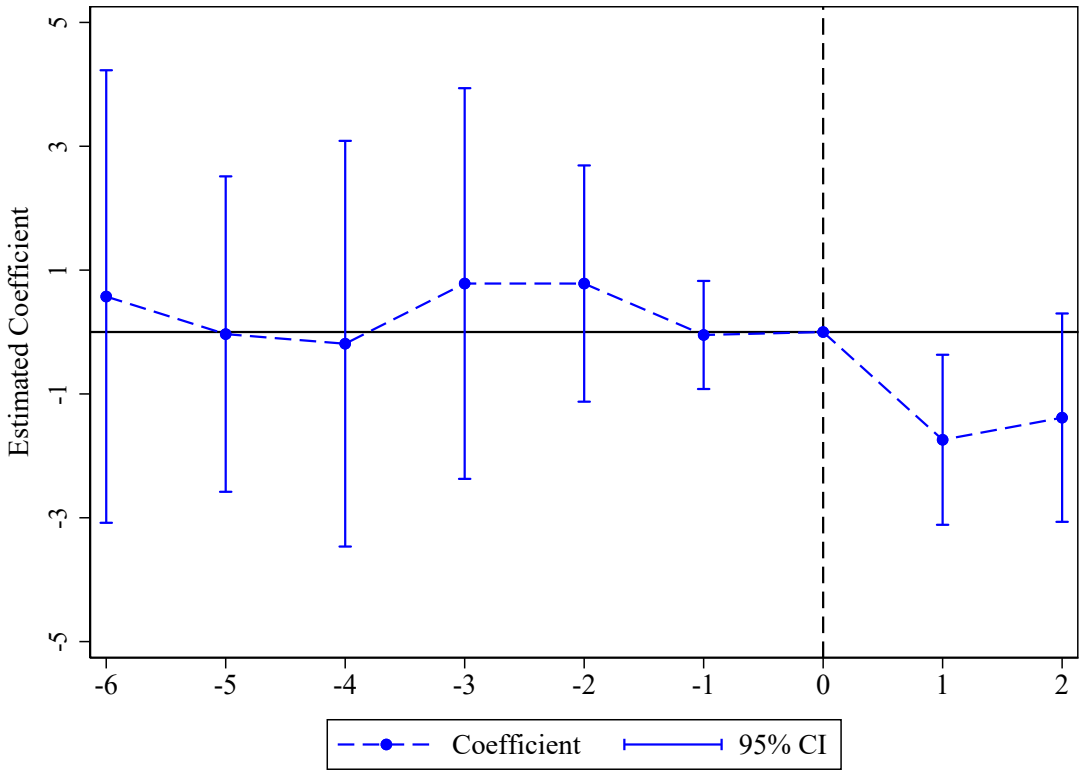
Note: Data on baseline characteristics is from the Development Statistics of the Pakistan Bureau of Statistics 1997-2010. The point estimates are from a regression of the respective covariates on a dummy that takes the value of one if the district lies in phase 1 and 2 of the digitization reform, and remains zero otherwise. The reference category are phase 3 districts. Intervals are 95% confidence intervals.

Figure 3: Balance test of baseline characteristics of bureaucrats in districts with digitized and manual land records



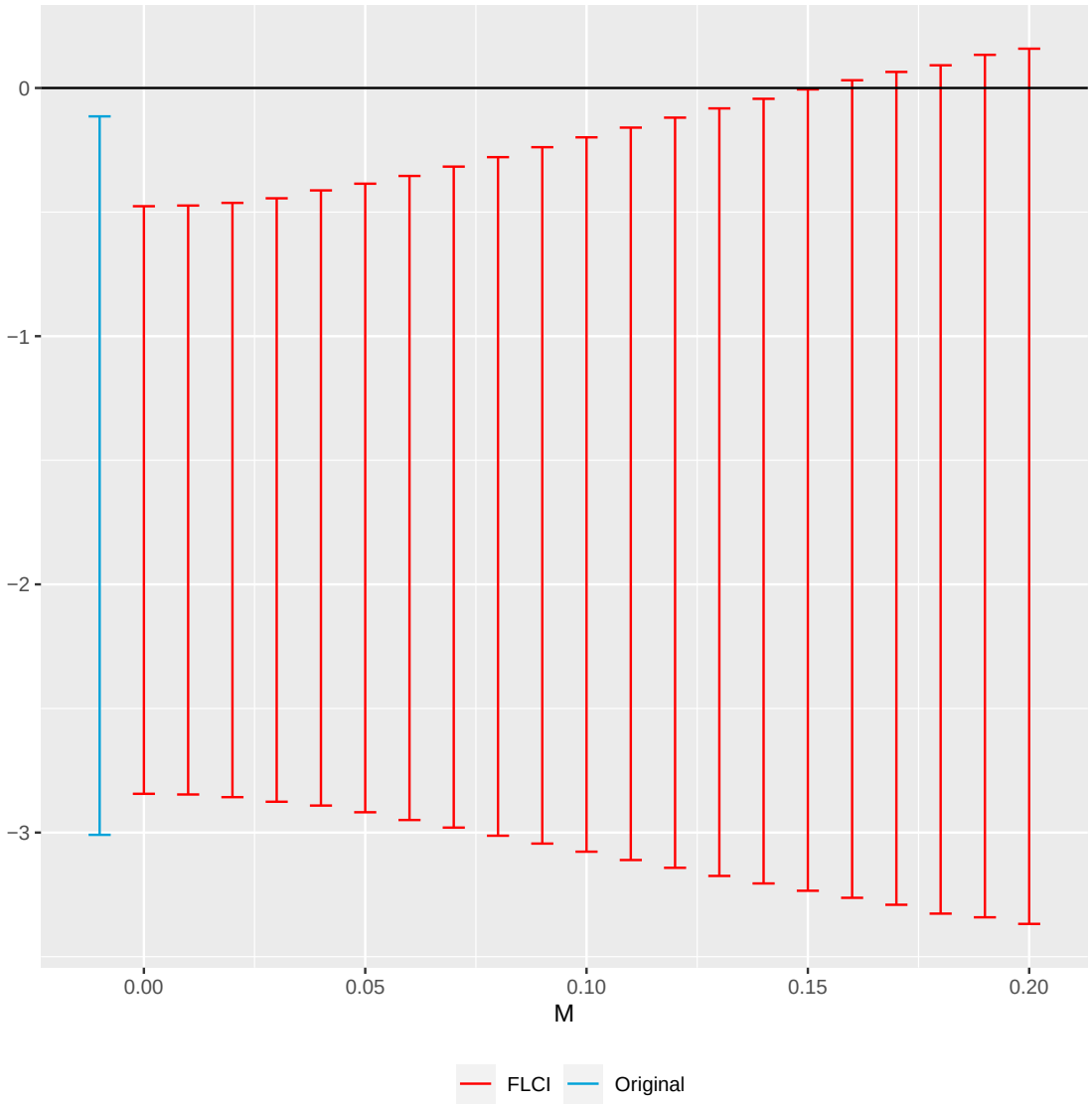
Note: Data on baseline characteristics is from the bureaucrat survey. The point estimates are from a regression of the respective covariates on a dummy that takes the value of one if the district lies in phase 1 and 2 of the digitization reform, and remains zero otherwise. The reference category are phase 3 districts. Tax collection is the amount of agricultural taxes they collected at baseline. Ability index is created from four measures: an incentivized matrix game and a memory game as in Hanna and Wang (2017), response to questions on general knowledge and revenue rules and regulations, respectively. Pro-sociality index is created from five measures: Inclusion of Others in Self scale (Aron et al., 2004; Ashraf et al., 2020), whether they have donated blood, money donated in public good game, whether they do volunteer work and whether they give charity. Politician friends are the number of friends of the bureaucrats that are either federal or provincial politicians. Dice game points is a proxy for dishonesty and it is the total in an incentivized dice game as in Hanna and Wang (2017). Intervals are 95% confidence intervals.

Figure 4: Event study plot for the district level IHS tax collection



Notes: Data is at the district-fiscal year level. Each coefficient is obtained from a set of indicator variables that take value one if, for phase 1 and 2 districts respectively in a fiscal year, the introduction of digitization of land records was k years away as described in Equation 1. The reference year is 2011 for phase 1 and 2012 for phase 2 districts. District and year fixed effects are included. Standard errors were clustered at the district level.

Figure 5: Effects on the inverse hyperbolic sine of tax collection - Robustness of the DID estimates using the "more credible approach" approach suggested by (Rambachan and Roth, 2023)



Notes. The specific approach used to create the plot is second differences (Δ^{SD}). FLCI refers to fixed length confidence intervals. The blue line is the confidence intervals for the coefficient obtained from the interaction term between digitization and post in an event-study regression similar to Equation 1. The only difference between this regression and Equation 1 is that in this version, only for years after 2011 for phase 1 and 2012 for phase 2 we let the digitization indicator interact with post. The reference year for phase 1 is 2011, while that for phase 2 is 2012.

Appendices

A Appendix tables

Table A.1: The correlation of planned and actual roll-out of digitization reform

Dependent Variables:	Number of villages digitized	
	(1)	
Digitization of land records	173.3***	
	(24.00)	
	[0.000]	
Phase 1		129.1***
		(20.50)
		[0.000]
Phase 2		243.2***
		(41.84)
		[0.000]
Control mean of outcome	34.14	34.14
District fixed effects	Yes	Yes
Fiscal year fixed effects	Yes	Yes
Observations	219	219

Notes: Data is at the district-fiscal year level. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. Phase 1 is a dummy that takes the value one after 2011 for all the districts that were planned to be digitized in phase 1, and remains zero otherwise. Phase 2 is a dummy that takes the value one after 2012 for all the districts that were planned to be digitized in phase, and remains zero otherwise. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.2: OLS regression of digitized villages and tax collection

Dependent variable:	IHS Tax collection
	(1)
IHS num of digitized villages	-0.27**
	(0.12)
	[0.021]
Control mean of outcome	9.05
District fixed effects	Yes
Fiscal year fixed effects	Yes
Observations	212

Notes: Data is at the district-fiscal year level. IHS num of digitized villages is the inverse hyperbolic sine of the number of villages that had digitized their land records. IHS tax collection is the inverse hyperbolic sine of the amount of taxes. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.3: Does digitization of land records affect district level tax collection? (robustness to the use of inverse hyperbolic sine of tax collection)

	Median regression	Levels of tax	Intensive Margin	Extensive Margin
Dependent variables:	IHS tax collection	Tax collection (000 PKR)	Log tax collection	Whether any tax collected?
	(1)	(2)	(3)	(4)
Digitization of land records	-0.780*	-8053.4*	-1.457***	-0.0349
	(0.429)	(4120.2)	(0.440)	(0.0376)
	[0.0069]	[0.028]	[0.0026]	[0.32]
Control mean of outcome	9.05	16278.9	9.09	0.93
District fixed effects	Yes	Yes	Yes	Yes
Fiscal year fixed effects	Yes	Yes	Yes	Yes
Observations	212	212	201	212

Notes: Data is at the district-fiscal year level. IHS tax collection is the inverse hyperbolic sine of the amount of taxes. Tax collection is the amount of taxes collected by the government in 000 PKR. Log tax collection is the natural log of tax collection. *Whether any tax collected* is a dummy variable that takes the value 1 if the tax collected in a district-year is positive, and remains zero otherwise. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.4: Does digitization of land records affect district level tax collection? (robustness to different thresholds defining the beginning of a phases of digitization)

Dependent variables:	IHS Tax collection		
Timing of phases:	First 1% villages digitized in a phase	First 2% villages digitized in a phase	First 5% villages digitized in a phase
	(1)	(2)	(3)
Digitization of land records	-0.571	-0.980*	-1.827***
	(0.572)	(0.508)	(0.634)
	[0.27]	[0.03]	[0.01]
Control mean of outcome	9.03	9.03	9.05
District fixed effects	Yes	Yes	Yes
Fiscal year fixed effects	Yes	Yes	Yes
Observations	178	212	212

Notes: The unit of observation is a district-fiscal year. In Column (1) and (2) *Digitization of land records* is a dummy variable that takes value 1 for phase 1 and 2 districts starting in years 2010 and 2011 respectively, and remains zero otherwise. However, Phase 3 is categorized to start in 2013 in Column (1) and in 2014 in Column (2). Column (3) replicates Table 1 and categorizes *Digitization of land records* as a dummy variable that takes value 1 for phase 1 and 2 districts starting in years 2012 and 2013 respectively, and remains zero otherwise. Phase 3 is categorized to start in 2014 in Column (3). IHS tax collection is the inverse hyperbolic sine of the amount of taxes in (000) PKR that were collected by the government in a district-fiscal year. Standard errors are clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.5: Comparison of districts whose land records were digitized in phase 1 and 2 with the districts in phase 3 (never treated group)

Dependent variables:	IHS Tax collection	
	(1)	(2)
Digitization of land records (Phase 1 vs. Phase 3)	-2.008**	
	(0.905)	
Digitization of land records (Phase 2 vs. Phase 3)		-1.766***
		(0.588)
Control mean of outcome	8.91	9.57
District fixed effects	Yes	Yes
Fiscal year fixed effects	Yes	Yes
Observations	126	156

Notes: Data is at the district-fiscal year level. IHS tax collection is the inverse hyperbolic sine of the amount of taxes in (000) PKR that were collected by the government in a district-fiscal year. Phase 1, 2 and 3 are districts that were planned to be digitized in phase 1, 2 and 3, respectively, and remains zero otherwise. Phase 2 is a dummy that takes the value one for all the districts that were planned to be digitized in phase, and remains zero otherwise. Phase 3 are those districts that are never treated in the sample period 2006-2013. Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.6: Does digitization of land records affect agricultural tax base? (robustness using Callaway and Sant'Anna (2021) estimator)

(a) Callaway and Sant'Anna (2021) estimator based on event

Dependent variables:	IHS Tax collection
	(1)
Pre-treatment average	0.18
	(0.30)
Post-treatment average	-1.96**
	(0.76)

(b) Callaway and Sant'Anna (2021) estimator based on event

Dependent variables:	IHS Tax collection
	(1)
Group average	-1.68***
	(0.54)

Notes: Standard errors clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) p-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.7: Does digitization of land records affect district level tax collection? (placebo test based on different years of start of digitization)

Dependent variable:	IHS Tax collection	
Placebo:	Phase 1 2006 Phase 2 2007	Phase 1 2009 Phase 2 2010
	(1)	(2)
Digitization of land records	-0.475	-0.394
	(1.827)	(0.790)
	[0.78]	[0.59]
Control mean of outcome	9.27	9.27
District fixed effects	Yes	Yes
Fiscal year fixed effects	Yes	Yes
Observations	212	212

Notes: The unit of observation is a district-fiscal year. The placebo is based on categorising digitization of land records as a dummy variable that takes value 1 for phase 1 and 2 districts in years 2006 (2009) and 2007 (2010) respectively, and remains zero otherwise. IHS tax collection is the inverse hyperbolic sine of the amount of taxes in (000) PKR that were collected by the government in a district-fiscal year. Standard errors are clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

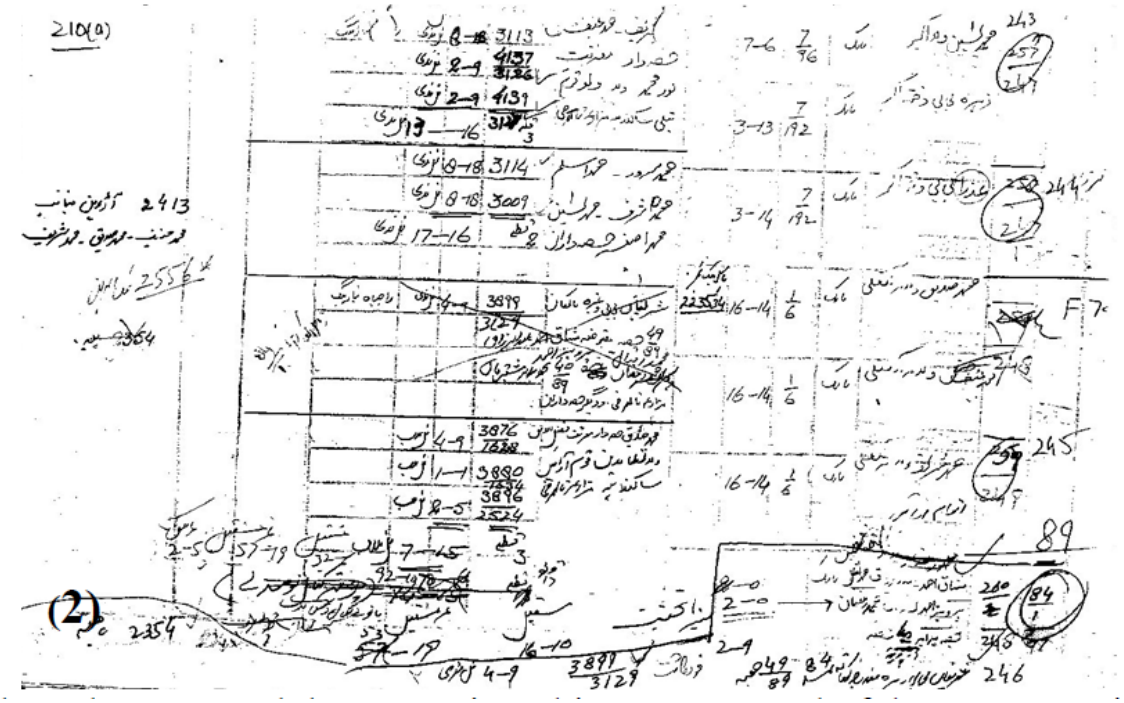
Table A.8: Does digitization of land records affect district level tax collection? (Randomization inference *p*-value)

Dependent variable:	IHS Tax collection
	(1)
Digitization of land records	-1.826***
	(0.634)
	[0.008]
Randomization inference *p*-val	0.000
Control mean of outcome	9.267
District fixed effects	Yes
Fiscal year fixed effects	Yes
Observations	212

Notes: The unit of observation is a district-fiscal year. Digitization of land records is a dummy variable that takes value 1 for phase 1 and 2 districts in years after 2011 and 2012 respectively, and remains zero otherwise. IHS tax collection is the inverse hyperbolic sine of the amount of taxes in (000) PKR that were collected by the government in a district-fiscal year. Standard errors are clustered at district level in parentheses. Wild clustered bootstrapped (with 10,000 replications) *p*-values are in brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

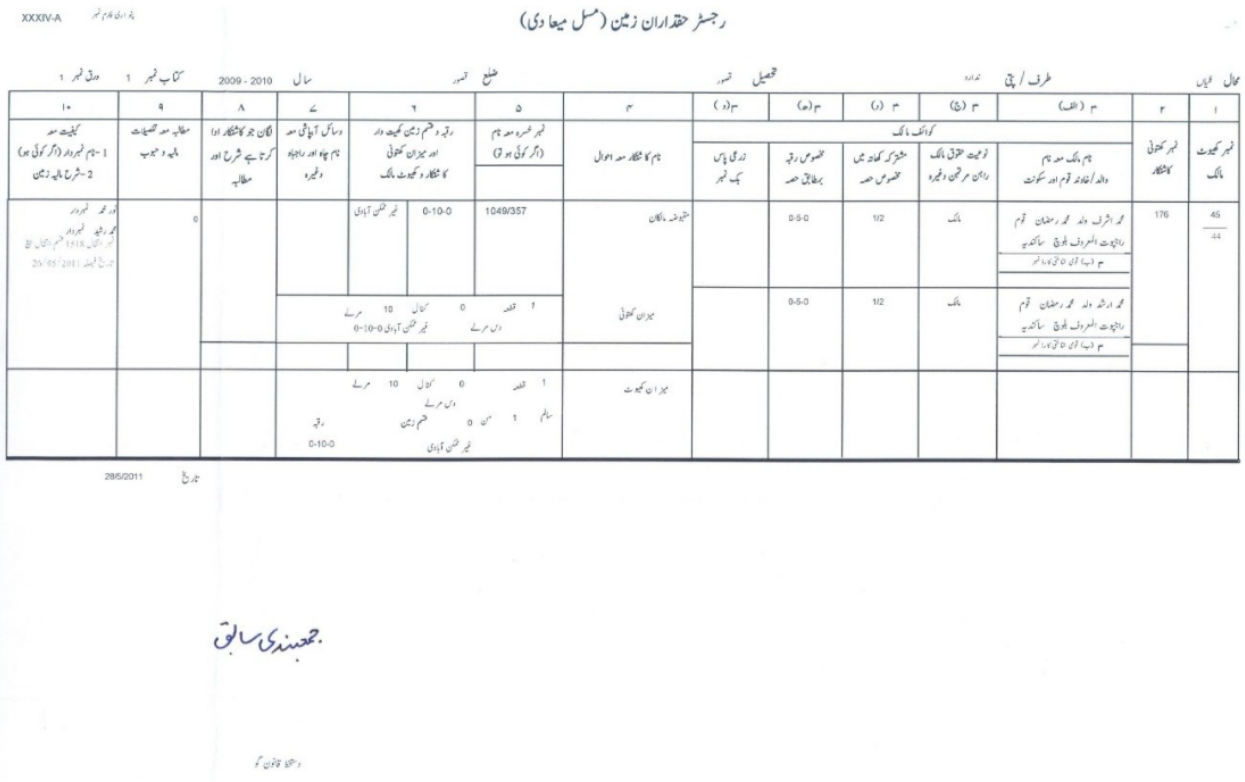
B Appendix: Figures

Figure B.1: An example of a manually entered land record



Notes. The source of the figure is Adeel (2010), Figure 2. Before digitization such land records were maintained by the tax collectors.

Figure B.2: An example of a computerized land record



رجسٹر حقداران زمین (مسل میعادی)

XXXIV-A

۱	۲	۳ (الف)	۳ (ب)	۳ (ج)	۳ (د)	۳ (ہ)	۳ (و)	۴	۵	۶	۷	۸	۹	۱۰
نمبر کھیوٹ مالک	نمبر کھتونی کاشتکار	نام مالک معہ نام والد/خاوند قوم اور سکونت	کوائف مالک: نوعیت حقوق مالک زمین مرہون وغیرہ	کوائف مالک: مشترکہ کھاتہ میں مخصوص حصہ	کوائف مالک: مخصوص رقبہ برطابق حصہ	کوائف مالک: زرعی یا کیا نمبر	تفصیل نمبر	نام کاشتکار معہ احوال	نمبر خسرہ معہ نام (اگر کوئی ہو تو)	رقبہ و قسم زمین کھیوٹ وار اور میزان کھتونی کاشتکار و کھیوٹ مالک	وسائل آبپاشی معہ نام چاہ اور راجباہ وغیرہ	لگان جو کاشتکار ادا کرتا ہے شرح اور مطالبہ	مطالبہ معہ تصفیات مالیہ و جیب	کیفیت معہ 1-نام نمبردار (اگر کوئی ہو) 2-شرح مالیہ زمین
45/44	176	محمد اشرف ولد محمد رمضان قوم راجپوت المعروف لوہرچ ساکنہ م (ب) لوئی نا لوئی وارثہ نمبر	مالک	1/2	0-5-0			متبوضہ مالکان میزان کھتونی	1049/357	0-10-0 غیر نہری آبادی			0	محمد نمبردار
		محمد رشید ولد محمد رمضان قوم راجپوت المعروف لوہرچ ساکنہ م (ب) لوئی نا لوئی وارثہ نمبر	مالک	1/2	0-5-0				۱ قطعہ 0 کنال 10 مرلے دس مرلے غیر نہری آبادی 0-10-0					
								میزان کھیوٹ	۱ قطعہ 0 کنال 10 مرلے دس مرلے سلام 1 من 0 قسم زمین رقبہ غیر نہری آبادی 0-10-0					

جمعبندی سابق

29/5/2011 تاریخ

Notes. The source of the figure is: http://cadastraltemplate.org/pakistan.php

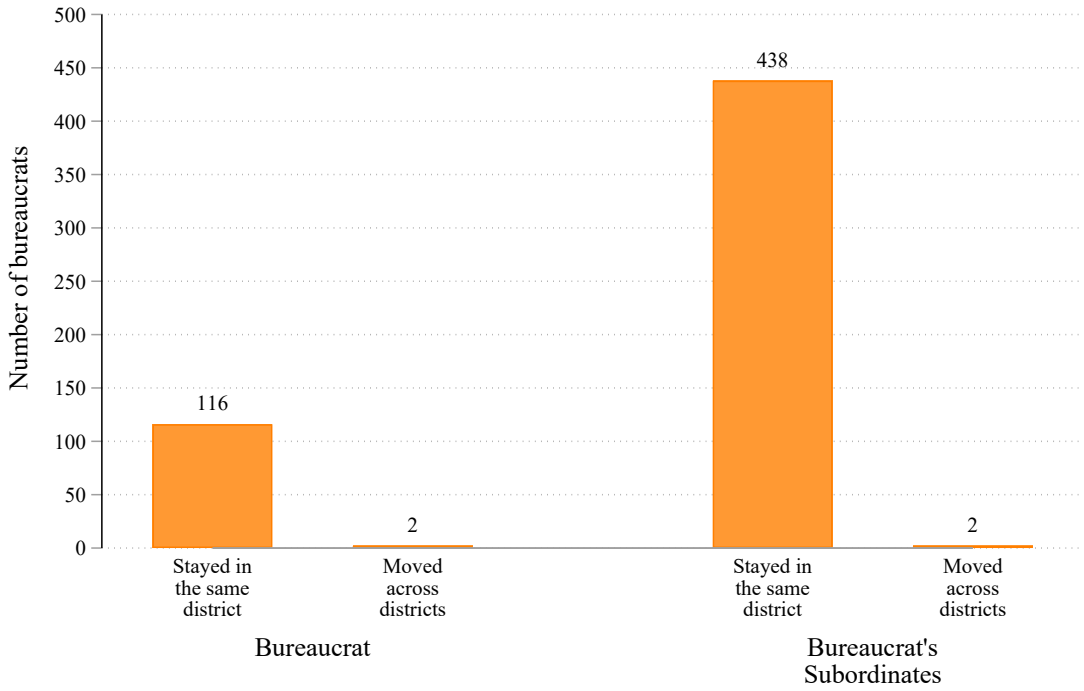
Figure B.3: A new bureaucracy set up to handle digitized land records



Notes. The source of the image is the World Bank (2017). New centers were

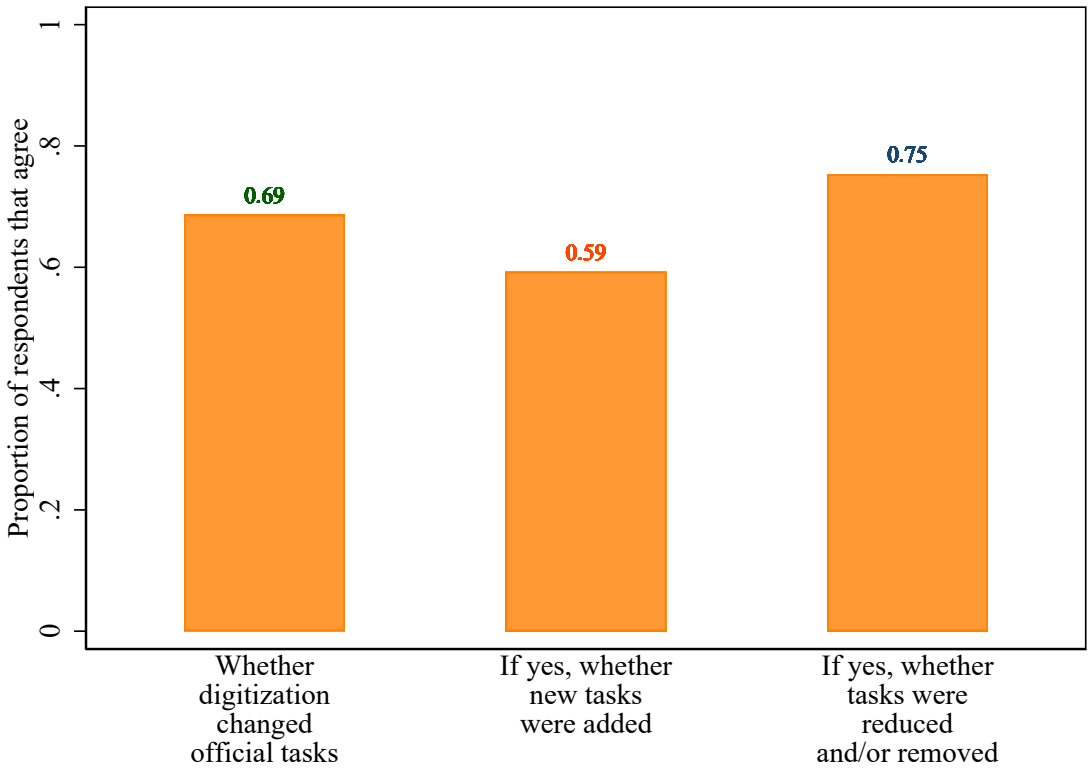
set up across Punjab to deliver computerized land record services.

Figure B.4: Movement of bureaucrats across districts



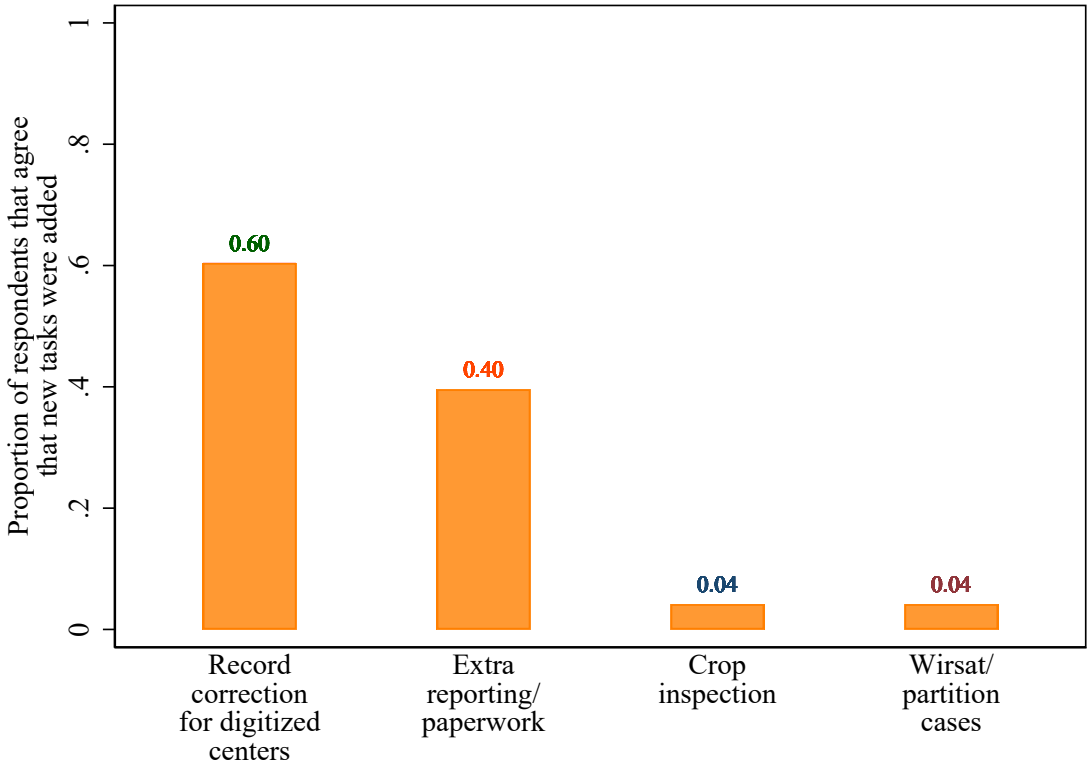
Notes. The figure is based on the bureaucrat survey. On the left hand side are the bureaucrats that are the focus of this paper. On the right hand side are careers of their subordinates.

Figure B.5: Changes in the tasks of the bureaucrats after digitization of land records



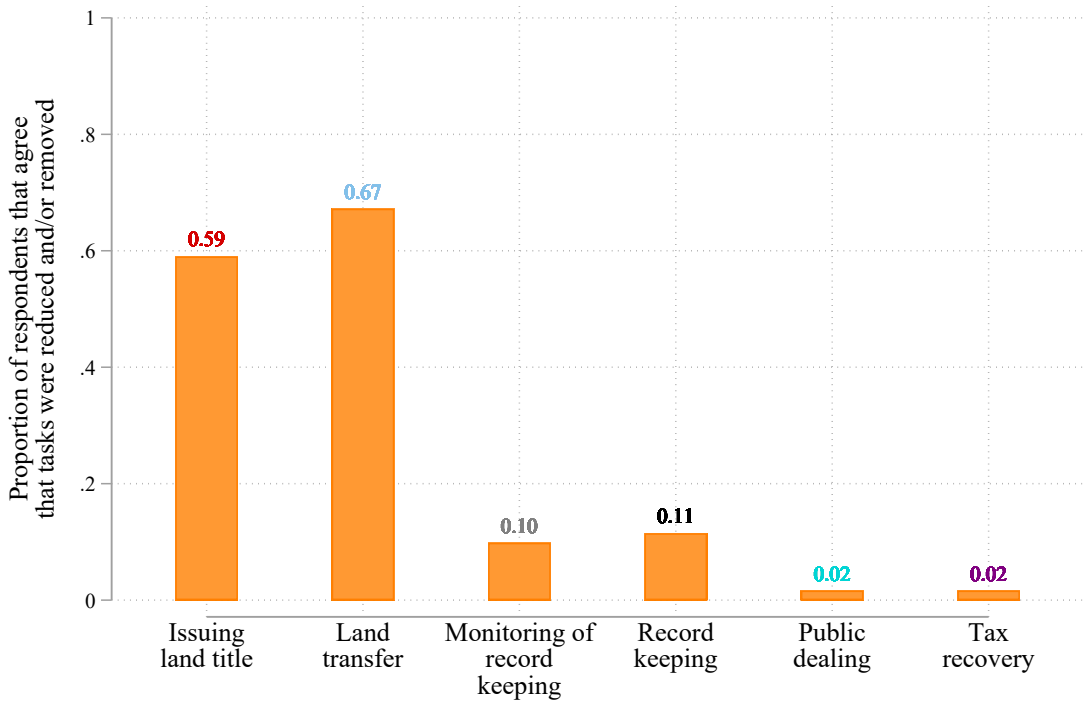
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013. The first bar plots the proportion of bureaucrats that agree with "Do you think LRMIS (digitization reform) changed the official tasks that you are supposed to do?"

Figure B.6: Additional tasks that bureaucrats carry out after digitization of land records



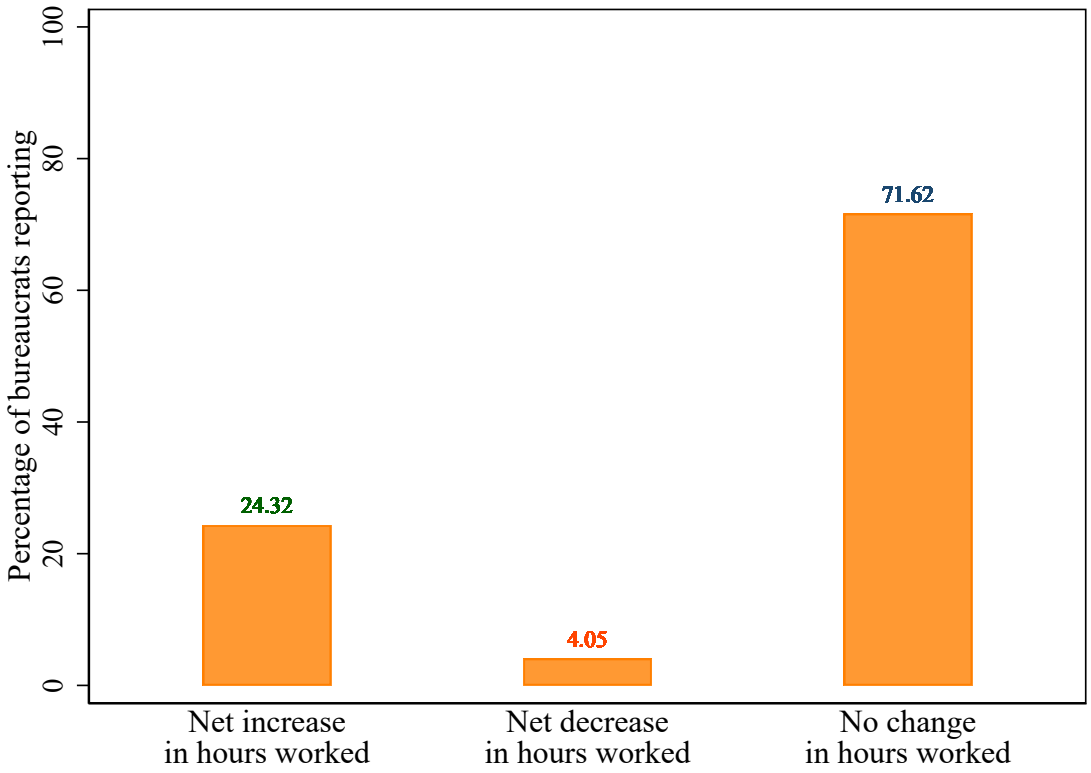
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013. The figure is based on the responses of bureaucrats who agreed with the question "Do you think LRMIS (digitization reform) changed the official tasks that you are supposed to do?"

Figure B.7: Bureaucrats' tasks that were removed and reduced after digitization of land records



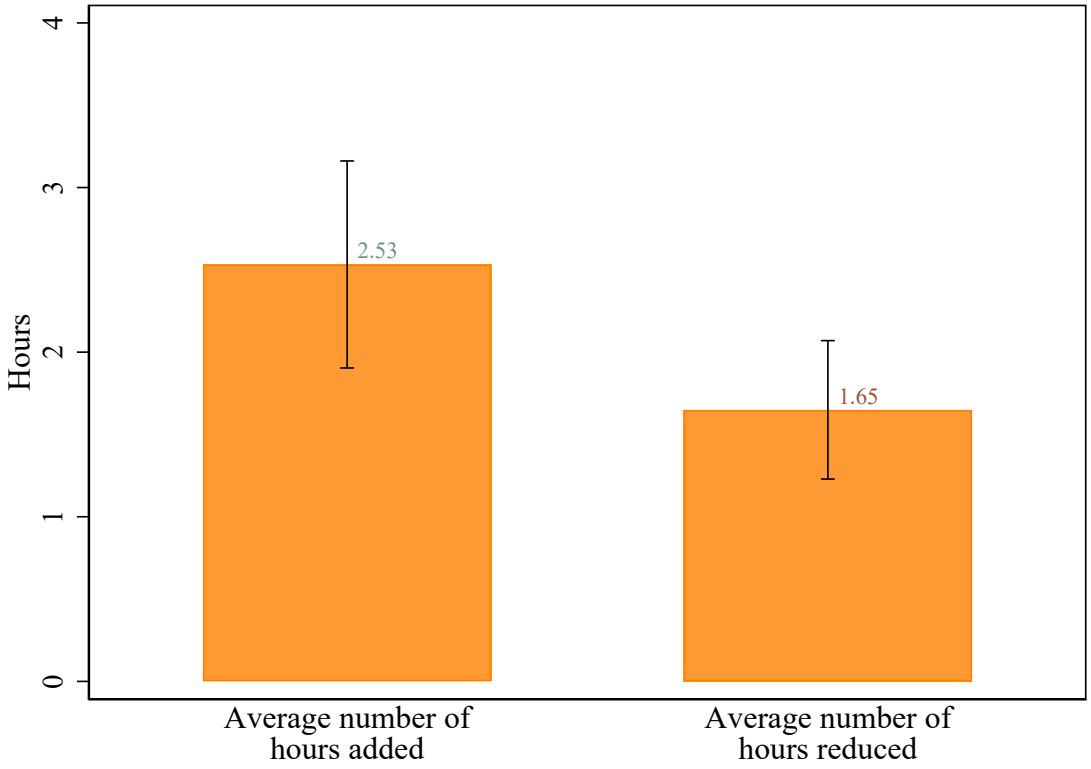
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013. The figure is based on the responses of bureaucrats who agreed with the question "Do you think LRMIS (digitization reform) changed the official tasks that you are supposed to do?"

Figure B.8: Changes in hours worked by bureaucrats after digitization of land records



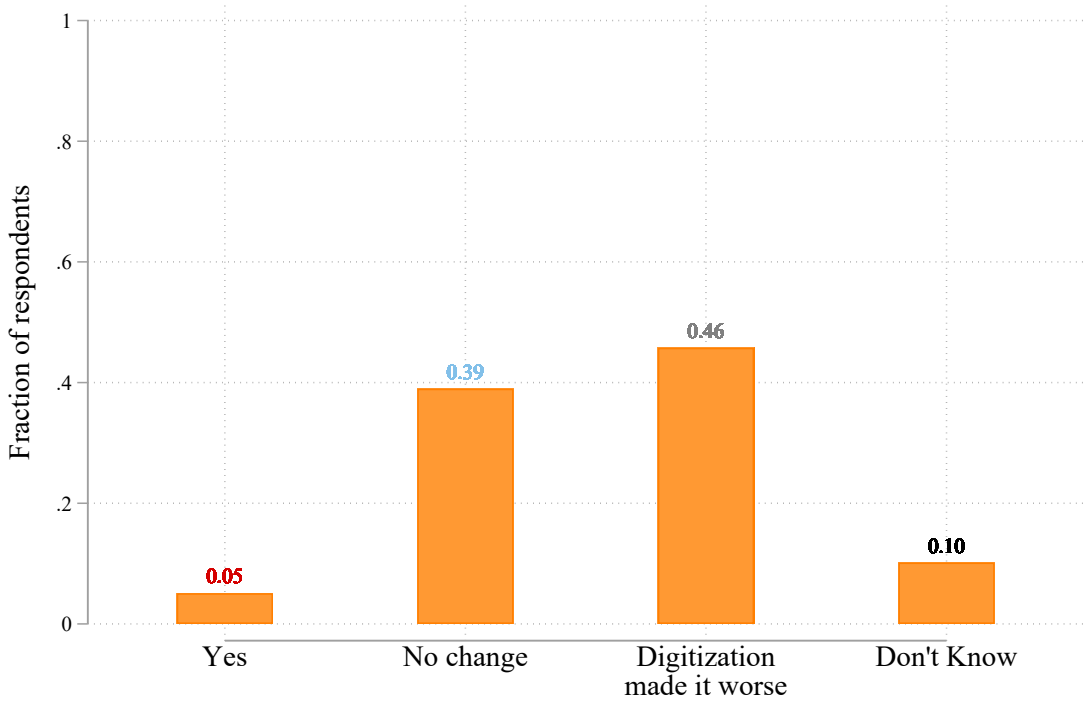
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013. The figure is based on the responses of bureaucrats who agreed with the question "Do you think LRMIS (digitization reform) changed the official tasks that you are supposed to do?"

Figure B.9: Changes in number of hours worked by bureaucrats after digitization of land records



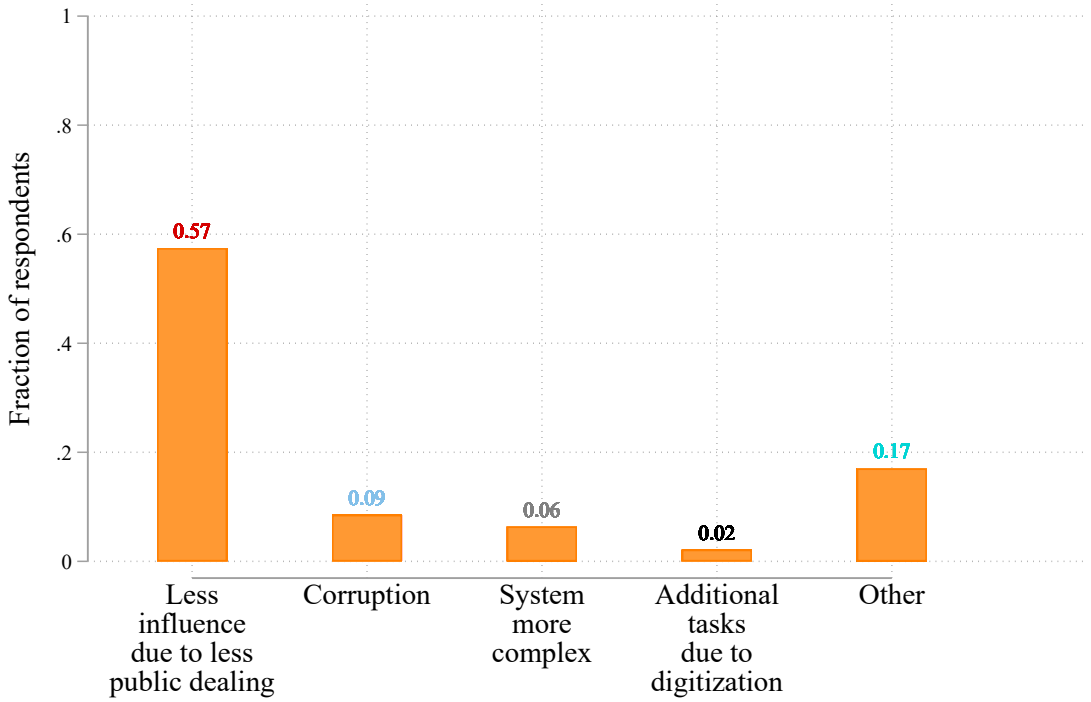
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013. The figure is based on the responses of bureaucrats who agreed with the question "Do you think LRMIS (digitization reform) changed the official tasks that you are supposed to do?"

Figure B.10: Bureaucrats' response to the question "Do you think digitization has improved overall tax collection?"



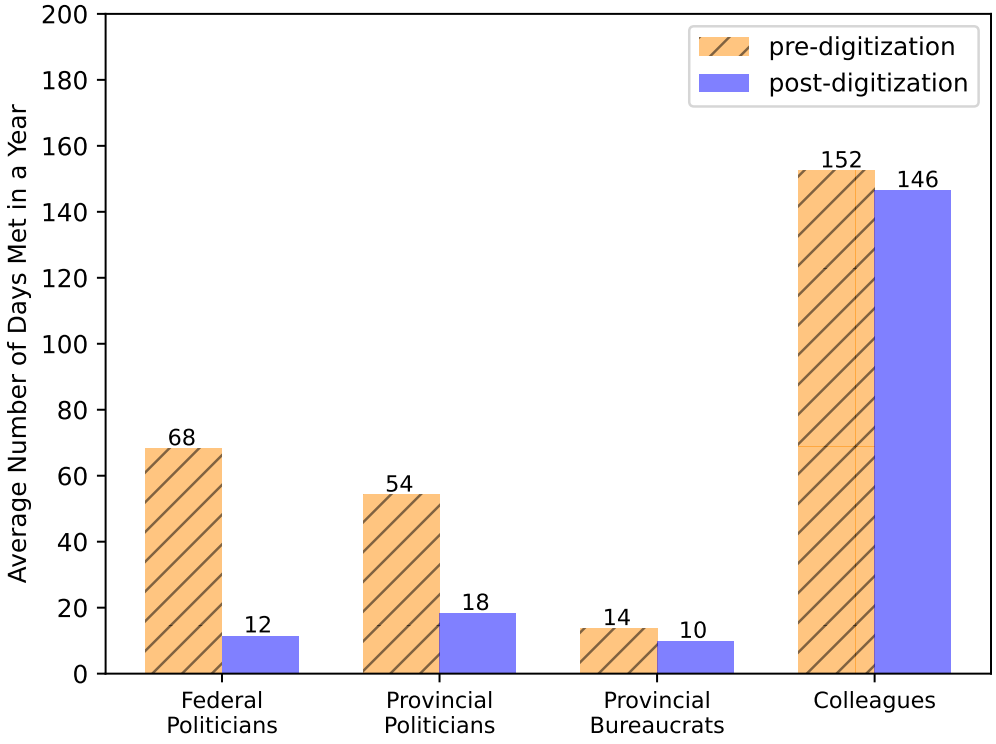
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013.

Figure B.11: Bureaucrats' views on why digitization did not improve tax collection?



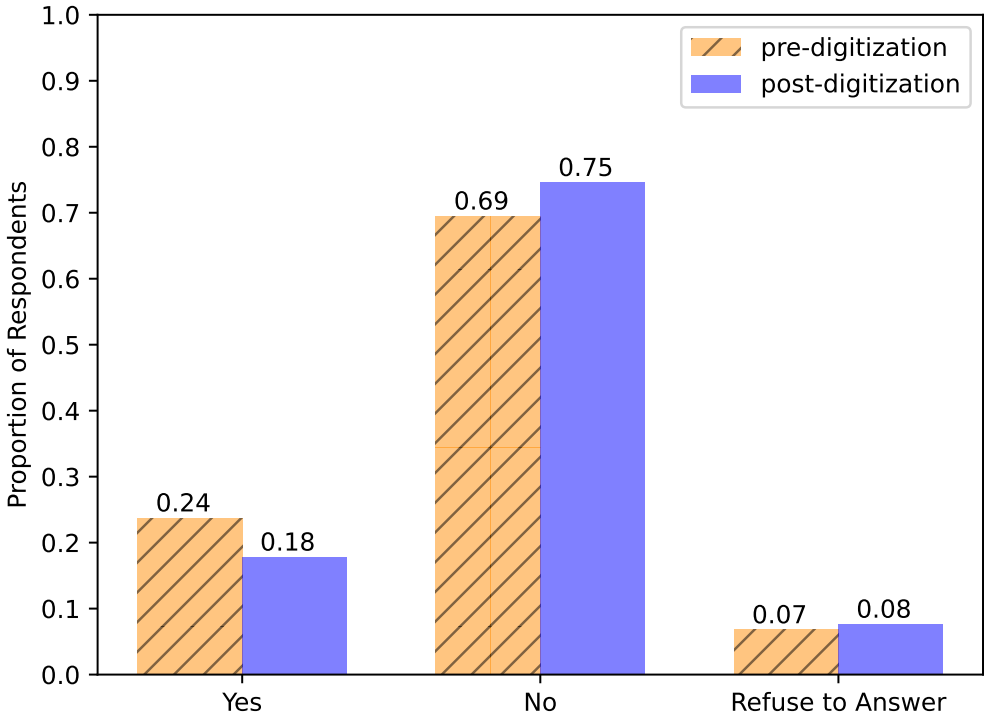
Notes. The figure is based on the bureaucrat survey. Data is restricted to the 118 bureaucrats that served as *Qanungo* between 2006-2013 and those that responded no to the question "Do you think digitization has improved overall tax collection?"

Figure B.12: Bureaucrats' social interactions with politicians and other bureaucrats before and after the digitization reform



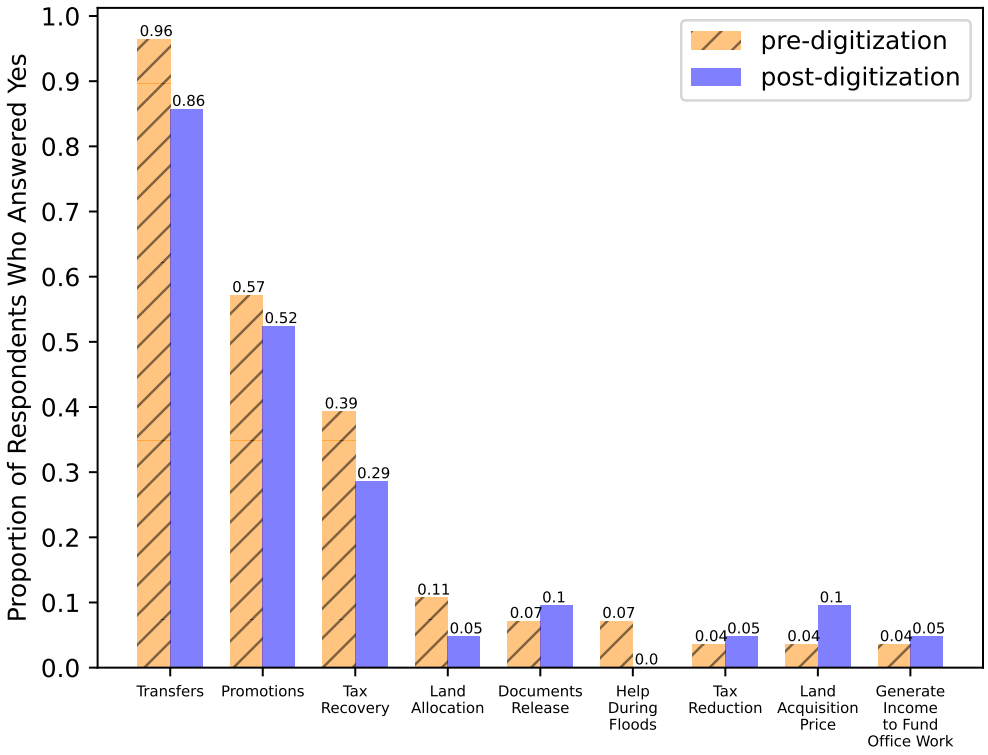
Notes. The figure is based on the bureaucrats survey. We used a Likert scale and asked them about the frequency of interactions with politicians and other bureaucrats before and after the reform. The Likert scale options were as follows: daily, twice a week, weekly, bi-monthly, monthly, quarterly, bi-annually, annually, less than once per year and never. For each bureaucrat the number of days of interactions in a year were calculated from these responses.

Figure B.13: Political interference in the work of bureaucrats



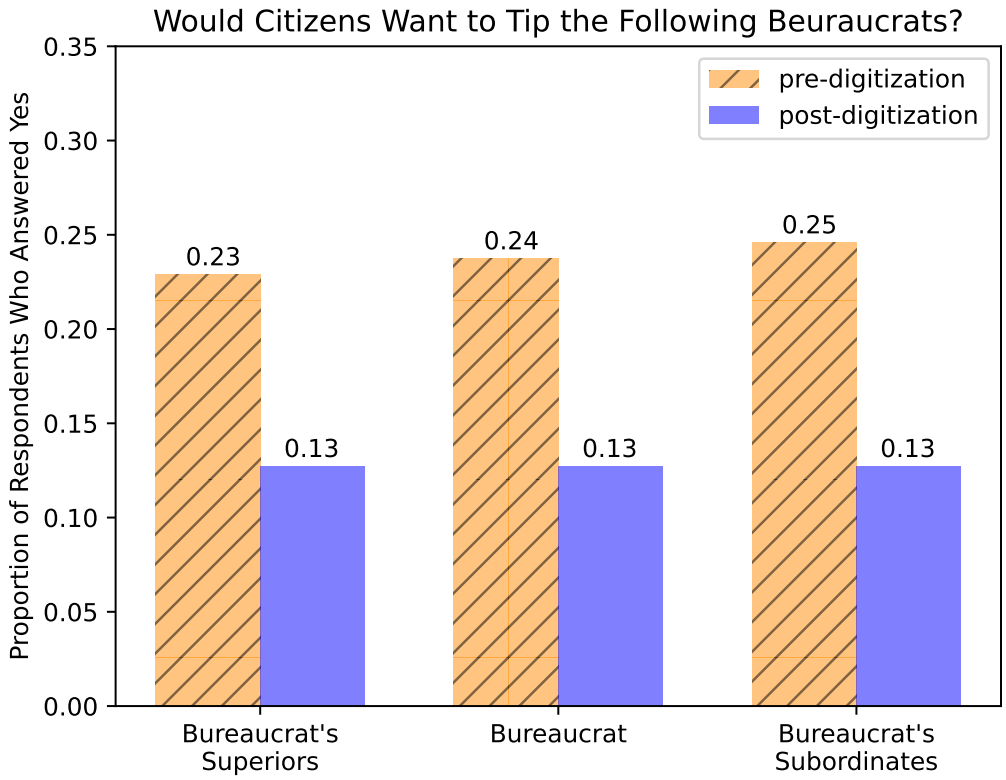
Notes. The figure is based on the bureaucrats survey. The responses are to the question: "In general, would you say that politicians interfere with the work of revenue officials in this revenue circle?"

Figure B.14: Matters in which politicians interfere



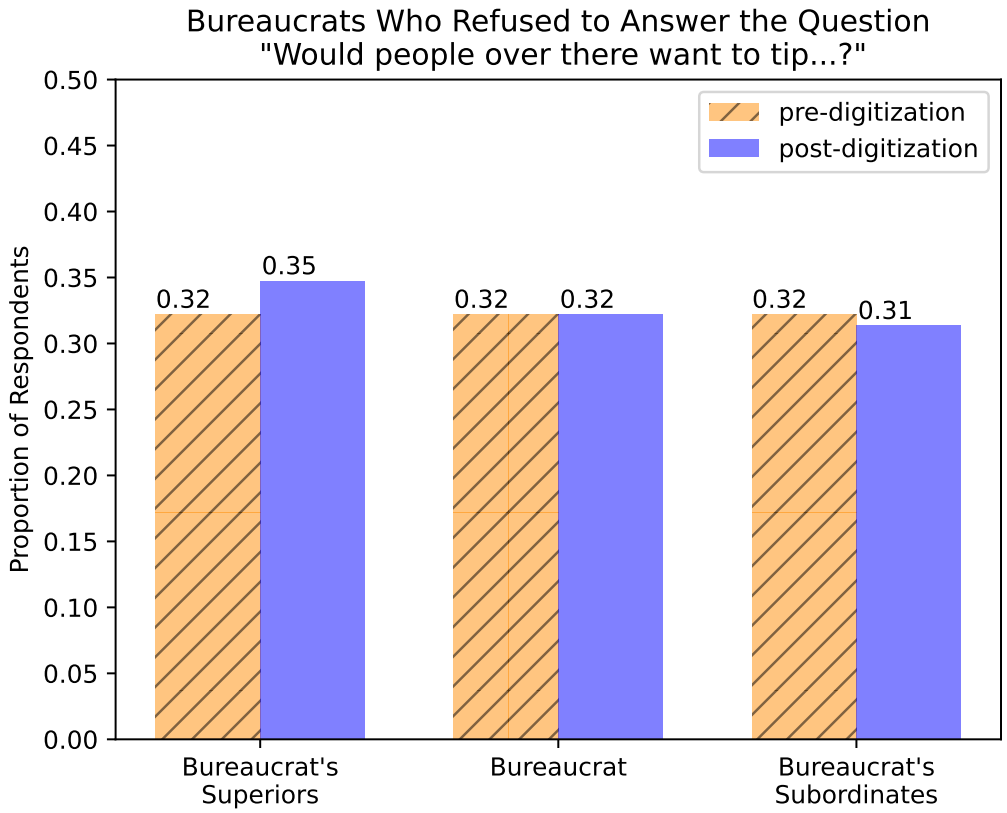
Notes. The figure is based on the bureaucrats survey. Those bureaucrats that responded yes to the question: "In general, would you say that politicians interfere with the work of revenue officials in this revenue circle?" were further asked "On which matters politicians usually interfere with work?" The matters listed above were suggested by the research team along with the category of “others”.

Figure B.15: Probability of bribes or "tips" to the bureaucrats



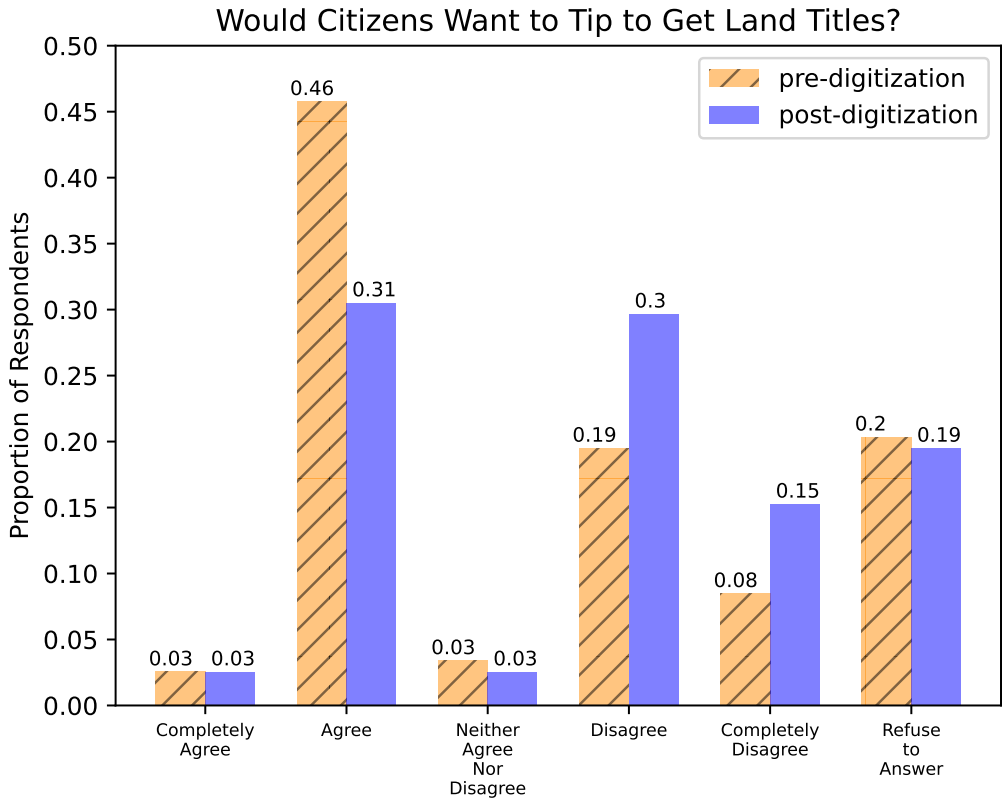
Notes. The figure is based on the bureaucrats survey. The figure shows the proportion of respondents that said yes to, "Would people over there (in a revenue circle) tip or want to tip a (the bureaucrat) in a normal month." Bureaucrat's superiors are "Tehsildars" and "Naib Tehsildars" and their subordinates are "Patwaris".

Figure B.16: Bureaucrats' non-response to the question on the probability of bribes or "tips"



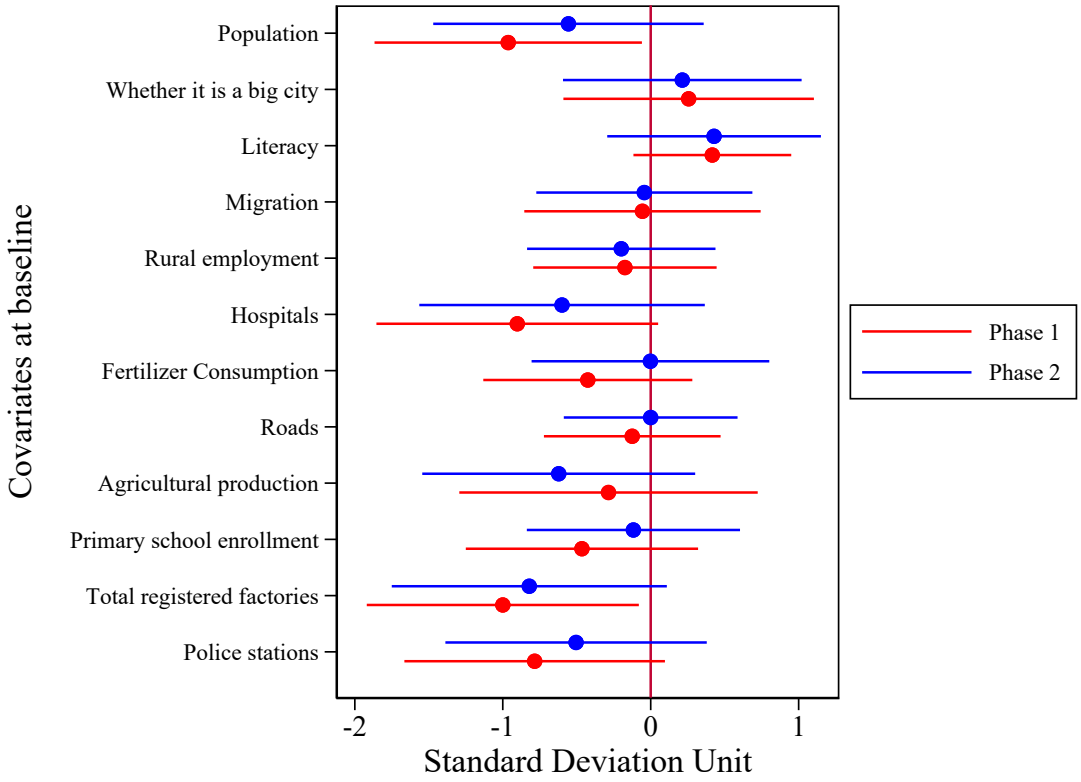
Notes. The figure is based on the bureaucrats survey. The figure shows the proportion of respondents that refused to answer, "Would people over there (in a revenue circle) tip or want to tip a (the bureaucrat) in a normal month." Bureaucrat's superiors are "Tehsildars" and "Naib Tehsildars" and their subordinates are "Patwaris".

Figure B.17: Probability of bribes or "tips" to the bureaucrat's subordinates for the issuance of land titles



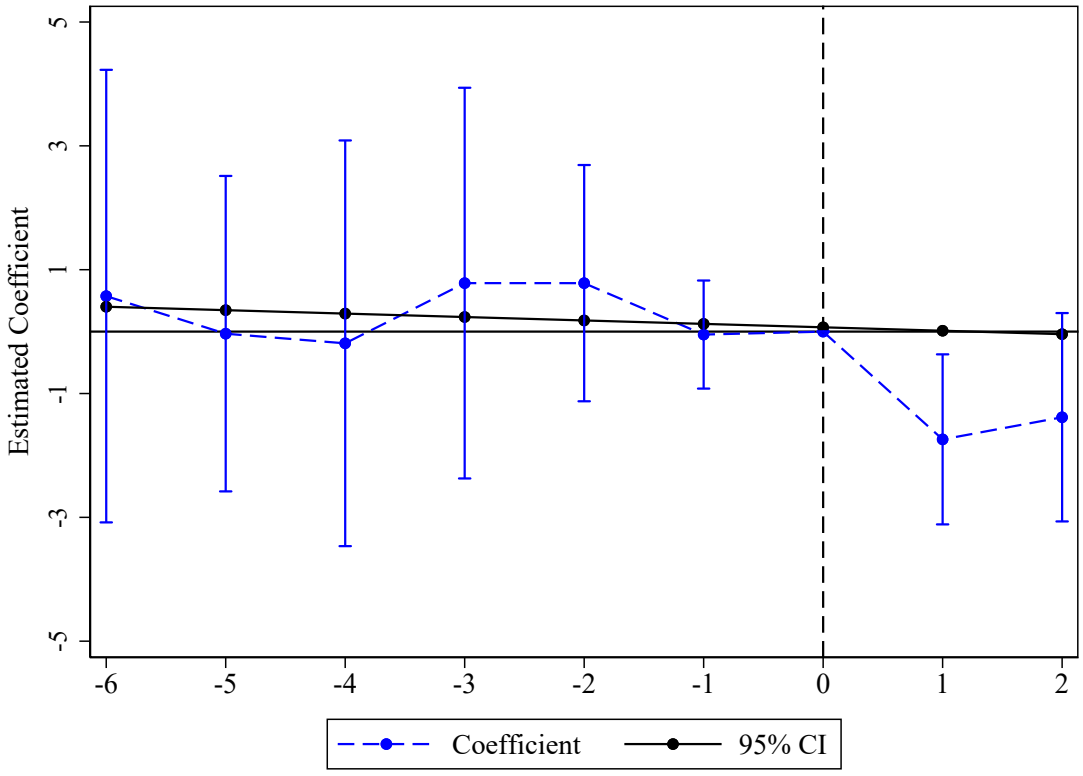
Notes. The figure is based on the bureaucrats survey. The figure shows the proportion of respondents that responded to, "People over there (in a revenue circle) would tip or want to tip a Patwari (bureaucrat's subordinates) for issuing Fard (land title)" measured on a Likert scale.

Figure B.18: Balance test of baseline characteristics of districts in phase 1 and 2 of the digitization reform



Note: Data on baseline characteristics is from the Development Statistics of the Pakistan Bureau of Statistics 1997-2010. The point estimates are from a regression of the respective covariates on a dummy that takes the value of one if the district lies in phase 1 or 2 of the digitization reform, and remains zero otherwise. The reference category are phase 3 districts. Intervals are 95% confidence intervals.

Figure B.19: Event study plot for the district level IHS tax collection with a trend line using pre-digitization data



Notes: Data is at the district-fiscal year level. Each coefficient is obtained from a set of indicator variables that take value one if, for phase 1 and 2 districts respectively in a fiscal year, the introduction of digitization of land records was k years away as described in Equation 1. The reference year is 2011 for phase 1 and 2012 for phase 2 districts. District and year fixed effects are included. Standard errors were clustered at the district level. The trend line is based on pre-digitization data (2006-2011). It connects predicted yearly outcomes generated from regressing the pre-treatment coefficients (from Equation 1) on fiscal years 2006-2011.

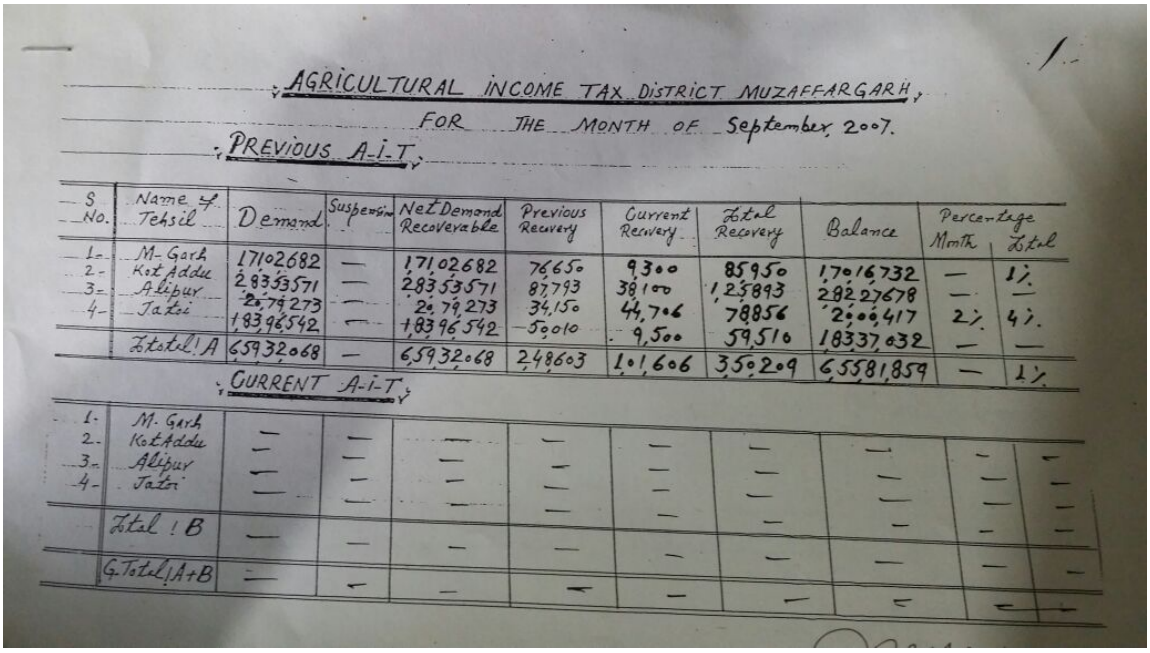
C Data Sources



Figure C.20: The BOR record room



Figure C.21: The BOR record room



AGRICULTURAL INCOME TAX DISTRICT MUZAFFARGARH

FOR THE MONTH OF September 2007.

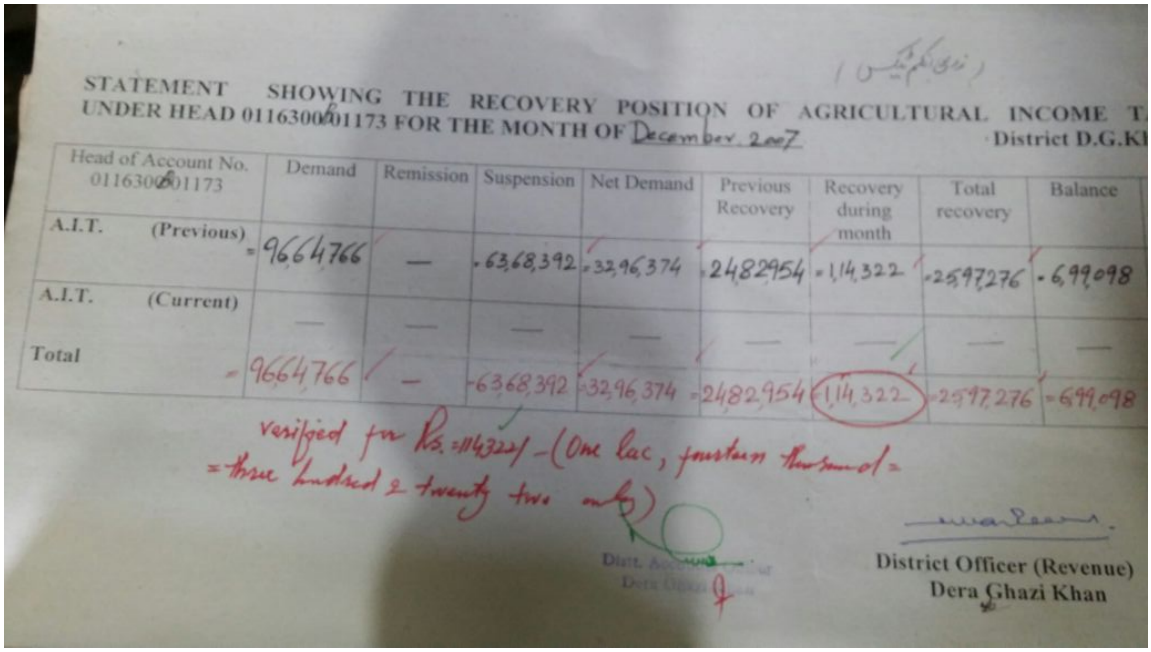
PREVIOUS A.I.T.

S No.	Name of Tehsil	Demand	Suspension	Net Demand Recoverable	Previous Recovery	Current Recovery	Total Recovery	Balance	Percentage Month	Percentage Total
1-	M. Garh	17102682	—	17102682	76650	9300	85950	17016732	—	1%
2-	Kot Addu	28353571	—	28353571	87793	38100	125893	28227678	—	—
3-	Alipur	2679273	—	2679273	34150	44706	78856	2600417	2%	4%
4-	Jatoi	18396542	—	18396542	50010	9500	59510	18337032	—	—
	Total A	65932068	—	65932068	248603	101606	350209	65581859	—	1%

CURRENT A.I.T.

1.	M. Garh	—	—	—	—	—	—	—	—	—
2.	Kot Addu	—	—	—	—	—	—	—	—	—
3.	Alipur	—	—	—	—	—	—	—	—	—
4-	Jatoi	—	—	—	—	—	—	—	—	—
	Total B	—	—	—	—	—	—	—	—	—
	G.Total (A+B)	—	—	—	—	—	—	—	—	—

Figure C.22: The BOR tax collection pro forma



STATEMENT SHOWING THE RECOVERY POSITION OF AGRICULTURAL INCOME T.
UNDER HEAD 01163001173 FOR THE MONTH OF December 2007
District D.G.K

Head of Account No. 01163001173	Demand	Remission	Suspension	Net Demand	Previous Recovery	Recovery during month	Total recovery	Balance
A.I.T. (Previous)	9664766	—	6368392	3296374	2482954	114322	2597276	699098
A.I.T. (Current)	—	—	—	—	—	—	—	—
Total	9664766	—	6368392	3296374	2482954	114322	2597276	699098

verified for Rs. 114322/- (One lac, fourteen thousand three hundred & twenty two only)

District Officer (Revenue)
Dera Ghazi Khan

Figure C.23: The BOR tax collection pro forma verified by District Accounts Officer

D Sampling for the bureaucrat survey

The retrospective survey was carried out in 2020, with the main aim to rebuild career trajectories of bureaucrats between 2006-2013 (the years when the tax data is available). Our sampling frame therefore, included people who were in-charge of revenue circles (*Qanungo*) as well as people who in the recent past had worked as a *Qanungo*. These included bureaucrats that had risen through the ranks via promotions and were in-charge of the tehsils: (*Tehsildars* and *Naib-tehsildars*).

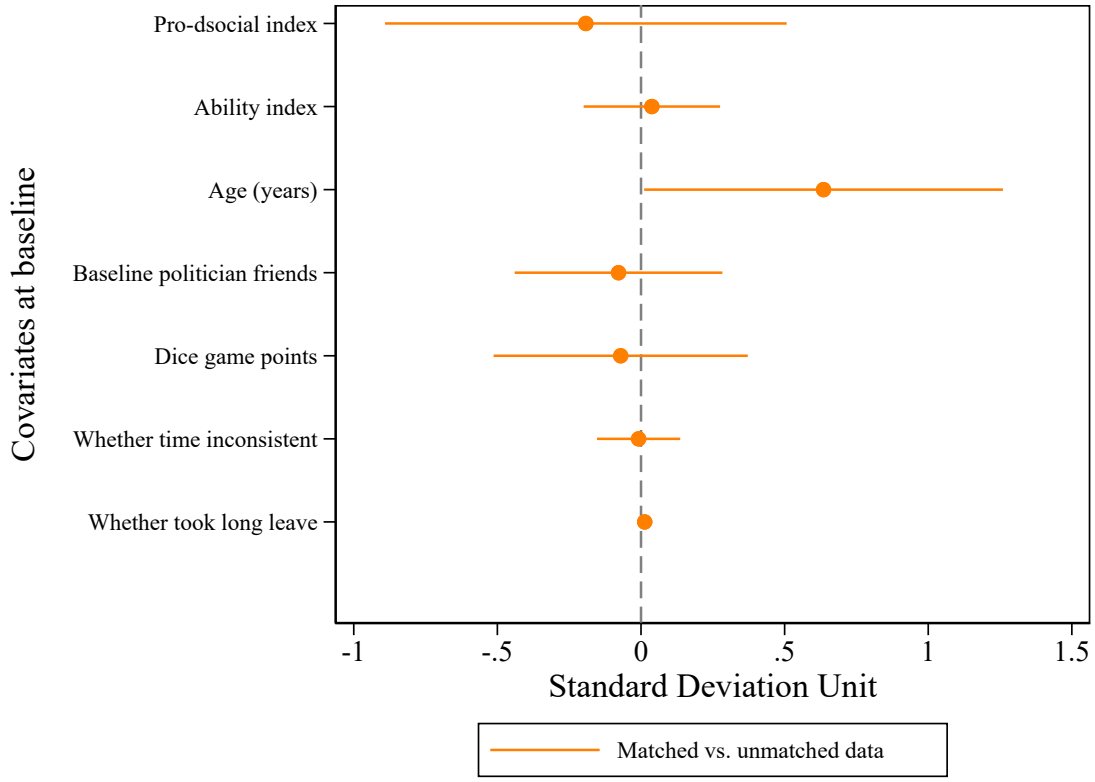
We stratified on districts and randomly sampled tehsils within each district.[34] We next created a sampling frame by contacting the local offices. Using that sampling frame we selected the universe of *Tehsildars* and *Naib-Tehsildars* working in the selected tehsils in Punjab. One *Qanungo* working with each of the *Naib-tehsildars* was randomly selected for the survey. We found 118 respondents who worked as Qanungos between 2006-2013.[35]

We could string-match the revenue circle name for 105 of those 118 respondents to match the survey data with the tax collection data. Of those the tax performance was missing for 27, so our final data set includes 78 respondents whose tax performance is observed between 2006-2013.

[34]Out of 141 tehsils in Punjab, we were able to survey bureaucrats from 138 tehsils. We were unable to survey the bureaucrats from the following three tehsils: Nishtar Town (Lahore districts), Shahkot (Nankana Sahib district), Ahmed Pur (Sheikhupura district).

[35]To find these, we started by surveying a total of 610 bureaucrats across different levels of hierarchy. Of those, 488 responded to the second round of telephonic survey about their career trajectory. The telephonic survey was used to recap the career paths of the bureaucrats, while their perceptions of digitization as well as their traits were measured in-person. In the pilot this was suggested by the field team as the best way to get maximum response rate, since the length of the two together were running to above 1 hour 30 min.

Figure D.24: Characteristics of the bureaucrats in the sample that were matched with the tax data



Note: Data on bureaucrats' characteristics is from the bureaucrat survey. The point estimates are from a regression of the respective covariates on a dummy that takes the value of one if the bureaucrats in the survey data were matched with the tax data, and remains zero otherwise. Pro-sociality index is created from five measures: Inclusion of Others in Self scale (Aron et al., 2004; Ashraf et al., 2020), whether they have donated blood, money donated in public good game, whether they do volunteer work and whether they give charity. Ability index is created from four measures: an incentivized matrix game and a memory game as in Hanna and Wang (2017), response to questions on general knowledge and revenue rules and regulations, respectively. Politician friends are the number of friends of the bureaucrats that are either federal or provincial politicians. Dice game points is a proxy for dishonesty and it is the total in an incentivized dice game as in Hanna and Wang (2017). Intervals are 95% confidence intervals.

E Details on string matching revenue circles

We carried out an extensive string matching exercise to merge the tax, digitization and bureaucrats' careers datasets. We took the following steps to merge the three sets of data:

- As a first step we manually checked each revenue circle, tehsil and district in the tax data against their counterparts recorded in the digitization data from the Punjab Land Records Authority (PLRA) and allocated a unique ID to each.[36] There were 1125 revenue circles in total, out of which 838 were given IDs using this process.
- The district, tehsil and revenue circle names in the bureaucrat survey data was manually cleaned. There were 690 unique revenue circles-tehsil-districts in this data, out of these we were able to give IDs to 458.
- Finally, we merged all the three data on revenue circle, tehsil and district names.

We next checked the veracity of these data using further records from the government on details of revenue circle, tehsil and district names across Punjab. These were personally obtained from the government in 2020.

[36] Digitization data from PLRA contained details of names of most of the tehsils and revenue circles except the following 19 (out of 141 in total) tehsils: Gujranwala Sadar, Kabirwala, Kharian, Shorkot, Khushab, Quaidabad, Jauharabad Lahore city, Nishtar Town, Muzaffargarh, Depalpur, Renala khurd, Arifwala, Khanpur, Murree, Rawalpindi city, Rawalpindi Sadar, Rawalpindi Cantt and Daska.

F Satellite vegetation cover data

We used NASA's MODIS land products to observe a satellite based vegetation cover index. MODIS vegetation indices provide consistent spatial and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, chlorophyll and canopy structure. The normalized difference vegetation index (NDVI) are derived from atmospherically-corrected reflectance in the red, near-infrared, and blue wavebands. NDVI ranges from -1 to +1. If the NDVI values are negative it is highly likely that it's water. On the other hand, values close to +1 suggest that there is a high possibility that there are dense green leaves.[37]

NASA's MODIS land products rely on the Sinusoidal Tile Grid System, which divides earth into 36 x 18 sinusoidal grids to locate a particular area on earth.[38]. Since we are only interested in the Punjab province of Pakistan, our first step was to locate the tiles where Punjab is located. Given the shapefile of districts of Punjab, we uniformly sample 10,000 points within each district and calculated their locations on the sinusoidal tile.[39] We find that all 36 $\times$ 10,000 points lie within three tiles, namely (horizontal 24, vertical 5), (horizontal 24, vertical 6), and (horizontal 23, vertical 5). There are a total of 4800 $\times$ 4800 pixels within each of the three tiles mentioned above, with each pixel having a 250 m $\times$ 250 m size. Moreover, data for each year is divided into time intervals of 16 days. This results in 23 different time intervals in a given year. For each 16-day time interval in a year, and for each of the 36 districts of Punjab, we obtained the NDVI values of all pixels belonging to that particular district and take the average to get the NDVI value for that particular district in that particular 16-day interval. Since each year has 23, 16-day intervals, we end up having a list of 23 different NDVI values for a particular district in a particular year. Following the method used in Beg (2020), we use the maximum value of that list as the NDVI value for that district for that year.

[37] Details accessed at https://gisgeography.com/ndvi-normalized-difference-vegetation-index/

[38] Accessed at: https://modis-land.gsfc.nasa.gov/MODLAND_grid.html

[39] We decided to rely on this method rather than just consider the district's center and calculate their locations on the sinusoidal tile. This allows us a more holistic view of the vegetation cover of the district.